

# The Second Data Release of the INT Photometric H $\alpha$ Survey of the Northern Galactic Plane (IPHAS DR2)

Geert Barentsen<sup>1\*</sup>, H. J. Farnhill<sup>1</sup>, J. E. Drew<sup>1</sup>, E. A. González-Solares<sup>2</sup>, R. Greimel<sup>3</sup>, M. J. Irwin<sup>2</sup>, B. Mizalski<sup>99</sup>, C. Ruhland<sup>1</sup>, P. Groot<sup>4</sup>, A. Mampaso<sup>5</sup>, S. E. Sale<sup>6</sup>, plus many contributors.

<sup>1</sup>School of Physics, Astronomy & Mathematics, University of Hertfordshire, College Lane, Hatfield, Hertfordshire, AL10 9AB, U.K.

<sup>2</sup>Institute of Astronomy, Cambridge University, Madingley Road, Cambridge, CB3 OHA, U.K.

<sup>3</sup>IGAM, Institute of Physics, University of Graz, Universitätsplatz 5, Graz, Austria.

<sup>4</sup>Afdeling Sterrenkunde, Radboud Universiteit Nijmegen, Faculteit NWI, Postbus 9010, 6500 GL Nijmegen, The Netherlands.

<sup>5</sup>Instituto de Astrofísica de Canarias, 38200 La Laguna, Tenerife, Spain.

<sup>6</sup>Rudolf Peierls Centre for Theoretical Physics, Keble Road, Oxford, OX1 3NP, U.K.

<sup>99</sup>Please add your current affiliation.

Current draft typeset 13 March 2014

## ABSTRACT

The INT/WFC Photometric H $\alpha$  Survey of the Northern Galactic Plane (IPHAS) is a 1800 square degrees imaging survey covering the entire northern Milky Way at  $|b| < 5^\circ$  in the  $r$ ,  $i$  and H $\alpha$  filters using the Wide Field Camera (WFC) on the 2.5-meter Isaac Newton Telescope (INT). This data release presents the first uniformly-calibrated source catalogue to have been extracted from the images, providing single-epoch photometry for 219 million unique sources across 92% of the survey. The observations were carried out between 2003 and 2012 at a median seeing of 1.1 arcsec to a depth of  $r = 21.2 \pm 0.5$ ,  $i = 20.0 \pm 0.3$  and H $\alpha = 20.3 \pm 0.3$  ( $5\sigma$  limits, Vega system). We explain the data reduction and quality control procedures, describe and test the new uniform photometric calibration, and detail the construction of the source catalogue and its quality warning flags. We find that the new calibration and the photometric repeatability are accurate to  $\sigma = 0.03$  mag. Finally, we demonstrate the ability of the catalogue’s unique  $r$ - $i$ / $r$ -H $\alpha$  colour-colour diagram to (i) characterise stellar populations and extinction regimes towards different Galactic sightlines and (ii) select candidate H $\alpha$ -emission line objects. The catalogue which accompanies this paper provides the much-needed visible-light counterpart to several infrared surveys of the Galactic Plane, including many poorly-studied star-forming regions and “bubbles”. It also provides images for some of the most crowded regions to be faced by Gaia.

**Key words:** catalogues, surveys, stars: emission line, Be, Galaxy: stellar content

## 1 INTRODUCTION

Since the first data release in 2008, the INT/WFC Photometric H $\alpha$  Survey of the Northern Galactic Plane (IPHAS) has provided new insights into the contents and structure of our own backyard, the Milky Way. The original motivation for undertaking this large-scale programme of observation – spanning almost a decade and using more than 300 nights at the Isaac Newton Telescope (INT) in La Palma – was to provide the digital update to the photographic northern H $\alpha$

surveys of the mid-20th century. By increasing the sensitivity with respect to these previous surveys by a factor  $\sim 1000$  (7 magnitudes), it was envisaged that IPHAS would allow the limited bright samples of Galactic emission line objects available at the outset (e.g. Kohoutek & Wehmeyer 1999), to be extended into larger, deeper, more statistically-robust samples that in turn could better inform our understanding of the early and late stages of stellar evolution. This aim has begun to be realised through a range of published studies including: a preliminary catalogue of candidate emission line objects (Witham et al. 2008); discoveries of new northern symbiotic stars (Corradi et al. 2008, 2010); new cataclysmic

\* E-mail:geert@barentsen.be

variables (Witham et al. 2007); new groups of young stellar objects (Vink et al. 2008; Barentsen et al. 2011); along with discoveries of new and remarkable planetary nebulae (Mampaso et al. 2006; Corradi et al. 2011; Viironen et al. 2011).

Over the years it has become apparent that the legacy of IPHAS will extend beyond the traditional H $\alpha$  applications of identifying emission line stars and nebulae. Through the provision of  $r$ ,  $i$  broadband photometry alongside H $\alpha$  data, IPHAS has created the opportunity to study Galactic Plane populations in a new way. For example, the surveys unique  $r - H\alpha$  colour, when combined with  $r - i$ , has been shown to provide simultaneous constraints on intrinsic stellar colour and interstellar extinction (Drew et al. 2008). This has opened the door to a wide range of Galactic science applications, including the mapping of extinction across the Plane in three dimensions and the probabilistic inference of stellar properties (Sale et al. 2009, 2010; Giannanco et al. 2011; Sale 2012; Barentsen et al. 2013). In effect, the availability of narrowband H $\alpha$  alongside  $r$ ,  $i$  magnitudes provides coarse spectral information for huge samples of stars which are otherwise too faint or numerous to be targeted by spectroscopic surveys. For such science applications to succeed however, it is vital that the imaging data is transformed into a homogeneously calibrated photometric catalogue, in which quality problems and duplicate detections of the same source are flagged.

The first release of IPHAS data, covering roughly half the survey footprint, was made in late 2007 (González-Solares et al. 2008). At the time the data were insufficiently complete to support a homogeneously calibrated source catalogue. The goal of this paper is to present the next release that takes the coverage up to over 90 percent of the survey area and includes a uniform calibration. In this work we aim to (i) explain the data reduction and quality control procedures that were applied, (ii) describe and test the new global photometric calibration, and (iii) detail the construction and demonstrate the use of the source catalogue that has been extracted from the newly re-calibrated data.

In §2 we start by recapitulating the key points of the survey design and strategy. In §3 we describe the data reduction and quality control procedures. In §4 we explain the uniform re-calibration and test our results against the Sloan Digital Sky Survey (SDSS). In §5 we explain how the source catalogue was compiled. In §6 we discuss the properties of the catalogue, and finally in §7 we demonstrate the health of the release by demonstrating its scientific exploitation. In §8 we conclude and outline our future ambitions.

## 2 OBSERVATIONS

The detailed properties of the IPHAS observing programme have been presented before by Drew et al. (2005) and González-Solares et al. (2008). To set the stage for this release, we briefly remind of some key points. IPHAS is a 1800 sq. deg. imaging survey of the northern Galactic Plane, providing images and photometry in Sloan  $r$ ,  $i$  along with narrowband H $\alpha$ . It is carried out using the Wide Field Camera (WFC) on the 2.5-meter Isaac Newton Telescope (INT) in La Palma. It is the first digital survey to offer comprehensive CCD photometry of point sources in the Galactic

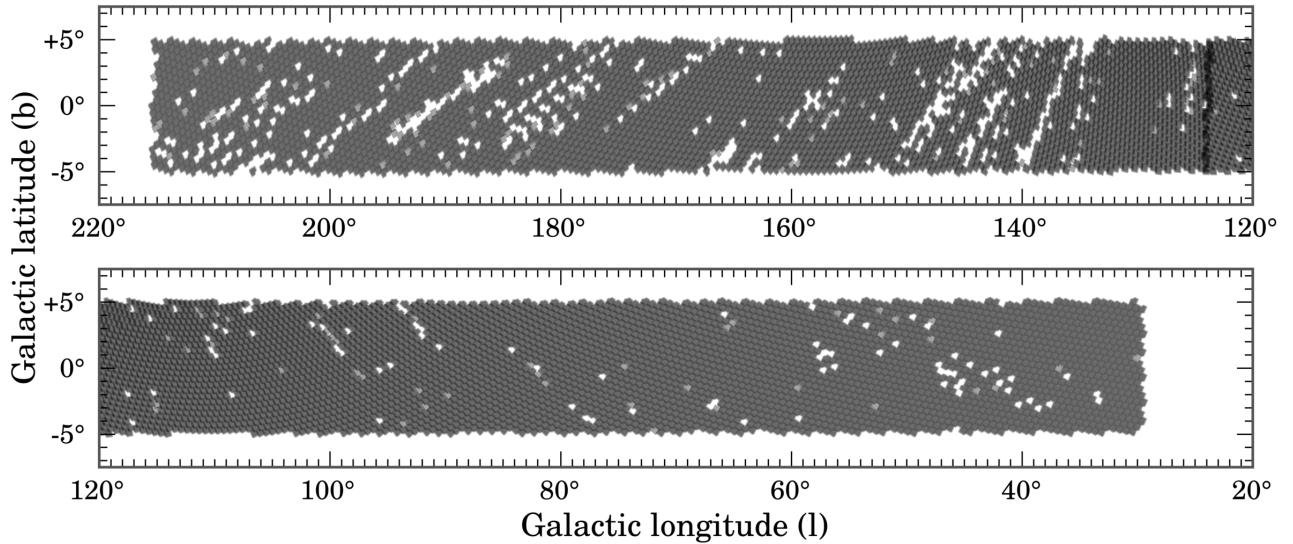
Plane at visible wavelengths, and does so down to a limiting magnitude of  $\sim$ 20th. The IPHAS footprint on the northern sky spans a box of roughly 180 by 10 degrees, taking in the entire northern Galactic Plane at latitudes  $-5^\circ < b < +5^\circ$  and longitudes  $30^\circ < l < 215^\circ$ .

The Wide Field Camera is a mosaic of 4 CCDs that captures a sky area of close to 0.29 square degrees. To cover the entire northern Plane with some overlap, the survey area was divided into 7635 telescope pointings. Each of these pointings is accompanied by an offset position at a displacement of +5 arcmin in declination and +5 arcmin in right ascension, to deal with inter-CCD gaps, detector imperfections, and to enable quality checks. The basic unit of observation hence amounts to  $2 \times 3$  exposures, in which each of the 3 survey filters is exposed at 2 offset sky positions, within an elapsed time of 10 minutes. We shall refer to the unit of 3 exposures at the same position as a *field*, and the combination of two fields at a small offset as a *field pair*. The survey hence contains 15270 fields grouped into 7635 field pairs. To achieve the desired survey depth of 20th magnitude or fainter, the filter exposure times were set at 120 sec (narrowband H $\alpha$ ), 30 sec ( $r$ ) and 10 sec ( $i$ ) in the majority of the survey observations.<sup>1</sup>

Data-taking began in the second half of 2003, and every field had been observed at least once by the end of 2008. At that time only 76 percent of the field pairs satisfied our minimum quality criteria however, often due to the effects of clouds, poor seeing, or technical faults (the quality criteria will be detailed in the next section). Since then, a programme of repeat observations has been in place to improve data quality. As a result, 92% of the survey now benefits from quality-approved data. The most recent observations which are part of this release were obtained in November 2012.

Figure 1 shows the footprint of the quality-approved observations included in this work. The fields which remain missing – covering 7 percent of the survey area – are predominantly located towards the Galactic anti-center at  $l > 120^\circ$ . Fields at these longitudes can only be accessed from La Palma in the months of November–December, which is when the weather and seeing conditions are often poor at the INT and observing attempts have failed repeatedly. To enable the survey to be brought to completion, a decision was made recently to limit repeats in this area to individual fields requiring replacement, i.e. fresh observations in all 3 filters may only be obtained at one of the two offset positions, if the data for the partner offset has already passed quality control. The catalogue is structured such that it is clear where a contemporaneous observation of both halves of a field pair is not available.

<sup>1</sup> In 2003 the  $r$ -band exposure time was 10 sec instead of 30 sec, and since Oct 2010 the  $i$ -band exposure time has been increased from 10 sec to 20 sec.



**Figure 1.** Survey area showing the footprints of all the quality-approved IPHAS fields which have been included in this data release. The area covered by each field has been coloured black with a semi-transparent opacity of 20%, such that regions where fields overlap are darker. The IPHAS strategy is to observe each field twice with a small offset, and hence the vast majority of the area is covered twice (dominant gray colour). There are small overlaps between all the neighbouring fields which can be seen as a honeycomb-style pattern of dark gray lines across the survey area. Regions with incomplete data are apparent as white gaps (no data) or as the lightest shade of gray (denoting that only the offset position is missing).

### 3 DATA REDUCTION AND QUALITY CONTROL

#### 3.1 Initial pipeline processing

All raw data obtained with the INT were transferred to the Cambridge Astronomical Survey Unit (CASU) for initial processing and archival. The procedures used by CASU were originally devised for the INT Wide Field imaging Survey (WFS; McMahon et al. 2001; Irwin et al. 2005), which was a  $200 \text{ deg}^2$  survey programme carried out between 1998 and 2003 after the WFC was commissioned. Because IPHAS uses the same telescope and camera combination, we have been able to benefit from the existing WFS pipeline. A detailed description of the processing steps is found in Irwin & Lewis (2001). Its application to IPHAS has previously been described by Drew et al. (2005) and González-Solares et al. (2008) and much of the source code is available on line<sup>2</sup>. In brief, the pipeline takes care of bias subtraction, linearity correction, flat-fielding, gain correction and de-fringing.

The reduced images are then stored in a multi-extension FITS file with a primary header describing the characteristics (position, filter, exposure time, etc.) and four compressed image extensions corresponding to each of the four CCDs. Source detection and characterisation is then carried out using the IMCORE tool (Irwin 1985, 1997). The flux of each source is measured using both the peak pixel height (i.e. a square  $0.33 \times 0.33''$  aperture) as well as a series of circular apertures of increasing diameter ( $1.2'', 2.3'', 3.3'', 4.6''$  and  $6.6''$ ).

The local background levels are estimated by computing the sigma-clipped median flux in a grid of  $64 \times 64$  pixels

( $21 \times 21''$ ) across the image, which is then interpolated to obtain an estimate of the background level at each pixel. These sky levels are subtracted from the aperture photometry and – when required – a deblending routine is applied which also attempts to remove the contamination from any other nearby sources. Whilst this approach works very well across the vast majority of the survey area, the Galactic Plane unavoidably contains crowded regions with large numbers of overlapping sources or rapidly spatially-varying nebulosity, in which case aperture photometry must always be interpreted with caution. In §5 we will explain that overlapping sources to which the deblending routine was applied are flagged in the catalogue using the *deblend* warning flag.

Finally, an astrometric solution is determined based on the 2MASS point source catalog (Skrutskie et al. 2006), which itself is calibrated in the International Celestial Reference System (ICRS). A provisional photometric calibration is also provided based on the average zeropoint determined from a set of standard stars observed in the same night. Sources are classified morphologically – stellar, galaxy or noise – based on the curve-of-growth determined from measuring the source intensity in a series of growing apertures. Finally, the resulting source detection tables are also stored in multi-extension FITS files.

At the time of preparing DR2, the CASU pipeline had processed 74 195 IPHAS exposures in which a total of 1.9 billion *candidate sources* were detected at the sensitive default detection level of  $1.25\sigma$  – unavoidably including spurious detections, artefacts and duplicate detections (in §5 we will explain how these have been removed or flagged). The pipelined data set – comprising 2.5 terabyte of FITS files – was then transferred to the University of Hertfordshire for the purpose of transforming the raw detection tables into a reliable source catalogue which is (i) quality-controlled, (ii)

<sup>2</sup> <http://casu.ast.cam.ac.uk/surveys-projects/software-release>

homogeneously calibrated, and (iii) contains user-friendly columns and warning flags. It is these post-processing steps which are explained below.

### 3.2 Quality control

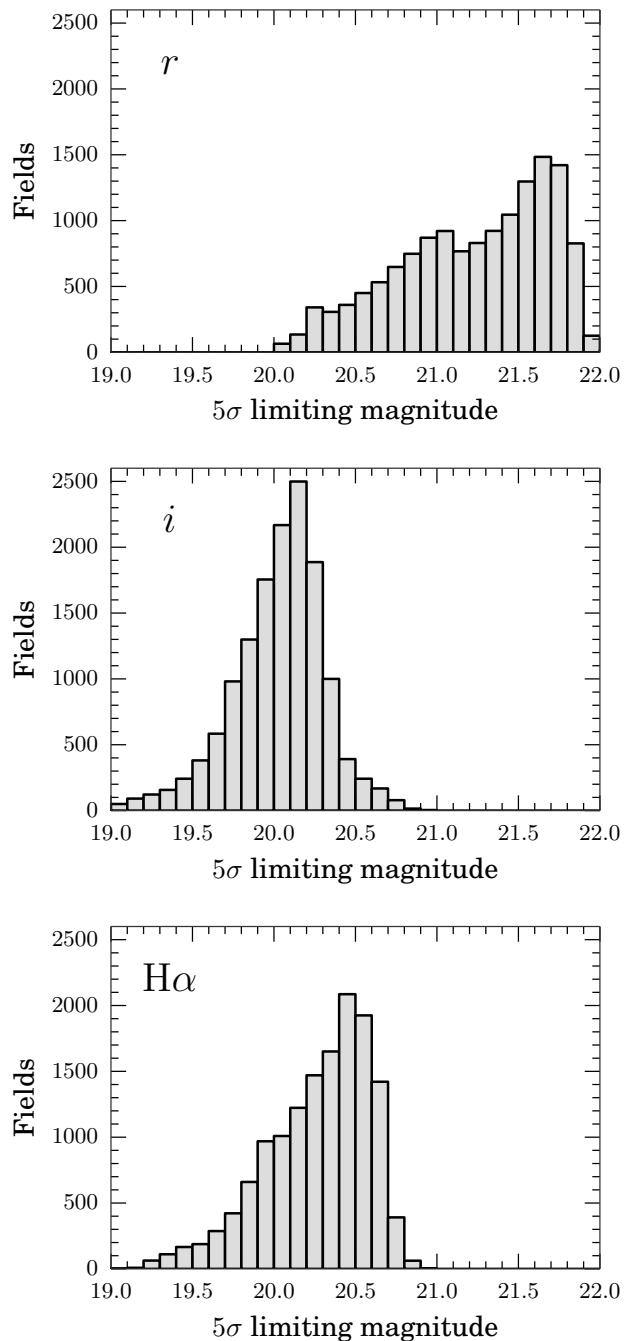
Observing time for IPHAS was obtained on a semester-by-semester basis through the traditional time allocation committees of the Isaac Newton Group of telescopes, which are competitive and invariably over-subscribed. For this reason, we attempted to utilise *all* the nights allocated to IPHAS, even those which were partially or entirely non-photometric or otherwise affected by technical problems (e.g. electronic noise or telescope tracking problems). Any unsuitable data that was taken as a result of this strategy was subsequently flagged and rejected using a series of seven quality criteria, which ensure a reliable and homogeneous level of quality across the data release:

(1) *Depth*. We discarded any exposures for which the  $5\sigma$  limiting magnitude was worse than 20th magnitude in the *r*-band or worse than 19th in *i* or H $\alpha$ . Such data were typically obtained during poor weather or full moon. Most observations fared significantly better than these limits. Figure 2 presents the distribution of limiting magnitudes for all quality-approved fields, which shows a mean depth of  $21.2 \pm 0.5$  (*r*),  $20.0 \pm 0.3$  (*i*) and  $20.3 \pm 0.3$  (H $\alpha$ ). We found that the depth achieved depended most strongly on the presence of the moon, which was above the horizon during 62 percent of our observations and explains the wide and bi-modal shape of the *r*-band limiting magnitude distribution (top panel in Fig. 2). In contrast, the depth attained in *i* and H $\alpha$  is less sensitive to moonlight and the distribution of their depths is hence more narrow (middle and bottom panel in Fig. 2).

Our minimum limiting magnitude criteria have led us to exclude 9% of the pipelined data. We note that much of the excluded data may nevertheless be useful for e.g. time-domain studies of bright stars. The detection tables for any such discarded data are made available through our website ([www.iphas.org](http://www.iphas.org)) but are ignored in what follows.

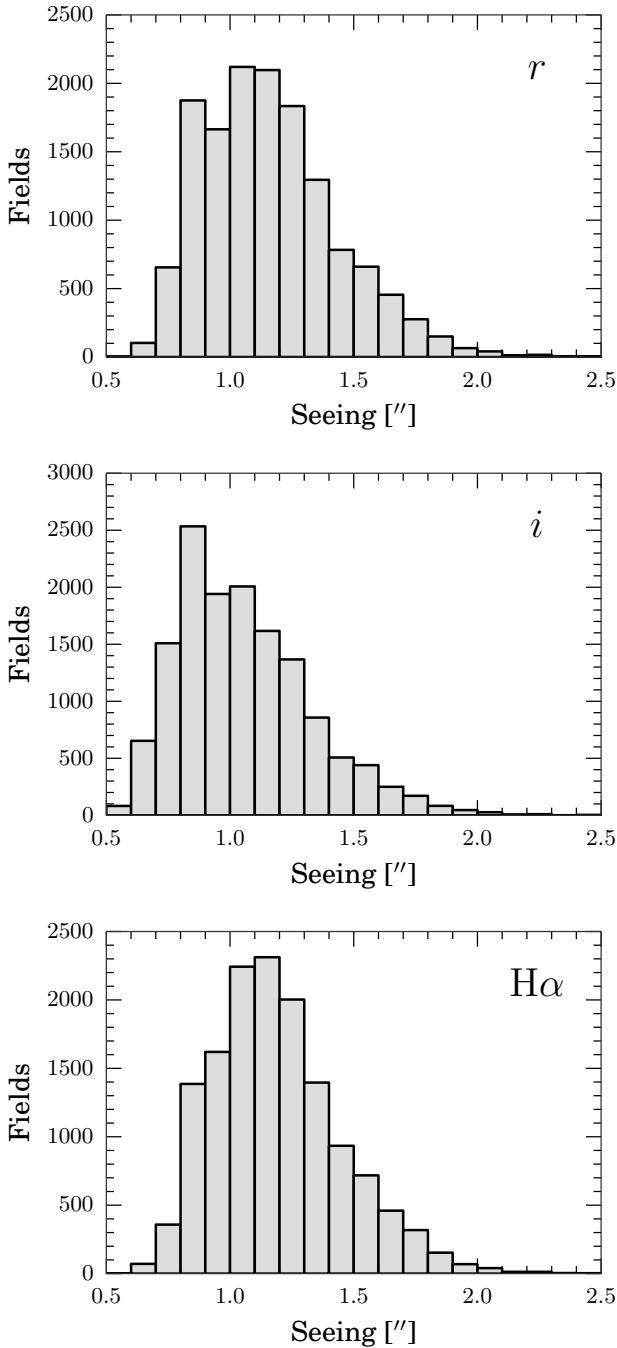
(2) *Ellipticity*. The ellipticity of a point source, defined as  $e = 1 - b/a$  with  $b$  the semi-minor and  $a$  the semi-major axis, is a morphological measure of the elongation of the point spread function. It is expected to be zero (circular) across the field in a perfect imaging system, but it is slightly non-zero in any real telescope data due to optical distortions and tracking errors. The mean ellipticity across a field in the IPHAS data set is  $0.09 \pm 0.04$ . There have been sporadic episodes with higher ellipticities however due to mechanical glitches in the telescope tracking system. For this reason, we rejected exposures in which the mean ellipticity across the detector exceeded  $e > 0.3$ , which is when the photometric measurements delivered by the pipeline were found to become degraded. Only 0.4% of the exposures were discarded on this basis.

(3) *Seeing*. The survey originally aimed to obtain data at seeing better than 1.7 arcsec. This target is currently attained across 86% of the footprint, in particular at early longitudes, e.g. 92% of the fields at  $l < 120^\circ$  are better than 1.7 arcsec. Figure 3 presents the distribution of the mean seeing for all the quality-approved fields. We find a median value of 1.1 arcsec in *r*/H $\alpha$  and 1.0 arcsec in *i*. In the *r*-band,



**Figure 2.** Distribution of the  $5\sigma$  limiting magnitude across all quality-approved survey fields for *r* (top), *i* (middle) and H $\alpha$  (bottom). Fields with a limiting magnitude brighter than 20th (*r*) or 19th (H $\alpha$ /*i*) were rejected from the data release. The *r*-band depth is most sensitive to the presence of the moon above the horizon, which is evidenced by the wide and bi-modal shape of its distribution.

90% of the data is better than 1.5 arcsec and 10% is better than 0.8 arcsec. To improve the area covered by this data release, we have decided to include the small fraction of data that was obtained under seeing up to 2.5 arcsec, so that only 1 per cent of the pipelined exposures had to be excluded. In §5 we will explain that the catalogue is structured such



**Figure 3.** Seeing distribution across all quality-approved survey fields for  $r$  (top),  $i$  (middle) and  $H\alpha$  (bottom).

that the information on the seeing is included, and that the photometry listed is based on the exposures with the best-available seeing.

(4) *Photometric repeatability.* The IPHAS field-pair observing strategy ensures that every pointing is immediately followed by an offset pointing at a displacement of +5 arcmin in Dec and +5 arcmin in RA. This allows pairs of images to be checked for the presence of clouds or electronic noise. To exploit this information, the overlap regions of all field pairs were systematically cross-matched to verify the consistency of the photometric measurements for stars observed in both

pointings. We rejected field pairs in which more than 2% of the stars showed an inconsistent measurement at the level of 0.2 mag, or more than 26% were inconsistent at the level of 0.1 mag. These limits were determined empirically by inspecting the images and photometry by eye. 11% of the data was rejected as part of this step.

(5) *Source density mapping.* Spatial maps showing the number density of the detected sources down to 20th magnitude were created to verify the health of the data and to check for unexpected artefacts. In particular, we created density maps which showed the number of *unique* sources obtained by cross-matching the detection tables of all three bands with a maximum matching distance of 1 arcsec. The success of such a cross-matching procedure crucially depends on the accuracy of the astrometric solution in each band, and hence we were able to detect and correct exposures with poor astrometry by inspecting the density map for spurious density variations.

(6) *Visual examination.* All images and their associated photometric colour/magnitude diagrams were inspected by a team of 20 survey team members, such that each image in the data release was looked at by at least three different pairs of eyes. Images deemed unsuitable were flagged, investigated and excluded from the release if necessary. 6% of the attempts to observe a field were placed on a *black-list* for this reason, most commonly due to the obvious presence of clouds or extreme levels of scattered moonlight. Such fields were often rejected as part of one or more of the above quality criteria as well.

(7) *Contemporaneous field data.* Finally, only exposures which are part of a sequence of three consecutive images of the same field ( $H\alpha/r/i$ ) were considered for inclusion in the release. This ensures that the three bands for a given field are observed at nearly the same time – essentially always within 5 minutes. An exception was made for 9 fields where the three exposures could not be obtained within the same night but for which the time gap did not exceed 48 hours. We note that the exact epoch of the magnitude in each band is included in the source catalogue (columns  $rMJD$ ,  $iMJD$ ,  $haMJD$ ).

The above criteria were satisfied for 14115 out of the 15270 fields (92%). In a few cases more than one successful attempt to observe a field was available due to the fact that slightly stricter quality criteria were adopted in the initial years of the survey. In such cases, only the attempt with the best seeing and depth was selected for inclusion in the release, because the focus of this work is to deliver the most reliable measurement at a single epoch. Those interested in any of the rejected data may nevertheless access the full set of detection tables on line.

## 4 PHOTOMETRIC CALIBRATION

Having obtained a quality-approved set of observations, we now turn to the problem of placing the data onto a uniform photometric scale.

### 4.1 Provisional nightly calibration

For the purpose of providing an initial calibration of the  $r$  and  $i$  broadband fluxes, photometric standard fields

were observed every night. The standards were chosen from a list based on the Landolt (1992) and Stetson (<http://cadcwww.dao.nrc.ca/standards>) objects. Two or three standard fields were observed during the evening and morning twilight, and at intervals of 2-3 hours throughout the night. The CASU pipeline automatically identified the observed standards and used them to determine a sigma-clipped average zeropoint MAGZPT for each night and filter, such that the number counts  $DN$  in the pipeline-corrected CCD frames relates to a magnitude  $m$  as:

$$m = \text{MAGZPT} - 2.5 \log_{10}(DN/\text{EXPTIME}) - \text{EXTINCT} \cdot (\text{AIRMASS} - 1) - \text{APCOR} - \text{PERCORR} \quad (1)$$

where EXPTIME is the exposure time in seconds, EXTINCT is the atmospheric extinction coefficient (typically 0.09 for  $r$  and 0.05 for  $i$  in La Palma), AIRMASS is the normalised optical path length through the atmosphere and APCOR is a correction for the flux lost outside of the aperture used. Finally, PERCORR is a correction based on the difference between the median dark sky for a CCD against the median for all the CCDs and as such is an ancillary correction to account for sporadic gain variations. All these quantities correspond to header keywords in the multi-extension FITS files produced by the CASU pipeline.

The zeropoint was determined such that the resulting magnitude system is in the WFC system that uses the SED of Vega as the zero colour reference object. Colour equations were used to transform between the IPHAS passbands and the Johnson-Cousins system of the published standard star photometry. The entire procedure has been found to deliver zeropoints which are accurate at the level of 1-2% in photometric conditions (González-Solares et al. 2011).

Unlike the broadbands, standard star photometry is not available in the literature for the  $H\alpha$  passband and hence there is no formally recognised flux scale for the narrowband. We can specify here however that the integrated in-band energy flux for Vega in the IPHAS  $H\alpha$  filter is  $1.52 \times 10^{-7}$  ergs  $\text{cm}^{-2} \text{s}^{-1}$  at the top of the Earth's atmosphere, which is the flux obtained by folding Vega's SED with the filter throughput curve corrected for atmosphere and detector quantum efficiency (using the method explained by Drew et al. 2005). This is 3.14 magnitudes less than the flux captured by the much broader  $r$  band which includes the  $H\alpha$  band. Hence to assure zero colour relative to the broadbands, we set the default zeropoint for the narrowband to be:

$$\text{MAGZPT}_{H\alpha} = \text{MAGZPT}_r - 3.14. \quad (2)$$

## 4.2 Uniform re-calibration

Despite the best efforts made to obtain a nightly calibration, large surveys naturally possess field-to-field variations at the level of 0.1 mag due to atmospheric fluctuations during the night and imperfections in the pipeline or the instrument (e.g. the WFC is known to suffer from sporadic glitches in the timing of exposures). Such variations need to be corrected for during a global re-calibration procedure. Notable past examples include the global re-calibration of the Two Micron All Sky Survey (2MASS; Nikolaev et al. 2000), the Sloan Digital Sky Survey (SDSS; Padmanabhan et al. 2008) and the Panoramic Survey Telescope and Rapid Response

System survey (Pan-STARRS; Schlafly et al. 2012), which all achieved photometry that is globally consistent to within 1-2%.

Surveys which observe identical stars at different epochs can use the repeat measurements to ensure a uniform calibration. For example, 2MASS attained its global calibration by observing six standard fields each hour, allowing zeropoint variations to be tracked over very short timescales (Nikolaev et al. 2000). Alternatively, the SDSS and PanSTARRS surveys could benefit from revisiting regions in their footprint to carry out a so-called *ubercalibration* procedure, in which multiple measurements of stars at different epochs are used to fit the calibration parameters (Padmanabhan et al. 2008; Schlafly et al. 2012).

Unfortunately these schemes cannot be applied directly to IPHAS for two reasons. Firstly, the survey was carried out in competitive observing time on a non-dedicated telescope, rendering the 2MASS approach of observing standards at a very high frequency prohibitively expensive – in part because standard fields are very scarce within the Galactic Plane. Secondly, the aim of IPHAS is to obtain magnitudes at a single epoch and hence stars are not normally observed at more than one epoch, unless they happen to fall on the narrow overlap region between two neighbouring fields.

Although the IPHAS data does contain a significant number of inter-field repeat measurements due to the small overlaps between neighbouring fields, we have found the information contained in these regions to be insufficient to constrain the calibration parameters. This is because photometry at the extreme edges of the WFC – where neighbouring fields overlap – is itself prone to systematics at the level of 1-2%. The cause of these errors is thought to include the use of twilight sky flats in the pipeline, which are known to be imperfect for calibrating stellar photometry due to stray light and vignetting (e.g. Manfroid 1995). Moreover, the illumination correction in the overlap regions is affected by a radial geometric distortion, which causes the pixel scale towards the edges to increase (González-Solares et al. 2011). Although these systematics are reasonably small within a single field, they can combine to cause artificial zeropoint gradients across the survey when they are used to constrain a global calibration without external constraints.

For these reasons, we have decided not to depend on an ubercalibration-type scheme alone, and have instead opted to exploit an external reference survey – where available – to bring the majority of our data onto a uniform calibration. This is explained next.

### 4.2.1 Correcting zeropoints using APASS

We have been able to benefit from the AAVSO Photometric All-Sky Survey (APASS; <http://www.aavso.org/apass>) to bring the vast majority of the survey onto a uniform scale. Since 2009, APASS has been using two 20 cm-astrographs to survey the entire sky down to  $\sim 17$ th magnitude in five filters which include Sloan  $r$  and  $i$ . The most recent catalogue available at the time of preparing this work was APASS DR7, which provided data across roughly half of the IPHAS footprint (Fig. 4). The photometric accuracy of APASS is currently estimated to be at the level of 3% (Henden, private communication), which is significantly better than the provisional calibration of IPHAS for which we estimate the

$1\sigma$ -error to be  $\sim 10\%$ . APASS achieves its uniform accuracy by measuring each star at least two times in photometric conditions – along with ample standard fields – using the large  $3 \times 3$  square degrees field of view of its detectors.

With the aim of bringing IPHAS to a similar accuracy of  $\sim 3\%$ , we used the APASS catalogue to identify and adjust all IPHAS fields which showed a magnitude offset larger than 3% against APASS. For this purpose, the  $r$ - and  $i$ -band detection tables of each IPHAS field were cross-matched against the APASS DR7 catalogue using a maximum matching distance of 1 arcsec. The magnitude range was limited to  $13 < r_{\text{APASS}} < 16.5$  and  $12.5 < i_{\text{APASS}} < 16.0$  in order to avoid sources brighter than the IPHAS saturation limit on one hand, and to avoid sources near the faint detection limit of APASS on the other hand.

The resulting set of  $\sim 220\,000$  cross-matched stars were then used to derive APASS-to-IPHAS magnitude transformations using a linear least-squares fitting routine, which iteratively removed  $3\sigma$ -outliers to improve the fit. The solution converged to:

$$r_{\text{IPHAS}} = r_{\text{APASS}} - 0.121 + 0.032(r - i)_{\text{APASS}} \quad (3)$$

$$i_{\text{IPHAS}} = i_{\text{APASS}} - 0.364 + 0.006(r - i)_{\text{APASS}} \quad (4)$$

The root mean square (rms) residuals of these transformations are 0.041 and 0.051, respectively. The small colour terms in the equations indicate that the  $r$  and  $i$  filters are very similar in both surveys. The transformations include a large fixed offset which is simply due to the fact that APASS magnitudes are given in the AB system and IPHAS uses magnitudes relative to Vega. Separate transformations were derived for sightlines with varying extinction properties to investigate the robustness of the transformations with respect to different reddening regimes, but the variations at these different sightlines were found to be insignificant. This is not surprising because heavily reddened objects are naturally scarce at  $r < 16$ .

Having transformed APASS magnitudes into the IPHAS system, we then computed the median magnitude offset for each field which contained at least 30 cross-matched stars. This was the case for 48% of the fields. The mean offset was found to be  $0.014 \pm 0.104$  for  $r$  and  $0.007 \pm 0.108$  for  $i$  (Table 1). A total of 4596 fields showed a median offset exceeding  $\pm 0.03$  mag in either  $r$  or  $i$ .

We then applied the most important step in our re-calibration scheme, which is to adjust the provisional zeropoints of the 4596 aberrant fields such that their offset is brought to zero. This allowed the mean IPHAS-to-APASS offset to be brought down to  $0.000 \pm 0.011$  in both  $r$  and  $i$  (Table 2). The procedure of fitting magnitude transformations and correcting the zeropoints was repeated a few times to ensure convergence, which was essentially reached after the first iteration.

#### 4.2.2 Adjusting fields not covered by APASS

At the time of writing, the APASS catalogue did not provide sufficient coverage for 7359 of the fields in our data release. Fortunately, these fields are pre-dominantly located in the early part of the Galactic Plane (Fig. 4), which were typically observed during the summer months when photometric conditions are more prevalent at the telescope. These remaining fields have nevertheless been brought onto the same

Before re-calibration	Mean	$\sigma$
$r$ (IPHAS - APASS)	+0.014	0.104
$i$ (IPHAS - APASS)	+0.007	0.108
$r$ (IPHAS - SDSS)	+0.016	0.088
$i$ (IPHAS - SDSS)	+0.010	0.089

**Table 1.** Mean magnitude offsets for objects crossmatched between IPHAS and APASS/SDSS before the uniform re-calibration. Eqns. 3–10 were applied to the APASS/SDSS magnitudes to bring them into the Vega-based IPHAS system prior to computing the offsets.)

After re-calibration	Mean	$\sigma$
$r$ (IPHAS - APASS)	+0.000	0.011
$i$ (IPHAS - APASS)	+0.000	0.011
$r$ (IPHAS - SDSS)	-0.001	0.029
$i$ (IPHAS - SDSS)	-0.002	0.032

**Table 2.** Same as Table 1 but computed after the uniform re-calibration was carried out.

uniform scale by employing an ubercalibration-style scheme, explained below, which minimises the magnitude offsets between stars located in the overlap regions with neighbouring fields. Although we explained earlier that these overlap regions are prone to systematics, the use of APASS enabled us to keep the zeropoints of roughly half of the fields fixed, which avoids these systematics from combining to introduce artificial gradients across the survey. In the next section we will show this to be true by validating our calibration against SDSS.

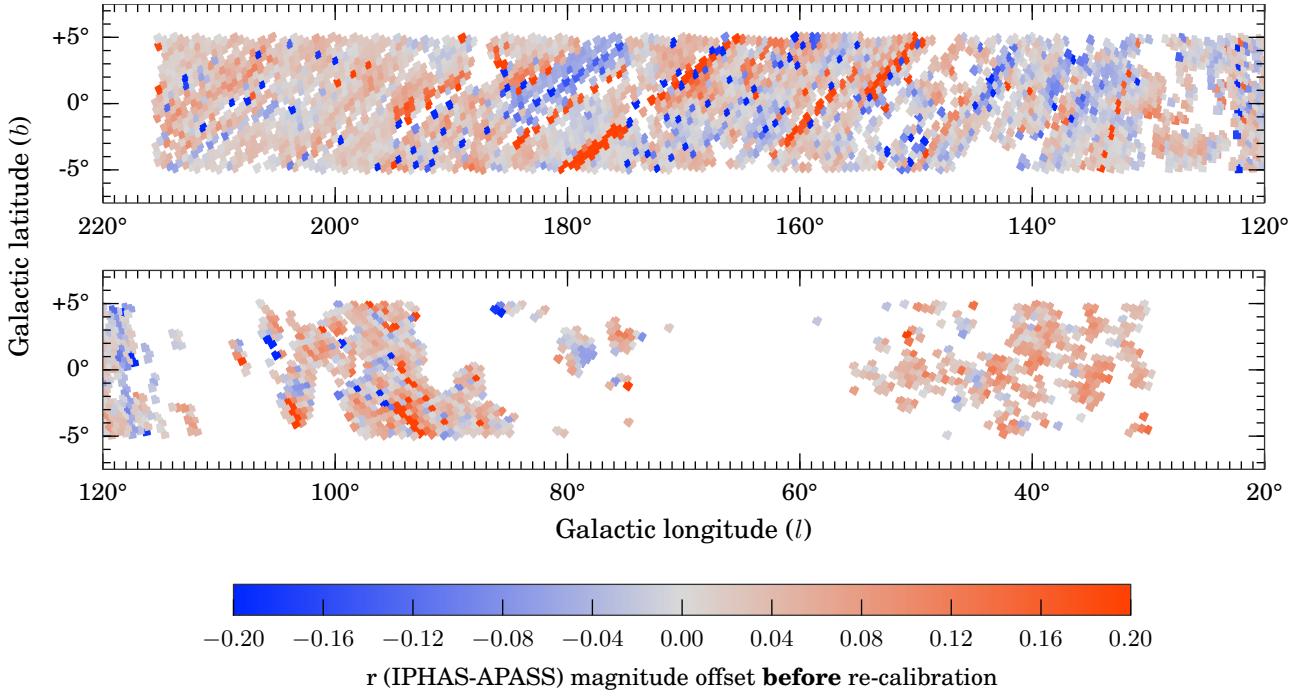
A general solution to the problem of minimising the magnitude differences between overlapping frames has previously been described by Glazebrook et al. (1994). In brief, there are two fundamental quantities to be minimised between each pair of overlapping exposures, denoted by the indices  $i$  and  $j$ . Firstly, the mean magnitude difference between stars in the overlap region  $\Delta_{ij} = \langle m_i - m_j \rangle = -\Delta_{ji}$  is a local constraint. Secondly, to ensure the solution does not stray far from the existing calibration, the difference in zeropoints  $\Delta ZP_{ij} = -\Delta ZP_{ji}$  between each pair of exposures must also be minimised.

Minimisation of these two quantities is a linear least squares problem because the magnitude  $m$  depends linearly on the ZP (Eqn. 1). Hence we can find the ZP shift to be applied to each field by minimising the sum:

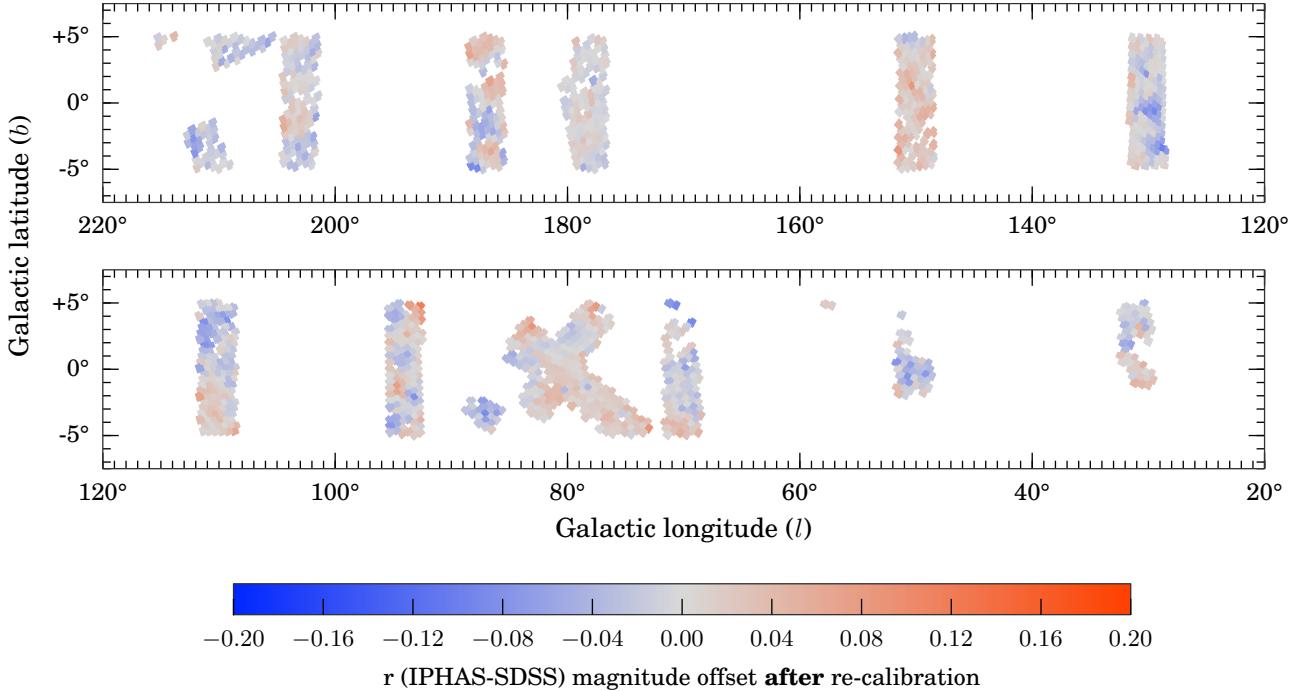
$$S = \sum_{i=1}^N \sum_{j=1}^N w_{ij} \theta_{ij} (\Delta_{ij} + a_i - a_j)^2 \quad (5)$$

where  $i$  denotes the exposure of interest,  $j$  an overlapping exposure,  $N$  the number of exposures,  $a_i$  the ZP to solve for and  $a_j$  the ZP of an overlapping field ( $\Delta ZP_{ij} = a_i - a_j$ ),  $w_{ij}$  are weights set to the uncertainty in  $\Delta_{ij}$  and  $\theta_{ij}$  is an overlap function equal to either 1 if exposures  $i$  and  $j$  overlap or 0 otherwise. Solving for  $a_i$  is equivalent to solving  $\partial S / \partial a_i = 0$  which gives the matrix equation:

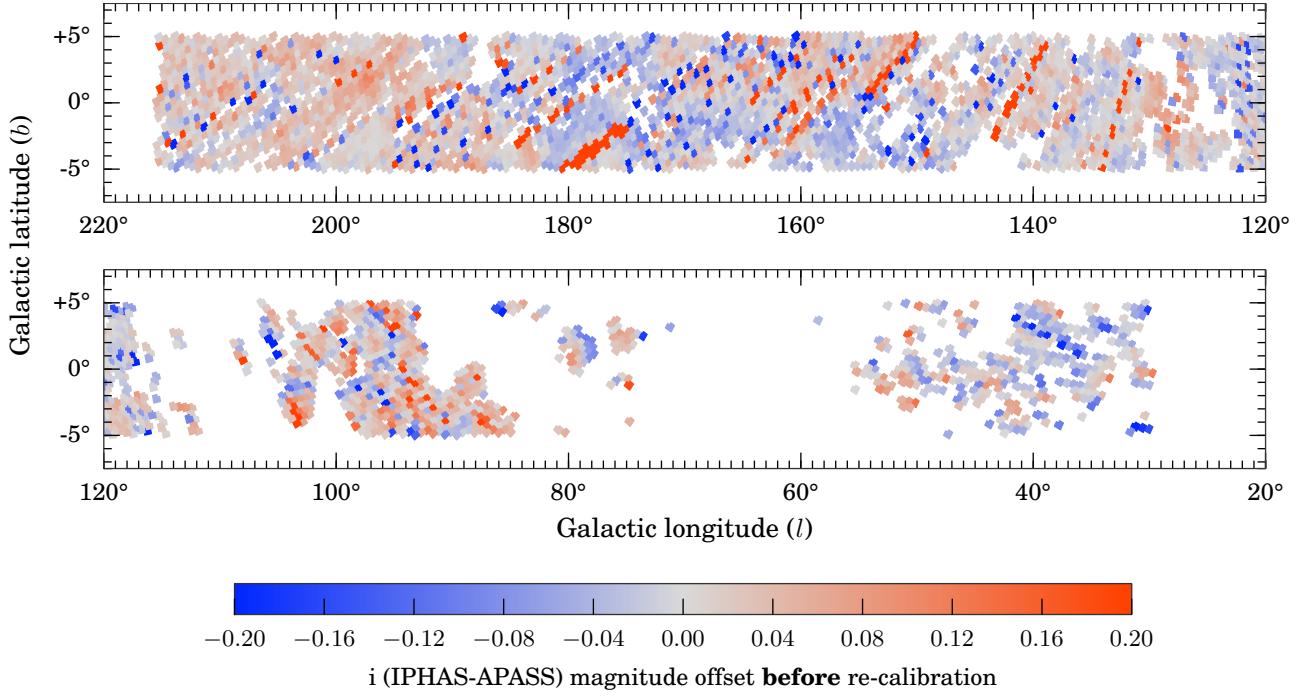
$$\sum_{j=1}^N A_{ij} a_j = b_j \quad (6)$$



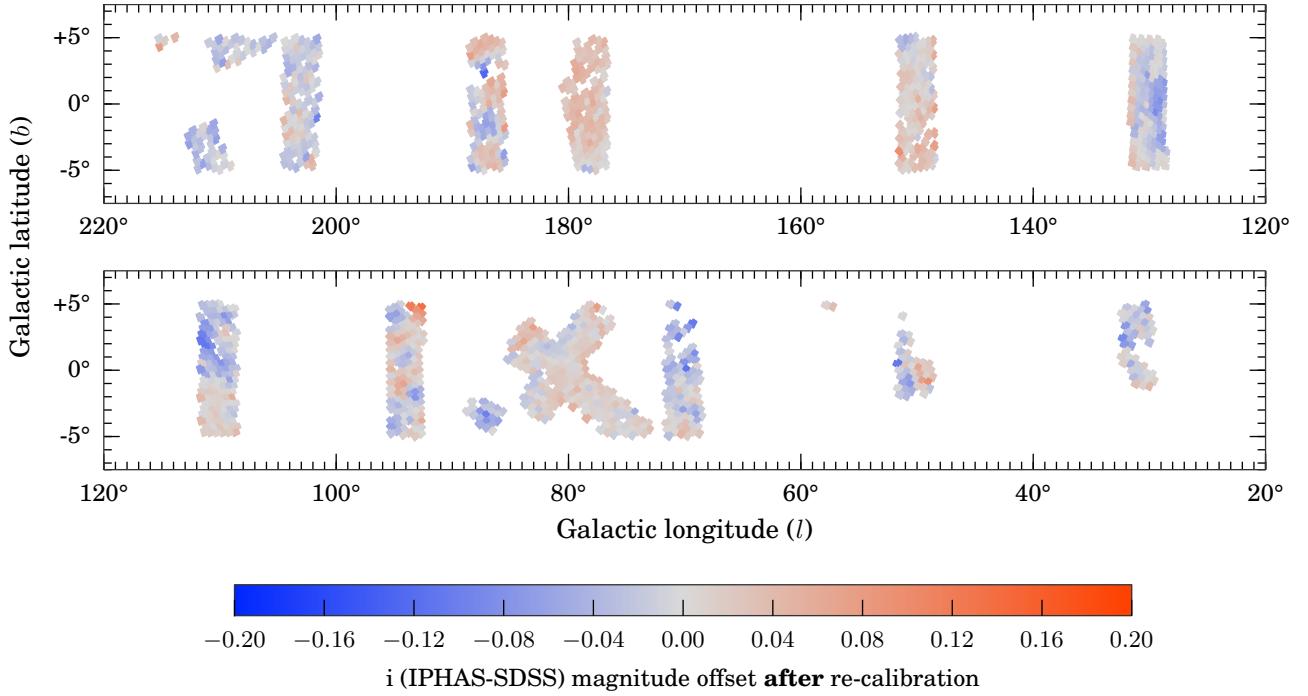
**Figure 4.** Median magnitude offset in the  $r$  band between IPHAS and APASS, plotted on a field-by-field basis prior to the re-calibration procedure. Each square represents an IPHAS field which contains at least 30 stars with a counterpart in the APASS DR7 catalogue. The colours denote the median IPHAS-APASS magnitude offset in each field, which was computed after applying the APASS-to-IPHAS transformation to the APASS magnitudes (Eqn. 3).



**Figure 5.** Median magnitude offset in the  $r$  band between IPHAS and SDSS after the re-calibration procedure was applied. Again, each square represents an IPHAS field which contains at least 30 stars with a counterpart in the SDSS catalogue. The colours denote the median IPHAS-SDSS magnitude offset in each field, which was computed after applying the SDSS-to-IPHAS transformation to the SDSS magnitudes (Eqn. 9).



**Figure 6.** Same as Figure 4 for the  $i$ -band.



**Figure 7.** Same as Figure 7 for the  $i$ -band.

where

$$A_{ij} = \delta_{ij} \sum_{k=1}^N w_{jk} \theta_{jk} - w_{ij} \theta_{ij}, \quad (7)$$

$$b_i = \sum_{j=1}^N w_{ij} \theta_{ij} \Delta_{ji} = - \sum_{j=1}^N w_{ij} \theta_{ij} \Delta_{ij}. \quad (8)$$

This prescription is essentially identical to Glazebrook et al. (1994).

As explained above, we enforce a strong external constraint on the solution by keeping the zeropoint fixed for the 6756 fields which have been compared and calibrated against APASS earlier. We will hereafter refer to these fields as *anchors*. It is asserted that the zeropoints  $a_i$  of these anchor fields are known and not solved for, though they do appear in the vector  $b_j$  as constraints. In addition to the APASS-based anchors, we selected 3273 additional anchor fields by hand to provide additional constraints in regions not covered by APASS. These extra anchors were deemed to have reliable zeropoints based on (i) the information contained in the observing logs, (ii) the stability of the standard star zeropoints during the night, and (iii) photometricity statistics provided by the Carlsberg Meridian Telescope, which is located at  $\sim 500$  meter from the INT.

We then solved Eqn. 6 for the  $r$  and  $i$  bands separately using the least-squares routine in Python's SCIPY.SPARSE module for sparse matrix algebra. This provided us with corrected zeropoints for the remaining fields, which were shifted on average by  $+0.02 \pm 0.11$  in  $r$  and  $+0.01 \pm 0.12$  in  $i$  compared to their provisional calibration.

We then turned to the uniform calibration of the  $H\alpha$  data. It is not possible to re-calibrate the narrowband in the same way as the broadbands, because the APASS survey does not offer  $H\alpha$  photometry. We can reasonably assume however, that the corrections required for  $r$  and  $H\alpha$  are identical because the  $H\alpha$  zeropoints have been derived directly from the  $r$ -band zeropoints during the provisional calibration (Eqn. 2). Moreover, the IPHAS data-taking pattern ensured that a field's  $H\alpha$  and  $r$ -band exposures were taken at essentially the same time, separated only by the  $\sim 30$ -second read-out time required by the WFC. We have hence corrected the  $H\alpha$  zeropoints by re-using the zeropoint adjustments that were derived for the  $r$  band in the earlier steps. An exception was made for 3101 fields for which our quality-control routines revealed zeropoint variations which exceeded 0.03 mag between consecutive fields, which indicates non-photometric conditions. For good practice, the  $H\alpha$  zeropoints of these fields were adjusted by solving Eqn. 6 rather than linking them directly to the  $r$ -band shift.

#### 4.3 Testing the calibration against SDSS

Having re-calibrated all fields to an expected accuracy of 3%, we then used an independent survey to validate the results. We have been able to exploit SDSS Data Release 9 (Ahn et al. 2012) for this purpose. SDSS DR9 provides several strips at low Galactic latitudes, providing data across 18% of the fields in our data release. We cross-matched IPHAS

fields against stars marked as reliable in the SDSS catalogue<sup>3</sup> in the same way as we did for APASS earlier, with the exception of using fainter magnitude ranges of  $15 < r_{\text{SDSS}} < 18.0$  and  $14.5 < i_{\text{SDSS}} < 17.5$ . This provided us with a set of 1.2 million cross-matched stars.

Colour transformations were obtained using a sigma-clipped linear least squares fit:

$$r_{\text{IPHAS}} = r_{\text{SDSS}} - 0.093 - 0.044(r - i)_{\text{SDSS}} \quad (9)$$

$$i_{\text{IPHAS}} = i_{\text{SDSS}} - 0.318 - 0.095(r - i)_{\text{SDSS}}. \quad (10)$$

The rms residuals of these transformations are 0.045 and 0.073, respectively. The equations are similar to the ones previously determined for APASS, although the colour terms are slightly larger; the throughput curve of the SDSS  $i$ -band filter appears to be somewhat more sensitive at longer wavelengths compared to both the IPHAS and APASS filters.

These global transformations were deemed adequate for the purpose of validating our uniform calibration in a statistical sense. Separate equations were derived towards different sightlines to investigate the effects of varying reddening regimes. The transformation coefficients were found to show some variation towards lowly reddened areas, which have relatively few numbers of (intrinsically) red objects at  $r - i > 1$  which can skew the colour term. The vast majority of red objects in the global sample are those in highly reddened areas however, which agree well with the global transformations and dominate the statistical appraisal of our calibration.

Having transformed SDSS magnitudes into the IPHAS system, we then computed the median magnitude offset for each IPHAS field which contained at least 30 objects with a counterpart in the SDSS catalogue. This was the case for 2602 fields. The median offsets for each of these fields are shown in Figs. 5–7. The mean offset and standard deviation found was  $-0.001 \pm 0.029$  for  $r$  and  $-0.002 \pm 0.032$  for  $i$  (Table 2). In comparison, offsets computed in the identical way before our re-calibration showed means of  $+0.016 \pm 0.088$  and  $+0.010 \pm 0.089$  (Table 2). We conclude that our re-calibration procedure has been successful in improving the uniformity of the calibration by a factor three and has achieved our aim of bringing the accuracy down to the level of  $\sigma = 0.03$  mag.

We warn that the SDSS comparison reveals a number of fields with offsets exceeding 0.05 mag (523 fields) or even 0.1 mag (18 fields). Such sporadic outliers are consistent with the tails of a Gaussian with mean  $\sim 0$  and  $\sigma = 0.03$ .

In future work, we hope to draw upon the PanSTARRS survey (Schlafly et al. 2012) to further improve the accuracy of our calibration. At the time of preparing this work data from PanSTARRS had not been made public yet.

## 5 SOURCE CATALOGUE GENERATION

Having obtained a quality-checked and re-calibrated data set, we now turn to the problem of transforming the observations into a user-friendly catalogue. The aim of this catalogue is to detail the best-available information for each

<sup>3</sup> We used the CasJobs facility located at <http://skyserver.sdss3.org/CasJobs> to obtain photometry from the SDSS PHOTOPRIMARY table with criteria  $\text{TYPE} = \text{STAR}$ ,  $\text{CLEAN} = 1$  and  $\text{SCORE} > 0.7$ .

unique source in a convenient format, including flags to warn about quality issues such as source blending and saturation. Compiling the catalogue essentially required four steps:

- (i) the single-band detection tables produced by the CASU pipeline were augmented with new columns and warning flags;
- (ii) the detection tables were merged into multi-band field catalogues;
- (iii) the overlap regions of the field catalogues were cross-matched to flag duplicate measurements and identify the best detection of each unique source; and
- (iv) these primary detections were compiled into the final source catalogue.

Each of these four steps are now explained.

### 5.1 User-friendly columns and warning flags

As the first step, the detection tables were enhanced by creating new columns. This is necessary because the tables generated by the CASU pipeline summarise the detections in their original CCD units, e.g. source positions are given in pixel coordinates and photometry in number counts. To transform these measurements into user-friendly fields, we have largely adopted the units and naming conventions which are in use at the WFCAM Science Archive (WSA; Hambly et al. 2008) and the VISTA Science Archive (VSA; Cross et al. 2012). These archives curate the near-infrared data from both the UKIDSS Galactic Plane Survey (GPS; Lucas et al. 2008) and the VISTA Variables in the Via Lactea survey (VVV; Minniti et al. 2010). Both these surveys provide high-resolution JHK photometry in the Galactic Plane. There is a significant degree of overlap between the footprints of UKIDSS/GPS and IPHAS, and hence by adopting a similar catalogue format we hope to encourage scientific applications which combine both data sets.

A detailed description of each column in our source catalogue is given in Appendix A. In the remainder of this section we highlight the main features.

Firstly, we note that each source is uniquely identified by an IAU-style designation of the form “IPHAS2 JHH-MMSS.ss+DDMMSS.s” (cf. column *name* in Appendix A), where “IPHAS2” refers to the present data release and the remainder of the string denotes the J2000 coordinates in sexagesimal format. For convenience, the coordinates are also included in decimal degrees (columns *ra* and *dec*) and in the Galactic coordinate system (columns *l* and *b*). We have also included an internal object identifier string of the form “#run-#ccd-#detection” (e.g. “64738-3-6473”), which documents the INT exposure number (#run), the CCD number (#ccd), and the row number in the CASU detection table (#detection) – see columns *rDetectionID*, *iDetectionID*, *haDetectionID*.

Photometry is provided based on the 2.3-arcsec diameter circular aperture by default (columns *r*, *i*, *ha*). The choice of this aperture size as the default is based on a trade-off between concerns about small number statistics and centroiding errors for small apertures on one hand, and diminishing signal-to-noise ratios and source confusion for large apertures on the other hand. The user is not restricted to this choice because the catalogue also provides magnitudes using three alternative aperture sizes: the peak pixel height

(columns *rPeakMag*, *iPeakMag*, *haPeakMag*), a circular 1.2-arcsec-diameter aperture (*rAperMag1*, *iAperMag1*, *haAperMag1*) and a 3.3-arcsec-diameter aperture (*rAperMag3*, *iAperMag3*, *haAperMag3*).

Each of these magnitude measurements have been corrected for the flux lost outside of their respective apertures, using a correction term which is inferred from the mean shape of the PSF measured locally in the CCD frame. In the case of a point source, the four alternative magnitudes are expected to be consistent with each other within the photon noise uncertainties (which are given in columns *rErr*, *rPeakMagErr*, *rAperMag1Err*, *rAperMag3Err*, etc). When this is not the case, it is likely that the source is either an extended object or that it has been incorrectly measured as a result of source blending or a rapidly spatially-varying nebulous background. In §6.1 we will explain that the consistency of the magnitude measurements in the different apertures can be used as a criterion for selecting stellar objects with reliable photometry from the catalogue.

The brightness of each object as a function of increasing aperture size is also used by the CASU pipeline to provide a discrete star/galaxy/noise classification flag (*rClass*, *iClass*, *haClass*) and a continuous stellarness-of-profile statistic (*rClassStat*, *iClassStat*, *haClassStat*). For convenience, we have combined these single-band morphological measures into band-merged class probabilities and flags (*pStar*, *pGalaxy*, *pNoise*, *mergedClass*, *mergedClassStat*) using the merging scheme in use at the WSA<sup>4</sup>.

Information on the quality of each detection is included in a series of additional columns. We draw attention to three useful flags which warn about the likely presence of systematics:

- (i) The *saturated* column is used to flag sources for which the peak pixel height exceeds 55000 counts, which is typically the case for stars brighter than 12-13th magnitude. Although the pipeline attempts to extrapolate the brightness of saturated stars based on the shape of their PSF, such extrapolation is prone to errors and we do not recommend their use.
- (ii) The *deblend* column is used to flag sources which partially overlap with a nearby neighbour. Although the pipeline applies a deblending procedure to such objects, the procedure is currently applied separately in each band and hence the *r-i* and *r-Ha* colours of such objects are prone to errors.
- (iii) The *brightNeighb* column is used to flag sources which are located within 5 arcmin from an object brighter than  $V = 7$  according to the Bright Star Catalogue (BSC; Hoffleit et al. 1991), or within 10 arcmin if the neighbour is brighter than  $V = 4$ . Such very bright stars are known to cause systematic errors and spurious detections as a result of stray light and diffraction spikes.

In addition to the above, we also created warning flags for internal bookkeeping. For example, we flagged detections which fell in the strongly vignetted regions of the focal plane, which were truncated by CCD edges, or which were otherwise affected by bad pixels in the detector. We will explain

<sup>4</sup> Explained at [http://surveys.roe.ac.uk/wsa/www/gloss\\_m.html#gpssource\\_mergedclass](http://surveys.roe.ac.uk/wsa/www/gloss_m.html#gpssource_mergedclass)

below that none of such detections have been included in the catalogue – an alternative detection was available in essentially all these situations because of the IPHAS field pair strategy – and hence these internal warning flags do not appear in the final source catalogue.

Finally, we note that basic information on the observing conditions is included (*fieldID*, *fieldGrade*, *night*, *seeing*). A table containing more detailed quality control information, indexed by *fieldID*, is made available on our website.

## 5.2 Band-merging the detection tables

The second step in compiling the source catalogue is to merge the contemporaneous trios of *r*, *i*, H $\alpha$  detection tables into multi-band field catalogues. This required a positional matching procedure to link sources between the three bands based on their position on the sky. We used the TMATCHN function of the STILTS software for this purpose, which allows rows from multiple tables to be matched (Taylor 2006). The result of the procedure is a band-merged catalogue in which each row corresponds to a group of linked *r/i/H $\alpha$*  detections which satisfy a maximum matching distance criterion in a pair-wise sense. Sources for which no counterpart was identified are retained in the catalogue as single-band detections.

We employed a maximum matching distance of 1 arcsec, which was chosen based on a trade-off between completeness and reliability. On one hand, a matching distance larger than 1 arcsec was found to allow too many spurious and unrelated sources to be linked. On the other hand, a value smaller than 1 arcsec would pose problems for very faint sources with large centroiding errors, and would occasionally fail to link detections near CCD corners where the astrometric solution can show systematic errors which exceed 0.5 arcsec. The position offsets between the *r* and *i/H $\alpha$*  detections have been included in the catalogue and can hence be tightened by the user if necessary (columns *iXi*, *iEta*, *haXi*, *haEta*). We note that also UKIDSS/GPS adopted a maximum matching distance of 1 arcsecond for similar reasons (Hambly et al. 2008).

The resulting band-merged catalogues were found to be reliable for the vast majority of fields. We warn that source blending and confusion is unavoidable for faint objects in the Galactic Plane however; in §6 we will show that 19% of the sources in our catalogue are flagged as blended objects (column *deblend*) and their band-merged data should be treated with care because they may have fallen victim to source confusion during the band-merging step.

## 5.3 Selecting the primary detections

We explained earlier that the survey contains repeat observations of identical sources as a result of overlaps in the data-taking pattern. Amongst all sources in the reliable magnitude range  $13 < r < 19$ , we find that 65% were detected twice and 25% were detected three times or more. Only 9% were detected once. Unsurprisingly, their spatial distribution traces the inter-CCD gaps and footprint edges.

The principal aim of this data release is to provide reliable photometry at a single epoch, and hence we have decided to focus on providing the magnitudes and coordinates using only the *best-available* detection of each object

– hereafter referred to as the *primary* detection. Although overlapping fields could have been co-added to gain a small improvement in depth, we have decided against this for two reasons. Firstly, combining the information from multiple epochs would make the photometry of variable stars difficult to interpret. Secondly, co-adding would cause the image quality to degrade towards the mean, which is a draw-back for crowded fields.

Anyone interested in the alternative detections of a source – hereafter called the *secondary* detections – can nevertheless obtain this information in two ways. To begin with, whenever a secondary detection was observed within 10 minutes of the primary, the magnitudes of that secondary detection have been included in the catalogue (columns *sourceID2*, *fieldID2*, *r2*, *i2*, *ha2*, *rErr2*, *iErr2*, *haErr2*, *errBits2*). This is the case for 66% of the sources brighter than  $r < 20$  due to the IPHAS field pair observing pattern. In addition, the complete set of detection tables – one for each exposure – is made accessible on our website to allow other uses of the data. A user-friendly catalogue of secondary detections has not been compiled at present but may be part of a future data release.

The primary detection is defined as the entry in the set of band-merged field catalogues which provides the most reliable information for a unique source. Primary detections have been selected using a so-called *seaming* procedure which has been adapted from the algorithm developed for the WSA<sup>5</sup>. In brief, the first step of the procedure is to identify all the duplicate detections by cross-matching the overlap regions of all field catalogues, again using a maximum matching distance of 1''. The duplicate detections for each unique source are then ranked according to (i) filter coverage, (ii) quality score (column *errBits*), and (iii) the average seeing of stars in the CCD frame rounded to 0.2 arcsec. If this ranking scheme reveals multiple ‘winners’ of identical quality, then the one that was observed closest to the optical axis of the camera is chosen.

## 5.4 Compiling the final source catalogue

As the final step, the primary detections that have been selected above are compiled into the 98-column source catalogue that is described in Appendix A and made available online. The entire list of sources naturally includes a significant number of spurious entries as a result of the very sensitive detection levels that are employed by the CASU pipeline by default. To limit the size of the source catalogue, we have decided to enforce three basic criteria which must be met for a candidate source to be included in the catalogue:

- (i) the source must have been detected at  $\text{SNR} > 5$  in at least one of the bands, i.e. it is required that at least one of *rErr*, *iErr* or *haErr* is smaller than 0.2 mag;
- (ii) the shape of the source must not be an obvious cosmic ray or noise artefact, i.e. we require either *pStar* or *pGalaxy* to be greater than 20%;
- (iii) the source must not have been detected in one of the strongly vignetted corners of the detector, not have had any known bad pixels in the aperture, and not have been on the

<sup>5</sup> <http://surveys.roe.ac.uk/wsa/dboverview.html#merge>

edge of one of the CCDs, i.e. we require the *errBits* quality score to be smaller than 64.

A total of 219 million primary detections satisfied the above criteria and have been included in the catalogue. Amongst these, 158 million objects are detected in all three bands (72%), 30 million are detected in two bands (14%), and 31 million entries are single-band detections (14%). Roughly half of the single-band detections were made in the *i*-band. This is likely explained by the fact that the *i*-band is least affected by interstellar extinction and can occasionally pick up highly-reddened objects which are otherwise lost in *r/H $\alpha$* .

## 6 DISCUSSION

Having explained how the catalogue was generated, we now offer an overview of its properties by discussing (i) the recommended quality criteria, (ii) the typical photometric uncertainties, and (iii) the source densities and the frequency of source blending.

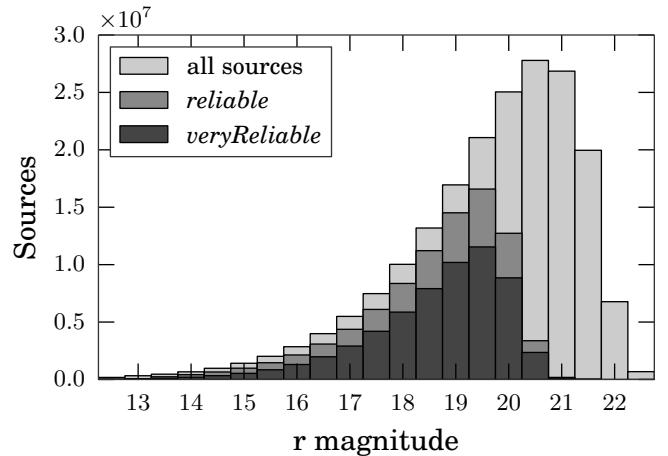
### 6.1 Recommended quality criteria

Like any other photometric survey, the majority of the objects in our catalogue are faint sources observed near the detection limits: 55% of the entries in the catalogue are fainter than  $r > 20$  and 18% are even fainter than  $r > 21$ . The measurements of such faint objects are naturally prone to large random and systematic uncertainties, for example, an inaccurately subtracted background will introduce a proportionally larger systematic error in a faint object. Most scientific applications will require a set of quality criteria to be enforced for the purpose of removing such faint and low-quality objects.

The choice of quality criteria is often a complicated trade-off between completeness on one hand and accuracy on the other. To aid users we have listed two sets of recommended quality criteria in Tables 3 and 4.

Firstly, Table 3 details a set of minimum quality criteria which should benefit most applications which require reliable  $r - i$  and  $r - \text{H}\alpha$  colours without removing more than  $\sim 80\%$  of the sources brighter than  $r < 19$ . The listed criteria are designed to (i) remove low-SNR sources, (ii) remove saturated sources, and (iii) remove objects for which the 2.3 arcsec diameter aperture magnitudes are inconsistent with the alternative 1.2 arcsec diameter aperture measurements. The last criterion is a proxy for identifying objects which are affected by inaccurate background subtraction or failed source deblending. A total of 86 out of 219 million sources (39%) satisfy all the criteria listed in Table 3 and are hereafter referred to as “reliable”. For convenience, the catalogue contains a boolean column named *reliable* which flags these objects and makes their selection easy.

For applications which require an even higher standard of reliability, a further set of additional quality criteria are listed in Table 4. These criteria are designed to ensure that (i) the object appeared as a perfect point source, (ii) the object was not blended with a nearby neighbour, and (ii) the object was not located near a very bright star. 59 million sources (27%) satisfy these additional criteria and are



**Figure 8.** r-band magnitude distribution for all objects in the catalogue (light grey), for objects flagged as *reliable* according to the criteria set out in Table 3 (grey), and for objects flagged as *veryReliable* following Table 4 (dark grey). The magnitude distributions for *i* and  $\text{H}\alpha$  look identical, apart from being shifted by about 1 and 0.5 mag towards brighter magnitudes, respectively.

hereafter referred to as “very reliable”. Again, the catalogue contains a boolean column named *veryReliable* which flags these objects.

Figure 8 compares the *r*-band magnitude distribution for objects with and without the *reliable* and *veryReliable* criteria applied. We find that 81% of the sources in the magnitude range  $13 < r < 19$  are considered *reliable*, which drops to 72% in the range  $19 < r < 20$  and 9% at  $r > 20$ . In contrast, only 54% of the sources in the magnitude range  $13 < r < 19$  are considered *veryReliable*. The stricter criteria filter out a lot of objects at early Galactic longitudes where source blending is a common problem (we will demonstrate this in §6.3). The *veryReliable* flag should hence only be used in applications which require very reliable photometry at the expense of completeness, which might be the case for e.g. spectroscopic target selection.

It is easy to see how the quality criteria may be adapted to be more tolerant. For example, by raising the allowed photometric uncertainties from 0.1 mag to 0.2 mag one can retrieve 42 million candidate sources.

### 6.2 Photometric uncertainties and repeatability

Figure 9 shows the mean photometric uncertainty as a function of magnitude for each band. We find that the uncertainty typically reaches 0.1 mag near 20.5 in *r* and 19.5 in *i/H $\alpha$*  when the default 2.3'' aperture is used. At this point we note that the average colour across all sources in the catalogue is  $1.06 \pm 0.12$  for  $(r-i)$  and  $0.44 \pm 0.03$  for  $(r-\text{H}\alpha)$ . The better depth of *r* is hence compensated by the fact that stars tend to have brighter magnitudes in both *i* and  $\text{H}\alpha$ .

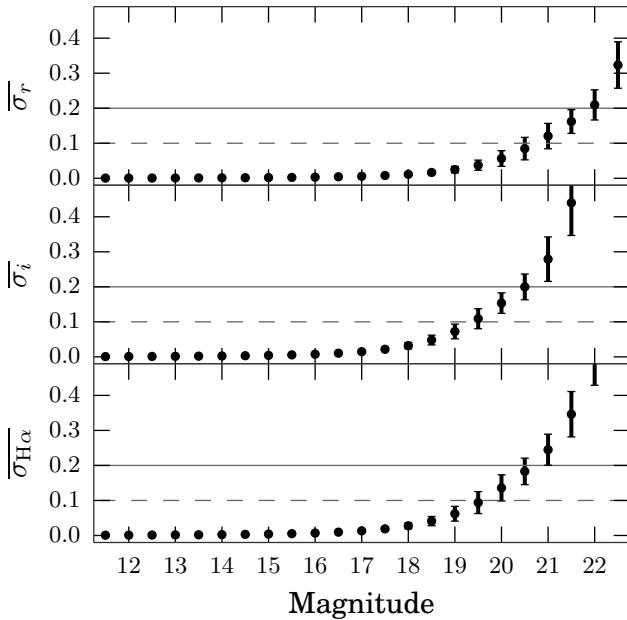
The uncertainties shown in Fig. 9 are the random errors based on the expected Poissonian photon noise. Systematics, such as calibration and deblending errors, are not included. To appraise the extent to which our photometry is affected by such systematics, we can exploit the secondary measurements which were made as part of the field-pair observing

Quality criterion	Rows passed	Description
$r\text{Err} < 0.1 \text{ AND } i\text{Err} < 0.1 \text{ AND } h\alpha\text{Err} < 0.1$	109 million (50%)	Require the photon noise to be less than 0.1 mag in all bands (i.e. $\text{SNR} > 10$ ). This implicitly requires a detection in all three bands.
$r > 13 \text{ AND } i > 12 \text{ AND } H\alpha > 12.5 \text{ AND NOT saturated}$	158 million (72%)	The brightness must not exceed the nominal saturation limit and the peak pixel height must not exceed 55 000 counts. Again, this implicitly requires a detection in all three bands.
$ r - r\text{AperMag1}  < 3\sqrt{r\text{Err}^2 + r\text{AperMag1Err}^2} + 0.03$	176 million (80%)	Require the $r$ magnitude measured in the default $2.3''$ diameter aperture to be consistent with the measurement made in the smaller $1.2''$ aperture, albeit tolerating a 0.03 mag systematic error. This will reject sources for which the background subtraction or the deblending procedure was not performed reliably.
$ i - i\text{AperMag1}  < 3\sqrt{i\text{Err}^2 + i\text{AperMag1Err}^2} + 0.03$	183 million (84%)	Same as above for $i$ .
$ h\alpha - h\alpha\text{AperMag1}  < 3\sqrt{h\alpha\text{Err}^2 + h\alpha\text{AperMag1Err}^2} + 0.03$	158 million (72%)	Same as above for $H\alpha$ .
All of the above (flagged as <i>reliable</i> )	86 million (39%)	

**Table 3.** Recommended minimum quality criteria for selecting objects with reliable colours from the IPHAS DR2 source catalogue. 86 million entries in the catalogue (39%) satisfy all the criteria listed in this table. For convenience, these have been flagged in the catalogue using the column named *reliable*.

Quality criterion	Rows passed	Description
<i>reliable</i>	86 million (39%)	The object must satisfy the criteria listed in Table 3.
$p\text{Star} > 0.9$	145 million (66%)	The object must appear as a perfect point source, as inferred from comparing its Point Spread Function (PSF) with the average PSF measured in the same CCD.
NOT <i>deblend</i>	177 million (81%)	The source must appear as a single, unconfused object.
NOT <i>brightNeigh</i>	216 million (99%)	There is no star brighter than $V < 4$ within 10 arcmin, or brighter than $V < 7$ within 5 arcmin. Such very bright stars cause scattered light and diffraction spikes, which may add systematic errors to the photometry or even trigger spurious detections.
All of the above (flagged as <i>veryReliable</i> )	59 million (27%)	

**Table 4.** Additional quality criteria which are recommended for applications which require very reliable colours at the expense of completeness. For convenience, the sources which satisfy the criteria listed in this table have been flagged in the catalogue using the column named *veryReliable*.



**Figure 9.** Mean photometric uncertainties for  $r$  (top),  $i$  (middle) and  $H\alpha$  (bottom). Data points shown are the average values of columns  $rErr$ ,  $iErr$  and  $haErr$  in the catalogue, and the errorbars show the standard deviations. The dashed and solid lines indicate the  $10\sigma$  and  $5\sigma$  limits, respectively. These uncertainties are based only on the (Poissonian) photon noise and hence this figure does not show systematic or calibration uncertainties.

strategy and are available for 51% of the sources in the catalogue.

Figure 6.2 shows the mean residuals between the primary and secondary magnitudes – i.e. the average difference between catalogue columns  $r-r2$ ,  $i-i2$ ,  $ha-ha2$  – as a function of magnitude. We find that sources across the magnitude range  $13 < r < 17$  are consistent at the level of 5% (i.e.  $\sigma_{r-r2} \leq 0.05$  mag), with the best repeatability reached at  $r = 14$  ( $\sigma_{r-r2} = 0.041$  mag). We draw attention to the fact that brighter stars tend to show significantly larger residuals – e.g.  $\sigma_{r-r2} = 0.14$  mag at  $r = 12$  – which is due to saturation effects. At the faint end we find residuals which show significantly more scatter than would be expected from photon noise alone, that is, the effects of source blending and background subtraction appear to dominate from  $\sim 20$ th magnitude onwards.

In Figure 6.2 we show a similar comparison of the primary and secondary detections, but this time we have only include sources which are flagged as *veryReliable* in the catalogue (i.e. not saturated, not confused, not near bright stars, etc.) We find that the average residuals are significantly better for this subset of the catalogue. Sources across the magnitude range  $13 < r < 17$  are consistent at the level of 0.03 mag, and the best repeatability is again reached at  $r = 14$  with  $\sigma_{r-r2} = 0.028$  mag. We conclude that the *veryReliable* quality criteria are effective in reducing the systematic errors to the same level as the accuracy of the global photometric calibration (cf. §4). Moreover, the large systematics at the bright and faint end have disappeared.

### 6.3 Source densities and blending

The mean source density as a function of Galactic longitude is shown in Figure 11 (thick blue line). The densities were computed by counting the number of sources in 1°-wide longitude bins, each covering the entire latitude range  $-5^\circ < b < +5^\circ$ . Unsurprisingly, we find the average source densities to increase towards the Galactic centre. For example, the average source density near  $l \simeq 30^\circ$  is roughly 30 000 objects per square degree, which is five times more than the density found near  $l \simeq 180^\circ$ .

In addition to the global trend, there are significant variations in the source density on smaller scales. For example, we find a significant drop near the constellations of Cygnus ( $l \simeq 80$ ) and Aquila ( $l \simeq 40$ ), which are regions known to be affected by high levels of foreground extinction. In fact dark clouds are visible towards these constellation by eye, and they have often been referred to as “the Great Rift”.

We warn however that the densities reported here have not been corrected for survey completeness or differences in the observing conditions across the survey. For example, the dip in the density near  $l \simeq 140^\circ$  is an artificial feature caused by gaps in the footprint coverage (which are apparent in Fig. 1). In a forthcoming paper, we aim to calibrate the IPHAS-based source densities by injecting artificial stars into the IPHAS images and measuring their recovery rate (Farnhill et al., in preparation). IPHAS has the potential to offer calibrated, two-dimensional stellar density maps which can be used to constrain detailed models of our Galaxy.

In Figure 11 we also shows the density of non-blended sources (thin red line). These are sources for which the *deblend* flag is FALSE, i.e. sources for which the CASU pipeline did not have to apply a deblending procedure to separate the flux originating from two or more overlapping objects. We find that the fraction of sources affected by blending problems is strongly correlated with the local source density. For example, only 11% of the sources are blended at  $l > 90^\circ$ , whereas 24% are blended at  $l < 90^\circ$ .

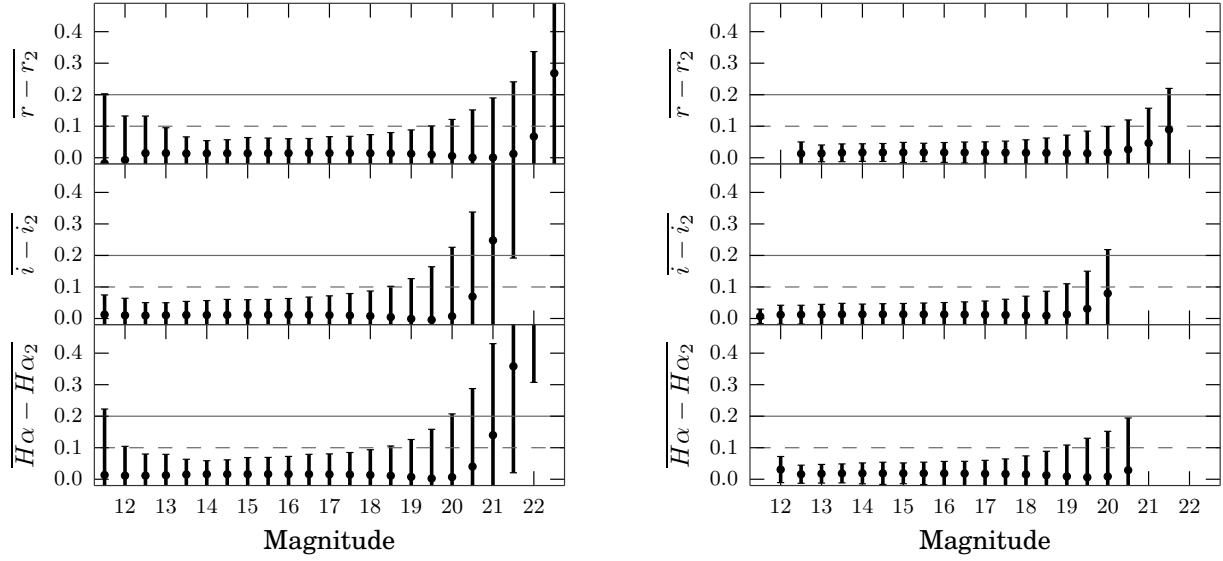
As we explained earlier, blended sources must be used with caution. In particular, we warn that blended sources are likely candidates to have fallen victim to source confusion during the band-merging procedure. In future work we hope to investigate the use of more advanced PSF-fitting routines in which sources are measured simultaneously across all bands, perhaps guided by an external catalogue such as the one that is to be provided by Gaia.

## 7 DEMONSTRATION

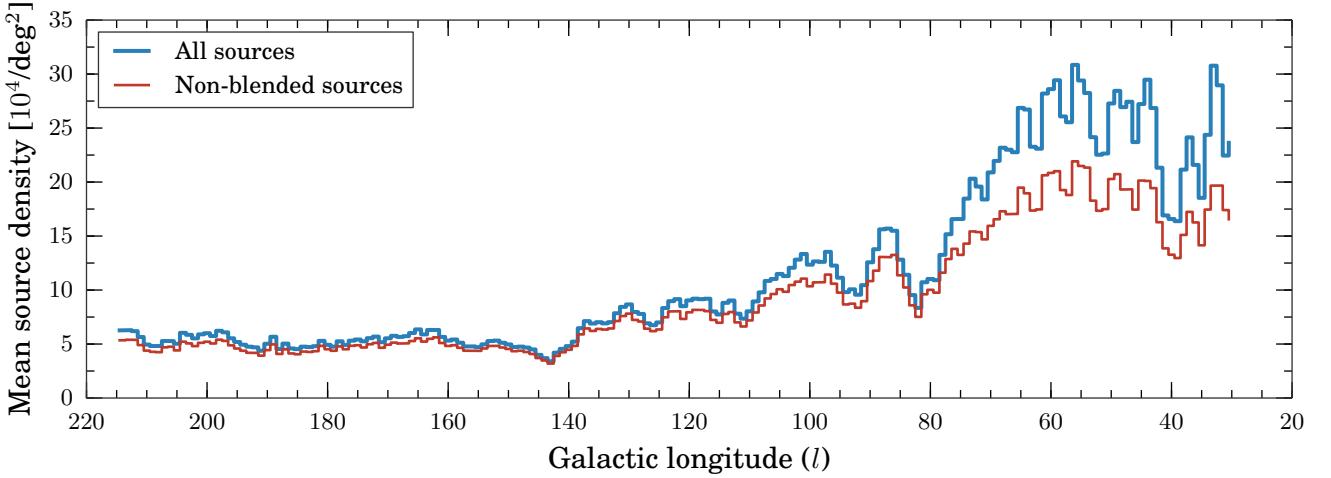
We conclude this paper by demonstrating how the unique  $r-i/r-H\alpha$  colour-colour diagram offered by this catalogue can readily be used to (i) characterise the extinction regime at different sightlines, and (ii) identify  $H\alpha$  emission-line objects.

### 7.1 Colour-colour and colour-magnitude diagrams

The survey’s unique  $r-H\alpha$  colour, when combined with  $r-i$ , has been shown to provide simultaneous constraints on intrinsic stellar colour and interstellar extinction (Drew et al. 2008). That is, the main sequence in the  $r-i/r-H\alpha$  diagram



**Figure 10.** Photometric repeatability as a function of magnitude for both all sources in the catalogue (left panel) and for the *veryReliable* sources alone (right panel). The values shown are the mean residuals between the primary and the secondary detections, while the errorbars show the standard deviations. Left panel: the best photometric repeatability is reached at  $r = 14$  with  $\sigma_{r-r_2} = 0.041$  mag. Bright stars at  $r < 13$  and  $i < 12$  show increasing uncertainties due to saturation effects. Right panel: we find that applying the quality criteria has improved the photometric repeatability significantly. The best repeatability is again reached at  $r = 14$  but has reduced to  $\sigma = 0.028$  mag. The quality criteria have also been successful at removing objects with large systematics at the bright and faint ends.



**Figure 11.** Mean number density of sources in the catalogue as a function of Galactic longitude, with and without blended sources included. The densities shown were computed by counting the sources at each longitude between  $-5^\circ < b < +5^\circ$  (upper blue line). We also show the densities based on only counting those sources for which the *deblend* flag is FALSE, i.e. unconfused sources for which the CASU pipeline did not have to apply a deblending procedure (lower red line).

runs in a direction that is different from the reddening vector, because the  $r$ - $H\alpha$  colour tends to act as a coarse proxy for spectral type and is less sensitive to reddening than  $r$ - $i$ . As a result, the distribution of a stellar population in the IPHAS colour-colour diagram can offer a handle on the properties and extinction regime along a line of sight.

This is demonstrated in Figures 12, 13 & 14, where we present three sets of IPHAS colour/magnitude diagrams towards three distinct sightlines located at Galactic longitudes  $180^\circ$ ,  $45^\circ$  and  $30^\circ$ , respectively. Each figure contains all the

sources which are flagged as *veryReliable* and are located in a region of one square degree centred on the coordinates indicated in the diagram (i.e. within a radius of  $0.564^\circ$  from the indicated sightline). For clarity, we have imposed the additional criterion that the photometric uncertainties must be smaller than 0.05 mag in each band (corresponding to a magnitude limit near  $\sim 19$ th magnitude).

Each of the diagrams reveals a well-defined locus, which demonstrates the health of the catalogue and the global calibration for investigating stellar populations across wide ar-

eas. We have annotated the colour-colour diagrams by showing the position of the unreddened main sequence (thin solid line), the unreddened giant branch (thick solid line), and the reddening track for an A0V-type star (dashed line) – all three are based on the Pickles (1998) library of empirical spectra transformed into the Vega-based IPHAS system by Drew et al. (2005). In the colour-magnitude diagrams we only show the reddening vector together with the unreddened 1 Gyr isochrone due to Bressan et al. (2012), which is made available for the IPHAS system through a popular on line tool hosted at the Observatory of Padova (<http://stev.oapd.inaf.it/cmd>). The isochrone and reddening vector have been placed at an arbitrary distance of 2 kpc.

Each of the sightlines reveal a stellar population with distinct characteristics. Firstly, towards the Galactic anti-centre at  $l = 180^\circ$  (Fig. 12) we find a population dominated by lowly-reddened main sequence stars. This is consistent with the estimated total sightline extinction of  $E(B-V) = 0.49$  given by Schlegel et al. (1998), following the correction of Schlafly & Finkbeiner (2011). Looking in more detail we can see that the stellar locus is narrower for M-type dwarfs than for earlier types; we do not observe M dwarfs experiencing the strongest reddening possible for this sightline. This implies that extinction is still increasing at distances beyond  $\sim 1\text{-}2$  kpc, where M dwarfs are too faint to be contained in the IPHAS catalogue. It is also clear that there are no unreddened stars earlier than  $\sim K0$  visible, such stars would be saturated if within a few hundred parsecs. This therefore suggests that there is a measurable increase in extinction locally. We also note a relative absence of late type giants which, due to the relative brevity of the corresponding phase of stellar evolution, would only account for a small proportion of a volume limited sample.

In contrast, closer towards the galactic centre at  $l = 45^\circ$  (Fig. 13) we find a wealth of reddened objects. In the colour-magnitude diagram it is clear that the stars are split into two distinct groups, with one significantly redder than the other. The bluer group is composed of main sequence stars, with the slope of this group in the colour-magnitude diagram attributable to the significantly increasing extinction. Meanwhile the redder group is principally composed of red giant stars. As these stars are intrinsically brighter, at some given apparent magnitude they will be substantially further away than their main sequence counterparts. Given that, along this sight-line, extinction continues to increase with distance, the red giants we observe will be subject to more reddening than a main sequence star exhibiting the same apparent magnitude. Therefore, the clear split between the two groups arises due to the combination of increasing extinction and the distinct absolute magnitude ranges of main sequence and giant branch stars.

Finally, in one of our earliest sightlines at  $l = 30^\circ$  we find a very high number of extremely reddened giants in addition to an unreddened population of foreground dwarfs. In contrast to the sightline at  $l = 45^\circ$ , there is no clear group of giant stars visible in the colour-magnitude diagram of Fig. 14, though the red clump stars are manifested as a slight over density which sits roughly 0.4 mags redder than the A0V reddening track. At  $(l, b) = (45^\circ, 2^\circ)$  the giant stars observed exhibit a relatively narrow range of redshifts as they lie beyond most or essentially all of the Galactic dust. At  $(l, b) = (30^\circ, 0^\circ)$  this is not the case: even at the substan-

tial distances at which we can observe reddened giant stars, extinction is continuing to rise and we still find ourselves well within the Galactic dust disc. It is also apparent that the  $(r-i)$  width of both the M dwarfs and early A dwarfs is greater than that in Fig. 13. This is indicative of a steeper rise in reddening, both within several hundred parsecs (M dwarfs) and within a few kpc (early A dwarfs).

Here we have presented only a short qualitative description of the information obtainable from IPHAS colour-colour and colour-magnitude plots. A more rigorous quantitative analysis of the IPHAS catalogue can be undertaken to estimate both the stellar density distribution in the Milky Way (Sale et al. 2010) and to create detailed three-dimensional maps of the extinction across several kpc (Sale et al. 2009; Sale 2012). An extinction map based on our catalogue is to be released in a separate paper that accompanies this data release (Sale et al., in preparation).

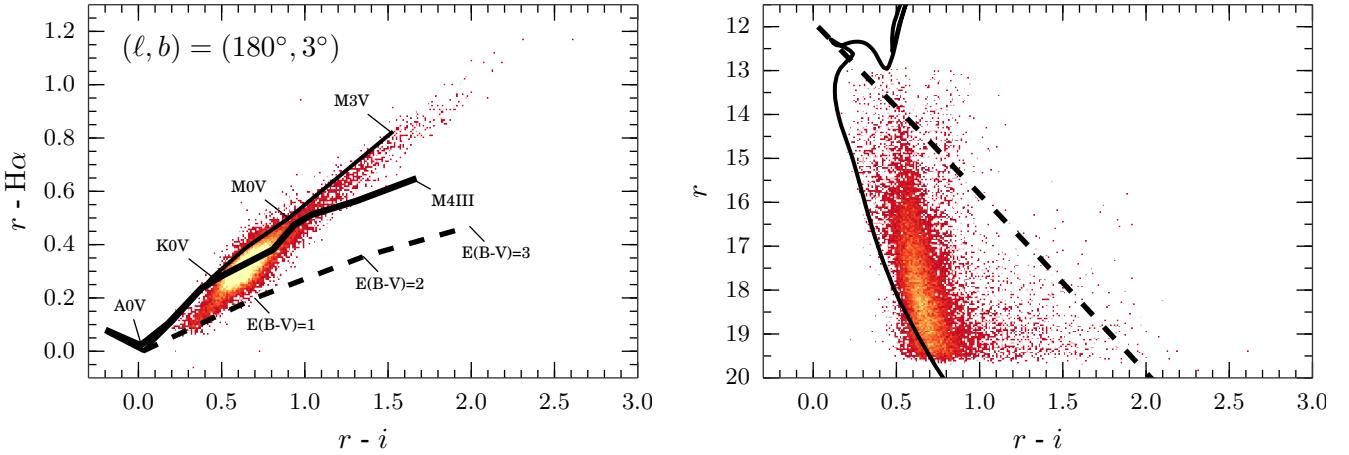
## 7.2 Identifying H $\alpha$ emission-line objects

A primary motivation for carrying out the survey was to enable the discovery of new emission-line objects across the Galactic Plane. H $\alpha$  in emission is a well-known tracer for stars in the short-lived pre- or post-main sequence stages of their evolution, and hence IPHAS allows larger, deeper and more statistically robust samples of such rare objects to be established. The survey-wide identification and analysis of emission-line objects is beyond the scope of the present work, but in this section we demonstrate how the catalogue may be used for this purpose.

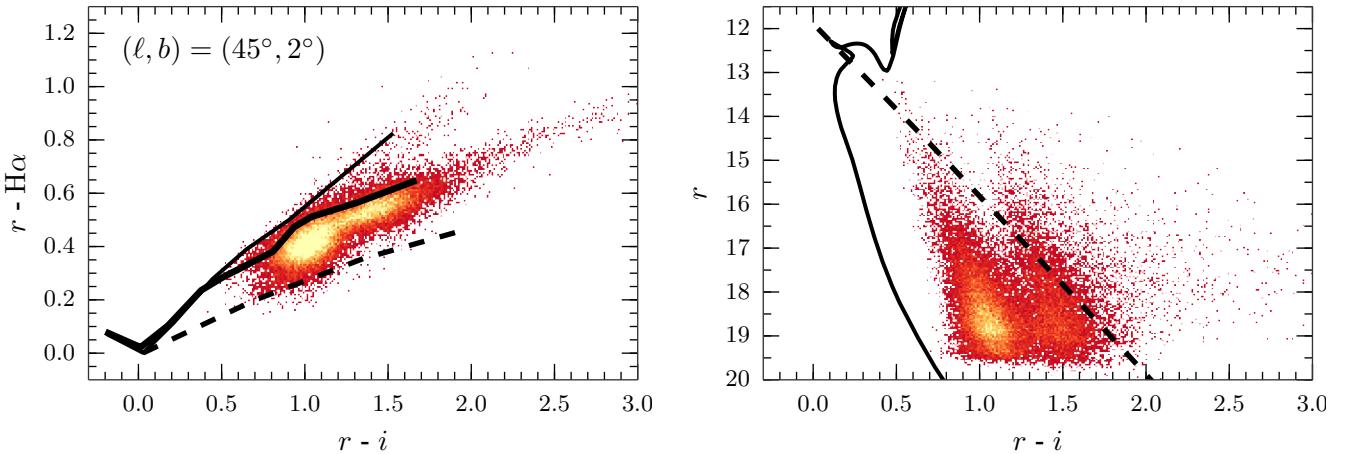
An initial list of candidate H $\alpha$ -emitters based on the first IPHAS data release has previously been presented by Witham et al. (2008). Because no global uniform calibration was available at the time, Witham et al. employed a sigma-clipping technique to select objects with large, outlying  $r$ -H $\alpha$  colours. In contrast, the new catalogue allows objects to be picked out from the  $r-i/r$ -H $\alpha$  colour-colour diagram using model-based colour criteria rather than a statistical procedure. In what follows we demonstrate this ability by selecting candidate emission-line objects towards a small region in the sky.

The target of our demonstration is Sh 2-82: a 5 arcmin-wide HII region located near  $(l, b) = (53.55^\circ, 0.00^\circ)$  in the constellation of Sagitta. Nicknamed by amateur astronomers as the “Little Cocoon Nebula”, Sh 2-82 is ionised by the  $\sim 10$ th magnitude star HD 231616 with spectral type B0V/III (Georgelin et al. 1973; Mayer & Macák 1973; Hunter & Massey 1990). The ionising star has been placed at a likely distance of 1.5–1.7 kpc based on its photometric parallax (Mayer & Macák 1973; Lahulla 1985; Hunter & Massey 1990).

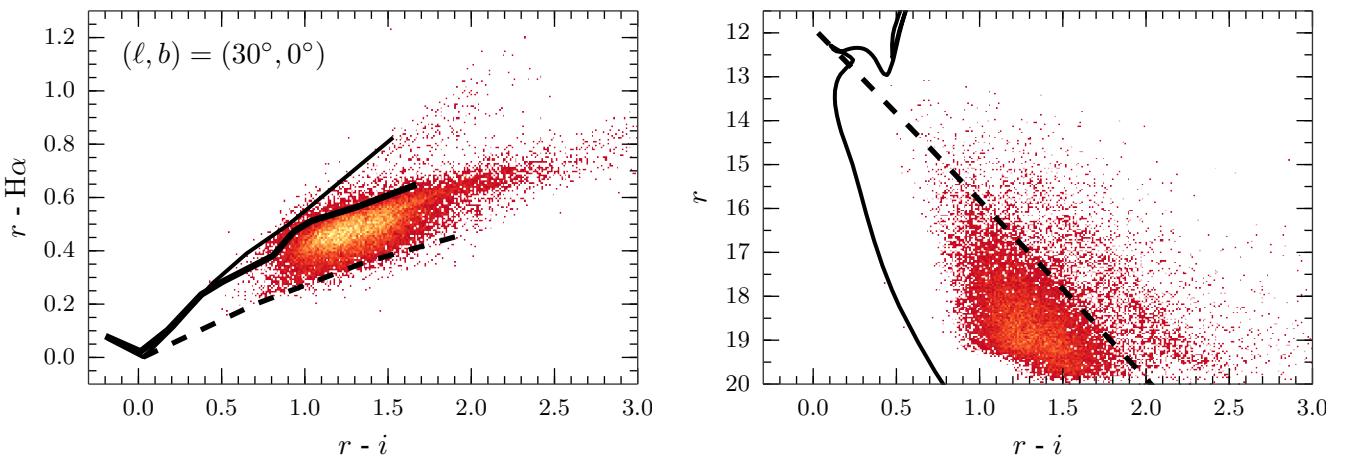
Figure 15 shows a 20-by-15 arcmin colour mosaic centred on Sh 2-82, composed of our H $\alpha$  (red channel),  $r$  (green channel), and  $i$  (blue channel) images. The ionising star can be seen as the brightest object in the centre of the HII region, which is surrounded by a faint reflection nebula and several dark cloud filaments. For comparison, Figure 16 shows a mosaic of the same region as seen in the mid-infrared by the Spitzer Space Telescope. The infrared image reveals a bubble-shaped structure of warm dust which surrounds the entire HII region, and is thought to contain a wealth of infrared-bright young stellar objects (Yu & Wang 2012).



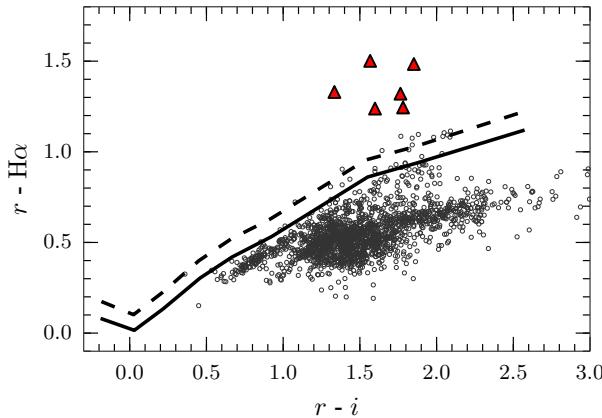
**Figure 12.** Colour-colour and colour-magnitude diagrams (left and right panel) showing sources flagged as *veryReliable* located in an area of one square degree centred near the Galactic anti-centre at  $(l, b) = (180^\circ, +3^\circ)$ . The diagrams are plotted as 2D-histograms which show the density of objects in bins of  $\sim 0.01 \times 0.01$  mag; bins containing 1-10 objects are coloured red, bins with 10-20 objects are orange, and bins with 20-25 objects are yellow. The left panel is annotated with the position of the main sequence (thin solid line), giant stars (thick solid line) and the reddening track for an A0V-type star (dashed line), which are based on the Pickles (1998) library of observed spectra. The right panel only shows the reddening vector along with the unreddened 1 Gyr isochrone due to Bressan et al. (2012), which has been placed at an arbitrary distance of 2 kpc. This is one of the least reddened sightlines in the survey and hence the observed stellar population appears to be dominated by lowly reddened main sequence stars (see text).



**Figure 13.** Same as above for  $(l, b) = (45^\circ, +2^\circ)$ , which is one of the highest-density sightlines in the survey, revealing two groups of stars in colour-magnitude space.



**Figure 14.** Same as above for  $(l, b) = (30^\circ, 0^\circ)$ , showing one of the most reddened sightlines in the survey.



**Figure 17.**  $r\text{-}i/r\text{-H}\alpha$  diagram for the rectangular region of 20-by-15 arcmin centred on the HII region Sh 2-82, which is the area shown in Figure 15. The diagram shows all objects in the catalogue which have been flagged as *reliable* and are brighter than  $r < 20$  (grey circles). The unreddened main sequence is indicated by a solid line, while the main sequence for stars with an  $\text{H}\alpha$  emission line strength of  $-10 \text{\AA}$  EW is indicated by a dashed line, following the colour simulations due to Barentsen et al. (2011). Red triangles indicate objects which have been identified as as highly likely  $\text{H}\alpha$ -emitters (see text).

Name [IPHAS2 ...]	$r$	$i$	$\text{H}\alpha$
J192954.40+181026.1	$17.69 \pm 0.01$	$16.12 \pm 0.01$	$16.19 \pm 0.01$
J193011.01+182051.2	$18.55 \pm 0.02$	$16.95 \pm 0.02$	$17.31 \pm 0.02$
J193021.52+181954.5	$19.72 \pm 0.05$	$17.94 \pm 0.03$	$18.47 \pm 0.04$
J193024.45+181938.3	$19.31 \pm 0.04$	$17.55 \pm 0.02$	$17.99 \pm 0.03$
J193033.00+181609.3	$18.25 \pm 0.01$	$16.91 \pm 0.01$	$16.92 \pm 0.01$
J193042.48+182317.4	$19.96 \pm 0.03$	$18.11 \pm 0.03$	$18.48 \pm 0.03$

**Table 5.** Candidate  $\text{H}\alpha$ -emitters towards Sh 2-82.

Many of these young objects appear in Fig. 16 as red- and pink-coloured objects, located predominantly in the north-western part of the bubble-shaped structure.

Figure 17 presents the IPHAS colour-colour diagram for the 20-by-15 arcmin region shown in the mosaics. Gray circles show all objects which are brighter than  $r < 20$  and have been flagged as *reliable* in our catalogue. The diagram also shows the unreddened main sequence (solid line) and the expected position of unreddened main-sequence stars with  $\text{H}\alpha$  in emission at a strength of  $\text{EW} = -10 \text{\AA}$  (dashed line), taken from the colour simulations due to Barentsen et al. (2011). Six stars are found to lie confidently above the dashed line at the level of  $3\sigma$ , that is, the distance between the objects and the dashed line is larger than three times the uncertainty on the  $r\text{-H}\alpha$  colour. These reliable candidate  $\text{H}\alpha$ -emitters are marked by red triangles in the colour-colour diagram and their photometry is detailed in Table 5. The spatial distribution of the six candidate emission-line objects is marked by yellow triangles in the colour mosaic (Fig. 15).

In previous work, we have shown how the majority of  $\text{H}\alpha$ -emitters seen towards an HII region are likely to be Classical T Tauri Stars (CTTS), which are young objects which show  $\text{H}\alpha$  in emission due to the presence of hot, infalling gas which is accreting onto the star from a circumstellar disk Barentsen et al. (2011). This is likely to be the case in this demonstration as well. Two of our candidates have recently been identified as Young Stellar Objects (YSO) in an investi-

gation of the region by Yu & Wang (2012). In their study, the authors used 2MASS and Spitzer data to select likely young stars by selecting objects with an infrared colour excess, consistent with the presence of a heated circumstellar disk. Although Yu & Wang did not identify four of our candidates as having an infrared excess, we note that each of them is detected at  $\text{S/N} > ?$  in the Spitzer  $8.0 \mu\text{m}$  image, which is more likely for a circumstellar-bearing object than it is for a field star.

Prior to IPHAS, this region was essentially unstudied at faint magnitudes in visible light. Lahulla (1985) reported magnitudes for 8 stars in the optical at  $V < 15$ . In contrast, the IPHAS catalogue offers photometry for 2197 objects in the region down to  $r < 20$ .

Sh 2-82 is one of a wealth of poorly-studied star-forming regions at low Galactic latitudes, which have started to become revealed in recent years through catalogues of ‘‘bubbles’’ seen in mid-infrared surveys (e.g. Churchwell et al. 2006; Simpson et al. 2012) and previously unknown clusters seen at near-infrared wavelengths (e.g. Bica et al. 2003). The optical data provided by IPHAS offers a handle (or lower limit) on the extinction and distance towards many of these regions, and can provide information on the optically-unveiled members of the region, as we demonstrated here.

## 8 CONCLUSIONS AND FUTURE WORK

A new data release for the IPHAS survey was presented, taking the coverage up to over 90% of the Northern Galactic Plane at  $|b| < 5^\circ$  and providing a uniform photometric calibration for the first time. We explained the data reduction and quality control procedures that were applied, described and tested the new global photometric calibration, and detailed the construction of the user-friendly source catalogue.

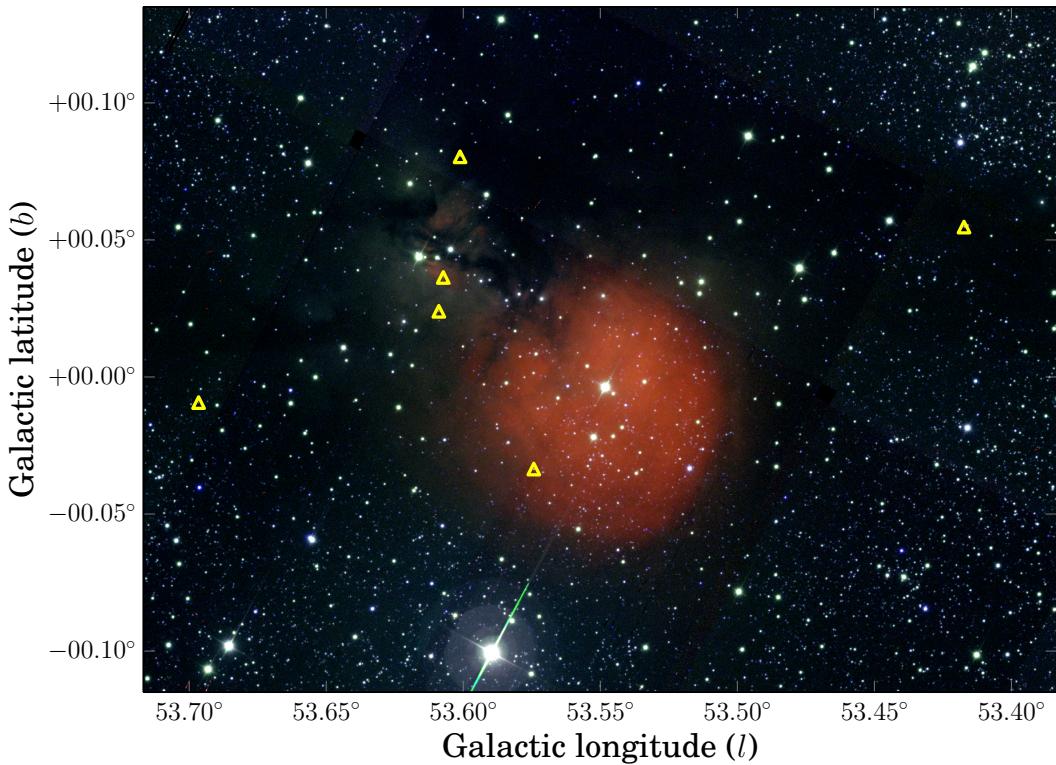
The observations included in this release achieved a median seeing of 1.1 arcsec and a  $5\sigma$ -depth of  $r = 21.2 \pm 0.5$ ,  $i = 20.0 \pm 0.3$ ,  $\text{H}\alpha = 20.3 \pm 0.3$ . The global calibration and photometric repeatability was found to be accurate at the level of  $\sigma = 0.03$  mag, providing a significant improvement over the previous data release. The source catalogue provides the best-available single-epoch astrometry and photometry for 219 million unique sources.

In future work, we aim to draw upon the PanSTARRS photometric survey to further improve the accuracy of our global calibration. We will also aim to correct the IPHAS photometry for the radial field distortions, and aim to produce a catalogue which details all the secondary detections to aid time-domain studies.

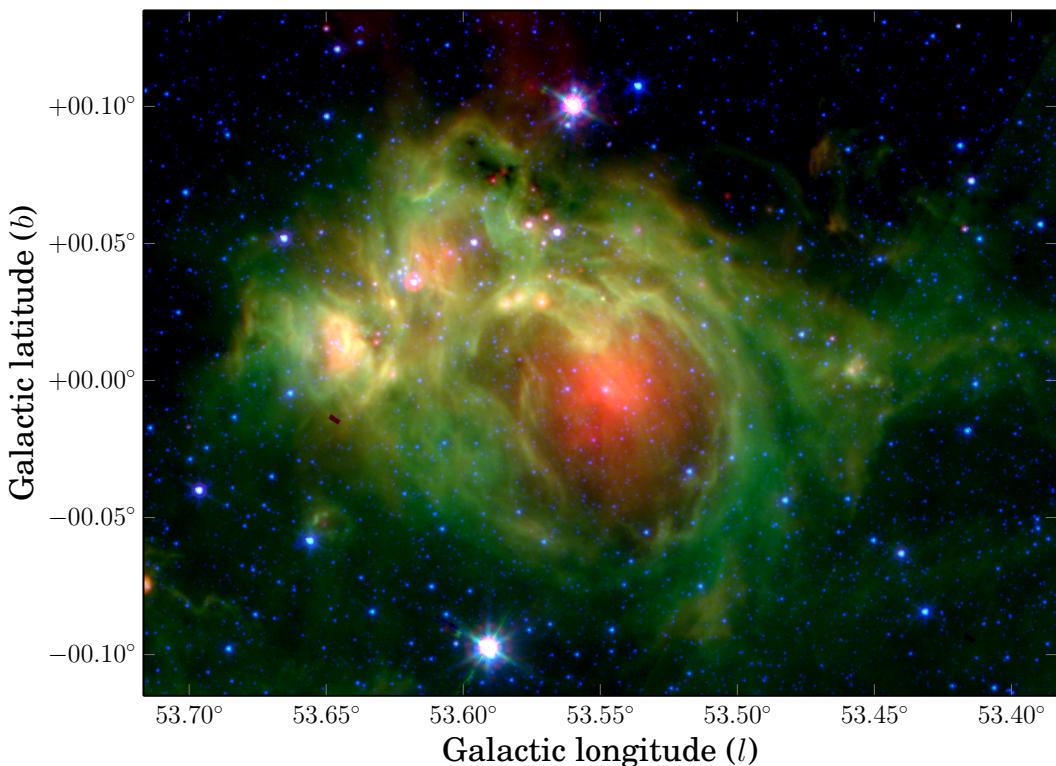
The data-taking strategy developed for IPHAS have since been reapplied to carry out a companion INT/WFC survey called UVEX in  $U/g/r/\text{He I}$  (Groot et al. 2009) and a southern companion survey in  $u/g/r/i/\text{H}\alpha$  called VPHAS+ (Drew et al. 2014). We intend to use the present data release as a template for future releases from these companion surveys.

## DATA ACCESS AND SOURCE CODE

The catalogue is made available through the Vizier catalogue tool (<http://vizier.u-strasbg.fr>), where it is known as the



**Figure 15.** IPHAS-based mosaic of HII region Sh 2-82, composed of H $\alpha$  (red channel),  $r$  (green channel) and  $i$  (blue channel). Yellow triangles show the position of candidate H $\alpha$ -emitters which have been selected from the colour-colour diagram in Figure 17. Note that the HII region is surrounded by a faint blue/green reflection nebula and dark cloud filaments.



**Figure 16.** Star-forming region Sh 2-82 as seen in the mid-infrared by the Spitzer Space Telescope. The mosaic is composed of the 24  $\mu\text{m}$  (red), 8.0  $\mu\text{m}$  (green) and 4.5  $\mu\text{m}$  (blue) bands. The image reveals a bubble-shaped structure which surrounds the HII region that is seen in the IPHAS mosaic of the same region (Figure 15).

“IPHAS DR2 Source Catalogue” (catalogue ID ???). The full catalogue can also be downloaded in its entirety from the IPHAS website ([www.iphas.org](http://www.iphas.org)) as a collection of binary FITS tables which comprise 50 GB, which is accompanied by a script to ingest the data into a PostgreSQL database. Our website also provides access to the pipeline-processed imaging data, which we have updated to include the re-calibrated DR2 zeropoint in the image headers.

The source code that was used to generate the catalogue is available at <https://github.com/barentsen/iphas-dr2>

## ACKNOWLEDGMENTS

The IPHAS survey was carried out at the Isaac Newton Telescope (INT). The INT is operated on the island of La Palma by the Isaac Newton Group in the Spanish Observatorio del Roque de los Muchachos of the Instituto de Astrofísica de Canarias. All data were processed by the Cambridge Astronomical Survey Unit at the Institute of Astronomy in Cambridge.

Preparation of the catalogue was eased greatly by a number of software packages, including the PostgreSQL database software, the TOPCAT and STILTS packages (Taylor 2005, 2006), and the Python modules AstroPy (Astropy Collaboration et al. 2013), NumPy/SciPy (Oliphant 2007), Matplotlib (Hunter 2007), IPython (Pérez & Granger 2007), and APLpy. We also made use of the Montage software maintained by NASA/IPAC, and the SIMBAD, Vizier and Aladin (Bonnarel et al. 2000) services operated at CDS, Strasbourg, France.

This research made extensive use of several complementary photometric surveys. Our global calibration was aided by the AAVSO Photometric All-Sky Survey (APASS), funded by the Robert Martin Ayers Sciences Fund. The calibration was tested against the Sloan Digitized Sky Survey (SDSS), funded by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The astrometric pipeline reduction made significant use of the Two Micron All Sky Survey (2MASS), which is a joint project of the University of Massachusetts and the Infrared Processing and Analysis Center/California Institute of Technology, funded by NASA and the NSF.

GB and JED acknowledge the support of a grant from the Science & Technology Facilities Council of the UK (STFC, ref ST/J001335/1). HJF is in receipt of an STFC postgraduate studentship.

## REFERENCES

- Ahn C. P. et al., 2012, ApJS, 203, 21  
 Astropy Collaboration et al., 2013, A&A, 558, A33  
 Barentsen G. et al., 2011, MNRAS, 415, 103  
 Barentsen G., Vink J. S., Drew J. E., Sale S. E., 2013, MNRAS, 429, 1981  
 Bica E., Dutra C. M., Soares J., Barbuy B., 2003, A&A, 404, 223  
 Bonnarel F. et al., 2000, A&AS, 143, 33  
 Bressan A., Marigo P., Girardi L., Salasnich B., Dal Cero C., Rubele S., Nanni A., 2012, MNRAS, 427, 127  
 Churchwell E. et al., 2006, ApJ, 649, 759  
 Corradi R. L. M. et al., 2008, A&A, 480, 409  
 Corradi R. L. M. et al., 2011, MNRAS, 410, 1349  
 Corradi R. L. M. et al., 2010, A&A, 509, A41  
 Cross N. J. G. et al., 2012, A&A, 548, A119  
 Drew J. E. et al., 2014, ArXiv e-prints  
 Drew J. E. et al., 2005, MNRAS, 362, 753  
 Drew J. E., Greimel R., Irwin M. J., Sale S. E., 2008, MNRAS, 386, 1761  
 Georgelin Y. M., Georgelin Y. P., Roux S., 1973, A&A, 25, 337  
 Giannamico C. et al., 2011, A&A, 525, A58  
 Glazebrook K., Peacock J. A., Collins C. A., Miller L., 1994, MNRAS, 266  
 González-Solares E. A. et al., 2011, MNRAS, 416, 927  
 González-Solares E. A. et al., 2008, MNRAS, 388, 89  
 Groot P. J. et al., 2009, MNRAS, 399, 323  
 Hambly N. C. et al., 2008, MNRAS, 384, 637  
 Hunter D. A., Massey P., 1990, AJ, 99, 846  
 Hunter J. D., 2007, Computing In Science & Engineering, 9, 90  
 Irwin M., Lewis J., 2001, New Astronomy Reviews, 45, 105  
 Irwin M., McMahon R., Walton N., González-Solares E., Hodgkin S., Irwin J., Lewis J., 2005, The Newsletter of the Isaac Newton Group of Telescopes, 9, 8  
 Irwin M. J., 1985, MNRAS, 214, 575  
 Irwin M. J., 1997, Detectors and data analysis techniques for wide field optical imaging., Rodríguez Espinosa J. M., Herrero A., Sánchez F., eds., pp. 35–74  
 Kohoutek L., Wehmeyer R., 1999, A&AS, 134, 255  
 Lahulla J. F., 1985, A&AS, 61, 537  
 Landolt A. U., 1992, AJ, 104, 340  
 Lucas P. W. et al., 2008, MNRAS, 391, 136  
 Mampaso A. et al., 2006, A&A, 458, 203  
 Manfroid J., 1995, A&AS, 113, 587  
 Mayer P., Macák P., 1973, Bulletin of the Astronomical Institutes of Czechoslovakia, 24, 50  
 McMahon R. G., Walton N. A., Irwin M. J., Lewis J. R., Bunclark P. S., Jones D. H., 2001, New Astronomy Reviews, 45, 97  
 Minniti D. et al., 2010, New A, 15, 433  
 Nikolaev S., Weinberg M. D., Skrutskie M. F., Cutri R. M., Wheelock S. L., Gizis J. E., Howard E. M., 2000, AJ, 120, 3340  
 Oliphant T. E., 2007, Computing in Science & Engineering, 9, 10  
 Padmanabhan N. et al., 2008, AJ, 674, 1217  
 Pérez F., Granger B. E., 2007, Computing in Science and Engineering, 9, 21  
 Pickles A. J., 1998, PASP, 110, 863  
 Sale S. E., 2012, MNRAS, 427, 2119  
 Sale S. E. et al., 2010, MNRAS, 402, 713  
 Sale S. E. et al., 2009, MNRAS, 392, 497  
 Schlafly E. F., Finkbeiner D. P., 2011, ApJ, 737, 103  
 Schlafly E. F. et al., 2012, ApJ, 756, 158  
 Schlegel D. J., Finkbeiner D. P., Davis M., 1998, ApJ, 500, 525  
 Simpson R. J. et al., 2012, MNRAS, 424, 2442  
 Skrutskie M. F. et al., 2006, AJ, 131, 1163

- Taylor M. B., 2005, in Astronomical Society of the Pacific Conference Series, Vol. 347, Astronomical Data Analysis Software and Systems XIV, Shopbell P., Britton M., Ebert R., eds., p. 29
- Taylor M. B., 2006, in Astronomical Society of the Pacific Conference Series, Vol. 351, Astronomical Data Analysis Software and Systems XV, Gabriel C., Arviset C., Ponz D., Enrique S., eds., p. 666
- Viironen K. et al., 2011, A&A, 530, A107
- Vink J. S., Drew J. E., Steeghs D., Wright N. J., Martin E. L., Gänsicke B. T., Greimel R., Drake J., 2008, MNRAS, 387, 308
- 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
- Yu N.-P., Wang J.-J., 2012, Research in Astronomy and Astrophysics, 12, 651

## APPENDIX A: CATALOGUE FORMAT

Table A1: Definition of columns in the IPHAS DR2 source catalogue.

Column	Type	Unit	Description
name	string		Sexagesimal, equatorial position-based source name in the form: JHH-MMSS.ss+DDMMSS.s. The full naming convention for IPHAS DR2 sources has the form "IPHAS2 JHHMMSS.ss+DDMMSS.s", where "J" indicates the position is J2000. The "IPHAS2" prefix is not included in the column.
ra	double	degrees	J2000 Right Ascension with respect to the 2MASS PSC reference frame (which is consistent with ICRS to within 0.1 arcsec). The coordinate given is obtained from the astrometric measurement in the r'-band exposure. If the source is undetected in r', then the i' or H $\alpha$ -band coordinate is given.
dec	double	degrees	J2000 Declination. See comments above.
sourceID	string		Unique identification number of the detection. Identical to rDetectionID if the source was detected in the r-band; identical to iDetectionID or haDetectionID otherwise.
posErr	float	arcsec	Astrometric fit error (RMS). Be aware that the error might be significantly larger than the RMS near CCD edges.
l	double	degrees	Galactic longitude $\ell$ converted from ra/dec (IAU 1958 system).
b	double	degrees	Galactic latitude $b$ converted from ra/dec (IAU 1958 system).
mergedClass	short		Image classification flag based on all bands (1=galaxy, 0=noise, -1=star, -2=probableStar, -3=probableGalaxy, -9=saturated). Computed using the UKIDSS scheme.
mergedClassStat	float		Merged N(0,1) stellarness-of-profile statistic. Computed using the UKIDSS scheme.
pStar	float		Probability that the source is a star (value between 0 and 1).
pGalaxy	float		Probability that the source is a galaxy (value between 0 and 1).
pNoise	float		Probability that the source is noise (value between 0 and 1).
pSaturated	float		Probability that the source is saturated (value between 0 and 1).
rmi	float	mag	( $r - i$ ) colour, formed by subtracting columns r and i. Included in the catalogue for convenience only. To obtain the uncertainty, take the root of the sum of the squares of columns rErr and iErr.
rmha	float	mag	( $r - H\alpha$ ) colour, formed by subtracting columns r and ha. See comments above.
r	float	mag	Default r-band magnitude using a 2.3 arcsec diameter aperture. Calibrated in the Vega system.
rErr	float	mag	Uncertainty for r. Does not include systematic errors.
rPeakMag	float	mag	Alternative r-band magnitude derived from the peak pixel height (i.e. a 0.3x0.3 arcsec square aperture). Calibrated in the Vega system.
rPeakMagErr	float	mag	Uncertainty in rPeakMag. Does not include systematics.
rAperMag1	float	mag	Alternative r-band magnitude using a 1.2 arcsec diameter aperture. Calibrated in the Vega system.
rAperMag1err	float	mag	Uncertainty in rAperMag1. Does not include systematics.
rAperMag3	float	mag	Alternative r-band magnitude using a 3.3 arcsec diameter aperture. Calibrated in the Vega system.
rAperMag3err	float	mag	Uncertainty in rAperMag3. Does not include systematics.
rGauSig	float	pixels	RMS of axes of ellipse fit in r.
rEll	float		Ellipticity in the r-band.
rPA	float	degrees	Position angle in the r-band.
rClass	short		Discrete image classification flag (1=galaxy, 0=noise, -1=star, -2=probableStar, -3=probableGalaxy, -9=saturated).
rClassStat	float		N(0,1) stellarness-of-profile statistic.
rErrBits	short		Bitmask used to flag a bright neighbour (1), source blending (2) and saturation (8).
rMJD	double	days	Modified Julian Date at the start of the r-band exposure.
rSeeing	float	arcsec	Average Full Width at Half Maximum (FWHM) of stars in the same CCD frame.
rDetectionID	string		Unique identifier of the r-band detection in the format "#run-#ccd-number", i.e. composed of the INT telescope run number, the CCD number and a sequential source detection number.
rX	float	pixels	Pixel coordinate of the source in the r-band exposure, in the coordinate system of the CCD.
rY	float	pixels	Pixel coordinate of the source in the r-band exposure, in the coordinate system of the CCD.
i	float	mag	Default i-band magnitude using a 2.3 arcsec diameter aperture. Calibrated in the Vega system.
iErr	float	mag	Uncertainty for i. Does not include systematic errors.

**Table A1 – continued**

Column	Type	Unit	Description
iPeakMag	float	mag	Alternative i-band magnitude derived from the peak pixel height (i.e. a 0.3x0.3 arcsec square aperture). Calibrated in the Vega system.
iPeakMagErr	float	mag	Uncertainty in iPeakMag. Does not include systematics.
iAperMag1	float	mag	Alternative i-band magnitude using a 1.2 arcsec diameter aperture. Calibrated in the Vega system.
iAperMag1err	float	mag	Uncertainty in iAperMag1. Does not include systematics.
iAperMag3	float	mag	Alternative i-band magnitude using a 3.3 arcsec diameter aperture. Calibrated in the Vega system.
iAperMag3err	float	mag	Uncertainty in iAperMag3. Does not include systematics.
iGauSig	float	pixels	RMS of axes of ellipse fit.
iEll	float		Ellipticity.
iPA	float	degrees	Position angle.
iClass	short		Discrete image classification flag (1=galaxy, 0=noise, -1=star, -2=probableStar, -3=probableGalaxy, -9=saturated).
iClassStat	float		$N(0,1)$ stellarness-of-profile statistic.
iErrBits	short		Bitmask used to flag a bright neighbour (1), source blending (2) and saturation (8).
iMJD	double	days	Modified Julian Date at the start of the single-band exposure.
iSeeing	float	arcsec	Average Full Width at Half Maximum (FWHM) of stars in the same CCD frame.
iDetectionID	string		Unique identifier of the r-band detection in the format "#run-#ccd-number", i.e. composed of the INT telescope run number, the CCD number and a sequential source detection number.
iX	float	pixels	Pixel coordinate of the source, in the coordinate system of the CCD.
iY	float	pixels	Pixel coordinate of the source, in the coordinate system of the CCD.
iXi	float	arcsec	Position offset of the i-band detection relative to the ra column. The original i-band coordinates can be obtained by computing (ra+iXi/3600, dec+iEta/3600).
iEta	float	arcsec	Position offset of the i-band detection relative to the dec column. See comments above.
ha	float	mag	Default H-alpha magnitude using a 2.3 arcsec aperture. Calibrated in the Vega system.
haErr	float	mag	Uncertainty for ha. Does not include systematic errors.
haPeakMag	float	mag	Alternative H-alpha magnitude derived from the peak pixel height (i.e. a 0.3x0.3 arcsec square aperture). Calibrated in the Vega system.
haPeakMagErr	float	mag	Uncertainty in haPeakMag. Does not include systematics.
haAperMag1	float	mag	Alternative H-alpha magnitude using a 1.2 arcsec diameter aperture. Calibrated in the Vega system.
haAperMag1err	float	mag	Uncertainty in haAperMag1. Does not include systematics.
haAperMag3	float	mag	Alternative H-alpha magnitude using a 3.3 arcsec diameter aperture. Calibrated in the Vega system.
haAperMag3err	float	mag	Uncertainty in haAperMag3. Does not include systematics.
haGauSig	float	pixels	RMS of axes of ellipse fit.
haEll	float		Ellipticity
haPA	float	degrees	Position angle.
haClass	short		Discrete image classification flag (1=galaxy, 0=noise, -1=star, -2=probableStar, -3=probableGalaxy, -9=saturated).
haClassStat	float		$N(0,1)$ stellarness-of-profile statistic.
haErrBits	short		Bitmask used to flag a bright neighbour (1), source blending (2) and saturation (8).
haMJD	double	days	Modified Julian Date at the start of the single-band exposure.
haSeeing	float	arcsec	Average Full Width at Half Maximum (FWHM) of stars in the same CCD frame.
haDetectionID	string		Unique identifier of the r-band detection in the format "#run-#ccd-number", i.e. composed of the INT telescope run number, the CCD number and a sequential source detection number.
haX	float	pixels	Pixel coordinate of the source, in the coordinate system of the CCD.
haY	float	pixels	Pixel coordinate of the source, in the coordinate system of the CCD.
haXi	float	arcsec	Position offset of the H-alpha detection relative to the ra column. The original Ha-band coordinates can be obtained by computing (ra+haXi/3600, dec+haEta/3600).
haEta	float	arcsec	Position offset of the H-alpha relative to the ra column. See comments above.
brightNeighb	boolean		True if a very bright star is nearby (i.e. errBits $\geq 1$ ). This indicates that the source might be spurious, or the photometry unreliable.

**Table A1 – continued**

Column	Type	Unit	Description
deblend	boolean		True if the source was blended with a nearby neighbour due to crowding (i.e. errBits $\geq 2$ ). Although a deblending procedure is applied when measuring the photometry, the result may be unreliable (e.g. colours should not be trusted).
saturated	boolean		True if the source is saturated (i.e. peak pixel $> 55000$ counts) in one or more bands (i.e. errBits $\geq 8$ ). The photometry of saturated stars is affected by systematic errors.
errBits	short		Maximum value of (rErrBits, iErrBits, haErrBits).
nBands	short		Number of bands in which the source is detected (equals 1, 2 or 3).
reliable	boolean		True if: errBits $\leq 2$ & nBands == 3 & r > 13 & i > 12 & ha > 12.5 & rErr < 0.1 & iErr < 0.1 & haErr < 0.1 & (abs(r-rAperMag1) < 3*hypot(rErr,rAperMag1Err)+0.03) & (abs(i-iAperMag1) < 3*hypot(iErr,iAperMag1Err)+0.03) & (abs(ha-haAperMag1) < 3*hypot(haErr,haAperMag1Err)+0.03).
veryReliable	boolean		True if: reliable & pStar > 0.9 & errBits = 0.
fieldID	string		Human-readable IPHAS field number and observing run (e.g. 0001o_aug2003).
fieldGrade	string		Internal quality control score of the field. One of A, B, C or D.
night	integer		Night of the observation (YYYYMMDD). Refers to the UT date at the start of the night.
seeing	float	arcsec	Maximum value of (rSeeing, iSeeing, haSeeing).
ccd	short		CCD-chip number on the Wide Field Camera (WFC) of the Isaac Newton Telescope (INT). 1, 2, 3 or 4.
nObs	short		Number of repeat observations of this source in the survey.
sourceID2	string		SourceID of the object in the partner exposure (if obtained within 10 minutes of the primary detection).
fieldID2	string		FieldID of the partner detection (e.g. 0001o_aug2003).
r2	float	mag	r-band magnitude in the dithered partner field, i.e. the dithered repeat measurement obtained within 10 minutes (if available).
rErr2	float	mag	Uncertainty for r2.
i2	float	mag	i-band magnitude in the dithered partner field, i.e. the dithered repeat measurement obtained within 10 minutes (if available).
iErr2	float	mag	Uncertainty for i2.
ha2	float	mag	H-alpha magnitude in the dithered partner field, i.e. the dithered repeat measurement obtained within 10 minutes (if available).
haErr2	float	mag	Uncertainty for ha2.
errBits2	integer		Error bitmask for the partner detection. Used to flag a bright neighbour (1), source blending (2), saturation (8), vignetting (64), truncation (128) and bad pixels (32768).