

¹⁶Imperial College, London. ¹⁷University of Cardiff

The UKIDSS Galactic Plane Survey

P.W.Lucas¹, M.G.Hoare², A.C.Schroeder³, C.J.Davis⁴, A.Adamson⁴,
 A.Longmore⁵, R.M.Bandyopadhyay⁶, R. de Grijs⁷, M.Smith⁸, S.Mitchison⁸,
 A.Gosling⁹, A.Gaspar¹⁰, M.Coe¹¹, M.Tamura¹², Q.Parker¹³, M.Irwin¹⁴,
 N.Hambly¹⁵, D.Wyn Evans¹⁴, S.Hodgkin¹⁴, E.Gonzalez-Solares¹⁴, N.Cross¹⁵, M.Read¹⁵, S. W.

email: p.w.lucas@herts.ac.uk

²*School of Physics and Astronomy, University of Leeds, Leeds, LS2 9JT, UK.*

³*Department of Physics & Astronomy, University of Leicester, University Road, Leicester LE1 7RH, UK.*

⁴*Joint Astronomy Centre, 660 North A'ohoku Place, University Park, Hilo, HI 96720, USA.*

⁵*Astronomy Technology Centre, Royal Observatory, Blackford Hill, EH9 3HJ, Edinburgh, UK*

⁶*Department of Astronomy, University of Florida, 211 Bryant Space Science Center, Gainesville, FL 32611.*

⁷*Department of Physics & Astronomy, The University of Sheffield, Hicks Building, Hounsfield Road, Sheffield S3 7RH, UK.*

⁸*Centre for Astrophysics & Planetary Science, The University of Kent, Canterbury CT2 7NH, UK.*

⁹*Department of Physics, University of Oxford, Keble Road, Oxford, OX1 3RH, UK.*

¹⁰*Department of Astronomy/Steward Observatory, University of Arizona, 933 N Cherry Ave., Tucson AZ 85721-0065, USA.*

¹¹*School of Physics and Astronomy, Southampton University, Highfield, Southampton SO17 1BJ, UK.*

¹²*National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan.*

¹³*Department of Physics, Macquarie University, Sydney 2109, Australia.*

¹⁴*Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA UK.*

¹⁵*Scottish Universities' Physics Alliance (SUPA), Institute for Astronomy, School of Physics, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK.*

15 July 2007

ABSTRACT

The UKIDSS Galactic Plane Survey (GPS) is one of the five near infrared Public Legacy Surveys that are being undertaken by the UKIDSS consortium, using the Wide Field Camera on the United Kingdom Infrared Telescope. It is surveying 1868 deg² of the northern and equatorial Galactic plane at Galactic latitudes $-5^\circ < b < 5^\circ$ in the J, H and K filters and a ~ 300 deg² area of the Taurus-Auriga-Perseus molecular cloud complex in these three filters and the 2.12 μm (1-0) H₂ filter. It will provide data on $\sim 3 \times 10^9$ sources and will therefore be the largest Galactic database for the foreseeable future. Here we describe the properties of the existing 530 deg² dataset available in UKIDSS DR2 and provide a user's guide for its exploitation. We also present brief Demonstration Science results from DR2 and from the Science Verification programme. These results illustrate how GPS data will frequently be combined with data taken in other wavebands to produce scientific results. The Demonstration Science includes studies of (i) M17; (ii) the ρ Ophiuchi dark cloud; (iii) X-ray sources in the Galactic Centre; (iv) external galaxies in the zone of avoidance; (v) IPHAS-GPS optical-infrared spectrophotometric typing; and (vi) GPS-GLIMPSE near/mid-infrared characterisation of star formation regions.

Key words: survey; methods: data analysis; stars: formation; (stars:) circumstellar matter; Galaxy: stellar content

1 INTRODUCTION

The United Kingdom Infrared Deep Sky Survey (UKIDSS) is a suite of five public surveys with varying depth and area coverage which began in May 2005 and are expected to continue until 2012. These surveys all use the Wide Field Camera (WFCAM) on the United Kingdom Infrared Telescope (UKIRT). The design and the principal aims of the five

UKIDSS surveys are described by Lawrence et al. (2007), along with a brief example of data from each survey. The data are reduced in Cambridge, England using a dedicated software pipeline (Irwin et al., in prep) and are then transferred to the WFCAM Science Archive (hereafter the WSA, Hambly et al., in prep) in Edinburgh. There, the data in the

different passbands are merged and ingested into a sophisticated SQL database.

The UKIDSS data are released as SQL databases at intervals of six months to a year. The area coverage and data quality of these releases are described by Dye et al. (2006, the Early Data Release), Warren et al. (2007, the First Data Release) and Warren et al. (2007, the Second Data Release (non-refereed).) In addition, the pipeline processed images and their associated FITS binary source catalogues for individual passbands are released in smaller batches on a faster timescale, but with no quality control. The UKIDSS data are presently public to all professional astronomers in any ESO member state. World access to the various data releases will begin shortly, starting 18 months after each data release to ESO. Users can obtain a WSA account through their local Community Administrator at their home institution. If the institution has no UKIDSS community then they should contact “wsa-support@roe.ac.uk” via email.

The aim of this paper is to provide a guide to the GPS database in the WSA. We describe the properties of the data in detail, illustrating usage of the WSA tools and giving examples of how to combine GPS data with data from other wavebands for maximum scientific benefit. This should be regarded as essential reading for those wishing to exploit the data. The authors include the GPS implementation group, members of WSA staff and members of the Cambridge Astronomical Survey Unit (CASU) which reduces the data. Some of the information provided here is available in the several publications mentioned above but much of it is not. We aim to bring together the main GPS-specific aspects of the WSA in one publication. We note that neither the pipeline nor the WSA were designed explicitly for the GPS but rather for the UKIDSS surveys as a whole. Some compromises were made to produce an efficient database in a reasonable period of time and we describe their effects here.

The paper is structured as follows. In §2 we summarize the survey design, including recent changes from the original plan described by Lawrence et al. In §3 we show how to use the parameters available in the GPS database to best effect and describe the strengths and weaknesses of the data that are presently available in DR2. In §4 we describe the properties of the GPS dataset (depth and stellar populations) as a function of Galactic location and source confusion. In §5 we present some brief Demonstration Science results from the Science Verification programme and from DR2 that are designed to illustrate the variety of uses of GPS data, in combination with data from other wavebands.

2 SUMMARY OF GPS STRATEGY, AREA AND INTEGRATION TIME

The GPS is designed to cover the entire northern and equatorial Galactic plane that is accessible to UKIRT in a 10 degree wide band around the sky. There is also a narrower southern extension to the Galactic Centre and a 300 deg² survey of the Taurus-Auriga-Perseus (TAP) molecular cloud complex, which lies off the plane and follows the contours of the ¹³CO gas mapped by Ungerechts & Thaddeus (1986). In Galactic coordinates, the in-plane region is: $141 < l < 230$, $-5 < b < 5$; $15 < l < 107$, $-5 < b < 5$; and $-2 < l < 15$, $-2 < b < 2$. The section of the plane at $l=107$ to 141 is not

covered because it lies north of Dec= $+60^\circ$, a region of the sky that UKIRT cannot reach due to fundamental features of the telescope design.

The total on-source integration times are 80s, 80s and 40s in the J, H and K bandpasses respectively. Data in the three filters are generally collected within an interval of 20 minutes for any point on the sky, in order to minimise the effects of any photometric variability on source colours and derived extinction. Exceptions can occur (beginning in DR2) when data which failed quality control in only 1 or 2 filters were repeated in only those filters. In these cases there will always be at least 2 near contemporaneous fluxes which can be used to estimate intrinsic source colours and extinction.

The survey depth is spatially variable owing to source confusion in the more crowded parts of the plane (see §4.x). Hence, in many regions of the plane the GPS represents the deepest survey possible for ground based, seeing limited observations.

The sky coverage in the various data releases so far is described in the references given in §1. We note that the most recent release, DR2, is the first to include a sufficiently large area of three band data to be useful for large scale Galactic studies. In particular, DR2 covers 50% of the Spitzer GLIMPSE and GLIMPSE-II mid-IR survey regions at longitudes $-2 < l < 70$, $-1 < b < 1$.

2.1 Cuts to the GPS Programme

The original GPS design also included two additional epochs of data in the K band only, which were intended to provide a long time baseline for proper motions and to detect high amplitude variable stars. Following a review of the UKIDSS project in early 2007, further progress on these K-only scans of the survey area have been deferred until at least 2010. Some 500 deg² of the plane have already been observed in the K-only mode during conditions of thin cirrus cloud (avoiding regions which have been observed in the main three band survey). It remains very possible that one of the full K-only scans will be completed within the seven year timeframe that was originally planned for UKIDSS (2005-2012). It should be noted that all UKIDSS photometry is calibrated using the numerous 2MASS sources in every field (see §3.3; Hewett et al. 2006; Irwin et al., in prep), so there is only a very small loss of photometric precision (at the millimagnitude level) even when observing through thin cirrus. The main JHK survey is observed in photometric conditions however, in part to ensure maximum depth.

The TAP molecular cloud component of the survey also includes observations in the 2.12 μm (1-0) S(1) H₂ filter. It was originally planned to observe this 300 deg² region three times: once in the J, H, K and H₂ filters and twice more in the K and H₂ filters, in order to measure the velocities and accelerations of protostellar jets. However this part of the survey was cancelled in the 2007 review of UKIDSS, at which point ~ 150 deg² of the region had been observed. Details will be provided in a future UKIDSS paper on the DR3 release. Nonetheless, the remainder of the TAP region will be observed in the Z, Y, J, H and K filters as part of the ongoing UKIDSS Galactic Clusters Survey, which included the same region. It is hoped that the full TAP region will ultimately be observed at least once with WFCAM at 2.12 μm .

3 GUIDE TO THE WFCAM SCIENCE ARCHIVE FOR GPS SCIENCE

3.1 Fundamentals of the GPS data, archive structure and useage

In this guide to the GPS we do not assume that the reader has previously used the WSA, nor that he has any knowledge of SQL (the Structured Query Language which is used to construct and interrogate the WSA). This guide is not exhaustive, partly because there is extensive documentation at the WSA and CASU websites and partly for reason of brevity.

3.1.1 Pawprints and Multiframe: the basic unit of data

WFCAM has four spatially separated 2048×2048 HAWAII 2 infrared array detectors laid out in a 2×2 square pattern. Each array covers a 13.65 arcminute field of view and the gaps between them are 12.83 arcmin, or 94% of a detector width. A single observation therefore covers a partially filled area of sky known as a **Multiframe** (in WSA parlance) or a **pawprint** (in UKIRT parlance). In the GPS, a series of 4 pawprints with offsets of 13.24 arcmin in RA and Dec is used to produced a filled **tile** with an area of 0.75 deg^2 . The Multiframe (not the tile) is the fundamental unit around which the WSA is designed. Every Multiframe has a reference number in the WSA. By contrast, the tiles are not catalogued. The WFCAM arrays are always oriented approximately north-south (to within $\sim 1^\circ$, so the GPS tiles the Plane in strips of constant Declination, always slightly overfilling the desired Galactic latitude boundaries with array corners and edges.

The term Multiframe is also applied to a coadded sequence of observations which include spatial offsets of a few arcsec to aid removal of bad pixels and fully sample the spatial resolution of the telescope. Every pipeline processed pawprint image is stored in the WSA as a multiframe (both individual integrations and a coadded series of observations). However, only the coadded images have associated source catalogues.

3.1.2 GPS observation sequence

All GPS data are taken with a 2×2 “microstepping” sequence in order to fully sample the spatial resolution of the telescope, which is frequently < 2 pixels Full Width Half Maximum, i.e. < 0.8 arcsec. The microsteps used in the GPS are spatial shifts of 4.6 arcsec, which is approximately 11.5 pixels (depending on telescope focus and the location within the focal plane). This “N+0.5” microstepping allows the four separate images to be interleaved in order to produce images with a 0.2 arcsec pixel scale. After going through the microsteps, the telescope is then moved through 3.2 arcsec (8 pixels) in RA and/or Dec before repeating the microstep sequence. This jitter procedure provides independent pixels to aid the removal of any transient bad pixels that were not included in the bad pixel mask. The number of jitters changed between the EDR and DR1, but is now settled at four jitters for J and H band data, and two jitters for K band and H_2 data.

3.1.3 GPS images in the WSA

The multiframe images are stored as FITS images with a header unit and four image extensions, one for each spatially separated array. There are three main types of reduced GPS multiframe image. (1) Individual integrations, which are listed as *normal* multiframe and have a pixel scale of 0.4 arcsec. (2) Interleaved images, listed as *leav* multiframe, which are formed from four images taken in the microstepping sequence and have a 0.2 arcsec pixel scale. (3) Final coadded images (also with a 0.2 arcsec pixel scale) that are formed by coadding two or more interleaved images and are listed as *leavstack* multiframe in the WSA. For most purposes GPS users will only need to use the *leavstack* multiframe.

The user can make a list of the GPS multiframe images for download, and view high quality jpeg preview images, by selecting “Archive Listing” at the top level of the WSA and then selecting the GPS from the Programme drop down menu. The user will normally use the most recent UKIDSS release by default (presently “UKIDSSDR2PLUS”), but there are also options to use either the earlier releases (which are a subset of the most recent release but can be searched slightly more quickly) or “WSA” in order to include images which failed quality control but may still be useful for some scientific purposes. The three types of reduced science image described above will be returned if the user uses the default settings of “object” in the Observation Type drop down menu and “all” in the Frame Type drop down menu. If the latter menu is set to “stack” then only “*leavstack*” images will be returned. The desired images can be then selected by date, by filter and by Equatorial coordinates (FK5 system, defined by a rectangular box). Two additional types of GPS image are the H_2 -K *difference* images provided in DR2 for the off-plane TAP region and *Confidence* images, which exist for all *leav* and *leavstack* images and can be used to check for defects in the data. *Confidence* images will only be returned if the Observation Type is changed from “object” to “confidence”. Pixels in the *Confidence* images have a median value of 100. Lower values correspond to either reduced sensitivity measured in the flat fields, hot pixels detected in the darks or saturated pixels. Other images types from the reduction process, such as *Sky* images and *Dark* images can also be obtained with the appropriate setting of the Observation Type. *Sky* images are generally constructed using a sequence of frames taken throughout the night. These may be useful in order to assess the quality of the background subtraction in Galactic plane regions with bright nebulosity.

N.B. The coordinates returned by the Archive Listing for a multiframe are those of the centre of the WFCAM focal plane, which lies between the four spatially separated arrays and is therefore a region with no data! Hence if the full array image nearest a particular location is desired, it is necessary to either use the Image tools described below to find the **Multiframe** number or else search for all images in a box 0.8° wide, centred on the target location, to find the desired one.

More user friendly image access is provided by the “GetImage”, “MultiGetImage” and Region Search tools at the top level of the WSA. The “GetImage” tool allows the user to find and download cut-out images in all available

filters, centred on the desired Galactic or Equatorial coordinates. The image size is limited by the user-specified dimensions or by the boundaries of the 13.6 arcminute array field of view. There may be more than one image in each filter if the coordinates lie in the small overlap region at the edge of an array, or if the field has been observed more than once (though the latter should not occur until UKIDSS DR4 for the GPS.) The “MultiGetimage” tool is similar but it allows a large number of images to be downloaded in a single filter, each centred around a list of Equatorial coordinates provided by the user. The Region Search is described below.

3.1.4 GPS catalogues

A list of FITS binary source catalogues associated with individual coadded multiframe images (*leavstacks*) is provided along with the image lists described above in an Archive Listing. They may be viewed with freely available software such as TOPCAT or FV (FitsViewer). These catalogues contain only single filter data and source fluxes are given in data numbers (normalised to the exposure time of individual integrations) rather than magnitudes. The catalogues include the aperture corrections, and the FITS image headers include the photometric zero points and integration times, so it is straightforward to calculate magnitudes or to conduct additional photometry, eg. on very faint sources which may have been missed by the source detection algorithm and are therefore not in the SQL tables described below. N.B. the FITS binary catalogues have less precise R.A. and Dec coordinates than the SQL source catalogues described below, owing to a limitation on the number of data bits which corresponds to a precision of ~ 0.36 arcsec. Details of the columns in the FITS binary catalogues (which provide additional information relevant to the SQL source catalogues) are available from the CASU website (Irwin et al.2007) and in Irwin et al.(in prep).

Most users will wish to use the SQL source catalogues to search for astronomical objects. **gpsSource** is the most commonly used catalogue. It is the merged catalogue containing most of the important data for all sources detected in at least one of the J, H and K passbands. The pairing scheme starts with the shortest available wavelength (usually J band) and searches for counterparts in the H band and then the K band that are located < 1 arcsec from the J band detection. Sources are only merged as a single entry in *gpsSource* if each is the closest match to the other at all wavelengths. The other main table, **gpsDetection**, contains all the information for all sources detected in every passband but with no association between detections of the same object in different filters. While **gpsSource** is more convenient to use, the caveats are: (i) that it includes only a subset of the photometric aperture sizes available in **gps-Detection** or the FITS binary catalogues; and (ii) that in future releases (DR3 onward) most of the extragalactic flux estimates such as Petrosian magnitude will be removed. The details of how to write an SQL query to extract these missing attributes from the **gpsDetection** table are given in the “SQL Cookbook” at the top level of the WSA.

The available columns or “attributes” in each table are seen when conducting a Menu Query (see below), which optionally provides a one line description of each attribute. We recommend that the user reads the more detailed en-

tries for the **gpsSource** and **gpsDetection** parameters in the Schema Browser (accessible from the WSA top level) under “WSA UKIDSS” and then “Tables”.

The GPS SQL catalogues can be searched using the Region Search, the Menu Query (recommended for new users) or the Freeform SQL page, all of which are accessible from the top level of the WSA. When using these tools, users of the DR2 release and subsequent releases have an option (in the drop down menu of tables) to view a subset of **gpsSource** which includes only areas which are complete in all three filters. This can slightly improve search speeds where the desired area is known to be complete (eg. from an Archive Listing). If previous data releases are being searched, the user can similarly search a dataset with full filter coverage by selecting eg. the DR1 database, as opposed to the larger DR1PLUS database.

The simplest way to search **gpsSource** table is with the Region Search. This enables a simple search of a circular area around the desired Equatorial or Galactic coordinates and it returns all information in the table for every object therein. In addition, the sources listed can be clicked on to provide jpeg and FITS format cut-out images around their location, in all available filters (similar to the GetImage results).

More flexible and sophisticated queries can be made with the Menu Query or a FreeForm SQL query. Menu Queries are constructed by clicking boxes. They return only the desired attributes and restrict their values (eg. magnitudes, colours, coordinates etc.) in order to refine a search. This is especially important for the GPS because a maximum of only 15 million data entries can be returned in a single query (eg. 1.5 million sources with 10 data attributes each), which can quickly be exceeded when searching more than a square degree in crowded fields. The **gpsSource** attributes that are highlighted in the Menu Query list have their own indices in the table, which means that searches based on these parameters will proceed much more rapidly than those which do not. Hence searches based on eg. specified coordinates (Equatorial or Galactic) and a range of magnitudes will proceed rapidly if the preferred photometric aperture (APERMAG3, see below) is used, but will take much longer if a different aperture size is used. Again, this is a serious issue since the size of the GPS catalogues means that searches can often time out (the maximum duration is presently 80 minutes) when searches are based on non-highlighted attributes. Merely including non-highlighted attributes among the data to be returned (as opposed to specifying a range of values) does not cause this problem.

FreeForm SQL queries can perform the same tasks as a Menu Query and also manipulate the data returned from the table or operate on two or more tables (eg. GPS data and Spitzer GLIMPSE data) to produce more complicated outputs. Menu Queries are transformed into ascii format SQL queries before processing. A useful feature of the WSA is that this SQL query can be viewed using the “Show and Edit SQL” button during a Menu Query. This enables the user to learn to perform basic FreeForm SQL queries with no prior knowledge of the language. The SQL Cookbook provides further training.

3.1.5 Simultaneous multi-aperture fitting photometry and *AperMag3*

The WFCAM pipeline performs aperture photometry, as opposed to profile fitting photometry. However, any overlapping apertures are simultaneously fitted and the fluxes in areas of overlap are assigned to each target in proportion to the total flux within each aperture. This generally produces good photometry even in the crowded GPS fields (see §3.2), though there will be cases where the photometry is imperfect, such as faint sources on the wings of a bright or saturated star. We note that profile fitting photometry may be included in the final processing of all UKIDSS data when the survey is finished, but tests with an algorithm designed for the irregular image profiles often generated by ground based microstepped data have shown no significant gain in precision.

Photometry is output for a series of circular apertures whose diameters have a ratio of $\sqrt{2}$. The smallest aperture, *Apermag1*, has a diameter of 1 arcsec. The default aperture that is used for GPS point source colours is *Apermag3*, which has a diameter of 2 arcsec. Only *Apermag3*, *Apermag4* and *Apermag6* and the much larger *Hallmag* aperture are available in **gpsSource** in the existing data releases (though all apertures are available in **gpsDetection**). We recommend that *Apermag3* is used for all GPS science unless there is good reason to desire a smaller or larger aperture. For future UKIDSS releases it is planned to change the available apertures in **gpsSource** to *Apermag2*, *Apermag3* and *Apermag4*, with the rationale that the slightly smaller aperture may be useful in the most crowded fields and the slightly larger aperture may yield more precise photometry in the off-plane TAP region, where relatively poor seeing is accepted.

3.1.6 Limitations and unusual features of the GPS data

Here we describe some peculiar features of the GPS data which the user may notice but whose origin may not be immediately obvious.

- In many images the profiles of stars in the *leav* and *leavstack* images may appear to be slightly “speckled” with departures from the smooth monotonic image profiles usually seen in astronomical data. This is due to the changes in the natural seeing during the microstepping procedure referred to above, which can cause adjacent rows and columns to show significant differences in flux. Comparison of fields with and without this effect have shown that it has no significant effect on photometric precision.

- Electronic cross talk between adjacent array read out channels causes the appearance of spurious sources located at multiples of 128 pixels (51.2 arcsec) away from all bright stars along the axis of array readout (which is oriented differently for each quadrant of an array). This is a general feature of Hawaii 2 arrays that is most serious for pixels which are close to saturation (> 30000) DN in the WSA images. More recent versions of the WFCAM reduction pipeline (applied to data with 2006 observation dates which were included in DR2) include a cross-talk removal procedure which has reduced the problem. However, the procedure is imperfect so some artifacts remain. This is particularly so in regions

with spatially variable background, where a flaw in the procedure actually generated additional cross talk artifacts that do not always lie 128N pixels from a bright star. (This error caused some images to be removed from DR2, including several H band images near the Galactic Centre. The error will be fixed for future releases and the lost images will be recovered). The archive also addresses the problem from the DR2 release onward, but in the catalogues rather than the images. Possible cross talk artifacts are flagged in the data given in the **ppErrbits** columns in **gpsSource** and **gpsDetection**. This identifies almost all cross talk artifacts at 128N pixels a bright star by using the 2MASS Point Source Catalogue (? check this with Nigel). Usage of the **ppErrbits** flags to remove suspect data is described in §3.2.

The brighter artifacts can be seen to be spurious since they have positive counts on one side of the profile and negative counts on the other (relative to the sky level). This is because the cross talk artifacts are produced as the Nth spatial derivative of the profile of the bright star that generated them, in the readout direction, where N is the relevant multiple of 128 pixels. Fainter stars can sometimes produce artifacts which look more like real stars (and generally have similar colours). It is advisable to guard against this by checking the images for bright stars near objects of interest. Over-saturated stars produce larger, obviously non-stellar artifacts with a ring-like structure.

- Saturated stars have $\gtrsim 35000$ counts in *leavstack* images, but the centre of an oversaturated star has a lower count level (often below the sky background). This is another general feature of data from Hawaii 2 arrays. Saturated stars exist in most GPS fields. They are usually included in the source catalogues but the fluxes given are not meaningful. Again, the **ppErrbits** column in the catalogues can be used to efficiently remove these bad data. Heavily oversaturated stars cause multiple spurious detections in each filter (and in **gpsSource**), since the saturated region is larger than the size of the stellar image profile used by the pipeline source detection algorithm. These are often listed in the archive as resolved sources rather than stars (**mergedclass=+1**).

- The photometric errors given in magnitudes and colours are only notional errors based on source counts in the apertures, the FWHM of the image and the background flux level. A plot of errors vs. magnitude yields a smooth curve for any given field. The WSA is similar to the 2MASS Point Source Catalogue in this respect. Errors specific to individual sources, eg. bad pixels and the presence of adjacent stars or cross talk artifacts, are not included in the error budget. Hence users should always inspect the image data (and perhaps also the *Confidence Images* to be sure that the error in a particular source is not underestimated.

- Nebulosity can also lead to photometric errors, since the sky background is evaluated with a relatively coarse grid in each field. More seriously, the source detection algorithm fails in regions of very bright nebulosity, since the nebula is interpreted as a single extended source. At present this effect has been seen only in parts of the M17 star formation region (see § x.x) which is one of the largest regions of bright nebulosity in the sky. In such regions, it is necessary for the user to do their own photometry (see §x.x)

3.2 Optimising the results of GPS catalogue searches

Here we present three different choices of catalogue parameters that are likely to be useful to generate different types of source list, ranging from a complete but unreliable selection of the stellar sources in a region to a highly reliable but seriously incomplete selection.

Generating a complete list of unique Galactic sources is a little more complicated than simply returning all sources in a region of the sky. This is for two reasons. (1) **gpsSource** contains sources classified as “noise”, via the attribute “*mergedClass* = 0”, in addition to the more numerous real sources which are variously classified as stars (*mergedClass* = -1), probable stars (*mergedClass* = -2), galaxies (*mergedClass* = +1) and probable galaxies (*mergedClass* = -3), on the basis of the image profile. We recommend that sources classified as noise be excluded from GPS searches, since most of them are not real sources, and they would inflate the star counts in any given field by ~10%. In general this will lead to a more accurate star count in any given field than a search on all types of source, at the price of excluding only a very small number of real sources. By contrast, it is necessary to include galaxies and probable galaxies in a complete search, since most of these are marginally resolved pairs of stars which have been misclassified in **gpsSource**. (2) Sources located in the regions of overlap between different arrays in a WFCAM tile are duplicated in the WSA. These duplicates can be identified and removed using the *PriOrSec* attribute, which describes Primary or Secondary detections (see Hambly et al., in prep). This consideration also affects ~10% of sources.

An illustrative SQL query to return a selection of useful attributes for the complete list of real unique sources located near $l = 31$, $b = 0$ is given below. Comment lines in the query begin with “/*” and end with “*/”.

```
select sourceID, ra, dec, l, b, jmhPnt,
jmhPntErr, hmk_1Pnt, hmk_1PntErr, mergedClass,
pStar, jAperMag3, jAperMag3Err, hAperMag3,
hAperMag3Err, k_1AperMag3, k_1AperMag3Err from
gpsSource
where l between 30.9 and 31.1
and b between -0.1 and 0.1
/* Exclude detections classified as noise*/
and mergedClass != 0
/* Exclude multiple detections of the same
source */
and (PriOrSec=0 or PriOrSec=framesetID)
```

It should be noted that this query must be typed or copied into the Freeform SQL interface, since the last line cannot be entered with the simple mathematical operators available in the Menu Query interface. The two colour diagram for the complete search at these coordinates is shown in Figure 1(a).

It is frequently desirable to restrict the sample to sources with small photometric errors, in order to identify the main stellar populations more easily in two colour and colour magnitude diagrams. A restricted search can be done entirely with the WSA but since the user will often wish to vary the search parameters while investigat-

ing the data it is often preferable to download the results of a complete search and then alter the parameters interactively. This can be done by loading the data into a suitable software package. We strongly recommend TOPCAT, a user friendly software package designed to manipulate and display data tables written in a variety of Virtual Observatory and text formats.¹ TOPCAT is available from www.starlink.ac.uk/topcat.

The following SQL query returns all the parameters needed to construct a more reliable but less complete photometric sample:

```
select sourceID, ra, dec, l, b, jmhPnt,
jmhPntErr, hmk_1Pnt, hmk_1PntErr, mergedClass,
pStar, jAperMag3, jAperMag3Err, jEll, jppErr-
rBits, jXi, jEta, hAperMag3, hAperMag3Err,
hEll, hppErrBits, hXi, hEta, k_1AperMag3,
k_1AperMag3Err, k_1Ell, k_1ppErrBits, k_1Xi,
k_1Eta from gpsSource
where l between 30.9 and 31.1
and b between -0.1 and 0.1
/* Exclude noise detections */ and merged-
Class != 0
/* Exclude multiple detections of the same
source */
and (PriOrSec=0 or PriOrSec=framesetID)
```

Figure 1(b) is the two colour diagram for the subset of this download with errors < 0.05 mag on each axis, i.e. $|jmhPntErr| < 0.05$ and $|hmk_1PntErr| < 0.05$. The majority of the sources are located in a band running from the lower left to the upper right, caused by the reddening of the main sequence and giant star populations. However, a large number of photometric outliers are apparent and some sources lie outside the boundaries of the plot. A smaller subsample with better photometry and > 50% completeness can be selected with the following additional cuts:

```
pstar > 0.99 && jppErrbits < 256 && hppErrbits
< 256 && k_1ppErrbits < 256
```

Here ‘&&’ is the logical operator ‘AND’ written in TOPCAT format. (We note that column numbers can be used in TOPCAT INSTEAD OF PARAMETER NAMES FOR BREVITY). THE “pstar” PARAMETER IS THE WSA ESTIMATE OF THE PROBABILITY THAT THE SOURCE IS A STAR, AS OPPOSED TO A GALAXY, OR NOISE. IN PRACTICE THIS SELECTION MAINLY REMOVES CLOSE STELLAR PAIRS WHICH ARE RESOLVED IN ONLY ONE OR TWO OF THE THREE PASSBANDS. THE “ppErrbits” PARAMETERS IN EACH WAVEBAND CAN BE USED TO IDENTIFY SOURCES WHICH ARE EITHER CROSS TALK ARTIFACTS (SEE X.X.X) OR STARS WITH LESS RELIABLE PHOTOMETRY DUE TO SATURATION, DEBLENDING, BAD PIXELS OR LOCATION AT THE EXTREME EDGE OF THE DETECTOR. WE NOTE THAT THE ALTERNATIVE “psat” PARAMETER FOR SATURATION PROBABILITY IS NOT FUNCTIONAL AT PRESENT IN THE WSA. THE CORRESPONDING PLOT IN FIGURE 1(C) IS MUCH CLEANER:

¹

MOST OF THE UNRESOLVED STELLAR PAIRS HAVE BEEN REMOVED BY THE `pstar` SELECTION AND NUMEROUS SATURATED STARS AND THEIR ASSOCIATED SPURIOUS MULTIPLE DETECTIONS HAVE BEEN REMOVED BY THE `ppErrbits` SELECTION. IN THIS CASE THE COMPLETENESS OF THE SAMPLE IS 72%, RELATIVE TO FIGURE 1(B). THE VERTICAL SEPARATION BETWEEN THE REDDENED POPULATIONS OF DWARFS AND GIANT IS NOW READILY APPARENT AT $(J-H) > 1.75$. THE GIANTS HAVE ~ 0.25 MAG LARGER $(J-H)$ COLOURS THAN DWARFS DUE TO THE EFFECT OF THEIR LOW SURFACE GRAVITIES ON H^- OPACITY.

A SAMPLE WITH STILL MORE RELIABLE PHOTOMETRY, BUT WITH LOWER COMPLETENESS, CAN BE SELECTED AS FOLLOWS:

```
pstar > 0.99 && jppErrbits < 256 && hppErrbits
< 256 && k_1ppErrbits < 256 &&
sqrt(hXi2 + hEta2) < 0.3 && sqrt(k_1Xi2 +
k_1Eta2) < 0.3 && jEll < 0.2 && hEll < 0.2 &&
k_1Ell < 0.2
```

The completeness of this selection is also typically $> 50\%$ in the less crowded parts of the GPS but is usually only 25-50% for fields near the mid-plane in the first quadrant. The 'Ell' parameters give the ellipticity, which can be used to remove unresolved stellar pairs in a similar manner to `pstar`. The 'Xi' and 'Eta' parameters are the distances in arcseconds of each detection from the shortest wavelength detection, in RA and Dec respectively. The maximum pair matching radius is 1 arcsec for the GPS, but since the typical separation of genuine matches is $\lesssim 0.2$ arcsec, these parameters can also be used to eliminate unresolved binaries and other false matches. Figure 1(d) shows the corresponding plot, which is now only 34% complete relative to Figure 1(b). Only a few sources with poor photometry now remain.

4 PROPERTIES OF THE DATA AS A FUNCTION OF GALACTIC COORDINATES

In this section we present magnitude histograms, colour magnitude diagrams and two colour diagrams for representative fields throughout the survey region. We discuss the survey depth as a function of Galactic coordinates and then briefly describe the nature of the main stellar populations that are detected. We then compare the data with the predictions of the Besançon Galactic stellar population model.

Figure 2, 3, 4, 5 and 6 shows the results for 12 GPS fields. The fields at $l < 90^\circ$ are square boxes that are 0.2 degrees on a side in l and b . The fields at $l > 90^\circ$ are square boxes that are 0.5 degrees on a side in l and b , in order to provide similar sized samples to reveal the nature of the populations in these less crowded regions. All fields include data from at least 2 separate WFCAM pawprints and usually more. Any errors in global photometric calibration would add to the scatter in these plots. However the quality of the 2MASS-based photometric calibration is such that we see no obvious increase in scatter compared to data from individual WFCAM arrays.

4.1 Survey sensitivity and the effect of source confusion

Figure 2 shows the magnitude histograms. Visual inspection of the images shows that the fields at $l > 90^\circ$ have less source confusion than the other fields and source confusion is insignificant for the two fields near $l=170^\circ$. The corresponding histograms represent the maximum intrinsic depth of the survey. We see that the modal depths in uncrowded fields are approximately $K=17.75$ to 18.0 , $H=18.5$ to 18.75 , $J=19.4$ to 19.65 . The variation is due to the variable seeing, sky brightness, thermal background and atmospheric transparency during the survey. By visually extrapolating the histograms and assuming that the decline at faint magnitudes is entirely due to limited sensitivity we estimate that the typical 90% completeness limits in uncrowded fields are located at $K=18.0$, $H=18.75$, $J=19.5$, with uncertainties of about 0.2 magnitudes. Individual fields in the outer galaxy should always be within 0.3 magnitudes of these typical values, since only a modest range of observing conditions (seeing and transparency) is permitted. We do not attempt a more precise quantification of the mode and spread of the sensitivity at this stage since there is limited data for the outer Galaxy in DR2, and the sensitivity in future data releases is likely to change slightly. The seeing conditions for observations at $l > 107^\circ$ were relaxed from 0.8 arcsec to 1.0 arcsec from February 2007 in order to speed the survey progress.

It is clear that the survey is much less sensitive near the Galactic centre than in uncrowded fields. However the 90% completeness limits and modal magnitudes are usually degraded by no more than 1 magnitude in fields at $l > 30^\circ$, even in the mid-plane at $b=0^\circ$.

The effects of extinction on the histograms are apparent in the change in the relative numbers of J, H and K band detections in different parts of the Galaxy. The K band source density is highest in the fields at $l < 30^\circ$. At $l > 30^\circ$ the H band source density is higher (at least for the fields at $|b| \leq 1^\circ$ for which multi-band data presently exist). The source density in the J band becomes comparable to the H band at $l > 90^\circ$. *N.B. the overdensity of sources in the H band at $l=171$ appears to be due to multiple detections in the overlap regions, which will have to be fixed when the PriOrSec bug is fixed.*

4.2 Stellar populations

The two colour diagrams and colour magnitude diagrams in Figures 3-6 allow us to clearly identify the main stellar populations detected in the GPS. They can also be used to detect changes in extinction as a function of distance along different lines of sight. The two colour diagrams all show samples selected using the least complete, most reliable selection described in section 3, with errors < 0.05 mag in the colours on each axis, in order to distinguish the closely spaced stellar populations. Including fainter sources with less precise photometry simply increases the scatter in the diagrams, as shown in Figure 1(a), and extends the dwarf and giant sequences to slightly higher reddening values at the upper right.

By contrast, the colour magnitude diagrams show all the three band detections within the chosen axis ranges.

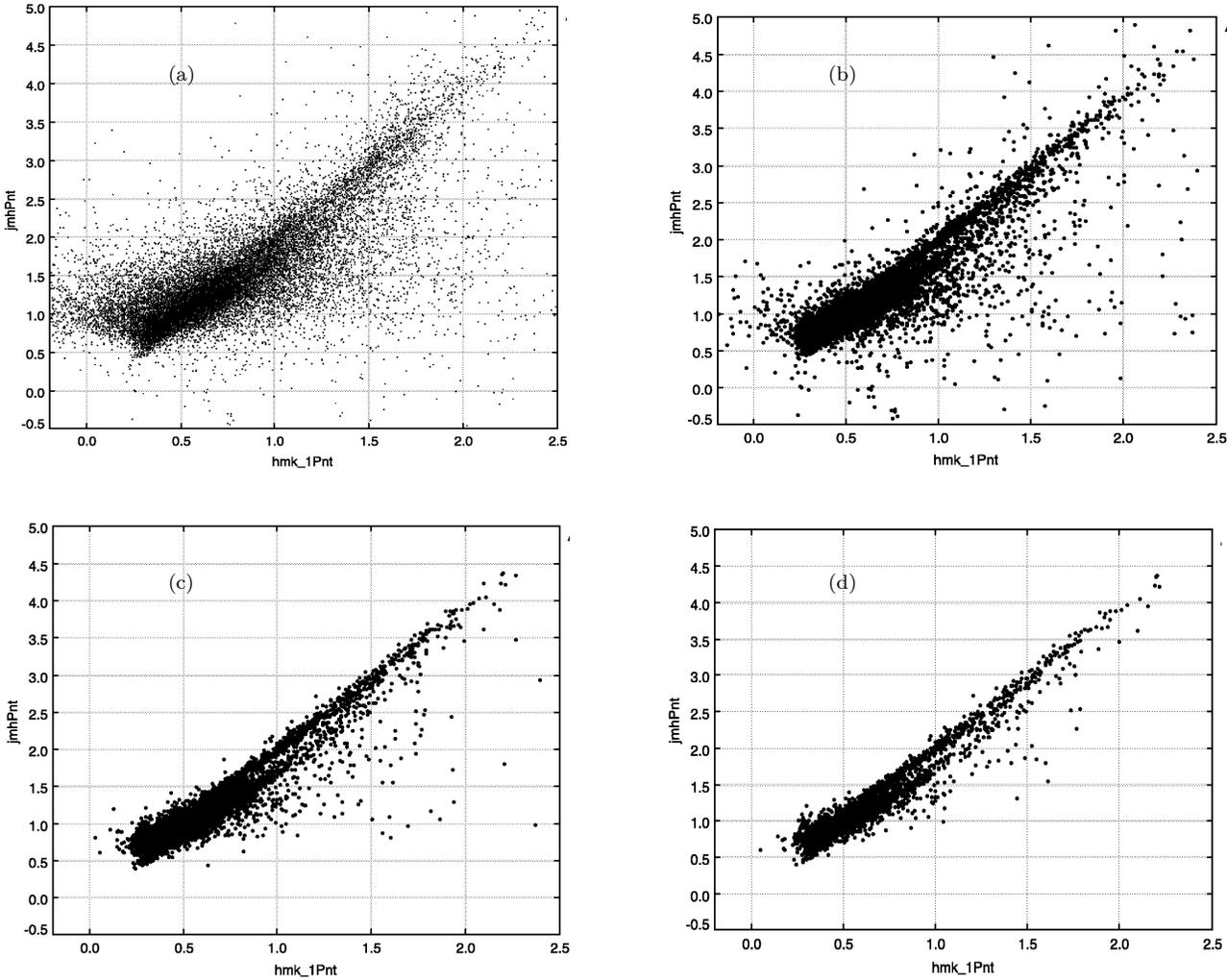


Figure 1. (J-H) vs (H-K) two colour diagrams for a 0.04 deg^2 area at $(l, b)=(31, 0)$ illustrate the trade off between completeness and reliability. (a) Complete selection of 51780 sources, excluding only “noise” sources. (b) Selection of the 8183 sources with errors < 0.05 mag in each colour. (c) Selection of 5914 sources with colour errors < 0.05 mag and more reliable photometry. (d) Selection of 2787 sources with colour errors < 0.05 mag and the most reliable photometry.

Sources with more reliable photometry are shown in black, while sources with less precise photometry are shown in blue (appearing as a lighter shade in the paper version of the journal). Sources are put in the “less precise” selection if they fail to satisfy any of the following criteria. (1) Photometric uncertainties < 0.2 mag in the two bands used in each diagram. The corresponding less precise sources are faint objects lying in a group at the bottom of each diagram. (2) $\text{pperrbits} < 256$. Most sources failing this criterion do so because they are flagged as close to saturation, and therefore lie in a group at the top of each diagram. Inspection of the diagrams shows that they nonetheless often play a vital role in defining the different stellar populations, indicating that the errors in their photometry need not be extreme. (3) $\text{pstar} > 0.9$. Sources only fail this criterion if they appear significantly non-stellar in at least one of the three passbands, usually because they are part of a marginally resolved stellar pair. Such sources are evenly spread among the sources with more reliable photometry.

The colour magnitude diagrams all show at least two principal sequences of stars, corresponding to luminosity class V dwarfs at the left of the diagrams and class III giants to the right. The majority of the giants are K2III giants, known as red clump giants. These occupy a sharp and well defined peak in the absolute magnitude function of Class III giants and are therefore a useful tracer of stellar distance, extinction and population density, eg. Lopez-Corredoira et al.(2001). As noted in section 3.2, the two colour diagrams also allow us to distinguish the dwarfs from the giants in most cases, since the giants lie 0.25-0.3 mag above the dwarfs in (J-H).

The two colour diagrams show an obvious trend of decreasing extinction with increasing Galactic longitude and latitude. The plot for $(l, b)=(171.1, 4.75)$ shows a field with negligible reddening. Nearly all the sources in this diagram are dwarfs, according to the Besançon model. The densest cluster of sources near $(J-H)=0.4$ are F types. In the adjacent plot for $(l, b)=(173.65, 0.65)$, we see that the F types

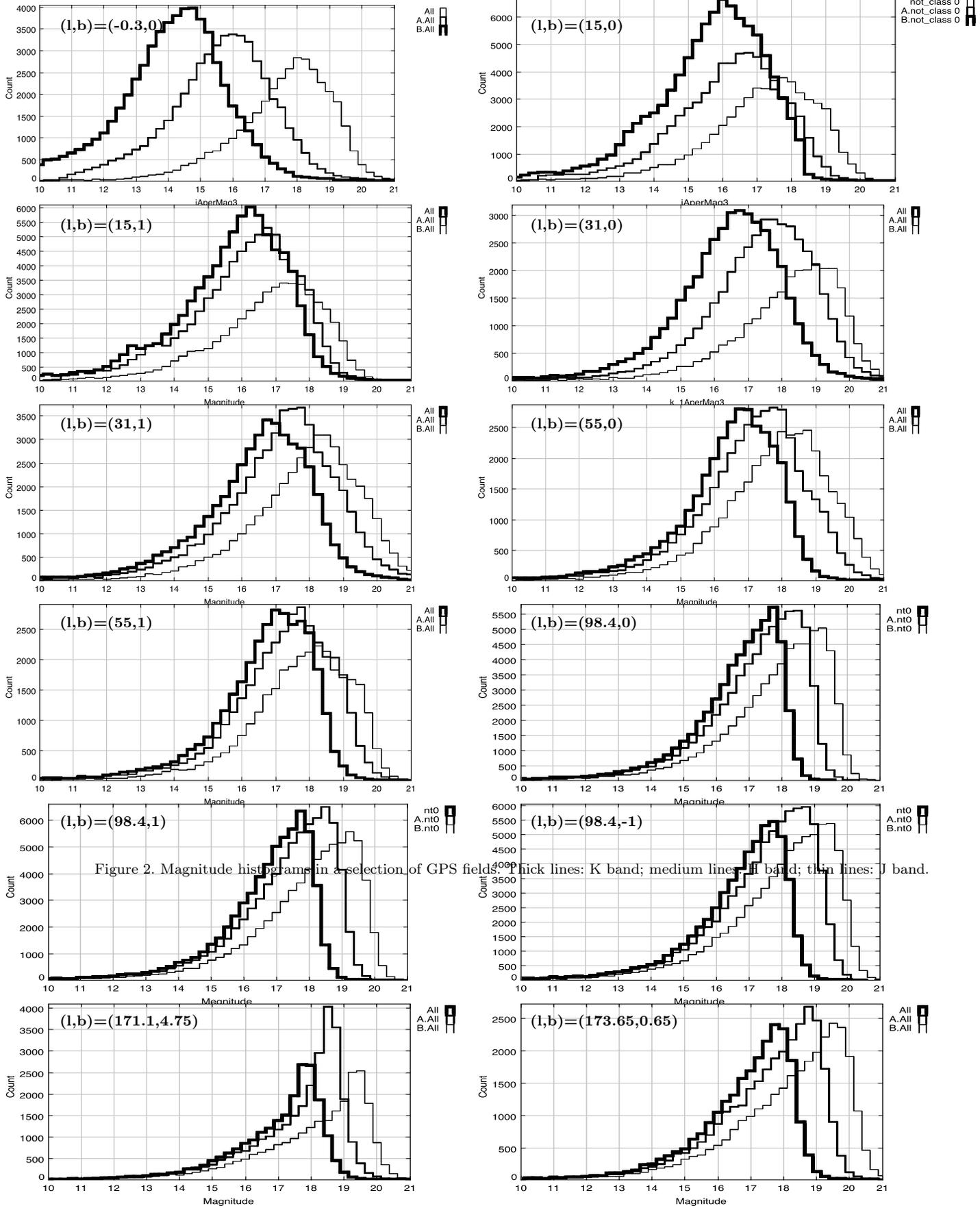
have moved up to $(J-H)=0.5$, owing to the significant extinction present in the mid-plane toward these relatively distant sources (1 to 6 kpc). By contrast the K and M type dwarfs with $(J-H)=0.6$ to 0.7 lie in the same place in both diagrams, owing to the absence of significant extinction toward these nearer and intrinsically fainter sources (<3 kpc).

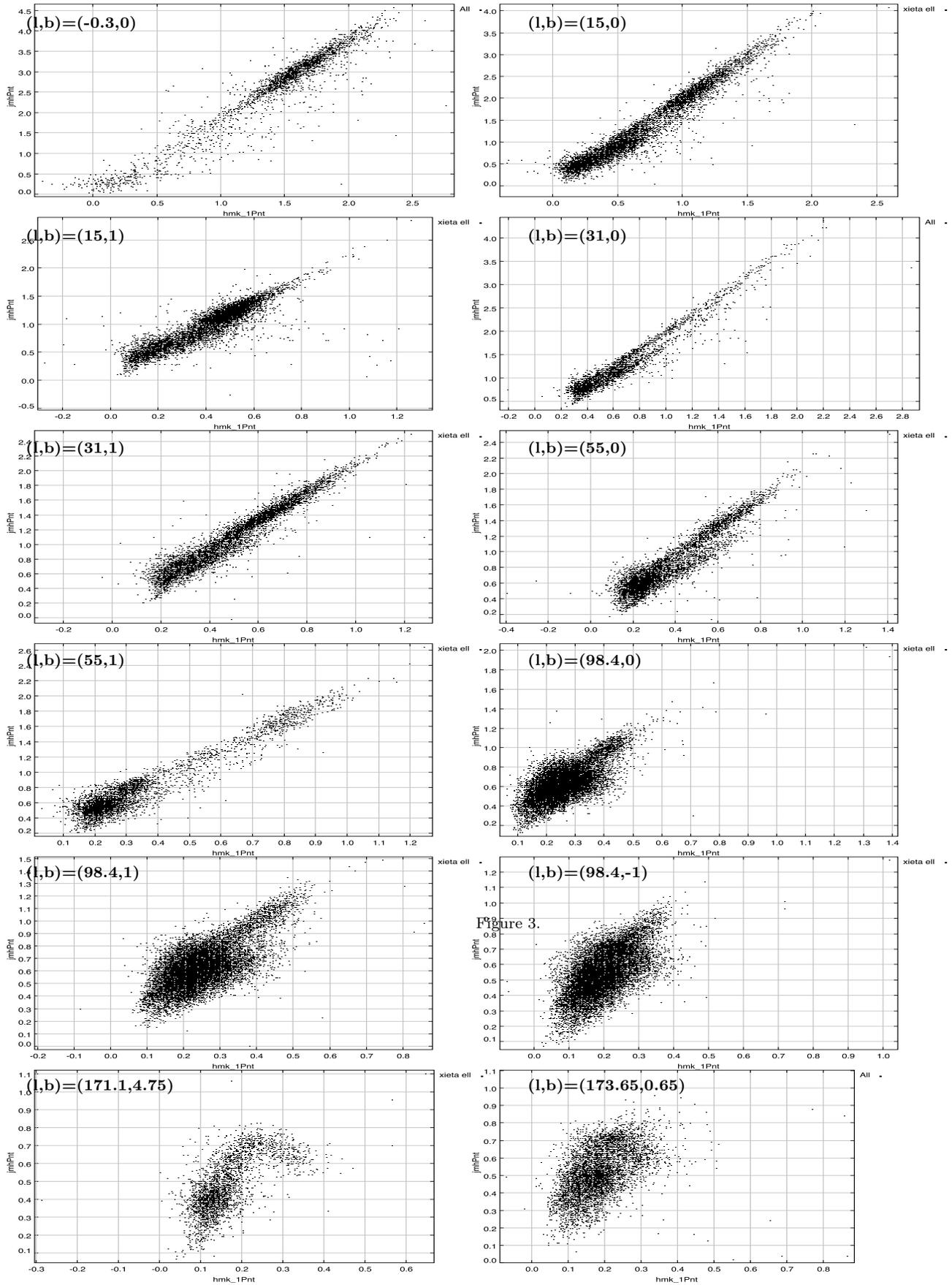
I have work to do to understand the various stellar sequences. The colour mag plots for near the Galactic Centre at $(l,b)=(-0.3,0)$ show three populations. The Besançon models confirm my interpretation that these are disc dwarfs (left sequence), disc red clump giants of types K1III to K2III (faint middle sequence) and Bulge giants (the big population on the right hand side, which covers the full range of giant spectral types apparently). By contrast the two populations seen in the colour mag diagrams for the outer Galaxy seem to be both mostly dwarfs according to Besançon (left hand sequence being F-G types and right hand sequence being K-M types). This surprises me, since I had thought they were simply dwarfs and giants respectively.

5 DEMONSTRATION SCIENCE

References

Irwin M., 2007, on-line,
www.ast.cam.ac.uk/vdfs/docs/catalogues.pdf





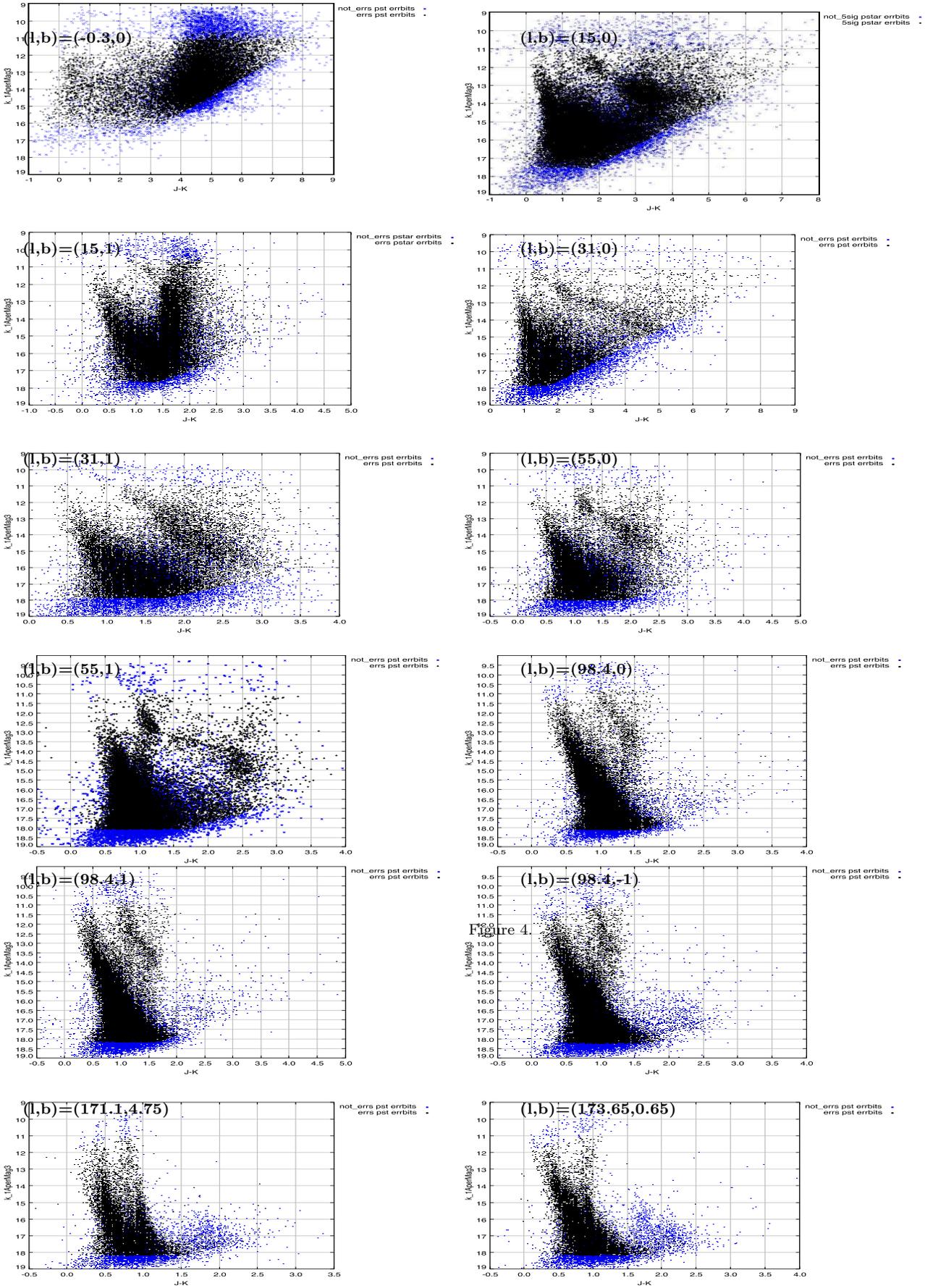


Figure 4.

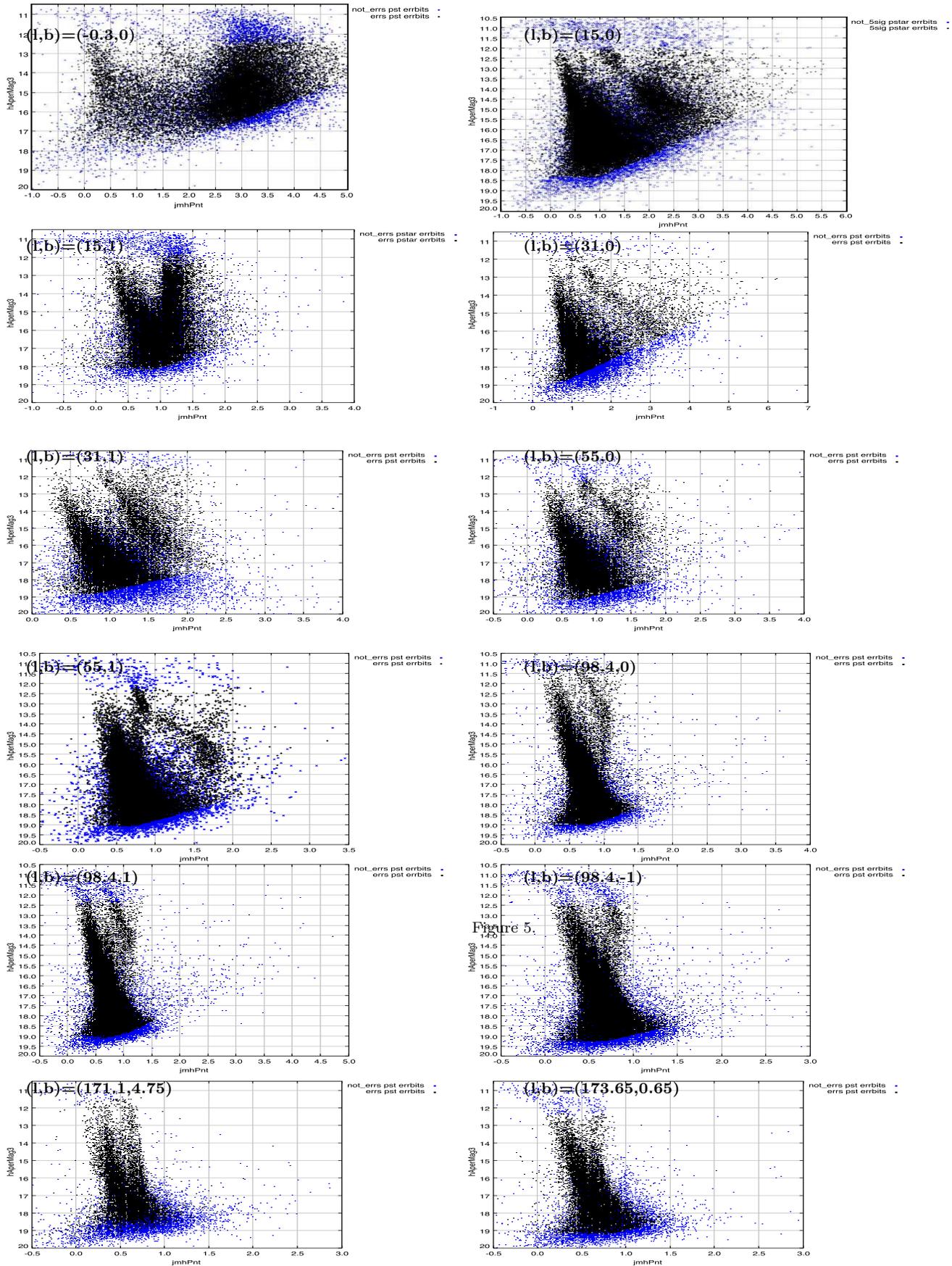


Figure 5.

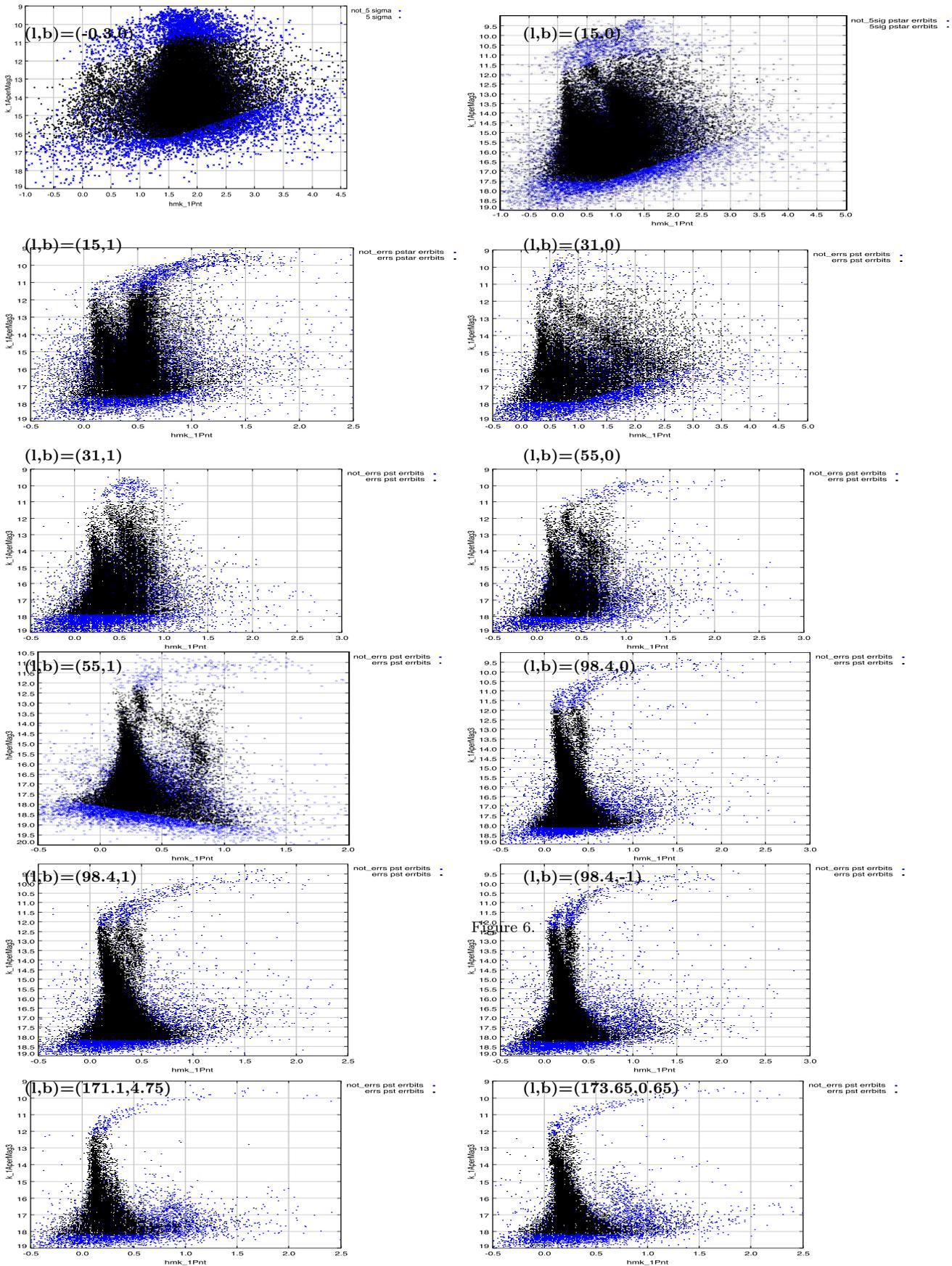


Figure 6.