

# HELP: A catalogue of 170 million objects, selected at 0.36—4.5 $\mu\text{m}$ , from 1270 $\text{deg.}^2$ of prime extragalactic fields

Raphael Shirley,<sup>1,2\*</sup> Yannick Roehly,<sup>1,3,4</sup> Peter D Hurley,<sup>1</sup> Veronique Buat,<sup>3</sup> María del Carmen Campos Varillas,<sup>1</sup> Steven Duivenvoorden,<sup>1</sup> Kenneth J Duncan,<sup>5</sup> Andreas Efstathiou,<sup>6</sup> Duncan Farrah,<sup>7,8</sup> Eduardo González Solares,<sup>9</sup> Katarzyna Małek,<sup>3,10</sup> Lucia Marchetti,<sup>11,12,13</sup> Ian McCheyne,<sup>1</sup> Andreas Papadopoulos,<sup>6</sup> Estelle Pons,<sup>9</sup> Roberto Scipioni,<sup>1</sup> Mattia Vaccari,<sup>12,13</sup> and Seb Oliver.<sup>1</sup>

<sup>1</sup>*Astronomy Centre, Department of Physics and Astronomy, University of Sussex, Brighton, UK, BN1 9QH*

<sup>2</sup>*Instituto de Astrofísica de Canarias, E-38205 La Laguna, Tenerife, Spain;*

*Universidad de La Laguna, Dpto. Astrofísica, E-38206 La Laguna, Tenerife, Spain*

<sup>3</sup>*Aix Marseille Université, CNRS, CNES, LAM, Marseille, France*

<sup>4</sup>*Université de Lyon, ENS de Lyon, CNRS, Centre de Recherche en Astrophysique de Lyon UMR5574, 69230 Saint-Genis-Laval, France*

<sup>5</sup>*Sterrewacht Leiden, Universiteit Leiden, Leiden, Netherlands*

<sup>6</sup>*School of Sciences, European University Cyprus, Diogenes street, Engomi, 1516 Nicosia, Cyprus*

<sup>7</sup>*Department of Physics and Astronomy, University of Hawaii, 2505 Correa Road, Honolulu, HI 96822, USA*

<sup>8</sup>*Institute for Astronomy, 2680 Woodlawn Drive, University of Hawaii, Honolulu, HI 96822, USA*

<sup>9</sup>*Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge, UK*

<sup>10</sup>*National Centre for Nuclear Research, ul. Hoża 69, 00-681 Warszawa, Poland*

<sup>11</sup>*Department of Astronomy, University of Cape Town, 7701 Rondebosch, Cape Town, South Africa*

<sup>12</sup>*Department of Physics and Astronomy, University of the Western Cape, 7535 Bellville, Cape Town, South Africa*

<sup>13</sup>*INAF - Istituto di Radioastronomia, via Gobetti 101, 40129 Bologna, Italy*

September 10, 2019

## ABSTRACT

We present an optical to near-infrared selected astronomical catalogue covering 1270  $\text{deg.}^2$ . This is the first attempt to systematically combine data from 23 of the premier extragalactic survey fields – the product of a vast investment of telescope time. The fields are those imaged by the *Herschel* Space Observatory which form the *Herschel* Extragalactic Legacy Project (HELP). Our catalogue of 170 million objects is constructed by a positional cross match of 51 public surveys. This high resolution optical, near-infrared, and mid-infrared catalogue is designed for photometric redshift estimation, extraction of fluxes in lower resolution far-infrared maps, and spectral energy distribution modelling. It collates, standardises, and provides value added derived quantities including corrected aperture magnitudes and astrometry correction over the *Herschel* extragalactic wide fields for the first time. *grizy* fluxes are available on all fields with *g* band data reaching  $5\sigma$  point-source depths in a 2 arcsec aperture of 23.5, 24.4, and 24.6 (AB) mag at the 25th, 50th, and 75th percentiles, by area covered, across all HELP fields. It has *K* or *K<sub>s</sub>* coverage over 1146  $\text{deg.}^2$  with depth percentiles of 20.2, 20.4, and 21.0 mag respectively. The IRAC Ch 1 band is available over 273  $\text{deg.}^2$  with depth percentiles of 17.7, 21.4, and 22.2 mag respectively. This paper defines the “*masterlist*” objects for the first data release (DR1) of HELP. This large sample of standardised total and corrected aperture fluxes, uniform quality flags, and completeness measures provides large well understood statistical samples over the full *Herschel* extragalactic sky.

**Key words:** catalogues – surveys – astronomical data bases: miscellaneous – galaxies: statistics

## 1 INTRODUCTION

Galaxy catalogues play a fundamental role in modern astronomy and cosmology. Combining catalogues and images from different instruments is a significant challenge and will be increasingly important as deeper and wider surveys are conducted in the coming years. Huge efforts have gone in to creating large homogeneous data sets such as for the SXDS (Furusawa et al. 2008), the Optical-NIR catalogue of the AKARI-NEP Deep Field (Oi et al. 2014), COSMOS (e.g. Scoville et al. 2007; Ilbert et al. 2013; Laigle et al. 2016), and GAMA (Driver et al. 2011). However, creating a consistent data set over the *Herschel* extragalactic fields is a challenge due to the wide variety of projects that have studied these fields. Each survey is at a different depth, with different source extraction pipelines, different astrometric solutions and units used, and different quality in terms of seeing, point spread function consistencies and exposure times. In this paper we collate and combine a large number of public astronomical catalogues to produce a single catalogue of general use to the astronomical community.

Access to such data sets over a wide area extends the scientific value for e.g. the discovery of rare objects, reducing sampling variance, studying environmental factors and statistical studies in fine sub-samples of populations. Dedicated large-area surveys such as the Dark Energy Survey (DES; Abbott et al. 2018) and the upcoming Large Synoptic Survey Telescope (LSST; Ivezić et al. 2008) typically only provide five or six optical bands. To access a wide multi-wavelength range for physical modelling and to exploit the deepest data requires the combination of many data sets from different telescopes. The premier extra-galactic fields<sup>1</sup> represent many hundreds of nights of the best ground based telescopes and thousands of hours of space telescope time and yet have never been effectively collected together.

The *Herschel*<sup>2</sup> Extragalactic Legacy Project (HELP<sup>3</sup>) (Oliver et al., in preparation) brings together key multi-wavelength surveys. It homogenises them; exploits prior information from the optical, near-infrared and mid-infrared surveys to deblend the low-resolution long wavelength maps using `xid+` (Hurley et al. 2017); provides well-calibrated photometric redshifts (Duncan et al. 2018a; Duncan et al. 2018b); and conducts energy-balanced spectral energy distribution (SED) modelling through `CIGALE` (Buat et al. 2018; Małek et al. 2018) to provide physical parameter estimations.

The first stage in this process is to compile all the ancillary data that is available across the HELP fields into a “*masterlist*” of objects. The principal aim is to produce a catalogue with consistent photometry from the best photometric catalogues that are available. A secondary aim is to

characterise the depth of the selected data to enable meaningful statistical analysis with standard empirical measures such as luminosity functions.

Here, we describe in detail the production and the extensive validation, flagging and characterisation of the *masterlist* across the full HELP area. We do this in part by discussing an example field, the European Large Area ISO Survey field North 1 (ELAIS-N1; Oliver et al. 2000), in detail. Although we use an example field for discussing the combination methods we provide detailed quantitative measures of the depths and number counts for all the fields individually and combined over the entire area.

The *masterlist* presented here is the basis for the first public HELP data release (HELP DR1) and defines the HELP identifiers and positions which will be used in subsequent data products for DR1.

Full details of decisions, validation and characterisation are available in the JUPYTER notebooks<sup>4</sup> (Kluyver et al. 2016) that were used to run all the PYTHON code used in the pipeline and are made available as part of the HELP documentation.

This work is of general use for galaxy formation studies by providing larger statistical samples than have been available before. The provision of tools to model selection functions in particular permit the combination of narrow deep fields with wide shallow fields to investigate both the bright and faint end of the luminosity function simultaneously. It facilitates the use of large numbers of different surveys for such research. For its specific purpose of deblending far-infrared maps, computing photometric redshifts, and modelling spectral energy distributions of galaxies, it is of critical importance. Constructing a prior list of objects for the application of Bayesian forced photometry with `xid+` requires a uniformly defined selection which is well correlated with far-infrared flux. This catalogue can be consistently used for such a purpose across all *Herschel* imaging, from the heavily observed and deep COSMOS field to the hundreds of square degrees of HATLAS-SGP.

The format of this paper is as follows. First, in section 2 the HELP fields are described and an overview of the various input data sets is given and a detailed overview of our example, pilot field, ELAIS-N1, is given to demonstrate the method of data combination. The data reduction pipeline is described in section 3 with some examples from the example field, ELAIS-N1. Validation of the data and the depth maps used to quantify survey depths are described in sections 4 and 5 respectively. In section 5.1 we compare the depths and number accounts across all the fields both combined and individually, for critical detection bands. Finally, the data presented is summarised in section 6.

## 2 THE INPUT DATA SETS

In this section, we first give an overview of the fields and their location on the sky. Then, using ELAIS-N1 as a demonstrative example, we describe the method of data combination that has been applied on all the fields. There are 51 public surveys combined in this work. A full list of all

<sup>1</sup> AKARI-NEP, AKARI-SEP, Bootes, CDFS-SWIRE, COSMOS, EGS, ELAIS-N1, ELAIS-N2, ELAIS-S1, GAMA-09, GAMA-12, GAMA-15, HDF-N, *Herschel*-Stripe-82, Lockman-SWIRE, NGP, SA13, SGP, SPIRE-NEP, SDF, xFLS, XMM-13hr, and XMM-LSS.

<sup>2</sup> *Herschel* Space Observatory (Pilbratt et al. 2010)

<sup>3</sup> <https://herschel.sussex.ac.uk/>

<sup>4</sup> [https://github.com/H-E-L-P/dmu\\_products](https://github.com/H-E-L-P/dmu_products)

these surveys with information regarding the instruments and bands used is given in table A1.

## 2.1 HELP fields

There are 23 HELP fields. Figure 1 shows their location on the sky. Creating statistical samples from these various fields in order to study galaxy formation and evolution first requires collecting data across the fields. In order for these samples to be scientifically useful they must have a well described selection function. It is therefore necessary to combine and manipulate data in a reproducible manner and to quantify depth on every patch of sky.

The data available on a given field is highly variable in terms of number of public surveys, the bands measured by those surveys, the coverage of each band on the fields, and the depths of those surveys. On XMM-LSS for example there are 23 surveys available although their respective coverages are not uniform across the field. On HATLAS-SGP however, covering 295 deg.<sup>2</sup>, there are just 6 surveys and no mid-infrared coverage by *Spitzer*. This is where the depth maps presented in section 5 become crucial for understanding selection effects leading to differences in effective area depending on a given sample selection. Also in section 5, we will discuss the depths available from specific filters on specific instruments in terms of the standard broad band filters:  $u$ ,  $g$ ,  $r$ ,  $i$ ,  $z$ ,  $y$ ,  $J$ ,  $H$ ,  $K$ ,  $K_s$ , IRAC Ch 1 ( $3.6\mu\text{m}$  - ‘ $i1$ ’), IRAC Ch 2 ( $4.5\mu\text{m}$  - ‘ $i2$ ’), IRAC Ch 3 ( $5.6\mu\text{m}$  - ‘ $i3$ ’), and IRAC Ch 4 ( $8.0\mu\text{m}$  - ‘ $i4$ ’). In total there are 150 specific filters used in the HELP data. 29 of these are narrow band filters. In the following sections we will describe the typical broad band filters used. All the filters have transmission curves available in the database, as illustrated in figure 2. The filters were taken from the Spanish Virtual Observatory (SVO) (Rodrigo et al. 2012, 2017) or from the original survey databases. These are corrected for atmospheric extinction and CCD quantum efficiency, and are as used in all subsequent HELP data processing.

### 2.1.1 Optical data

We define the optical region as between  $0.36\text{--}1.5\mu\text{m}$ . All of the HELP fields have some coverage by optical surveys in the  $g$ ,  $r$ ,  $i$ ,  $z$ , and  $y$  bands. Some areas of HELP are also covered by the broad band  $u$  filter in addition to narrow bands. Later in the paper we will describe tools for determining what area is covered by each of these bands. General descriptions of all the optical data and its coverage of HELP area are given in appendix A.

### 2.1.2 Near-infrared data

We define the near-infrared region as between  $1.5\text{--}3.0\mu\text{m}$ . 1146 deg.<sup>2</sup> of HELP is covered by either  $K$ , or  $K_s$ . All HELP fields have some coverage excluding the smaller fields ( $< 10\text{ deg.}^2$ ) ELAIS-N2, SA13, SPIRE-NEP, xFLS, and XMM-13hr. There are fluxes from one or more of the  $J$ ,  $H$ ,  $K$ , and  $K_s$  bands from 10 instruments. Full descriptions of the telescopes, instruments and bands are given in appendix A.

### 2.1.3 Mid-infrared data

We define mid-infrared data as between  $3.0\text{--}10.0\mu\text{m}$ . Photometry from this part of the spectrum exclusively comes from the IRAC camera on the *Spitzer* space telescope. We have measurements in IRAC bands over 273 deg.<sup>2</sup> of HELP.

A summary of the datasets available across all HELP fields is given in table A1. They are each based on different selection criteria corresponding to features of the instrument, observations, and extraction software and have varying depths and coverage.

### 2.1.4 Multi-wavelength Catalogues

On some HELP fields we have multi-wavelength catalogues produced with forced photometry from  $K$  selected catalogues (where the positions from a  $K$ -band catalogue are used to extract fluxes from the other imaging bands), or from other methods such as stacked images as in the COSMOS2015 catalogue (Laigle et al. 2016) (where a  $\chi^2$  sum of the four near-infrared bands and the optical  $z$  is used to extract positions and forced photometry is then conducted on those positions). Where catalogues contain narrow bands in addition to the common broad bands, we include all bands available.

## 2.2 An example field in detail: ELAIS-N1

In this section we will demonstrate some of the general data properties and the combination method on an example field. We chose the ELAIS-N1 field to be representative of the HELP coverage in general. ELAIS-N1 was the first *Spitzer* Wide InfraRed Extragalactic (SWIRE; Lonsdale et al. 2003) field observed. This field is used to demonstrate the full HELP pipeline on areas with InfraRed Array Camera (IRAC) coverage, important for the HELP deblending with xID+ (Hurley et al. 2017). It has previously been used to test another part of the HELP pipeline (Malek et al. 2018) and for early studies using HELP data (Ocran et al. 2017).

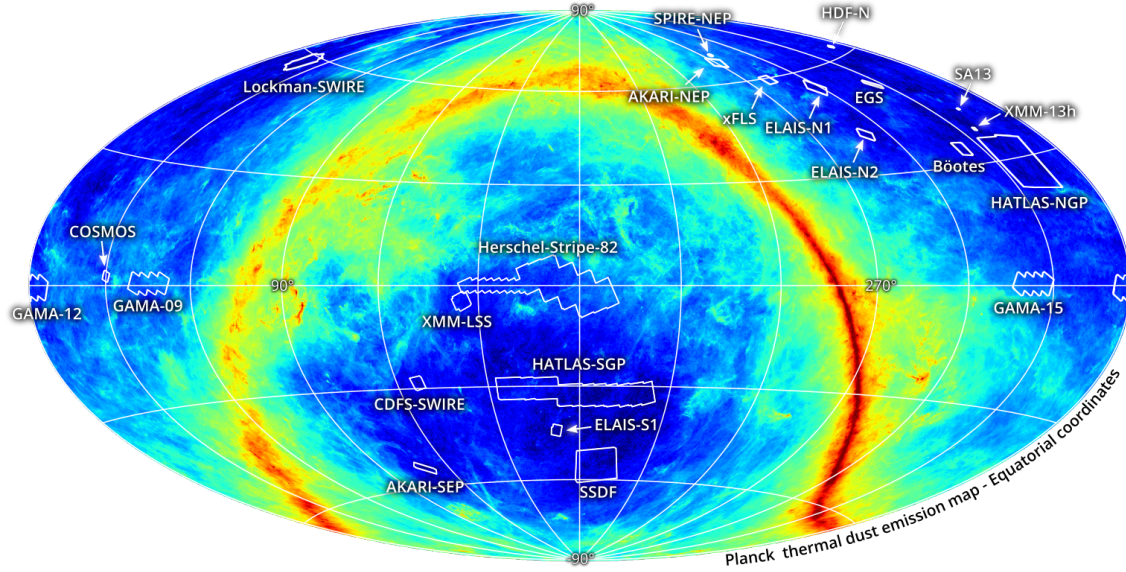
Table 1 shows all the catalogues that are available on the ELAIS-N1 field. The spectral responses of the bands available in these surveys are shown in figure 2. Figure 3 shows the various observation coverages overlaid on the SPIRE 250  $\mu\text{m}$  map<sup>5</sup> and corresponding variance map.

### 2.2.1 Optical data

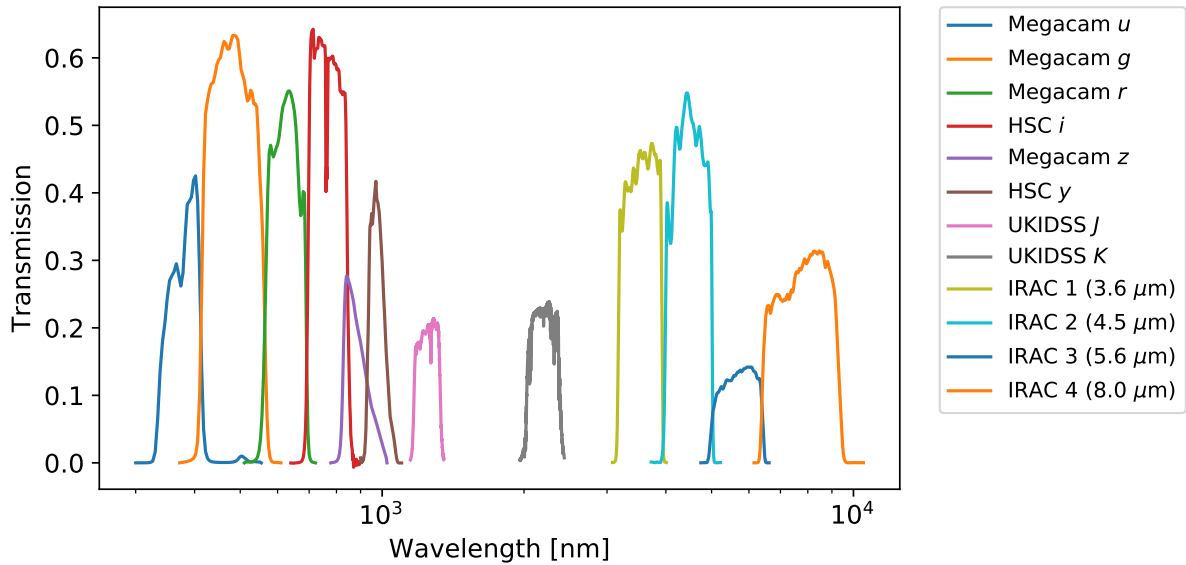
Observations from the Isaac Newton Telescope/Wide Field Camera (INT/WFC) cover 93% of the ELAIS-N1 field (González-Solares et al. 2011). This survey comprises  $u, g, r, i, z$  band imaging with magnitude limits in  $u, g, r, i, z$  of 23.9, 24.5, 24.0, 23.3, 22.0 respectively (AB,  $5\sigma$  point source).

The Subaru Telescope/Hyper-Suprime-Cam Strategic Program Catalogues (HSC-SSP) wide area survey covers

<sup>5</sup> We display SPIRE here because the motivation for producing the database was to provide relevant ancillary data to the *Herschel* images; for our fields, PACS data has generally the same coverage as it was acquired in parallel mode.



**Figure 1.** The boundaries of the HELP fields overlaid on the Planck Galactic thermal dust emission map. Field areas range from less than one  $\text{deg.}^2$  (HDF-N, SA13, SPIRE-NEP, and XMM-13hr) to several hundred  $\text{deg.}^2$  (e.g. HATLAS-SGP).



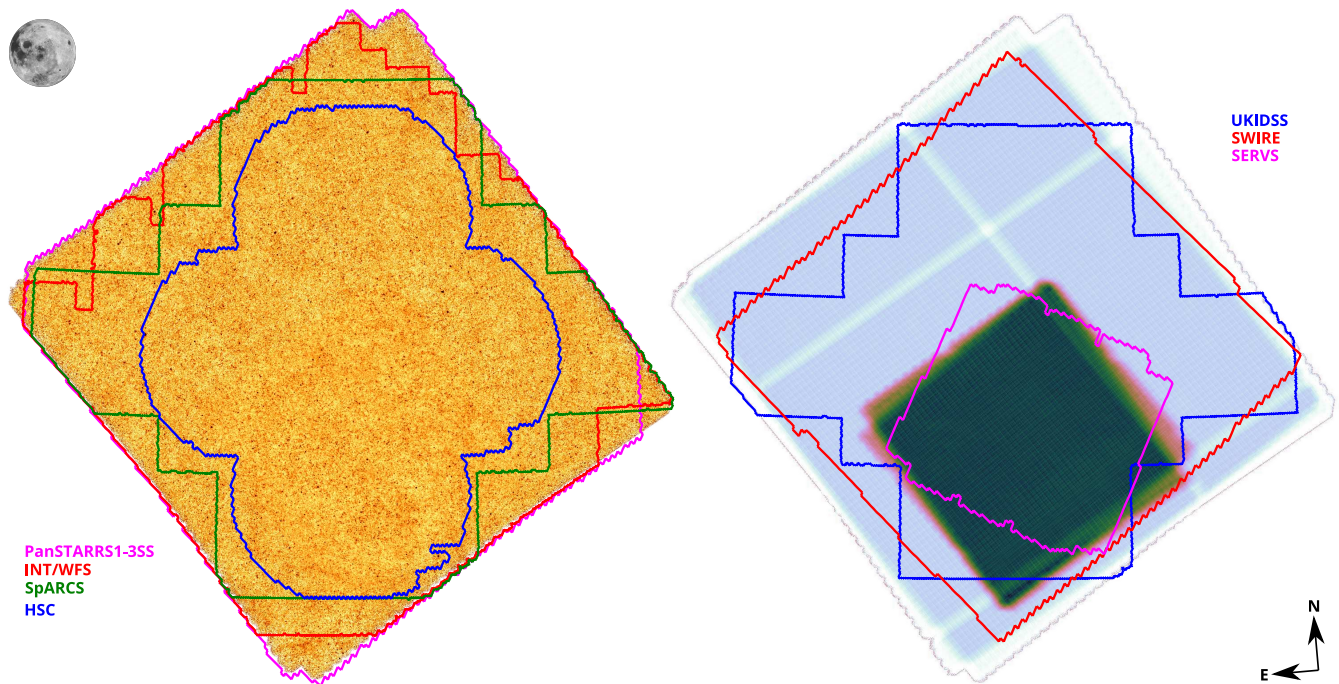
**Figure 2.** Filter transmission curves for a sample of filters present on the example field ELAIS-N1. Some of the optical bands have multiple measurements from different instruments with similar filters. We only show one for each band to aid clarity. These responses include quantum efficiency of the camera and atmospheric extinction as measured by the respective telescopes. Wavelength coverage varies significantly across the HELP fields. Figures for each individual field are provided in the notebooks.

57% of the ELAIS-N1 field (Aihara et al. 2018). The survey contains imaging in five broad bands ( $g, r, i, z, y$ ), with a  $5\sigma$  AB point-source depth of  $r \approx 26$ .

The *Spitzer* Adaptation of the Red-sequence Cluster Survey (SpARCS) contains fluxes from the Canada France Hawaii Telescope’s MegaCam instrument in  $ugryz$  (Tudor-ica et al. 2017) down to a mean AB  $5\sigma$   $z$  depth of 24.

### 2.2.2 Near-infrared data

The UK Infrared Telescope Deep Sky Survey - Deep Extragalactic Survey (UKIDSS-DXS) provides  $J$  and  $K$  band data to a  $5\sigma$  point source AB depth of  $K = 21.0$  (Lawrence et al. 2007; Swinbank 2013).



**Figure 3.** Coverage of the various pristine catalogues on the example field ELAIS-N1 with Moon shown for scale. The image on the left is the SPIRE 250  $\mu\text{m}$  map with the optical survey regions overlaid. The image on the right shows the SPIRE 250  $\mu\text{m}$  error map with the near and mid-infrared survey regions overlaid.

### 2.2.3 Mid-infrared data

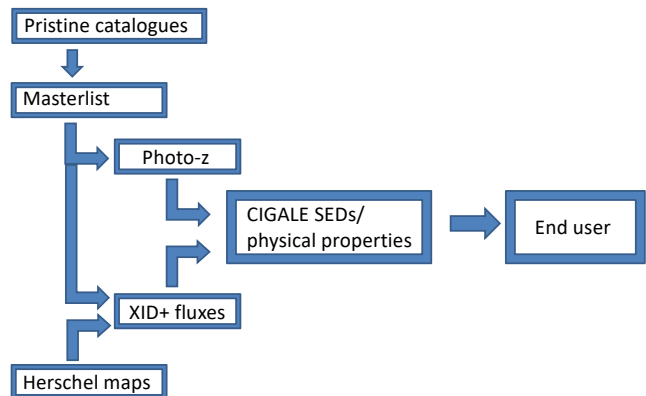
There are two separate surveys providing mid-infrared fluxes from IRAC on the *Spitzer* Space Telescope. The *Spitzer* Extragalactic Representative Volume Survey (SERVS; Mauduit et al. 2012) provides mid-infrared fluxes to a greater depth,  $5\sigma$  AB point-source depth of  $\approx 23$ , over a smaller area of 2 deg.<sup>2</sup> The *Spitzer* Wide InfraRed Extragalactic survey (SWIRE; Lonsdale et al. 2003) provides fluxes in all four IRAC bands over 9.65 deg.<sup>2</sup> but to  $5\sigma$  AB point-source depth of  $\approx 22$ . In our work we use the *Spitzer* data fusion products for SERVS and SWIRE as presented in Vaccari (2015).

Our pipeline starts with *pristine* catalogues provided by independent survey teams; standardises these to produce consistently formatted *added value* catalogues; and merges these together to produce multi-wavelength *masterlist* catalogues. These then feed in to later stages of the HELP pipeline. An overview of the full HELP pipeline is shown in figure 4.

## 3 THE MASTERLIST PIPELINE

The entire pipeline is written in PYTHON making extensive use of ASTROPY for the cross matching (The Astropy Collaboration et al. 2018). Appendix B gives details of how to access and run the code including all the code used to produce the figures presented here.

In addition to processing the data the pipeline performs quality analysis and produces diagnostic plots that helped us identify errors or misunderstandings of the pristine data (e.g. Vega magnitudes reported as AB) and provide a useful



**Figure 4.** Schematic diagram of the full HELP pipeline. This paper concerns the production of the *masterlist* which impacts all the later data processing.

assurance for the user. These are discussed more in Section 4.

### 3.1 Pristine catalogues

We collect *pristine* catalogues from public data repositories, usually as delivered by the original survey teams. We provide these original tables such that one can return to the original data as it was before any corrections or conversions were carried out. This is especially useful when a particular survey contains extra information not present on all surveys and so

**Table 1.** Overview of data included on the ELAIS-N1 field. We chose the deepest public data available. Shallower data such as SDSS and 2MASS are not included because they don't provide useful extra data compared to deeper surveys. SDSS indexes are however included in order to facilitate quick identification of SDSS objects. Coverage is the percentage of the full *Herschel* field observed by the given survey. A summary of the data used across all HELP fields is given in table A1.

Input survey name	Coverage	bands	reference
Isaac Newton Telescope / Wide Field Camera (INT-WFC)	93%	<i>u, g, r, i, z</i>	González-Solares et al. (2011)
UKIRT Infrared Deep Sky Survey / Deep Extragalactic Survey (UKIDSS/DXS)	65%	<i>J, K</i>	Swinbank (2013)
Hyper Suprime-Cam Subaru Strategic Program Catalogues (HSC-SSP)	57%	<i>g, r, i, z, y, N921, N816</i>	Aihara et al. (2018)
Pan-STARRS1 - 3pi Steradian Survey (3SS) data	100%	<i>g, r, i, z, y</i>	Chambers et al. (2016)
<i>Spitzer</i> Adaptation of the Red-sequence Cluster Survey (SpARCS)	81%	<i>u, g, r, z</i>	Tudorica et al. (2017)
<i>Spitzer</i> Data Fusion (SERVS)	20%	IRAC i1, i2	Vaccari (2015)
<i>Spitzer</i> Data Fusion (SWIRE)	73%	IRAC i1, i2, i3, i4	Vaccari (2015)

not propagated through our pipeline such as morphology information or the results of modelling.

### 3.2 Value-added catalogue preparation

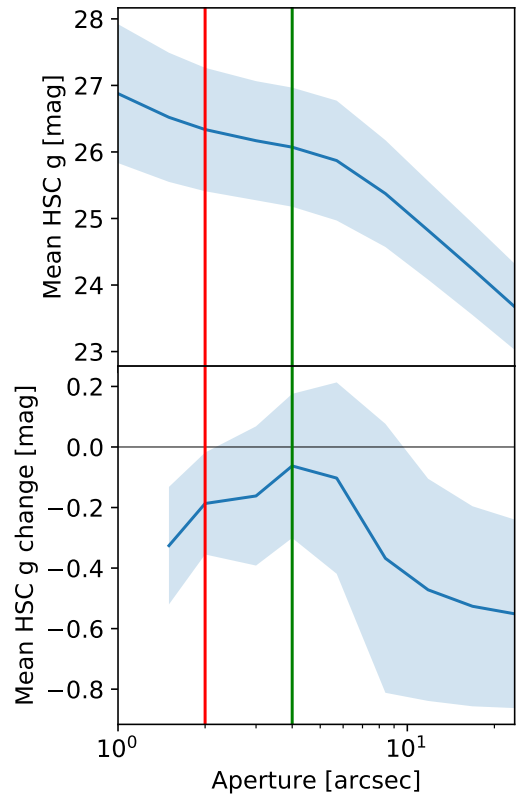
The first stage required to produce the *masterlist* is to standardise the individual surveys. They must be converted into a format with consistent metadata; column headings, units, and column descriptions. Data are also set to the same astrometric reference frame and flux and magnitudes are converted to our standard of  $\mu\text{Jy}$  and AB magnitude respectively.

#### 3.2.1 Standardising fluxes and magnitudes

In the final catalogue we provide both fluxes and magnitudes. This is done partly to make the catalogues more user friendly as both are still widely used and partly because we want to retain all information from the initial catalogues that is provided. We convert any Vega magnitudes to AB, and provide fluxes in units of  $\mu\text{Jy}$ .

We record both total and aperture fluxes as these have different scientific uses. The aperture fluxes are used to compute photometric redshifts (see Duncan et al. (2018a); Duncan et al. (2018b) for an overview of the HELP photometric redshift pipeline which utilises the Easy and Accurate Z(photometric redshifts) from Yale code (EAZY; Brammer et al. 2008)). Total fluxes are used to fit the spectral energy distributions (SEDs) using the method presented in Małek et al. (2018), which uses the Code Investigating GALaxy Emission (Noll et al. 2009; Roehly et al. 2013; Boquien et al. 2019, CIGALE;). For point sources total and aperture fluxes should be the same but diverge for extended objects.

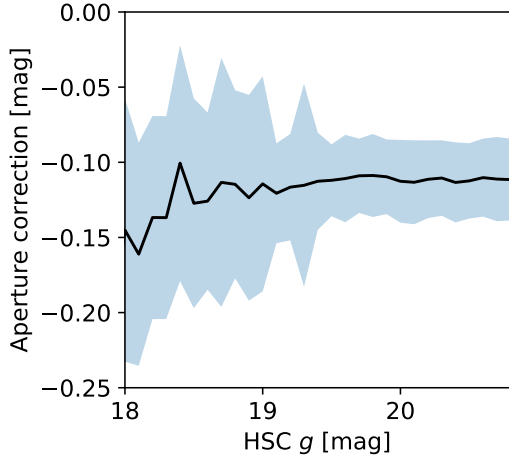
Total magnitudes are either Kron magnitudes or the SExtractor AUTO magnitudes (Bertin & Arnouts 1996) chosen in that order if both are available. 2 arcsec diameter aperture magnitudes are calculated and corrected if necessary from the closest aperture to 2 arcsec. 2 arcsec is chosen as it provides a good compromise between high signal to noise and capturing a significant fraction of the total flux for a typical optical point spread function, while avoiding blended nearby objects. It is also the most commonly available aperture and consistency is a fundamental aim. Nevertheless, the choice of 2 arcsec will not be ideal in all situations. At low redshift it will yield lower signal to noise than a larger aperture for faint extended objects. It will impact selection effects and biases, through to the calculation of redshifts. Any use of these samples will have to account for this choice of total and aperture flux properties.



**Figure 5.** Top: Hyper-Suprime-Cam uncorrected *g*-band aperture magnitude as a function of aperture size on the example field ELAIS-N1. Bottom: change in magnitude from previous aperture for point source objects. Errors are shown by the shaded region. Both figures represent the same set of objects. The vertical red line to the left shows the 2 arcsec value that is used by HELP for all surveys (or nearest available). The vertical green line to the right shows the target aperture used here (4 arcsec) to compute the average aperture correction to be applied to all objects. The curve starts to level off before showing monotonically decreasing behaviour due to contaminating background sources.

#### 3.2.2 Aperture correction

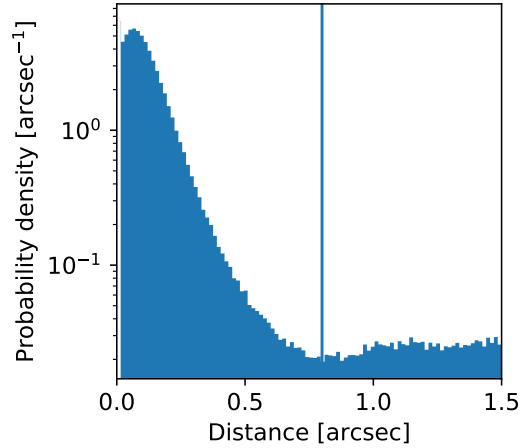
We provide aperture magnitudes in a standard 2 arcsec aperture. We provide ‘corrected’ aperture magnitudes. The correction is to account for the fact that a 2 arcsec aperture will not encapsulate the full point spread function. The fluxes are therefore divided by the proportion of flux from a point source that is captured in the aperture. This means that for a point source they should be equal to the total mag-



**Figure 6.** HSC aperture correction (difference between target, 4 arcsec, aperture magnitude and 2 arcsec aperture magnitude) as a function of  $g$  band magnitude on the example field ELAIS-N1. The shaded region shows the 25th to 75th percentiles. We check it is relatively stable and take an average over the magnitude range where the correction is level; above the high corrections in very bright objects, and apply this to all objects. For the above graph point sources between 19 and 21 have a constant aperture correction which should be representative of point like galaxies at higher magnitudes. At the bright end, the sample may become subject to small number statistics and the divergence from the constant value seen between 17.5 mag and 20.0 mag may be due to point spread function features or saturation which are negligible in faint sources.

nitide. However, for extended sources they will have higher aperture magnitude than total magnitude. If these corrected aperture magnitudes are not provided by the original survey, we calculate an aperture correction by using curve of growth analysis (Stetson 1990). Figure 5 shows how the parameters for the corrections are calculated for a typical example, HSC. This correction was applied to SpARCS and HSC data on the example field ELAIS-N1 and all surveys on other fields where corrected aperture fluxes are not provided. We look at the mean object flux in the aperture as a function of the aperture size. For a perfect zero background image in an empty field this would level off to a constant when the largest aperture included the entire point spread function. In actuality the averages start to grow as the large apertures start to include other background sources. We thus look for the point where the flux has started to level off but before it starts growing linearly with area due to unrelated sources and use that point to set a target aperture which is used to measure the ratio of the total flux for a point source to that captured in the 2 arcsec aperture. This correction, is then uniformly applied to every aperture flux. Figure 6 shows the correction as a function of magnitude. We use these figures to check that there is no significant magnitude dependence or high noise.

Fluxes and magnitudes are not corrected for galactic extinction. However, we include an E(B-V) column which can be used to apply a correction. This is done using the E(B-V) values from the Schlegel et al. (1998) dust map



**Figure 7.** Number density of objects from the *masterlist* within a given distance from a GAIA object with  $g$  magnitude above 18. For bright stars there are large numbers of artefacts. This figure demonstrates that 0.8 arcsec is effective as a standard cross match distance. Taking all objects within 0.8 arcsec will include the majority of true matches (peak at left) and a minority of background erroneous matches (linear slope). This value is chosen for every catalogue based on the minimum in the above figure. For the majority of catalogues 0.8 arcsec (vertical line on plot) is appropriate.

and a scaling of 0.86 is applied to the values to reflect the re-calibration by Schlafly & Finkbeiner (2011).<sup>6</sup>

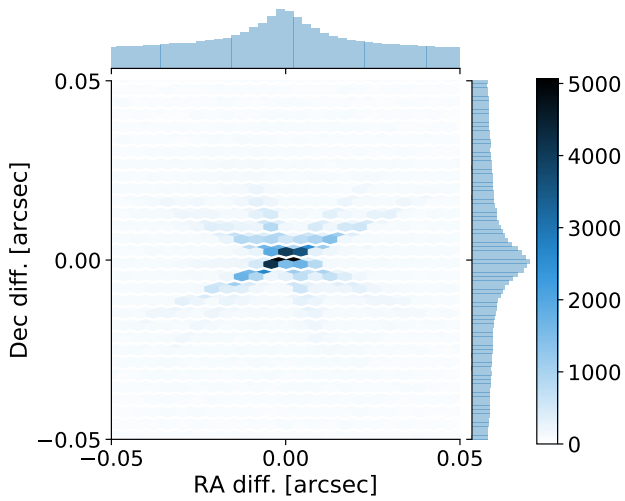
### 3.2.3 Removing duplicate objects

Some pristine catalogues included objects extracted from overlapping image tiles. This can lead to duplicates. We identify duplicates as objects within 0.4 arcsec of one and another in the same catalogue. This threshold is intended to catch most true duplicates, without excluding close pairs of intrinsically different objects. We remove the object with the highest error and flag the remaining object as having been “cleaned”. This procedure will be influenced by the deblending algorithm used in the production of the pristine catalogue and may remove close pairs that have been correctly deblended in highly resolved catalogues. For our main aim of deblending *Herschel* objects these offsets are negligible compared to the *Herschel* beam but will influence photometric redshifts and SED fitting. In the deepest and highest resolution *Hubble* CANDELS fields this method led to the removal of thirty objects out of over one hundred thousand.

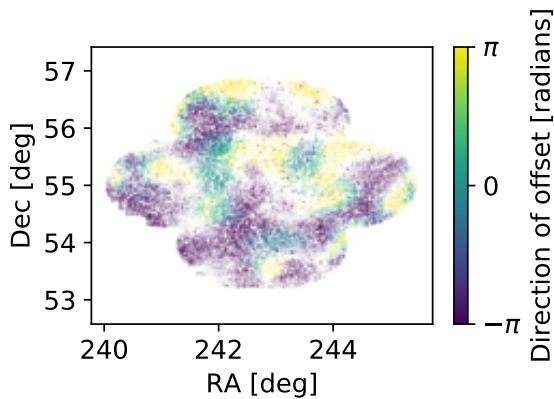
### 3.2.4 Astrometry correction

We convert all astrometry to the GAIA DR1 (Gaia Collaboration et al. 2016) astrometric reference frame. We estimate the mean offset by performing a positional cross match with a 0.6 arcsec radius. Offset plots are produced which can be inspected to diagnose anomalies such as significantly non-Gaussian errors. Figure 8 shows the offsets of *masterlist* objects compared to GAIA objects with  $g$  magnitude greater

<sup>6</sup> Dust map taken from <https://github.com/kbarbary/sfdmap>.



**Figure 8.** Offsets of *masterlist* positions to GAIA reference for objects with GAIA  $g$  magnitude larger than 18.



**Figure 9.** Direction of residual astrometric offsets between HSC positions and the GAIA reference frame, after correction for mean offset, for the example field ELAIS-N1. Colour indicates the direction and intensity indicates the size of residual offset.

than 18. We only use the fainter objects because the bright stars often cause a large number of artefacts in the galaxy imaging surveys. The spikes in the figure and the heavy tails in the distributions are due to these. Nevertheless, the distribution is centred and tight.

We also produce a map showing the average direction of the offset for every field and every input catalogue in the diagnostic notebooks. This can reveal regular shaped regions which could result from individual tiles which should be corrected separately. Figure 9 gives an overview of astrometry offsets for HSC data before correction on the example field ELAIS-N1. These broadly Gaussian errors are typical of all the data sets but automatically plotted for every pristine catalogue to check for errors or inconsistencies.

We also take a unique identifier from the pristine catalogue such that any object from the final catalogue is associated with the original data set via a cross identification table and our internal HELP ID. This means that any additional information such as morphological metrics or flags for object type beyond stellarity are still available. It also means that

a user with experience with one of the pristine data sets can quickly find corresponding HELP IDs for target objects to retrieve HELP data products.

### 3.3 Merging the value-added catalogues

Once all the original data has been standardised, the catalogues are merged to produce a single *masterlist*. The *masterlist* provides the fundamental objects fed in to all subsequent parts of the HELP workflow.

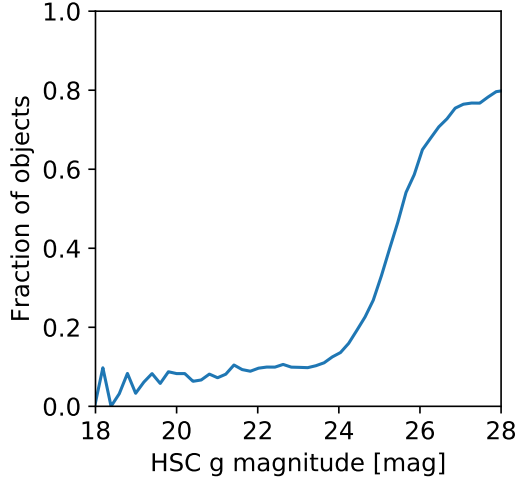
Combining data with a positional cross match clearly comes with some limitations compared to matched aperture photometry from homogenised imaging. These include the possibility of mis-associations, and differences in selection criteria meaning selection effects are difficult to model. In the following section we investigate the performance of the data and include some metrics to aid the user in characterising the quality of their sample.

#### 3.3.1 The positional cross matching

After all the pristine catalogues have been standardised we merge them together into the overall list of objects – the “*masterlist*”. This is done non-destructively: a cross-identification table can be used to return to the original tables meaning no information is discarded.

We start with the highest resolution optical data and add in subsequent surveys in order of increasing positional error. We first plot the number of pairs as a function of separation. We use this to determine a maximum cross match radius. This radius varies from data set to data set but is typically chosen to be 0.8 arcsec. The threshold is chosen to capture the majority of the true associations shown by the initial bulge in figure 7. Figure 7 shows the offsets of *masterlist* positions with respect to GAIA, which should reflect positional errors in general, and shows that most matches will be within 0.8 arcsec.

Where an object has multiple associations both are included, and flagged as possibly spurious objects, but the closest match by angular separation is associated. Two percent of objects across all HELP fields have multiple associations. Magnitudes were not used as a further matching criteria. These can be included as a further criteria by selecting objects that are flagged as merged, cross matching to each other to associate merged objects and replacing unmatched objects where the magnitudes are closer. Future data releases may use a more sophisticated cross-matching or homogenised imaging and forced photometry. Tight correlations between fluxes in similar bands from different instruments show this method produces accurate cross matches (see figure C1). Figure 10 shows the fraction of these merged objects which are unmatched as a function of magnitude. It increases rapidly with the faintest objects, which are most likely to be artefacts or undetected in another survey. The majority of objects flagged as merged will be these unmatched faint objects or the pair that they were within 0.8 arcsec of.



**Figure 10.** Fraction of HSC objects which are flagged as merged (associated to multiple other objects) but rejected as a counterpart based on the closest match criteria as a function of HSC  $g$  magnitude. These objects are unmatched and only have HSC fluxes. We expect the faintest HSC objects to be unmatched since HSC is often the deepest survey available over most of the area it exists.

**Table 2.** Summary of the HELP *masterlist* catalogue in each field.

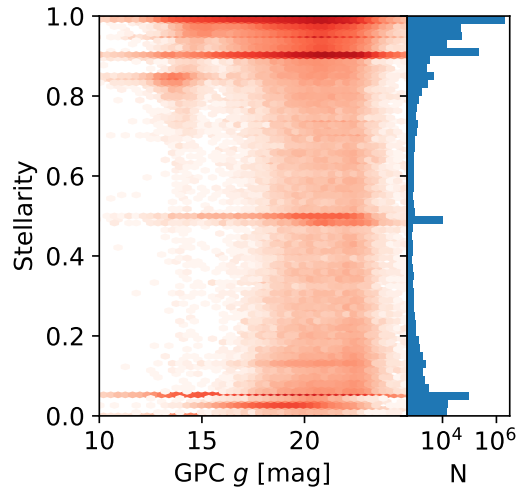
Field id	HELP field name	Objects	Area (deg. <sup>2</sup> )
1	AKARI-NEP	531 746	9.2
2	AKARI-SEP	844 172	8.7
3	Bootes	3 367 490	11
4	CDFS-SWIRE	2 171 051	13
5	COSMOS	2 599 374	5.1
6	EGS	1 412 613	3.6
7	ELAIS-N1	4 026 292	14
8	ELAIS-N2	1 783 240	9.2
9	ELAIS-S1	1 655 564	9.0
10	GAMA-09	12 937 982	62
11	GAMA-12	12 369 415	63
12	GAMA-15	14 232 880	62
13	HDF-N	130 679	0.67
14	Herschel-Stripe-82	50 196 455	363
15	Lockman-SWIRE	4 366 298	22
16	HATLAS-NGP	6 759 591	178
17	SA13	9 799	0.27
18	HATLAS-SGP	29 790 690	295
19	SPIRE-NEP	2 674	0.13
20	SSDF	12 661 903	111
21	xFLS	977 148	7.4
22	XMM-13hr	38 629	0.76
23	XMM-LSS	8 704 751	22
Total:		171 570 436	1270

### 3.3.2 Summary of the masterlist detections

A summary of the numbers of objects in each field in the *masterlist* is given in Table 2. Table 3 gives an overview of the numbers in the final merged catalogue across all HELP in relation to the selection criteria of objects for subsequent HELP processing.

**Table 3.** Overview of object detections across all of HELP. The main list of `XID+` prior objects must be IRAC 1 detected and satisfy one of either having at least two optical detections, or at least 2 near-infrared detections. This criteria is only applied on areas observed by IRAC 1. On other areas an SED modelling method is used and will be discussed in Oliver et al. (in prep).

Total number of objects	171 570 436
Observed in all wavelength regimes	56 618 577
$\geq 2$ optical detections	118 750 890
$\geq 2$ near-infrared detections	26 499 180
$\geq 2$ optical and $\geq 2$ near-infrared detections	22 065 144
<code>XID+</code> prior objects	11 594 158



**Figure 11.** The distribution of stellerity index, which should reflect point-sourceness (definitely point like = 1, definitely extended = 0) with GPC1  $g$  flux for GAIA objects. Approximately one percent are incorrectly labeled as extended.

### 3.3.3 Stellerity indices and quality control flags

The catalogue contains flags to identify GAIA stars, objects that have had duplicates cleaned and objects that may have a degenerate cross match pairing.

The stellerity is computed by taking the largest value from all the pristine catalogues. This is done on a scale where 0 represents a definitely extended object and 1 represents a definite point source. This conservative approach ensures we mark objects which are point-source like in any band. Figure 11 shows the distribution of stellerity values for GAIA objects (which should all be point like and have stellerity equal to one) with GPC1  $g$  flux. This shows that approximately one percent are incorrectly labelled as extended.

Two further flags are added to aid removal of artefacts. First we include a flag to indicate which wavelength regimes were observed at this position. This flag records whether this position on the sky was observed by any optical survey, any near-infrared survey and any mid-infrared survey. This is necessary to determine whether an absence of measurement is because the source is too faint to be detected or a given position has not been observed.

Another flag describes whether the object was detected in at least two optical bands, at least two near-infrared bands

and at least two mid-infrared bands. These flags are used in later HELP products to remove artefacts.

### 3.3.4 Dealing with multiple measurements

In the *masterlist* catalogue we want exactly one measurement per instrument-band. We decided to choose the lowest flux error measurement rather than combining. Our rationale is that it is not often obvious how to combine measurements rigorously and in some cases the data may not be independent. This approach means we also preserve a clear record of which survey was used for each individual object.

An example of multiple measurements can be found on ELAIS-N1 where there are two IRAC surveys available. We investigated the distribution of errors and decided to use the deeper SERVS where it is available and the SWIRE otherwise. The depth maps described later allow us to deal with the varying depth of the resulting catalogue.

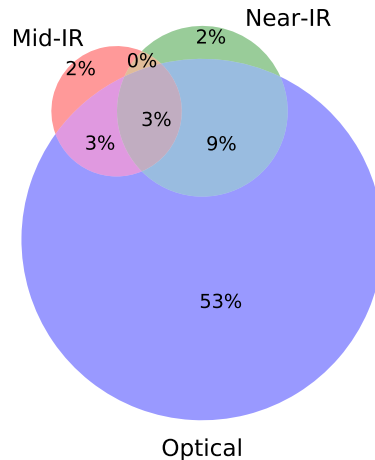
## 4 VALIDATING THE FINAL CATALOGUE

In this section we will provide details of the quality of the data along with a description of some flags used to warn the user about questionable objects or photometry values. The checks and diagnostics are automatically run following the production of the *masterlist*. They were used extensively in the debugging and testing of the pipeline, consist of thousands of figures and tables and are all available on GitHub. We will show a small number of examples from ELAIS-N1 to demonstrate their utility for understanding the data.

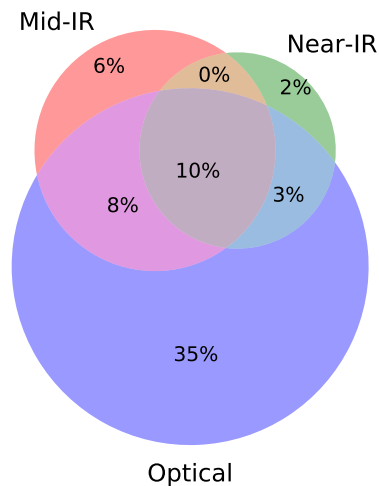
### 4.1 Photometry and stellarity

After the production of the *masterlist* a number of checks and diagnostics are performed. During development of the code these facilitated the fixing of errors and validation of the data. Here, we present a discussion of each check performed and how it revealed data issues as we were constructing the software pipeline. This final stage was a crucial aspect of debugging earlier stages in the pipeline and code was developed over many iterations particularly where a full description of input data was hard to find or incomplete. We look at the numbers of objects which have detections in each combination of wavelength regimes as shown in figure 12 for all objects and in figure 13 for objects in regions with surveys in all wavelength regimes. The majority of objects have optical detections only due to the relative depths of each band compared to typical galaxy SEDs.

The most fundamental check is to compare magnitudes where multiple surveys give photometry for similar bands. We compare every possible combination to look for data sets that might not be compatible. On all the northern fields we have as a minimum SDSS fluxes to compare to. An example of comparing fluxes in similar bands is shown in figure C1 in appendix C. For a small number of southern fields we may not have duplicate measurements for every band so comparisons are impossible. We also compare total magnitudes to aperture magnitudes as shown in figure C2 to check that point sources are in strong agreement. This also functions as



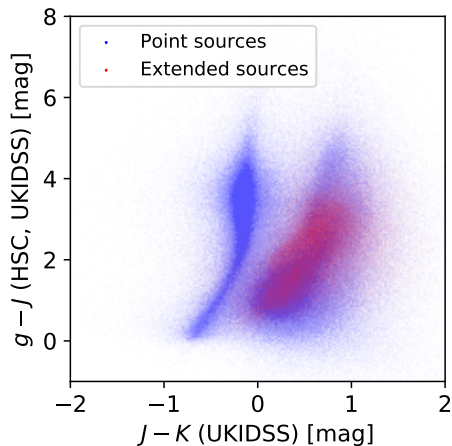
**Figure 12.** Overview of object detections across all of HELP. Optical and near-infrared objects must have at least two detections in the given wavelength regime.



**Figure 13.** Overview of objects in the area, across all HELP fields, that has been observed by an optical, a near-infrared and a mid-infrared survey that have been detected in each of those regimes. The majority of objects surveyed in all wavelength regimes are only detected in an optical survey, which tend to be the deepest.

a check for the stellarity measure which we take as the highest value of the stellarity of each input survey. If the stellarity measures were poor, we would not see a clear distinction between point sources (stellarity  $> 0.7$ ) and extended objects (stellarity  $< 0.7$ ). We also plot basic number counts for every band as a further check of units and numbers of objects.

In the *masterlist* the stellarity parameter gives the probability that the source is extended or point-like. In order to check the accuracy of this parameter we can compare the total and aperture magnitude. As we can see in figure C2, point-source objects tend to have similar aperture and total magnitudes, while for extended objects (i.e. galaxies), we get a much larger scatter. However, the stellarity parameter



**Figure 14.** Colour-colour plot showing star galaxy separation for all objects with HSC  $g$  and UKIDSS  $K$  and  $J$  fluxes across all HELP. This also serves as a check on data quality for all sets of  $g$ ,  $J$ , and  $K$  during catalogue production. These are plotted for a large number of colour-colour pairs in the validation section where large outliers can reveal problems with one of the bands.

only gives an indication of the morphological shape of the objects. The separation between stars and quasars/galaxies is usually done using optical colours. The addition of near-infrared colours allows us to reach higher redshift. We have used up to five different colour-colour diagrams, when all wavebands were available to look at the classification of the *masterlist*:

- Stars and galaxies can be separated using  $J - K$  vs  $g - i$  colours (Baldry et al. 2010).
- Stars can be distinguished from quasars/galaxies based on optical colours only with  $g - r$  vs  $u - g$  (Chiu et al. 2007) or with a mix of optical and near-infrared colours with  $g - J$  vs  $J - K$  (Maddox et al. 2008) seen in figure 14.
- Stars, galaxies and quasars can also be identified on a  $g - i$  vs  $i - W1$  colour-colour diagram (Tie et al. 2017).
- Finally, a standard mid-infrared colour-colour diagram can be used to identify active galactic nuclei (Stern et al. 2005; Donley et al. 2012; Lacy et al. 2013).

As shown in figure 14, we observe two distributions on the  $g - J$  vs  $J - K$  colour-colour diagram. On one side, we have the point-source objects which correspond to the stars and on the other side we have the extended objects with some point-like sources which are galaxies and quasars respectively.

## 4.2 Flags

We provide some flags to aid the user in determining samples that can be used for science purposes. These are not exhaustive and merely provide some well defined metrics that will be correlated with measurement quality. We compute  $\chi^2$ ,

$$\chi^2 = \frac{(M_1 - M_2)^2}{(\sigma_1^2 + \sigma_2^2)} \quad (1)$$

where  $M_x$  are the magnitudes of similar bands in two different surveys and  $\sigma_x$  are their corresponding errors. We flag

as outliers all objects more than  $5\sigma$  from the mean. If using fluxes we recommend rerunning the notebooks with flux comparisons rather than magnitude comparisons. It can be shown that this criteria is equivalent to any pair of measurements in comparable bands having:

$$\chi^2 > P_{75\text{th}} + 3.2 \times (P_{75\text{th}} - P_{25\text{th}}) \quad (2)$$

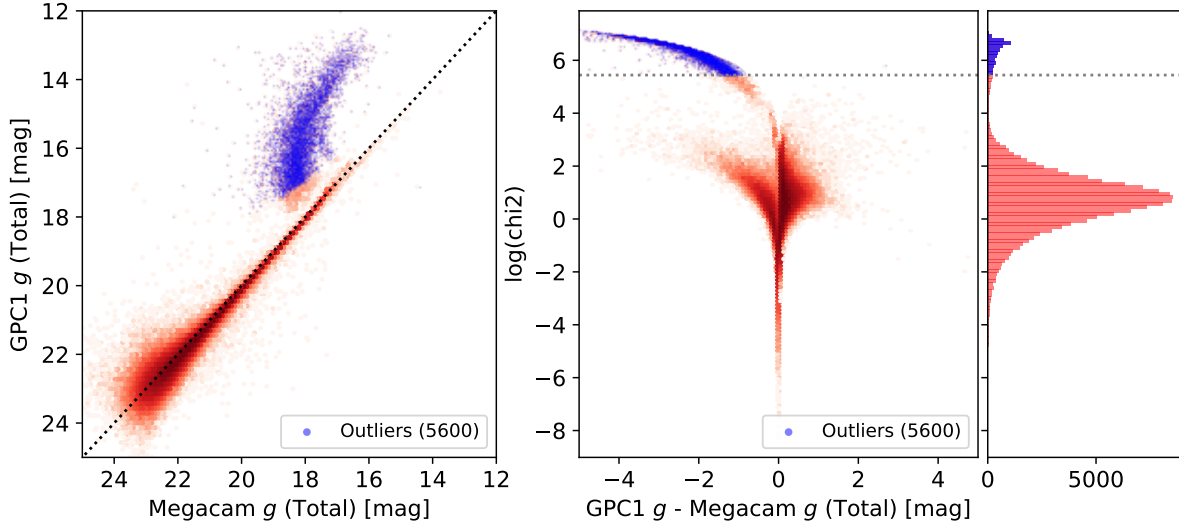
where  $P_{75\text{th}}$  and  $P_{25\text{th}}$  are the 75th and 25th percentiles respectively. Bright sources tend to have their errors underestimated with values as low as  $10^{-6}$  mag, which is unrealistic. To avoid high  $\chi^2$  due to these unrealistically small errors we clip the error to get a minimum value of  $10^{-3}$  mag. An example of this method being applied is shown in figure 15.

The vast majority of flagged objects (50 thousand compared to less than ten) are due to disagreements between deep MegaCam and DECam surveys, and the wide and shallow PanSTARRS. Since in these cases there is no disagreement between PanSTARRS fluxes and numerous other surveys we believe these to be due to saturated pixels in the MegaCam and DECam surveys. We therefore recommend that for all objects with magnitude less than 16 that DECam or MegaCam fluxes are rejected in favour of fluxes from the shallower PanSTARRS if available. If this is not done then the fluxes are systematically underestimated and errors are inaccurate. Figure 16 which shows the PanSTARRS fraction of flagged objects as a function of magnitude shows how, above 18, zero objects are flagged. Since the flagging procedure is blind to which of the two fluxes has the inaccurate flux and error the PanSTARRS fluxes are also flagged in this situation.

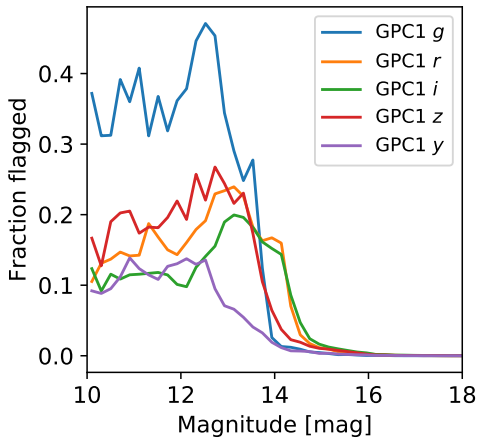
We also implement some specific flags based on finding anomalies in the the pristine catalogues. For instance in the PanSTARRS catalogue we noticed that a number of objects had exactly the same error which is a concern. These objects are flagged. One of the major concerns with a cross matched catalogue will be the number of mis-associations. In addition to the flagging based on  $\chi^2$  between measurements in similar bands we also apply a flag to objects that have multiple associations. We take the closest object as the true association and flag all objects within 0.8 arcsec as possible mis-associations.

## 5 DEPTH MAPS AS SELECTION FUNCTIONS

One area of complexity is that each individual survey is subject to a different selection function; the set of criteria which determine whether a given galaxy will be in the final catalogue. Characterising this selection function from the pristine catalogues will aid the understanding of the sample of objects required for a given statistical analysis; such as using completeness to weight objects in the computation of luminosity functions (e.g. Loveday et al. 2015). As a means to describe the catalogue selection function we present ‘depth maps’ over the field allowing one to select a sub sample according to some criteria of depth in a given band from a given instrument. For individual surveys, authors may provide measures of depth and completeness based on simulations or comparisons to deeper surveys. However, in general we may not have access to such information and a means to



**Figure 15.** Comparison between Canada France Hawaii Telescope MegaCam and PanSTARRS1 GigaPixel Camera (GPC1)  $g$ -band magnitudes on the example field ELAIS-N1. The blue points show those objects flagged as being outliers. They form a coherent group amongst the bright objects. We believe these are due to saturated pixels for bright objects in the SpARCS survey. The two figures to the right show the distribution of  $\chi^2$  values as a function of magnitude and how the problematic fluxes form a distinct population.



**Figure 16.** Fraction of flagged measurements in each of the five PanSTARRS1 GigaPixel Camera (GPC1) *grizy* bands as a function of GPC1  $g$  magnitude across all HELP. We believe the vast majority of these flags to be due to saturated pixels in the deep DECam and MegaCam surveys, which are not optimised for bright objects.

estimate and model survey depths and completeness from the catalogues only is highly advantageous.

To model the selection function of the catalogue we would like to know the completeness in a given band at a given location on the sky as a function of object flux. This is done using regions described by Hierarchical Equal Area Iso-Latitude pixelation of the sphere (HEALPix Gorski et al. 1999), which describes a pixelisation of the sky at varying scales or ‘orders’. The sphere is divided into twelve parallelograms at order zero and each subsequent order divides each parallelogram into four. The choice of order is a compromise between producing a high resolution map and having enough objects in each pixel to achieve reliable statistics.

To facilitate this, and as a first step, we provide a depth at every order 10 HEALPix cell ( $0.003 \text{ deg.}^2$ ) which we take as the average error on flux on that pixel. This is done for every band for both total and aperture magnitudes. This assumes that the errors are dominated by the low flux objects, which is consistent with typical number counts but clearly there are differences between these average errors and the error on a zero flux object. Objects are typically selected according to some signal to noise criteria. Therefore errors in the total flux of an object will clearly be related to selection criteria. Nevertheless actual selection in source extraction software will depend on individual pixel measurements which are not available to us. These non-linear effects are difficult to model when we only have access to fluxes and flux errors. Since only the bands used for detection impact the selection function, the depth maps produced for bands that are not detection bands will be only correlated with the true depths to the extent that the fluxes between the bands are correlated. Further, for regions where detection is performed on a  $\chi^2$  image, the depth maps produced here will only reflect correlations between flux in a given band and the  $\chi^2$  value.

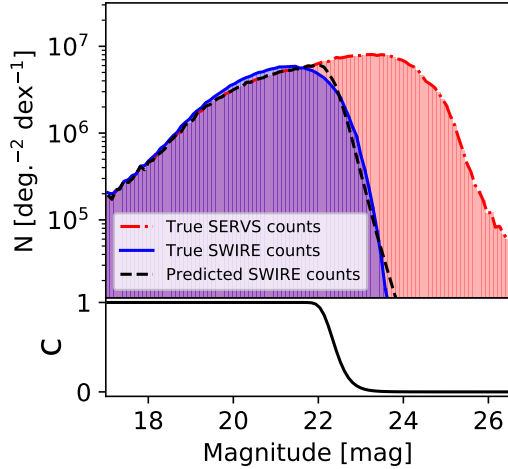
In the method presented here completeness is given by the probability of detecting an object given the true flux,  $f_{\text{true}}$ , in terms of the measured flux,  $f_{\text{measured}}$  and the signal to noise cut,  $n\sigma_{\text{mean}}$ .

$$P(\text{detection} | f_{\text{true}}) = P(f_{\text{measured}} > n\sigma_{\text{mean}}) \quad (3)$$

where  $n$  is determined by the survey. This measured flux is modelled by assuming Gaussian errors on the true flux such that the completeness is given by:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt \quad (4)$$

where  $\Phi(x)$  is the completeness as a function of the measured



**Figure 17.** IRAC *i1* number counts on the example ELAIS-N1 field for the deeper SERVS survey and the shallower SWIRE survey in addition to predicted SWIRE counts based on our depth map selection function from SWIRE depths. Shown beneath is the SWIRE completeness,  $C$ , as calculated using the depth model in equation 4. By comparing SWIRE against the deeper SERVS we see that this method produces reasonable completeness curves.

flux in terms of a standard normal distribution such that:

$$x = \frac{n\sigma_{\text{mean}} - f_{\text{true}}}{\sigma_{\text{mean}}} \quad (5)$$

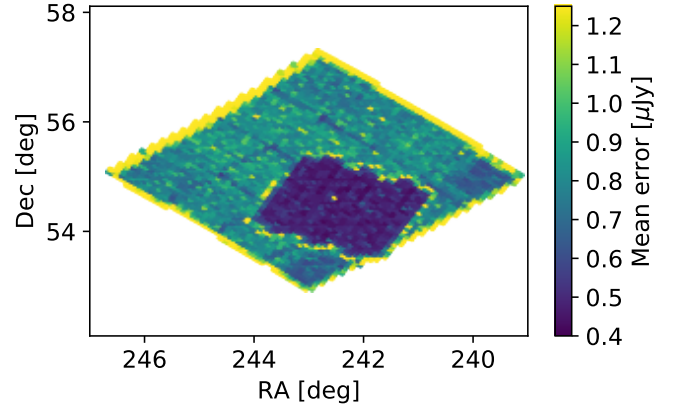
and the dummy variable,  $t$ , is given by

$$t = \frac{f_{\text{measured}} - f_{\text{true}}}{\sigma_{\text{mean}}} \quad (6)$$

We experimented with various values of  $n$  between 3 and 5. Fitting  $n$ , by applying a single Gaussian cumulative distribution function to the SERVS number counts to recreated SWIRE yields  $n \approx 5$ . This is reassuring and, especially given the relative insensitivity to  $n$ , demonstrates that in general we can set it to 5 under the assumption that was the criteria used by the survey.

Investigations into the region covered by both SERVS and SWIRE indicate that we can successfully model the completeness of the shallower SWIRE by comparing to the ‘true’ SERVS number counts, which should not severely suffer from incompleteness at typical SWIRE depths. Figure 17 shows a comparison of SERVS number counts to those of SWIRE. By applying our modelled SWIRE completeness to the SERVS number counts we roughly recreate the SWIRE number counts. Moreover by freely fitting the depth criterion between  $3\sigma$  and  $5\sigma$  we see that completeness is relatively insensitive to the value such that assuming a  $5\sigma$  cut and taking the mean errors allows the computation of reasonable completeness curves. Completeness models based on simulations which add artificial sources to real images and test retrieval will be superior to these estimates based purely on catalogues; however this method does allow one to roughly characterise depths and to set conservative values for producing magnitude limited samples.

Figure 18 shows a map of the example field ELAIS-N1 with colour signifying mean error as a proxy for depth. Depths measured with this method on the COSMOS field are consistent with the depths shown in figure 3 in (Laigle



**Figure 18.** Map of the example ELAIS-N1 field showing the mean error on the IRAC *i1* flux over order 10 HEALPix cells. The mean error, which can be used as a proxy for depth, makes it possible to take account of selection effects over large areas with highly inhomogeneous observation depths. The smaller and deeper SERVS area is clearly visible against the wider and shallower SWIRE area.

et al. 2016), which were computed from empty apertures. These depth maps are currently being used to develop a method for computing the comoving volume over which the galaxy could be detected,  $V_{\text{max}}$ , used for calculating a luminosity function that uses a different maximum redshift at which a galaxy can be detected,  $z_{\text{max}}$ , for every position on the sky.

The differences between filter transmission profiles for individual surveys will introduce systematics. Using the filter transmission curves it is possible to calculate galaxy fluxes as a function of redshift and position on the sky given which survey has observed that position and thus account for this to some extent. The decision to use 2 arcsec apertures will also feed through to the depth map values and have an effect on systematics where the completeness is far from unity. The depth maps can usefully be used in this context to set magnitude limits for the construction of samples with high completeness.

## 5.1 Overview of depths and number counts

The final HELP selection function depends on optical, near-infrared and mid-infrared detections. We therefore present a number of graphs generated from the depth maps here to summarise the key depths available across HELP. We also show summaries of the number counts in  $g$  and  $K$  or  $K_s$  to verify the depth maps reflect the numbers detected. Figure 19 shows the distribution of depths over HEALPix cells for each of the main broad bands available across all of HELP. Figure 20 shows the distribution of depths of area on the sky for each field individually given an overview of the variation in the depths and coverage by a given type of broad band filter. We also provide individual summaries of important detection bands; figure 21 shows the cumulative area  $g$  coverage. This is made from the depth maps by taking the deepest  $g$  for every HEALpix cell. Figure 22 shows the cumulative area  $K$  or  $K_s$  coverage. This is made from

the depth maps by taking the deepest  $K$  or  $K_s$  for every HEALpix cell. Together, these have coverage over the 1146 deg.<sup>2</sup> of HELP, and some coverage on all but five fields. Figure 23 shows the cumulative area IRAC  $i1$  coverage. The depth percentiles quoted in the abstract correspond to the distribution of depths by area covered on the sky as displayed in these figures.

We also provide cumulative depth plots for bands  $g$  (figure 21),  $K$  or  $K_s$  (figure 22) and IRAC  $i1$  (figure 23) for a more detailed overview of the depths available in these bands.

We also provide figures showing the number counts for each field. Figure 24 highlights the COSMOS field and compares to the counts from the following sources McCracken et al. (2009); Aihara et al. (2011); Bielby et al. (2012); Fontana et al. (2014); Ilbert et al. (2015); Laigle et al. (2016). Figures 25 and 26 which show the differential number counts on each field for  $g$  and either  $K$  or  $K_s$ . The areas used to compute these are the total area over which a given band is available on a given field. For this reason there can be two peaks where a given survey has different areas with varying depths.

## 6 CONCLUSIONS

We have presented a new multi-wavelength catalogue across the well known and well studied extragalactic fields that were targeted by the *Herschel* Space Observatory. This new catalogue will be of general use to extragalactic astronomers and forms the basis of all data products currently being produced for the *Herschel* Extragalactic Legacy Project (HELP). This catalogue defines the list of objects that will comprise the first data release from the project (HELP-DR1). We have described a method for producing and testing multi-wavelength catalogues from disparate and inhomogeneous surveys for use in wide area astronomy. We discussed a number of problems involved with collating data of differing quality and production methods. The resulting catalogue will be used in various upcoming projects including physical modelling and forced photometry from low resolution *Herschel* imaging.

To summarise, the paper presents:

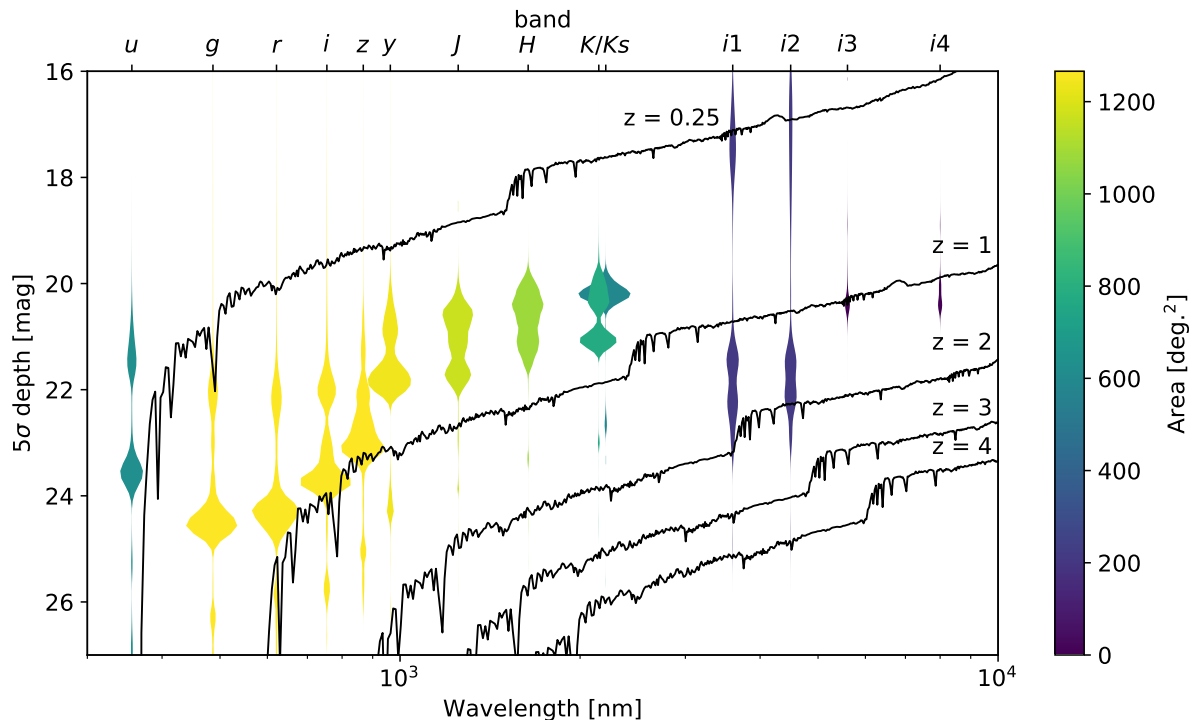
- A new multi-wavelength catalogue on all HELP fields covering 1270 deg.<sup>2</sup>, with astrometry corrected to the GAIA frame and a fully documented positional cross match, that is reproducible and extensible in an open science framework.
- A full description of the catalogue production methods that have been applied across all HELP fields. The code used to perform this reduction in addition to thousands of diagnostic plots are provided in JUPYTER notebooks that are available to download. We provide summary plots showing data quality across the whole HELP coverage. Equivalent plots for individual fields are available in the open access notebooks.
- Diagnostics and depth maps that demonstrate and measure the quality and limitations of the data in addition to flags to warn the user about possibly spurious objects. The depth maps can be used to model selection functions, facilitating the construction of a well understood sample of objects for statistical analysis.

## ACKNOWLEDGEMENTS

The research leading to these results has received funding from the European Union Seventh Framework Programme FP7/2007-2013/ under grant agreement no.607254. This publication reflects only the authors' view and the European Union is not responsible for any use that may be made of the information contained therein. We are grateful to the anonymous referee for comments that have lead to a significant improvement in the presentation of these results. Raphael Shirley acknowledges support from the Daphne Jackson Trust. Katarzyna Małek has been partly supported by the National Science Centre (grant UMO-2018/30/E/ST9/00082). Lucia Marchetti and Mattia Vaccari have been partially supported by the South African Department of Science and Technology (DST/CON 0134/2014).

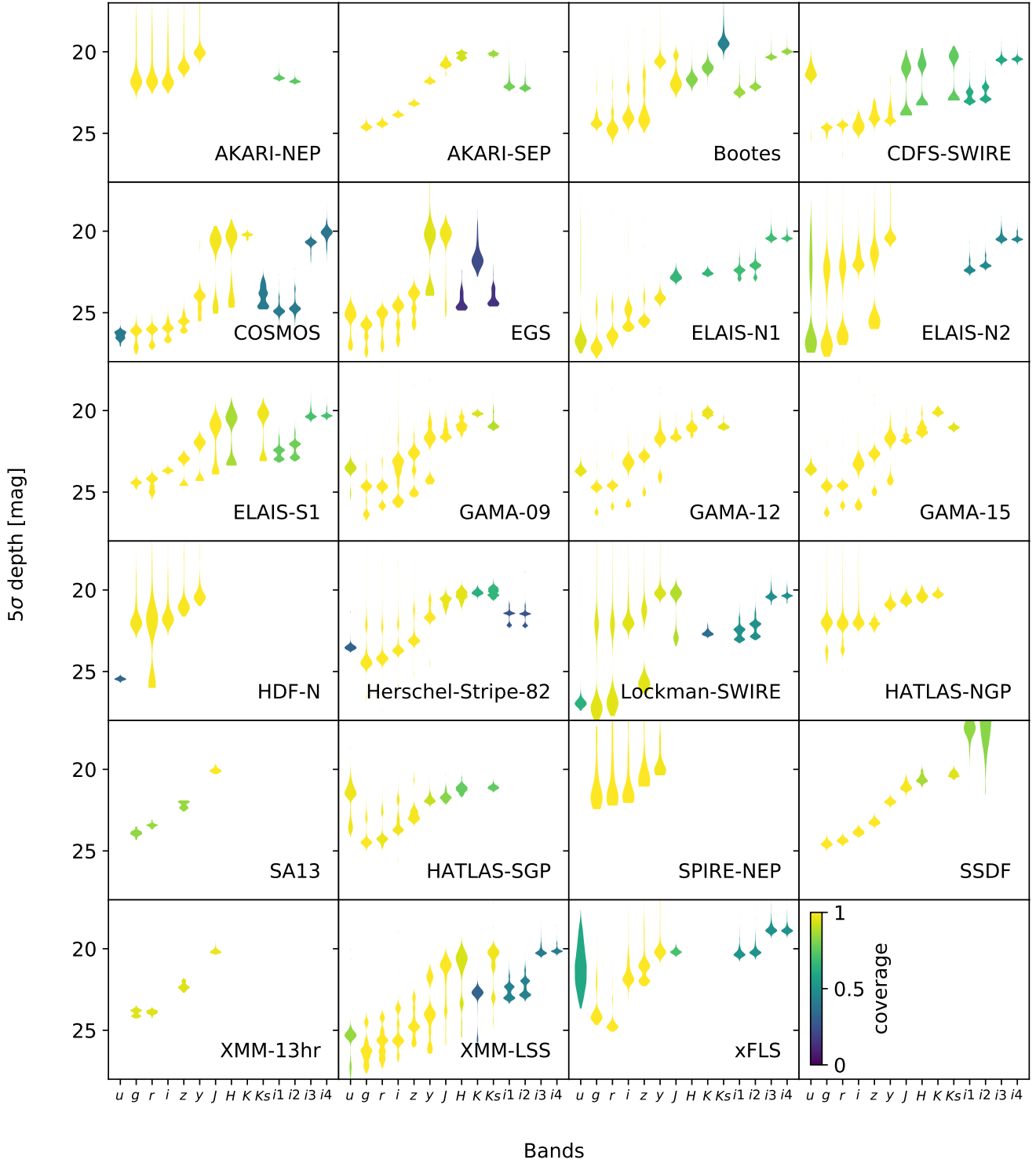
## References

- Abbott T., et al., 2018, *The Astrophysical Journal Supplement Series*, 239, 18
- Aihara H., et al., 2011, *The Astrophysical Journal Supplement Series*, 193, 29
- Aihara H., et al., 2018, *Publications of the Astronomical Society of Japan*, 70, S4
- Albareti F. D., et al., 2017, *ApJS*, 233, 25
- Almaini O., et al., 2007, in Metcalfe N., Shanks T., eds, *Astronomical Society of the Pacific Conference Series Vol. 379, Cosmic Frontiers*. p. 163
- Ashby M. L. N., et al., 2009, *ApJ*, 701, 428
- Ashby M. L. N., et al., 2013, *ApJS*, 209, 22
- Baldry I. K., et al., 2010, *MNRAS*, 404, 86
- Baronchelli I., et al., 2016, *The Astrophysical Journal Supplement Series*, 223, 1
- Barro G., et al., 2011, *ApJS*, 193, 13
- Bertin E., Arnouts S., 1996, *A&AS*, 117, 393
- Bielby R., et al., 2012, *A&A*, 545, A23
- Blum R. D., et al., 2016, in *American Astronomical Society Meeting Abstracts #228*. p. 317.01
- Boquien M., Burgarella D., Roehly Y., Buat V., Ciesla L., Corre D., Inoue A., Salas H., 2019, *Astronomy & Astrophysics*, 622, A103
- Brammer G. B., van Dokkum P. G., Coppi P., 2008, *ApJ*, 686, 1503
- Buat V., Boquien M., Małek K., Corre D., Salas H., Roehly Y., Shirley R., Efstathiou A., 2018, *Astronomy & Astrophysics*, 619, A135
- Bundy K., et al., 2006, *ApJ*, 651, 120
- Capak P., et al., 2004, *AJ*, 127, 180
- Chambers K. C., et al., 2016, preprint, ([arXiv:1612.05560](https://arxiv.org/abs/1612.05560))
- Chiu K., Richards G. T., Hewett P. C., Maddox N., 2007, *MNRAS*, 375, 1180
- Coil A. L., Newman J. A., Kaiser N., Davis M., Ma C.-P., Kocevski D. D., Koo D. C., 2004, *ApJ*, 617, 765
- Cool R. J., 2007, *ApJS*, 169, 21
- Cutri R. M., et al., 2003, *VizieR Online Data Catalog*, 2246
- Demleitner M., Neves M. C., Rothmaier F., Wambsganss J., 2014, *Astronomy and Computing*, 7, 27
- Dey A., et al., 2019, *The Astronomical Journal*, 157, 168
- Donley J., et al., 2012, *The Astrophysical Journal*, 748, 142
- Driver S. P., et al., 2011, *Monthly Notices of the Royal Astronomical Society*, 413, 971
- Duncan K. J., et al., 2018a, *MNRAS*, 473, 2655
- Duncan K. J., Jarvis M. J., Brown M. J., Röttgering H. J., 2018b, *Monthly Notices of the Royal Astronomical Society*, 477, 5177

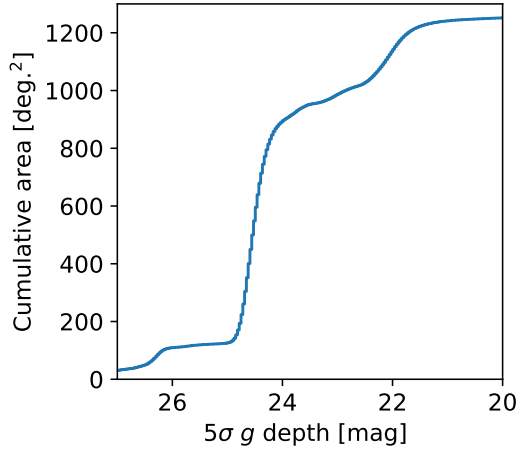


**Figure 19.** The distribution of  $5\sigma$  point-source depths in a 2 arcsec aperture for each broad band type (taking the deepest specific band available in a given HEALPix cell). The colour of each area is determined by the total area that data for that band is available. The areas of each probability density function are also weighted by the total area available for that band so that area in a given curve is proportionally related to area on the sky that a given band is available at a given depth. The combination of areas covered by either the  $K$  or  $K_s$  bands is the full  $1270 \text{ deg.}^2$  of HELP. We also plot a typical HELP spectral energy distribution for a galaxy with star formation rate of  $200 M_{\odot}/\text{yr}$  and a stellar mass of  $10^{10} M_{\odot}$  at various redshifts.

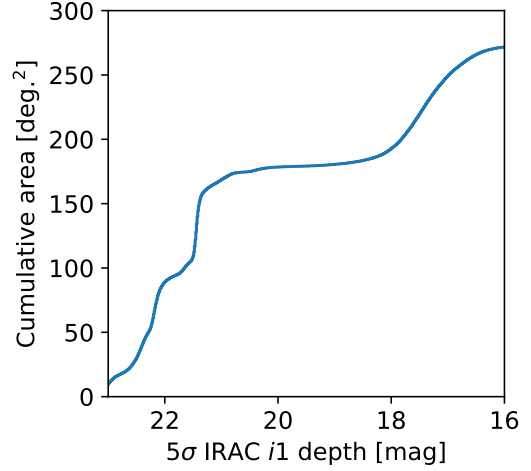
- Dunlop J., et al., 2007, A Spitzer Public Legacy survey of the UKIDSS Ultra Deep Survey, Spitzer Proposal
- Dye S., et al., 2018, *MNRAS*, **473**, 5113
- Edge A., Sutherland W., Kuijken K., Driver S., McMahon R., Eales S., Emerson J. P., 2013, *The Messenger*, **154**, 32
- Erben T., et al., 2013, *MNRAS*, **433**, 2545
- Fadda D., Jannuzi B. T., Ford A., Storrie-Lombardi L. J., 2004, *AJ*, **128**, 1
- Fliri J., Trujillo I., 2016, *MNRAS*, **456**, 1359
- Fontana A., et al., 2014, *Astronomy & Astrophysics*, 570, A11
- Furusawa H., et al., 2008, *ApJS*, **176**, 1
- Gaia Collaboration et al., 2016, *A&A*, **595**, A2
- Geach J. E., et al., 2017, *ApJS*, **231**, 7
- Giavalisco M., et al., 2004, *ApJ*, **600**, L93
- González-Solares E. A., et al., 2011, *MNRAS*, **416**, 927
- Gorski K. M., Wandelt B. D., Hansen F. K., Hivon E., Banday A. J., 1999, *ArXiv Astrophysics e-prints*,
- Grogin N. A., et al., 2011, *ApJS*, **197**, 35
- Heymans C., et al., 2012, *MNRAS*, **427**, 146
- Hildebrandt H., et al., 2016, *MNRAS*, **463**, 635
- Hudelot P., et al., 2012, *VizieR Online Data Catalog*, 2317
- Hurley P. D., et al., 2017, *Monthly Notices of the Royal Astronomical Society*, 464, 885
- Ilbert O., et al., 2013, *VizieR Online Data Catalog*, 355
- Ilbert O., et al., 2015, *Astronomy & Astrophysics*, 579, A2
- Ivezic Z., et al., 2008, preprint, ([arXiv:0805.2366](https://arxiv.org/abs/0805.2366))
- Jannuzi B. T., Dey A., NDWFS Team 1999, in *American Astronomical Society Meeting Abstracts*. p. 1392
- Jannuzi B. T., et al., 2004, in *American Astronomical Society Meeting Abstracts*. p. 1478
- Jarvis M. J., et al., 2013, *MNRAS*, **428**, 1281
- Kluyver T., et al., 2016, in Loizides F., Schmidt B., eds, *Positioning and Power in Academic Publishing: Players, Agents and Agendas*. IOS Press, pp 87–90, <https://eprints.soton.ac.uk/403913/>
- Lacy M., et al., 2013, *The Astrophysical Journal Supplement Series*, 208, 24
- Laigle C., et al., 2016, *ApJS*, **224**, 24
- Lawrence A., et al., 2007, *Monthly Notices of the Royal Astronomical Society*, 379, 1599
- Lonsdale C. J., et al., 2003, *Publications of the Astronomical Society of the Pacific*, 115, 897
- Loveday J., et al., 2015, *MNRAS*, **451**, 1540
- Maddox N., Hewett P. C., Warren S. J., Croom S. M., 2008, *MNRAS*, **386**, 1605
- Malek K., et al., 2018, *Astronomy & Astrophysics*, 620, A50
- Mauduit J.-C., et al., 2012, *Publications of the Astronomical Society of the Pacific*, 124, 714
- McCracken H., et al., 2009, *The Astrophysical Journal*, 708, 202
- McMahon R. G., Banerji M., Gonzalez E., Kuposov S. E., Bejar V. J., Lodieu N., Rebolo R., VHS Collaboration 2013, *The Messenger*, **154**, 35
- Mehta V., et al., 2017, preprint, ([arXiv:1711.05280](https://arxiv.org/abs/1711.05280))
- Moutard T., et al., 2016, *A&A*, **590**, A102
- Noll S., Burgarella D., Giovannoli E., Buat V., Marcellac D., Munoz-Mateos J., 2009, *Astronomy & Astrophysics*, 507, 1793
- Ocran E. F., Taylor A. R., Vaccari M., Green D. A., 2017, *MNRAS*, **468**, 1156
- Oi N., et al., 2014, *VizieR Online Data Catalog*, 356
- Oliver S., et al., 2000, *MNRAS*, **316**, 749
- Papovich C., et al., 2016, *ApJS*, **224**, 28
- Pilbratt G. L., et al., 2010, *A&A*, **518**, L1



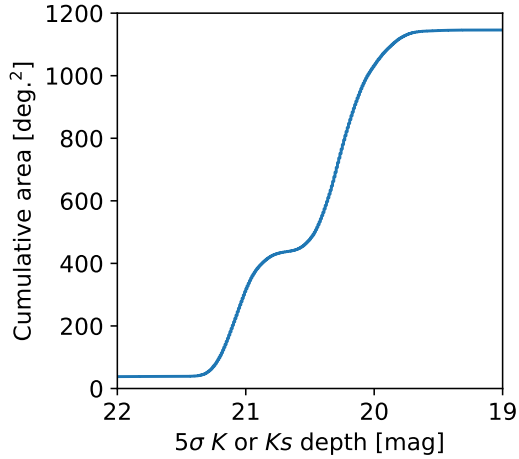
**Figure 20.** Overview of  $5\sigma$  point-source aperture depths on each respective field showing the variation in coverage and depth over the fields. The violin plots show the distribution of pixel depths and thus show distributions by area. For each pixel we take the deepest specific band available. The depth maps are available for each specific band. These bands are evenly spaced for clarity and not positioned by wavelength as in figure 19. Colour represents coverage and is given as a fraction of field area. All fields have *grizy* coverage everywhere.



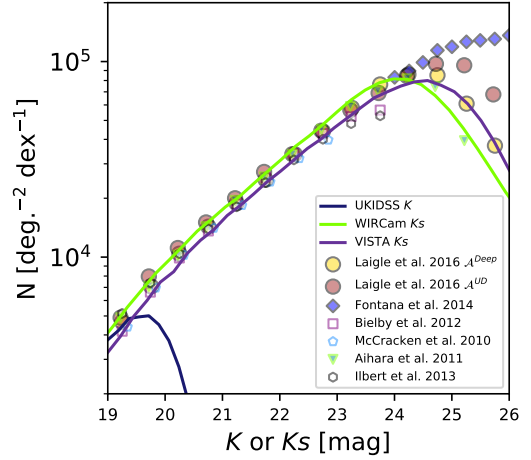
**Figure 21.** The area of HELP coverage with  $5\sigma$  point-source  $g$  depth above a given value. The queries for generating this plot from the Virtual Observatory at SusseX (VOX) are available on GitHub.



**Figure 23.** The area of HELP coverage with  $5\sigma$  point-source IRAC  $i1$  band depth above a given value. IRAC  $i1$  is available over 273 deg.<sup>2</sup>. Building the  $\chi_{ID}+$  priors on these regions requires flux prediction techniques that will be discussed in the upcoming paper by Oliver et al. Figures 21 to 23 play a key role in the selection of objects for HELP processing and are a necessary component of modelling selection.



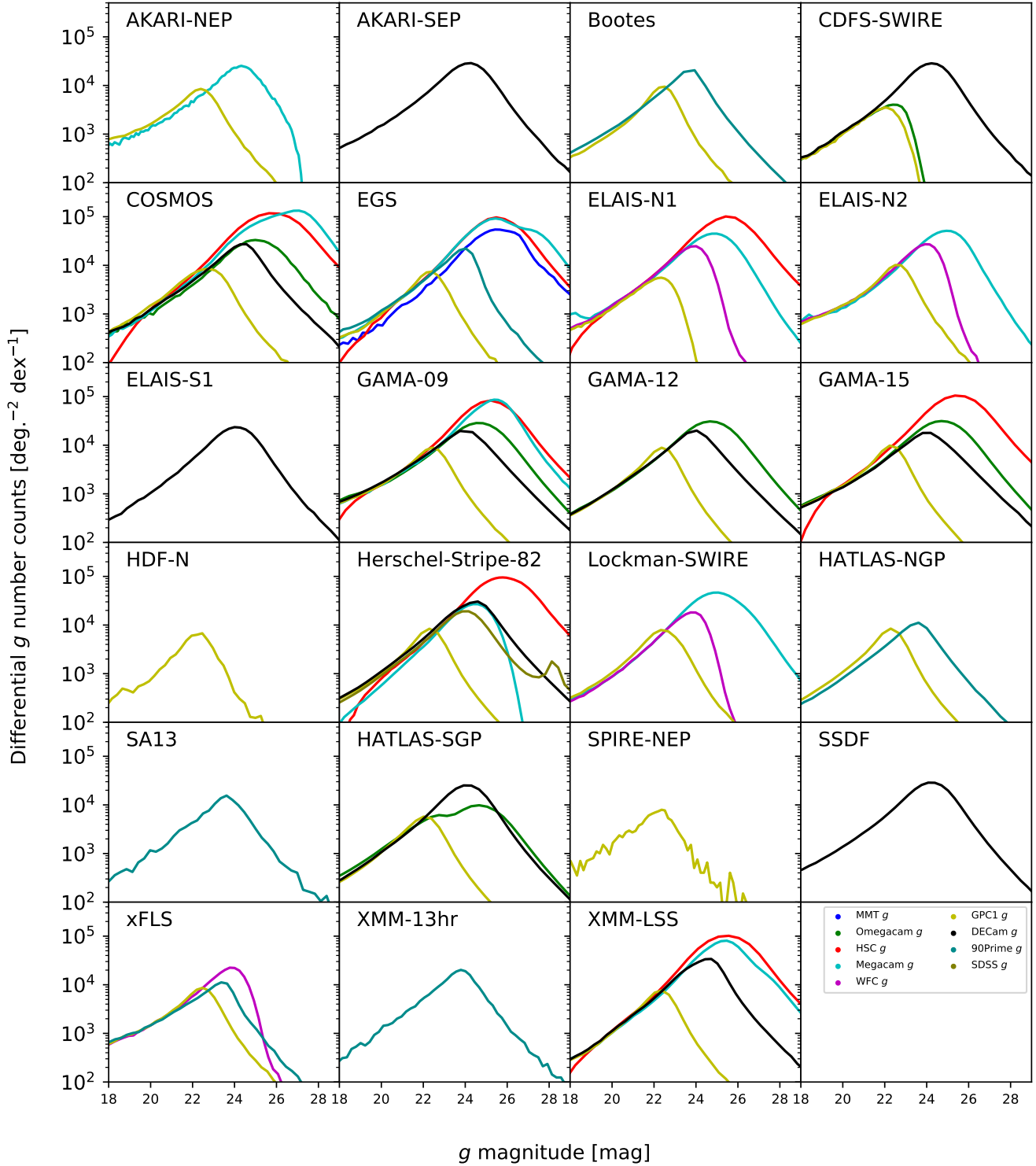
**Figure 22.** The area of HELP coverage with any  $5\sigma$  point-source  $K$  or  $Ks$  depth above a given value. The depth map product has the mean errors for all  $K$  and  $Ks$  bands allowing us to query the lowest value on every HEALpix cell.



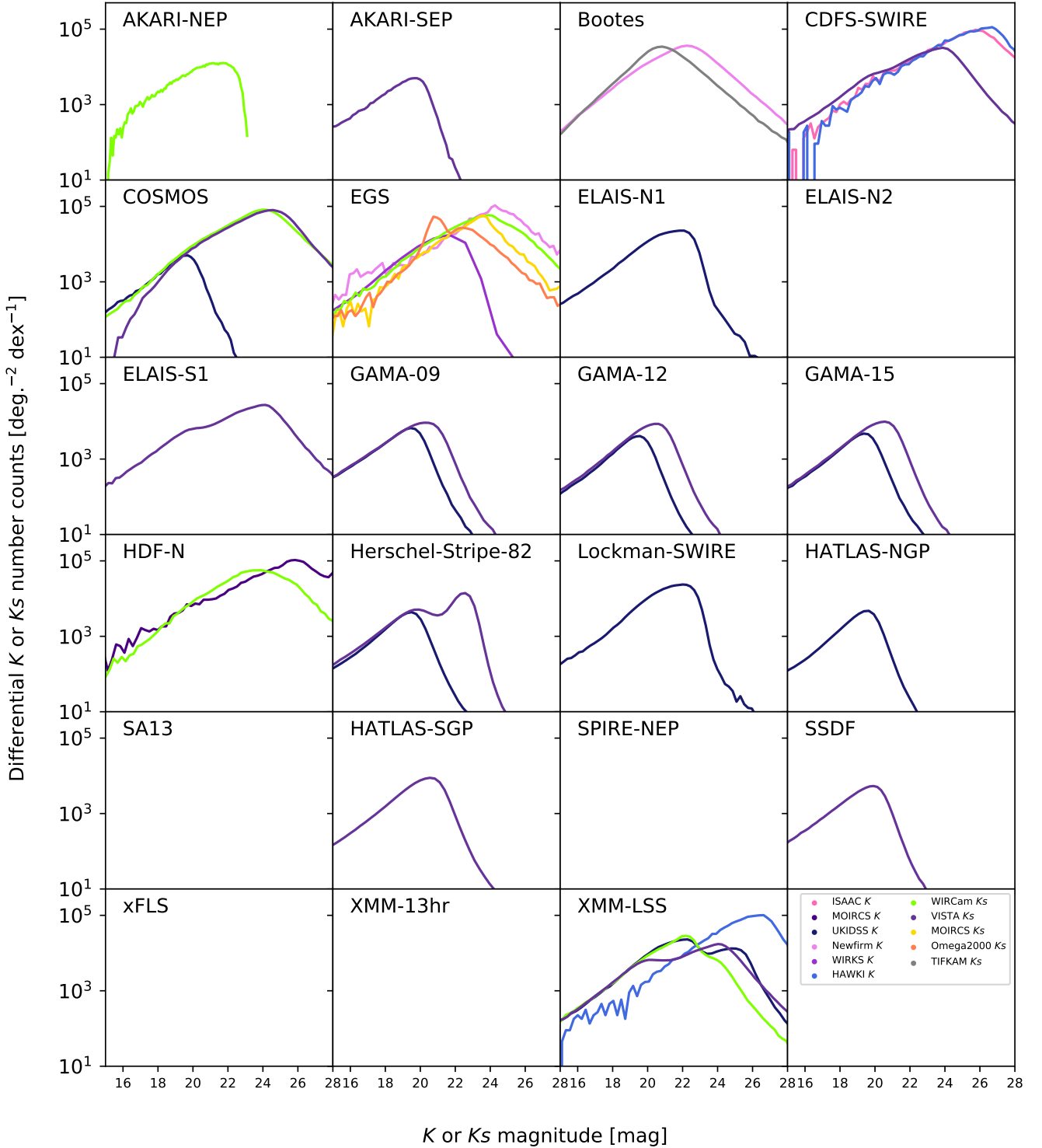
**Figure 24.**  $K$  or  $Ks$  selected number counts for the COSMOS field. HELP numbers (this study) are shown by the lines. The points show the number counts from McCracken et al. (2009); Aihara et al. (2011); Bielby et al. (2012); Fontana et al. (2014); Ilbert et al. (2015); Laigle et al. (2016).

Rodrigo C., Solano E., Bayo A., 2012, IVOA Working Draft, 15  
 Rodrigo C., Solano E., Bayo A., 2017, The SVO Filter Profile Service  
 Roehlly Y., Burgarella D., Buat V., Boquien M., Ciesla L., Heinis S., 2013, arXiv preprint arXiv:1309.6366  
 Schlafly E. F., Finkbeiner D. P., 2011, The Astrophysical Journal, 737, 103  
 Schlegel D. J., Finkbeiner D. P., Davis M., 1998, The Astrophysical Journal, 500, 525  
 Scoville N., et al., 2007, *ApJS*, 172, 1  
 Shanks T., et al., 2015, *MNRAS*, 451, 4238  
 Silva D. R., et al., 2016, in American Astronomical Society Meeting Abstracts #228. p. 317.02  
 Skelton R. E., et al., 2015, VizieR Online Data Catalog, 221  
 Skrutskie M. F., et al., 2006, *AJ*, 131, 1163  
 Stern D., et al., 2005, *ApJ*, 631, 163  
 Stetson P. B., 1990, *PASP*, 102, 932

Swinbank A. M., 2013, in Adamson A., Davies J., Robson I., eds, Astrophysics and Space Science Proceedings Vol. 37, Thirty Years of Astronomical Discovery with UKIRT. p. 299, doi:10.1007/978-94-007-7432-2\_28  
 The Astropy Collaboration et al., 2018, preprint, (arXiv:1801.02634)  
 Tie S. S., et al., 2017, *AJ*, 153  
 Timlin J. D., et al., 2016, *ApJS*, 225, 1  
 Tudorica A., et al., 2017, *A&A*, 608, A141  
 Vaccari M., 2015, in The Many Facets of Extragalactic Radio Surveys: Towards New Scientific Challenges. p. 27 (arXiv:1604.02353)



**Figure 25.** The differential  $g$  (total magnitude) number counts for the most frequently encountered wide area  $g$  broad band fluxes available in the catalogue. These are computed using the area over which a given band is available on a given field. Occasionally two peaks can be seen where a given survey has areas with different depths. This effect, which corresponds to using an overestimated area for the deeper part of the survey, can be seen on Herschel-Stripe-82 and HATLAS-SGP for instance.



**Figure 26.** The differential  $K$ , or  $K_s$  (total magnitude) number counts for the most frequently encountered wide area  $K$ , or  $K_s$  broad band fluxes available in the catalogue. Occasionally, where a survey has regions in the field to different depths there are multiple peaks. These are computed using the area over which a given band is available on a given field. There are five fields without any  $K$  or  $K_s$  data. These are ELAIS-N2, SA13, SPIRE-NEP, xFLS, and XMM-13hr.

- Vaccari M., et al., 2016, in Proceedings of the 4th Annual Conference on High Energy Astrophysics in Southern Africa (HEASA 2016). 25-26 August. p. 26 ([arXiv:1704.01495](#))
- Viero M. P., et al., 2014, The Astrophysical Journal Supplement Series, 210, 22
- Wang W.-H., Cowie L. L., Barger A. J., Keenan R. C., Ting H.-C., 2010, *ApJS*, 187, 251
- Wolf C., et al., 2004, *A&A*, 421, 913
- Wuyts S., Labbé I., Förster Schreiber N. M., Franx M., Rudnick G., Brammer G. B., van Dokkum P. G., 2008, *ApJ*, 682, 985
- Zou H., et al., 2017, *AJ*, 153, 276
- de Jong J. T. A., et al., 2013, The Messenger, 154, 44

## APPENDIX A: SUMMARY OF INPUT DATA SETS AND COVERAGE

Here we include further details of the 51 surveys included in the *masterlist* and their respective coverage of HELP fields. Table 2 shows the HELP fields with a basic overview of the field size and number of catalogue objects. A full description of the HELP fields from the perspective of *Herschel* imaging and downstream HELP data processing will be provided in the upcoming paper by Oliver et al. Table A1 gives an overview of all data included here. Table A2 gives an overview of the coverages of each survey on each field including the mean depth. Table A3 gives an overview of the main broad band photometry available on each field including the median depths. Further details regarding the input datasets, including where they can be downloaded, pre-processing we performed and documentation regarding their units etc are provided in the database documentation on GitHub in data management unit 0 (dmu0). The code used to produce the multi-wavelength catalogues created from these is presented in dmu1. All other numbered dmu folders contain details of aspects of the HELP pipeline which are not presented here.

**Table A1.** Overview of data included on all HELP fields. We chose the deepest public data available.

Survey id	Survey	Telescope	Instrument	Filters	Reference
1	Two Micron All Sky Survey (2MASS)	2MASS dedicated telescopes	Dedicated instrument	J, H, Ks	Skrutskie et al. (2006); Cutri et al. (2003)
2	All-wavelength Extended Groth strip International Survey (AEGIS)	Palomar Observatory	WIRC	J, Ks	Bundy et al. (2006)
3	Optical-NIR catalog of AKARI NEP Deep Field	Canada France Hawaii Telescope (CFHT)	MegaCam, WIRCam	u, g, r, i, z, Y, J, Ks	Oi et al. (2014)
4	VST ATLAS SURVEY	VLT Survey Telescope (VST)	OmegaCAM	u, g, r, i, z	Shanks et al. (2015)
5	3D-HST+CANDELS catalog	Hubble Space Telescope (HST)	Wide Field Camera 3 (WFC3)	5 HST bands (125nm–606nm)	Stelton et al. (2015)
6	Cosmic Assembly Near-IR Deep Extragalactic Legacy Survey (CANDELS)	HST	WFC3, Advanced Camera for Surveys (ACS)	10 HST bands (98nm–850nm)	Grogin et al. (2011)
7	WIRCam Deep Survey (WIRDS)	CFHT	WIRCAM	J, H, Ks	Bielby et al. (2012)
8	Canada-France-Hawaii Telescope Legacy Survey (CFHTLS)	CFHT	MegaCam	u, g, r, i, z	Hudlot et al. (2012)
9	The CFHT Lensing Survey (CFHTLenS)	CFHT	MegaCam	u, g, r, i, z	Heymans et al. (2012); Erben et al. (2013)
10	Classifying Objects by Medium-Band Observations 17-filter survey (COMBO-17)	MPG/ESO 2.2m Telescope	Wide Field Imager (WFI)	UBVRI, 12 narrow bands (420nm–914nm)	Wolf et al. (2004)
11	DESI Legacy Imaging Surveys: DECam Legacy Survey (DECaLS)	Bok Telescope, Mayall Telescope, Blanco Telescope	90Prime camera, MOSAIC-3 camera, Dark Energy Camera (DECam)	u, g, r, i, z, Y	Blum et al. (2016); Dey et al. (2019)
12	Beijing–Arizona Sky Survey, and the Mayall z-band Legacy Survey	Bok Telescope, Mayall Telescope	90Prime camera, MOSAIC-3 camera	u, g, r, i, z, Y	Zou et al. (2017); Silva et al. (2016); Dey et al. (2019)
13	DEEP2 Galaxy Redshift Survey	CFHT	CFH12K	B, R, I	Coil et al. (2004)
14	Dark Energy Survey (DES)	Blanco Telescope	Dark Energy Camera (DECam)	u, g, r, i, z, Y	Abbott et al. (2018)
15	Spitzer Data Fusion Catalogues (SERVS and SWIRE)	Spitzer	IRAC, MIPS	IRAC1234, MIPS123	Vaccari (2015)
16	ESO-Spitzer Imaging extragalactic Survey (ESIS)	MPG/ESO 2.2m Telescope	Wide Field Imager (WFI)	B, V, R	Vaccari et al. (2016)
17	FIREWORKS photometry of GOODS CDF-S	MPG/ESO 2.2m Telescope, HST, VLT, Spitzer	ESO/WFI, HST/ACS, VLT/ISAAC, Spitzer/IRAC	16 bands UV to FIR	Wuyts et al. (2008)
18	GOODS ACS Treasury program	HST	ACS	b, i, v, z	Giavalisco et al. (2004)
19	Hyper Suprime-Cam Subaru Strategic Program (HSC-SSP)	Subaru	Hyper Suprime-Cam (HSC)	g, r, i, z, y, N921, N816	Aihara et al. (2018)
20	Hawaii Hubble Deep Field North (Hawaii-HDFN)	Mayall Telescope, Subaru, UH 2.2m Telescope	MOSAIC-1, Suprime-Cam, QUIRC	U, B, V, R, I, z', HK'	Capak et al. (2004)
21	Infrared Bootes Imaging Survey (IBIS)	Mayall Telescope	NEWFIRM	J, H, Ks	Jannuzi et al. (2004)
22	Wide-field optical imaging on ELAIS N1, ELAIS N2, First Look Survey and Lockman Hole	Isaac Newton Telescope (INT)	Wide Field Camera (WFC)	u, g, r, i, z	González-Solares et al. (2011)
23	UV-to-FIR Analysis of Spitzer/IRAC Sources in the Extended Groth Strip	Multiple	Multiple	Multiple	Barro et al. (2011)
24	Kilo-Degree Survey (KiDS)	VLT Survey Telescope (VST)	OmegaCAM	u, g, r, i	de Jong et al. (2013)
25	Spitzer First Look Survey (FLS) - NOAO Extragalactic - R	Mayall Telescope	MOSAIC-1	R	Fadda et al. (2004)
26	NOAO Deep Wide-Field Survey (NDWFS)	Mayall Telescope	MOSAIC-1	Bw, R, I, Ks	Jannuzi et al. (1999)
27	Panoramic Survey Telescope and Rapid Response System (PanSTARRS1) 3pi Steradian Survey (3SS)	PanSTARRS1 (PS1)	Gigapixel Camera 1 (GPC1)	g, r, i, z, y	Chambers et al. (2016)
28	Red Cluster Sequence Lensing Survey (RCSLenS)	CFHT	MegaPrime/MegaCam	g, r, i, y, z	Hildebrandt et al. (2016)
29	Sloan Digital Sky Survey (SDSS) - DR13	Sloan Foundation 2.5m Telescope	SDSS camera	u, g, r, i, z	Albareti et al. (2017)
30	SDSS - Stripe 82	Sloan Foundation 2.5m Telescope	SDSS camera	u, g, r, i, z	Viero et al. (2014)
31	Instituto de Astrofísica de Canarias (IAC) Stripe 82 Legacy Project	Sloan Foundation 2.5m Telescope	SDSS camera	u, g, r, i, z	Fliri & Trujillo (2016)
32	Spitzer Deep, Wide-Field Survey (SDWFS)	Spitzer	IRAC	IRAC1234	Ashby et al. (2009)
33	Spitzer/HETDEX Exploratory Large-Area (SHELA) survey	Spitzer	IRAC	IRAC12	Papovich et al. (2016)
34	Spitzer-IRAC/MIPS Extragalactic Survey (SIMES)	Spitzer, Herschel	IRAC, MIPS, SPIRE	IRAC12	Baronchelli et al. (2016)
35	SPLASH-SXDF Multi-Wavelength Photometric Catalog	Multiple	Multiple	Multiple	Mehta et al. (2017)
36	Spitzer-South Pole Telescope Deep Field (SSDF)	Spitzer	IRAC	IRAC12	Ashby et al. (2013)
37	Subaru/XMM-Newton Deep Survey (SXDS)	Subaru	Suprime-Cam	B, V, r, i, z	Furusawa et al. (2008)
38	Spitzer Adaptation of the Red-sequence Cluster Survey (SpARCS)	CFHT	MegaCam	u, g, r, y, z	Tudorica et al. (2017)
39	Spitzer IRAC Equatorial Survey (SpIES)	Spitzer	IRAC	IRAC12	Timlin et al. (2016)
40	Spitzer UKIDSS Ultra Deep Survey (SpUDS)	Spitzer	IRAC	IRAC1234	Dunlop et al. (2007)
41	UKIRT Hemisphere Survey (UHS)	United Kingdom Infrared Telescope (UKIRT)	WFCAM	J	Dye et al. (2018)
42	UKIDSS Deep eXtra-galactic Survey (DXS)	UKIRT	WFCAM	J, K	Swinbank (2013)
43	UKIDSS Large Area Survey (LAS)	UKIRT	WFCAM	U, J, H, K	Lawrence et al. (2007)
44	UKIDSS Ultra-Deep Survey (UDS)	UKIRT	Wide Field Infrared Camera (WFCAM)	J, H, K	Almaini et al. (2007)
45	Ultradeep Ks Imaging in the GOODS-N	CFHT	WIRCam	Ks (plus IRAC crossmatches)	Wang et al. (2010)
46	VISTA-CFHT Stripe 82 Survey (VICSS82)	CFHT, VISTA	CFHT/WIRCam, VISTA/VIRCAM	J, Ks	Geach et al. (2017)
47	VIPERS Multi-Lambda Survey (MLS)	CFHT + others	WIRCam (+ others)	Ks selected multiband	Moutard et al. (2016)
48	Visible and Infrared Survey Telescope for Astronomy (VISTA) Hemisphere Survey (VHS)	VISTA	VISTA InfraRed CAMera (VIRCAM)	Y, J, H, Ks	McMahon et al. (2013)
49	VISTA Deep Extragalactic Observations (VIDEO) Survey	VISTA	VIRCAM	Y, J, H, Ks	Jarvis et al. (2013)
50	VISTA Kilo-Degree Infrared Galaxy Survey (VIKING)	VISTA	VIRCAM	Y, J, H, Ks	Edge et al. (2013)
51	zBootes - z-band Observations of the NOAO Deep Wide-Field Survey Bootes Field	Bok Telescope	90Prime imager	z	Cool (2007)

**Table A2.** Availability on a given HELP field of each input survey. Table 2 describes the id number used for each field. Survey names in there expanded form with references are given in Table A1 based on the survey id.

Survey id	Survey	area (deg. <sup>2</sup> )	No. fields	Area of survey coverage for each field in deg. <sup>2</sup> (Use Table 2 for key)																						
				1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	2MASS	1270 (100%)	23	9.2	8.7	11	13	5.1	3.6	14	9.2	9.0	62	63	62	0.7	363	22	178	0.3	294	0.1	111	7.4	0.8	22
2	AEGIS	0.7 (0.1%)	1						0.7																	
3	AKARI-NEP-OptNIR	1.1 (0.1%)	1	1.1																						
4	ATLAS	308 (24%)	2				13																			
5	CANDELS-3D-HST	0.3 (0.02%)	5				0.1	0.1	0.1																	0.1
6	CANDELS	0.2 (0.02%)	4				0.1		0.1																	0.1
7	CFHT-WIRDS	2.3 (0.2%)	3					1.0	0.5											0.1						0.8
8	CFHTLS	24 (1.9%)	4					1.1	3.6					4.3												15
9	CFHTLenS	21 (1.7%)	3						3.2					3.8												14
10	COMBO-17	0.3 (0.02%)	1				0.3																			
11	DECaLS	670 (53%)	8				8.5																			
12	Legacy Survey	90 (7.1%)	9				11																			
13	DEEP2	2.5 (0.2%)	2							1.4	13	8.9														
14	DES	595 (47%)	7			8.7		13																		
15	DataFusion-Spitzer	66 (5.2%)	8				9.8	8.3																		
16	ESIS-VOICE	5.1 (0.4%)	1								9.9	4.5	7.3													
17	FIREWORKS	0.1 (0.01%)	1					0.1																		
18	GOODS-ACS	0.1 (0.01%)	2					0.1																		
19	HSC	85 (6.7%)	8																							
20	Hawaii-HDFN	0.2 (0.02%)	1					5.1	1.3	7.7				19	13	17										14
21	IBIS	9.4 (0.7%)	1				9.4																			
22	INTWFC	42 (3.3%)	4																							
23	IRAC-EGS	0.5 (0.04%)	1							0.5																
24	KIDS	269 (21%)	6					0.3	1.1																	
25	KPNO-FLS	9.1 (0.7%)	2																							
26	NDWFS	9.1 (0.7%)	1								2.3															
27	PanSTARRS1-3SS	921 (72%)	18																							
28	RCSLenS	177 (14%)	4																							
29	SDSS-DR13	646 (51%)	15																							
30	SDSS-S82	115 (9.0%)	1																							
31	IAC-S82	113 (8.9%)	1																							
32	SDWFS	9.8 (0.8%)	1																							
33	SHELA	35 (2.8%)	1																							
34	SIMES	6.9 (0.5%)	1																							
35	SPLASH-SXDF	4.4 (0.3%)	1																							
36	SSDF	96 (7.6%)	1																							
37	SXDS	1.3 (0.1%)	1																							
38	SpARCS	33 (2.6%)	4																							
39	SpIES	74 (5.9%)	2																							
40	SpUDS	0.8 (0.1%)	1																							
41	UHS	65 (5.1%)	9																							
42	UKIDSS-DXS	23 (1.8%)	3																							
43	UKIDSS-LAS	602 (47%)	6																							
44	UKIDSS-UDS	0.9 (0.1%)	1																							
45	Ultradeep	0.4 (0.03%)	1																							
46	VICS82	81 (6.4%)	1																							
47	VIPERS-MLS	15 (1.1%)	1																							
48	VISTA-VHS	429 (34%)	7																							
49	VISTA-VIDEO	14 (1.1%)	3																							
50	VISTA-VIKING	436 (34%)	6																							
51	zBootes	6.7 (0.5%)	1																							
No surveys				4	4	11	14	10	14	12	9	6	11	8	8	8	15	10	6	3	6	3	4	8	3	23

**Table A3.** Overview of depths available in a given band. For each band we provide the area,  $a$  in square degrees and the median value of  $\sigma$  averaged over the objects in every order 10 HEALPix cell in the field. These are  $1\sigma$  depth values in  $\mu\text{Jy}$  in 2 arcsec apertures. Some K or Ks bands are empty because those fields are based on catalogues which only provide total fluxes. The small field SA13 (0.27 square degrees) is the only field with no K band coverage.

field	u		g		r		i		z		y		J		H		K		Ks		i1		i2		i3		i4				
	a	$\sigma$	a	$\sigma$	a	$\sigma$	a	$\sigma$	a	$\sigma$	a	$\sigma$	a	$\sigma$	a	$\sigma$	a	$\sigma$	a	$\sigma$	a	$\sigma$	a	$\sigma$	a	$\sigma$	a	$\sigma$			
AKARI-NEP			9.6	1.5	9.6	1.6	9.6	1.5	9.6	3.2	9.6	7.2									7.2	1.7	7.2	1.4							
AKARI-SEP			9.3	0.1	9.3	0.1	9.3	0.2	9.2	0.4	9.2	1.4	9.1	3.7	8.6	5.9					7.9	6.5	7.3	1.0	7.2	1.0					
Bootes			11.9	0.1	11.9	0.1	11.9	0.2	11.9	0.2	11.9	4.3	11.8	1.4	9.7	1.5	9.7	3.0	5.3	12.1	10.1	0.8	10.1	1.0			10.0	5.4	10.0	7.4	
CDFS-SWIRE	13.2	2.2	13.5	0.1	13.5	0.1	13.5	0.1	13.5	0.2	13.5	0.2	13.5	0.2	10.6	2.3	10.2	0.9	10.2	0.9	8.2	0.5	8.2	0.6			8.1	4.7	8.0	4.8	
COSMOS	2.2	0.0	5.4	0.0	5.4	0.0	5.4	0.0	5.4	0.0	5.4	0.2	5.3	3.8	5.3	4.7	5.3	5.9	2.2	0.2											
EGS	3.8	0.1	3.9	0.0	3.9	0.1	3.9	0.1	3.9	0.2	3.7	5.1	3.8	6.3	0.6	0.1	0.9	1.5	0.6	0.1											
ELAIS-N1	13.2	0.0	13.9	0.0	13.9	0.0	13.9	0.1	13.9	0.0	13.8	0.2	9.3	0.5			9.3	0.7			9.7	0.8	9.7	1.0	9.6	4.9	9.7	4.9			
ELAIS-N2	8.3	0.0	9.5	0.0	9.5	0.0	9.5	1.3	9.5	0.1	9.5	5.1									4.5	0.8	4.5	1.1	4.4	4.8	4.4	4.6			
ELAIS-S1			9.5	0.1	9.4	0.1	9.5	0.2	9.5	0.4	9.5	1.1	9.4	2.6	8.2	3.7					9.2	5.3	7.2	0.7	7.4	1.0	6.7	5.3	6.7	5.5	
GAMA-09	59.1	0.3	63.5	0.1	63.5	0.1	63.5	0.2	63.5	0.5	63.5	1.4	63.5	1.9	63.4	3.1	60.0	6.1	56.6	3.1											
GAMA-12	62.1	0.2	64.2	0.1	64.2	0.1	64.2	0.4	64.2	0.5	64.1	1.4	64.1	1.7	64.0	2.7	62.4	6.3	61.8	2.9											
GAMA-15	62.0	0.3	63.2	0.1	63.3	0.1	63.2	0.3	63.2	0.6	63.2	1.4	63.0	1.5	63.1	2.5	62.1	6.6	60.6	2.8											
HDF-N	0.3	0.0	0.8	1.2	0.8	1.2	0.8	1.5	0.8	2.8	0.8	5.0																			
Herschel-Stripe-82	115.8	0.3	366.9	0.1	366.9	0.2	366.9	0.3	367.0	0.4	368.0	1.7	356.1	4.0	351.3	5.5	233.9	6.3	241.2	6.5	98.9	1.9	99.0	1.8							
Lockman-SWIRE	14.8	0.0	22.1	0.0	22.1	0.0	22.1	1.2	22.1	0.0	22.0	6.2	20.4	5.5			8.6	0.6			11.6	0.7	11.6	0.9	11.5	5.0	11.5	5.3			
NGP			179.6	1.1	179.6	1.1	179.6	1.2	179.6	1.1	179.6	3.3	177.2	4.1	177.3	5.1	178.4	5.7													
SA13			0.3	0.2	0.3	0.3			0.3	1.0		0.3	6.8																		
SGP	291.8	1.8	298.7	0.1	290.1	0.1	292.8	0.3	294.7	0.5	272.4	1.3	244.2	1.5	226.6	2.4			226.2	2.6											
SPiRE-NEP			0.2	1.6	0.2	1.6	0.2	1.6	0.2	3.9	0.2	7.1																			
SSDF			112.8	0.1	112.8	0.1	112.8	0.2	112.8	0.4	112.7	1.2	111.7	2.7	98.0	4.1			107.5	5.5	94.3	79.7	93.0	86.9							
xFLS	4.7	2.4	7.8	0.2	7.9	0.1	7.8	1.5	7.8	2.5	7.8	6.2	5.6	6.1							4.1	5.4	4.1	6.0	4.0	21.0	4.0	20.4			
XMM-13hr			0.8	0.2	0.8	0.2			0.8	0.8		0.9	6.2																		
XMM-LSS	18.9	0.1	22.4	0.0	22.5	0.0	22.5	0.0	22.4	0.1	22.4	0.2	22.3	2.6	21.1	3.6	7.1	0.6	22.0	5.1	10.0	0.7	9.8	0.8	9.3	5.8	9.3	6.4			

**APPENDIX B: DATA ACCESS**

All the notebooks used to perform the reduction are available here:

[https://github.com/H-E-L-P/dmu\\_products/](https://github.com/H-E-L-P/dmu_products/)

These notebooks make use of the following Python package:

[https://github.com/H-E-L-P/herschelhelp\\_internal](https://github.com/H-E-L-P/herschelhelp_internal)

Some code to aid accessing and using the database is available here:

[https://github.com/H-E-L-P/herschelhelp\\_python](https://github.com/H-E-L-P/herschelhelp_python)

The database is structured in line with the code on GitHub in order to make relative links persistent. The folder `dmu_products/dmu0/` contains the pristine catalogues and `dmu_products/dmu1/` contains the standardised and merged catalogues. The data is available to download here:

<http://hedam.lam.fr/HELP/dataproducts/>

The data is also available via Virtual Observatory (VO; Demleitner et al. 2014) protocols at the dedicated server the Virtual Observatory at susseX (VOX):

<https://herschel-vos.phys.sussex.ac.uk/>

This permits quick searching and downloading of small subsets including programmatic access. A full description of the over one thousand columns headings can be found here:

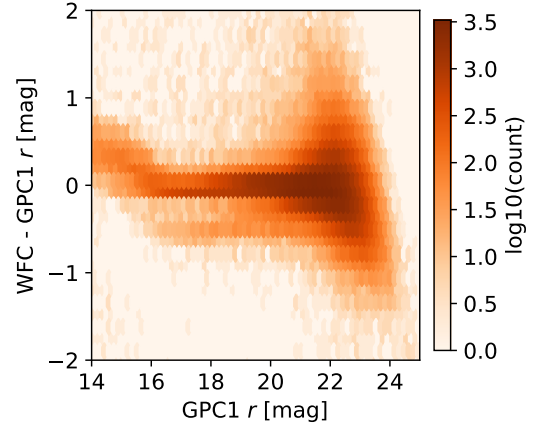
<https://herschel-vos.phys.sussex.ac.uk/herschelhelp/q/cone/info>

Examples of accessing the database with Python, including all the code used to produce the figures presented in this paper, can be found here:

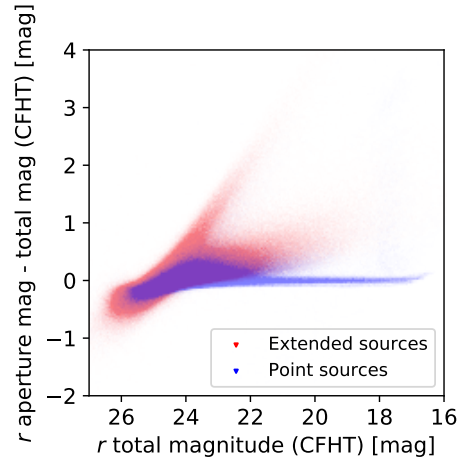
[https://github.com/H-E-L-P/dmu\\_products/tree/master/dmu31/dmu31\\_Examples](https://github.com/H-E-L-P/dmu_products/tree/master/dmu31/dmu31_Examples)

There is also an all sky viewer for viewing the catalogue over Herschel imaging or other imaging surveys:

[http://hedam.lam.fr/HELP/dataproducts/dmu31/dmu31\\_HiPS/viewer/](http://hedam.lam.fr/HELP/dataproducts/dmu31/dmu31_HiPS/viewer/)



**Figure C1.** Comparison between PanSTARRS GPC1 r band magnitudes and INT-WFC r band magnitudes on the example field ELAIS-N1. We see errors increasing for fainter objects.



**Figure C2.** Comparison between aperture and total magnitude for CFHT r band on the example field ELAIS-N1. It is shown separately for extended and point sources. Unsurprisingly, point sources tend to equality and extended objects show differences. The small number of spurious objects at the bright end are due to saturation in the CFHT imaging. A feature which is also captured in figures 15, 16, and 25.

**APPENDIX C: DIAGNOSTIC PLOTS**

We here provide figures C1 and C2 which are described in section 3 and used for validating the final catalogue. These are automatically generated for every field and many combinations of bands and can be seen in the notebooks on GitHub. The diagnostics for the example field ELAIS-N1, for instance, are viewable here:

[http://hedam.lam.fr/HELP/dataproducts/dmu1/dmu1\\_ml\\_ELAIS-N1/notebook\\_output/3\\_Checks\\_and\\_diagnostics.html](http://hedam.lam.fr/HELP/dataproducts/dmu1/dmu1_ml_ELAIS-N1/notebook_output/3_Checks_and_diagnostics.html)

This paper has been typeset from a  $\text{\TeX}/\text{\LaTeX}$  file prepared by the author.