



## Research Article

# SkyMapper Southern Survey: Data release 4

Christopher A. Onken<sup>1,2</sup>, Christian Wolf<sup>1,2</sup>, Michael S. Bessell<sup>1</sup>, Seo-Won Chang<sup>3,4</sup>, Lance C. Luvaul<sup>1</sup>,  
John L. Tonry<sup>5</sup>, Marc C. White<sup>1</sup>, and Gary S. Da Costa<sup>1</sup>

<sup>1</sup>Research School of Astronomy and Astrophysics, Australian National University, Canberra, ACT, Australia, <sup>2</sup>Centre for Gravitational Astrophysics, Australian National University, Canberra, ACT, Australia, <sup>3</sup>Department of Physics and Astronomy, Seoul National University (SNU), Seoul, Republic of Korea, <sup>4</sup>SNU Astronomy Research Center, Seoul National University (SNU), Gwanak-gu, Seoul, Republic of Korea and <sup>5</sup>Institute for Astronomy, University of Hawaii, Honolulu, HI, USA

### Abstract

We present the fourth data release (DR4) of the SkyMapper Southern Survey (SMSS), the last major step in our hemispheric survey with six optical filters:  $u$ ,  $v$ ,  $g$ ,  $r$ ,  $i$ ,  $z$ . SMSS DR4 covers 26 000 deg<sup>2</sup> from over 400 000 images acquired by the 1.3 m SkyMapper telescope between 2014-03 and 2021-09. The 6-band sky coverage extends from the South Celestial Pole to  $\delta = +16^\circ$ , with some images reaching  $\delta \sim +28^\circ$ . In contrast to previous DRs, we include all good-quality images from the facility taken during that time span, not only those explicitly taken for the public Survey. From the image dataset, we produce a catalogue of over 15 billion detections made from  $\sim 700$  million unique astrophysical objects. The typical  $10\sigma$  depths for each field range between 18.5 and 20.5 mag, depending on the filter, but certain sky regions include longer exposures that reach as deep as 22 mag in some filters. As with previous SMSS catalogues, we have cross-matched with a host of other imaging and spectroscopic datasets to facilitate additional science outcomes. SMSS DR4 is now available to the worldwide astronomical community.

**Keywords:** Optical astronomy; sky surveys; catalogs

(Received 3 February 2024; revised 29 July 2024; accepted 29 July 2024)

### 1. Introduction

As originally proposed in 2002, the Stromlo Southern Sky Survey (S4) aimed to create the first digital map of the southern sky, providing a database of a billion objects. While it was always expected that the database would be used by the entire community, the four key areas driving the S4 design at the Australian National University (ANU) were: ‘studying the creation of our solar system through a census of distant asteroids, exploring how stars and planets form by observing nearby young stars, probing the shape and extent of the Galaxy’s dark matter halo, and discovering when the first stars in the Universe formed.’

In January 2003, bushfires destroyed the original S4 telescope at Mount Stromlo Observatory: the 1.27 m ‘Great Melbourne Telescope’, which had recently completed a seven-year survey of the Galactic Bulge and Magellanic Clouds in search of microlensing by MAssive Compact Halo Objects, i.e. the MACHO Project (Alcock et al. 2000). Pursuit of the original S4 science goals required a new facility, and with the new telescope came a new survey name and new survey plan.

The 1.3 m SkyMapper telescope, located at Siding Spring Observatory (SSO), has been conducting the SkyMapper Southern Survey (SMSS) since early 2014. The telescope has a 5.7 deg<sup>2</sup> field-of-view and a 32-CCD mosaic camera (10 $\times$  the field-of-view of the original facility, with a slightly smaller pixel scale). The SMSS

includes multiple visits of varying depth in six optical filters:  $u$ ,  $v$ ,  $g$ ,  $r$ ,  $i$ , and  $z$  (Bessell et al. 2011). Full 6-band coverage of the survey now extends from the South Celestial Pole to  $\delta = +16^\circ$ , with some fields of partial coverage reaching as far north as  $\delta \sim +28^\circ$ .

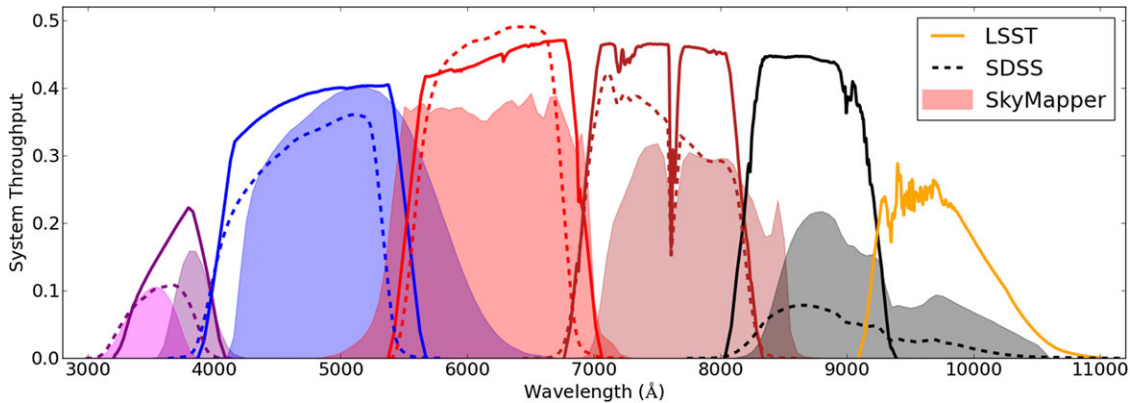
SMSS Data Release 1 (DR1, and its photometric recalibration, DR1.1; Wolf et al. 2018b) presented a shallow pass over the hemisphere, with  $10\sigma$  point-source depths<sup>a</sup> of  $\sim 18$  mag. In DR2 (Onken et al. 2019), we began introducing longer images that extend the depth by 1.5–3 mag. With DR3, we nearly doubled the number of images and detections, as the coverage of the deeper component of the survey expanded.

SMSS data have been used for a diverse array of scientific investigations, ranging from Earth-impacting asteroids closer than the Moon (namely, the last ex-atmospheric measurements of 2018 LA on UT 2 June 2018; Jenniskens et al. 2021) to the most luminous quasars at redshift  $\sim 5$  (including the most UV-luminous quasar known, SMSS J2157-3602, and the most complete survey of the bright end of the  $z \sim 5$  quasar luminosity function; Wolf et al. 2018a; Onken et al. 2022). However, the greatest impact of SMSS, in publications and citations, has been in the field of Extremely Metal Poor star searches, one of the principal science goals underpinning the SkyMapper project and the design of its unique filter set (e.g. Nordlander et al. 2019; Da Costa et al. 2019; Yong et al. 2021; Oh et al. 2023).

Here, we present SMSS DR4, which, compared to DR2, nearly doubles the time baseline from 4 to 7.5 yr, more than triples the number of images (to over 400 000), expands the sky coverage

<sup>a</sup>All magnitudes are given on the AB scale (Oke & Gunn 1983) unless otherwise indicated.

**Corresponding author:** Christopher A. Onken; Email: [christopher.onken@anu.edu.au](mailto:christopher.onken@anu.edu.au)  
**Cite this article:** Onken CA, Wolf C, Bessell MS, Chang S-W, Luvaul LC, Tonry JL, White MC and Da Costa GS. (2024) SkyMapper Southern Survey: Data release 4. *Publications of the Astronomical Society of Australia* 41, e061, 1–35. <https://doi.org/10.1017/pasa.2024.53>



**Figure 1.** SkyMapper bandpasses: throughput curves are shown for the six SMSS passbands *uvgriz* relative to SDSS *ugriz* and LSST *ugriz*. The SkyMapper curves describe the end-to-end throughput including atmosphere, all optical components and the detector, at airmass 1.3 in good weather with a recently cleaned main mirror; note, that the *u*-band sensitivity varies among the mosaic CCDs. These curves were calibrated from the count rates of standard stars in a range of survey images. According to the SDSS documentation, the SDSS passbands do not show the total throughput from atmosphere to detector. The LSST passbands have been modelled to include the full system throughput (optical elements, detectors, and an airmass 1.0 atmosphere with aerosols; version 1.5 from [this GitHub page](#)).

by 5 000 deg<sup>2</sup> (to over 26 000 deg<sup>2</sup>), and improves the astrometry and photometry of the dataset. DR4 is immediately available to the worldwide community.

Section 2 provides an overview of the SkyMapper facility and operations. We describe the SMSS design, nightly operations, and data release history in Section 3. In Section 4, we describe the SMSS data and its DR4 processing, highlighting the differences from prior releases. Section 5 describes the distillation process to go from photometric detection lists to astrophysical object lists. Section 6 details the properties of the DR4 dataset, from the selection of the input images to the final catalogue. In Section 7, we describe how to access the DR4 dataset, as well as the data format. Section 8 describes our plans for augmentation of DR4, as well as future aspirations. In Section 9, we summarise the data release.

## 2. SkyMapper overview

The SkyMapper telescope operates at SSO, near Coonabarabran, New South Wales, and was inaugurated in 2009. In the subsections below, we describe the facility itself, some of its operational constraints, the history of the telescope operations, and the absolute calibration of its passbands.

### 2.1 The facility

The SkyMapper telescope is a modified Cassegrain design, with a 3-element corrector lens assembly, providing a system *f*-number of *f*/4.79 and a delivered field-of-view 3.4 deg in diameter (Rakich et al. 2006). The primary mirror has a diameter of 1.35 m, with an unobstructed aperture of 1.30 m, and features a protective coating that may be washed (rather than having to re-aluminise the mirror). The telescope has an alt-az design with an image rotator, and sits inside an 11.5 m tall, 3-level enclosure. Both the telescope and dome were designed and constructed by Electro Optic Systems (EOS). The mean geographical coordinates of the telescope focus (in the WGS-84 frame) have been measured by GPS to be (Latitude, Longitude) = (−31.272147 ± 0.000012, 149.061416 ± 0.000014) deg, with a Height Above Ellipsoid of 1 165.5 ± 1.3 m. The location has

been registered with the International Astronomical Union (IAU) Minor Planet Center<sup>b</sup> and given observatory code Q55.

The custom-designed SkyMapper filter set (Bessell et al. 2011) is shown in Fig. 1 alongside those of the Sloan Digital Sky Survey (SDSS) and the Vera C. Rubin Observatory/Large Synoptic Sky Survey (VRO/LSST). The filters are 31 × 31 cm in size, and are primarily of coloured-glass construction (*uvgz*), although the *r* and *i* filters utilise dielectric coatings (*r* on both wavelength extremes, and *i* on the long-wavelength side). The *u*-band is known to have a red leak centred at 717 nm with ≈0.7% throughput relative to the transmission peak in the main bandpass at 350 nm (together with the CCD response, the effective throughput of the leak is twice as large). Similarly, the *v*-band was found to have a red leak at 690 nm, but 10 times lower amplitude than that of *u*-band (≈0.05%; Rocci 2013) and has not been included in the passband models.

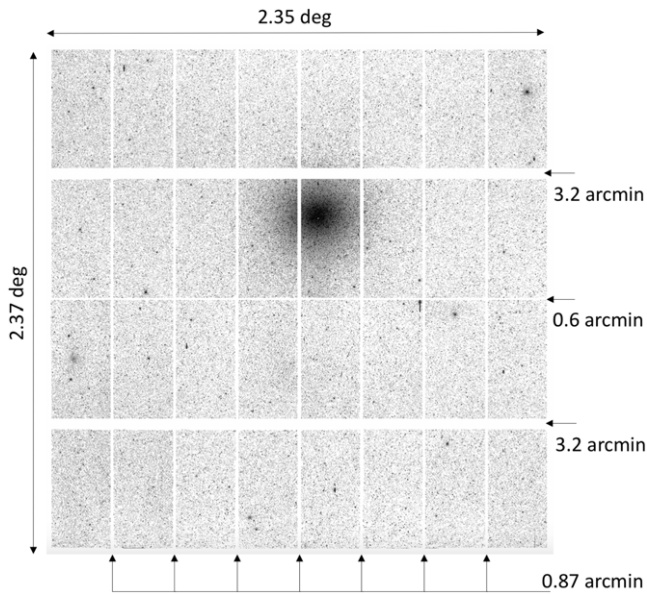
The six filters<sup>c</sup> are housed in a slide system, with filter pairs residing in one of three levels. The inactive filters are moved to either side of the optical path and provide extra baffling of stray light. The filters are locked into place with pneumatic pins, which are activated with a dry-air system that also supplies a flow of low-humidity air over the camera window and detector controllers to mitigate condensation. Filter-free observations are possible, but no such images are included in SMSS DR4. Absolute calibration of the passbands is described in Section 2.4.

The camera shutter was manufactured by a group from the Argelander Institute for Astronomy of the University of Bonn (now the private company, Bonn-Shutter GmbH), and consists of two moving blades that work together to expose and then obscure the detectors with high precision (±200 μs variation in the effective exposure time across the field of view). While providing a high uniformity in the exposure time, the time at which the exposure begins then becomes a function of location within the mosaic.<sup>d</sup>

<sup>b</sup>See <https://minorplanetcenter.net>.

<sup>c</sup>An additional filter, centred around H $\alpha$ , has also been fabricated, although its use has been limited and its calibration has not been extensively studied, thus no H $\alpha$  images are included in SMSS DR4.

<sup>d</sup>The blades travel horizontally, so the exposure timing is simply related to the mosaic *x*-axis coordinate.



**Figure 2.** SkyMapper detector mosaic, with sky coverage and gaps between CCDs indicated. Each CCD is approximately  $17 \times 34$  arcmin in size. The background is a 100-s *i*-band image of the region around the Milky Way globular cluster, Omega Centauri.

The travel time is 658 ms, and each successive exposure sees the direction of blade travel reversed.

At the focal plane of the telescope is the ANU-built mosaic camera (Granlund et al. 2006). The 32 CCDs in the SkyMapper mosaic (e2v CCD44-82-1-D03 deep-depletion, back-illuminated devices) are each  $2048 \times 4096$  pixels, giving a total of 268 million on-sky pixels with a plate scale of  $\approx 0.497$  arcsec  $\text{pix}^{-1}$ . Variations in the pixel area are seen across the field-of-view, with two, opposing corners exhibiting areas approximately 2% smaller than in the mosaic centre (the other pair of corners show a smaller change).

Allowing for the gaps between detectors illustrated in Fig. 2, the camera delivers a 90% fill factor over  $2.35 \times 2.37$  deg. The mosaic is read out by 64 amplifiers that are driven by a Scalar Topology Architecture of Redundant Gigabit Readout Array Signal Processors (STARGRASP) controller system developed by the Institute for Astronomy at the University of Hawai'i (Onaka et al. 2008). Using the methodology of Robertson (2021), measurements of the detector gain from pre-flash images<sup>e</sup> taken in 2015 were found to be  $0.74 \text{ ADU } e^{-1}$ , with a root-mean-square (RMS) scatter of  $0.03 \text{ ADU } e^{-1}$ . A repeat of the measurement in 2023 found  $0.76 \pm 0.03 \text{ ADU } e^{-1}$ . Thus, we assume a universal gain value of  $\approx 0.75 \text{ ADU } e^{-1}$ . Each amplifier includes 50 pre-scan and 50 post-scan pixels, and each  $1124 \times 4096$  read-out is stored in a separate extension of a single multi-extension FITS file.

## 2.2 Facility constraints

The images obtained by SkyMapper are affected by curvature in the focal plane, with significant point spread function (PSF) variations with radius (for detail, see Fig. 4 in the DR1 paper:

<sup>e</sup>The pre-flash images use an array of light emitting diodes inside the camera to impose an illumination pattern on the mosaic without opening the camera shutter, giving rise to a highly consistent count level well above the level of the bias.

Wolf et al. 2018a). The focal position has been selected to balance the image quality across the mosaic, but the corners exhibit a trefoil shape that is particularly visible in good seeing conditions (Fig. 3). On average, however, the FWHM is no more than 0.5 pixels worse in the corner CCDs than the mosaic centre (with the outermost 1000 pixels reaching 1–1.5 pixels wider), and standard seeing-informed convolution kernels for object detection ensure that such PSFs are not deblended into separate sources. The degraded PSF quality compared to expectations is likely due to the mechanical pressure applied to the primary mirror at three locations around the perimeter. These clamps were installed in 2013 to mitigate the vibrations of the primary mirror, which had contributed to an effective seeing as bad as 8 arcsec across the whole mosaic. Residual vibrations in the secondary mirror can be seen in the satellite trails present in some DR4 images (Fig. 4), and sets a floor in the seeing statistics of  $\sim 1$  arcsec.

The vibrations in the system are driven by the closed-cycle gaseous helium cooling system used to maintain the CCDs at their operating temperature of 155 K. The two single-stage Gifford-McMahon coldheads on either side of the camera use free-floating displacers that cannot be mechanically coupled. This leads to vibrational impulses from the two coldheads that are constantly changing in relative phase, and which drive vibrations through the telescope structure. A series of mechanical modifications to stiffen the telescope supports and damp the vibrations were enacted between 2010 and 2013, which eventually improved the image quality sufficiently to begin the Survey in March 2014.

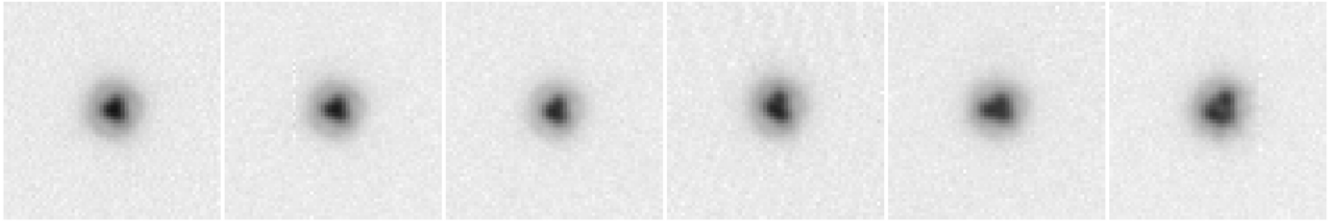
The cable drape that connects the fixed components of the dome to the rotating telescope and camera imposes limitations on the on-sky position angle (PA) that may be reached. Thus, exposures that approach the cable drape limits may be forced to rotate by  $180^\circ$  at the same RA/Dec position, imposing an overhead of  $\sim 40$  s. The images for the public Survey are typically acquired with a PA of  $0^\circ$  (meaning north is aligned with increasing *y*-axis pixel). In SMSS DR4, such images constitute 92% of the dataset, with 75% of the remainder having PA =  $180^\circ$  (giving the same mosaic footprint on the sky). The strong preference for PA =  $0^\circ$  results from that being the default setting, if available. Images with PA =  $180^\circ$  are required on the eastern side of the meridian, but these were disallowed for most Survey images from 2016-12-20 onwards. It was found that images just east of the meridian exhibited worse PSFs than those just west of it, with the FWHM in *g*-band, e.g. increasing from an average of 2.6 to 3.2 arcsec.

## 2.3 Operational History

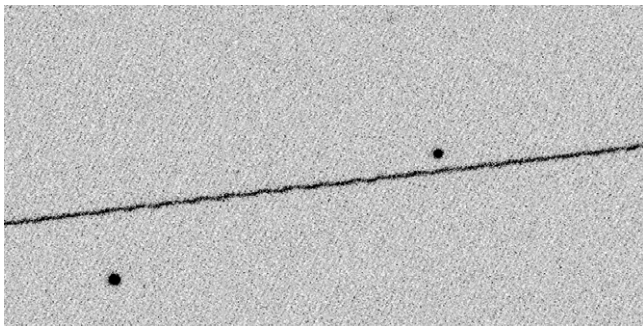
*'The best-laid schemes o' mice an' men/Gang aft agley'*  
- Robert Burns, 1785

The original description of the SkyMapper facility (Keller et al. 2007) was published before the telescope had achieved first light at the factory. In this section, we describe the operational history of the facility, particularly any modifications to plans previously published.

SkyMapper was officially opened in May 2009, but underwent a long commissioning period. Much of the effort during that time was devoted to improving the image quality, which was significantly worse than expected for the site. The main culprit was found to be the vibrations noted above (Section 2.2), which were mitigated through several rounds of mechanical engineering interventions to stiffen the structure and damp the vibrations of the primary and secondary mirrors.



**Figure 3.** An extreme case of trefoil distortions in the point spread function (PSF) with increasing distance from the mosaic centre (left-to-right: radii of  $\sim 500$ , 1 500, 2 500, 3 500, 4 500, and 5 500 arcsec). Each cutout is  $30 \times 30$  arcsec and comes from the same 100-s *i*-band image (IMAGE\_ID=20160916141036), which shows one of the largest FWHM differences between mosaic centre and corner. The greyscale uses a logarithmic stretch to highlight the extended emission of the PSF but the typical source FWHM varies by less than 0.5 pixels from the central CCDs to the corner CCDs.



**Figure 4.** Satellite trail showing the residual vibrations in the system, which contribute to a floor in the PSF FWHM distribution. The portion shown is  $10 \times 5$  arcmin, with north up and east left, and comes from a 100-s *u*-band exposure (IMAGE\_ID=20141019163153) with a mean FWHM of 2.3 arcsec.

Another unanticipated factor degrading the quality of SkyMapper images was residue on the telescope axis encoders arising from infestations of the dome by ladybird beetles (taxonomic family *Coccinellidae*). The contamination of the optical encoders caused repeatable tracking errors. These were largely resolved by delicate cleaning of the encoder tapes, but the difficulty (for humans) in accessing the encoders and lasting damage to the surfaces by the insects has continued to impact SkyMapper tracking and is thought to be responsible for the gradual accumulation of pointing errors ( $\sim 1$  arcmin  $d^{-1}$ ), but reset by homing the telescope, which became regular practice).

In light of the image quality issues early on, the Shack-Hartmann wavefront sensing system was of little utility and has not been used in subsequent operations. The off-axis auto-guider was also never implemented.

Additional delays to the start of Survey operations were incurred because of a major bushfire which hit SSO in January 2013. In addition to the loss of 53 homes and the SSO Lodge, the Wambelong fire burned more than 95% of the Warrumbungle National Park that surrounds SSO, and 55 000 hectares overall. SkyMapper was subjected to intrusion of ash into the dome, but suffered no lasting physical damage. However, because of the potentially corrosive nature of the ash, an extensive cleaning process was undertaken, including the careful washing and baking out of most circuit boards in the dome. Ultimately, 10 weeks were spent on the bushfire cleanup process.

One of the elements which was exposed to the ash of the Wambelong fire was SkyDICE (the SkyMapper Direct Illumination Calibration Experiment; Rocci 2013; Regnault et al.

2015). This set of calibrated photodiodes was intended to provide absolute measurements of the system throughput at 23 wavelengths (with LED emission widths of 20–50 nm) across the SkyMapper filter set, while also sampling any spatial variations thereof. Following the fire, the module was not put into regular deployment, and in 2016, the system was disassembled and shipped back to the team at Laboratoire de Physique Nucléaire et de Hautes-Énergies in Paris. Lack of available personnel precluded its recalibration and return to SSO.

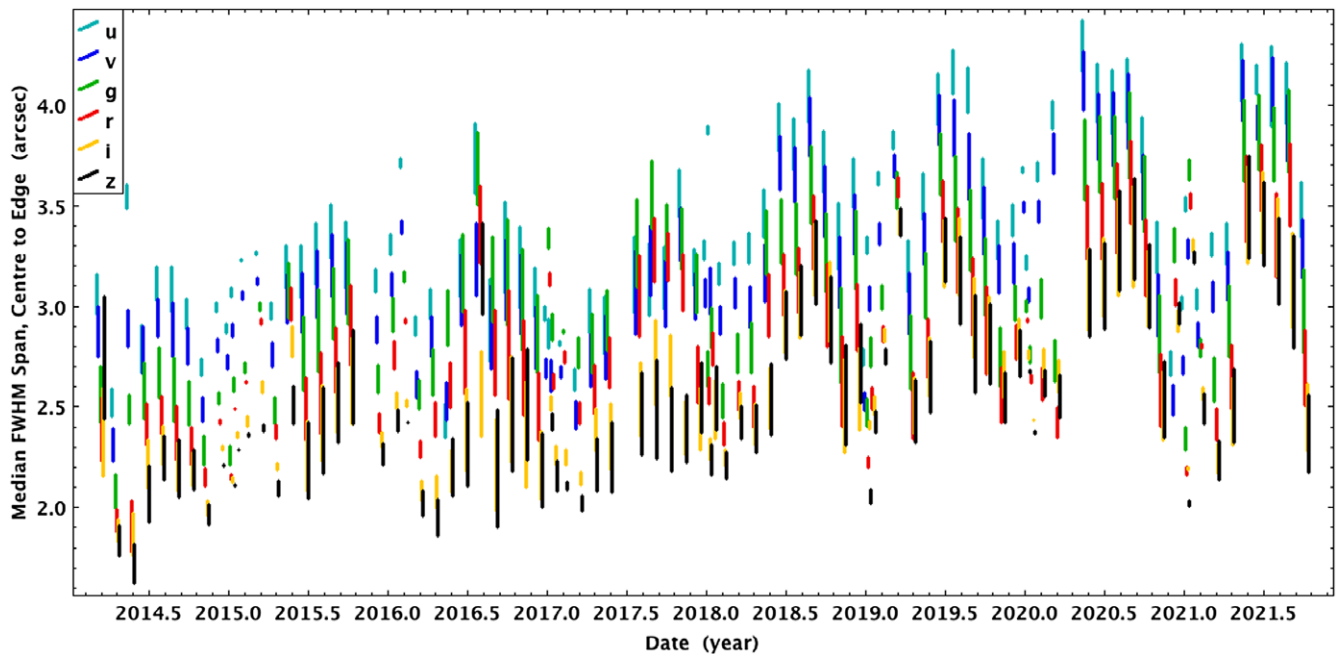
Despite these teething issues with the telescope, scientific observations prior to the commencement of the Survey had already demonstrated the power of the bandpass design for selecting low metallicity stars. These discoveries included the most iron-deficient star known at that time (Keller et al. 2014), and the first large collection of metal-poor stars in the Galactic Bulge (Howes et al. 2015).

Ahead of the start of the Survey in early 2014, the strategy was reassessed to take account of the intervening scientific progress. Priority was given to the Shallow Survey (see Section 3.1) for the first year, to obtain a uniform dataset across the full RA range for the scientific community. In addition, the Shallow Survey exposure times were increased for both the bluest filters and the reddest filters, to deliver a similar depth of  $\sim 18$  ABmag.

In contrast, the exposure time for the Main Survey was reduced from 110 to 100 s. While this came at the cost of a small amount of depth, it was much less than the effects of the image quality being worse than expected, and helped to improve the rate of progress for the Survey. When the Main Survey became fully activated in April 2015, the cadence of images was significantly relaxed compared to the original schedule – the loss of depth having reduced the importance of the sampling of RR Lyrae light curves, and the start of *Gaia* observations having made Trans-Neptunian Object proper motions less compelling to measure with SkyMapper. As a result, the telescope was able to place a stronger emphasis on observations close to the meridian.

When the Shallow Survey was prioritised early in the Survey, it also involved a reconsideration of the photometric calibration strategy. No longer was the Shallow Survey only undertaken in photometric conditions – ultimately, photometricity was not actively measured in real-time. Thus, the fields with spectrophotometric standard stars (see Section 3.1) were not used to anchor the overall Survey calibration, but external all-sky data sources were relied upon, culminating with the *Gaia*-based solution described below (Section 4.8).

The early months of the Survey operations also revealed other properties of the data, which motivated certain alterations of



**Figure 5.** Monthly PSF FWHM in each filter (with small horizontal offsets for visibility), showing the span between the median mosaic centre CCD (lower end of each line) and median edge CCD (upper end of line). The overall seeing and centre-to-edge differences are smallest late in the southern summer, and a long-term degradation in the seeing over time can be seen; both features are likely due to imperfect focus settings.

the hardware configuration and nightly activities. For example, Section 4.3.3 of Wolf et al. (2018b) describes the changes to the detector voltages that were required to remove spatially varying curvilinear features in the images. However, the images obtained prior to that correction in July 2014 retain those artefacts. Similarly, the approach used to remove large-scale flux gradients from the twilight flatfields (see Section 3.2) was only developed near the end of 2014, and the data taken prior to November 2014 will be less well calibrated because of the flatfields having been obtained at fixed position angle.

The dome cooling systems installed in the SkyMapper enclosure proved unable to pre-cool the internal air to the expected nighttime ambient temperatures. Therefore, rather than performing the planned focus runs at the start of each night, a more dynamic focus system was enacted. Suitable focus control was found to be achievable by simple correction for the thermal expansion of the telescope truss along with an airmass-dependent flexure compensation with the secondary mirror (much improved after updates to the coefficients in Feb 2014). The focus equation is linear in temperature, with a slope of  $-46 \mu\text{m K}^{-1}$ .

In 2020, a gradual degradation of the mean PSF over the years of the Survey was noticed, leading to a mild revision of the focus equation. Pairs of out-of-focus ‘doughnut’ images bracketing the predicted focus setting by  $\pm 400 \mu\text{m}$  are obtained during twilight, and the ratio of their diameters indicates the offset between real and predicted focus.

Over the Survey years, the average PSF FWHM has evolved in a pattern that is common to all passbands (see Fig. 5). The focus offset reconstructed from doughnut images evolve broadly similarly to the mean PSF FWHM, indicating that an imperfect focus equation is the cause of the evolution. The worst image quality, from 2020, shows the strongest reconstructed focus offsets, of

up to  $50 \mu\text{m}$ , and equivalent to misjudging the relevant temperature for the telescope structure by  $\sim 1$  degree. The seeing records from the 3.9 m Anglo-Australian Telescope (AAT), also located at SSO, shows relatively stable seeing behaviour over the DR4 period. However, the median AAT seeing from the start of 2020 to the end of the DR4 date range increased from 1.5 to 1.75 arcsec (C. Ramage, priv. comm.).

Faults in the detector cooling system have been the primary source of extended periods of maintenance downtime since the start of Survey operations: each of the three helium compressor failures (in 2015, 2016, and 2017) has taken roughly 8 weeks to resolve. In addition, faults with some of the detector controllers have resulted in periods of partial mosaic operation.<sup>f</sup> SSO was closed from 25 March due to COVID-19, but SkyMapper was able to restart operations from 07 May 2020.

#### 2.4 Absolute passband calibration

We determined the end-to-end throughput of the SkyMapper passbands from DR4 data. The throughput of the glass filters is expected to change little in wavelength dependence over time. However, the atmospheric throughput fluctuates from day to day and even during the night. The reflectivity of the telescope mirrors tends to degrade over time between the less-than-annual mirror washing cycles; time series data of the image zeropoints show an average loss of reflectivity by  $\sim 1\%$  per month (see Section 6). We chose to evaluate the throughput from images of the southern spectrophotometric standard stars Feige 110, GD 50, GD 108 and LDS 749B, taken in good weather soon after a mirror

<sup>f</sup>Incomplete mosaic operation motivated the decision to limit the scope of images for DR4 to those taken before October 2021.

washing restored high system throughput. The photon flux of these stars arriving outside of the Earth's atmosphere is known from CALSPEC<sup>8</sup> (Bohlin, Gordon, & Tremblay 2014; Bohlin, Hubeny, & Rauch 2020) and can be compared to the electron count recorded by the CCD camera.

We start from laboratory measurements of the filter transmission curves (Bessell et al. 2011). In the  $u$ -band, the quantum efficiency varies between the detectors in the mosaic, so we use a mean CCD efficiency for the synthetic photometry of the standard stars. Given that we imaged the standard stars only on a subset of CCDs, our throughput estimation is only a rough average for the mosaic in  $u$ -band. The observations had an average airmass of 1.2, which we use to predict the wavelength dependence of the atmospheric transmission. We base our expectations on a reflective aperture area of 0.95 m<sup>2</sup> resulting from a primary mirror with 1.30 m unobstructed aperture diameter and an obstruction from a secondary mirror with 0.69 m diameter. A CCD gain of 0.75 ADU e<sup>-1</sup> is used. The resulting end-to-end throughput curves are shown in Fig. 1. At the blue end, they are comparable to SDSS, while the red-sensitive CCDs in SkyMapper provide better sensitivity at longer wavelengths.

Bessell et al. (2011) stated a need to recalibrate the filter transmission curves with evidence from on-sky measurements in the converging beam of the telescope, which is expected to make a difference, especially for the  $r$ - and  $i$ -bands that have dielectric coatings. In Section 6.6 we discuss what we can learn from the DR4 data and give an outlook to potential future calibration improvements.

### 3. SkyMapper southern survey

In the subsections below, we describe the overall design of the SMSS, the typical nightly operations, and the history of SMSS data releases.

#### 3.1 Survey design

In the first year of SMSS operations, the telescope was mainly focused on completing an initial, rapid pass around the sky in all filters. With exposure times between 5 and 40 s, the Shallow Survey component of the SMSS achieved a depth of  $\sim 18$  mag in all 6 filters. The Shallow Survey observations of each field were obtained sequentially in order of increasing filter wavelength, with any interruptions to the 4-min image set causing the full sequence to be repeated. This dataset was then processed into SMSS DR1 (Wolf et al. 2018b).

After significant sky coverage was obtained, the Shallow Survey was restricted to days around Full Moon (since the short exposure times leave the background levels at modest levels). This change was enacted on UT 2017-03-05.

Early in the Survey, a set of seven fields containing CALSPEC spectrophotometric standard stars (Bohlin et al. 2014, 2020) were observed multiple times per night, again in order of increasing filter wavelength. These stars (Table 2) were originally intended to form the basis for the photometric calibration of the entire SMSS, but the variable observing conditions and the eventual availability of all-sky photometric datasets of high uniformity and precision (culminating in the *Gaia* low-resolution spectroscopy described

**Table 1.** Properties of the SkyMapper filter set: name, central wavelength  $\lambda_{\text{cen}}$  and full width at half maximum (FWHM), peak system efficiency  $Q_{\text{peak}}$  as in Fig. 1, zeropoint loss ( $\Delta ZP$ ) per airmass (AM), median PSF FWHM among Quality 1 and 2 Shallow and Main Survey images, and magnitude limit of a typical Q1/2 100 sec exposure (peak of source count histogram). Note that the  $u$ -band has a 0.7% red leak; the  $\lambda_{\text{cen}}$  and FWHM refer to the main bandpass.

Filter	$\lambda_{\text{cen}}$ /FWHM (nm)	$Q_{\text{peak}}$	$\Delta ZP/AM$ (mag AM <sup>-1</sup> )	PSF FWHM (arcsec)	100sec limit ABmag
$u$	350/43	0.11	0.65	3.15	18.1
$v$	384/31	0.16	0.41	3.00	18.3
$g$	510/157	0.40	0.19	2.80	19.5
$r$	617/158	0.39	0.12	2.63	19.2
$i$	779/141	0.32	0.08	2.54	18.4
$z$	916/84	0.22	0.08	2.49	17.8

**Table 2.** CALSPEC standard star fields.

Name	RA/Dec (J2000)	DR4 $N_{\text{visits}}$ $u, v, g, r, i, z$
HD009051	22.3079/−24.2257	966, 990, 1 002, 1 005, 998, 1 000
HD031128	73.1554/−26.9493	1 181, 1 186, 1 177, 1 200, 1 176, 1 187
HD074000	130.3250/−16.2293	806, 812, 802, 820, 817, 811
HD111980	193.4275/−18.4076	1 002, 1 008, 999, 1 007, 1 015, 1 017
HD160617	265.8200/−40.2054	2 090, 2 086, 2 070, 2 057, 2 034, 2 022
HD200654	316.7583/−49.8487	1 152, 1 148, 1 168, 1 169, 1 171, 1 170
GJ754.1A	290.2604/−7.5521	603, 617, 621, 612, 618, 620

in Section 4.8), meant that the Standard fields were never used for that purpose. Because of the shorter exposure times (from 3 s in  $g$  and  $r$  to 20 s in  $u$ ), these seven fields are nearly the only regions in which sources brighter than  $\sim 9$  mag are unsaturated. The nightly visits (weather and season permitting) resulted in  $\sim 1$  000 visits per filter for each Standard field. However, after the Shallow Survey was limited to bright Moon phases, the Standard field observations were similarly restricted (from UT 2017-03-05), and then were halted altogether from UT 2021-05-01 onwards, because they consumed  $\sim 6\%$  of the observing time in a clear night.

The next major SMSS component is the Main Survey, wherein the images had a standard exposure time of 100 s. These were acquired in two modes: image pairs ( $u + v$ ,  $g + r$ , or  $i + z$ ; taken together to help protect against uncorrected cosmic rays implying spurious brightening) and colour sequences (10-image collections acquired over a span of 20 min in the filter order:  $uvgruvizuv$ ). Pooling the  $u + v$  exposures into a single visit was intended to enhance the depth for these two lowest-sensitivity filters. However, it does turn out that they are also useful for observing short-term variability, e.g. in compact eclipsing binaries (Li et al. 2022) and Blue Large-amplitude Pulsators (Chang et al. 2024).

Late in the Survey operations, a portion of the  $u + v$  image pairs were extended to 300 s each (with exposures south of  $\delta = -75^\circ$  further lengthened to 400 s, while that region of sky was also restricted to seeing conditions better than 1.9 arcsec from April 2018 onwards). The longer exposures in  $u + v$  were intended as a trade-off for a reduced number of visits (and thereby reduced overheads), and were still observationally suitable because the standard 100-s exposures in those filters were read-noise limited. However, the long exposures constitute less than 0.5% of the Main Survey images in those filters.

<sup>8</sup>See <https://www.stsci.edu/hst/instrumentation/reference-data-for-calibration-and-tools/astronomical-catalogs/calspec>.

In the early years of the Survey, the bad seeing time (with a threshold that evolved between 2 and 3 arcsec) was principally used by the SkyMapper Transient (SMT) survey (Scalzo et al. 2017; Möller et al. 2019), which searched  $\sim 2000 \text{ deg}^2$  for supernovae and other transients. In addition, a fraction of the good-seeing time was made available to Australia-based applicants, totalling over 500 h between 2014 and 2019.

### 3.2 Nightly operations

In this section, we describe a typical night's operations plan for the SkyMapper facility. The telescope is fully robotic, with no human involvement expected during standard operations. The details have evolved over the course of the Survey, but the description below reflects the current framework.

Each afternoon, a crontab process launches the Scheduler software, a Perl framework that controls the telescope's activities until an automatic shutdown following the morning twilight. In preparation for the night, it pre-selects available survey fields while excluding those containing bright planets. For each SkyMapper image, the Scheduler prepares an observation definition that is provided to the high-level interface software, the Telescope Automation and Remote Observing System (TAROS; Wilson et al. 2005), as implemented for SkyMapper (Vaccarella et al. 2008). TAROS coordinates the activities between low-level systems, including the Configurable Instrument Control and Data Acquisition software (CICADA; Young et al. 1997; Young, Roberts, & Sebo 1999) that interfaces to the camera hardware (filter selector, camera shutter, detector controllers, etc.), and the software that interfaces with the EOS telescope and dome control systems, which run on a separate pair of computers.

A typical night obtains a set of evening bias frames before sunset, and if the weather is suitable for observing, the dome is then opened after the Sun is down in order to obtain twilight flatfields. Working through a sequence of filters of increasing sky-level sensitivity ( $u$ ,  $v$ ,  $z$ ,  $i$ ,  $r$ ,  $g$ , and the filter-free clear aperture), the telescope takes 3 images at one PA, then rotates  $180^\circ$  to obtain another 3 images. (The rotation allows for the trivial correction for the large-scale gradient in sky illumination over the wide field-of-view.) The next filter begins from the same PA, then rotates back to the original PA for the second set of 3 images, and so on through all the filters until all of the individual exposure times exceed 60 s. The starting position is selected to be near an Hour Angle of  $-1 \text{ h}$  and a Declination of  $-35^\circ$ , while avoiding the Moon, Galactic Plane, and any bright planets (from Venus to Saturn, inclusive). Each observation is executed while tracking the sky, with 30 arcsec dithers in RA and Dec between exposures.

During astronomical twilight (Sun angles between 12 and 18 deg below the horizon), the sky in the redder filters has become faint enough to allow useful astronomical observations. Thus, we allow 100 s Main Survey exposures in  $i$  and  $z$  to be taken before full nighttime darkness is achieved.

In full darkness, the Scheduler then cycles through the available image types (which may depend upon the Moon's phase and current position relative to the horizon) until it finds a suitable observation to execute. The top priority is given to the Target-of-Opportunity (ToO) programs, including the follow-up of gravitational wave alerts (Chang et al. 2021). Next, any other non-Survey images are considered within the UT time boundaries defined by the user per exposure. Then, Shallow Survey, and Main Survey images are considered in turn. For Survey images,

each available field is given a weight that incorporates its current Hour Angle, position within a sequence/pair, and other priority levels. The field with the highest weight is translated into a TAROS observing block, which passes the observation definition to the hardware.

For most image types, TAROS is configured to hold two observation definitions, so it can reconfigure the system as soon as it records the completed exposure of the first image. This allows the multiple system components to be reconfigured during the  $\approx 22 \text{ s}$  overhead time between images (consisting of approximately 13 s of readout and 9 s of additional system overheads). Additional parameters affecting the time between images are the  $\sim 40 \text{ s}$  time for the instrument rotator to execute a  $180^\circ$  rotation, and the slew speeds of  $4 \text{ deg s}^{-1}$  in azimuth and  $2 \text{ deg s}^{-1}$  in elevation.

After each image has been written to disk, the QuickLook analysis process is run on two of the image's central amplifiers, one each from CCDs on opposite sides of the mosaic centre. Basic image parameters are recorded in the Scheduler's PostgreSQL database, including an estimate of the seeing. After normalising between filters (the SkyMapper seeing improves notably towards longer wavelengths) and to an airmass of 1 (using an empirically derived trend that seeing degrades as airmass to the power of 0.8), the last 30 images from within the past 30 min have their QuickLook seeing estimates medianed, which serves to establish the current seeing estimate used by the Scheduler in its next observation decision.

At the close of each night, twilight flatfields are obtained in the opposite filter order as in the evening (now beginning near an Hour Angle of  $+1 \text{ h}$ ), and a final set of 10 bias frames are taken. After observing has concluded, images are transferred from the telescope to the National Computational Infrastructure (NCI) on the ANU campus in Canberra. The images are stored there until ready to be processed with NCI's high-performance computing system. The ToO images and other high-priority data are typically processed in near-real-time, through a separate data pipeline that operates on the computer systems at Mount Stromlo Observatory, but are still copied to NCI for long-term archiving.<sup>h</sup> However, for inclusion in DR4, all such images are processed from a raw state as described in Section 4 below.

### 3.3 Previous data releases

Over the course of the Survey, data releases of increasing sky coverage, data volume, and photometric quality have been made available. The series of DR parameters, including the current DR4, is given in Table 3.

Experience with the instrument and dataset, as well as the availability of new auxiliary data from other surveys, has led to an evolution in the SMSS image processing (cf. Wolf et al. 2017; Luvaul et al. 2017; Wolf et al. 2018b; Onken et al. 2019). All previous data releases are now nearly obsolete: they include data that is excluded from DR4 on the grounds of low quality; such data might be useful if coverage is desired at a specific time. However, for static sources and for reliable statistical studies of variability, the DR4 data set is the best reference. In the following Section, we describe the processing approaches adopted in SMSS DR4.

<sup>h</sup>Because the general data processing did not keep pace with the observations, the steady trickle of data from the telescope that had been anticipated proved unnecessary.

**Table 3.** History of data releases of the SkyMapper Southern Survey.

Data release	Sky			Release		
	Images	area deg <sup>2</sup>	Objects ( $\times 10^6$ )	Detections ( $\times 10^9$ )	date	Reference
Test DR	138	60	0.1	0.001	2015-05	...
Early DR	19 147	6 700	49	0.5	2016-05	...
DR1	66 840	20 200	285	2.1	2017-06	Wolf et al. (2018b)
DR2	121 494	21 000	285	4.7	2019-02	Onken et al. (2019)
DR3	208 860	24 000	619	8.1	2020-02	...
<b>DR4</b>	<b>417 223</b>	<b>26 000</b>	<b>724</b>	<b>15.1</b>	<b>2024-02</b>	<b>This paper</b>

#### 4. SMSS DR4 data processing

Here we describe the main steps in the SMSS image reduction and extraction of photometric parameters in DR4. The images were processed on *Gadi*, NCI's peak supercomputer, utilising approximately 550 000 CPU-hours (including time for images which failed subsequent quality cuts). The image properties and derived photometry are stored in a PostgreSQL database (version 11.20), which also manages the data reduction flow of the pipeline by recording the ongoing and completed steps for each image, along with a status code for each step to determine how the image is treated by subsequent steps.

##### 4.1 Electronic noise filtering, overscan subtraction, and cross-talk correction

The SkyMapper electronics are subject to variable levels of high-frequency sinusoidal noise, which we filter from the images. Each amplifier is Fourier-transformed and we search for significant power corresponding to wavelengths between 6 and 8 pixels in the  $x$ -axis direction.

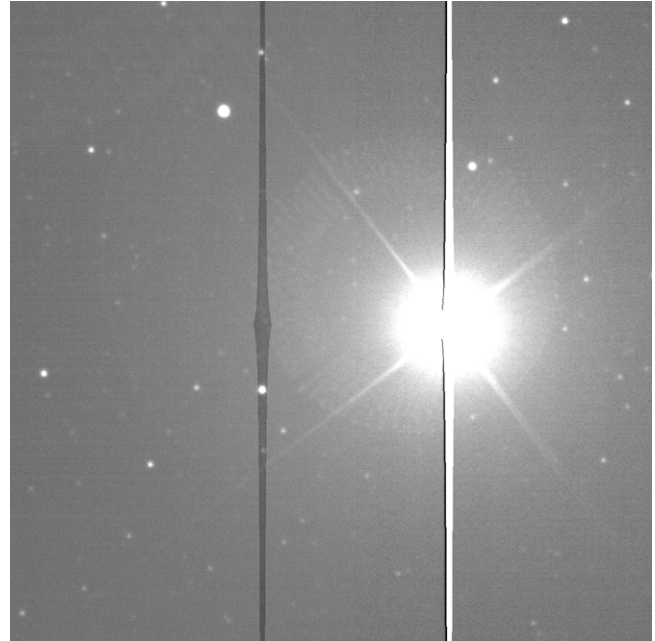
If any amplifier shows such evidence of sinusoidal variations, then all amplifiers for that image are run through a row-by-row fitting procedure. For each row, we ignore the overscan region and subtract a 30-pixel boxcar-smoothed copy of the row from itself in order to isolate variations of the intended frequency. We then perform a least-squares fit of a sine function to the row, allowing the wavelength, phase, and amplitude to vary, and taking the results from the FFT analysis as the starting value for the wavelength. The best-fitting sinusoid is then subtracted from the original row (including the overscan region).

Next, the data is analysed with SOURCE EXTRACTOR (version 2.19.5; Bertin & Arnouts 1996) in order to identify saturated pixels (taken as those above 58 000 counts), which are flagged in the pixel masks associated with each CCD. The description of all bits in the pixel mask is given in Table 4. The data from each CCD is then merged into a single FITS image from its two constituent amplifiers, while simultaneously subtracting the bias level using the post-scan region (which is more stable in its behaviour than the pre-scan) and trimming both overscan regions.

We also correct for cross-talk between the two amplifiers of each CCD. The typical fractional amplitudes are  $5 \times 10^{-4}$  and are subtracted from the neighbouring amplifier. Source pixels that are flagged in the previous step as saturated cannot have their cross-talk accurately corrected in the neighbour amplifier, and so the

**Table 4.** Pixel mask bit values.

Bit value	Description
1	Known bad pixels
2	Not used
4	Saturated (above 58 000 ADU in raw image)
8	Affected by uncorrectable cross-talk
16	Masking of full amplifier due to electronic fault
32	Corrected cosmic rays from L.A.Cosmic
64	Affected by correlated bias fluctuations



**Figure 6.** SkyMapper image of Altair showing uncorrected cross-talk from saturated sources between adjacent amplifiers of a CCD. The image is a 20 s  $v$ -band exposure. Counts in the adjacent amplifier are reduced, but those regions are flagged in the associated pixel masks, as are both the saturated pixels and the low-count ringing adjacent to saturation.

pixels in the latter are flagged as cross-talk-affected in the pixel masks (see Fig. 6). Pixels with full wells also induce amplifier ringing, where the next non-saturated pixel in the row has 0 counts and the subsequent two pixels have severely suppressed count levels. Such pixels are also flagged as saturated in the pixel mask. Additional cross-talk effects on other CCDs read out by the same controller are less than 5 ADU for a fully saturated pixel, and are not presently flagged.

##### 4.2 WCS solution

Previous SMSS data releases have utilised the astrometric software of Astrometry.net (Lang et al. 2010) to derive WCS solutions for each CCD by matching against the Fourth US Naval Observatory CCD Astrograph Catalog (UCAC4; Zacharias et al. 2013). The resulting coordinate system provided a good match to the positions of stars presented in *Gaia* DR2 (Gaia Collaboration et al. 2018), with typical offset smaller than 0.2 arcsec (Onken et al. 2019).



However, in some images, particularly in shorter  $u$ - and  $v$ -band exposures, there were insufficient stars matched to UCAC4 to produce a reliable coordinate system for certain CCDs, and the sources on those CCDs were then absent from the photometric catalogue. In DR3, 36% of images lost at least two CCDs, principally for lack of a WCS solution. This motivated a revised approach to recover those lost CCDs.

For DR4, we have adopted a mosaic-wide algorithm for determining the coordinate system. Based on a careful fitting of *Gaia* sources across the mosaic in a set of densely populated images, we improved the mapping of each CCD's location (offset and rotation) relative to the mosaic centre.

First, to get the overall image boresight, we run SOURCE EXTRACTOR on all CCDs of an image, and select the 30 brightest stars in each of the central 8 CCDs. We map the CCD  $x/y$  positions into mosaic  $x/y$  positions and run the (up to 240) stars through Astrometry.net's SOLVE-FIELD software (version 0.76), using the '5000-series' index files created from *Gaia* DR2 positions. The location of the mosaic centre and position angle of the mosaic system that is determined by that process is then used with the mosaic mapping to generate provisional RA/Dec positions for the full list of sources on each CCD.

For each CCD, the brightest 300 sources are matched to the nearest *Gaia* DR2 source within 10 arcsec, where the *Gaia* source is required to have  $G < 14$  mag (Vega). The median shift in RA and Dec is determined and applied to all 300 sources when a second round of matching is performed, now with a maximum allowed offset of 5 arcsec. The *Gaia* positions are then adopted as the 'true' coordinates for those 300 stars.

From the (up to) 9 600 stars with *Gaia* DR2 coordinates over 32 CCDs, we then fit a mosaic-wide WCS solution, allowing polynomial corrections to the tangent projection of up to 3rd-order (but excluding radial terms). The resulting coordinate system is then overlaid on a dense grid of  $x/y$  positions for each CCD using the XY2SKY routine from the WCSTools package<sup>1</sup> (version 3.8.7; Mink 1996, 2019). The grid of CCD-based  $x/y$  points with associated RA/Dec coordinates is then re-fit for each CCD to yield the final WCS solution, adopting a TPV convention.<sup>2</sup> This two-step fitting procedure ensures that each CCD has a coordinate system that is referenced to its individual CCD centre, and is well defined in relation to its native  $x/y$  axes, regardless of any small rotations relative to the overall mosaic. (In practice, the largest rotation of any CCD is less than  $0.08^\circ$ , but this still translates to a shift of up to 3 pixels at the CCD corners.)

Finally, the WCS solution is saved in the header of each CCD image and its corresponding image mask. In Section 6.4, we quantify the accuracy and precision of the resulting WCS solutions.

### 4.3 Bias correction

On each night of SkyMapper telescope operations, between 10 and 20 bias exposures are obtained. These are treated as in Section 4.1 and then mean-combined with outlier clipping (to omit cosmic rays) to produce a 2D bias image. If no bias frames are available on a given night, those of adjacent nights are used. The 2D bias image is subtracted from the science frame.

Next, the pipeline addresses variable bias patterns in each row through the use of principal components analysis (PCA). The

bias level during readout fluctuates in a manner that, for a given row, can be different from image to image, but which is effectively drawn from a small family of patterns. We first generate a set of principal components (PCs) by subtracting the 2D bias image from each of the input bias exposures, and then determining the top 10 PCs that describe the residual bias variations for each of the 4 096 rows in that CCD (and performed separately for each of the two amplifiers).

In applying the PCs to the science image, we first use SOURCE EXTRACTOR to determine background and object maps. The former is subtracted and the latter is slightly broadened (using a  $\sigma = 0.5$  pixels Gaussian kernel) before being used to mask data in the science image. Up to 10 of the PCs generated from the bias images are then fit to the background-subtracted and masked science image on a row-by-row basis, where the number of PCs varies in proportion to the unmasked pixel percentage,  $p_{\text{unm}}$ , of each row (10 PCs for  $p_{\text{unm}} > 25\%$ , 5 PCs for  $p_{\text{unm}} > 10\%$ , only the first PC for  $p_{\text{unm}} > 2\%$ , and no correction below that). The best-fit pattern for each row is then subtracted from the original science frame.

While this approach largely works well to model and remove the bias level, it is known to function less than optimally in the regions around extended sources. We return to this issue in Section 6.7.4.

### 4.4 Flatfield correction

The large SkyMapper field-of-view creates challenges for uniformly illuminating internal screens for producing dome flats, so the SMSS relies on twilight flats, which are observed whenever the weather conditions allow. While some surveys utilise long-running mean flatfields for a first-pass calibration, as nightly variations in flatfield illumination are often larger than seasonal or long-term variations (e.g. Drlica-Wagner et al. 2018), SkyMapper suffers from temporally varying sensitivity changes that are more impactful than localised effects arising from changes in the pattern of dust motes.

The SkyMapper camera features an evolving pattern of sensitivity changes most profoundly observed around the edges of the mosaic. The pattern varies in a systematic way over the period of time following a warm-up of the camera, with changes occurring most rapidly soon after returning to the operating temperature, and then asymptoting to a persistent pattern over long timescales. The effect is wavelength-dependent, with  $u$ -band showing the strongest decreases in sensitivity at the mosaic edges relative to the centre,  $g$ -band showing very minor evolution, and  $z$ -band showing the inverse behaviour of increasing edge sensitivity compared to the mosaic centre. In  $u$  band, the four corner CCDs in the mosaic may change their average sensitivity relative to central CCDs as rapidly as 1% per day before they converge to an aggregate sensitivity loss of 10–20%. To mitigate this effect, we gather twilight flatfields from  $\pm 10$  days around each observing night in order to approximate the behaviour in the middle of that span.

Twilight flats are obtained in two opposing position angles (PAs) during each twilight period, so that the large scale gradient in the sky emission can be cancelled out. Within the  $\pm 10$ -day span (with hard cutoffs imposed for detector warm-ups or other configuration changes), the potential input frames are grouped by PA, and a tolerance-testing procedure is applied to remove flatfield affected by patchy clouds. Within each PA, the valid inputs are median-combined after rescaling the counts by the mean of the

<sup>1</sup>See <http://tdc-www.harvard.edu/wcstools/index.html>.

<sup>2</sup>See <https://fits.gsfc.nasa.gov/registry/tpvwcs/tpv.html>.

central 8 CCDs. For the opposing PAs within each twilight, the PA-medians are then mean-combined with equal weighting. Finally, the twilight-means are combined in a weighted mean, where the weights are taken as the number of contributing input frames, resulting in the master flatfield.

After dividing the science frame by the master flat, a small additive shift is applied between the two halves of a given CCD to ensure a smooth background level across the image.

#### 4.5 Fringing correction

Similar to the bias PCA method described in Section 4.3, to correct for fringing in the *i*- and *z*-band images, we generate a set of fringing PCs for each filter and each CCD, this time treating the whole CCD at once. The inputs for the PC creation were  $\sim 5\,000$  Main Survey images in each filter processed as part of SMSS DR2. We employed 3 PCs for *i*-band images and 10 PCs for *z*-band images, fitting to a background-subtracted and object-masked science image, and subtracting the resulting fringe pattern. The fringe PCs remain the same as for DR2 and DR3, and more details can be found in Onken et al. (2019).

#### 4.6 Additional pixel masking and image compression

During a portion of the survey operations (MJD = 57 290–58 323), ground loops in the detector electronics gave rise to correlated fluctuations in the bias level across the entire mosaic, which took the form of a spike in counts with adjacent count-depressed pixels on either side along the same row. To mask the affected pixels, we median-stack each of the four 8-CCD groups that shared a particular readout timing (after masking any detected astronomical sources). Sources with  $\geq 7$ -sigma positive fluctuations were flagged in the pixel masks (see Table 4), as was one pixel on either side.

Cosmic rays are then identified using the LACOSMICX software package,<sup>k</sup> a PYTHON implementation of the L.A.Cosmic routine (van Dokkum 2001). Each amplifier is treated separately to allow for the use of specific read noise settings. The affected pixels are replaced with values typical of the local background, but the modified pixels are also flagged in the pixel masks.

The calibration process transforms the original 16-bit integer image data into 32-bit floating point values. However, because of the significant readnoise in the SkyMapper images, we are able to round the data back to 16-bit integers without suffering from significant degradation in data fidelity. To avoid truncation of the noise around low sky count levels, the allowed integer range is  $-100$  to  $65\,435$  (by setting the BZERO header keyword to  $32\,668$ ). This transformation then allows us to reduce the data storage footprint by losslessly compressing the images using CFITSIO's FPACK routine (Pence et al. 2009). Compared to the 32-bit floating point images, the conversion and compression amounts to a factor of  $\sim 4$  reduction in disk space. The pixel masks, natively having 8-bit integer format, are also compressed, reducing the footprint by a factor of  $\sim 32$  because of the sparse nature of the flagged pixels.

#### 4.7 Photometric measurements

We run SOURCE EXTRACTOR on each CCD, with a detection threshold of  $1.5\sigma$  and a Gaussian filtering function adapted to the

median image FWHM. We measure photometry in a series of circular apertures (diameters of 2, 3, 4, 5, 6, 8, 10, 15, 20, and 30 arcsec). We provide SOURCE EXTRACTOR with the CCD-specific pixel mask as a Flag image and a version of the global bad pixel mask as a Weight image (to un-weight bad pixels).

When SOURCE EXTRACTOR measured the photometry, the input gain value was provided as 1, rather than the  $0.75\text{ ADU e}^{-1}$  determined from the preflash images (see Section 2.1). The consequence is a slight overestimate of the magnitude errors for each individual measurement, which in the limit of high source counts and low background, asymptotes to a value 15% too large. Because of the significant time that would be required to revise the gain value used by SOURCE EXTRACTOR in DR4, we leave the photometric errors unmodified. In practice, the impact for each *object* (those in the master table) is not nearly as significant, because the final photometric errors in each filter are derived from the outlier-clipped median absolute deviation (as described in Section 5.5).

##### 4.7.1 Aperture corrections and PSF variation

The sequence of aperture magnitudes provides a growth curve, which depends on object morphology. For point sources, the growth curve has a fixed shape, which can be used to infer total point source photometry from any aperture magnitude. However, the PSF shape drifts across the focal plane and affects the required aperture correction. Thus, we determine the PSF correction for each aperture in each image as a function of position using unsaturated bright stars. We also use this information also to estimate a PSF magnitude for each source from the 1D sequence of aperture magnitudes (see next section).

In other surveys, PSF magnitudes are commonly obtained by PSF fitting to sources in 2D image data. Our process works on table data and is equally robust for isolated sources. However, blended sources are not correctly deblended by our PSF magnitude calculation; instead, we provide warning flags based on the brightness difference and distance to any neighbouring objects, which specify (per filter) whether the PSF magnitude is likely compromised (see the description of the FLAGS\_PSF column in Section 5.5).

For each CCD, we derive aperture corrections for each aperture smaller than 15 arcsec by fitting the flux ratio between the  $15''$  -aperture and the aperture in question as a function ( $x, y$ ) pixel coordinates. The 2D linear gradient is intended to mitigate the large-scale variations in the PSF shape across the mosaic (see Section 2.2). The fit is performed iteratively with 4 cycles of  $2.5\text{-}\sigma$  clipping. In this process, we preference sources that have counterparts in the photometric zeropoint catalogue (described further below), but relax that requirement when the number of such matches is less than 8 per CCD. If fewer than 5 stars are available to fit the ( $x, y$ ) plane, we apply just the median aperture correction to all sources in the CCD.

##### 4.7.2 PSF magnitude calculation

During the preparation of DR4, it was discovered that the aperture corrections were not being correctly propagated to the associated magnitude errors in previous DRs, leading to underestimates of the uncertainties for both the aperture-corrected magnitudes (E\_MAG\_APCnn for aperture *nm* between 02 and 10 arcsec) and the 1D PSF magnitude derived therefrom (E\_MAG\_PSF, along with the flux versions of the latter, E\_FLUX\_PSF). Consequently, the per-object mean magnitudes ( $\{f\}$ \_PSF for filter  $\{f\}$ ) and

<sup>k</sup><https://github.com/cmccully/lacosmicx>.

uncertainties ( $E_{\{f\}}_{\text{PSF}}$ ) were incorrectly weighted in the master table, and the estimates of photometric variability (the per-epoch CHI2VAR and the per-object  $\{f\}_{\text{RCHI2VAR}}$ ) were overestimated in a manner that worsened for brighter stars. These deficiencies have been rectified by suitably propagating the aperture corrections and their uncertainties.<sup>1</sup>

The DR4 approach to constructing the one-dimensional point-spread function (PSF) magnitudes is also different from previous DRs. In DR4, we consider each *annulus* of aperture-corrected flux for the 7 smallest apertures (diameters of 2–10 arcsec), rather than a curve-of-growth approach that uses the entire flux within each aperture. We calculate the PSF magnitude and its uncertainty from the weighted mean annulus-corrected magnitude and the error in the weighted mean, as well as the  $\chi_{\text{red}}^2$  value relative to the expectations for that CCD's aperture corrections. Fitting annuli to the PSF model – other surveys often do this per-pixel from the image data, whereas we do it in the tabulated fluxes per-annulus by differencing the nested aperture fluxes – allows us to more properly propagate the photometric errors than had been done in previous SMSS DRs.

#### 4.8 Photometric Zeropoint

For DR4, we adopt an entirely new photometric zeropoint (ZP) catalogue, based on synthetic photometry derived from the low-resolving-power ( $R \sim 50$ ) *Gaia* DR3 spectroscopic data (Gaia Collaboration et al. 2023a; De Angeli et al. 2023) and the SkyMapper photometric bandpasses (Bessell et al. 2011, and also available through the Spanish Virtual Observatory's Filter Profile Service<sup>m</sup>). The *Gaia* team used the GAIAXPY PYTHON package on the set of over 200 million BP/RP spectra to produce synthetic photometry in the SkyMapper filters (Gaia Collaboration et al. 2023b). By default, the *Gaia* fluxes are convolved with the filter throughput and mean CCD response, as well as the typical atmospheric transmission for an airmass of 1 (as given in the unprimed columns of Table 2 of Bessell et al. 2011).

For DR2 and DR3, the SMSS utilised the ATLAS All-Sky Stellar Reference Catalog (known as Refcat2; Tonry et al. 2018), which brought together a variety of survey data to provide a well calibrated catalogue of *griz* photometry between 6 and 19 mag. However, this left the calibration for the SMSS *u*- and *v*-bands relying on extrapolations to shorter wavelengths. Recalibrations of the *uv* photometry have been derived based on stellar-colour regression models (Huang et al. 2021), but the available spectroscopic data did not sample the full footprint of the SMSS, and still extrapolated the short-wavelength properties based on their stellar classifications.

In contrast, the *Gaia* spectroscopic sensitivity to wavelengths as short as 330 nm anchors the *u*- and *v*-band data in a way that was not possible for the calibration of previous DRs. However, initial testing with the synthetic photometry for *u*-band revealed that the uncertainties in the short-wavelength *Gaia* sensitivity led to undesirably large errors in the predicted *u*-band data for stars with high-quality CALSPEC data. For example, the predicted photometry for CALSPEC stars with  $(B_p - R_p) < 0$  mag was too faint by  $\approx 0.25$  mag.

As a result, the *Gaia* team kindly reprocessed their spectra with a modified *u*-band throughput model having a cutoff at 340 nm instead of 330 nm. While this left an even larger fraction of the *u*-band throughput shortward of the *Gaia* cutoff,<sup>n</sup> it substantially reduced the resulting photometric scatter of the comparison stars. We determined a residual colour-term as a difference between the synthetic photometry of CALSPEC stars using our full *u* bandpass and the *Gaia* prediction that included the cutoff (both included the red leak). The *u* band magnitudes in the zeropoint catalogue were then corrected with the residual, represented by the following 4-part relation:

$$u_{\text{pred}} = u_{340} + \begin{cases} -0.045 + 0.125 \times \alpha & \text{if } \alpha \leq -0.2 \\ -0.015 + 0.250 \times \alpha & \text{if } -0.2 \leq \alpha \leq 0.2 \\ 0.000 + 0.175 \times \alpha & \text{if } 0.2 \leq \alpha \leq 0.6 \\ 0.097 + 0.8 \times (\alpha - 0.5)^2 & \text{if } 0.6 \leq \alpha \end{cases} \quad (1)$$

where  $\alpha = (B_p - R_p) - E(B_p - R_p)_{\text{gspphot}}$ , the extinction-corrected colour, which was only permitted over the interval  $-0.6 \leq \alpha \leq 1.0$ , and the latter term is the colour excess inferred from *Gaia*'s best-fitting GSP-Phot Aeneas library<sup>o</sup>.

The ZP catalogue was further restricted to *Gaia* sources having the following characteristics:

- synthetic magnitude < 16 mag
- synthetic signal-to-noise > 20
- RUWE < 1.4
- PHOT\_VARIABLE\_FLAG  $\neq$  "VARIABLE"
- IPD\_FRAC\_MULTI\_PEAK < 7
- IPD\_FRAC\_ODD\_WIN < 7
- $|C^*/\sigma_{C^*}| < 2$

where the central 4 constraints refer to the standard columns from the `gaiadr3.gaia_source` table, and  $C^*$  is the colour-corrected indicator of extended source flux from Riello et al. (2021), which we require to be less than  $2\text{-}\sigma$  from the 0-value of point sources. In addition, for *u*-band only, we apply the following condition:

- $E(B_p - R_p)_{\text{gspphot}} \leq E(B - V)_{\text{SFD}} + 0.1$   
OR
- $(E(B_p - R_p)_{\text{gspphot}} \leq 1.2)$   
AND  
 $(E(B_p - R_p)_{\text{gspphot}} \leq 1.2 \times E(B - V)_{\text{SFD}})$ ,

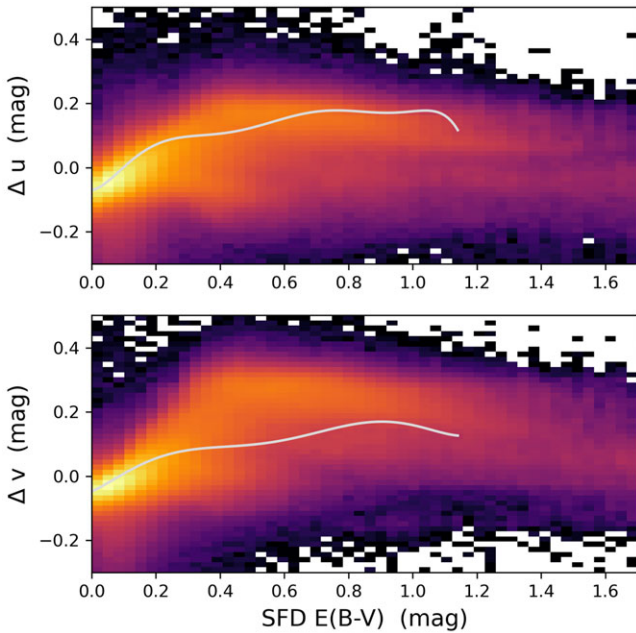
where the SFD colour excess is drawn from the maps of Schlegel, Finkbeiner, & Davis (1998). Based on the reddening coefficients of Casagrande et al. (2019), we expect a mean relation of  $E(B_p - R_p) = (2.905 - 1.75) \times 0.86 \times E(B - V)_{\text{SFD}} \approx E(B - V)_{\text{SFD}}$ . After applying these conditions, we have an all-sky ZP catalogue with between 6.3 million and 95.9 million sources per filter (from *u*- to *z*-band, growing in number by a factor of  $\sim 1.8$  with each progressively redder filter). We note that the synthetic photometry of the

<sup>1</sup>We make note here of the convention adopted throughout the paper, in which the available tables of the data release are indicated in lower-case script as, e.g. `master`, while columns within each table are indicated as upper-case script, e.g. `OBJECT_ID`.

<sup>m</sup><http://svo2.cab.inta-csic.es/theory/fps/index.php?mode=browse&name=SkyMapper>.

<sup>n</sup>The *u*-band's throughput, relative to its peak at 350 nm, is 80% at 340 nm, 50% at 330 nm, and 20% at 320 nm.

<sup>o</sup>See the *Gaia* DR3 Documentation, section 11.3.3, 'General Stellar Parametrizer from Photometry (GSP-Phot)'.



**Figure 7.** Photometric differences (DR4 – DR2) in  $u$ -band (top) and  $v$ -band (bottom) as functions of  $E(B - V)$  drawn from the Schlegel et al. (1998) reddening maps. Lighter colours indicate logarithmically higher density of points. The polynomial corrections derived by Huang et al. (2021, 2022) are overplotted with white lines. The new ZP catalogue naturally corrects for the biases present in the SMSS DR2/DR3 photometric calibration and reproduces the trends found by Huang et al. (2021).

ZP catalogue has not yet undergone the standardisation process applied by the *Gaia* team to other bandpasses (Gaia Collaboration et al. 2023b), but the worst effects of the ‘hockey stick’ feature, arising from a magnitude-dependent background overestimation, will be avoided by the cutoff at 16 mag.

Previous work by Huang et al. (2021, 2022) used stellar-colour regression (SCR) techniques to predict SkyMapper photometry based on precisely measured stellar parameters from the third data release of the Galactic Archaeology with HERMES survey (GALAH+; Buder et al. 2021). Based on SMSS DR2 and sparsely sampled spectroscopic data from GALAH+, Huang et al. fit high-order polynomials to describe the inferred  $u$ - and  $v$ -band magnitude biases as functions of Galactic reddening (using the  $E(B - V)$  maps of Schlegel et al. 1998) as well as spatial position.

By comparing the SMSS DR4 magnitudes to those of DR2,<sup>P</sup> we can investigate whether the new ZP catalogue has corrected the features identified by Huang et al. (2021). Fig. 7 shows a logarithmic density plot of SMSS DR4–DR2 photometry as a function of  $E(B - V)$  for 7 million well measured stellar sources in  $u$ -band and  $v$ -band (restricted to magnitudes 12–16, with photometric errors less than 0.05 mag, and CLASS\_STAR values above 0.9). We also plot the 7th- and 6th-order polynomials with the updated parameters of Huang et al. (2022) for the two filters, where we apply zeropoint shifts of  $(u, v) = (-0.064, -0.042)$  mag to the polynomials to match them to the bulk of our low-reddening data; such a shift was previously unconstrained as the polynomial solution only represented trends between different reddening levels. The new ZP catalogue recalibrates the SMSS photometry in a manner that naturally reproduces the SCR trends with reddening, but

which extends to higher reddening areas and accounts for localised variations.

Similarly, the photometric bias trends identified by Huang et al. (2021) as a function of sky position ( $(\alpha, \delta)$  for  $uv$ ,  $(l, b)$  for  $gr$ ) are well matched by the updated ZP catalogue. The six panels of Fig. 8 present the sky maps of the photometry differences in each filter between SMSS DR4 and DR2. They show spatial patterns that trace Galactic structure as well as other regions of high source density – precisely the sky areas in which the previous calibrations were known to be least reliable.

The improvements in photometric calibration will greatly enhance the utility of the  $u$ - and  $v$ -band photometry from SkyMapper for a variety of scientific purposes. In Section 6.5, we compare the DR4 photometric results to various external datasets.

#### 4.9 Photometric calibration of images

Using the instrumental PSF magnitudes derived above, we then derive the photometric calibration for each mosaic image by fitting a 2D plane of ZPs vs.  $(x_{\text{mosaic}}, y_{\text{mosaic}})$  to account for atmospheric transmission gradients across the large FoV. In a photometric night, the most extreme transmission gradients are expected in  $u$  band images taken near the South Celestial Pole at airmass  $\sim 2$ : across the image diagonal, the airmass may change by up to  $\sim 0.3$ , which will cause a throughput change of 0.2 mag (see also Table 1).

As with the aperture corrections, we perform 4 iterations of  $2.5\sigma$ -clipping. If the number of remaining stars falls below 6, we discard the image from consideration for the DR. Mostly, these ZPs produce better fits with lower root-mean-square (RMS) scatter than ZPs that are flat across the image, but in rare cases the gradient procedure converges to a bad fit. Hence, we generally prefer the gradient solution but we choose the flat ZP instead whenever it produces a significantly better RMS ( $\sigma$ ), i.e. when  $\sigma_{\text{flat}} < \sigma_{\text{gradient}} - 0.01$ . The final ZP fit is applied to all source magnitudes in the image.

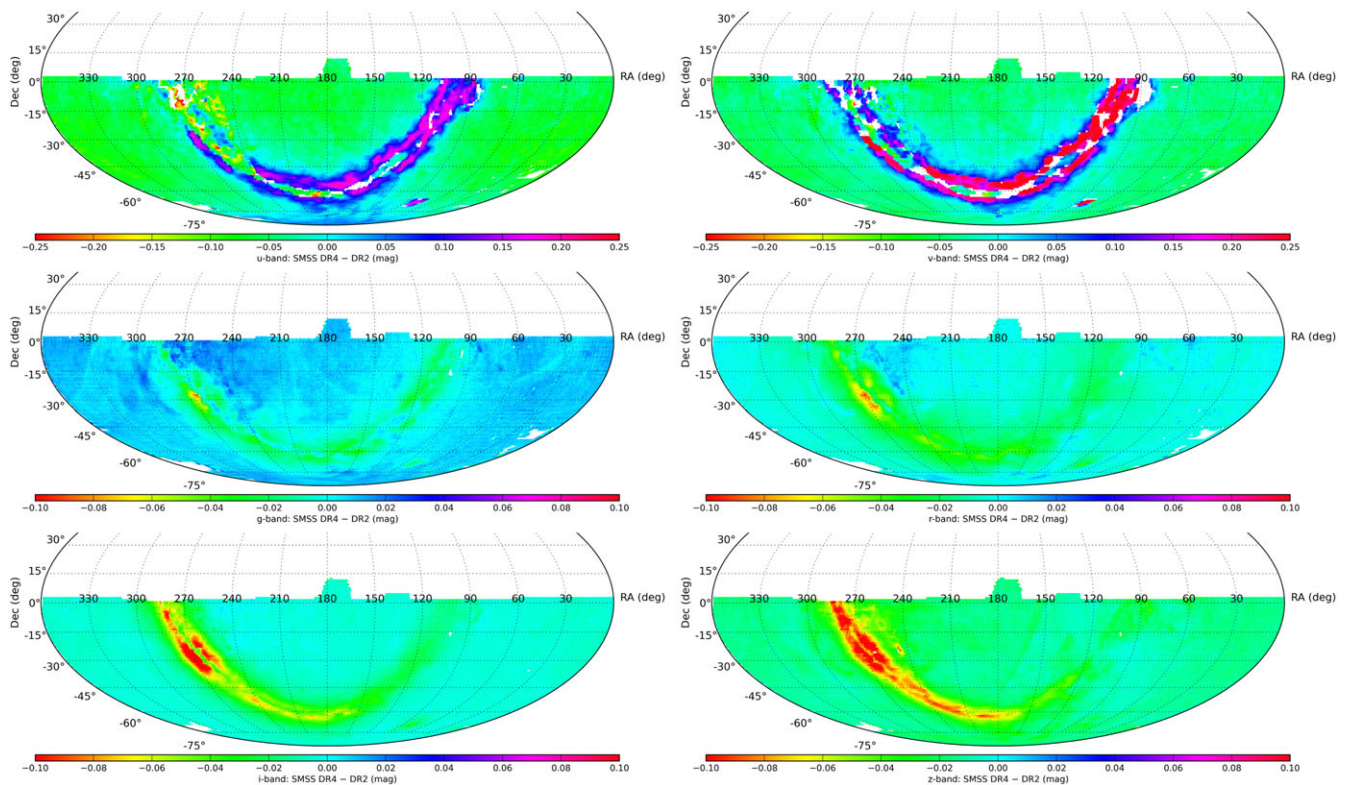
### 5. Distilling the master catalogue

Some applications of DR4 will involve a time-domain analysis and consider individual per-image detections of objects as described above and recorded in the photometry table. Many other applications would start better from a catalogue of unique astrophysical sources, here called the master table, and only use best estimates of per-object properties, or escalate into a time-domain analysis only after filtering the master table for relevant objects. In the master table, we record mean positions and mean per-filter magnitudes of objects by ‘distilling’ them from the sample of individual detections, while assuming that the objects are not variable. In addition, we record quality flags, numbers of good-quality images in each filter, information on neighbouring objects, and cross-match IDs with external catalogues. Here, we describe our intentions for the properties of the master table and the distillation procedures.

#### 5.1 Distill platform

In anticipation of this process, we transferred the  $\sim 15$  billion photometric measurements from the database server at NCI to a SkyMapper-dedicated server at ANU’s Mount Stromlo

<sup>P</sup>The SMSS DR3 photometry was calibrated in the same way as DR2.



**Figure 8.** Median photometric differences (DR4–DR2) per square degree for each of the 6 SMSS filters (*uvgriz*). Note that the colourbars cover the largest range for *u* and *v*, and smaller ranges for the other four filters. The new ZP catalogue has resolved the known systematic biases present in the DR2/DR3 calibration.

Observatory.<sup>q</sup> This 64-core machine, with over 500 GB of RAM, runs a similar PostgreSQL database (version 9.6.24) on which we have performed the final distillation of mean object properties for each SMSS DR to date. A master shell script launches a sequence of PostgreSQL scripts (parallelised where appropriate) that builds the master table via the steps below.

From the point-of-view of operational robustness, it is advantageous to move data into databases as soon as feasible and execute any further processing within the database (Thakar 2008). This allows one to take advantage of the fact that database engines are well-tested and robust enterprise-grade systems with built-in mechanisms to ensure integrity even during drastic failures such as power interruptions. However, our database instance is limited in its multiplex factor for parallel operations, and in a couple of instances, bottlenecks meant that some processes were prohibitively slow compared to running them through standard Linux shell commands in the multi-processor environment at NCI. For this reason, a few steps in the DR4 process were shifted to the latter. End-to-end the distill process for DR4 has still taken six months.

## 5.2 Selection of images and detections, flags

We first discard images that fail to meet any of the three quality levels in Section 6.2, and further remove any individual CCDs for

<sup>q</sup>The images and pixel masks remained on disk at NCI, from where they will be made available through the data access tools described in Section 7.1.

which the WCS solution yields a distance between opposite corners that differs from the expected value by more than 1 arcsec in either direction.<sup>r</sup>

For each of the 6 filters, we restrict the photometry tables to those images and CCDs which remain, while simultaneously applying a number of bitmasks to the FLAGS column (where bits up to values of 128 are retained from SOURCE EXTRACTOR). In the following, we describe the meaning of the full list of extra flags, irrespective of whether the flags are determined per detection or, later in the distill process, per object:

- 256 – indicating the multiple detections in a single image that were spatially linked to a single master table object (applied in a later step of the distill process). This occurs as variations in image quality may sometimes render double sources separate and sometimes merged. The rate varies from 1 in 45 000 detections in the larger-FWHM *u*-band images to 1 in 800 detections ( $\sim 5$  million out of  $\sim 4$  billion) in the smaller-FWHM *z*-band imagery.
- 512 – indicating very faint detections that are considered dubious and a potential source of error. This flagging ensures that very faint detections in short exposures are ignored when the Main Survey or other deep observations can better define their distilled properties, while faint

<sup>r</sup>Because the parameters of the WCS solutions are not numerically bounded, the best high-order fit to the brightest  $\sim 10$  000 stars in a mosaic image sometimes produces badly extrapolated coordinates, especially in the mosaic corners.

detections in deep images are still included into the master table (applied before distilling detections).

- 1 024 – indicating detections that appear much more concentrated than point sources, where the smallest aperture magnitude (2 arcsec) is significantly brighter than the largest aperture magnitude (15 arcsec), suggesting they may be affected by uncorrected cosmic rays or transient hot pixels (applied before distilling detections).
- 2 048 – indicating detections close to bright stars, which may suffer from scattered light or represent unreal sources. Using the stars from the ATLAS Refcat2 catalogue (Tonry et al. 2018), we apply a flag to sources with  $\log_{10}(\text{radius}) < -0.2 \times m$ , where  $m$  is the PS1-to-SkyMapper transformed magnitude in the filter in question, down to limits of (4, 5, 8.5, 8.5, 6, 5) mag for ( $u, v, g, r, i, z$ ). In contrast to previous DRs, such objects are now allowed to enter the master table, as many of them will have useful measurements.
- 4 096 – indicating objects where all existing detections were affected by a bug in the distill code: the true flag values for individual images were unintentionally overwritten by NULL values, when their aperture correction failed, most often because the smallest (2 arcsec) aperture would have negative fluxes; these objects have no single detection without aperture-correction issues (3 263 075 objects, applied to the global FLAGS column in the master table after the distill).
- 8 192 – indicating detections from the very corners of the mosaic (beyond a radius of  $\approx 1.6$  deg), where the WCS solution was found to be systematically biased. Cross-matches to *Gaia* DR3 positions shows mean offsets larger than 1 pixel, with opposing corners showing consistent behaviour, and the other pair of corners being offset in the opposite sense. The trends appear to primarily be associated with the camera position angle (PA), where images acquired with PA=180 deg reverse the behaviour in each corner. The flag setting suppresses the inclusion of such detections in distilled mean properties unless the measurements are all bad (applied before distilling detections).
- 16 384 – indicating objects that were never detected on any image other than Standard field images; these objects are typically extended and of low surface brightness so that their centroid positions are not well constrained in some of the shallowest Standard field images and thus show up as separate objects not matched properly to the correct astrophysical object seen more clearly in deeper images (applied to the global FLAGS column as well as the per-filter  $\{f\}$ \_FLAGS columns in the master table after the distill; note that the FLAGS column for such objects is set to 16 384, rather than being OR-combined, and thus no longer reflects the OR-combined bits from all filters).

We also apply the final version of the photometric zeropoint – either with the spatial gradient or flat across the image – to all of the magnitude columns.

### 5.3 Merging detections into unique astrophysical objects

Within each filter, we then identify a set of ‘primary’ sources by spatially matching the full list of detected sources against itself;

in order to preference robust detections in good seeing, we use a maximum linking distance of 1.5 arcsec or  $1.5 \times$  the FWHM of the object, whichever is greater. From the detections in this neighbourhood we select as a *primary* counterpart the highest-ranked detection by first sorting detections by the binary choice of  $\text{FLAGS} < 8$  (i.e. no corruption of the SOURCE EXTRACTOR photometry), then ranking images with IMG\_QUAL of 1 or 2 higher than images with IMG\_QUAL of 3, and then sorting for the highest peak counts divided by the larger of 4 pixels or the measured object FWHM.

The per-filter primary detections are collated into a single all-filter table which is then spatially cross-matched against itself. In this step, a linking distance of 1.5 arcsec is adopted and the same high-count-good-seeing metric is applied to identify the master source which seeds the eventual master table. Based on those master source positions, any object identifiers that can be retained from SMSS DR3 are adopted (using a spatial cross-match distance of 2 arcsec, and keeping only the closest match if multiple new master sources were matched to a DR3 object). Objects newly identified in DR4 (nearly 122 million sources) are assigned OBJECT\_ID values beginning from 2E9 (while sources new to DR3 were assigned OBJECT\_ID values beginning from 1E9).

The list of individual detections are then associated with those master OBJECT\_ID identifiers by adopting the closest master source within 10 arcsec (allowing a larger distance to account for any cumulative offsets between links spanning the six filters). This procedure leaves only a very small number of detections unassociated with a master table entry: a total of 1 376 detections, for a rate of 1 in 10 million. If, conversely, a single master object is associated with multiple detections in a given image, all of those detections have a bit value of 256 added to their FLAGS column in the photometry table.

### 5.4 Mean properties per object

To calculate the mean object properties in each filter, we then select those OBJECT\_ID-associated detections that appear to be of good quality. We first consider all detections for which  $\text{FLAGS} < 4$  and NIMAFLAGS recorded  $< 5$  masked pixels within the source’s isophotal area. The number of good detections thus defined in the filter  $\{f\}$  is recorded in the column  $\{f\}$ \_NGOOD. We then omit all  $\text{IMG\_QUAL} = 3$  detections from the mean properties calculation when quality level 1 or 2 data is available for a source and set the USE\_IN\_CLIPPED column of the low-quality entries in the photometry table to  $-1$ . We also add up the number of good detections over all six filters for the total number of good detections recorded in column NGOOD.

Then we make a further distinction: for the determination of, e.g. mean magnitudes of an object, we assume *a priori* that the object is not variable. We enhance signal-to-noise by clipping potential outlier measurements, even if they have good-quality flags. Such outliers may be faulty measurements that are not recognised as such by our data processing, or they may be genuine astrophysical signals such as a rare binary eclipse affecting only one in several available magnitude measurements. Thus, the USE\_IN\_CLIPPED column uses three values to indicate data levels for the calculation of mean properties:  $-1$  denotes bad data excluded from a distilled mean, 0 denotes good data clipped to reduce scatter, and 1 means data retained after clipping and used for the calculation of the final mean and error.

We separately average properties of sources for which *only* saturated detections or detections with otherwise bad flags exist in a given filter, to ensure they are not omitted from the master table by the restrictive quality selections above. The photometry of these objects may be useful but unreliable, and the errors on their photometry will be unknown; we indicate this further by listing their distilled magnitude errors as NULL. Their coordinates are listed with useful errors derived from the scatter among the detections, but they may still be on average less accurate than those of well-measured objects.

### 5.5 Mean photometry per object

For each object in each filter, we then compute the weighted median PSF magnitudes and their median absolute deviations (MADs, measured with respect to the weighted median). The USE\_IN\_CLIPPED column of the photometry table is set to 1 for magnitude values which are less than 3 MADs (or 3 times the square-added errors of the PSF magnitude and ZP RMS, if that is larger) from the weighted median PSF magnitude, and 0 otherwise.

From the detections that survive the clipping process, we calculate for each filter a summary of characteristics for each object: the mean properties (with the magnitude estimates weighted by their individual errors, but other entries unweighted), as well as the maximum CLASS\_STAR and FLUX\_MAX values, the OR-combined FLAGS bits, and the sum of the NIMAFLAGS entries (capped at 32 767 to limit the column to a 16-bit representation). The weighting for the magnitude columns includes (by square-adding) the RMS of the photometric ZP fit to the image.

In contrast to previous DRs, which tabulated a  $\chi^2$ -like estimator of whether a source was consistent with being non-variable ( $\{f\}$ \_RCHI2VAR for filter  $\{f\}$ ), for DR4 we adopt a simpler metric of the minimum-to-maximum range of PSF magnitudes (with no clipping considered):  $\{f\}$ \_MMVAR. This excludes detections in IMG\_QUAL=3 images when better images exist, as well as any that were flagged as bad.

The master table has also been expanded from previous DRs to record the weighted mean 5" aperture magnitude (corrected for aperture losses as described above). This small-aperture photometry,  $\{f\}$ \_APC05, is useful for galaxy colours as it focuses on well measured inner parts, i.e. core- or bulge-dominated regions in nearby galaxies and near-total colours in distant galaxies. As in past DRs, Petrosian magnitudes are included (using standard SOURCE\_EXTRACTOR parameters) that integrate out to poorly measured peripheral regions of galaxies and thus produce noisier colours.

From the per-filter summarised characteristics, we generate the final source descriptors for the master table: weighted-mean positions and uncertainties; mean and RMS MJD of the observations; OR-combined FLAGS; summed NIMAFLAGS and number of good detections (i.e. those with USE\_IN\_CLIPPED > -1); mean FWHM (recorded in MEAN\_FWHM) and *r*-band RADIUS\_PETRO<sup>5</sup> values; and maximum FLUX\_MAX and CLASS\_STAR values.

A special case is the CHI2\_PSF parameter (determined from the light profile in a 15 arcsec aperture): the values in the photometry

table are first averaged per filter and then the largest value among the filters is used in the master table, because blended sources of different colours may only be recognised in some bands but not others. However, the largest-value selection implies that the CHI2\_PSF values in the master table are expected to be larger than 1 on average, as they are sampled from the larger-than-average tail of the distribution.

A cleaning is then applied to sources close to the RA = 0/360 deg divide to ensure appropriate calculation of the mean position and uncertainty.

We finally create a FLAGS\_PSF entry for each object to indicate whether their PSF magnitude may be biased by close neighbours adding flux to their apertures. We construct a bitmask by testing whether any of the neighbours are closer in arcsec than  $5 + 2 \times (m_{\text{object}} - m_{\text{neighbour}})$ . Neighbours closer than this limit are expected to brighten the PSF magnitude by over 0.01 mag, with no upper limit. The flagging is performed independently for each filter, with *u*-band to *z*-band encoded as descending bit values from 32 to 1. The procedure differs from previous DRs, where only the single closest source was considered, while other, potentially brighter, sources had been ignored.

### 5.6 Cross-matches to external catalogues

With the final coordinates for the master table established, we cross-match SMSS DR4 against the variety of other catalogues described in Section 7, as well as the next three closest sources within the master table itself.

One of the new cross-matches is the *Planck* satellite map of interstellar reddening, based on the generalised needlet internal linear combination (GNILC) method<sup>4</sup>. (Planck Collaboration et al. 2016) Comparisons of the GNILC reddening with the Schlegel et al. (1998) map from an earlier generation of infrared/microwave satellites (*COBE* and *IRAS*) have found that near-IR-bright galaxies show a smaller colour scatter with the GNILC maps, especially at low Galactic latitudes (Schröder et al. 2021). Thus, we provide the Schlegel et al. (1998)  $E(B - V)$  estimate alongside the GNILC  $E(B - V)$  estimate and its error in the master table.

Finally, to make the data available to the community, we transfer the master table, the modified photometry tables (now containing the OBJECT\_ID labels, additional FLAGS bits, and USE\_IN\_CLIPPED indicators), the images and ccds tables to the PostgreSQL database (version 9.5.25) underlying the SkyMapper node of the All-Sky Virtual Observatory (ASVO).

## 6. SMSS DR4 properties

In this section, we describe the image selection criteria for inclusion in SMSS DR4, and the resulting distribution of image properties and sky coverage. We then assess the astrometric and photometric performance of the data.

### 6.1 DR4 input images

SMSS DR2 and DR3 incorporated images taken up to March 2018 and October 2019, respectively. With DR4, we bring the processing of images up to September 2021. (Faults in the detector controller electronics on UT 2021-09-16 resulted in periods of time with incomplete mosaic operation, which will require

<sup>5</sup>During review of this manuscript, it came to our attention that the RADIUS\_PETRO values compiled in the master table had not been converted from the number of semi-major/semi-minor axis lengths to units of pixels. Petrosian radii in pixels can be computed from the photometry table as  $\text{RADIUS\_PETRO} \times \sqrt{A \times B}$ , and converted to arcseconds with the average plate scale of  $0.497 \text{ arcsec pix}^{-1}$ .

<sup>4</sup>We note a typo in Section 6 of Planck Collaboration et al. (2016), which reports the optical depth scale factor as  $1.49 \times 10^{-4}$  rather than  $1.49 \times 10^4$  (cf. Planck Collaboration et al. 2014).

**Table 5.** Requirements for image quality levels: Constraints for quality 3 also apply to quality 2, and those of quality 2 also apply to quality 1. All numbers are upper limits except for the number of ZP stars.

Property	Units	Passband					
		<i>u</i>	<i>v</i>	<i>g</i>	<i>r</i>	<i>i</i>	<i>z</i>
Image quality 1							
Zeropoint RMS	mag	0.05	0.033	0.015	0.015	0.015	0.02
Gradient ZP offset	mag			0.01			
Gradient ZP slope	mag px <sup>-1</sup>	6e-6		3e-6			
Image quality 2							
Zeropoint RMS	mag	0.07	0.06	0.04	0.04	0.055	0.055
PSF FWHM	arcsec	4.00	4.00	4.00	3.90	3.75	3.60
Background	counts	300	300	1 000	1 000	3 000	3 000
Gradient ZP offset	mag			0.025			
Gradient ZP slope	mag px <sup>-1</sup>			1.2e-5			
Airmass				2.02			
Image quality 3							
Zeropoint RMS	mag	0.12	0.10	0.08	0.08	0.08	0.08
Gradient ZP slope	mag px <sup>-1</sup>			2.1e-5			
PSF FWHM	arcsec			5.0			
PSF elongation				1.4			
Background	counts			5 000			
Number of ZP stars				>20			
Throughput loss	mag			1.0			
WCS RMS	arcsec			0.5			
Compressor noise	fraction			0.2			

pipeline modifications to properly calibrate, and thus defined a natural cutoff for DR4.)

In addition to the Survey image types released previously (Shallow Survey and Main Survey), DR4 has been allowed to consider the SMT survey and other non-Survey images acquired since the start of the SMSS, provided they meet the various quality metrics indicated in Section 6.2. Finally, the Standard fields containing the CALSPEC standard stars (see Section 3.1) have also been included in DR4. The incorporation of these additional images means that the observational cadence is enhanced, even within the range of dates covered by previous DRs.

## 6.2 Image selection criteria

We use three image quality categories with requirements that depend on the filter, and reject all images that do not conform with minimum requirements. The requirements consider the RMS ( $\sigma$ ) of the ZP determination that is used for the frame (either flat or gradient) and also the gradient fit parameters themselves when the ZP gradient solution is adopted for an image. Furthermore, high airmass and PSF FWHM beyond a threshold affect the image quality level irrespective of the ZP quality, and are also constrained (see Table 5).

Images that do not meet the criteria for quality level 3 are not included in the data release. When a unique object in a unique filter has images from both quality level 3 and the better quality levels (1 and 2), then the images of quality level 3 are not included for distilling quantities in the master table of mean object properties

(further details are given in Section 5), although the data is still present in the photometry table and useful for probing behaviour of variable sources at additional epochs. Thus, in the interest of better characterisation of non-variable sources, data at quality level 3 only work their way into the master table if no better-quality images exist in that filter for a given object.

In Table 6, we present the distribution of DR4 images by filter and image quality. Overall, more than 80% of the DR4 images have image quality 1 or 2.

The distribution of images over time is shown in the two panels of Fig. 9, which present the number of images as well as the total exposure time, in intervals of 10 nights. A number of periods of technical downtime are evident as gaps in the time coverage. The evolution of the system throughput is seen in the *g*-band zeropoint over time (Fig. 10), with periodic improvements coming from the cleaning of the telescope optics.

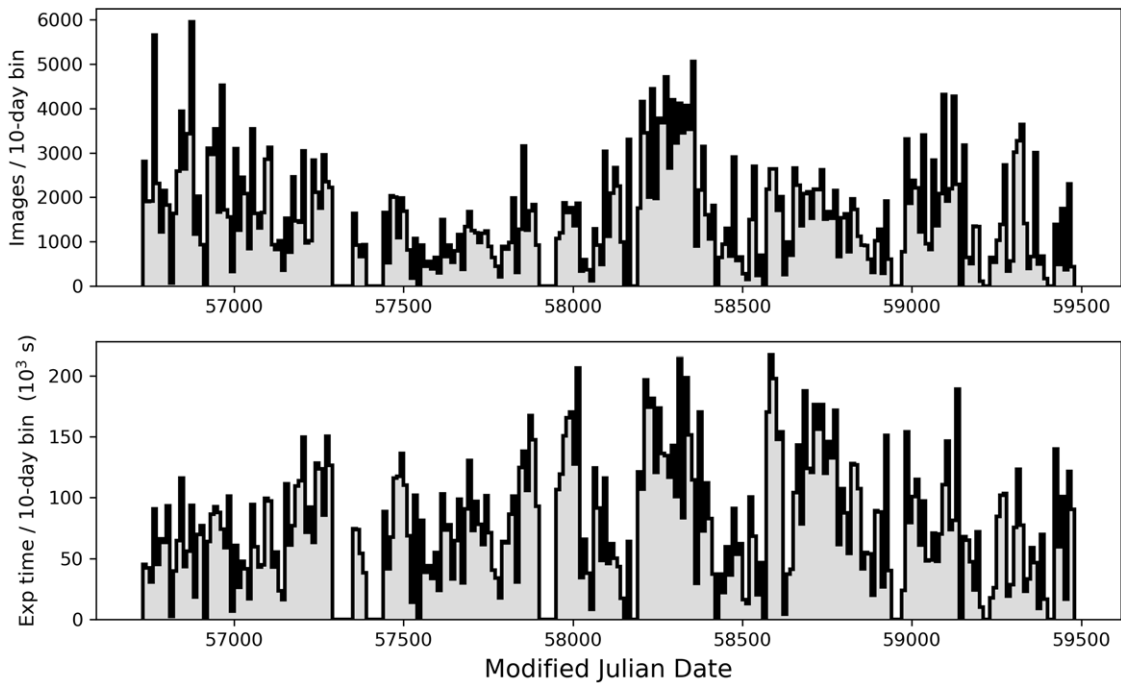
## 6.3 Sky coverage and completeness

The sky area covered in DR4 is essentially the complete sky south of declination +16 deg plus a few special fields north of that line. The most intense coverage in terms of depth and repeat observations is south of the celestial equator. However, it is not guaranteed that every position in the southern sky has fallen onto good pixels in all filters, because the bad pixel patterns, the telescope dither pattern, the pointing accuracy of the telescope, and the de-selection of bad images for the release all affect the coverage of the data set.

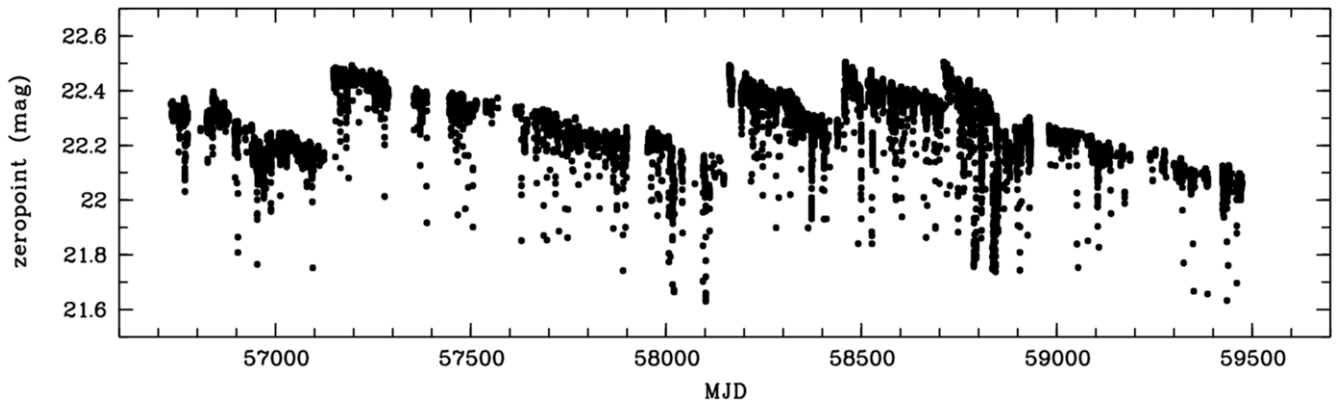


**Table 6.** DR4 includes all images taken with SkyMapper that pass the quality criteria, irrespective of original purpose. Image properties are: final survey depth as peak in histogram of sources with  $\sim 0.1$  mag errors in the PSF magnitude (source detection was made on individual images); saturation magnitude in the Shallow Survey; distribution of exposure times ( $t_{\text{exp}}$ ) across all images and for the Shallow and Main Survey specifically; fraction ( $f_{\text{image}}$ ) of Shallow and Main Survey images among the DR4 set; and image quality distributions. A small fraction ( $< 0.5\%$ ) of  $u$  and  $v$  images have 300- or 400-s durations.

Filter	10 $\sigma$ depth	Shallow saturation	Min–Max $t_{\text{exp}}$	Mean $t_{\text{exp}}$	Shallow/Main $t_{\text{exp}}$	Shallow/Main $f_{\text{image}}$	Number of images		
	ABmag	ABmag	(s)–(s)	(s)	(s)/(s)	(%)/(%)	Qual 1	Qual 2	Qual 3
$u$	18.6	8.9	15–600	61.6	40/100	51/31	35 300	7 740	16 750
$v$	18.9	8.2	15–600	52.5	20/100	49/30	23 077	26 779	13 535
$g$	20.5	9.4	0.1–600	43.1	5/100	46/22	25 616	28 874	15 478
$r$	20.2	9.4	1–400	40.2	5/100	48/19	20 827	34 590	13 875
$i$	19.4	9.5	1–600	48.0	10/100	41/35	31 584	39 635	11 385
$z$	18.6	9.6	1–300	50.2	20/100	47/38	28 615	32 472	11 091



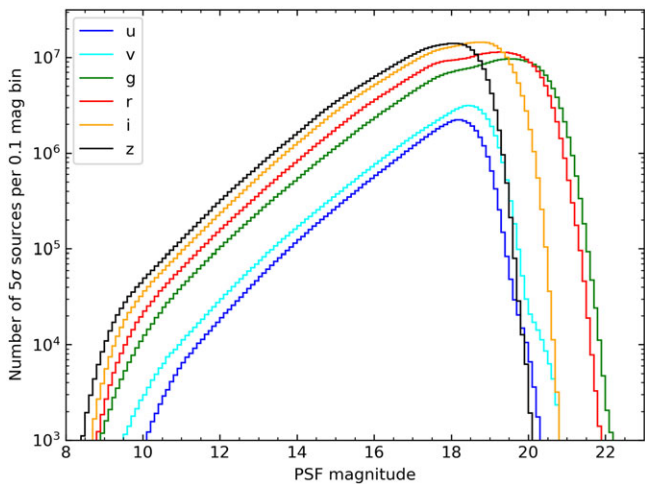
**Figure 9.** For the images included in DR4, we show the number of images (*top*) and their total exposure time (*bottom*), in bins of 10 nights. Longer periods of downtime include three failures of the cooling system (around MJD 57300, 57400 and 57900).



**Figure 10.** Time series of  $g$ -band image zeropoints in DR4, normalised by airmass and exposure time, from 15 March 2014 to 16 Sep 2021 (MJD 56733 to 59473). The gradual degradation of the ZP stems from loss of mirror reflectivity of  $\sim 1\%$  per month. Sudden improvements appear after mirror cleaning (on MJD 57147, 58150, 58448 and 58710).

**Table 7.** Galactic Plane Coverage: DR4 contains overall  $\sim 3.5\times$  as many images as DR2, but at low Galactic latitudes DR4 has  $\sim 7\times$  and  $\sim 12\times$  as much data in the  $u$  and  $v$  bands, respectively. Astrometry methods used in earlier releases often failed at calibrating  $uv$  data in crowded fields.

Filter	DR4 CCDs		DR2 CCDs	
	$ b  < 5$ deg	$ b  = 5\text{--}10$ deg	$ b  < 5$ deg	$ b  = 5\text{--}10$ deg
$u$	137 911	183 870	20 559	40 895
$v$	141 604	189 137	11 647	26 622
$g$	137 738	177 893	42 467	46 338
$r$	143 541	181 825	42 796	46 911
$i$	171 576	208 026	54 018	58 697
$z$	174 562	215 288	55 254	61 200



**Figure 11.** Distribution of mean magnitudes in each SMSS filter for sources below  $0^\circ$  Dec having  $5+\sigma$  photometry in that filter.

Compared to earlier releases, DR4 has much improved coverage of well calibrated data near the Galactic Plane (see Table 7). Such data was previously missing, especially in the  $u$  and  $v$  filters, where the number of images in DR4 has increased by a factor of  $\sim 7$  and  $\sim 12$ , respectively (counted as individual CCDs). This is partly due to improved astrometry methods in DR4, where previously several CCDs in an image were lost for lack of reliable per-CCD astrometry; the mosaic-global astrometry in DR4 recovers those CCDs as embedded in a whole-mosaic solution. Another important reason is the new zeropoint star catalogue, which provides a higher density of sources, especially in the  $u$  and  $v$  band in regions with higher foreground reddening or source density.

Among the 700 million distinct objects contained in DR4's master table, we have magnitude estimates in the six SMSS filters for 9, 13, 59, 73, 85, and 72% of the objects in  $u$ ,  $v$ ,  $g$ ,  $r$ ,  $i$ , and  $z$ , respectively. The histogram of source magnitudes ( $5\sigma$  or higher) is shown in Fig. 11.

### 6.3.1 Completeness versus DES

We assess the completeness of the DR4 dataset by comparison to the deeper DES DR2 catalogue (Abbott *et al.* 2021), focusing on the  $r$ -band. In practice, the fraction of faint sources detected by SMSS will depend upon the particular mix of image seeing and photometric zeropoint for any given sky position, but we illustrate the typical behaviour.

To avoid any biases arising from sky location, we sample the DES data in a series of 30-arcmin-radius circles around every other SMSS field centre (which form a regular grid on the sky) in alternating Declination stripes, thus giving a spacing of about 4 deg in each direction. The SMSS field centres considered were also limited to those with at least three 100 s Main Survey exposures in DR4, to avoid artificially underestimating the completeness from those rare sky areas only observed with short exposures. From the DES catalogue, we select those objects identified as unsaturated point sources (`EXTENDED_CLASS_COADD=0` and `FLAGS_R<4`; see Abbott *et al.* 2021). Corresponding SMSS sources were required to be less than 1 arcsec from the DES positions.

Because of differences in the DECam and SkyMapper  $r$ -band filter curves, there are colour terms when comparing the DES and SMSS photometry. We correct for these by restricting the DES-measured colour range to  $0 < (g - r) < 1.3$  mag and fitting the SMSS-minus-DES magnitude difference,  $\Delta r$ . We adopt a colour correction relation of  $(-0.007 + 0.118 \times (g - r))$  that is applied to the DES  $r$ -band photometry to predict corresponding SMSS  $r$ -band magnitudes. In Fig. 12, we show the improvement in the photometric comparison before ( $\Delta r$ , top panel) and after ( $\Delta r_{corr}$ , second panel) applying the colour correction.

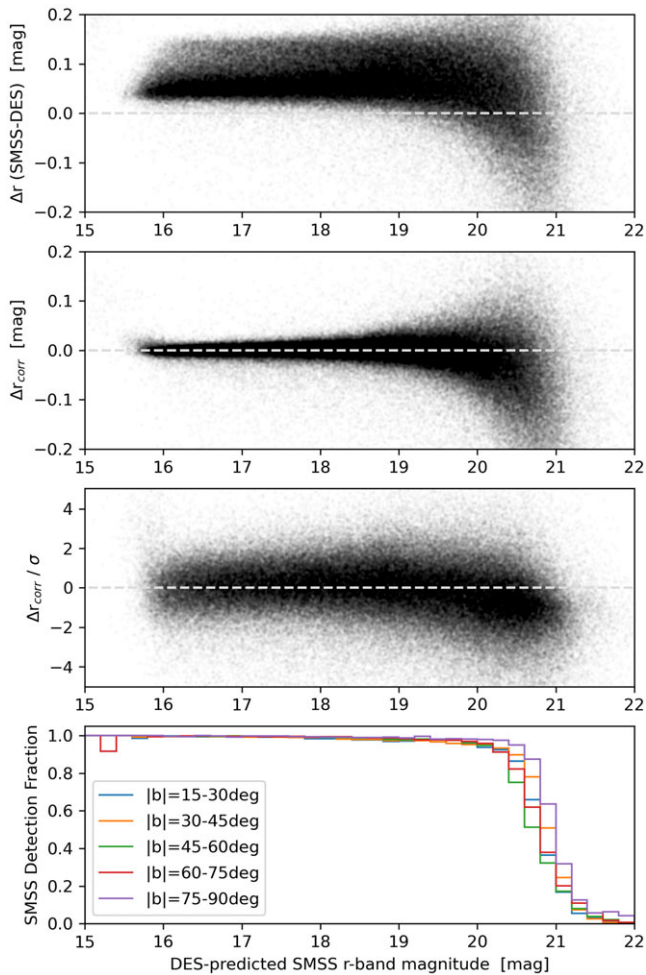
The third panel of Fig. 12 shows the colour-corrected magnitude difference normalised by the photometric errors (with the SMSS and DES errors being square-added). Until  $r \sim 20$  mag, the  $\Delta r_{corr}$  distribution is consistently distributed around zero, with fainter sources showing a slight bias towards brighter magnitudes in DR4. At  $r = 20.7$  mag, the median offset to brighter magnitudes begins to exceed  $1\sigma$ . The bottom panel of Fig. 12 shows the fraction of DES sources having corresponding SMSS matches as a function of  $r$ -band magnitude, and divided into different bins of Galactic latitude,  $|b|$ . The variations as a function of  $|b|$  are small, but the high-latitude, low-extinction region of the sky is the most complete in DR4. The 50% completeness level is typically around  $r = 20.9$  mag in these fields, a brightness at which the mean DR4 magnitude error for sources with just a single detection (which cannot take advantage of multiple measurements to reduce the uncertainty) is  $\sim 0.1$  mag.

Finally, Fig. 13 shows distribution of error-normalised, colour-corrected magnitude differences for sources with DES  $r$ -band magnitudes between 16 and 19. The dashed line overplotted is a Gaussian of unit standard deviation, which is an excellent representation of the distribution. The long tails in the histogram likely reflect the brightness changes of variable objects. Approximately 13% of the sources lie above the Gaussian shown in the Figure. This illustrates that the SMSS photometric errors are reliable indications of the magnitude uncertainty.

### 6.4 Astrometric performance

The WCS solutions derived for the SMSS DR4 images typically involved 800 stars from across the mosaic, with quartiles of  $\sim 500$  and  $\sim 1\,500$  stars. While no significant differences between filters are seen in the number of stars fit, the resulting RMS in the WCS fit<sup>u</sup> did vary from median values of 0.155 arcsec in  $u$ -band to 0.110 arcsec in  $z$ -band, likely due to the wavelength-dependent trends in seeing.

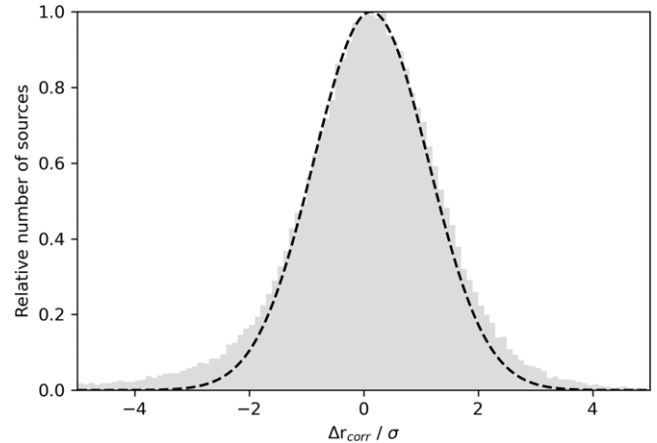
<sup>u</sup>Catalogued in the `WCS_RMS` column of the `images` table.



**Figure 12.** Difference in  $r$ -band photometry between SMSS DR4 and DES DR2 as a function of DES-predicted magnitude in the SMSS system. Grey dashed lines in the first three panels are drawn at a magnitude difference of zero. (*Top panel*) Raw photometric difference prior to colour correction (in this panel only, the  $x$ -axis shows the original DES  $r$ -band magnitudes). (*Second panel*) Photometric difference after application of the colour correction, significantly tightening the relation. (*Third panel*) Colour-corrected photometric difference in units of photometric error (square-adding the DES and SMSS errors). The SMSS errors are consistently representing the photometric differences until magnitudes fainter than  $r \sim 20$  mag, beyond which the SMSS photometry exhibits a slight bias to brighter measurements. (*Bottom panel*) SMSS detection fraction as a function of magnitude, plotted in different ranges of Galactic latitude, showing high completeness to  $r \sim 20$  mag.

We also see a date-related trend in the WCS RMS, worsening by  $\approx 50\%$  over the MJD range included in DR4. This effect may be driven by underlying trends in the median seeing, which also increased by  $\sim 1$  arcsec between 2014 and 2021 across all 6 filters. In addition, stellar proper motions since the *Gaia* DR2 epoch of 2015.5 (which anchor our WCS solutions) will contribute a small level of scatter for more recent observations.

Overall, amongst the master table sources with *Gaia* DR3 matches within 1 arcsec, the median astrometric offset was 0.13 arcsec, or 0.26 pixels. The sky map of median *Gaia* offset distance per square degree is shown in Fig. 14. The only regions in which the median offsets exceed 0.5 pixels (0.25 arcsec) are some highly crowded regions (e.g. the centre of the Large Magellanic Cloud) and at the edges of the SMSS DR4 sky coverage (where



**Figure 13.** Distribution of differences in colour-corrected  $r$ -band photometry between SMSS DR4 and DES DR2 in units of photometric error (square-adding the DES and SMSS errors). The sample of point sources was limited to DES  $r$ -band magnitudes between 16 and 19. The dashed line indicates a Gaussian of standard deviation equal to one, which fits the bulk of the population, indicating the reliability of the SMSS photometric errors. The sources at larger magnitude differences include variable sources.

the numbers of sources in the square degree is small and the WCS solution in a given image is least constrained).

We show in Fig. 15 how the offset from *Gaia* positions depends upon both the local source density (number of *Gaia* sources within 15 arcsec) and the source magnitude (SMSS DR4  $i$ -band). Aside from the extremes in  $i$ -band magnitude, the offsets are well behaved, with mild increases at both fainter magnitudes and denser fields.

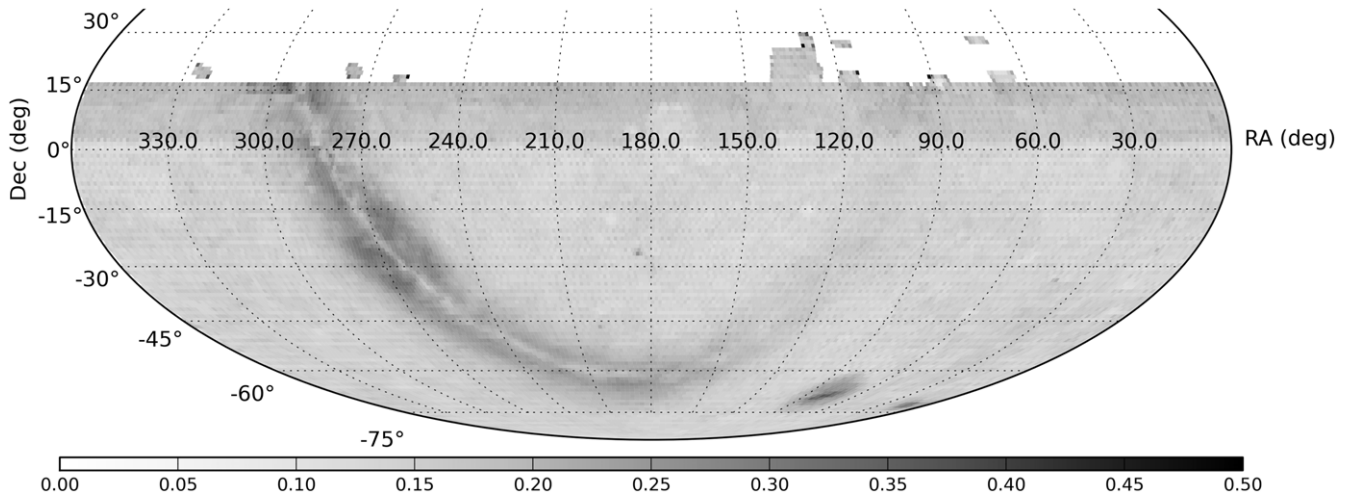
Fig. 16 shows the cumulative distribution of *Gaia* position offsets within illustrative slices of  $i$ -band magnitude and Galactic latitude. The distributions are broadest at faint magnitude and low Galactic latitude. The left-hand panel shows a steady improvement in the offsets towards brighter magnitudes, as expected from S/N considerations. The right-hand panel shows very consistent behaviour between  $|b|=20^\circ$  and  $80^\circ$ , indicating that once extreme values of source density are avoided, the SMSS positions are consistently estimated across the sky.

### 6.5 Photometric performance

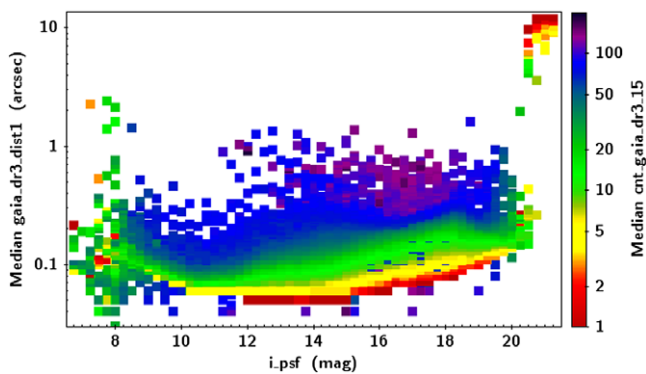
We first confirm that the calibration of the master table photometry yields results that match those of the ZP catalogue. Fig. 17 shows that the DR4 mean photometry of the ZP stars is well matched to the *Gaia* synthetic photometry. The small residuals in regions of high density arise because of the effects of crowding on the SkyMapper photometry.

In contrast to previous DRs – where the per-image ZP fits had a median RMS values in  $u$ - and  $v$ -band of  $\approx 0.1$  mag – with the revised ZP catalogue in DR4, more than 99% of  $u$  and  $v$  images have ZP RMS values smaller than 0.1 mag. The median RMS values in DR4 are 0.03 mag for  $u$  and  $v$ , 0.02 mag in  $z$ , and 0.01 mag in the other filters.

Returning to the full *Gaia* DR3 synthetic photometry catalogue (not only those retained in the ZP catalogue), we find the well known ‘hockey stick’ pattern of underestimated *Gaia* fluxes at faint magnitudes (Montegriffo et al. 2023; Gaia Collaboration et al. 2023b). However, between 9 and 17 mag, the median differences for filters  $griz$  remain within  $\pm 0.01$  mag, with no restrictions



**Figure 14.** Sky map of the median astrometric offset between SMSS DR4 and *Gaia* DR3; the colourbar indicates the offset in units of arcsec. The median offset per square degree ranges between 0.04 and 0.54 arcsec, with all values larger than 0.25 arcsec (0.5 pixels) occurring at the edges of SMSS DR4’s sky coverage or in regions of high density and source blending.



**Figure 15.** Median astrometric offset between SMSS DR4 and *Gaia* DR3 as a function of SMSS *i*-band magnitude; the colourbar indicates the median number of *Gaia* sources within 15 arcsec. Very bright sources (many of which have significant proper motion, uncorrected here) and very faint sources (for which the cross-match may be spurious) deviate from the expected trends of smoothly, but mildly increasing offset at both fainter source magnitude and increasing source density (where SMSS is subject to blending).

on colour, reddening, or crowding; the only requirements on the DR4 sources were to be unsaturated and for an absence of particularly bright close neighbours (bit value for the band in question being zero in `FLAGS_PSF`), which yielded between 35 million (*u*-band) and 120 million (*i*-band) matches to DR4. For *v*-band, the same  $\pm 0.01$  mag median behaviour holds true for a more limited magnitude range of 12.5–16.5 mag.

In *u*-band, we consider the *Gaia* synthetic photometry with the modified bandpass model that cuts off at 340 nm. We restrict the comparison to those *Gaia* sources falling within the valid range of Equation (1), namely,  $(B_p - R_p) - E(B_p - R_p)_{\text{gsp}}^{\text{phot}}$  between  $-0.6$  and  $1.0$  mag, so that the correction can be applied.<sup>v</sup>

<sup>v</sup>Omitting the correction of Equation (1) to the synthetic *u*-band photometry yields a near-linear trend of median magnitude difference between  $u=10$ –17 mag having a slope of 14 mmag mag<sup>-1</sup> and an offset of +85 mmag at  $u=10$  mag.

This retains 70% of the original *u*-band sample, and places the median magnitude difference within  $\pm 0.01$  mag between  $u=10.6$  to 17 mag.

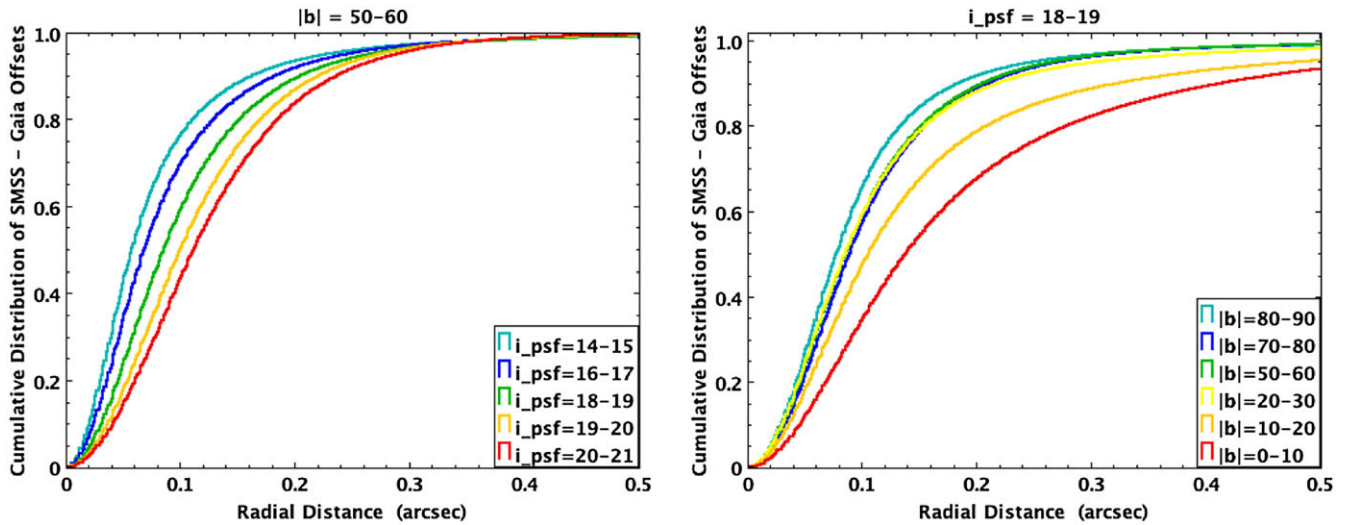
The 16–84-percentile half-range (akin to  $1\sigma$  for a Gaussian distribution) is less than 0.05 mag for all magnitude bins containing at least 100 sources down to 16.1, 17.0, 18.7, 18.2, 17.6, and 17.0 mag for *uvgriz*, respectively. Only *u*-band violates that constraint on the bright end, reaching a peak range of 0.06 mag in the range  $u=8.3$ –9.4 mag.

We emphasise that these magnitude-dependent median trends are most likely attributable to the SMSS background estimation and source confusion in regions of high density, and to the background subtraction in the *Gaia* spectroscopy in low-density regions. Sky maps of the magnitude differences show larger offsets in the middle of the Galactic Plane at otherwise reliable magnitudes, and large offsets for the whole sky when considering bins fainter than  $\sim 17$  mag. We anticipate that modelling of the differences between SMSS DR4 and the *Gaia* synthetic photometry will enable a robust correction of the hockey stick effect and the creation of standardised SkyMapper bandpasses within GAIAXPY.

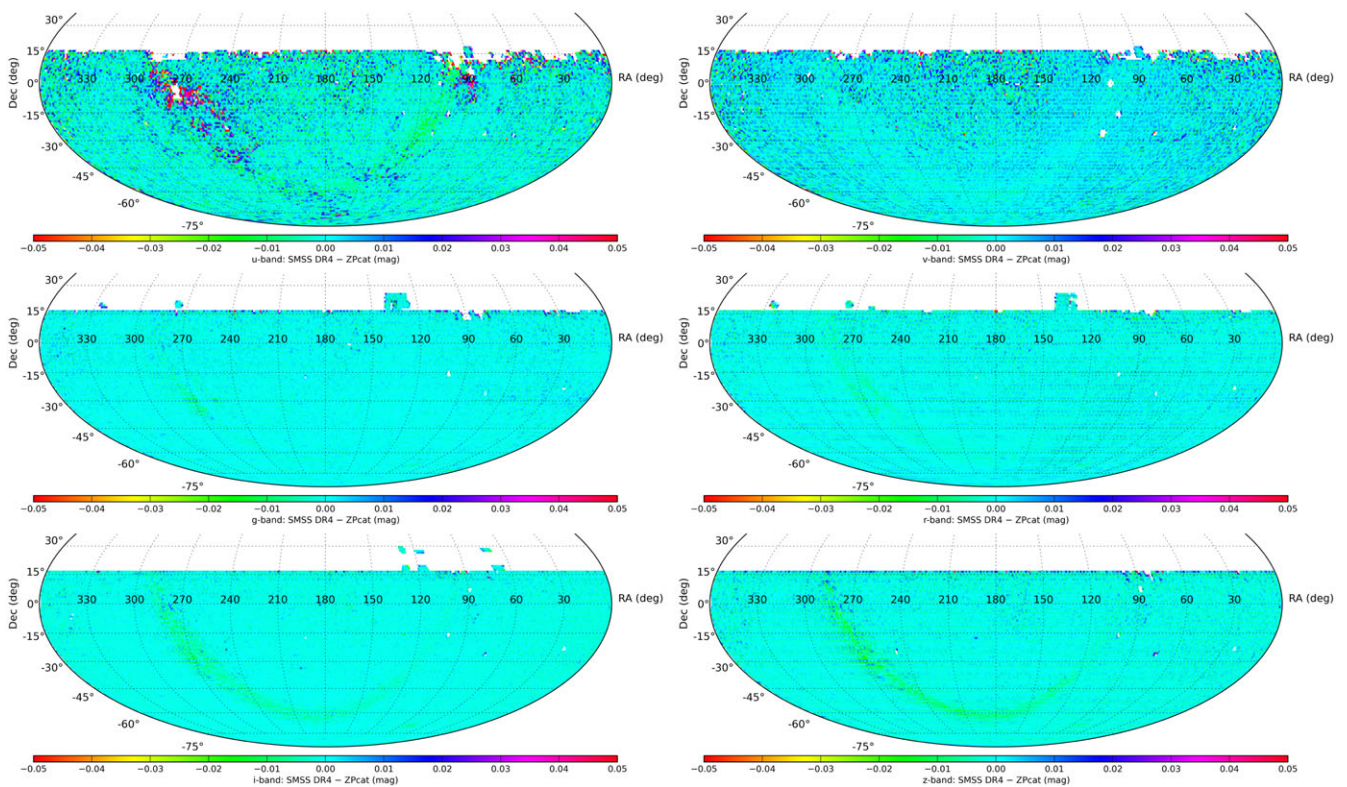
### 6.5.1 PS1 comparison

We next compare the DR4 photometry with that of Pan-STARRS (Panoramic Survey Telescope & Rapid Response System, hereafter PS1; Chambers et al. 2016). Although the region of overlap only covers the Declination range of  $-30$  to  $+15$  deg, and the comparable filters are limited to *griz*, PS1 is exceptional in its knowledge of its bandpasses and its photometric homogeneity across the  $3\pi$  steradians of sky coverage.

We apply the magnitude transformations of Tonry et al. (2018) and find that they continue to provide reliable conversions between the SMSS and PS1 photometric systems. For well measured SMSS stars (unsaturated sources between 14 and 18 mag without bright neighbours) with PS1 counterparts within 1 arcsec, we compute the median magnitude difference per square degree. Across the 15 000 deg<sup>2</sup> of overlap, the median per-square-degree differences in *griz* are  $-2$ ,  $-14$ ,  $-12$ , and  $-13$  mmag



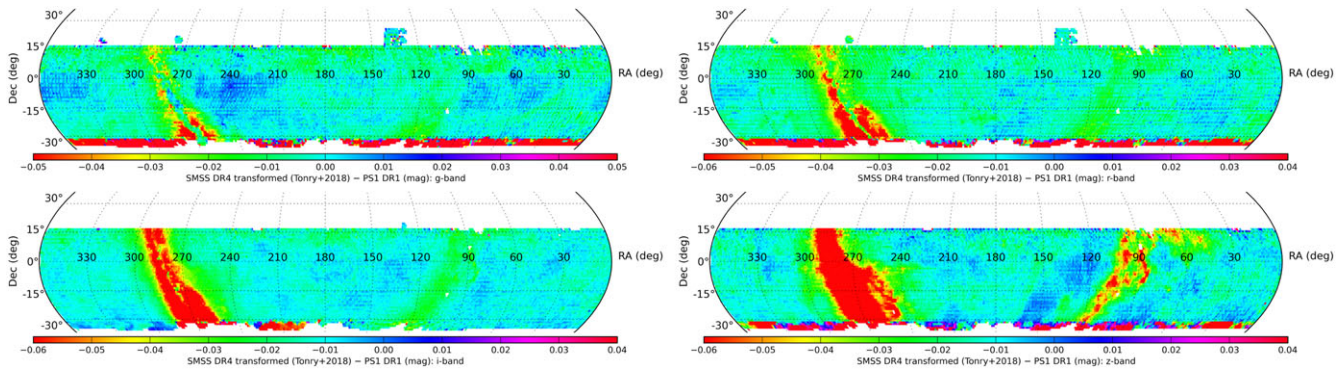
**Figure 16.** Cumulative distribution of astrometric offsets between SMSS DR4 and *Gaia* DR3 as a function of (left) SMSS *i*-band magnitude in a fixed range of  $|b|$ , and (right) Galactic latitude in a fixed range of *i*-band magnitude. The green line in each panel corresponds to the slice examined in the alternative panel. For these distributions, the SMSS CLASS\_STAR was required to be >0.7 to avoid contamination by extended sources and the offset distance was limited to 1 arcsec to avoid spurious matches.



**Figure 17.** Median photometric differences (DR4-observed minus ZP-catalogue-predicted) per square degree for each of the 6 SMSS filters (*uvgriz*). In contrast to Figure 8, the colourbars now cover a uniform range of  $\pm 0.05$  mag.

(SMSS DR4 – PS1 DR1), and the scaled median absolute deviation (SMAD) values are 6, 6, 4, and 7 mmag, respectively. The sky maps of the median difference are shown in Fig. 18.

As an indication of how our *Gaia*-based ZP catalogue might be expected to perform with an idealised SkyMapper observation, we next compare PS1 to the



**Figure 18.** Median magnitude differences (DR4-transformed minus PS1) per  $\text{deg}^2$  for *griz*, using the photometric transformations of Tonry et al. (2018).

synthetic *Gaia* DR3 photometry for stars selected from our ZP catalogue.<sup>w</sup>

We selected a large region of low-reddening sky, bounded by  $\text{RA}=0\text{--}15^\circ$  and  $\text{Dec}=-28\text{--}+15^\circ$ , yielding  $\sim 200\,000$  sources matched between the two samples (using a 0.5 arcsec threshold). We find a slope in the median *Gaia*–PS1 magnitude difference of  $3\text{--}4\text{ mmag mag}^{-1}$  for the magnitude range 14–16 mag (bounded by PS1 saturation on the bright end and the ZP catalogue requirement for at least one SMSS band to be brighter than 16 mag), with approximate offsets at 16 mag of  $+1.5, -1.5, 0, +2\text{ mmag}$  in *griz*, respectively. These trends appear to be somewhat steeper than those found by Gaia Collaboration et al. (2023b) from two  $15 \times 15^\circ$  patches at high Galactic latitude. The sample described here contains 20–30 times more sources in the 14–16 mag range, so a wider examination in the future could be valuable for refining the GAIAXPY standardisation process for PS1.

### 6.5.2 ATLAS Refcat2 comparison

SMSS DR1.1 provided one of the inputs to the ATLAS All-Sky Stellar Reference Catalog (Refcat2, hereafter; Tonry et al. 2018), a multi-survey synthesis of optical point sources brighter than 19 mag. The Refcat2 photometry was transformed onto the PS1 system, as that dataset formed the anchor over the vast majority of the sky.

As with PS1 above, applying the transformations of Tonry et al. (2018) to the SMSS DR4 photometry produces good agreement with Refcat2. North of  $\text{Dec}=-28^\circ$ , where PS1 underpins the photometric calibration, the statistics closely match the values quoted for PS1 above. South of the PS1 coverage, where a previous generation of SMSS calibration provided much of the key photometry, we find more significant differences. The southern median *griz* differences per square degree are  $-7, -18, -25,$  and  $-26\text{ mmag}$  (SMSS DR4 – PS1 DR1), with SMAD values of 6, 9, 6, and 11 mmag.

We expect that the change in behaviour south of  $\text{Dec}=-30^\circ$  reflects an improvement in the SMSS calibration arising from access to the spectrally resolved data in *Gaia* DR3, which was unavailable when Refcat2 was constructed.

<sup>w</sup>The photometry was generated natively in the PS1 bandpasses via GAIAXPY v2.1.1, with the photometric standardisation applied by using the *PanSTARRS1\_Std* bandpass set (see Gaia Collaboration, et al. 2023b). Results without the standardised bandpasses were broadly consistent with those described below, although we were obliged to fit and remove colour terms ourselves.

### 6.5.3 NSC comparison

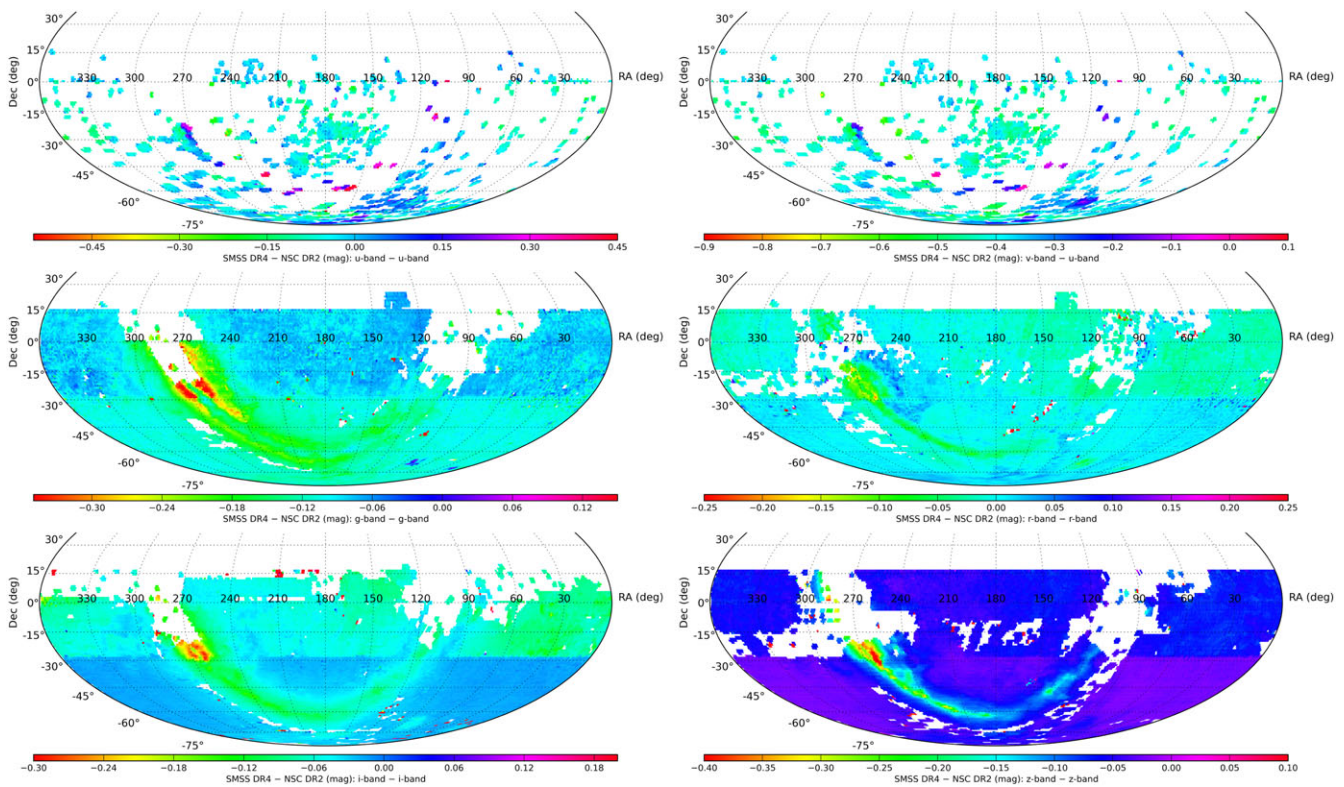
In terms of hemispheric coverage in the south, one of the largest and deepest optical datasets available for comparison to SMSS DR4 is the NOIRLab Source Catalog (NSC) DR2 (Nidever et al. 2021). NSC has compiled a broad set of images, which, for the southern hemisphere, primarily come from the NOIRLab Blanco 4m telescope, and processed them in a homogeneous way. We show the sky maps of the median magnitude differences per square degree in Fig. 19. The maps are driven by two factors:

1. Primarily, they illustrate bandpass differences between SkyMapper and DECam. The SkyMapper *u* and *v* filters have no equivalent bandpass in the NSC dataset and are both compared with the NSC *u*-band. The SkyMapper *griz* filters have eponymous NSC bands with different transmission curves. The AB colours for an unreddened M0V star, e.g. are  $(g-r, r-i, i-z) = (0.83, 0.76, 0.31)$  for SkyMapper passbands, while they are  $(g-r, r-i, i-z) = (1.28, 0.62, 0.31)$  for the DECam passbands in NSC.
2. A secondary factor is actual calibration differences between the two data sets. The sparse NSC DR2 data in *u*-band limits the visual comparison between the photometry of the two surveys in Fig. 19. However, because SMSS DR1.1 provided the *u*-band photometric calibration of NSC, one can anticipate significant spatial residuals that reflect the improvement of the SMSS photometric zeropoints. For the *griz* sky maps, we see a clear change in behaviour at  $\text{Dec} \sim -30^\circ$ , which marks the southern limit of Pan-STARRS DR1's coverage and the transition to where the primary NSC calibration source, ATLAS Refcat2 (Tonry et al. 2018), more heavily relies on SkyMapper DR1.1 (in conjunction with other surveys).

Continued efforts to explore the optimal photometric calibration of optical surveys will be extremely valuable.

## 6.6 Revisiting the passband transmission curves

After the production of the full DR4 data products, we compared the measurements of CALSPEC stars against the prediction from synthetic photometry and noticed imperfections hinting at inaccurate bandpass definitions. Already before the start of the survey,



**Figure 19.** Median photometric differences (DR4-observed minus NSC DR2) per square degree for each of the 6 SMSS filters (*uvgriz*) compared to the *ugriz* of NSC DR2. The SMSS *u*- and *v*-band maps are both made in comparison to the NSC *u*-band, and have larger colourbar ranges than *griz*.

Bessell et al. (2011) argued that empirical data will eventually facilitate improved bandpass definitions.

For CALSPEC stars with AB colours in the range of  $g - i = [0, 1]$ , the offsets between measured and predicted magnitudes are mostly less than 0.02 mag (see Fig. 20). The strongest offsets are seen in hot blue stars of  $g - i < -0.4$ , which can reach 0.05 mag. The current evidence suggests that these colour terms could be explained by shifting the mean wavelength of the passbands by between 1 and 5 nm.

A full explanation of these effects is beyond the scope of this paper, however, we briefly discuss the empirical evidence. First, existing records revealed an ambiguity around the mean CCD quantum efficiency curve. Using the alternative CCD curve may nearly remove the colour trends in the *u*, *g* and *z* bandpasses, which are in areas where the CCD efficiency is not flat across the width of the filter. Shifts in the *r* and *i* filter might be explained by the differences between convergent beams in the telescope and a parallel beam in the 2010 laboratory measurements that produced the existing bandpass definitions. A 0.05 mag offset seen in the *v* band for hot stars might be explained by a bandpass shift, although that is not expected in a glass filter. It may also reflect systematic uncertainties in the *Gaia* spectral reconstructions at the wavelengths of *v*-band.

In future work, we will revisit these questions by exploiting more evidence from DR4 data, especially considering many more objects with well-known spectra. There may also be an opportunity to repeat laboratory measurements of the filters to determine whether their transmission has changed due to ageing.

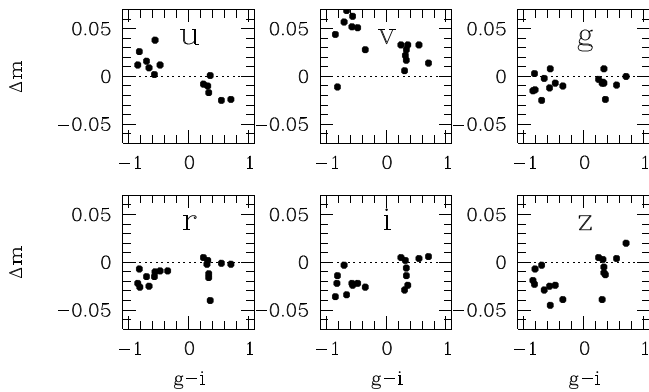
## 6.7 Notes on source types

In this section, we describe some important features of the DR4 dataset, broken down by source type.

### 6.7.1 Moving sources

Sources with significant sky motion, principally Solar System objects, will show up with a range of properties depending on their apparent speed:

1. Slow-moving objects such as stars with moderate proper motion will show large position uncertainties associated with a single OBJECT\_ID that captures all detections of the object over the years of the survey.
2. Stars with high proper motions and very distant Solar System objects will see their detections broken up into more than one OBJECT\_ID. The example of Pluto was presented in Sect. 3.7 of Onken et al. (2019).
3. Distant Solar System objects are bound to show a separate OBJECT\_ID for each night of their observation, and may also show large position uncertainties when a sequence of images was observed.
4. Typical Main-Belt asteroids may appear broken up into several OBJECT\_ID entries already during a Shallow Survey (4 min duration) or Main Survey (20 min duration) colour sequence.



**Figure 20.** Magnitude differences, DR4 (measured) minus predicted, of CALSPEC spectrophotometric standard stars. Two of the hot stars and two of the cool ones do not appear in the  $u$ -band panel, where they have  $\Delta m \approx -0.1$  for reasons that remain under investigation.

5. Near-Earth asteroids and artificial satellites will produce streaks covering one or more CCDs (see, e.g. Fig. 4), which could each seed their own OBJECT\_ID (or multiple, if split into separate SOURCE\_EXTRACTOR detections).

In any case, there is a risk of such objects being associated with the OBJECT\_ID of a persistent source from beyond the Solar System and being blended with it; one such case is discussed in detail in Section 3.7 of Onken et al. (2019). In this case, also the  $\{f\}$ \_MMVAR column for the filter  $\{f\}$  observed during and affected by the transient blending may show an inflated value above a true variability level of the persistent source, although the rare outlier measurements may be clipped from the distilled magnitude estimates. Finally, the fastest and faintest moving objects may pass so quickly that the flux is too diffuse to appear in the photometry table at all.

The main challenge for selecting moving or transient sources from the master table is to differentiate between genuine objects and spurious detections. In persistent astrophysical sources, the multiple detections indicated by NGOOD values above 1 tend to weed out spurious imposters. But among single detections, a variety of indicators need to be consulted for a pure sample of real objects: good FLAGS ( $<4$ ) and sensible MEAN\_FWHM values should be required.

### 6.7.2 Small-separation sources

Some objects, such as those in crowded areas or binary/triple stars or merging galaxies, require special attention: it is possible that two objects are physically present, but three unique objects appear in the master table. This happens when, in some images, the two are separately detected, and in others, a blended object is detected at an intermediate position. The two sets would appear in mutually exclusive image sets. In such cases, the spatially resolved photometry remains useful for the two genuine sources, and the blended detection should be ignored. Recognising which is which requires consulting the images or the photometry table with an eye on positions and FWHM values of the detections.

Similarly, at intermediate separations, no third object may appear with its own OBJECT\_ID, but the blended version may be associated with one of two true sources, bringing brighter photometry into the detection list; the  $\{f\}$ \_MMVAR column for each

filter  $\{f\}$  may show an inflated value far above the true variability; to what extent measurements might be deemed outliers and thus be clipped from the distilled magnitude estimates depends on the individual measurement statistics.

Finally, tight binary objects will usually not be separated, and thus they will have blended photometry in most or all images, which then dominates the available photometry and the clipping process retains the blended measurement.

### 6.7.3 Variable sources

Variable objects are easily found in the master table by looking for objects with a  $\{f\}$ \_MMVAR value that is atypically high for the object's magnitude, which points to excess variability above the expected noise variation. Chance projections with asteroids, however, can enhance the brightness of otherwise non-variable objects in a small number of detections and thus make them appear variable.

As discussed in Section 2.3, the SMSS has not followed a strict cadence pattern, but collected the observations over time to accrue the desired depth. One specific cadence that was retained, however, is that of the Main Survey Colour Sequence, which completed a  $uvgruvizuv$  pattern on one field within a span of 20 min. Such data have proven particularly useful for measuring short-term variability in the  $uv$  bands (see e.g. Li et al. 2022).

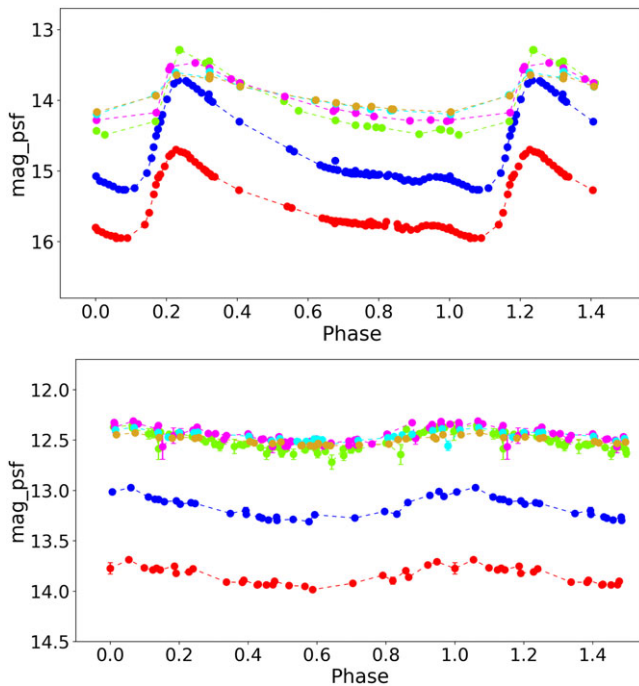
Around 90% of the moderately bright objects in DR4, with  $g_{\text{PSF}} \sim 17$ , have less than ten observations per band with good flags. While the total number of visits by the telescope has been larger, that includes observations with bad quality flags and images not even included in DR4. The median number of good detections in the  $g$  and  $r$  filters is  $\{f\}$ \_NGOOD = 6. As DR4 includes all data taken with the SkyMapper Telescope, irrespective of the original purpose, some areas in the sky have many more visits, including the area of the SkyMapper Transient Survey (Scalzo et al. 2017), the Standard Fields, and smaller non-Survey programs. As a result, about 10% of the objects have 10 or more good detections, 1% have 25 or more, 0.3% have over 100 detections, and finally 0.1% of the objects have over 500–600  $g$ - and  $r$ -band detections. Therefore, some specific sky areas have a strong time-domain coverage, while most of the hemisphere has few visits. In Fig. 21 we show the phase-folded light curves of two known variable stars.

### 6.7.4 Extended and point-like sources

Marginally extended sources can be distinguished from point-like sources using one of three indicators (see Fig. 22):

1. The CLASS\_STAR parameter measured by SOURCE\_EXTRACTOR on individual images is recorded in the photometry table, and the master table records the largest value detected for any distilled object. Down to a magnitude of  $r \approx 19$ , nearly all point sources show CLASS\_STAR values  $> 0.95$ .
2. The CHI2\_PSF parameter is determined from light profiles in 15 arcsec apertures; the values of individual detections in the photometry table get averaged per filter and then the largest value among the filters is listed in the master table (and hence the average value in the master table will be larger than 1 even if the mean value was  $\sim 1$  among the detections). For point sources, typical values of CHI2\_PSF range from 1 to 3, but they show little correlation to the





**Figure 21.** SMSS DR4 phase-folded lightcurves of two known variable stars: (top) the fundamental-mode RR Lyr (RRab) star, SX Gru (OBJECT\_ID 489799903), with a period of 0.5934 d; (bottom) the  $\delta$  Scuti pulsating variable, ATO J136.5509-00.8024 (OBJECT\_ID 89456216), with a period of 0.1206 d. In each panel, the points are colour-coded by filter:  $u$ =red,  $v$ =blue,  $g$ =green,  $r$ =magenta,  $i$ =cyan,  $z$ =gold. The phase coverage is more complete for these two stars than the average DR4 source (even the redder four filters in the top panel have 2–3 $\times$  more measurements than typical), but a small fraction of sources will have even more detections in DR4 than the  $\sim 80$  for  $u$ - and  $v$ -band in the top panel.

CLASS\_STAR parameter. Towards faint point sources, the CLASS\_STAR values decrease, while the CHI2\_PSF values remain at the same level just with increased scatter. Some galaxies with bright nuclei show high, point-like CLASS\_STAR values although their CHI2\_PSF values are large, indicating a significant envelope of light.

3. The difference,  $\{f\}_{\text{APC05}} - \{f\}_{\text{PETRO}}$ , between the PSF-corrected  $5''$ -aperture magnitude and the Petrosian magnitude in filter  $\{f\}$  can be calculated for individual detections in the photometry table but will have the least noisy value from distilled values in the master table. It is well-correlated with the CHI2\_PSF parameter.

Of course, this is no reliable indication for the physical nature of an object: AGN and QSOs may show point-like appearance especially when the nucleus is significantly brighter than the host, and stars could appear extended when they are in projected or real binary systems that are only marginally resolved by SkyMapper.

### 6.7.5 Detection of extended sources

Spatially extended sources can prove challenging in catalogue-building for a number of reasons: variations in seeing and/or image depth can alternatively split and merge emission features; the centroid may shift with the wavelength of the bandpass; in the case of marginal detections, the centroid position may not be well constrained.

In Fig. 23, we show three examples of galaxies from each of two magnitude bins ( $r_F = 10$  and 12 mag) from the 6dF Galaxy Survey (6dFGS; Jones et al. 2009), overlaid with the nearby DR4 master table entries. The red circles, indicating sources not matched to previous SMSS data releases, are often dubious, with  $\text{NGOOD}=0$  (as is the case for all such sources in the figure, aside from the top-middle panel discussed below). The cyan circles indicate sources retained from previous SMSS datasets.

The top-middle panel of Fig. 23 shows the case of 6dFGS g2245317-392031, a bright edge-on spiral galaxy at  $z=0.008$ . The red square indicates the DR4 source that best matched the reported galaxy position from 6dFGS, which differs from the OBJECT\_ID used in SMSS DR3. The offset between those two DR4 sources is 2.1 arcsec, just larger than the 1.5-arcsec linking distance described in Section 5.3. Amongst the three objects at the centre of the galaxy, only one is present in any given image. And while all three objects have  $vgr$  image associations, one object accounts for all  $u$ -band data, while the other two account for all  $i$ - and  $z$ -band measurements. Fortunately, such cross-match confusion is rare, with only 1% of 6dFGS galaxies having a different OBJECT\_ID than in DR3.

We also find that over-deblending of extended sources is not a widespread issue. Amongst nearly 1 million objects from the Two Micron All Sky Survey Extended Source Catalog (2MASS XSC; Jarrett et al. 2000), 94% of those with DR4 counterparts have a single DR4 object within  $k_{r\_eff}$  (the  $K$ -band half-light radius). For 6dFGS, only 8% of galaxies have more than one good DR4 source (i.e. with  $\text{NGOOD}>0$ ) within 5 arcsec.

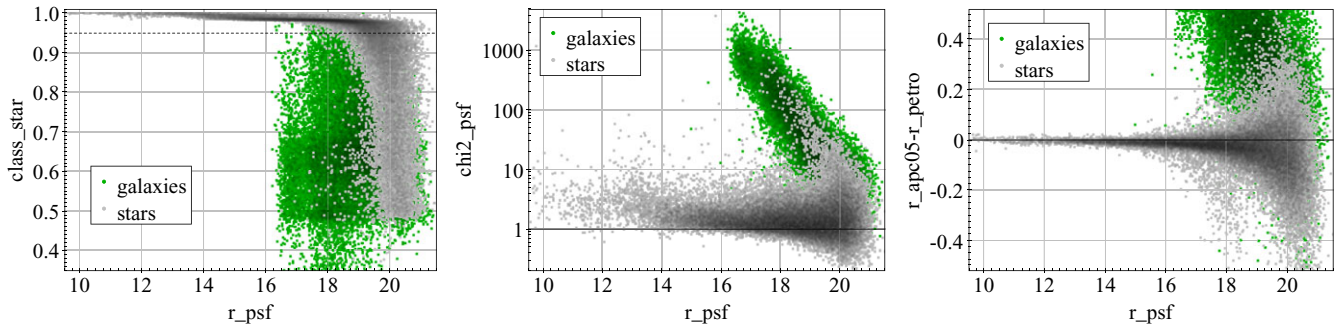
### 6.7.6 Photometry of extended sources

We explore the extended source photometry by comparison against the Petrosian magnitudes from DES DR2 Abbott et al. (2021), which shares the same underlying measurement algorithm in SOURCE\_EXTRACTOR. As with the point-source comparison in Section 6.3.1, we utilise  $r$ -band to minimise colour terms. From the Astro Data Lab (Fitzpatrick et al. 2014; Nikutta et al. 2020), we retrieved the DES Petrosian magnitudes and radii for high-confidence galaxies (EXTENDED\_CLASS\_COADD=3) brighter than  $r = 19$  mag and having  $r$ -band flags (FLAGS and NIMAF\_LAGS\_ISO) equal to zero, and then cross-matched those 1 140 431 sources against SMSS DR4, finding 1 134 200 counterparts.

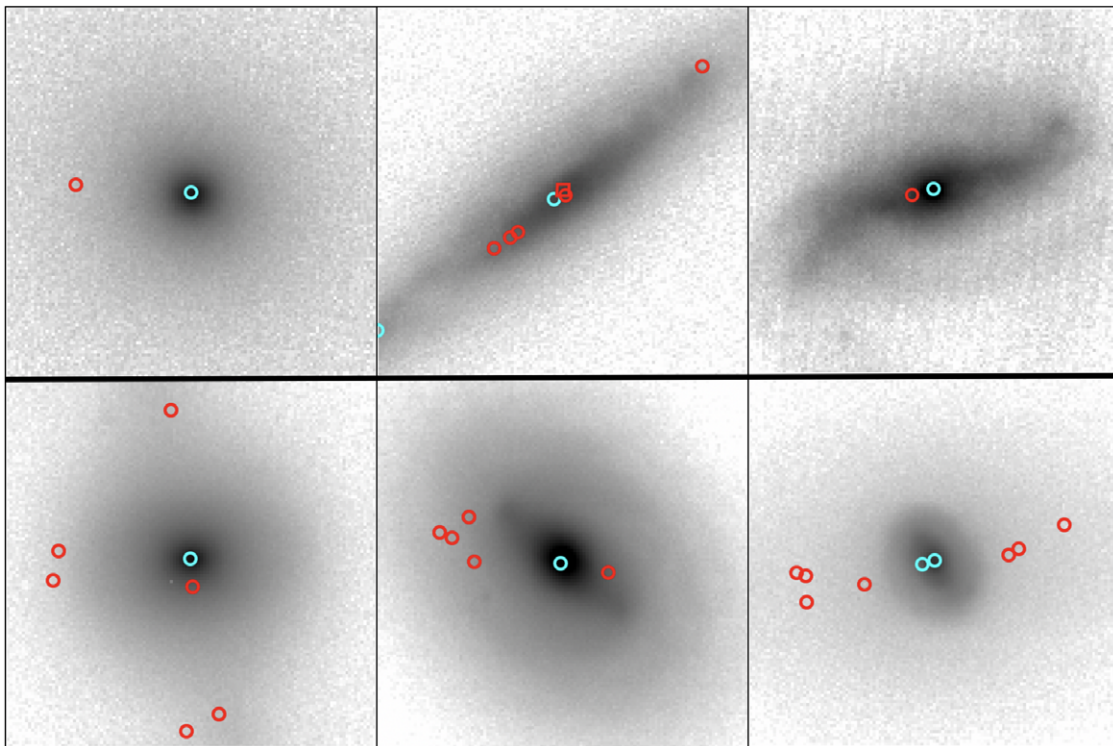
We further sub-select those objects having a single good  $r$ -band visit, to explore the limiting case of the photometric performance from single images without the benefit of multiple measurements to reduce statistical noise. This subsample of 109 744 objects is then corrected for a colour term, adding  $(0.065 + 0.1 \times (g - r))$  to the DES magnitudes, based on the median Petrosian magnitude difference as a function of the DES galaxy colour.

We find that the median Petrosian magnitude difference is consistent with zero within  $0.3\sigma$  for all DES Petrosian magnitudes between 17 and 19 (i.e. excluding the brightest 1% of the sample). However, the distribution relative to the combined photometric uncertainties is wider than the expected normal distribution. Increasing the SMSS Petrosian magnitude errors by a factor of 1.8 restores the expected Gaussian width, while further reducing the small bias in the median magnitude difference.

Compared to DES DR2, we find a residual trend in the Petrosian radius (as recorded in the photometry table, owing to the issue discussed in footnote s), with the SMSS estimates



**Figure 22.** Star-galaxy separation as probed by three indicators: the SOURCE EXTRACTOR CLASS\_STAR (left) is  $> 0.95$  for point sources; the CHI2\_PSF (centre) ranges mostly between 1 and 3 for point sources; and the difference between the APC05 magnitude and the Petrosian magnitude (right) is close to zero for point sources. Stars are a subset of *Gaia* sources with parallax and proper motion signals of at least  $5\sigma$  significance and no neighbours seen within 15 arcsec radius, and compares them to galaxies from the 2dFLenS Survey (Blake *et al.* 2016). The double fan seen among stars in the top right panel results from objects with only Shallow Survey imagery vs the majority that has also Main Survey data.



**Figure 23.** SMSS DR4 *i*-band cutouts (1 arcmin square) of example 6dFGS sources having  $r_f$  magnitudes of (top) 10 and (bottom) 12, with DR4 sources from the master table overlaid as circles. Cyan circles indicate DR4 sources matched to previous SMSS data releases, while red circles are objects newly identified in DR4 (OBJECT\_ID values  $> 2E9$ ). For the top-middle panel, the source which is recorded as the DR4 match within the ext. *spec\_6dFGS* database table (see Section 7.2.1) is indicated as a red square. Except for the central two sources in that panel, all other sources indicated in red have NG00D=0. Only 1% of 6dFGS galaxies well-matched to DR4 sources have a new OBJECT\_ID assigned in DR4 compared to DR3. The 6dFGS identifiers are (left-to-right, top-to-bottom): g2203270-345631, g2245317-392031, g2224218-334139, g0338521-262016, g1249022-083952, and g0342015-471319.

underpredicting the DES radii at small values. This is likely due to the poorer SkyMapper seeing and extended PSF halo. We find the DES Petrosian radii are reasonably well reproduced by applying a correction to the SMSS values of  $R_{\text{DES}} = R_{\text{SMSS}} - 30/R_{\text{SMSS}}$ , with all quantities in arcseconds, e.g.  $R_{\text{SMSS}} = \text{RADIUS\_PETRO} * \sqrt{A * B} * 0.497$ .

Returning to the larger sample, we find that the repeat measurements in SMSS DR4 help to improve the reliability of the Petrosian

magnitude errors. The overall scaling required to produce a zero-centred Gaussian is a factor of 1.4 increase to the reported SMSS Petrosian magnitude errors. Furthermore, when the Petrosian magnitude errors are smaller than 0.015 mag (constituting  $\sim 1\%$  of the galaxy sample considered here), the Petrosian magnitudes exhibit a bias to brighter values that exceeds the (inflated) uncertainties by  $1\sigma$  in the median. Thus, we suggest caution when the magnitude uncertainties fall below that level.

As noted in footnote s, the Petrosian radii recorded in the `master` table have not been scaled to seeing-insensitive units, and thus do not take proper advantage of the multiple measurements when computing the mean value.

### 6.7.7 Colours of extended sources

Integrated colours of extended objects depend on the choice of aperture due to potential colour gradients. However, the light profiles of galaxies imply that the photometry of outer regions is noisy. Hence, colours show lower scatter when calculated from the PSF-corrected  $5''$ -aperture magnitude rather than the Petrosian magnitude; the aperture colours are also corrected for PSF variation between and within images. For example, the bright red-sequence galaxies at  $z = 0.10 \pm 0.01$  in 2dFLenS (with  $R_{\text{PETRO}} = [15.9, 16.7]$ ) show observed-frame APC05 colours of  $\nu - r = 2.41 \pm 0.115$  (mean and RMS scatter) but Petrosian colours of  $\nu - r = 2.89 \pm 0.315$ . Propagating formal magnitude errors, which are dominated by the  $\nu$ -band errors, suggests typical errors for the APC05 colour of 0.086 mag and for the Petrosian colour of 0.13 mag. After square-subtracting these from the colour scatter, we derive an intrinsic colour scatter (not corrected for the slope of the colour-magnitude relation) of 0.075 mag from APC05 colours and 0.29 mag from Petrosian colours, i.e. four times larger. Although the two results come from two different-sized physical footprints on the galaxies, and the red sequence is contaminated with red spirals with modest colour gradients, the inflated scatter in the Petrosian  $\nu - r$  colour is likely due to mostly larger noise that is not captured in the formal errors of the Petrosian magnitude.

### 6.7.8 Large extended sources

One known deficiency in the current bias-PCA algorithm (Section 4.3) is inadequate masking around the edges of large extended objects. This particularly affects the  $u$ - and  $\nu$ -band images, which are naturally faint in the outskirts of galaxies. The consequence is that the extended emission is treated as part of the bias level, and so is fit and subtracted by the PCA. Furthermore, because the PCs tend to be slowly varying functions with pixel position, the enhanced count level in the outer regions can force the PC fit higher than appropriate, resulting in dark wings on either side of the galaxy.

## 6.8 Other known issues

- Around 50 million sources have  $\text{NGOOD}=0$  in the `master` table, which are a combination of spurious and unreliable sources. 95% of them have  $\text{FLAGS} \geq 4$ . The remainder have  $\text{NIMAFLAGS} \geq 5$ , which excluded them from providing magnitude data to the distilling process. The spurious sources are most often ( $\sim 50\%$ ) single spurious detections close to CCD borders and not real objects (with bit value 16 set in `FLAGS`). Nearly as common, comprising over 40% of the sources, is the case of real objects in the vicinity of a nearby bright star, and so having bit value 1 024 set in `FLAGS` because their photometry could be affected by scattered light.
- Slightly over 6 million  $\text{MEAN\_FWHM}$  values are less than or equal to zero. In  $\sim 95\%$  of the cases these are objects with single spurious detections: in  $\sim 70\%$  of cases, they reside

in the vicinity of bright stars (bit value 1 024 in `FLAGS`), and in  $\sim 25\%$  of cases they are detected at the edges of a CCD (bit value 16 in `FLAGS`). Even when they have good flags, they have at most one detection and usually no counterparts in the *Gaia* DR3 catalogue; they are often image artefacts that have not been flagged or masked.

- Some detections in the photometry table show identical PSF magnitudes for all apertures, and thus  $\text{CHI2\_PSF}$  values equal to 0. This phenomenon occurs on CCDs where five or fewer stars were available to determine the PSF map; in these cases, the adopted PSF aperture corrections is chosen to be the median of the available stars, and thus, unintentionally, when the number of PSF stars is odd, one of them ends up with a seemingly perfect PSF shape by construction. In most cases, this problem is irrelevant, as multiple detections will exist with proper  $\text{CHI2\_PSF}$  values, the largest of which gets distilled into the entry for the `master` table. However, the `master` table does contain 48 objects with good flags and a  $\text{CHI2\_PSF}$  value of 0; all of these objects have only a single detection in one single filter. Some of these objects appear genuine in the images, but as they are not associated with a source known from another survey such as *Gaia* DR3, they may be asteroids. However, the majority appear to be unflagged cosmic rays and image artefacts near the central amplifier boundaries of the CCDs.
- Extended sources in shallow images may have poorly determined central coordinates. In certain circumstances, this can result in multiple `OBJECT_ID` values being created for the same astrophysical object, which were unintentionally left unmerged during the calculation of the mean object properties (Section 5). Among the consequences of this proliferation of sources in the `master` table, we find 1 592 sources that have  $\text{SELF\_DIST1} < 0.1$  arcsec, four that have  $\text{SELF\_DIST2} < 0.2$  arcsec, and one that has  $\text{SELF\_DIST3} < 0.3$  arcsec. More than half of these cases are clustered around low-redshift galaxies in the 6dF Galaxy Survey (Jones et al. 2009), especially those that fall within the Standard fields (Section 3.1 and Table 2), were visited thousands of times with short exposures. As indicated in Section 5.2, most of these cases arose from the Standard Field images, which were subsequently assigned a `FLAGS` value of 16 384.
- When aperture corrections for a given CCD did not yield valid results for a given aperture size, then certain corrected aperture magnitudes will be missing in the photometry table, although the uncorrected aperture fluxes remain. As a result, no determination of the  $\text{MAG\_PSF}$  could be made for the detections in that CCD. This implies that the detection is not considered for distilling and  $\text{USE\_IN\_CLIPPED} = -1$ . However, the other magnitude columns – the larger-aperture (15, 20, and 30 arcsec) magnitudes that do not receive aperture corrections, and those of Petrosian (1976) and Kron (1980) – will exist in the photometry table.

## 7. SMSS DR4 data access and format

In this section, we describe how to access the DR4 data, and the format of the available data.

## 7.1 Data access

The SMSS DR4 dataset is available through both the SkyMapper website (<https://skymapper.anu.edu.au/>) and the SkyMapper node of the All-Sky Virtual Observatory (ASVO). The website provides documentation for DR4 and previous releases; interfaces for cone-search, image cutout, and catalogue queries; summary pages for each SMSS DR4 astrophysical object; and a User Forum to obtain additional information from the SMSS user community or the SkyMapper Team.

The SkyMapper ASVO node provides API access to the cone-search, image cutout, and catalogue query (Astrophysical Data Query Language [ADQL]) interfaces. Further details are available here: <https://skymapper.anu.edu.au/how-to-access/>.

Additional data access pathways for DR4 are being established through the Astro Data Lab (<https://datalab.noirlab.edu>) at NOIRLab's Community Science and Data Center (CSDC), the VizieR catalogue service (<https://vizier.cds.unistra.fr/viz-bin/VizieR>) at the Centre de Données astronomiques de Strasbourg (CDS), the SciServer science platform (<https://sciserver.org/>) operated by the Institute for Data Intensive Engineering and Science (IDIES) at the Johns Hopkins University (JHU), and the Barbara A. Mikulski Archive for Space Telescopes (MAST, <https://archive.stsci.edu>) hosted by the Space Telescope Science Institute (STScI).

## 7.2 DR4 data format

We describe here the format of the information released in DR4, both the table data and the image data.

### 7.2.1 Table data

The DR4 catalogue consist of five distinct tables:

- `dr4.master` – main catalogue, listing the distinct astrophysical sources with their mean properties;
- `dr4.photometry` – per-image detection properties, which can be joined to the `master` table with the `OBJECT_ID` column;
- `dr4.images` – image-level properties, which can be joined to the `photometry` table with the `IMAGE_ID` column;
- `dr4.ccds` – CCD-level properties, which can be joined to the `photometry` table with the (`IMAGE_ID`, `CCD`) columns;
- `dr4.mosaic` – mapping between a detection's CCD and ( $x_{\text{CCD}}$ ,  $y_{\text{CCD}}$ ) position to the ( $x_{\text{mosaic}}$ ,  $y_{\text{mosaic}}$ ) position.

A description of the `master` table columns is provided in the Appendix. The `master` table contains a number of cross-match identifiers and spatial offset distances (in arcseconds) for other major catalogues. These external catalogues are stored in the `ext` schema, and can be utilised to extract the other data by matching on the unique identifier relevant for each table. Matches are provided with offset distances up to 15 arcsec. The cross-matched tables include:

- `ext.allwise` – AllWISE Source Catalog, Cutri et al. (2013)
- `ext.catwise2020*` – CatWISE2020 Catalog, Marocco et al. (2021)

- `ext.des_dr2*` – Dark Energy Survey (DES) DR2 Catalog Abbott et al. (2021)
- `ext.refcat2` – ATLAS All-Sky Stellar Reference Catalog, Tonry et al. (2018)
- `ext.gaia_dr3` – *Gaia* DR3 Source Catalogue, Gaia Collaboration et al. (2023a)
- `ext.galex_guvcat_ais` – *GALEX* Catalog of UV sources from the All-sky Imaging Survey (GUVcat\_AIS\_fov055), Bianchi, Shiao, & Thilker (2017)
- `ext.ls_dr9*` – DESI Legacy Imaging Surveys (LS) DR9 Sweep Catalogs, Dey et al. (2019)
- `ext.nsc_dr2*` – NOIRLab Source Catalog (NSC) DR2, Nidever et al. (2021)
- `ext.ps1_dr1*` – Pan-STARRS1 (PS1) DR1 Catalog, Chambers et al. (2016) (Note: only selected columns, restricted to `decMean<30deg` and `nDetections>1`)
- `ext.splus_dr3*` – Southern Photometric Local Universe Survey (S-PLUS) DR3 Catalog, Mendes de Oliveira et al. (2019)
- `ext.twomass_psc` – Two Micron All Sky Survey (2MASS) Point Source Catalog (PSC), Skrutskie et al. (2006)
- `ext.vhs_dr6` – VISTA Hemisphere Survey (VHS) DR6 Source Table, McMahon et al. (2013) (Note: VISTA Science Archive copy of DR6, includes data through UT 2017-04-01)

where the asterisk indicates tables that have been newly added or expanded for SMSS DR4.

In addition, we store SMSS DR4 cross-match information in a number of smaller (primarily spectroscopic) tables in the `DR4_ID` and `DR4_DIST` columns, which record the `OBJECT_ID` from the `master` table and the offset distance in arcseconds, respectively. Since these object lists are orders of magnitudes smaller than the `DR4 master` table, it is more efficient to find and store suitable DR4 matches in those tables. However, these 'reverse matches' are affected by the issue that some unique astrophysical objects appear as two separate and non-identical entries in the `master` table, typically with one of the entries distilling most detections and the other one distilling only few. The best match is thus determined not only based on proximity but also on the volume of good-quality information. Hence, we allow reverse-matching only to sources with `FLAGS < 16384` and choose the match as the `DR4 master` source that minimises the value of  $(0.5'' + d)/(1 + \text{NGOOD})$ , for distance  $d$  in arcsec.

As above, these cross-matches extend up to 15 arcsec. This set of tables includes:

- `ext.kids_dr4p1*` – Kilo-Degree Survey (KiDS) DR4.1, Kuijken et al. (2019)
- `ext.milliquas_v8*` – Million Quasars Catalog (Milliquas), version 8 (2023), Flesch (2023)
- `ext.spec_2dfgrs` – 2dF Galaxy Redshift Survey (2dFGRS) Final Data Release, Colless et al. (2003)
- `ext.spec_2df lens` – 2dF Gravitational Lens Survey (2dFLenS), Blake et al. (2016)
- `ext.spec_2qz6qz` – 2dF and 6dF QSO Redshift Surveys (2QZ/6QZ) Final Catalogue, Croom et al. (2004)

- `ext.spec_6dfgs` – 6dF Galaxy Survey (6dFGS) DR3, Jones et al. (2009)
- `ext.spec_anu2p3` – spectroscopic classifications from the ANU 2.3 m telescope for quasar candidates, extremely metal poor stars, and other programs (Onken et al. 2022, 2023; Da Costa et al. 2019)
- `ext.spec_galah_dr3*` – GALAH+ DR3, Buder et al. (2021)
- `ext.spec_gama_dr3` – Galaxy Mass and Assembly(GAMA) Survey DR3, Baldry et al. (2018)
- `ext.spec_hesqso` – Hamburg/ESO survey for bright QSOs, Wisotzki et al. (2000)
- `ext.spec_ozdes_dr2*` – Australian Dark Energy Survey (OzDES) DR2 Lidman et al. (2020)
- `ext.spec_rave_dr6` – Radial Velocity Experiment(RAVE) DR6, combining the `sparv`, `classification`, and `obsdata` tables, Steinmetz et al. (2020)
- `ext.spec_twomrs` – 2MASS Redshift Survey (2MRS), Huchra et al. (2012)
- `ext.viking_dr5` – VISTA Kilo-degree Infrared Galaxy (VIKING) Survey DR5 Source Table, Edge et al. (2013) (Note: VISTA Science Archive copy of DR5, includes data through UT 2018-02-15)
- `ext.vsx_20230626*` – AAVSO International Variable Star Index, version 2023-06-26, Watson, Henden, & Price (2006)

where the asterisk again indicates tables that have been newly added or expanded for SMSS DR4.

### 7.2.2 Image data

The pipeline-processed image data and associated pixel masks that underlie the photometry in the tables are also made available to users. The images and masks are available in FITS or PNG formats via the Image Cutout service of the ASVO. At the initial release of DR4, each individual CCD is presented independently, with no stitching together of the separate CCDs in each mosaic image or coadding of multiple exposures. Such processes can be undertaken by the user, tuned to the needs of their own scientific analysis.

The images provided by the Image Cutout service have been bias-subtracted, flat-fielded, and defringed (where appropriate), but the sky level has not been subtracted. In the header of each FITS image, the `ZPAPPROX` keyword reports the centre-of-mosaic ZP, but users should be aware that the 2D ZP gradient applied to photometry in the tables has not been applied to the image cutouts (neither to renormalise the ZP to the centre of the cutout nor to apply a gradient in the count levels). However, because of the limits placed on ZP gradients in order for images to be included in DR4, application of the simple relation,

$$mag = ZPAPPROX - 2.5 \log_{10}(\text{counts}), \quad (2)$$

is likely to provide a calibration better than 0.1 mag for photometry measured by users directly from the image data.

## 8. Future activities

The wealth of DR4 data will provide the opportunity to update the passband definitions from on-sky data, and thus tidy up the overall

calibration of the catalogue. We anticipate such an update as part of a refined DR4.1 release.

We intend to produce co-added images in large sky tiles, taking advantage of dithered pointings to cover the gaps in the CCD mosaic, bad pixels, and pixels masked by cross-talk. Different image depths allow us to ignore pixel values that are saturated or masked for other reasons, including pixel overflow. The result will be deeper images with wider dynamic range. These will be produced with native coadded PSFs but also after convolution to a common PSF FWHM, where the latter will also allow the creation of robust colour maps. This imagery will then be used to update the ALADIN LITE-based SkyViewer on the SkyMapper website.<sup>x</sup>

Future work may explore additional methods of correcting for the high-frequency noise (Section 4.1) and slower bias level fluctuations (Section 4.3), which would improve the photometry in and around extended sources. To address the WCS solution biases in the mosaic corners (see Section 5), DR4 images with high source density could be used to create a fundamental high-fidelity, high-order WCS mapping, which could then be used for improved corner astrometry while allowing just subtle adjustments based on the observed sources in each image.

## 9. Summary

We present the 4th data release of the SMSS. The data set is nearly complete relative to the original public Survey plans; for the first time, the release includes data taken with SkyMapper for other science programmes, such as the SkyMapper Transient Survey (SMT; Scalzo et al. 2017), and ~6 000 very short exposures on each of seven standard fields. Improvements in sky coverage and data processing should enhance its usefulness compared to previous releases.

In summary, we highlight the following properties:

1. Compared to the previous public data release (DR2; Onken et al. 2019), SMSS DR4 more than triples the number of images, adds three years of additional time baseline (now 2014–2021), and 2 000 deg<sup>2</sup> of additional sky coverage.
2. WCS solutions are now based on *Gaia* astrometry, and mosaic-wide solutions have improved the sky coverage near the Galactic Plane, especially in the short-wavelength *u* and *v* filters.
3. A new photometric zeropoint catalogue uses synthetic photometry derived from *Gaia* DR3 low-resolution spectra. The precision in the *u* and *v* filters is much improved over previous releases and naturally obviates the need for the Huang et al. (2021) corrections. The calibration quality now appears limited by our knowledge of the quantum efficiency of the CCDs and filter throughput curves; for the latter, we have characterised the total system throughput.
4. Calibration comparisons with other surveys are now also limited by knowledge of the exact colour terms that follow from the system efficiency curves.
5. Source completeness at the faint end is limited by object detection on individual images and ranges from ~18 mag (AB) in *u*- and *z*-band to nearly 20 mag in *g*- and *r*-band.

<sup>x</sup>The SkyViewer utilises ALADIN LITE version 2 (Boch & Fernique 2014).

However, photometric errors for non-variable objects are relatively small due to repeat measurements: the  $10\sigma$  depth ranges from 18.6 mag in  $u$  and  $z$  to 20.5 mag in  $g$  band.

6. Photometric errors are improved, especially for the PSF magnitudes, where they were previously underestimated.
7. OBJECT\_IDS from DR2/3 have been kept where appropriate, while new sources have been given new IDs starting with from a value of 2E9.
8. In total, the catalogues include measurements for over 15 billion detections from  $\sim 700$  million astrophysical sources over 26 000 deg<sup>2</sup> of sky; these were derived from over 417 000 on-sky images with exposures ranging from 0.1 to 600 s.
9. In some specific sky areas, hundreds of epochs are available in the  $g$  and  $r$  filters of the SkyMapper Transient Survey.
10. The average PSF FWHM of the images is 2.7 arcsec, but has steadily degraded from 2018 to the end of the survey period.
11. Updated cross-matches with additional multi-wavelength catalogues have been included.

We look forward to utilisation of DR4 by the astronomical community via the tools of the All-Sky Virtual Observatory and the SkyMapper website (<https://skymapper.anu.edu.au>), or through partner data hosts.

**Acknowledgement.** We acknowledge the Gamilaroi people as the traditional owners of the land on which the SkyMapper Telescope stands. We are indebted to the original PIs of the S4 and SMSS projects: Brian P. Schmidt, Paul J. Francis, and MSB; as well as to the numerous students, postdocs, and technical staff who devoted themselves to the success of SkyMapper. We thank Ian Adams and the other hard-working staff of Siding Spring Observatory, whose diligent efforts have supported SkyMapper's ongoing productivity. We thank Annino Vaccarella and Mike Ellis for their tireless dedication to resolving technical challenges. We thank Peter Onaka, Sidik Isani, and the STARGRASP team for critical assistance with the CCD controller system. We thank Andrew Robinson, Robert Cohen, Andrew Howard, James Fitzsimmons, George Seaton, and the rest of the staff from the National Computational Infrastructure (NCI) for their long-standing support of the SkyMapper node of the All-Sky Virtual Observatory (ASVO) and the SkyMapper project. We thank Chris Ramage for retrieving the AAT seeing records.

We thank the referee for suggestions that improved the manuscript, particularly in the discussion of extended sources.

We thank the *Gaia* team for their assistance with the generation of synthetic photometry used in our photometric calibration, particularly Francesca De Angeli and Nicholas Walton from the Institute of Astronomy at the University of Cambridge, and Paolo Montegriffo from the Osservatorio Astronomico di Bologna of the Italian National Institute for Astrophysics.

We thank the team from the NOIRLab Astro Data Lab, particularly Robert Nikutta, Adam Scott, Mike Fitzpatrick, and Benjamin Weaver, as well as David Nidever from Montana State University, for making the DES DR2 and NSC DR2 datasets available for cross-matching to this data release.

We thank Alice Jacques from NOIRLab for helping make SMSS DR4 available at the Astro Data Lab; Francois-Xavier Pineau from CDS for making DR4 available via VizieR and the CDS XMatch Service; Aniruddha Thakar, Suzanne Werner, and Victor Paul from JHU for helping to make DR4 available on the SciServer platform; and Susan Mullally and Brian McLean from STScI for making DR4 available through MAST.

The national facility capability for SkyMapper has been funded through ARC LIEF grant LE130100104 from the Australian Research Council (ARC), awarded to the University of Sydney, the Australian National University, Swinburne University of Technology, the University of Queensland, the University of Western Australia, the University of Melbourne,

Curtin University of Technology, Monash University and the Australian Astronomical Observatory. Parts of this project were conducted by the Australian Research Council Centre of Excellence for All-sky Astrophysics (CAASTRO), through project number CE110001020. We acknowledge support from the ARC Discovery Projects program, most recently through DP190100252.

SWC acknowledges support from the National Research Foundation of Korea (NRF) grants, No. 2020R1A2C3011091 and No. 2021M3F7A1084525 funded by the Ministry of Science and ICT (MSIT). This research was also supported by Basic Science Research Program through the NRF funded by the Ministry of Education (RS-2023-00245013).

Development and support for the SkyMapper node of the ASVO has been funded in part by Astronomy Australia Limited (AAL) and the Australian Government through the Commonwealth's Education Investment Fund (EIF) and National Collaborative Research Infrastructure Strategy (NCRIS), particularly the National eResearch Collaboration Tools and Resources (NeCTAR) and the Australian National Data Service Projects (ANDS). The NCI, which is supported by the Australian Government, has contributed resources and services to this project and hosts the SkyMapper node of the ASVO. We also thank the ANU Major Equipment Committee (via grant 14MEC25), the Research School of Astronomy & Astrophysics, and CAASTRO for financial contributions toward the purchase of the SkyMapper server on which the mean object properties were distilled. The NCI resources used for the image processing were granted through the ANU Merit Allocation Scheme.

This work has made use of data from the European Space Agency (ESA) mission *Gaia* (<https://www.cosmos.esa.int/gaia>), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, <https://www.cosmos.esa.int/web/gaia/dpac/consortium>). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement. This work has made use of the PYTHON package GAIAXPY, developed and maintained by members of the *Gaia* Data Processing and Analysis Consortium (DPAC), and in particular, Coordination Unit 5 (CU5), and the Data Processing Centre located at the Institute of Astronomy, Cambridge, UK (DPCI).

This publication makes use of data products from the Wide-field Infrared Survey Explorer, which is a joint project of the University of California, Los Angeles, and the Jet Propulsion Laboratory/California Institute of Technology, and NEOWISE, which is a project of the Jet Propulsion Laboratory/California Institute of Technology. WISE and NEOWISE are funded by the National Aeronautics and Space Administration.

This research uses services or data provided by the Astro Data Lab at National Science Foundation's National Optical-Infrared Astronomy Research Laboratory. NOIRLab is operated by the Association of Universities for Research in Astronomy (AURA), Inc. under a cooperative agreement with the NSF.

This research has made extensive use of TOPCAT: Tool for OPERations on Catalogues And Tables, developed by Mark Taylor, for data analysis and figure generation (Taylor 2005). This research has made use of 'Aladin sky atlas' developed at CDS, Strasbourg Observatory, France. This research made use of ASTROPY, a community-developed core PYTHON package for Astronomy (Astropy Collaboration 2013, 2018).

## References

- Abbott, T. M. C., et al. 2021, *ApJS*, **255**, 20
- Alcock, C., et al. 2000, *ApJ*, **542**, 281
- Astropy Collaboration, et al. 2018, *AJ*, **156**, 123
- Astropy Collaboration, et al. 2013, *A&A*, **558**, A33
- Baldry, I. K., et al. 2018, *MNRAS*, **474**, 3875
- Bertin, E., & Arnouts, S. 1996, *A&AS*, **117**, 393
- Bessell, M., et al. 2011, *PASP*, **123**, 789
- Bianchi, L., Shiao, B., & Thilker, D. 2017, *ApJS*, **230**, 24
- Blake, C., et al. 2016, *MNRAS*, **462**, 4240
- Boch, T., & Fernique, P. 2014, in *Astronomical Society of the Pacific Conference Series*, Vol. 485, *Astronomical Data Analysis Software and Systems XXIII*, ed. N. Manset, & P. Forshay, 277
- Bohlin, R. C., Gordon, K. D., & Tremblay, P. E. 2014, *PASP*, **126**, 711

- Bohlin, R. C., Hubeny, I., & Rauch, T. 2020, *AJ*, **160**, 21
- Buder, S., et al. 2021, *MNRAS*, **506**, 150
- Casagrande, L., et al. 2019, *MNRAS*, **482**, 2770
- Chambers, K. C., et al. 2016, arXiv e-prints, arXiv:1612.05560
- Chang, S.-W., et al. 2021, *PASA*, **38**, e024
- Chang, S.-W., Wolf, C., Onken, C. A., & Bessell, M. S. 2024, *MNRAS*, **529**, 1414
- Colless, M., et al. 2003, arXiv e-prints, astro-ph/0306581
- Croom, S. M., et al. 2004, *MNRAS*, **349**, 1397
- Cutri, R. M., et al. 2013, Explanatory Supplement to the AllWISE Data Release Products, by R. M. Cutri et al.
- Da Costa, G. S., et al. 2019, *MNRAS*, **489**, 5900
- De Angeli, F., et al. 2023, *A&A*, **674**, A2
- Dey, A., et al. 2019, *AJ*, **157**, 168
- Drlica-Wagner, A., et al. 2018, *ApJS*, **235**, 33
- Edge, A., et al. 2013, *Msngr*, **154**, 32
- Fitzpatrick, M. J., et al. 2014, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9149, Observatory Operations: Strategies, Processes, and Systems V, ed. A. B. Peck, C. R. Benn, & R. L. Seaman, 91491T
- Flesch, E. W. 2023, arXiv e-prints, arXiv:2308.01505
- Gaia Collaboration, et al. 2018, *A&A*, **616**, A1
- Gaia Collaboration, et al. 2023a, *A&A*, **674**, A1
- Gaia Collaboration, et al. 2023b, *A&A*, **674**, A33
- Granlund, A., et al. 2006, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 6269, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, ed. I. S. McLean, & M. Iye, 626927
- Howes, L. M., et al. 2015, *Natur*, **527**, 484
- Huang, Y., et al. 2021, *ApJ*, **907**, 68
- Huang, Y., et al. 2022, *ApJ*, **924**, 141
- Huchra, J. P., et al. 2012, *ApJS*, **199**, 26
- Jarrett, T. H., et al. 2000, *AJ*, **119**, 2498
- Jenniskens, P., et al. 2021, *MPS*, **56**, 844
- Jones, D. H., et al. 2009, *MNRAS*, **399**, 683
- Keller, S. C., et al. 2014, *Nature*, **506**, 463
- Keller, S. C., et al. 2007, *PASA*, **24**, 1
- Kron, R. G. 1980, *ApJS*, **43**, 305
- Kuijken, K., et al. 2019, *A&A*, **625**, A2
- Lang, D., Hogg, D. W., Mierle, K., Blanton, M., & Roweis, S. 2010, *AJ*, **139**, 1782
- Li, J., et al. 2022, *MNRAS*, **515**, 3370
- Lidman, C., et al. 2020, *MNRAS*, **496**, 19
- Luvaul, L. C., Onken, C. A., Wolf, C., Smillie, J. G., & Sebo, K. M. 2017, in Astronomical Society of the Pacific Conference Series, Vol. 512, Astronomical Data Analysis Software and Systems XXV, ed. N. P. F. Lorente, K. Shortridge, & R. Wayth, 393
- Marocco, F., et al. 2021, *ApJS*, **253**, 8
- McMahon, R. G., et al. 2013, *Msngr*, **154**, 35
- Mendes de Oliveira, C., et al. 2019, *MNRAS*, **489**, 241
- Mink, D. J. 1996, in Astronomical Society of the Pacific Conference Series, Vol. 101, Astronomical Data Analysis Software and Systems V, ed. G. H. Jacoby, & J. Barnes, 96
- Mink, J. 2019, in Astronomical Society of the Pacific Conference Series, Vol. 523, Astronomical Data Analysis Software and Systems XXVII, ed. P. J. Teuben, M. W. Pound, B. A. Thomas, & E. M. Warner, 281
- Möller, A., et al. 2019, in Southern Horizons in Time-Domain Astronomy, ed. R. E. Griffin, Vol. 339, 3–6
- Montegriffo, P., et al. 2023, *A&A*, **674**, A3
- Nidever, D. L., et al. 2021, *AJ*, **161**, 192
- Nikutta, R., Fitzpatrick, M., Scott, A., & Weaver, B. A. 2020, *A&C*, **33**, 100411
- Nordlander, T., et al. 2019, *MNRAS*, **488**, L109
- Oh, W. S., Nordlander, T., Da Costa, G. S., Bessell, M. S., & Mackey, A. D. 2023, *MNRAS*, **524**, 577
- Oke, J. B. & Gunn, J. E. 1983, *ApJ*, **266**, 713
- Onaka, P., et al. 2008, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7014, Ground-based and Airborne Instrumentation for Astronomy II, ed. I. S. McLean, & M. M. Casali, 70140D
- Onken, C. A., et al. 2019, *PASA*, **36**, e033
- Onken, C. A., et al. 2022, *MNRAS*, **511**, 572
- Onken, C. A., et al. 2023, *PASA*, **40**, e010
- Pence, W. D., Seaman, R., & White, R. L. 2009, *PASP*, **121**, 414
- Petrosian, V. 1976, *ApJ*, **210**, L53
- Planck Collaboration, et al. 2014, *A&A*, **571**, A11
- Planck Collaboration, et al. 2016, *A&A*, **596**, A109
- Rakich, A., et al. 2006, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 6267, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, ed. L. M. Stepp, 62670E
- Regnault, N., et al. 2015, *A&A*, **581**, A45
- Riello, M., et al. 2021, *A&A*, **649**, A3
- Robertson, J. G. 2021, *PASA*, **38**, e006
- Rocci, P.-F. 2013, Calibration of Wide Field Imagers - The SkyDICE Project, PhD thesis, Université Pierre et Marie Curie - Paris VI
- Scalzo, R. A., et al. 2017, *PASA*, **34**, e030
- Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, *ApJ*, **500**, 525
- Schröder, A. C., van Driel, W., & Kraan-Korteweg, R. C. 2021, *MNRAS*, **503**, 5351
- Skrutskie, M. F., et al. 2006, *AJ*, **131**, 1163
- Steinmetz, M., et al. 2020, *AJ*, **160**, 82
- Taylor, M. B. 2005, in Astronomical Society of the Pacific Conference Series, Vol. 347, Astronomical Data Analysis Software and Systems XIV, ed. P. Shopbell, M. Britton, & R. Ebert, 29
- Thakar, A. R. 2008, *CSE*, **10**, 65
- Tonry, J. L., et al. 2018, *ApJ*, **867**, 105
- Vaccarella, A., et al. 2008, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7019, Advanced Software and Control for Astronomy II, ed. A. Bridger, & N. M. Radziwill, 70192R
- van Dokkum, P. G. 2001, *PASP*, **113**, 1420
- Watson, C. L., Henden, A. A., & Price, A. 2006, Society for Astronomical Sciences Annual Symposium, **25**, 47
- Wilson, G., et al. 2005, in Astronomical Society of the Pacific Conference Series, Vol. 347, Astronomical Data Analysis Software and Systems XIV, ed. P. Shopbell, M. Britton, & R. Ebert, 563
- Wisotzki, L., et al. 2000, *A&A*, **358**, 77
- Wolf, C., et al. 2018a, *PASA*, **35**, e024
- Wolf, C., et al. 2018b, *PASA*, **35**, e010
- Wolf, C., Luvaul, L. C., Onken, C. A., Smillie, J. G., & White, M. C. 2017, in Astronomical Society of the Pacific Conference Series, Vol. 512, Astronomical Data Analysis Software and Systems XXV, ed. N. P. F. Lorente, K. Shortridge, & R. Wayth, 289
- Yong, D., et al. 2021, *MNRAS*, **507**, 4102
- Young, P. J., Brooks, M., Meatheringham, S. J., & Roberts, W. H. 1997, in Astronomical Society of the Pacific Conference Series, Vol. 125, Astronomical Data Analysis Software and Systems VI, ed. G. Hunt, & H. Payne, 385
- Young, P. J., Roberts, W. H., & Sebo, K. M. 1999, in Astronomical Society of the Pacific Conference Series, Vol. 172, Astronomical Data Analysis Software and Systems VIII, ed. D. M. Mehringer, R. L. Plante, & D. A. Roberts, 115
- Zacharias, N., et al. 2013, *AJ*, **145**, 44

## Appendix 1. Metadata for master table

In Table A1, we provide a description of the `dr4.master` table columns.

**Table A1.** Name, description, units, minimum value, and maximum value for each column in the SMSS DR4 master table.

Column Name	Description	Units	Minimum	Maximum
<code>object_id</code>	Global unique object ID in the master table	–	1	2724329343
<code>raj2000</code>	Mean ICRS Right Ascension of the object	deg	0	359.999998
<code>dej2000</code>	Mean ICRS Declination of the object	deg	–89.992883	29.182411
<code>e_raj2000</code>	RMS variation around the mean Right Ascension in milliarcseconds	mas	36	7 311
<code>e_dej2000</code>	RMS variation around the mean Declination in milliarcseconds	mas	36	7 562
<code>smss_j</code>	SkyMapper Southern Survey designation of the form SMSS Jhhmmss.ss±ddmmss.s derived from mean ICRS coordinates	–	000000.00+095415.5	235959.99-755208.7
<code>mean_epoch</code>	Mean MJD epoch of the observations	d	56 730.6305	59 473.7663
<code>rms_epoch</code>	RMS variation around the mean epoch	d	0	1 934.46
<code>glon</code>	Galactic longitude derived from ICRS coordinates. Not to be used as primary astrometric reference	deg	6.71978E-7	360
<code>glat</code>	Galactic latitude derived from ICRS coordinates. Not to be used as primary astrometric reference	deg	–89.9936	79.3879
<code>flags</code>	Bitwise OR of SOURCE EXTRACTOR flags across all observations	–	0	16 384
<code>nimaflags</code>	Total number of flagged pixels from bad or saturated or crosstalk-affected pixel masks across all observations	–	0	23069556
<code>ngood</code>	Number of observations used across all filters	–	0	11 306
<code>u_flags</code>	Bitwise OR of SOURCE EXTRACTOR flags from u-band measurements in photometry table	–	0	26 136
<code>u_nimaflags</code>	Number of flagged pixels from bad or saturated or crosstalk-affected pixels masks from u-band measurements in photometry table	–	0	32 767
<code>u_ngood</code>	Number of u-band observations used	–	0	1 604
<code>u_nclip</code>	Number of u-band observations with magnitudes clipped from the final PSF magnitude estimate	–	0	252
<code>v_flags</code>	Bitwise OR of SOURCE EXTRACTOR flags from v-band measurements in photometry table	–	0	26 136
<code>v_nimaflags</code>	Number of flagged pixels from bad or saturated or crosstalk-affected pixels masks from v-band measurements in photometry table	–	0	32 767
<code>v_ngood</code>	Number of v-band observations used	–	0	1 739
<code>v_nclip</code>	Number of v-band observations with magnitudes clipped from the final PSF magnitude estimate	–	0	362
<code>g_flags</code>	Bitwise OR of SOURCE EXTRACTOR flags from g-band measurements in photometry table	–	0	28 160
<code>g_nimaflags</code>	Number of flagged pixels from bad or saturated or crosstalk-affected pixels masks from g-band measurements in photometry table	–	0	32 767
<code>g_ngood</code>	Number of g-band observations used	–	0	1 914
<code>g_nclip</code>	Number of g-band observations with magnitudes clipped from the final PSF magnitude estimate	–	0	846
<code>r_flags</code>	Bitwise OR of SOURCE EXTRACTOR flags from r-band measurements in photometry table	–	0	28 176



**Table A1.** Continued

Column Name	Description	Units	Minimum	Maximum
r_nimaflags	Number of flagged pixels from bad or saturated or crosstalk-affected pixels masks from r-band measurements in photometry table	–	0	32 767
r_ngood	Number of r-band observations used	–	0	1 968
r_nclip	Number of r-band observations with magnitudes clipped from the final PSF magnitude estimate	–	0	833
i_flags	Bitwise OR of SOURCE EXTRACTOR flags from i-band measurements in photometry table	–	0	26 139
i_nimaflags	Number of flagged pixels from bad or saturated or crosstalk-affected pixels masks from i-band measurements in photometry table	–	0	32 767
i_ngood	Number of i-band observations used	–	0	2 381
i_nclip	Number of i-band observations with magnitudes clipped from the final PSF magnitude estimate	–	0	859
z_flags	Bitwise OR of SOURCE EXTRACTOR flags from z-band measurements in photometry table	–	0	26 136
z_nimaflags	Number of flagged pixels from bad or saturated or crosstalk-affected pixels masks from z-band measurements in photometry table	–	0	32 767
z_ngood	Number of z-band observations used	–	0	2 175
z_nclip	Number of z-band observations with magnitudes clipped from the final PSF magnitude estimate	–	0	854
class_star	Maximum stellarity index from photometry table (between 0=no star and 1=star)	–	0	1
chi2_psf	Maximum chi-squared from photometry table	–	0	3.2558e+06
flags_psf	Bitmask indicating whether photometry is likely biased by neighbours at >1%; bits 0-5 correspond to filters z-u	–	0	63
radius_petro	Mean r-band Petrosian radius factor	semi-major axis lengths	3.5	10.56
mean_fwhm	Mean FWHM of detections	pix	-2.27989e+08	3 696.12
u_psf	Weighted mean u-band PSF magnitude	mag	3.5168	22.4221
e_u_psf	Error in weighted mean u-band PSF magnitude	mag	0.0014	3.0543
u_petro	Weighted mean u-band Petrosian magnitude	mag	0.3224	31.7358
e_u_petro	Error in weighted mean u-band Petrosian magnitude	mag	0.0014	2 939.69
u_apc05	Weighted mean u-band aperture-corrected magnitude from flux in 5 arcsec diameter	mag	6.1501	31.6061
e_u_apc05	Error in weighted mean u-band aperture-corrected magnitude from flux in 5 arcsec diameter	mag	0.0015	2.4124
u_mmvar	Magnitude range between minimum and maximum unclipped u-band PSF magnitudes	mag	0	7.3489
v_psf	Weighted mean v-band PSF magnitude	mag	3.1073	22.2866
e_v_psf	Error in weighted mean v-band PSF magnitude	mag	0.0009	2.5687
v_petro	Weighted mean v-band Petrosian magnitude	mag	0.1233	35.1441
e_v_petro	Error in weighted mean v-band Petrosian magnitude	mag	0.0009	108 244
v_apc05	Weighted mean v-band aperture-corrected magnitude from flux in 5 arcsec diameter	mag	6.0923	30.1388
e_v_apc05	Error in weighted mean v-band aperture-corrected magnitude from flux in 5 arcsec diameter	mag	0.0009	2.3113
v_mmvar	Magnitude range between minimum and maximum unclipped v-band PSF magnitudes	mag	0	7.3253
g_psf	Weighted mean g-band PSF magnitude	mag	2.8979	24.0783
e_g_psf	Error in weighted mean g-band PSF magnitude	mag	0.0006	4.1083

**Table A1.** Continued

Column Name	Description	Units	Minimum	Maximum
g_petro	Weighted mean g-band Petrosian magnitude	mag	0.4787	36.1903
e_g_petro	Error in weighted mean g-band Petrosian magnitude	mag	0.0006	82992.5
g_apc05	Weighted mean g-band aperture-corrected magnitude from flux in 5 arcsec diameter	mag	6.0203	30.9735
e_g_apc05	Error in weighted mean g-band aperture-corrected magnitude from flux in 5 arcsec diameter	mag	0.0007	3.7124
g_mmvar	Magnitude range between minimum and maximum unclipped g-band PSF magnitudes	mag	0	11.3157
r_psf	Weighted mean r-band PSF magnitude	mag	2.5215	25.3334
e_r_psf	Error in weighted mean r-band PSF magnitude	mag	0.0006	4.2441
r_petro	Weighted mean r-band Petrosian magnitude	mag	0.5543	35.4118
e_r_petro	Error in weighted mean r-band Petrosian magnitude	mag	0.0006	23 988.6
r_apc05	Weighted mean r-band aperture-corrected magnitude from flux in 5 arcsec diameter	mag	5.8349	34.6417
e_r_apc05	Error in weighted mean r-band aperture-corrected magnitude from flux in 5 arcsec diameter	mag	0.0006	3.1773
r_mmvar	Magnitude range between minimum and maximum unclipped r-band PSF magnitudes	mag	0	9.7395
i_psf	Weighted mean i-band PSF magnitude	mag	2.4086	25.446
e_i_psf	Error in weighted mean i-band PSF magnitude	mag	0.0003	3.9294
i_petro	Weighted mean i-band Petrosian magnitude	mag	0.586	34.3901
e_i_petro	Error in weighted mean i-band Petrosian magnitude	mag	0.0003	186 518
i_apc05	Weighted mean i-band aperture-corrected magnitude from flux in 5 arcsec diameter	mag	5.4024	32.7377
e_i_apc05	Error in weighted mean i-band aperture-corrected magnitude from flux in 5 arcsec diameter	mag	0.0004	2.8266
i_mmvar	Magnitude range between minimum and maximum unclipped i-band PSF magnitudes	mag	0	9.1216
z_psf	Weighted mean z-band PSF magnitude	mag	1.3949	22.2704
e_z_psf	Error in weighted mean z-band PSF magnitude	mag	0.0007	3.8631
z_petro	Weighted mean z-band Petrosian magnitude	mag	0.6096	33.6571
e_z_petro	Error in weighted mean z-band Petrosian magnitude	mag	0.0007	359 127
z_apc05	Weighted mean z-band aperture-corrected magnitude from flux in 5 arcsec diameter	mag	4.7756	32.5408
e_z_apc05	Error in weighted mean z-band aperture-corrected magnitude from flux in 5 arcsec diameter	mag	0.0008	2.9392
z_mmvar	Magnitude range between minimum and maximum unclipped z-band PSF magnitudes	mag	0	9.0278
self_id1	Unique identifier (object_id) of closest SMSS DR4 source	–	1	2724329340
self_dist1	Distance on sky to closest SMSS DR4 source	arcsec	0.03	15
self_id2	Unique identifier (object_id) of second-closest SMSS DR4 source	–	2	2724329343
self_dist2	Distance on sky to second-closest SMSS DR4 source	arcsec	0.18	15
self_id3	Unique identifier (object_id) of third-closest SMSS DR4 source	–	40	2724327091
self_dist3	Distance on sky to third-closest SMSS DR4 source	arcsec	0.26	15
cnt_self_15	Number of SMSS DR4 sources within 15 arcsec	–	1	107
ebmv_sfd	E(B-V) from Schlegel+1998 extinction maps at the ICRS coordinates	mag	0.0006	152.148
ebmv_gnilc	E(B-V) from Planck satellite GNILC reddening maps at the ICRS coordinates	mag	0	164.918

**Table A1.** Continued

Column Name	Description	Units	Minimum	Maximum
ebmv_g_err	Error in the E(B-V) from GNILC	mag	0	52.6726
gaia_dr3_id1	Unique identifier (source_id) of closest Gaia DR3 source	–	4295806720	6917528997577384320
gaia_dr3_dist1	Distance on sky to closest Gaia DR3 source	arcsec	0	15
gaia_dr3_id2	Unique identifier (source_id) of second-closest Gaia DR3 source	–	34361129088	6917528997577384320
gaia_dr3_dist2	Distance on sky to second-closest Gaia DR3 source	arcsec	0.09	15
cnt_gaia_dr3_15	Number of Gaia DR3 sources within 15 arcsec	–	1	333
twomass_key	Unique identifier (pts_key) of closest 2MASS PSC source	–	40	1339356373
twomass_dist	Distance on sky to closest 2MASS PSC source	arcsec	0	15
allwise_cnr	Unique identifier (cntr) of closest ALLWISE source	–	1601351000001	3584116601351035683
allwise_dist	Distance on sky to closest ALLWISE source	arcsec	0	15
catwise_id	Unique identifier (id) of closest CatWISE2020 source	–	0000m016_b0-000001	3584p166_b0-080605
catwise_dist	Distance on sky to closest CatWISE2020 source	arcsec	0	15
refcat2_id	Unique identifier (objid) of closest ATLAS Refcat2 source (only useful for this copy of Refcat2)	–	1	992637833
refcat2_dist	Distance on sky to closest ATLAS Refcat2 source	arcsec	0	15
ps1_dr1_id	Unique identifier (objid) of closest Pan-STARRS1 DR1 source	–	65810810454945744	143011300804053051
ps1_dr1_dist	Distance on sky to closest Pan-STARRS1 DR1 source	arcsec	0	15
galex_guv_id	Unique identifier (objid) of closest GALEX GUVcat AIS source	–	6374012077942505627	6388191387412599267
galex_guv_dist	Distance on sky to closest GALEX GUVcat AIS source	arcsec	0	15
vhs_dr6_id	Unique identifier (objid) of closest Pan-STARRS1 DR1 source	–	472446402562	473820609749
vhs_dr6_dist	Distance on sky to closest Pan-STARRS1 DR1 source	arcsec	0	15
ls_dr9_id	Unique identifier (unique_id) of closest Legacy Survey DR9 source	–	100000236480000014	120004860270005003
ls_dr9_dist	Distance on sky to closest Legacy Survey DR9 source	arcsec	0	15
des_dr2_id	Unique identifier (coadd_object_id) of closest DES DR2 source	–	870245760	1700560060
des_dr2_dist	Distance on sky to closest DES DR2 source	arcsec	0	15
nsc_dr2_id	Unique identifier (id) of closest NSC DR2 source	–	100000_1000	99999_9999
nsc_dr2_dist	Distance on sky to closest NSC DR2 source	arcsec	0	15
splus_dr3_id	Unique identifier (id) of closest S-PLUS DR3 source	–	DR3.HYDRA-0011.000000	DR3.STRIPE82-0170.061552
splus_dr3_dist	Distance on sky to closest S-PLUS DR3 source	arcsec	0	15