The Astrophysical Journal Supplement Series, 225:5 (24pp), 2016 July	doi:10.3847/0067-0049/225/1/5

WISE × SuperCOSMOS PHOTOMETRIC REDSHIFT CATALOG: 20 MILLION GALAXIES OVER 3π STERADIANS

Maciej Bilicki[1,2,3], John A. Peacock[4], Thomas H. Jarrett[1], Michelle E. Cluver[5], Natasha Maddox[6], Michael J. I. Brown[7], Edward N. Taylor[8], Nigel C. Hambly[4], Aleksandra Solarz[3,9], Benne W. Holwerda[2], Ivan Baldry[10], Jon Loveday[11], Amanda Moffett[12], Andrew M. Hopkins[13], Simon P. Driver[12,14], Mehmet Alpaslan[15], and Joss Bland-Hawthorn[16]

[1] Department of Astronomy, University of Cape Town, Private Bag X3, Rondebosch, 7701, South Africa; maciek@ast.uct.ac.za
[2] Leiden Observatory, Leiden University, Niels Bohrweg 2, NL-2333 CA Leiden, The Netherlands
[3] Janusz Gil Institute of Astronomy, University of Zielona Góra, ul. Szafrana 2, 65-516, Zielona Góra, Poland
[4] Institute for Astronomy, University of Edinburgh, Royal Observatory, Edinburgh EH9 3HJ, UK
[5] Department of Physics, University of the Western Cape, Robert Sobukwe Road, Bellville, 7530, South Africa
[6] ASTRON, The Netherlands Institute for Radio Astronomy, Postbus 2, 7990 AA Dwingeloo, The Netherlands
[7] School of Physics & Astronomy, Monash University, Clayton, Victoria 3800, Australia
[8] School of Physics, University of Melbourne, VIC 3010, Australia
[9] National Centre for Nuclear Research, ul. Hoża 69, Warsaw, Poland
[10] Astrophysics Research Institute, Liverpool John Moores University, IC2, Liverpool Science Park, 146 Brownlow Hill, Liverpool, L3 5RF, UK
[11] Astronomy Centre, University of Sussex, Falmer, Brighton BN1 9QH, UK
[12] ICRAR, The University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia
[13] Australian Astronomical Observatory, P.O. Box 915, North Ryde, NSW 1670, Australia
[14] School of Physics and Astronomy, University of St Andrews, North Haugh, St Andrews, KY16 9SS, UK
[15] NASA Ames Research Center, N232, Moffett Field, Mountain View CA 94035, USA
[16] Sydney Institute for Astronomy, School of Physics A28, University of Sydney, NSW 2006, Australia

Received 2015 November 1; revised 2016 April 7; accepted 2016 April 22; published 2016 July 13

ABSTRACT

We cross-match the two currently largest all-sky photometric catalogs—mid-infrared *Wide-field Infrared Survey Explorer* and SuperCOSMOS scans of UKST/POSS-II photographic plates—to obtain a new galaxy sample that covers 3π steradians. In order to characterize and purify the extragalactic data set, we use external GAMA and Sloan Digital Sky Survey spectroscopic information to define quasar and star loci in multicolor space, aiding the removal of contamination from our extended source catalog. After appropriate data cleaning, we obtain a deep wide-angle galaxy sample that is approximately 95% pure and 90% complete at high Galactic latitudes. The catalog contains close to 20 million galaxies over almost 70% of the sky, outside the Zone of Avoidance and other confused regions, with a mean surface density of more than 650 sources per square degree. Using multiwavelength information from two optical and two mid-IR photometric bands, we derive photometric redshifts for all the galaxies in the catalog, using the ANNz framework trained on the final GAMA-II spectroscopic data. Our sample has a median redshift of $z_{\rm med} = 0.2$, with a broad dN/dz reaching up to $z > 0.4$. The photometric redshifts have a mean bias of $|\delta z| \sim 10^{-3}$, a normalized scatter of $\sigma_z = 0.033$, and less than 3% outliers beyond $3\sigma_z$. Comparison with external data sets shows no significant variation of photo-z quality with sky position. Together with the overall statistics, we also provide a more detailed analysis of photometric redshift accuracy as a function of magnitudes and colors. The final catalog is appropriate for "all-sky" three-dimensional (3D) cosmology to unprecedented depths, in particular through cross-correlations with other large-area surveys. It should also be useful for source preselection and identification in forthcoming surveys, such as TAIPAN or WALLABY.

Key words: catalogs – galaxies: distances and redshifts – large-scale structure of universe – methods: data analysis – methods: statistical – surveys

1. INTRODUCTION

Direct mapping of the three-dimensional (3D) distribution of galaxies in the universe requires their angular coordinates and redshifts. Dozens of such wide-angle galaxy redshift catalogs now exist, the most notable of which include the SDSS (York et al. 2000), the Two-degree Field Galaxy Redshift Survey (Colless et al. 2001), and the Six-degree Field Galaxy Survey (6dFGS; Jones et al. 2004).

For some applications, it is an advantage if the survey can cover the majority of the sky; for example, searches for a violation of the Copernican principle in the form of large-scale inhomogeneities or anisotropies (Gibelyou & Huterer 2012; Appleby & Shafieloo 2014; Alonso et al. 2015; Yoon & Huterer 2015) and coherent motions (Bilicki et al. 2011; Branchini et al. 2012; Carrick et al. 2015), as well as cross-correlations of galaxy data with external wide-angle data sets. Examples of the latter include studies of the integrated Sachs–Wolfe effect (see Nishizawa 2014, for a review), gravitational lensing of the cosmic microwave background (CMB) on the large-scale structure (Lewis & Challinor 2006), and searches for sources of the extragalactic γ-ray background (e.g., Xia et al. 2015), including constraints on annihilating or decaying dark matter (Cuoco et al. 2015). These analyses are limited by cosmic variance, and much of the signal frequently lies at large angular scales—both of which are factors that make it desirable to have the largest possible sky coverage.

But there is a practical limit to the number of spectroscopic redshifts that can be measured in a reasonable time. Spectroscopic galaxy catalogs covering the *whole* extragalactic sky, such as the IRAS Point Source Catalog Redshift Survey (PSCz;

Saunders et al. 2000) and the 2MASS Redshift Survey (2MRS; Huchra et al. 2012), tend to be relatively shallow ($z < 0.1$) and the same applies to hemispherical samples such as the 6dFGS. This problem can be addressed by using only rare tracers—as with the highly successful BOSS program (Dawson et al. 2013) or planned projects such as the Dark Energy Spectroscopic Instrument (Levi et al. 2013) or Wide Area VISTA Extragalactic Survey (Driver et al. 2015) within the 4MOST program—but for many applications it is desirable to have a fully sampled galaxy density field. For that reason, new wide-field surveys such as the Dark Energy Survey (The Dark Energy Survey Collaboration 2005), Pan-STARRS (Kaiser et al. 2002), and the Kilo-Degree Survey (de Jong et al. 2013) focus on measuring the photometric properties of objects, with only a partial spectroscopic follow-up. In the longer term, the same will apply to forthcoming multi-billion-object facilities including Euclid (Laureijs et al. 2011) and the Large Synoptic Survey Telescope (LSST Science Collaboration et al. 2009). Lying somewhat in between the spectroscopic and photometric surveys, the Javalambre-Physics of the Accelerated universe Astrophysical Survey (J-PAS, Benitez et al. 2014) is expected to reach sub-percent redshift precision on ~8000 $\deg^2$, thanks to the use of 56 narrow-band filters. Of a similar nature, but aiming to cover 100 $\deg^2$ to a greater depth than J-PAS, is the Physics of the Accelerating universe survey (PAU; Martí et al. 2014).

In order for such surveys to yield cosmological information of comparable or even better quality than from traditional spectroscopic samples, one needs to resort to the technique of *photometric redshifts* (photo-zs). In the near future, this approach will dominate those cosmological analyses where the benefit from larger volumes outweighs the loss of redshift accuracy. Although some small-scale analyses are not feasible with the coarse accuracy of photo-z estimation (typically a few percent precision), there are many applications where this level of measurement is more than adequate. This is particularly true when there is an angular signal that changes slowly with redshift, requiring a *tomographic* analysis in broad redshift bins (e.g., Francis & Peacock 2010); but until recently that necessary photo-z information has only been available for relatively shallow subsamples of all-sky catalogs.

To improve on this situation, in Bilicki et al. (2014; hereafter B14) we combined three all-sky photometric samples—optical SuperCOSMOS, near-infrared 2MASS, and mid-infrared *Wide-field Infrared Survey Explorer* (*WISE*)—into a multiwavelength data set. We used various spectroscopic calibration samples to compute photometric redshifts for almost 1 million galaxies over most of the extragalactic sky: the 2MASS Photometric Redshift catalog (2MPZ).[17] The 2MPZ is currently the deepest 3D full-sky galaxy data set, with a median redshift of $z \simeq 0.1$ and a typical uncertainty in photometric redshift of about 12% (scatter $\sigma_z = 0.013$). Ideally, these estimates should be superseded by actual spectroscopy—and recently prospects have emerged for this to happen, thanks to the new hemispherical TAIPAN survey in the south, which is starting in 2016 (Kuehn et al. 2014), as well as the recently proposed LoRCA (Comparat et al. 2016) in the north. These efforts, if successful, will provide spectroscopic information for all the 2MASS galaxies that do not have redshifts, although at their planned depths ($r \lesssim 18$ for the former and $K_s < 14$ for the latter) they will not replace the need for the catalog presented in the current paper. We note, however, the *SPHEREx* concept by Doré et al. (2014) to probe much deeper on most of sky.

The depth of 2MPZ is limited by the shallowest of the three photometric surveys combined for its construction, the 2MASS Extended Source Catalog (XSC; Jarrett et al. 2000; Jarrett 2004). However, as was shown in B14, one can go beyond the 2MASS data and obtain a much deeper all-sky photo-z catalog based on *WISE* and SuperCOSMOS only. In B14 we predicted that such a sample should have a typical redshift error of $\sigma_z \simeq 0.035$ at a median $z \simeq 0.2$ (median relative error of 14%). The construction of this catalog is the focus of the present paper, and indeed we confirm and even exceed these expectations on the photo-z quality. We note that in a related effort Kovács & Szapudi (2015) presented a wide-angle sample that is deeper than the 2MASS XSC, based on *WISE* and the 2MASS *Point* Source Catalog (PSC). However, its depth is still limited by 2MASS: PSC has an order of magnitude smaller surface density than *WISE* (T. H. Jarrett et al. 2016, in preparation). Overall, the Kovács & Szapudi (2015) sample includes 2.4 million sources at $z_{\rm med} \simeq 0.14$ over half the sky, of which 1/3 are in common with the 2MASS XSC. Here we map the cosmic web to much higher redshifts than can be accessed with 2MASS, yielding a third shell of presently available all-sky redshift surveys. The first, with exact spectroscopic redshifts at $z_{\rm med} = 0.03$, is provided by the 2MRS, flux-limited to $K_s \leqslant 11.75$ (Vega) and contains 44,000 galaxies at $|b| > 5^\circ$ ($|b| > 8^\circ$ by the Galactic Bulge). The second is the 2MPZ, which includes almost a million 2MASS galaxies at $K_s < 13.9$ with precise photo-zs at $z_{\rm med} = 0.07$, based on 8-band 2MASS × *WISE* × SuperCOSMOS photometry. This present work concerns 20 million galaxies with $z_{\rm med} = 0.2$, thus reaching three times deeper than 2MPZ, over 3π steradians of the sky outside the Galactic Plane.

This paper is organized as follows. In Section 2 we provide a detailed description of the catalogs contributing to the sample and their cross-matching. In Section 3 we analyze the properties of the input photometric data sets by pairing them up with GAMA spectroscopic data. Section 4 describes the use of external GAMA and SDSS spectroscopic information to remove quasars and stellar blends from the cross-matched catalog. The construction of the angular mask to be applied to the data is also presented in Section 4.3. Next, in Section 5 we show how photometric redshifts were obtained for the sample and discuss several tests of their performance; Section 5.2 discusses the properties of the final all-sky catalog. In Section 6 we summarize and list selected possible applications of our data set.

2. CONTRIBUTING CATALOGS

The galaxy catalog presented in this paper is a combination of two major photometric surveys of the whole celestial sphere: optical SuperCOSMOS scans of photographic plates (SCOS for short) and mid-IR *WISE*. Each of these two data sets includes about 1 billion sources, a large fraction of which are extragalactic. *WISE* is deeper than SCOS, but its poorer resolution and lack of morphological information (the latter available from SCOS) prevent the selection of galaxies without the optical criterion of an extended image. Pairing up these two data sets thus provides a natural means of obtaining a deep wide-angle extragalactic sample, as we proposed in B14. With appropriate spectroscopic calibration data, the wide wavelength

---

[17] Available for download from the Wide Field Astronomy Unit, Edinburgh, at http://surveys.roe.ac.uk/ssa/TWOMPZ.

range yields robust photometric redshifts for each of the *WISE* × SCOS galaxies.

In this Section we describe the properties of the underlying photometric catalogs and the preselections applied to them. We aim for the highest depth possible for the cross-matched sample, while optimizing its reliability, purity, and completeness. By reliability, we chiefly refer to the quality of the photometry; purity refers to the percentage of our sources that are indeed galaxies and not stars, high-redshift quasars, or blends thereof; completeness is the fraction of all galaxies that are included in the catalog, within adopted magnitude limits. As our focus in the present paper is to derive photometric redshifts for all the galaxies in our sample, which requires multiwavelength coverage, we select from the two catalogs only those sources that have detections in at least four bands: $W1$ and $W2$ (3.4 and 4.6 μm) in *WISE*, and B and R in SCOS. The additional bands available from the two surveys, $W3$ and $W4$ (12 and 23 μm)[18] from *WISE*, and I from SCOS, are not used due to their low sensitivity and non-uniform sky coverage.

This exercise cannot be expected to yield a fully all-sky catalog: both *WISE* and SCOS suffer at low Galactic latitudes from severe blending of stars with other stars and with galaxies, and high Galactic extinction levels effectively censor the optical bands. In Section 4 we discuss how to minimize such foreground contamination, and develop a mask within which the overall catalog has an acceptable completeness and purity. In practice, we find that this can be done over about 70% of the sky.

2.1. WISE

WISE (Wright et al. 2010) is a NASA space-based mission that surveyed the celestial sphere in four infrared bands: 3.4, 4.6, 12, and 23 μm ($W1$–$W4$), with an angular resolution of $6\farcs1$, $6\farcs4$, $6\farcs5$, and $12''$, respectively. We use the AllWISE full-sky release[19] (Cutri et al. 2013), which combines data from the cryogenic and post-cryogenic survey phases and provides the most comprehensive picture of the full mid-infrared sky currently available. The AllWISE Source Catalog and Image Atlas have enhanced sensitivity and accuracy compared with earlier *WISE* data releases, especially in its two shortest bands. This results in a larger effective depth than available from an earlier "All-Sky" release (Cutri et al. 2012), used for example, in B14. AllWISE includes more than 747 million sources (mostly stars and galaxies) detected with $S/N \geqslant 5$ in at least one band. The 5σ sensitivities in the four respective bands are approximately[20] 0.054, 0.071, 0.73, and 5 mJy, and the 95% completeness averaged over large areas of unconfused sky is about[21] $W1 < 17.1$, $W2 < 15.7$, $W3 < 11.5$, and $W4 < 7.7$ in the Vega system.[22] However, the depth of coverage does vary over the sky due to the survey strategy, being much higher in the ecliptic poles and the lowest near the ecliptic plane (Jarrett et al. 2011); there are also some anomalous stripes resulting from moon avoidance maneuvers and instrumental issues.

The *WISE* photometric pipeline was not optimized for extended sources and the online database does not include a formal extended source catalog. The basic magnitudes (which we use here) are the `w?mpro` mags, based on PSF profile-fit measurements, where "?" stands for the particular channel number, from 1 to 4. This information is available for all objects, whereas existing attempts to handle extended images are somewhat heterogeneous. For instance, the `w?gmags`, which are measured in elliptical apertures derived from associated 2MASS XSC sources, are available only for the 483,000 largest *WISE* galaxies. Circular aperture magnitudes are in fact provided for practically all sources, namely the `w?mag_n`, where $n = 1, 2, \ldots, 8$; these were obtained from the coadded Atlas images in a series of different fixed radii. But the angular sizes of the sources have not been determined; in addition, this photometry does not account for source ellipticities, is prone to contamination from nearby objects, and is not compensated for saturated or missing pixels in the images.

In any case, as we eliminate all the bright ($W1 < 13.8$) sources from our cross-matched catalog (see Section 4), we are thus left with galaxies that are typically smaller than the *WISE* resolution threshold, which are well-described by PSF magnitudes, although we note that their fluxes might be underestimated by *WISE*. This is supported by independent analyses showing that the eventual *WISE* XSC will mainly include 2MASS XSC galaxies and be limited to $W1 \lesssim 14$ (Cluver et al. 2014; T. H. Jarrett et al. 2016, in preparation). In any case, any residual biases in photometry for resolved sources, which may influence source colors, will not be propagated to the photometric redshifts derived via the neural network framework employed here, as such systematics are automatically accounted for in the empirical training procedure.

Initially, we selected AllWISE sources with signal-to-noise ratios larger than 2 in its two shortest bands. This selection, meaning that we use detections in the two bands and not upper limits (the latter having $S/N < 2$ in *WISE*), is practically equivalent to selecting objects with `w1snr` $\geqslant 5$, as those with low S/N in $W1$ but high in $W2$ are extremely rare. Having cleaned the sample of obvious artifacts (`cc_flags[1,2]` = `'DPHO'`) and saturated sources (`w?sat` > 0.1), we ended up with more than 603 million AllWISE objects over the whole sky. In order to optimize all-sky uniformity, we applied a global magnitude cut of $W1 < 17$. This removes ~20% of AllWISE (mostly around the ecliptic poles, where the *WISE* depth is greatest), leaving 488 million objects (pictured in Figure 1). From this image, it is apparent that low Galactic latitudes are entirely dominated by stars and blends thereof; as we will show below, stellar contamination remains significant even at high latitudes (see also Jarrett et al. 2011; T. H. Jarrett et al. 2016, in preparation). A minimal Galactic restriction to $|b| > 10°$ lowers the total to 340 million sources (see Table 1 for a summary), but we will show that the final masking needs to be more severe than this.

Note that some sources observed during the early three-band cryo survey phase are not captured by the above selection, as they have missing $W1$ magnitude uncertainties and are listed as upper limits in the database. This is discussed in detail in the

---

[18] The $W4$ channel effective wavelength was recalibrated from the original 22 μm by Brown et al. (2014).

[19] Available for download from NASA/IPAC Infrared Science Archive at http://irsa.ipac.caltech.edu.

[20] http://wise2.ipac.caltech.edu/docs/release/allwise/expsup/sec2_3a.html

[21] http://wise2.ipac.caltech.edu/docs/release/allwise/expsup/sec2_4a.html

[22] Conversions of *WISE* magnitudes from Vega to AB are provided by Jarrett et al. (2011); for the bands of interest in this paper, $W1$ and $W2$, one needs to add, respectively, 2.70 and 3.34 to the Vega magnitudes to switch to the AB system.

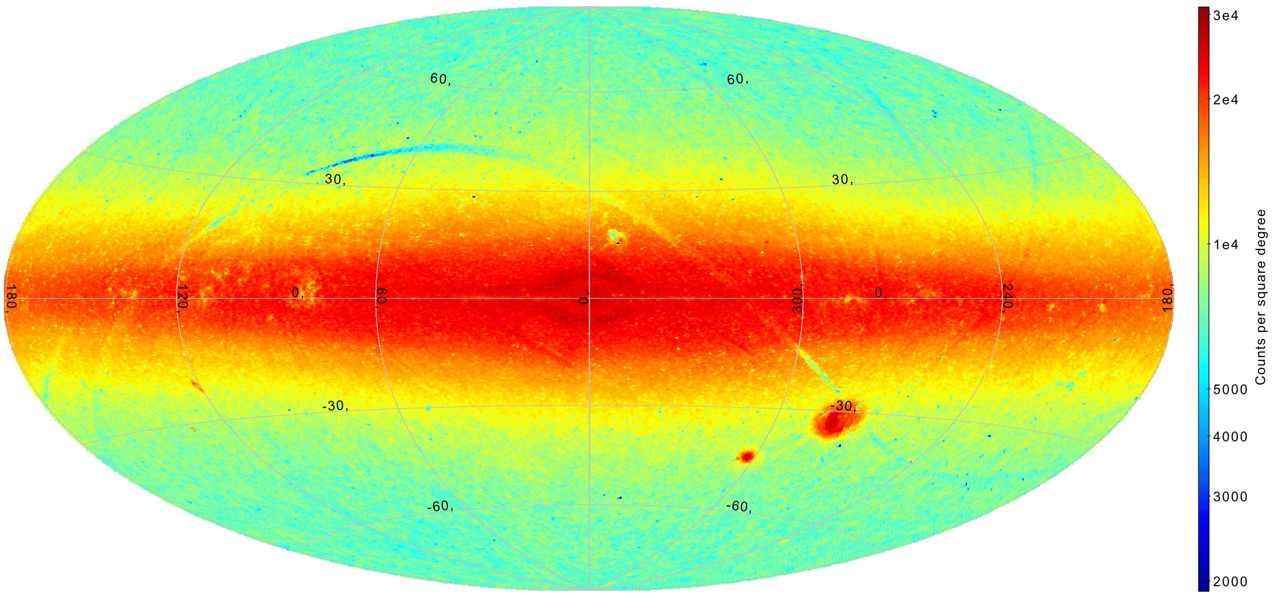


Figure 1. *WISE* all-sky Aitoff map, in Galactic coordinates, of 488 million sources preselected from AllWISE with $W1 < 17$, before cross-matching with SuperCOSMOS and purification. This sample contains both galaxies and stars, and the latter dominate at low latitudes. The missing data in a strip crossing the Galactic plane is due to saturation in $W1$ at the onset of the post-cryogenic phase and can be supplemented by using only data from the cryogenic stage in this region. The color bar shows counts per square degree at each pixel.

Table 1
Statistics of the Parent Photometric Catalogs and the Final *WISE* × SuperCOSMOS Cross-match Used in This Paper

Catalog	Flux Limit(s)	Sky Cut	# of Sources
WISE	none	none	604×10^6
(preselected in $W1$ and $W2$)	none	$\|b\| > 10°$	457×10^6
	$W1 < 17$	none	488×10^6
	$W1 < 17$	$\|b\| > 10°$	343×10^6
SuperCOSMOS XSC	none	none	288×10^6
(preselected in B and R)	none	$\|b\| > 10°$	158×10^6
	$B < 21$ and $R < 19.5$	none	208×10^6
	$B < 21$ and $R < 19.5$	$\|b\| > 10°$	85.1×10^6
WISE × SuperCOSMOS XSC	none	none	109×10^6
	none	$\|b\| > 10°$	78.3×10^6
	$W1 < 17$ and $B < 21$ and $R < 19.5$	none	77.9×10^6
	$W1 < 17$ and $B < 21$ and $R < 19.5$	$\|b\| > 10°$	47.7×10^6
after star and quasar cleanup	$13.8 < W1 < 17$ and $B < 21$ and $R < 19.5$	$\|b\| > 10°$ + Bulge, masked	18.8×10^6
galaxies in *WISE*, not in SCOS XSC	$W1 < 17$	$\|b\| > 10°$	$\sim 100 \times 10^6$

AllWISE Explanatory Supplement[23] and applies mostly to two strips within ecliptic longitudes of $44\overset{\circ}{.}7 < \lambda < 54\overset{\circ}{.}8$ or $230\overset{\circ}{.}9 < \lambda < 238\overset{\circ}{.}7$ (visible in Figure 1). This will be rectified in our final galaxy sample that is cross-matched with SuperCOSMOS by adding data from the earlier *WISE* data release, All-Sky (Cutri et al. 2012). Some other issues are caused by variable coverage due to moon avoidance maneuvers, which results in several under- or oversampled stripes crossing the Ecliptic.[24]

Galactic extinction corrections are very small in the *WISE* bands, over an order of magnitude smaller than in the optical, which does not mean they are totally negligible. Following Indebetouw et al. (2005) and Schlafly & Finkbeiner (2011), we use $A_{W1}/E(B-V) = 0.169$ and $A_{W2}/E(B-V) = 0.130$ as coefficients to be applied to the original Schlegel et al. (1998) maps; these values in part implement a general recalibration of the original $E(B-V)$ values, which need to be lowered by 14% (Schlafly & Finkbeiner 2011).

2.2. SuperCOSMOS

The SuperCOSMOS Sky Survey (SCOS, Hambly et al. 2001a, 2001b, 2001c) was a program of automated scanning and digitizing sky atlas photographic plates in three bands (B, R, I), using source material from the last decades of the 20th century obtained by the United Kingdom Schmidt Telescope (UKST) in the south and the Palomar Observatory Sky Survey-II (POSS-II) in the north. The data are stored in the SuperCOSMOS Science Archive,[25] with multicolor

---

[23] http://wise2.ipac.caltech.edu/docs/release/allwise/expsup/sec2_2.html#w1sat

[24] http://wise2.ipac.caltech.edu/docs/release/allwise/expsup/sec4_2.html#lowcoverage

[25] Available for download from http://surveys.roe.ac.uk/ssa/.

information provided for 1.9 billion sources, in the form of integrated quasi-total and point-source photometry (where available). The derived resolved-source data were accurately calibrated using SDSS photometry when possible, with the calibration extended over the full sky by matching plate overlaps and using the average color between the optical and 2MASS J bands as a constraint to prevent large-scale drifts in zero point (Francis & Peacock 2010; J. A. Peacock et al. 2016, in preparation). The typical resolution of SuperCOSMOS images is $\sim 2''$ (Hambly et al. 2001b) and the photometric depth is $R \simeq 19.5$, $B \simeq 21$ in a pseudo-AB system, in which SCOS and SDSS coincide for objects with the color of the primary SDSS standards; detailed color equations are given in J. A. Peacock et al. (2016, in preparation). The third band available from the catalog, I, offers shallower coverage and will not be used here.

For the present work, we are interested only in resolved images. SCOS supplies a classification flag for every image in each of the three bands, as well as a combined one, `meanClass`. These are equal to one if a source is non-stellar, two if it is consistent with unresolved, three if unclassifiable, and four if likely to be noise; the last two classes constitute a negligible fraction of all sources ($\ll 1\%$) in any given plate. The image classification is based on image morphology via a two-stage process (Hambly et al. 2001b, and references therein). The first stage uses image surface brightness, size, and shape to identify isolated, point-like images with good reliability and high completeness. The second stage takes this first-pass selection and analyses the 1D radial profile of those unresolved images as a function of plate position and source brightness. Finally, every image is assigned a profile statistic η—the probability distribution of which has zero mean and unit variance—to quantify the point-likeness. This continuously distributed statistic is used to define discrete classification codes when cut at fixed thresholds: sharper-than-point-like images with $\eta < -3$ are assigned class = 4 (noise), point-like images with $-3 < \eta < 2.5$ are assigned class = 2 (stellar), and resolved images having $\eta > 2.5$ are assigned class = 1 (non-stellar). Where data from two or more plates are available for the same image, the individual profile statistics are averaged to form a single, zero mean unit variance statistic via $\overline{\eta} = \Sigma\eta/\sqrt{N}$ for N plates with a discrete merged classification code, `meanClass,` assigned using the same ranges as above for the individual class codes. For brighter images with good detections in all bands, this increases the precision of classification, but for faint objects lacking good I-band data, this overall classification may be less reliable than the B or R plates individually. The data we use here have cuts that eliminate the faintest objects, so we use the `meanClass` parameter in all cases. We verified by comparison with SDSS test regions that this choice leads to better star-galaxy separation than using individual B and R classes.

Note that the classification flags also affect the photometric calibration procedure (Hambly et al. 2001b): separate calibrations were applied for stars and galaxies. This is mainly because of the limited dynamic range of SCOS when compared to, for example, some of the much slower modified PDS scanning machines or the highly optimized "flying spot" APM system (Hambly et al. 2001c and references therein). SCOS employed a linear CCD in the imaging system and a strip of emulsion was therefore illuminated to quickly scan lanes of ~ 1 cm width. When scanning over denser spots in an otherwise less dense emulsion, the core density measured was limited by light from the entire illuminated strip diffracting in the imaging lens. This was not the case for extended objects because the amount of light subject to diffraction was significantly reduced. The diffraction limit of the measurement process for stars occurred at much lower densities than any emulsion saturation in the photographic emulsions themselves. Hence the calibration curve of instrumental magnitude versus externally measured magnitude bifurcated into separate star and galaxy loci was only a magnitude or so above the plate limit, despite both the point and extended images being well exposed on the log–linear part of the photographic response curve. In any case, the galaxy calibration was performed at a later date (J. A. Peacock et al. 2016, in preparation), following the wider availability of SDSS photometry.

Because of a slight difference between the passbands of the UKST and POSS-II, there is in effect a small color-dependent offset in the SCOS magnitudes between the north and the south (here meaning above and below $\delta_{1950} = 2\overset{\circ}{.}5$). As discussed in B14, direct corrections were designed by comparison with SDSS to compensate for this effect. The following appropriate formulae (revised over B14) aim to correct the southern B and R data ($\delta_{1950} < 2\overset{\circ}{.}5$) to be consistent with the north[26]:

$$B_S^{\rm cal} = B + 0.03(B-R)^2 - 0.005(B-R), \tag{1}$$

$$R_S^{\rm cal} = R + 0.03(B-R)^2 - 0.06(B-R) + 0.015. \tag{2}$$

However, even these corrections may not fully guarantee N−S uniformity: within our fiducial flux limits, the mean high-latitude surface density in the north is up to 4% larger than in the south; it is hard to be sure whether this is a remaining very small calibration offset or a genuine cosmic variance. On the other hand, these offsets do not induce significant additional scatter to the corrected magnitudes. For typical galaxy colors, $B - R \sim 1$, by error propagation in Equations (1)–(2) we see that the random error in $B_S^{\rm cal}$ is increased by less than 6% with respect to the original values, while in $R_S^{\rm cal}$ there is a fortuitous cancellation and the error is not changed at all. For a general discussion of SCOS magnitude errors, see J. A. Peacock et al. (2016, in preparation).

We also revised the extinction corrections used in 2MPZ—a series of papers using SDSS (Schlafly et al. 2010; Schlafly & Finkbeiner 2011) and Pan-STARRS data (Schlafly et al. 2014) show that the original Schlegel et al. (1998) maps overestimate the $E(B-V)$ values by roughly 14%, and that one should use the Fitzpatrick (1999) reddening coefficients rather than the Cardelli et al. (1989) ones. Based on the revised extinction coefficients for the SDSS g and r bands (Schlafly & Finkbeiner 2011), the new corrections for the B and R SCOS bands are, respectively, $A_B/E(B-V) = 3.44$ and $A_R/E(B-V) = 2.23$ (J. A. Peacock et al. 2016, in preparation) for the full sky.[27] These numbers already incorporate the rescaling of the $E(B-V)$ values by Schlafly & Finkbeiner (2011); they should thus be applied to the original Schlegel et al. (1998) $E(B-V)$ to obtain band-dependent

---

[26] Unfortunately, the corresponding equations in B14 (Equations (1)–(2) therein) are incorrect, owing to an inadvertent swapping of north and south. A revised version of the 2MPZ catalog that incorporates this correction will be issued.

[27] Note that in B14 we incorrectly provided different extinction corrections for the two hemispheres; because the magnitudes are already calibrated N–S, one should use a single coefficient (N) in a given band for the full sky.

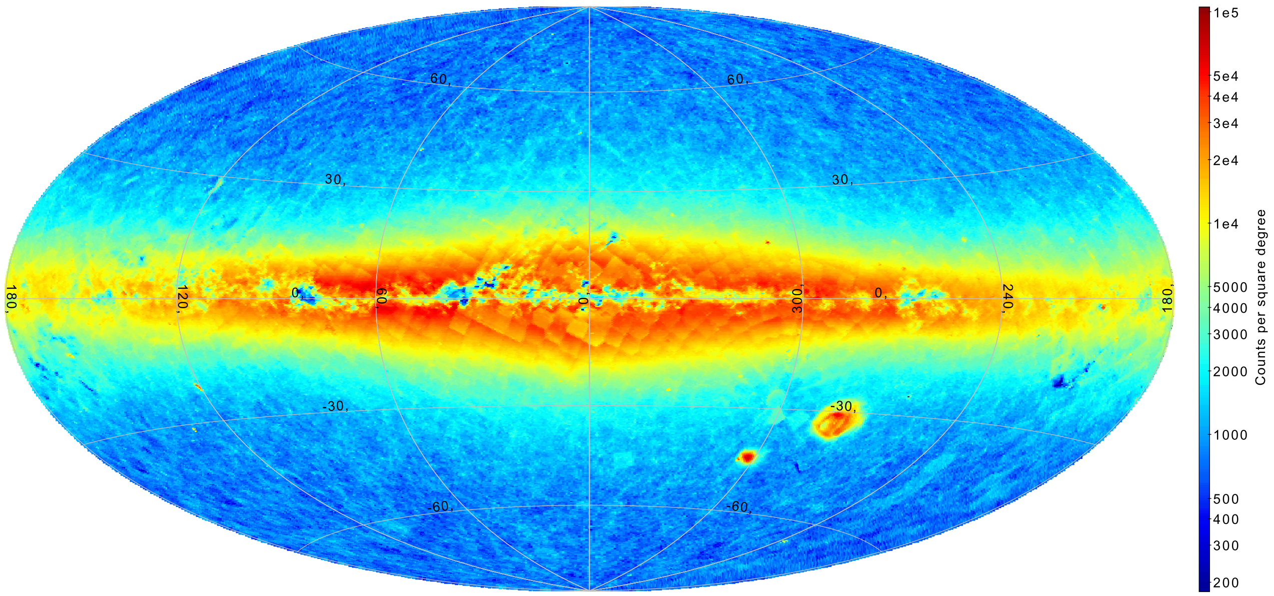


Figure 2. SuperCOSMOS all-sky Aitoff map, in Galactic coordinates, of 208 million extended sources preselected with $B < 21$ and $R < 19.5$, before the cross-match with *WISE* and purification. The spurious overdensities in the Galactic Plane and at the Magellanic Clouds arise due to star blending. The color bar shows counts per square degree at each pixel.

extinction corrections for a given galaxy in magnitudes. In what follows, all the quoted SCOS magnitudes refer to hemisphere-calibrated and extinction-corrected values in the AB-like system.

For the purposes of the present work, our requirements for SCOS preselection were that the sources be properly detected with aperture photometry in B and R bands: `gCorMagB` and `gCorMagR2` are not null in the database, quality flags `qualB` and `qualR2` < 2048 (no strong warnings nor severe defects: Hambly et al. 2001b). In addition, as described above, we used the sources with SCOS morphological classification flag `meanClass` $= 1$. This selection greatly enhances the *purity* of our final cross-matched sample by eliminating most of the stars from unconfused regions, as well as many quasars (see further discussion on these issues in Section 4). On the other hand, it only slightly reduces the *completeness* of the catalog, removing less than 1% of galaxies, which we estimated based on GAMA and SDSS galaxies cross-matched to our data. As with *WISE*, for SCOS we will also not be using low-latitude sources in the present work (almost 50% of SCOS "extended" sources are in the $|b| < 10°$ strip—mostly blends of stars). On the other hand, we have supplemented our catalog over what is publicly available by adding sources that were originally omitted from the SCOS catalogs due to areas excluded around stepwedges, which affected mostly plate corners (564,000 objects in our case).

The above selections in SCOS resulted in the "SCOS extended source catalog" (XSC), with about 158 million sources at $|b| > 10°$. Owing to the remaining low-latitude stellar blends, only part of these sources are actually extragalactic. Simple cross-matching of this catalog with AllWISE would give a highly contaminated sample, therefore extra effort was needed to derive the best possible purity and completeness criteria for our catalog. This is discussed in Section 4.

As far as reliability is concerned, the main limitation here and for the cross-matched catalog is the depth of the SCOS data. We decided to adopt $B < 21$ and $R < 19.5$ as the optical limits, motivated by our analysis of galaxy counts from direct comparison with very deep SDSS photometric data (Ahn et al. 2014; see also J. A. Peacock et al. 2016, in preparation). Applying these magnitude cuts to the $|b| > 10°$ sample removes almost 50% of the SCOS XSC there, leaving 85 million sources. Had we included the Galactic Plane data, the flux-limited sample would include almost 208 million objects (see Table 1). Their distribution is shown in Figure 2; in addition to the Galactic Plane, the Magellanic Clouds are also clearly dominated by spurious overdensities from star blends. The plate pattern is noticeable at low latitudes because the degree of blending varies with plate quality. Note also that the dynamic range of the counts is much wider than in the case of *WISE*.

2.3. *WISE* × *SuperCOSMOS Cross-match*

In the following, all the cross-matches will be performed within a radius of $2''$, unless otherwise specified.[28] In the case of the *WISE* × SCOS cross-match, the radius is motivated by the large beam of the former ($\sim 6''$ in the $W1$ band; Wright et al. 2010) and the angular resolution of the latter ($\sim 2''$). The mean matching radius for the resolved *WISE* × SCOS sources that pair up is $0''.54 \pm 0''.42$, and less than 14% of the cross-matched sources are separated by more than $1''$. It is important to note that both surveys offer comparable, sub-arcsecond astrometric accuracy: $\lesssim 0''.15$ for *WISE* (Wright et al. 2010) and $\lesssim 0''.3$ for SCOS (Hambly et al. 2001a). Thus, it is highly unlikely for a source identified in the two catalogs and detected in the four bands used here to be spurious.

As already mentioned, all the *WISE*-based magnitudes are in the Vega system, while the SCOS ones are AB-like. We will also keep this convention for source colors derived from the two catalogs. From this point on, all magnitudes are corrected for extinction as described earlier.

After selecting the AllWISE and SCOS objects as discussed above, the resulting cross-match at $|b| > 10°$ gave us more tha

---

[28] Catalog cross-matching was done using the TOPCAT/STILTS software (Taylor 2005, 2006) available for download from http://www.star.bristol.ac.uk/~mbt/.

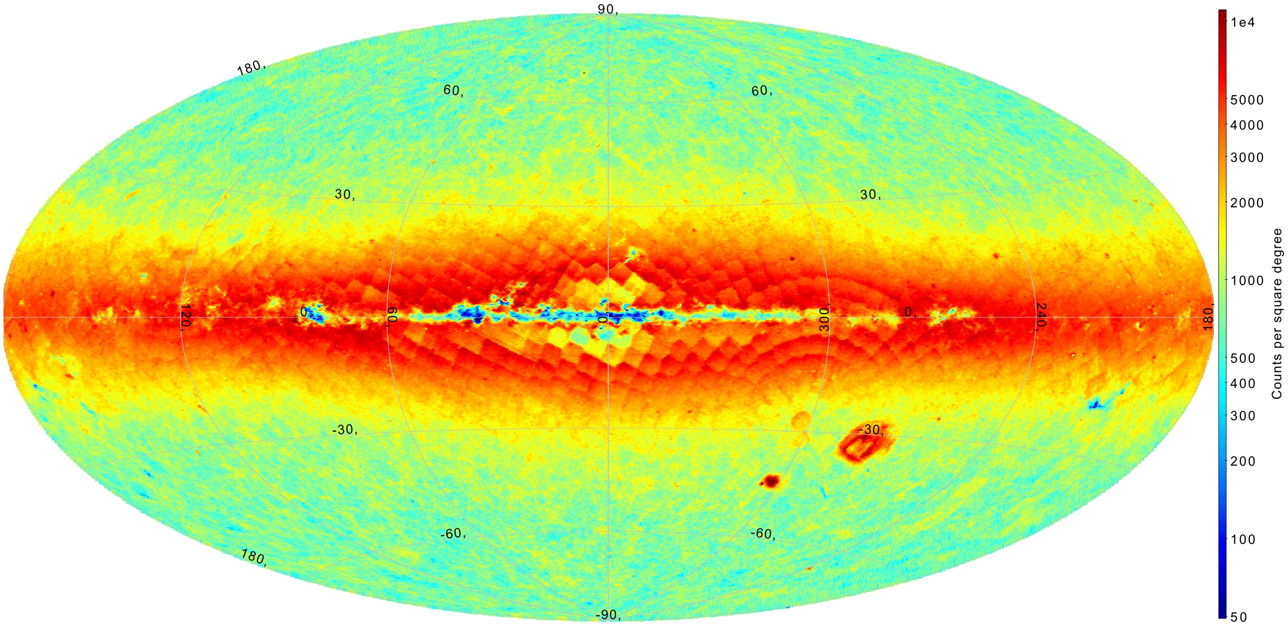


Figure 3. *WISE* × SuperCOSMOS cross-matched catalog of extended sources, before purification of stars and masking, in an all-sky Aitoff map in Galactic coordinates. The map contains 78 million objects flux-limited to $B < 21$, $R < 19.5$, and $W1 < 17$. Low latitudes and Magellanic Clouds are dominated by star blends mimicking extended sources. The color bar shows counts per square degree at each pixel.

78 million sources if no flux limits were applied, of these almost 48 million were within (extinction-corrected) magnitude cuts of $W1 < 17$, $B < 21$, and $R < 19.5$ (Table 1). These numbers include sources that were added to the sample from the earlier *WISE* release (“All-Sky”) to remove the incompleteness in AllWISE data visible as undersampled strips in Figure 1 and discussed in Section 2.1, as well as the SCOS objects lost through stepwedge exclusion. Figure 3 shows the sky distribution of this flux-limited sample. One expects the angular distribution of extended (extragalactic) sources to be relatively uniform on the sphere, whereas here the foreground Milky Way clearly dominates the counts at low latitudes and the Magellanic Clouds do the same at their respective positions. Although the contamination from stellar blends is much reduced with respect to the two parent catalogs considered individually, less than half of these sources are actually extragalactic, despite being classified by SCOS as *extended*.

To purify this sample, in Section 4 we present color cuts aimed at removing some problematic quasars (Section 4.1) and the remaining stars (Section 4.2). In Section 4.3 we describe the mask that needs to be applied to the data to remove regions where the stellar and other contaminations cannot be corrected. However, in Section 3 we first analyze the properties of the photometric catalogs used here by pairing them up with the GAMA spectroscopic sample. Table 1 summarizes the surveys contributing to our sample for different flux and sky cuts, including the cross-match after the removal of stars and quasars as described later in Section 4.

3. PROPERTIES OF THE INPUT PHOTOMETRIC CATALOGS: CROSS-MATCH WITH GAMA

To explore the properties of our input catalogs, we cross-matched them with the Galaxy And Mass Assembly (GAMA) data covering three equatorial fields. GAMA (Driver et al. 2009) is an ongoing multiwavelength spectroscopic survey of the low-redshift universe: its input catalog (including star and quasar removal) is discussed in Baldry et al. (2010), the tiling strategy is described in Robotham et al. (2010), and the spectroscopic pipeline is explained in Hopkins et al. (2013). Baldry et al. (2014) present a fully automatic redshift code (AUTOZ) developed to homogenize the redshift measurements and improve their reliability, and Liske et al. (2015) discuss the accuracy of these new measurements in context. The data set we use here, taken from GAMA-II (`TilingCat v43`, not publicly released yet), covers three GAMA Equatorial Regions (G09, G12, and G15) centered on 9, 12, and 14.5 hr in right ascension, respectively. Each of these fields spans across $5° \times 12°$, which gives $180\,\mathrm{deg}^2$ in total. This sample is preselected in the SDSS Petrosian r magnitude, and within the limit of $r_{\mathrm{Petro}} \leqslant 19.8$ its galaxy redshift completeness is 98.4% (Liske et al. 2015). This makes the catalog ideal for our purposes, because it is deeper and more complete in the fields it covers than our core flux-limited *WISE* × SCOS sample, and at the same time it is free from stellar and quasar contamination by construction. GAMA is also unique in comparison to other surveys because it offers a plethora of ancillary data and parameters derived by the team. Some of the intrinsic properties of galaxies presented by Taylor et al. (2011) and more recently by Cluver et al. (2014) are particularly useful. The latter paper focused on sources common to GAMA and *WISE* in the equatorial fields.

The GAMA-II sample we use includes almost 203,000 sources with redshift measurements (some fainter than $r = 19.8$). Of these, we have preselected confirmed galaxies ($z > 0.002$) with reliable redshifts (quality $\mathtt{NQ} \geqslant 3$). This gave us more than 193,500 sources with $z_{\mathrm{med}} = 0.22$; their redshift distribution is presented in Figure 4 (red line). This plot displays a dip at $z \simeq 0.23$, which is observed in all three equatorial fields at roughly the same redshift. We interpret this as a coincidence in cosmic variance, as the three areas are too widely separated to trace the same large-scale structures. In fact, it is a projection effect that is mostly due to filaments and walls present in the three fields at $z \sim 0.2$ and $z \sim 0.26$, as can

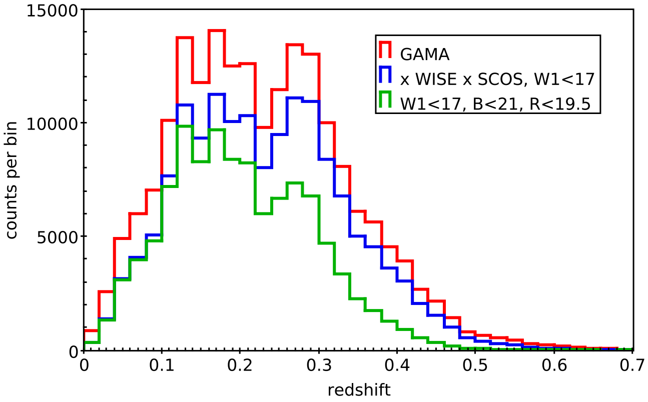


Figure 4. Redshift distributions of GAMA (red line) and of its cross-matches with the *WISE* × SuperCOSMOS extended source catalog. Two flux limits for the cross-matches are shown: *WISE*-based only (blue line) and *WISE*+optical (green line).

Table 2
Properties of Photometric Surveys in the GAMA Equatorial Fields and of their Cross-Matches with GAMA

Sample	Flux Limit(s)	# of Sources	$z_{\rm med}$
GAMA-II	none[a]	193,500[b]	0.23
	$r \leqslant 19.8$	183,000[b]	0.22
WISE[c]	none	2,000,000	N/A
WISE × GAMA	none	167,000	0.23
GAMA but not *WISE*	none	26,500	0.17
SCOS XSC[c]	none	484,000	N/A
	$B < 21$ and $R < 19.5$	183,000	N/A
SCOS × GAMA	$r \leqslant 19.8$ (GAMA)	167,000	0.21
	$B < 21$ and $R < 19.5$	117,000	0.19
GAMA but not SCOS	$r \leqslant 19.8$ (GAMA)	16,000	0.26
WISE × SCOS XSC[c]	$W1 < 17$	294,000	N/A
	$W1 < 17$ and $B < 21$ and $R < 19.5$	151,000	N/A
WISE × SCOS × GAMA	$W1 < 17$	153,000	0.22
	$W1 < 17$ and $B < 21$ and $R < 19.5$	109,000	0.19

Notes.
[a] Most of the sources are within the flux limit of $r \leqslant 19.8$.
[b] Preselected with $z > 0.002$ and `NQ` $\geqslant 3$.
[c] In the GAMA equatorial fields.

be seen in cone plots of Eardley et al. (2015), where environmental classification is also provided. In addition, this pattern is not observed in the southern GAMA fields (G02 and G23) for which the spectroscopy was processed in the same way as for the equatorial ones, so it cannot reflect an error in the redshift determination (cf. footnote #10 in Liske et al. 2015). The two additional fields available from GAMA-II are significantly less complete than the equatorial ones (Liske et al. 2015) and will not be used in this part of the present work; we will however employ them for photometric redshift quality tests discussed in Section 5.3.

A detailed analysis of *WISE* sources common with GAMA was presented in Cluver et al. (2014). Two of the three equatorial fields were studied there, and the *WISE* data originated from the earlier, "All-Sky" release (Cutri et al. 2012). Cluver et al. (2014) analyzed mid-infrared properties of GAMA galaxies, paying particular attention to characterizing and measuring resolved *WISE* sources. Many other issues were explored therein, particularly the empirical relations between optically determined stellar mass and the $W1$ and $W2$ measurements (using the synthetic stellar population models of Taylor et al. 2011).

In the present work, we use the updated AllWISE release as well as the complete information in the three GAMA equatorial fields. Out of more than 2 million AllWISE sources (of any kind) in these areas, our cross-match with GAMA gives almost 167,000 objects, which constitutes 86% of the GAMA galaxy sample (see Table 2 for these and other details). This is a similar percentage to the one reported by Cluver et al. (2014), where a larger matching radius ($3''$) was used. The GAMA sources with no AllWISE counterparts are mostly faint and at lower redshifts ($z_{\rm med} = 0.23, 0.17$, respectively, for the matches and non-matches); that is, they are more local low-luminosity galaxies. Some of the non-matches arise due to *WISE* blending GAMA galaxies at smaller angular separations than the beam of the former (T. H. Jarrett et al. 2016, in preparation).

The source density of AllWISE is some 10 times that of GAMA, and objects that are in AllWISE and not in GAMA belong to two general classes: either mostly bright, having colors consistent with stellar ones (e.g., $W1 - W2 \lesssim 0$)—stars filtered out by GAMA preselection—or those at the the faint end ($W1 > 16$), where galaxies dominate over stars (Jarrett et al. 2011; T. H. Jarrett et al. 2016, in preparation), with colors typical for an extragalactic population. Some are also quasars, which were eliminated from GAMA via morphological and color preselections (Baldry et al. 2010). All this leads to the conclusion that a significant fraction of the unmatched AllWISE sources will be galaxies too faint for GAMA, and that the $z_{\rm med}$ of the former should be significantly larger than that of the latter. This is further supported by the results of T. H. Jarrett et al. (2016, in preparation), which shows that the *WISE* × GAMA cross-match becomes incomplete for *WISE* galaxies that are fainter than $W1 = 15$ (0.3 mJy).

Next, we paired up the GAMA galaxy sample with the SCOS XSC. Here we used only the $r_{\rm Petro} \leqslant 19.8$ GAMA galaxies (183,000 with $z > 0.002$ and `NQ` $\geqslant 3$) so as to have a complete and unbiased sample. Not applying any flux limit on SCOS gave 9% of GAMA without SCOS counterparts. The unmatched GAMA sources were mostly at high redshifts, with $z_{\rm med} = 0.26$ (see Table 2), in contrast to the AllWISE case—confirming that the GAMA data are deeper than SCOS. The SCOS magnitudes for the fainter GAMA galaxies have a substantial random error; thus, one would need to go to SCOS $R \lesssim 21$ to capture most of the true $r \leqslant 19.8$ GAMA objects, which is beyond its reliability limit (J. A. Peacock et al. 2016, in preparation). Flux-limiting the SCOS sample to our fiducial values of $R < 19.5$ and $B < 21$ resulted in 64% of GAMA galaxies that were also found in the photographic data, with $z_{\rm med} = 0.19$ for the matched sources and more than 90% of unmatched GAMA galaxies having $r_{\rm Petro} > 19.2$.

Finally, we analyzed the *WISE* × SCOS cross-match in the three equatorial GAMA fields, focusing on the sources of interest for the present work, namely those resolved by SCOS. Out of 484,000 SCOS `meanClass` = 1 sources in these areas, roughly 294,000 (61%) had counterparts in AllWISE $W1 < 17$ if no magnitude cuts were applied to SCOS data. If we preselect SCOS as $R < 19.5$, $B < 21$, we end up with more than 150,000 *WISE* × SCOS XSC objects, which is 83% of the flux-limited extended SCOS sources. Of these two *WISE* × SCOS samples (with no SCOS magnitude limit and the flux-limited one), respectively, 51% and 71% have GAMA counterparts. If no SCOS flux limit is applied, the median redshift of the *WISE* × SCOS × GAMA sample is $z_{\rm med} = 0.22$ and decreases to $z_{\rm med} = 0.19$ if only the $R < 19.5$, $B < 21$ sources are used; see Figure 4 for relevant redshift distributions and Table 2 for a summary. The sources present in the flux-limited *WISE* × SCOS resolved sample and not identified among GAMA galaxies are mostly bright and have colors (especially $W1 - W2$) that are consistent with Milky Way stars, which illustrates the aforementioned fact that SCOS morphological classification is prone to misidentifying stellar blends as extended sources.

The analysis of this Section has confirmed that the present GAMA data are appropriate for the photometric redshift training of the wide-angle ("all-sky") *WISE* × SCOS catalog that we aim to produce. On the other hand, as is visible in Figure 4, we cannot hope to reach beyond $z \simeq 0.45$ with our present sample due to the depth of the SCOS data; but as shown in T. H. Jarrett et al. (2016, in preparation) *WISE* alone with no optical limit reaches up to $z \sim 1$. We plan to explore the latter property in future work.

4. PURIFYING THE WISE × SUPERCOSMOS GALAXY CATALOG

Despite preselecting the sources as extended in SCOS, our catalog will be contaminated with blended stellar images that masquerade as galaxies; this problem also affects *WISE*, and becomes more pronounced as we approach the Galactic plane. In addition, a number of high-z quasars projected on more local galaxies will be present in the all-sky data set, thus contaminating the colors of the galaxies. In this Section we propose relatively simple cuts to clean our data of this quasar and stellar contamination. To these one should also add an angular mask based on, for instance, Galactic extinction and star density, as well as encompassing such objects as the Magellanic Clouds, other large nearby galaxies, or very bright stars. We discuss such a mask in Section 4.3. A separate work (T. Krakowski et al. 2016, in preparation) will be devoted to another, machine-learning-based attempt at all-sky galaxy selection from the *WISE* × SCOS catalog.

As already mentioned, because GAMA includes practically no stars or quasars it cannot be used as a calibration set to identify them in our *WISE* × SCOS sample. We have thus employed Sloan Digital Sky Survey (SDSS; Eisenstein et al. 2011) spectroscopic data from Data Release 12 (DR12; Alam et al. 2015) for the purpose of star and quasar cleanup. At the moment, SDSS is the most appropriate deep and wide-angle data set that contains stars, galaxies, and quasars comprehensively identified based on their spectral properties (Bolton et al. 2012). SDSS assigns a class to spectroscopic sources while deriving their redshift (velocity),[29] which ensures far better reliability of this procedure over the photometric-only (morphological) classification. Thus, properly cleaned SDSS spectroscopic data form the best calibration sample for star-galaxy-quasar identification in wide-angle $z \lesssim 0.5$ photometric catalogs such as ours. The trade-off is the limited and variable depth of the spectroscopic sample, which is not as uniformly selected in SDSS as the photometric data.

The full SDSS DR12 spectroscopic catalog, which encompasses earlier releases (properly recalibrated where necessary), contains almost 3.9 million sources, of which 61% are classified as galaxies, 16% as quasars, and the remaining 23% as stars. However, not all of these objects have sufficient classification and redshift quality for our purposes. To maintain reliability, we cleaned this sample of `zWarning` $\neq 0$ sources (problematic redshifts), as well as those without a redshift error estimate ($\Delta z < 0$) or with low-accuracy redshifts ($\Delta z / z > 0.01$). This gave as more than 2.6M sources listed in SDSS DR12 as extragalactic (galaxies+quasars), including both the "Legacy" (Abazajian et al. 2009) and "BOSS" (Dawson et al. 2013) samples, along with 750,000 stars.

4.1. Quasar Removal

Our core data set of extended objects is expected to contain a number of AGN and quasars. These will occasionally be outliers in the size distribution, which are much rarer than stars, as well as blends. Low-luminosity, relatively low-redshift, morphologically extended AGN, which dominate the quasar population in our sample, are acceptable as long as their redshifts can be reliably reproduced photometrically. However, blends of a high-redshift quasar with a foreground star, which mimic extended sources and have peculiar colors, as well as quasar-galaxy projections that also can have compromised colors, will be problematic for the photo-z procedure. Such blends lying at high redshifts should preferably be removed from the catalog before the photometric redshift estimation, because their presence may contaminate the derived galaxy sample that is expected to reach up to $z \sim 0.5$. In what follows, we often use the terms "AGN" and "quasar" interchangeably.

Most of the quasars from the *WISE* × SCOS sample were eliminated through the morphological preselection of resolved SCOS sources, as well as through the flux limits in the optical and infrared bands: high-redshift quasars are typically fainter than low-redshift galaxies in terms of their apparent magnitudes. There are, however, some quasars bright enough to be captured in the sample, while still classified as extended: about 30% of the Sloan quasars/AGN (i.e., `CLASS=QSO` in SDSS) identified in our flux-limited sample have SCOS `meanClass` $= 1$ (30,000 sources). These are mostly at redshifts of $z < 0.6$, but some reach up to $z > 3.5$. The latter are blends of background point-like quasars with a low-redshift foreground galaxy or with a foreground Galactic star, and might be problematic for photometric redshift estimation, regardless of the method used to obtain the photo-zs. AGNs experiencing significant dust obscuration, such as type 2 AGNs, where the accretion disk and the broad-line region are completely obscured, have colors similar to galaxies. In broad-band photometry, quasars at $z \gtrsim 2.3$ can be mistaken for low-redshift galaxies because the Lyα spectral break can mimic the 4000 Å break. Additionally, at $z \sim 2.7$ and ~3.5, the optical colors of broad-line, unreddened quasars and Galactic stars are the same (Richards et al. 2006).

To remove as many of the remaining quasars in our sample as possible, we analyzed their multicolor properties, based on the SDSS spectroscopic data cross-matched with our catalog. Stern et al. (2012) proposed $W1 - W2 > 0.8$ as an efficient AGN finder in *WISE*, which was subsequently used in several other studies to select quasars (e.g., DiPompeo et al. 2014; Donoso et al. 2014). By looking additionally at GAMA sources (which are practically free of quasar contamination), we slightly revised this cut to preserve the completeness of the galaxy sample, and added a second criterion using the optical R

---

[29] http://www.sdss3.org/dr12/algorithms/redshifts.php

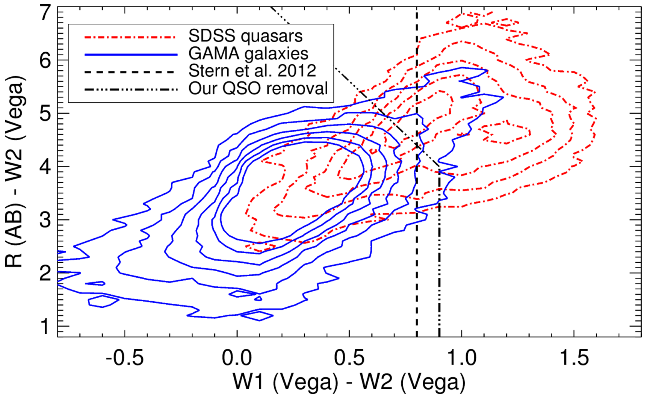


Figure 5. Color–color plot ($W1 - W2$ vs. $R - W2$) for GAMA galaxies (blue solid contours) and SDSS quasars (red dotted–dashed) classified as extended in *WISE* × SCOS, together with the cuts that are used to remove quasars. The quasars in this plot are either low-redshift AGN or blends of a high-z quasar with a star or galaxy. The contours are linearly spaced.

band (see Figure 5). Our color cuts to remove quasars are:

$$R - W2 > 7.6 - 4(W1 - W2) \ \text{or} \ W1 - W2 > 0.9, \quad (3)$$

where the *WISE* magnitudes are Vega and the SCOS one AB-like. These criteria remove 71% of SDSS quasars present in our extended source sample, while affecting less than 1% of GAMA galaxies. Nearly all the quasars with $0.5 < z < 2.1$ are eliminated through this cut; those that remain are mostly at low redshifts (peaking at $z \sim 0.2$), with some at $2.1 < z < 3.5$. An alternative way of selecting quasars by using only *WISE* information through a comparison of $W1 - W2$ and $W2 - W3$ colors (Jarrett et al. 2011; Mateos et al. 2012) cannot be applied here because the detection rate of our sources in the $W3$ band is too low.

The cut defined in Equation (3) removed nearly 300,000 quasar candidates from our flux-limited, $|b| > 10°$ photometric sample of *WISE* × SCOS extended sources. Rescaling from the SDSS-based numbers, we thus estimate that there are about 115,000 quasars remaining in the all-sky catalog, which is about 0.6% of the total number of galaxies. Thus our photo-zs derived in Section 5 will only be minimally affected by the high-z quasars that were not filtered out.

4.2. Star Removal

We also paired the SDSS DR12 stars with reliable spectra (`zWarning` = 0 and $0 < \Delta z < 0.001$) against our core sample, and used the result to derive typical stellar colors for star removal. Thanks to the morphological information from SCOS (`meanClass` = 1 only sources), many of the stars had already been eliminated and only 8% of those common to the Sloan spectroscopic and *WISE* × SCOS are present in our sample. These “stars” in the catalog of extended sources are expected to be blends, which might be the reason why the separations from their SDSS counterparts are usually larger than in the case of extragalactic sources: $0\farcs31 \pm 0\farcs26$ for SDSS galaxies, $0\farcs30 \pm 0\farcs29$ for quasars, but $0\farcs40 \pm 0\farcs24$ for stars. To avoid mismatches when deriving the color cuts for star removal, we used only the stars paired up within $1''$ with our photometric catalog. Note that poorer matching accuracy for the stars might also be partly due to proper motions between the epochs of SCOS photographic material and these of the SDSS (see Madsen & Gaensler 2013 for a related discussion).

To remove stellar contamination we examined the source distribution presented in Figure 3, which clearly shows spurious overdensities (caused by stellar blends) at low Galactic latitudes and at the Magellanic Clouds. We began by rejecting by hand regions in the Galactic Plane and Bulge where the contamination was too severe to contemplate reliable correction (an enhancement in surface density by a factor ~10). We applied a latitude cut depending on the distance from the Galactic Center (GC): it goes smoothly from $|b| < 17°$ at $\ell = 0°$ to $|b| < 10°$ at $\ell \sim 80°$ or $\ell \sim 280°$. Detailed equations are provided in the Appendix. This cut removed almost 6 million sources from the flux-limited sample at $|b| > 10°$. To this we added circular cutouts around the most prominent nearby galaxies, namely the Magellanic Clouds and M31.

We then investigated what other cuts need to be taken to purify the sample further. This is traditionally done in multicolor space, and we explored different combinations of the available bands based on the cross-match with SDSS spectroscopy. Stars are much more difficult to remove than quasars from our catalog without seriously compromising the completeness of the galaxy sample; this is due to blends of stars with other stars and with galaxies, especially at low redshift, where galaxies from our data set often have colors that are similar to stellar ones. In particular, the SCOS optical bands were found not to be useful for star identification. We were left with the option to use only *WISE* colors for star-galaxy separation, as had been discussed in earlier studies (Jarrett et al. 2011; Goto et al. 2012; Yan et al. 2013; Ferraro et al. 2015). An advantage of applying infrared-only cuts to our sample is less sensitivity to variations in plate zero points, or in extinction corrections and their errors.

The colors usually considered for *WISE* source identification are $W1 - W2$ and $W2 - W3$, and the former is particularly useful for this task (Jarrett et al. 2011). We found that using $W2 - W3$ does not add much information, mostly due to the low level of signal-to-noise in the $W3$ band, and a similar effect is observed in the automatic galaxy identification of T. Krakowski et al. (2016, in preparation). In a related effort, Ferraro et al. (2015) treated as stellar anything with $W1 - W2 < 0$ or ($W1 < 10.5$ and $W2 - W3 < 1.5$ and $W1 - W2 < 0.4$). However, once these conditions are applied to *WISE*, they leave a certain degree of contamination that is dependent on the distance from the GC—see Figure 1 in Ferraro et al. (2015). The same is found in our *WISE* × SCOS catalog, namely a fixed $W1 - W2$ color cut would give purity levels largely varying over the sky; this is also expected because we are using extinction-corrected magnitudes, so effective stellar colors will be correlated with the $E(B - V)$ map.

To account for star contamination changing with Galactic coordinates, we examined the source density and the $W1 - W2$ color as a function of distance from the GC and found that the stellar locus shifts as the GC is approached, which we interpret as a reflection of older stellar populations being located toward the Bulge. This lead us to design a position-dependent color cut, the details of which are provided in the Appendix. In brief, at high latitudes we remove sources with $W1 - W2 < 0$, while this cut is gradually shifted toward $W1 - W2 < 0.12$ closer to the GC. This adaptive star removal, together with the sky cuts discussed earlier, eliminated more than half the sample, mostly from low Galactic latitudes, as expected (90% of removed

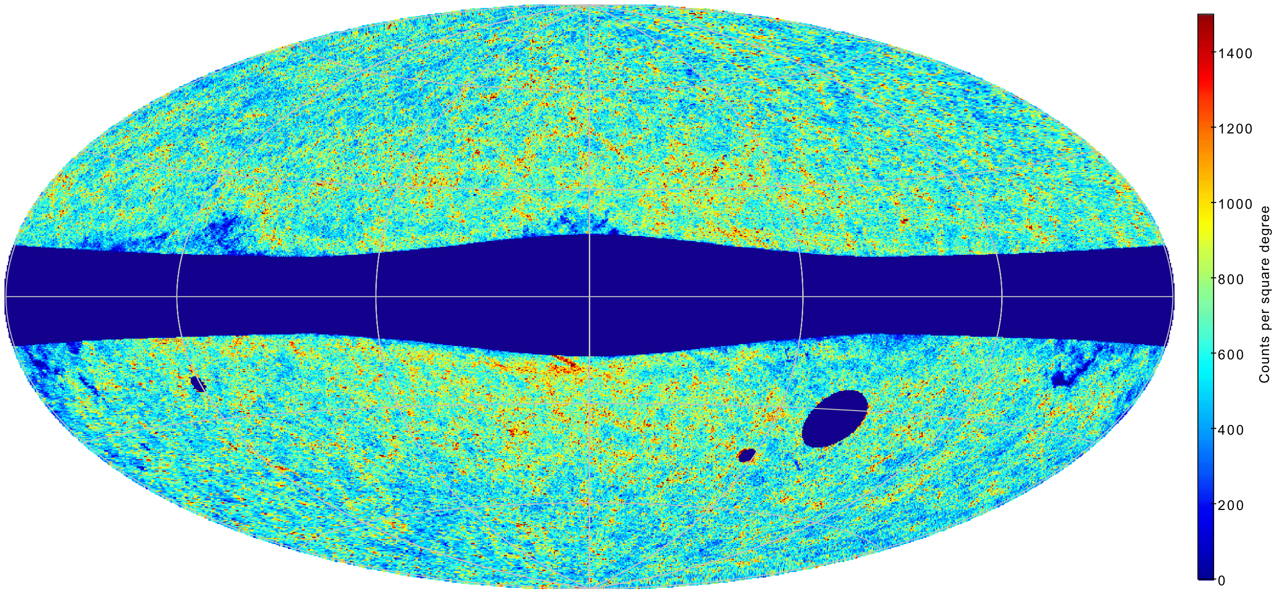


Figure 6. *WISE* × SuperCOSMOS galaxy catalog after star and quasar cleanup and manual cutouts of the Galaxy, Magellanic Clouds, and M31, but before final masking, in an all-sky Aitoff map in Galactic coordinates. The map contains 21.5 million sources flux-limited to $B < 21$, $R < 19.5$, and $13.8 < W1 < 17$.

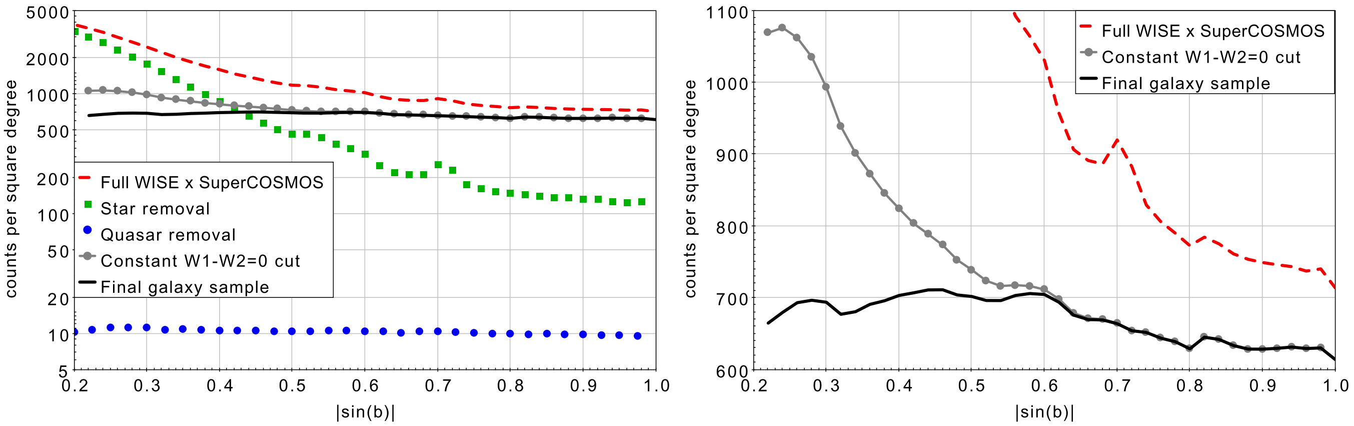


Figure 7. Source counts per square degree as a function of the sine of the Galactic latitude in the cross-matched *WISE* × SuperCOSMOS extended source catalog: full sample (red dashed), sources removed with our star (green squares) and quasar (blue dots) cleanup, and the final sample (black solid). For comparison, we also show the counts for a sample with a constant $W1 - W2$ color cut applied (gray solid-dotted). The right panel shows a zoom in on the two latter curves, in linear scaling.

sources are within $|b| < 34°$). This approach slightly degrades the completeness of the final galaxy sample: almost 6% of *WISE* × SCOS × GAMA galaxies are removed with this cut. On the other hand, a completeness level of ~90% in the final sample is preserved for $|b| \gtrsim 15°$, which gives almost 3π sr of the extragalactic sky comprehensively sampled with the catalog. A more detailed discussion of completeness and purity of the galaxy data set is provided in Section 4.4.

In addition, we removed the bright end of our sample ($W1 < 13.8$) for two main reasons. First, the galaxies that have counterparts in the 2MASS XSC $K_s < 13.9$ already have precise photometric redshifts derived in the 2MPZ (B14). At low redshifts the typical galaxy color is $K_s - W1 \simeq 0$, so most of these 2MASS sources are removed by applying this bright-end cut in *WISE*. Second, most of the bright *WISE* sources that are *not* present in 2MASS XSC are stars, because they dominate $W1$ number counts there (Jarrett et al. 2011) and are concentrated toward the Galactic plane. There were more than 5 million objects with $W1 < 13.8$ in the cross-matched catalog before the cleanup, of which 90% lay at $|b| < 50°$.

Figure 6 shows the all-sky distribution of our sources after the purification and manual cutouts but before final masking, which is addressed in the following Section. In Figure 7 we show source counts per square degree, as a function of the sine of Galactic latitude b, for the cross-matched sample: before and after the star and quasar cleanup, as well as for the sources removed with our cuts. A uniformly distributed (extragalactic) sample should have roughly constant counts in this scaling, which is approximately true for the final data set, as well as for the quasars removed. The bump at $|\sin b| \simeq 0.7$ in the removed sources is the LMC. For comparison, we also show the case of a constant $W1 - W2 > 0$ cut as in Ferraro et al. (2015). Stellar contamination becomes prominent from $|\sin b| = 0.5$ ($|b| = 30°$), that is, for half the sky, and the surface density of the sources close to the Galactic Plane is almost twice as large as in the Caps, as visualised in the right panel of Figure 7.

4.3. Final Mask

The above cuts helped improve the fidelity of the catalog, reducing the numbers of non-galaxy entries resulting from stellar blends and other problems. Nevertheless, a casual inspection of the sky distribution reveals clear imperfections, especially at low Galactic latitudes; Figure 6 exhibits some spurious source overdensities and a lack of extragalactic data behind Galactic molecular clouds such as Orion, Taurus/Perseus, and Ophiuchus. We thus need to develop a mask that excludes significantly affected regions. This is a common task, but not a trivial one: the human eye is highly adept at spotting artifacts of this sort, and it takes some effort to design an objective, automated process that performs as well. As a starting point, one can identify pixels where the surface density is discrepant, using the fact that the galaxy surface density very nearly obeys a log-normal distribution (Hubble 1934) and clipping pixels in the tails of this distribution. This approach can be made more effective if we perform it at a variety of resolutions: large-scale regions where the density is systematically slightly in error can be found more sensitively by using coarse pixels where the pixel-to-pixel variance is reduced. We therefore constructed a HEALPix[30] (Górski et al. 2005) map of the galaxy counts, initially at $N_{\rm side} = 256$, identified discrepant pixels, and then repeated the process degrading the resolution by successive factors of two. The final mask is the accumulation of flagged sky areas at all resolutions. However, this process requires an unsatisfactory compromise: in order to remove all apparent artifacts, the clipping threshold has to be set at a rather high probability ($p(\delta) \sim 0.001$), with the unacceptable result that the extreme regions of real cosmic structures are also removed.

To deal with this problem, we take the Bayesian approach of bringing in prior information. Most of the problems are associated with the Milky Way, so we can make a good guess in advance about whether a given region should be masked. We therefore consider two indicators of potential problems: extinction and stellar density, measured via $E(B-V)$ and the empirical total *WISE* density, Σ (to $W1 < 17$). The latter additionally brings in information on *WISE* coverage issues that cause spurious over- and underdensities in the source distribution. We use the first estimate of the mask derived from clipping to estimate a prior probability that a given pixel is masked as a function of these variables; this is shown in Figure 8. From this, it is clear that regions at $E(B-V) > 0.25$ should be completely clipped. We can now repeat the clipping analysis, but considering a full posterior probability that a given pixel is clean:

$$p_c = f_{\rm prior} \times p(\delta), \tag{4}$$

where $f_{\rm prior}$ is the fraction of pixels accepted at that Galactic location. It is now possible to clip with a more discerning threshold, $p_c \sim 10^{-5}$, which removes negligible amounts of real large-scale structures, while still remaining sensitive to anomalies at low latitudes. As a further precaution, we apply "guilt by association," and mask all pixels within 1 degree of a masked pixel. Finally, this process can be iterated, updating the prior when a revised mask has been generated. The final mask, shown in Figure 9, it removes 32% of the sky, leaving a satisfactorily clean galaxy sample on 28,000 deg^2. Applying the mask to the data gives us a final catalog of almost 18.7 million sources, as illustrated in Figure 10. The mean surface density of the sources is about 670 deg^{-2}, which is a more than 20-fold increase over 2MASS. We note, however, that for cosmological applications it might be more appropriate to repeat the above masking procedure on data that is first preselected in photo-z or other bins (e.g., magnitude).

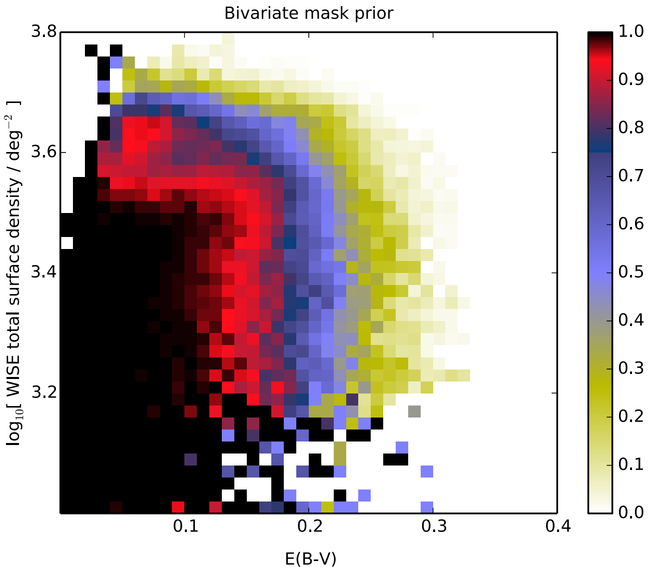


Figure 8. Initial prior for the mask, based on clipping regions of abnormal galaxy density. The probability of a pixel being accepted is shown as a function of extinction and of total *WISE* surface density at $W1 < 17$ as a proxy for stellar density.

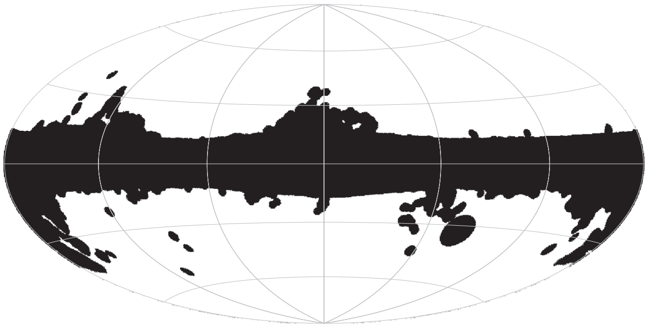


Figure 9. Final mask applied to the *WISE* × SuperCOSMOS galaxy catalog, presented here in Galactic coordinates with $\ell = 0$, $b = 0$ in the center. Black areas are masked and more than 68% of sky is retained for further analysis.

4.4. Completeness and Purity of the Final Catalog

Having applied all the cuts and the mask aiming at optimizing the reliability of the *WISE* × SCOS galaxy catalog, we now quantify its levels of completeness and purity. This was done using external data that will be treated as the "truth," ignoring any imperfections. Because our catalog was created by requiring detection in three independently surveyed wavebands, we assume that all our objects are genuine astronomical sources. We then need to measure the purity of the catalog (i.e., the fraction of our objects that are actually galaxies rather than stars) and its completeness (the fraction of all true galaxies that are included).

Purity is relatively easy to assess via cross-matching with SDSS. We selected a one-degree-wide strip centered at $\delta = 30°$ with magnitude limits much fainter than those of *WISE* × SCOS, which yielded more than 130,000 matched sources at $|b| > 12°$. From this, we found that at high latitudes,

---

[30] http://healpix.sourceforge.net/

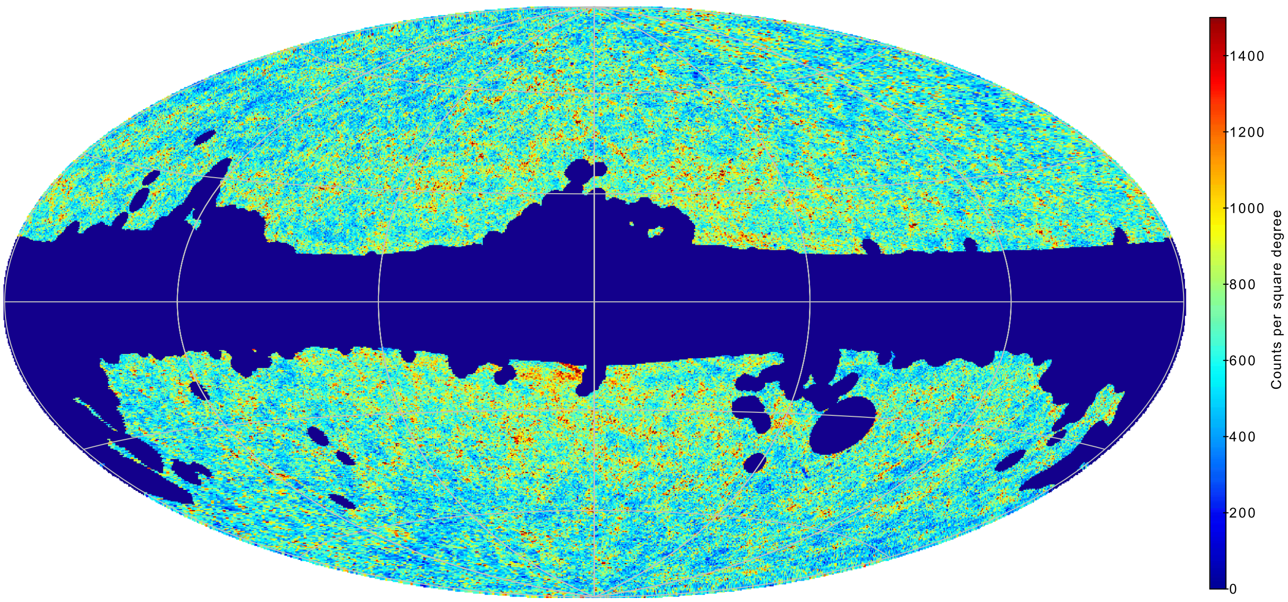


Figure 10. *WISE* × SuperCOSMOS galaxy catalog after star and quasar cleanup and masking in an all-sky Aitoff map in Galactic coordinates. The unmasked region (68% of the sky) contains 18.7 million sources flux-limited to extinction-corrected limits of $B < 21$, $R < 19.5$, and $13.8 < W1 < 17$.

approximately 95% of our sources are indeed galaxies. For latitudes of $|b| > 60°$ the stellar contamination of our catalog does not exceed 6%, and remains less than 10% down to almost $|b| = 40°$. One could further improve the purity at the expense of completeness; for instance, using a color cut of $W1 - W2 > 0.2$ instead of the fiducial one used in Section 4.2, would allow for a catalog with a stellar contamination of $\lesssim$3% down to $|b| \sim 40°$. Of course, this would also lead to a significant drop in completeness, because about 30% of galaxies have $W1 - W2 \leqslant 0.2$ (typically early types).

Completeness is slightly more complicated to assess thanks to magnitude errors. Even if our catalog was perfect in all other respects, it will still miss many galaxies that are really just brighter than our magnitude limits, and include many that are in reality just slightly fainter. Therefore, the assessment of completeness involves two questions that go beyond noisy magnitudes: (1) What fraction of true galaxies are incorrectly classified by SCOS as stars? (2) What further fraction of galaxies are lost as a result of the color cuts aimed at purifying the sample? The first question can be addressed by looking at a pairing of SDSS data with both the SCOS galaxy *and* star catalogs (the latter classified as `meanClass` = 2). We started with the same SDSS strip centered at $\delta = 30°$ as above, synthesized SCOS magnitudes (J. A. Peacock et al. 2016, in preparation), and cut the sample to $B_{\rm syn} < 21$ and $R_{\rm syn} < 19.5$, looking at the relative numbers that paired with SCOS galaxies and stars. The conclusion is that the overall misclassification incompleteness is about 15%, with some dependence on Galactic latitude. Averaged completeness exceeds 90% for $|b| > 40°$ and equals 88% for half the sky ($|b| > 30°$). Unsurprisingly, the limited image quality on Schmidt plates leads to compact galaxies being classified as stars. According to Baldry et al. (2010), the corresponding figure for SDSS is about 2% (GAMA does much better because it uses near-IR colors in addition to image width).

Finally, we can use GAMA to measure the effect of the additional color cuts applied in Sections 4.1 and 4.2. The depth and completeness of GAMA are large enough that we can assume that practically each *WISE* × SCOS galaxy should also be present in GAMA. We thus treat the cross-match of the two catalogs *before* star and quasar removal as the reference. As already mentioned, the quasar cutout criterion (Equation (3)) affects less than 1% of GAMA galaxies; thus the most important for the final completeness will be the color cut used to discard stellar contamination. Using our prescription for star removal (Section 4.2), we lose about 6% of galaxies at $|b| \geqslant 30°$, increasing to 10% for the lowest latitudes of $b \sim 22$ observed by GAMA. We can thus safely assume that, including the 1% drop in completeness due to quasar cutout, our catalog is about 93% complete for half the sky ($|b| > 30°$) and more than 90% complete over at least 2.4π steradians (in addition to the classification incompleteness).

This completeness analysis is ultimately limited by the fact that our catalog is a cross-match of two independent samples of different characteristics. Our data set will thus miss some sources that could not be detected by one of the parent surveys, or by both. For instance, *WISE* is not sensitive to low surface brightness galaxies, while SCOS is biased against the dusty ones that *WISE* detects very well. Quantification of these effects is beyond the scope of this paper and it would require using much deeper reference catalogs of otherwise very similar preselections than those employed here.

5. PHOTOMETRIC REDSHIFTS

This Section presents the derivation and analysis of photometric redshifts for our full galaxy sample. To compute the photo-zs, we used the ANNz package[31] developed by Collister & Lahav (2004), which is an artificial neural network algorithm that estimates photometric redshifts based on a training sample with photometric quantities and spectroscopic redshifts (see also Firth et al. 2003). The ANNz photo-z estimator has been shown to be one of the most accurate methods (e.g., Abdalla et al. 2011; Sánchez et al. 2014), so long as a sufficiently large and representative spectroscopic

---

[31] Available from http://www.homepages.ucl.ac.uk/~ucapola/annz.html.

sample is available for redshift calibration (the latter being generally true for empirical photo-z methods). In our case, such a sample is provided by the deep and complete GAMA data set. We also experimented with another photometric redshift code, GAz (Hogan et al. 2015),[32] which gave results similar to ANNz, albeit slightly poorer.[33]

As was the case for 2MPZ (B14), we have not used any template-fitting photo-z estimation methods. One cannot efficiently implement them when incorporating SCOS photometry, mostly because the photographic filter transmission curves are not known to sufficient accuracy. We note however that another independent technique of redshift estimation could in principle be employed for the catalog presented here, namely the "clustering redshifts" (e.g., Newman 2008; Ménard et al. 2013) recently applied to the SDSS (Rahman et al. 2015, 2016b) and 2MASS (Rahman et al. 2016a) samples.

The most optimal neural network architecture for the ANNz code is not known a priori and depends on such parameters as the number of photometric bands used and the size of the training sample. For each of the tests we have always tried a number of different architectures, limiting ourselves to no more than two intermediate layers (adding more layers does not increase photo-z accuracy). We used "committees" of at least six networks and in most cases the most accurate photo-zs were obtained by applying network architectures with one or two intermediate layers, 10–35 nodes in each.

Before computing photometric redshifts for our full sample, we performed extensive tests of their properties using GAMA redshifts as training and test samples. Such an approach gives the most comprehensive results, as GAMA is a highly complete, flux-limited sample deeper than our catalog (Section 4) and offers much auxiliary information that allowed us to examine photo-z performance as a function of both observed (apparent) and intrinsic properties of the sources. We also experimented with adding the SDSS DR12 spectroscopic data set to the GAMA calibration sample, but it was too non-uniformly selected at $r > 17.77$ (beyond the Main Sample) to be applicable for photometric redshift training.

Empirical photometric redshift estimators, such as the ANNz, generally provide better results when more photometric parameters are used in the photo-z derivation. It would thus be desirable to add more bands to the four basic ones employed to preselect *WISE* and SCOS data (B, R, $W1$, $W2$), but this is not currently possible for our full catalog. In B14 we showed that the SCOS I band does not significantly change the photo-z accuracy for the GAMA-based sample, mainly because it is shallower than B and R. The situation is no better with *WISE*, where two additional bands are in principle available: mid-IR $W3$ and $W4$ centered, respectively, on 12 and 23 μm. However, these bands were of much lower sensitivity than $W1$ and $W2$: only 30% of our sources have $\mathrm{S/N} > 2$ in $W3$ and fewer still are detected in $W4$ (compare also Cluver et al. 2014). The I, $W3$, and $W4$ bands will thus not be used in the derivation of all-sky photometric redshifts to preserve uniformity of the catalog.

5.1. Tests and Calibration on GAMA

In B14 we presented the potential of applying GAMA as a photo-z training sample for *WISE* × SCOS all-sky data. Using the shallower GAMA DR2 public release (complete to $r < 19.0$ in two of the equatorial fields and to $r < 19.4$ in the third one; Liske et al. 2015) together with the *WISE* "All-Sky" and SCOS extended source data, we obtained an accuracy of $\sigma_z \simeq 0.035$ at a median redshift of $z \sim 0.18$. Here we extend that study using a complete flux-limited $r \leqslant 19.8$ GAMA-II sample, along with deeper AllWISE data, corrected SCOS color calibrations, and Galactic extinction coefficients revised according to Schlafly & Finkbeiner (2011).

We start with the *WISE* × SCOS × GAMA sample as described in Section 3. We tested photo-z performance for various cuts applied to this data set, such as flux limits being the same as in the all-sky sample discussed in Section 2.3, star and quasar color cuts (Section 4), and so on. We found that photo-z statistics are practically independent of whether the cuts are first applied both on the training and test sample, or only on the latter once ANNz had been trained on the full sample; in other words, it is enough to train the neural networks on the most general training set possible and apply the required cuts on the test set only. The only cut applied a priori to all the GAMA samples discussed in this Section was $\delta_{1950} < 2^\circ\!.5$, to avoid residual passband mismatch between SCOS "north" and "south." More discussion of this issue will be provided in Section 5.2.

The *WISE* × SCOS × GAMA sample of 142,000 sources was divided randomly in proportion 1:9 into training and test sets, and a validation set was additionally separated out from the former (20% of the full training set). Summary statistics for the photo-z tests are provided in Table 3. Note that the outlier fraction is defined here relative to the scatter of each of the particular test sets, so the decrease in the scatter may lead to a slight increase in the outlier rate. The table also includes the statistics computed separately for the three GAMA equatorial patches (in the case of no flux limits in *WISE* × SCOS), showing that the variations in the photo-z accuracy between these fields are not significant. Comparing the general results with Table 2 of B14 shows that the current photo-z performance is very similar to the one achieved with GAMA DR2, taking into account the increased depth of the present sample. It is also worth noting that our uncertainty of $\sigma_z \simeq 0.03$ for the flux-limited case is comparable to the results of Christodoulou et al. (2012), where GAMA spectroscopy was applied to train a large SDSS $r < 19.4$ sample using five-band *ugriz* photometry. Last but not least, this accuracy is very close to the prediction for future surveys such as Euclid or LSST (Ascaso et al. 2015).

Figures 11–13 illustrate the general performance of our photometric redshifts trained and tested on GAMA for the flux-limited *WISE* × SCOS sample. A comparison of z_{spec} with z_{phot} (Figure 11) and z_{phot} with the difference between them (Figure 12) confirms the expected property of the photo-zs being unbiased in the true z_{spec} at a given z_{phot} (Driver et al. 2011; B14; Sadeh et al. 2015); a non-flat $N(z)$ must lead, however, to the redshifts being photometrically overestimated at the low end and underestimated at high z.

The redshift distribution (Figure 13) shows similar features to those in Figure 13 of Driver et al. (2011), where photo-zs were derived using ANNz for an earlier version of GAMA. The dN/dz_{phot} diagram is narrower than the spec-z one and the

---

[32] https://github.com/rbrthogan/GAz

[33] A new version of ANNz, dubbed ANNz2 (Sadeh et al. 2015), was released when the present work was in an advanced stage, and we postpone its possible application to future work.

Table 3
Statistics for the Photometric Redshift Estimation Generated for a Test Sample of *WISE* × SuperCOSMOS Sources in GAMA Equatorial Fields

# of Sources[a]	Mean $\langle z \rangle$ Spec[b]	Mean $\langle z \rangle$ Phot[c]	Median $\tilde{z}$ Spec[b]	Median $\tilde{z}$ Phot[c]	1σ Scatter[d] $\sigma_{\delta z/(1+z)}$	Scaled MAD[e]	Norm. SMAD[f]	Mean Bias[g] $\langle \delta z \rangle$	Median Error[h]	% of Outliers[i]
Trained and tested on GAMA										
no flux limits										
127,703	0.231	0.231	0.223	0.233	0.045	0.044	0.036	−7.3e−5	14.1%	2.8%
separate fields										
(G09) 37,810	0.242	0.240	0.242	0.244	0.044	0.045	0.037	−2.0e−3	14.0%	2.4%
(G12) 48,974	0.229	0.230	0.221	0.231	0.045	0.044	0.036	6.8e−4	14.5%	2.7%
(G15) 40,919	0.223	0.224	0.211	0.224	0.044	0.043	0.036	7.6e−4	14.6%	2.9%
flux limited $B < 21$ and $R < 19.5$ and $W1 < 17$										
86,516	0.200	0.200	0.191	0.202	0.041	0.040	0.033	−3.7e−4	14.7%	2.8%

Notes. Two cases are shown: (i) no flux limits applied to the photometric sample; (ii) fiducial flux limits as in the final data set. In the former case, we also show statistics for the particular GAMA fields.

[a] In the test set.

[b] Input (spectroscopic) redshift sample.

[c] Output (photometric) redshift sample.

[d] Normalized 1σ scatter between the spectroscopic and photometric redshifts, $\sigma_{\delta z/(1+z)}$; unclipped.

[e] Scaled median absolute deviation, $\text{SMAD}(\delta z) = 1.48 \times \text{med}(|\delta z - \text{med}(\delta z)|)$.

[f] Scaled median absolute deviation of the normalized bias, $\text{SMAD}(\delta z/(1 + z_{\text{sp}}))$.

[g] Mean bias of z_{phot}: $\langle \delta z \rangle = \langle z_{\text{phot}} - z_{\text{spec}} \rangle$; unclipped.

[h] Median of the relative error, $\text{med}(|\delta z|/z_{\text{sp}})$; unclipped.

[i] Percentage of outliers for which $|(z_{\text{ph}} - z_{\text{sp}})/(1 + z_{\text{sp}})| > 3\,\text{SMAD}(\delta z/(1 + z_{\text{sp}}))$.

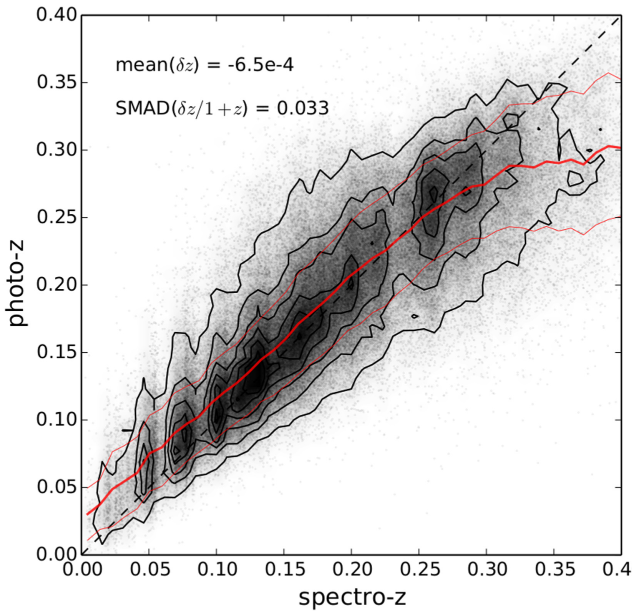


Figure 11. Comparison of GAMA spectroscopic redshifts with the photometric ones derived from *WISE* × SuperCOSMOS photometry, for the $B < 21$, $R < 19.5$, and $W1 < 17$ flux-limited sample of *WISE* × SCOS × GAMA sources. Red lines show the running median photo-z and its scatter (SMAD). The vertical striping of the density plot results from galaxy overdensities being radially diluted in the photometric redshift space.

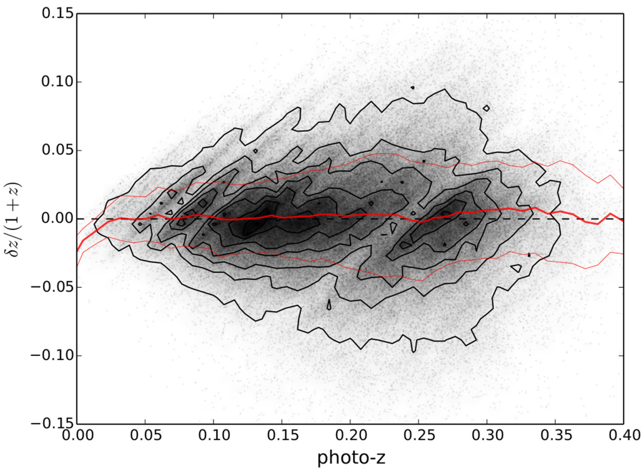


Figure 12. Photometric redshift accuracy as a function of the photo-z derived from *WISE* × SuperCOSMOS photometry, for the $B < 21$, $R < 19.5$, and $W1 < 17$ flux-limited sample of *WISE* × SCOS × GAMA sources. Red lines illustrate the running median and scatter (SMAD) of $\delta z/(1 + z)$.

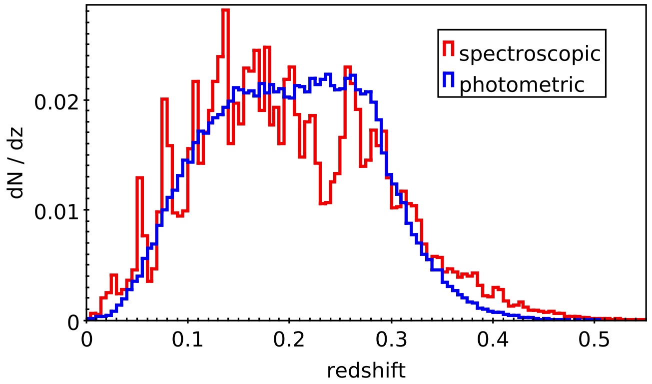


Figure 13. Comparison of the *WISE* × SCOS × GAMA spectroscopic redshift distribution (red) with the photometric one derived from *WISE* × SuperCOSMOS photometry (blue), for the $B < 21$, $R < 19.5$, and $W1 < 17$ flux-limited sample.

former is unable to reproduce sharp features in the latter, such as the dip at $z \sim 0.23$ and several peaks related to clusters and walls. However, the quality of our photometric redshifts is impressive given that we used only two optical bands. The latter limitation cannot currently be overcome when constructing nearly all-sky photo-z samples with the presently available data, because they are covering much more of the sky at $z \sim 0.2$ than available from, for example, SDSS only (D'Abrusco et al. 2007; Oyaizu et al. 2008; Brescia et al. 2014; Beck et al. 2016).

5.1.1. Dependence on Apparent Properties

We have examined the photometric redshift performance as a function of various observed and intrinsic properties of the galaxies. As far as the former are concerned, we found no alarming patterns in the photo-z accuracy as a function of apparent magnitudes in the four bands used in the procedure, other than a general deterioration in the photometric redshift quality as the sources become fainter, which is consistent with expectations. Note, however, that because our catalog is produced by combining optical and infrared preselections, in some magnitude bins, the fainter sources are not necessarily more distant. This, together with the dependencies of photometric redshifts on varying magnitude, is illustrated for the $W1$ band in Figure 14. Here we inverted the axes with respect to Figure 11 to emphasize possible systematics for samples selected in photo-z and magnitude bins, as will be practical for the full-sky sample where spectroscopic redshifts are not available. Some issues are evident, such as the lack of $z_{\rm phot} > 0.3$ galaxies for $W1 > 15.5$.

This analysis of photometric redshift properties can be made more detailed by binning the data further. In particular, in addition to $W1$ intervals, we also divided the test set into bins of the observed $B - R$ color (in $\Delta(B - R) = 0.5$ mag), as well as of $z_{\rm spec}$ and $z_{\rm phot}$ (in bins of $\Delta z = 0.1$). This gives two 3D "tables," each cell of which contains photo-z statistics as in Table 3. The extracts of these two tables are provided in Table 4 for a particular bin of $z_{\rm spec}$ ($z_{\rm phot}$) and of $B - R$. Full electronic versions of these tables can be made available on request.

Variations of the photometric redshift quality are also observed with *WISE* $W1$ and $W2$ signal-to-noise levels, which are already strongly correlated with the flux. In particular, there is a noticeable decrease in photo-z accuracy for sources with `w2snr` < 5, as compared with those with `w2snr` > 5: the former have an order of magnitude larger mean bias $\langle\delta z\rangle$ than the latter and a considerably larger scatter in δz. Interestingly, in the cross-matched *WISE* × SCOS × GAMA sample, the low-`w2snr` sources are on average located at *smaller* distances than the high-`w2snr` ones: the `w2snr` < 5 galaxies have $z_{\rm med} \simeq 0.13$ (spectroscopic) and practically never reach beyond $z = 0.4$. The low-`w2snr` sources, however, constitute a small fraction (less than 5%) of our galaxy catalog, and are mostly localized in several strips crossing the ecliptic, resulting from moon avoidance maneuvers.

5.1.2. Dependence on Intrinsic Properties

As an additional verification, we examined the photometric redshift performance as a function of source intrinsic properties, such as absolute magnitudes, rest-frame colors, and stellar masses. Such parameters are not available for the full-sky sample because they require spectroscopic redshifts, but these tests are useful to search for possible issues in the photo-z data set. The test were done "blindly," that is, the parameters had not been used in the photo-z procedure, and they were extracted a posteriori from two GAMA data management units (DMUs), namely `StellarMasses v16` (Taylor et al. 2011) and *WISE*-`GAMA v01` (Cluver et al. 2014). The first provides optical and near-infrared absolute magnitudes and rest-frame colors, as well as galaxy stellar masses and several other ancillary parameters, while the second offers *WISE*-derived mid-infrared photometry, including isophotal magnitudes for resolved sources. Cluver et al. (2014) also discuss the derivation of the absolute luminosities and stellar masses of *WISE* × GAMA galaxies, which we use here.

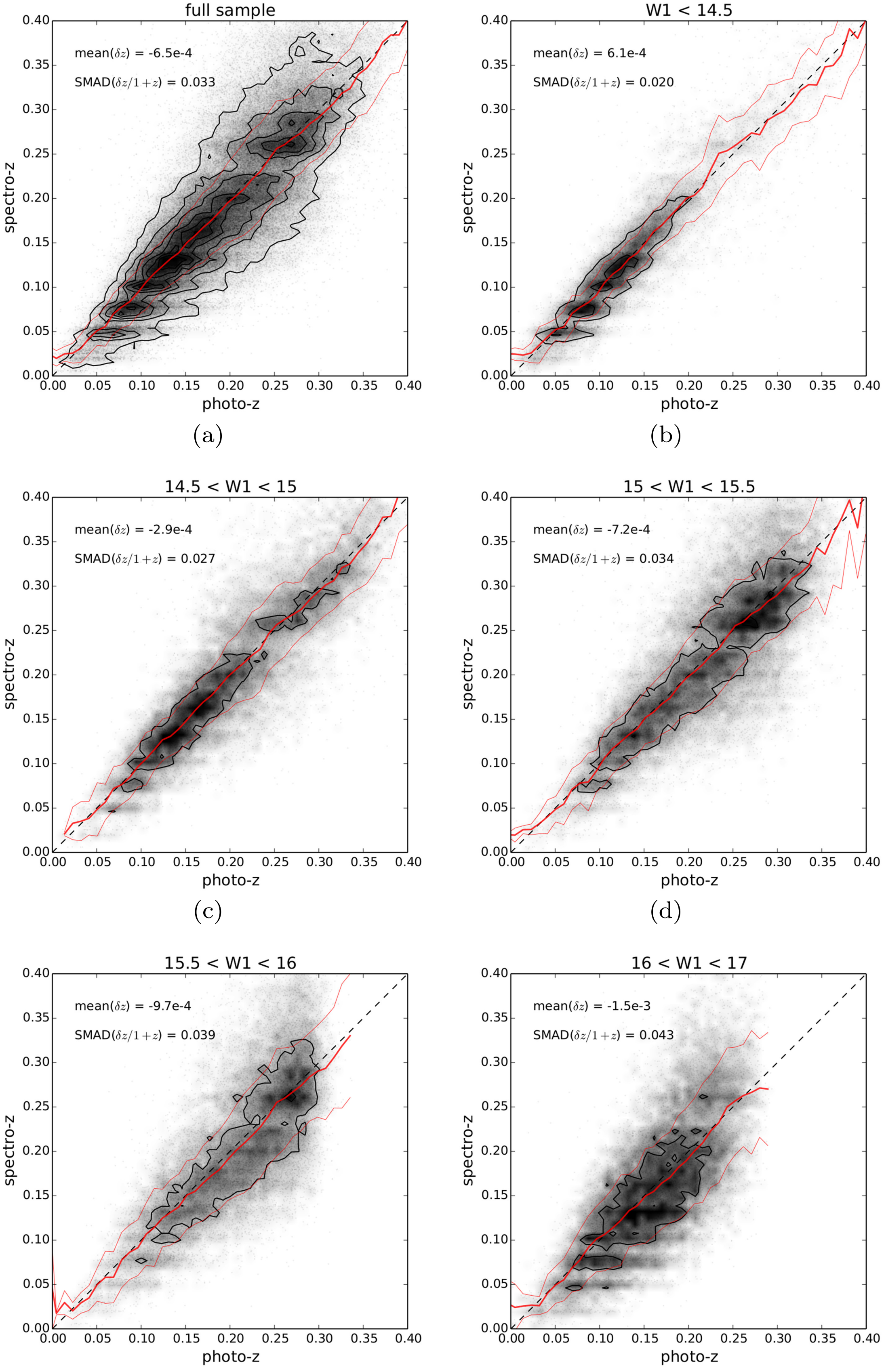
In practically all the bands available for analysis from these GAMA DMUs (from u up to $W3$), the photometric redshifts as a function of absolute magnitude are typically underestimated for bright galaxies and overestimated for faint ones. This is expected and related to the previously mentioned property that the photo-zs are not unbiased in the true redshift at a given $z_{\rm spec}$, but are unbiased at $z_{\rm phot}$. A similar dependence is found for stellar masses, which again is not unexpected because of the correlation between the galaxy's absolute luminosity and stellar mass (being the tightest in near-infrared bands). Interestingly, our photometric redshifts are relatively unbiased as a function of rest-frame colors, such as $u - r$ or $g - i$.

5.2. Final All-sky Catalog

After performing all the tests discussed above, we trained the ANNz algorithm on the full GAMA-south sample (142,000 sources, of which 98,000 fall within the flux limit of the core *WISE* × SCOS catalog) and applied the resulting networks to the cleaned data set described in Section 4. Figure 15 compares the normalized redshift distributions of the *WISE* × SCOS × GAMA spectroscopic input (red bars) with the all-sky *WISE* × SCOS photometric output (black line). In Figure 16 we present the absolute dN/dz, in logarithmic scaling, for three "all-sky" data sets: 2MASS Redshift Survey (2MRS; Huchra et al. 2012), 2MASS Photometric Redshift catalog (2MPZ; B14), and the *WISE* × SCOS photo-z sample. This clearly illustrates the great improvement in information that *WISE* × SCOS brings at $z > 0.1$ as compared with 2MASS.

As described in Section 2.2, and discussed earlier in Francis & Peacock (2010) and B14, the SCOS passbands between the north ("N," $\delta_{1950} > 2\overset{\circ}{.}5$) and the south ("S") were slightly different; if not accounted for, this can lead to a bias between the photometric redshifts in the two "hemispheres." The first step toward making N and S consistent was to calibrate the two parts using Equations (1)–(2) from Section 2.2. In B14 a possible residual photo-z bias due to imperfect calibration was avoided by training the neural networks separately for N and S, which was possible thanks to comprehensive training sets in both parts of the sky. At the depths of the present sample, using GAMA data for photo-z derivation in each hemisphere is not practical because the majority of the data set is below $\delta = 2\overset{\circ}{.}5$: only 9% of the GAMA sample is in the north. Such a training set of ~10,000 sources is too small for proper calibration of the photo-zs for $\sim 10^7$ *WISE* × SCOS objects.

In the absence of a large northern spectroscopic data set of GAMA's depth, there is no direct solution to this problem. We therefore rely, in the first instance, on the color corrections by which we attempt to place SCOS photometry in both hemispheres on a uniform basis. However, there exist remaining inconsistencies in the photo-zs between N and S, which we address in the same way as in Francis & Peacock (2010): examine the probability distribution of photo-zs in the two "hemispheres" and make an adjustment to the redshift scale so that these distributions are consistent. The result of the procedure is illustrated in Figure 17, where the derived offset



(a) (b)

(c) (d)

(e) (f)

Figure 14. Dependence of photometric redshift quality on the apparent $W1$ magnitude in the *WISE* × SuperCOSMOS sample calibrated on GAMA spectroscopic redshifts. Panel (a) shows the full sample, and the subsequent panels (b)–(f) present data binned in $W1$ intervals. Red lines illustrate the running median and scatter (SMAD) of z_{spec}.

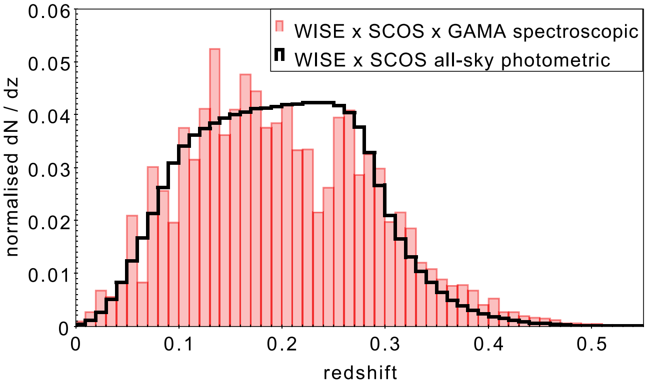


Figure 15. Normalized redshift distributions of the *WISE* × SCOS × GAMA spectroscopic training set (red bars) and of the final all-sky *WISE* × SuperCOSMOS photometric sample (black line).

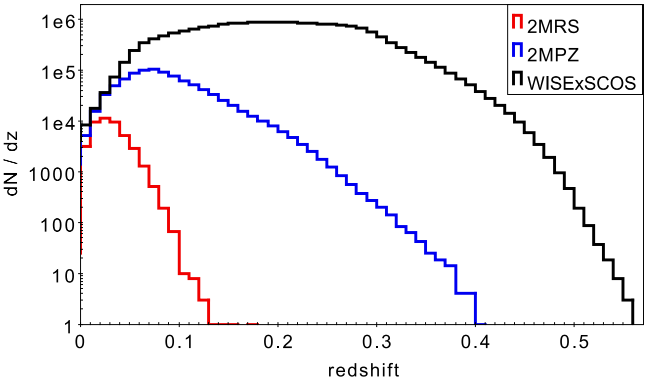


Figure 16. Redshift distributions for three major all-sky surveys: 2MRS spectroscopic (red), 2MPZ photometric (blue), and *WISE* × SCOS photometric (black).

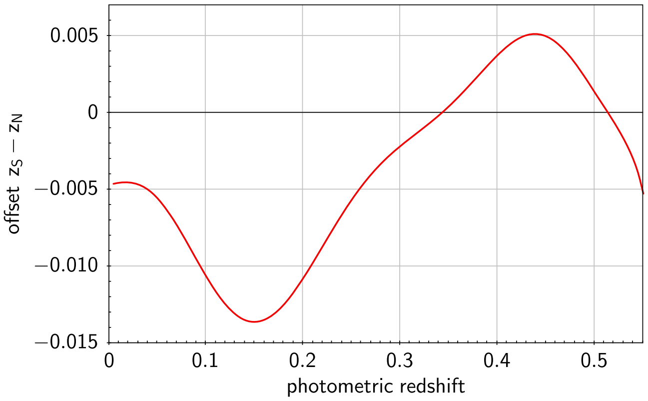


Figure 17. Average offset between the photometric redshifts in the north and in the south, plotted as a function of the redshift. For each galaxy located in the north ($\delta_{1950} > 2\overset{\circ}{.}5$), this correction is added to the photo-*z*s derived with ANN*z*.

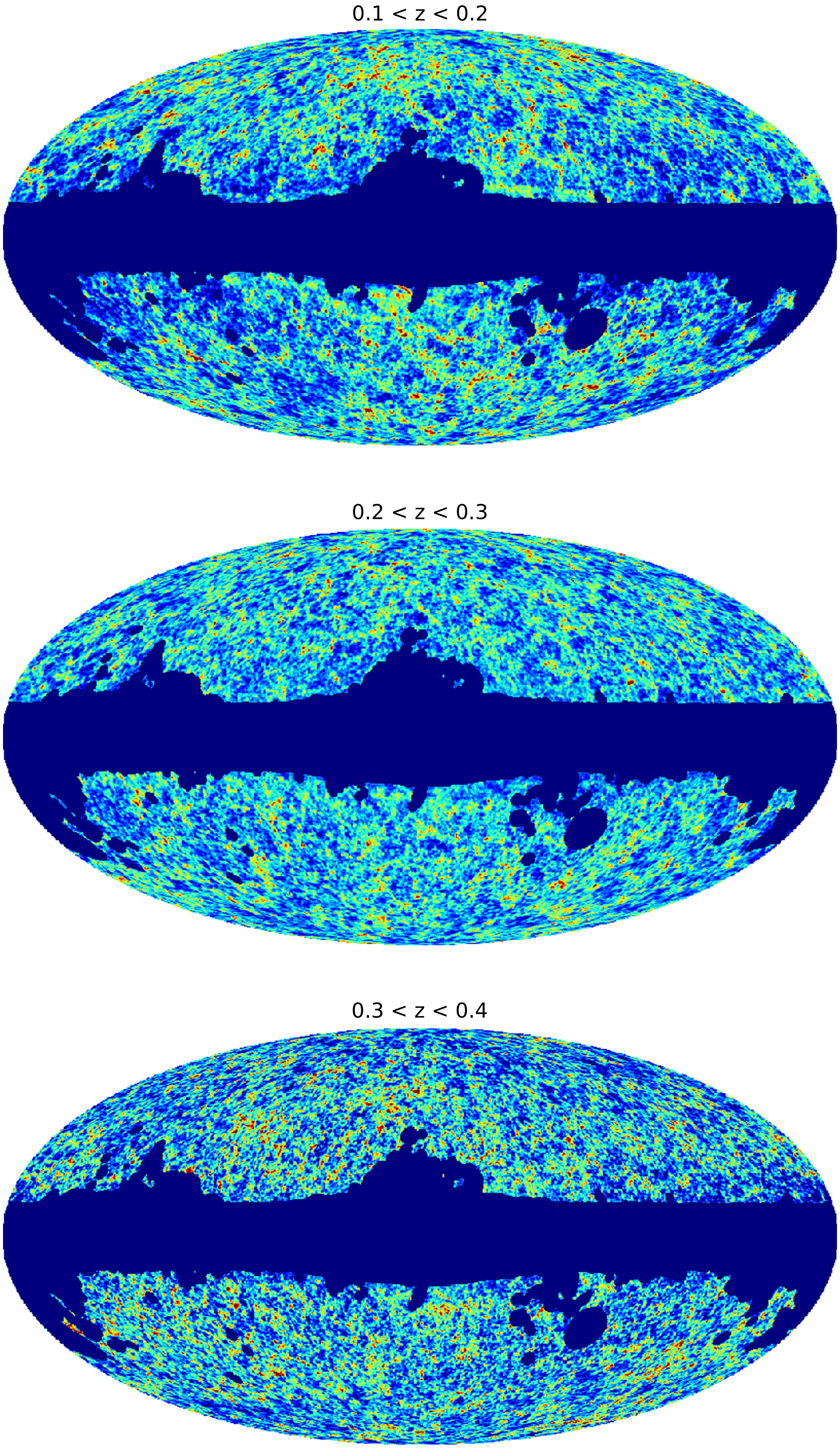
$z_S - z_N$ is plotted as a function of z_N. In all cases, such a correction is a small fraction of the photo-*z* precision, but it does seem to be real. Therefore, for the sources at $\delta_{1950} > 2\overset{\circ}{.}5$, we added this offset to the individual photo-*z*s derived with ANN*z*.

An additional test of photometric redshift quality comes from extending the purity analysis of Section 4.4. Examining photo-*z* distributions of the sources identified as galaxies and as stars in the cross-match with the photometric SDSS data, we found that the stars contaminating the full-sky sample are assigned photo-*z*s peaking at $z \sim 0.05$ rather than zero; because the training sample used to derive the photo-*z*s contains no stars, the neural networks will then assign them redshifts of galaxies that are closest in the parameter space. But interestingly, some stars are assigned very high redshifts: at $z_{\rm phot} > 0.4$, more than 40% of the *WISE* × SCOS photo-*z* sample is stellar contamination. However the absolute source numbers are very small at these redshifts: only 144,000 of the sources have such photo-*z*s (less than 1% of the full sample). In general, we conclude that the final sample can be purified further over what was discussed in Section 4 by removing the lowest- and highest-photo-*z* bins. The contamination is always smaller than 20% for a cutout of $0.085 < z_{\rm phot} < 0.345$, which is 90% of the full sample; the purity improves further if the lowest Galactic latitudes are discarded. We would like to emphasize that for cosmological studies benefiting from “tomographic” slicing in redshift bins, the $z < 0.1$ range is better sampled by 2MPZ (B14), rather than the present catalog owing to the much higher completeness, purity, and photo-*z* accuracy of the former.

With these reservations, Figure 18 shows examples of three photo-*z* shells of $\Delta z = 0.1$ extracted from the full *WISE* × SCOS sample, centered, respectively, on $z = 0.15$, $z = 0.25$, and $z = 0.35$, and including, respectively, 7.3, 7.4, and 1.7 million sources. To produce these images we made one final correction for the redshift-dependent stellar contamination discussed above, which serves to mitigate large-scale non-uniformities, especially in the lowest redshift shell. We correlated the surface density in the slice with the total *WISE* sky density (treated as a proxy for stellar density) and removed the appropriate scaled fraction to remove the stellar gradients. This process is successful in yielding redshift slices with no apparent large-scale artifacts. These are the most comprehensive illustrations of all-sky galaxy distribution at these redshifts available so far, revealing new large-scale structures. In particular, neither 2MASS, nor especially 2MRS or PSC*z*, could reach to depths of $z > 0.15$ in a comprehensive manner. Three-dimensional sampling of the cosmic web at these scales has been so far possible only with SDSS, covering three times less sky than our catalog, and being less complete beyond $r > 17.77$ as far as the spectroscopic data are concerned.

5.3. Comparison with External Redshift Samples

As a final “blind” test of the photometric redshifts in our catalog, and to verify whether they exhibit noticeable variations in performance over the sky, we cross-matched the sample with a number of external redshift catalogs. Such an exercise is only meaningful for auxiliary data sets that are complete to a depth at least similar to the present sample. For that reason, the cross-matched catalogs mainly cover small fields: other than GAMA, there are currently no other wide-angle spectroscopic data sets complete to its depth (otherwise we would have used them for the photo-*z* training). In fact part of the GAMA data is included in this test, as some of them were not used in the photo-*z* derivation and tests described above. These include the equatorial data from the G09 and G15 fields above $\delta_{1950} > 2\overset{\circ}{.}5$ (cf. Section 5.1), as well as the southern G02 and G23 fields (which are less complete than the equatorial ones; Liske et al. 2015), and the much deeper G10/COSMOS field (Davies et al. 2015). Except for the very deep but small (~2 deg^2) G10 field, all of these have high matching rates with *WISE* × SCOS because they have the same (G09 and G15) or



$0.1 < z < 0.2$

$0.2 < z < 0.3$

$0.3 < z < 0.4$

Figure 18. The large-scale structure at $z = 0.1$–0.4: all-sky projections of $\Delta z = 0.1$ photo-z slices from the *WISE* × SuperCOSMOS galaxy catalog, illustrating the power of this data set in mapping the cosmic web at these redshifts.

Table 4
Extract from Tables Showing Photometric Redshift Statistics in Bins of Redshift (Separately Spectroscopic and Photometric), Observed $B - R$ Color and Apparent $W1$ Magnitude

redshift bin	$B - R$ color bin	$W1$ mag bin	% of sources[a]	mean bias[b]	1σ scatter[c]
Trained and tested on GAMA					
binned in spectro-z					
$0.2 < z_{\rm spec} < 0.3$	$1.5 < B - R < 2.0$	$W1 < 14$	113	−0.0135	0.0251
		$14 < W1 < 15$	2667	0.0111	0.0245
		$15 < W1 < 16$	4001	0.0152	0.0236
		$16 < W1 < 17$	92	−0.0131	0.0284
binned in photo-z					
$0.2 < z_{\rm phot} < 0.3$	$1.5 < B - R < 2.0$	$W1 < 14$	112	−0.0135	0.0287
		$14 < W1 < 15$	2591	−0.0002	0.0261
		$15 < W1 < 16$	4179	−0.0015	0.0336
		$16 < W1 < 17$	144	−0.0025	0.0437

Notes.
[a] In the test set.
[b] Mean bias of $z_{\rm phot}$: $\langle \delta z \rangle = \langle z_{\rm phot} - z_{\rm spec} \rangle$; unclipped.
[c] Normalized 1σ scatter between the spectroscopic and photometric redshifts, $\sigma_{\delta z/(1+z)}$; unclipped.

very similar (G02 and G23) preselections as the "fiducial" GAMA data used for the photo-z training.

We paired several other publicly available data sets[34] with *WISE* × SCOS; however, only a fraction had a significant matching rate, and we discuss here only those that had at least 500 common sources with our catalog. In all cases, we used a $2''$ matching radius, which is a compromise between minimizing spurious cross-matches that could result from imprecise astrometry, and maximizing the matching rate for cases of not well defined centroids.

Details regarding the fields' central coordinates, areas, and the *WISE* × SCOS photo-z statistics (calculated with respect to the redshifts provided in the external data sets) are provided below and listed in Table 4. The samples included are:

1. GAMA data in the equatorial G09 and G15 fields located at $\delta_{1950} > 2^{\circ}.5$; this is part of the sample comprehensively described in Section 3, removed from the photo-z training and test phase due to the SuperCOSMOS north–south band difference (cf. Section 2.2); this sample includes 17,523 galaxies with $z_{\rm med} = 0.217$ after cuts of $z > 0.002$ and redshift quality `NQ` $\geqslant 3$.
2. GAMA G02 field (`TilingCat v04`): a spectroscopic redshift survey of a ~56 deg^2 field centered at $\alpha = 34^{\circ}.5$, $\delta = -7^{\circ}$, currently not publicly available (part of GAMA-II, Liske et al. 2015), with targets preselected from SDSS DR8 (Aihara et al. 2011) and CFHTLenS (Heymans et al. 2012) photometric data sets to a limiting magnitude of $r < 19.8$, although not fully complete to this limit over the whole field (Liske et al. 2015); this sample includes 33,677 galaxies with $z_{\rm med} = 0.226$ after cuts of $z > 0.002$ and redshift quality `NQ` $\geqslant 3$.
3. GAMA G23 field (`TilingCat v11`): a spectroscopic redshift survey of a ~87 deg^2 field centered at $\alpha = 345^{\circ}$, $\delta = -32^{\circ}.5$, currently not publicly available (part of GAMA-II, Liske et al. 2015), with targets preselected from KiDS (de Jong et al. 2013) to a limiting magnitude of $i < 19.2$, although not fully complete to this limit over the whole field (Liske et al. 2015); this sample includes 47,489 galaxies with $z_{\rm med} = 0.208$ after cuts of $z > 0.002$ and redshift quality `NQ` $\geqslant 3$.
4. AGES (AGN and Galaxy Evolution Survey; Kochanek et al. 2012): a spectroscopic redshift survey covering ~10 deg^2, centered at $\alpha = 217^{\circ}.8$, $\delta = 34^{\circ}.3$, with targets preselected to a limiting magnitude of $I < 20$ from several data sets at various wavelengths; this sample includes 21,805 galaxies with $z_{\rm med} = 0.342$ after a cut of $z_{\rm spec} > 0$ (measured spectro-z).
5. SHELS (Smithsonian Hectospec Lensing Survey; Geller et al. 2014): a spectroscopic redshift survey of a ~5 deg^2 field centered at $\alpha = 140^{\circ}$, $\delta = 30^{\circ}$, with targets preselected from the Deep Lens Survey (Wittman et al. 2006), complete to a limiting magnitude of $R \leqslant 20.6$ with part of the sources beyond this limit; this sample includes 15,591 galaxies with $z_{\rm med} = 0.317$ after a cut on the redshift error $e_z < 0.001$ and using only unmasked sources.
6. G10/COSMOS data set: a publicly available redshift catalog (Davies et al. 2015), part of GAMA, covering 2 deg^2 in the COSMOS field ($\alpha = 150^{\circ}.1$, $\delta = 2^{\circ}.2$), obtained by re-reducing archival spectroscopic zCOSMOS-bright data (Lilly et al. 2007, 2009), together with input from other sources (PRIMUS, SDSS, VVDS); the sample includes 16,128 galaxies with $z_{\rm med} = 0.533$ after applying the flag `Z_USE` = 1 (reliable high resolution spectroscopic redshift) and a cut on redshift $z > 0$.
7. PRIMUS (PRIsm MUlti-object Survey; Coil et al. 2011; Cool et al. 2013): a low-resolution spectroscopic redshift survey covering a total of ~10 deg^2 in nine fields to a depth of $i_{\rm AB} \sim 23.5$, preselected from several imaging data sets; here we used only the `ZQUALITY` = 4 sources (highest quality redshifts) that gave 87,742 objects in total with $z_{\rm med} = 0.476$; one should bear in mind, however, that even for these sources, the PRIMUS *spectroscopic* redshift precision is $\sigma_{\delta z/(1+z)} = 0.005$ (Cool et al. 2013).
8. COSMOS photometric redshift catalog (Ilbert et al. 2009): providing very accurate photo-zs based on 30-band photometry from UV through mid-IR; this is the same field as in the case of G10/COSMOS GAMA release, but the number of sources is much larger because

---

[34] A comprehensive list of galaxy (redshift) surveys is provided at http://www.astro.ljmu.ac.uk/~ikb/research/galaxy-redshift-surveys.html.

there is no requirement of spec-z availability; we used version 1.5 of the catalog, magnitude-limited to $I < 25$, which includes 304,999 objects with photo-zs measured ($0 <$ `zp_best` < 9.99 in the catalog), with $z_{\rm med} = 0.888$; at the bright end of interest for this exercise, the COSMOS photo-zs have a scatter of $\sigma_{\delta z/(1+z)} = 0.007$ (Ilbert et al. 2009), which is comparable to the accuracy of PRIMUS low-res spectro-z.

Several other data sets were tested, but they either gave less than 500 cross-matches with *WISE* × SCOS each, or were too shallow for meaningful photo-z statistics. The latter case includes the SDSS spectroscopic data: while it does provide a very high matching rate with our catalog (more than 1/3 of SDSS DR12 spectroscopic galaxies are also in the *WISE* × SCOS fiducial sample), the specific preselections of SDSS targets leads to biased photo-z statistics. In order to circumvent this, one could try applying some weighting procedures, such as those discussed in Lima et al. (2008), which is beyond the scope of the present work.

The mean and median redshifts quoted in Table 4 differ from those in Table 2 (flux-limited case) either because of different preselections of the external fields and/or, as in the case of GAMA equatorial data at $\delta_{1950} > 2.5$, due to the bright-end flux cut of $W1 = 13.8$ applied to the fiducial *WISE* × SCOS sample, but not to the one referred to in Table 2. This is a minor detail that does not influence the conclusions.

Overall, the statistics provided in Table 4 do not indicate any significant variations of the *WISE* × SCOS photo-z quality among the tested samples, and hence over the sky. In particular, in most cases the scatter in $\delta z/(1 + z)$ (as measured through SMAD) is within 0.035, the mean bias of δz is <0.01 and the median error in $|\delta z|/z$ is within 15%. These numbers are very similar to those obtained in the GAMA equatorial fields in the test phase summarized in Table 2. On the other hand, the apparently worse results for the G10/COSMOS, PRIMUS, and COSMOS catalogs should be taken lightly. These data sets have either a very low matching rate with our catalog (being very deep but covering small areas), and/or have low-resolution spectroscopy, or provide only photometric redshifts, even if they have very high accuracy. In particular, the PRIMUS scatters in the spectroscopic redshifts of ~0.005 and the COSMOS scatter in photo-z of ~0.007 are comparable to the mean bias of *WISE* × SCOS photo-zs and are less than an order of magnitude smaller than the SMAD of the latter. Thus for these samples, the comparison of the "true" redshift with the *WISE* × SCOS photometric one brings in uncertainty in the former parameter, which leads to apparent deterioration of the photo-z statistics.

6. SUMMARY, CONCLUSIONS, FUTURE PROSPECTS

In this paper we presented a novel photometric redshift galaxy catalog based on the two largest existing all-sky photometric surveys, *WISE* and SuperCOSMOS. A union of these two samples, once cleaned of stellar contamination, provides access to redshifts of $z < 0.4$ on unmatched angular scales. Its angular coverage ($\simeq 3\pi$ sr) is a major advance with respect to existing surveys covering these redshifts (e.g., D'Abrusco et al. 2007; Oyaizu et al. 2008; Brescia et al. 2014; Beck et al. 2016).

We envisage manifold possible applications of our catalog, including being able to improve the statistics on analyses of shallower data sets, such as 2MPZ (Appleby & Shafieloo 2014; Xu et al. 2014; Alonso et al. 2015), 2MRS and 2MASS (Gibelyou & Huterer 2012), or the 2MASS PSC—*WISE* combination (Yoon et al. 2014). Our catalog can also be used as a testbed for current or forthcoming more precise and deeper wide-angle samples, such as from DES, SKA, or Euclid. A particularly interesting class of application involves "tomography": slicing the data set into redshift bins. Cross-correlations with other wide-angle astrophysical probes at various wavelengths should be especially fruitful, owing to their insensitivity to any remaining small systematics. Such analyses include, for instance, CMB temperature maps for the integrated Sachs–Wolfe effect searches (e.g., Giannantonio et al. 2008; Francis & Peacock 2010), CMB lensing measurements to constrain non-Gaussianity (Giannantonio & Percival 2014) or neutrino mass (Pearson & Zahn 2014), or the gamma-ray background provided by the Fermi satellite to constrain the sources of this emission (Xia et al. 2015) or to search for dark matter (Cuoco et al. 2015). In addition, we expect the *WISE* × SCOS sample to be useful for studies on Faraday rotation of extragalactic sources (Vacca et al. 2015), identification of galaxies in the SKA pathfinder WALLABY (Popping et al. 2012), or in planned CMB missions such as CoRE+ (De Zotti et al. 2015). It should also be appropriate in searches for electromagnetic counterparts of extragalactic gravitational wave sources, because—together with 2MPZ (cf. Antolini & Heyl 2016)—it extends well beyond the catalogs currently used for that purpose (e.g., White et al. 2011). In addition, both 2MPZ and the present catalog provide two crucial parameters for such studies: the B-band magnitude (a proxy for black hole and neutron star merger rates), and the $W1$ magnitude, which is directly related to the galaxy's stellar mass. In the near future, its bright end ($R \lesssim 18$) may be employed as one of the input catalogs for the forthcoming TAIPAN survey (Kuehn et al. 2014).

The fact that the median redshift of *WISE* galaxies is much higher than that of SCOS (Section 3; T. H. Jarrett et al. 2016, in preparation) makes it desirable to extend the present analysis beyond the latter sample. However, *WISE* on its own will not allow for precise photometric redshifts, because at its full depth it provides only two mid-IR bands. To obtain photo-z coverage beyond the SDSS area, it will be necessary to combine *WISE* with forthcoming catalogs, such as Pan-STARRS (Kaiser et al. 2002) or the VISTA Hemisphere Survey (McMahon et al. 2013). A supplementary approach to derive redshift estimates for *WISE* could be that of Ménard et al. (2013), which is already being undertaken (A. Mendez 2016, private communication). One of the requirements for such studies to succeed will be the ability to reliably separate galaxies from stars and quasars in *WISE*. A report on ongoing machine-learning efforts toward this goal is presented in Kurcz et al. (2016).

The *WISE* × SuperCOSMOS photometric redshift catalog is made publicly available through the Wide Field Astronomy Unit at the Institute for Astronomy, Edinburgh at http://ssa.roe.ac.uk/WISE×SCOS.

We thank the Wide Field Astronomy Unit at the Institute for Astronomy, Edinburgh, for archiving the *WISE* × SuperCOSMOS catalog.

Table 5
Statistics for the Photometric Redshift Estimation, Calculated for *WISE* × SuperCOSMOS Sources Cross-matched with External Redshift Catalogs

External Data set	Area (deg^2)	Coords. (α, δ)[a]	# of Sources[b]	Mean $\langle z \rangle$ Spec[c]	Mean $\langle z \rangle$ Phot[d]	Median $\bar{z}$ Spec[c]	Median $\bar{z}$ Phot[d]	1σ sc.[e] $\sigma_{\delta z/(1+z)}$	Scaled MAD[f]	Norm. SMAD[g]	Bias[h] $\langle \delta z \rangle$	Median Error[i]
lGAMA equatorial $\delta_{1950} > 2.5$	16.8	(135, 2.65) and (217.5, 2.65)	9,404	0.213	0.203	0.202	0.205	0.041	0.041	0.034	−9.9e−3	14.8%
G02	55.9	(34.5, −6.95)	17,812	0.211	0.205	0.207	0.207	0.039	0.038	0.031	−6.2e−3	13.4%
G23	87.2	(345, −32.5)	29,270	0.200	0.202	0.201	0.205	0.039	0.039	0.033	+1.6e−3	14.4%
AGES	10.4	(217.9, 35.9)	4,770	0.209	0.199	0.195	0.198	0.042	0.038	0.032	−9.7e−3	13.8%
SHELS	4.8	(139.9, 30)	2,272	0.222	0.208	0.216	0.209	0.043	0.043	0.035	−9.8e−3	14.6%
G10/COSMOS	2	(150.1, 2.2)	603	0.226	0.202	0.218	0.201	0.068	0.039	0.033	−2.4e−2	15.6%
PRIMUS[j]	~ 10	$-44 < \delta < 3$[k]	3,222	0.227	0.215	0.215	0.217	0.059	0.046	0.038	−1.2e−2	15.3%
COSMOS[l]	2	(150.1, 2.2)	918	0.214	0.207	0.220	0.209	0.045	0.045	0.037	−7.5e−3	15.8%

Notes. See text for details on the auxiliary data sets.

[a] Central coordinates of the field(s) in degrees.

[b] In the cross-match with the *WISE* × SCOS fiducial sample.

[c] External redshifts for a sample cross-matched with *WISE* × SCOS.

[d] *WISE* × SCOS photometric redshifts for a sample cross-matched with the external data set.

[e] Normalized 1σ scatter between the spectroscopic and photometric redshifts, $\sigma_{\delta z/(1\,+\,z)}$; unclipped.

[f] Scaled median absolute deviation, $\mathrm{SMAD}(\delta z) = 1.48 \times \mathrm{med}(|\delta z - \mathrm{med}(\delta z)|)$.

[g] Scaled median absolute deviation of the normalized bias, $\mathrm{SMAD}(\delta z/(1 + z_{\mathrm{spec}}))$.

[h] Mean bias of z_{phot}: $\langle \delta z \rangle = \langle z_{\mathrm{phot}} - z_{\mathrm{spec}} \rangle$; unclipped.

[i] Median of the relative error, $\mathrm{med}(|\delta z|/z_{\mathrm{spec}})$; unclipped.

[j] Low-resolution spectroscopy, $\sigma_{\delta z/(1\,+\,z)} = 0.005$.

[k] Nine fields over this declination range.

[l] Accurate photometric redshifts, $\sigma_{\delta z/(1\,+\,z)} = 0.007$.

We acknowledge assistance from Eddie Schlafly with extinction corrections. We also benefited from discussions with Malcom Fairbairn, Robert Hogan, and Mike Read.

Special thanks to Mark Taylor for the TOPCAT (Taylor 2005) and STILTS (Taylor 2006) software, and for his assistance. Some of the results in this paper were derived using the HEALPix package (Górski et al. 2005).

Financial assistance from the South African National Research Foundation (NRF) toward this research is hereby acknowledged. M Bilicki was supported by the Netherlands Organization for Scientific Research, NWO, through grant number 614.001.451, and through FP7 grant number 279396 from the European Research Council. M Bilicki and A Solarz were partially supported by the Polish National Science Center under contract #UMO-2012/07/D/ST9/02785. A Solarz was supported by the Polish National Science Center under contract #UMO-2015/16/S/ST9/00438.

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 has made use of data obtained from the SuperCOSMOS Science Archive, prepared and hosted by the Wide Field Astronomy Unit, Institute for Astronomy, University of Edinburgh, which is funded by the UK Science and Technology Facilities Council.

GAMA is a joint European-Australasian project based around a spectroscopic campaign using the Anglo-Australian Telescope. The GAMA input catalog is based on data taken from the Sloan Digital Sky Survey and the UKIRT Infrared Deep Sky Survey. Complementary imaging of the GAMA regions is being obtained by a number of independent survey programs including *GALEX* MIS, VST KiDS, VISTA VIKING, *WISE*, Herschel-ATLAS, GMRT, and ASKAP providing UV to radio coverage. GAMA is funded by the STFC (UK), the ARC (Australia), the AAO, and the participating institutions. The GAMA website is http://www.gama-survey.org/.

Funding for SDSS-III was provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, and the U.S. Department of Energy Office of Science. The SDSS-III website is http://www.sdss3.org/. SDSS-III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration including the University of Arizona, the Brazilian Participation Group, Brookhaven National Laboratory, Carnegie Mellon University, University of Florida, the French Participation Group, the German Participation Group, Harvard University, the Instituto de Astrofisica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max Planck Institute for Astrophysics, Max Planck Institute for Extraterrestrial Physics, New Mexico State University, New York University, Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia, University of Washington, and Yale University.

APPENDIX
POSITION-DEPENDENT CUTS TO REMOVE STARS FROM THE CROSS-MATCHED SAMPLE

Here we provide the details on the cuts applied to our data set to clean it from stellar contamination (Section 4.2). This procedure was then followed by more sophisticated masking of problematic areas that persisted after the cleanup procedure (Section 4.3).

We start with the modified longitude given by

$$\ell_{\mathrm{mod}} = \begin{cases} \ell & \text{for} \quad 0^\circ \leqslant \ell \leqslant 180^\circ, \\ 180^\circ - \ell & \text{for} \quad \ell > 180^\circ. \end{cases} \tag{5}$$

Next we define the limiting latitude for the Bulge cutout (in degrees) as

$$b_{\mathrm{Bulge}} = 6 + \frac{11}{1 + (\ell_{\mathrm{mod}}/60)^2}. \tag{6}$$

Anything with $|b| < b_{\mathrm{Bulge}}$ is removed from the sample. This limiting latitude would go from $|b| = 17^\circ$ at the GC to $|b| = 7^\circ$ at the Anticenter; however, in our sample we have only $|b| > 10^\circ$ sources, so this cut is effective up to $\sim 80^\circ$ in longitude from the Center.

In addition, we masked out by hand the three most prominent nearby galaxies (LMC, SMC, and M31) by applying circular cuts of radius 8°, 2°, and 2°, respectively.

The next step is to define a position-dependent color cut for star removal, such that it would be equal to $W1 - W2 = 0$ at high Galactic latitudes and be gradually increased to $W1 - W2 = 0.12$ near the Galactic Plane. Using

$$b_{\mathrm{lim}} = 5 + \frac{10}{1 + (\ell_{\mathrm{mod}}/60)^2} \tag{7}$$

and taking $\Delta b = |b| - b_{\mathrm{lim}}$, we define the following threshold:

$$W12_{\mathrm{lim}} = 0.12 \exp\left[-\left(\frac{\Delta b}{15}\right)^2\right]. \tag{8}$$

Anything with $W1 - W2 < W12_{\mathrm{lim}}$ is removed from the sample. The results of this procedure are illustrated in Figure 7 (Section 4).

REFERENCES

Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, ApJS, 182, 543
Abdalla, F. B., Banerji, M., Lahav, O., & Rashkov, V. 2011, MNRAS, 417, 1891
Ahn, C. P., Alexandroff, R., Allende Prieto, C., et al. 2014, ApJS, 211, 17
Aihara, H., Allende Prieto, C., An, D., et al. 2011, ApJS, 193, 29
Alam, S., Albareti, F. D., Allende Prieto, C., et al. 2015, ApJS, 219, 12
Alonso, D., Salvador, A. I., Sánchez, F. J., et al. 2015, MNRAS, 449, 670
Antolini, E., & Heyl, J. S. 2016, arXiv:1602.07710
Appleby, S., & Shafieloo, A. 2014, JCAP, 10, 70
Ascaso, B., Mei, S., & Benítez, N. 2015, MNRAS, 453, 2515
Baldry, I. K., Robotham, A. S. G., Hill, D. T., et al. 2010, MNRAS, 404, 86
Baldry, I. K., Alpaslan, M., Bauer, A. E., et al. 2014, MNRAS, 441, 2440
Beck, R., Dobos, L., Budavári, T., Szalay, A. S., & Csabai, I. 2016, arXiv:1603.09708
Benitez, N., Dupke, R., Moles, M., et al. 2014, arXiv:1403.5237
Bilicki, M., Chodorowski, M., Jarrett, T., & Mamon, G. A. 2011, ApJ, 741, 31
Bilicki, M., Jarrett, T. H., Peacock, J. A., Cluver, M. E., & Steward, L. 2014, ApJS, 210, 9
Bolton, A. S., Schlegel, D. J., Aubourg, É., et al. 2012, AJ, 144, 144
Branchini, E., Davis, M., & Nusser, A. 2012, MNRAS, 424, 472
Brescia, M., Cavuoti, S., Longo, G., & De Stefano, V. 2014, A&A, 568, A126
Brown, M. J. I., Jarrett, T. H., & Cluver, M. E. 2014, PASA, 31, 49
Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, ApJ, 345, 245
Carrick, J., Turnbull, S. J., Lavaux, G., & Hudson, M. J. 2015, MNRAS, 450, 317
Christodoulou, L., Eminian, C., Loveday, J., et al. 2012, MNRAS, 425, 1527
Cluver, M. E., Jarrett, T. H., Hopkins, A. M., et al. 2014, ApJ, 782, 90
Coil, A. L., Blanton, M. R., Burles, S. M., et al. 2011, ApJ, 741, 8
Colless, M., Dalton, G., Maddox, S., et al. 2001, MNRAS, 328, 1039
Collister, A. A., & Lahav, O. 2004, PASP, 116, 345
Comparat, J., Chuang, C. H., Rodriguez-Torres, S., et al. 2016, MNRAS, in press (arXiv:1510.00147)
Cool, R. J., Moustakas, J., Blanton, M. R., et al. 2013, ApJ, 767, 118
Cuoco, A., Xia, J.-Q., Regis, M., et al. 2015, ApJS, 221, 29
Cutri, R. M., et al. 2012, Explanatory Supplement to the WISE All-Sky Data Release Products, Tech. Rep.
Cutri, R. M., et al. 2013, Explanatory Supplement to the AllWISE Data Release Products, Tech. Rep.
D'Abrusco, R., Staiano, A., Longo, G., et al. 2007, ApJ, 663, 752
Davies, L. J. M., Driver, S. P., Robotham, A. S. G., et al. 2015, MNRAS, 447, 1014
Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2013, AJ, 145, 10
de Jong, J. T. A., Verdoes Kleijn, G. A., Kuijken, K. H., & Valentijn, E. A. 2013, ExA, 35, 25
De Zotti, G., Castex, G., González-Nuevo, J., et al. 2015, JCAP, 6, 18
DiPompeo, M. A., Myers, A. D., Hickox, R. C., Geach, J. E., & Hainline, K. N. 2014, MNRAS, 442, 3443
Donoso, E., Yan, L., Stern, D., & Assef, R. J. 2014, ApJ, 789, 44
Doré, O., Bock, J., Ashby, M., et al. 2014, arXiv:1412.4872
Driver, S. P., Davies, L. J., Meyer, M., et al. 2015, arXiv:1507.00676
Driver, S. P., Norberg, P., Baldry, I. K., et al. 2009, A&G, 50, 12
Driver, S. P., Hill, D. T., Kelvin, L. S., et al. 2011, MNRAS, 413, 971
Eardley, E., Peacock, J. A., McNaught-Roberts, T., et al. 2015, MNRAS, 448, 3665
Eisenstein, D. J., Weinberg, D. H., Agol, E., et al. 2011, AJ, 142, 72
Ferraro, S., Sherwin, B. D., & Spergel, D. N. 2015, PhRvD, 91, 083533
Firth, A. E., Lahav, O., & Somerville, R. S. 2003, MNRAS, 339, 1195
Fitzpatrick, E. L. 1999, PASP, 111, 63
Francis, C. L., & Peacock, J. A. 2010, MNRAS, 406, 2
Geller, M. J., Hwang, H. S., Fabricant, D. G., et al. 2014, ApJS, 213, 35
Giannantonio, T., & Percival, W. J. 2014, MNRAS, 441, L16
Giannantonio, T., Scranton, R., Crittenden, R. G., et al. 2008, PhRvD, 77, 123520
Gibelyou, C., & Huterer, D. 2012, MNRAS, 427, 1994
Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759
Goto, T., Szapudi, I., & Granett, B. R. 2012, MNRAS, 422, L77
Hambly, N. C., Davenhall, A. C., Irwin, M. J., & MacGillivray, H. T. 2001a, MNRAS, 326, 1315
Hambly, N. C., Irwin, M. J., & MacGillivray, H. T. 2001b, MNRAS, 326, 1295
Hambly, N. C., MacGillivray, H. T., Read, M. A., et al. 2001c, MNRAS, 326, 1279
Heymans, C., Van Waerbeke, L., Miller, L., et al. 2012, MNRAS, 427, 146
Hogan, R., Fairbairn, M., & Seeburn, N. 2015, MNRAS, 449, 2040
Hopkins, A. M., Driver, S. P., Brough, S., et al. 2013, MNRAS, 430, 2047
Hubble, E. 1934, ApJ, 79, 8
Huchra, J. P., Macri, L. M., Masters, K. L., et al. 2012, ApJS, 199, 26
Ilbert, O., Capak, P., Salvato, M., et al. 2009, ApJ, 690, 1236
Indebetouw, R., Mathis, J. S., Babler, B. L., et al. 2005, ApJ, 619, 931
Jarrett, T. 2004, PASA, 21, 396
Jarrett, T. H., Chester, T., Cutri, R., et al. 2000, AJ, 119, 2498
Jarrett, T. H., Cohen, M., Masci, F., et al. 2011, ApJ, 735, 112
Jones, D. H., Saunders, W., Colless, M., et al. 2004, MNRAS, 355, 747
Kaiser, N., Aussel, H., Burke, B. E., et al. 2002, Proc. SPIE, 4836, 154
Kochanek, C. S., Eisenstein, D. J., Cool, R. J., et al. 2012, ApJS, 200, 8
Kovács, A., & Szapudi, I. 2015, MNRAS, 448, 1305
Kuehn, K., Lawrence, J., Brown, D. M., et al. 2014, Proc. SPIE, 9147, 10
Kurcz, A., Bilicki, M., Solarz, A., et al. 2016, arXiv:1604.04229
Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv:1110.3193
Levi, M., Bebek, C., Beers, T., et al. 2013, arXiv:1308.0847
Lewis, A., & Challinor, A. 2006, PhR, 429, 1
Lilly, S. J., Le Fèvre, O., Renzini, A., et al. 2007, ApJS, 172, 70
Lilly, S. J., Le Brun, V., Maier, C., et al. 2009, ApJS, 184, 218
Lima, M., Cunha, C. E., Oyaizu, H., et al. 2008, MNRAS, 390, 118
Liske, J., Baldry, I. K., Driver, S. P., et al. 2015, MNRAS, 452, 2087
LSST Science Collaboration et al. 2009, arXiv:0912.0201
Madsen, G. J., & Gaensler, B. M. 2013, ApJS, 209, 33
Martí, P., Miquel, R., Castander, F. J., et al. 2014, MNRAS, 442, 92
Mateos, S., Alonso-Herrero, A., Carrera, F. J., et al. 2012, MNRAS, 426, 3271
McMahon, R. G., Banerji, M., Gonzalez, E., et al. 2013, Msngr, 154, 35
Ménard, B., Scranton, R., Schmidt, S., et al. 2013, arXiv:1303.4722
Newman, J. A. 2008, ApJ, 684, 88
Nishizawa, A. J. 2014, PTEP, 2014, 060000
Oyaizu, H., Lima, M., Cunha, C. E., et al. 2008, ApJ, 674, 768
Pearson, R., & Zahn, O. 2014, PhRvD, 89, 043516
Popping, A., Jurek, R., Westmeier, T., et al. 2012, PASA, 29, 318

Rahman, M., Ménard, B., & Scranton, R. 2016a, MNRAS, 457, 3912

Rahman, M., Ménard, B., Scranton, R., Schmidt, S. J., & Morrison, C. B. 2015, MNRAS, 447, 3500

Rahman, M., Mendez, A. J., Ménard, B., et al. 2016b, MNRAS, 460, 163

Richards, G. T., Strauss, M. A., Fan, X., et al. 2006, AJ, 131, 2766

Robotham, A., Driver, S. P., Norberg, P., et al. 2010, PASA, 27, 76

Sadeh, I., Abdalla, F. B., & Lahav, O. 2015, arXiv:1507.00490

Sánchez, C., Carrasco Kind, M., Lin, H., et al. 2014, MNRAS, 445, 1482

Saunders, W., Sutherland, W. J., Maddox, S. J., et al. 2000, MNRAS, 317, 55

Schlafly, E. F., & Finkbeiner, D. P. 2011, ApJ, 737, 103

Schlafly, E. F., Finkbeiner, D. P., Schlegel, D. J., et al. 2010, ApJ, 725, 1175

Schlafly, E. F., Green, G., Finkbeiner, D. P., et al. 2014, ApJ, 789, 15

Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525

Stern, D., Assef, R. J., Benford, D. J., et al. 2012, ApJ, 753, 30

Taylor, E. N., Hopkins, A. M., Baldry, I. K., et al. 2011, MNRAS, 418, 1587

Taylor, M. B. 2005, in ASP Conf. Ser. 347, Astronomical Data Analysis Software and Systems XIV, ed. P. Shopbell, M. Britton, & R. Ebert (San Francisco, CA: ASP), 29

Taylor, M. B. 2006, in ASP Conf. Ser. 351, Astronomical Data Analysis Software and Systems XV, ed. C. Gabriel et al. (San Francisco, CA: ASP), 666

The Dark Energy Survey Collaboration 2005, arXiv:astro-ph/0510346

Vacca, V., Oppermann, N., Ensslin, T., et al. 2015, arXiv:1509.00747

White, D. J., Daw, E. J., & Dhillon, V. S. 2011, CQGra, 28, 085016

Wittman, D., Dell'Antonio, I. P., Hughes, J. P., et al. 2006, ApJ, 643, 128

Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, AJ, 140, 1868

Xia, J.-Q., Cuoco, A., Branchini, E., & Viel, M. 2015, ApJS, 217, 15

Xu, W., Wen, Z., & Han, J. 2014, SCPMA, 57, 2168

Yan, L., Donoso, E., Tsai, C.-W., et al. 2013, AJ, 145, 55

Yoon, M., & Huterer, D. 2015, ApJL, 813, L18

Yoon, M., Huterer, D., Gibelyou, C., Kovács, A., & Szapudi, I. 2014, MNRAS, 445, L60

York, D. G., Adelman, J., Anderson, J. E., Jr., et al. 2000, AJ, 120, 1579