

Towards automatic classification of all WISE sources

A. Kurcz^{1,2}, M. Bilicki^{3,2,4}, A. Solarz^{5,2}, M. Krupa^{1,2}, A. Pollo^{1,5,2} and K. Małek^{5,2}

¹ Astronomical Observatory of the Jagiellonian University, ul.Orla 171, 30-244 Cracow, Poland
e-mail: kurcz.agnieszka@gmail.com

² Janusz Gil Institute of Astronomy, University of Zielona Góra, ul. Szafrana 2, 65-516 Zielona Góra, Poland

³ Leiden Observatory, Leiden University, Niels Bohrweg 2, NL-2333 CA Leiden, the Netherlands

⁴ Astrophysics, Cosmology and Gravity Centre, Department of Astronomy, University of Cape Town, Rondebosch, South Africa

⁵ National Centre for Nuclear Research, ul.Hoża 69, 00-681 Warszawa, Poland

Received ; accepted

ABSTRACT

Context. The Wide-field Infrared Survey Explorer (WISE) has detected hundreds of millions sources over the entire sky. Classifying them reliably is however a challenging task due to degeneracies in WISE multicolour space and low levels of detection in its two longest-wavelength bandpasses. Simple colour cuts are often not sufficient and for satisfactory levels of completeness and purity, more sophisticated classification methods are needed.

Aims. Here we aim at obtaining comprehensive and reliable star, galaxy and quasar catalogues based on automatic source classification in full-sky WISE data. This means that the final classification will employ only parameters available from WISE itself, in particular those which are reliably measured for a majority of sources.

Methods. For the automatic classification we applied a supervised machine learning algorithm, support vector machines (SVM). It requires a training sample with relevant classes already identified, and we chose to use the SDSS spectroscopic dataset (DR10) for that purpose. We tested the performance of two kernels used by the classifier, determined the minimum number of sources in the training set required to achieve stable classification, as well as the minimum dimension of the parameter space. We also tested SVM classification accuracy as a function of extinction and apparent magnitude. Thus calibrated classifier was finally applied to all-sky WISE data, flux-limited to 16 mag (Vega) in the 3.4 μm channel.

Results. By calibrating on the test data drawn from SDSS, we first established that a polynomial kernel is preferred over a radial one for this particular dataset. Next, using three classification parameters ($W1$ magnitude, $W1 - W2$ colour, and a differential aperture magnitude) we obtained very good classification efficiency in all the tests. At the bright end, the completeness for stars and galaxies reaches $\sim 95\%$, deteriorating to $\sim 80\%$ at $W1 = 16$ mag, while for quasars it stays at a level of $\sim 95\%$ independently of magnitude. Similar numbers are obtained for purity. Application of the classifier to full-sky WISE data and appropriate a posteriori cleaning allowed us to obtain reliably-looking catalogues of star and galaxy candidates. However, the sources flagged by the classifier as ‘quasars’ are in fact dominated by dusty galaxies but also exhibit contamination from sources located mainly at low ecliptic latitudes, consistent with Solar System objects.

Key words. Cosmology: observations, infrared; methods: statistical, surveys; automatic classification: galaxies, stars, quasars

1. Introduction

The Wide-field Infrared Survey Explorer (WISE, Wright et al. 2010) is a space-borne telescope that scanned the entire sky in four infrared (IR) bands (3.4 – 23 μm) and delivered one of the largest catalogues of astronomical objects to date. It detected almost 750 million sources, compiled in the publicly released AllWISE Source Catalogue (Cutri et al. 2013). WISE provides at present the most comprehensive census of the entire sky in the IR, and offers large advancement in comparison to earlier all-sky IR surveys, such as IRAS (Neugebauer et al. 1984), 2MASS (Skrutskie et al. 2006) or AKARI (Murakami et al. 2007).

Such a vast amount of data, giving access to all-sky information in unprecedented volumes, has found multiple astronomical applications, starting from our closest neighbourhood (Near-Earth Objects, nearby stars and brown dwarfs), through the galaxies in the local volume, up to the largest possible distances of high-redshift quasars (Wright et al. 2010). All such types of source are present in the WISE database (within its sensitivity limits) but reliably extracting them in large numbers is challenging. Shortly speaking, WISE data release does not pro-

vide separate catalogues of different objects. What is more, at present there is even no separate point- and extended-source catalogues extracted from this survey, although efforts towards the latter are underway (Cluver et al. 2014).

There are several reasons for the lack of comprehensive object identification in WISE. Firstly, this survey is mostly a near-IR selected one. Practically all the sources listed in the WISE catalogue have $S/N > 2$ detections in the $W1$ band (3.4 μm), and 83% of them have $W2$ (4.6 μm) measured with this accuracy; the detection rates in $W3$ (12 μm) and $W4$ (23 μm) are much lower¹. The light emitted at 3.4 and 4.6 μm comes mainly from the photospheres of evolved stars. This means that in the low-redshift universe, where a large part of extragalactic WISE sources are located, the $W1 - W2$ colour of galaxies will be very similar to that of Galactic stars, and cannot in general provide a comprehensive criterion to separate out ones from the others. This is readily seen in WISE colour-colour diagrams such as provided by Jarrett et al. (2011). These diagrams also show that

¹ http://wise2.ipac.caltech.edu/docs/release/allwise/expsum/sec2_1.html#stats

even adding the W3 band, which traces dust such as silicates and PAHs, is not sufficient to unambiguously separate out stars from elliptical galaxies. Last but not least, such colour-colour plots assume negligible photometric errors, while in reality significant scatter will occur, as is the case for the $W3$ and $W4$ bands, for which most of the WISE sources have only upper limits or are not even detected.

Another hope to separate out galaxies from stars in WISE could be by identifying extended sources in the sample. Similarly as in the case of 2MASS Extended Source Catalogue (XSC, Jarrett et al. 2000), such sources in WISE are expected to be mainly extragalactic, especially at sufficiently high Galactic latitudes. However, despite much better sensitivity of WISE with respect to 2MASS (e.g. 0.054 mJy in $W1$ vs. 2.7 mJy in K_s) and lack of atmospheric nuisances in the former, lower angular resolution of WISE and higher background levels mean that the eventual all-sky WISE XSC is expected to contain a similar number of sources as the 2MASS one does (~ 30 per square degree; Jarrett et al., submitted). This will be a very small percentage of all WISE galaxies: Cluver et al. (2014) showed that only 3–4% of WISE sources matched with the Galaxy And Mass Assembly (GAMA, Driver et al. 2011) survey are resolved in $W1$, while already the WISE×GAMA sample itself is much shallower than what one can expect from the full WISE galaxy catalogue (Bilicki et al., submitted; Jarrett et al., submitted). One possible avenue to obtain a WISE-based all-sky catalogue of galaxies of a similar depth as that of GAMA is to cross-match WISE sources with SuperCOSMOS data (Hambly et al. 2001), as was first discussed by Bilicki et al. (2014). Such a catalogue has been recently compiled (Bilicki et al., submitted; Krakowski et al., in prep.), but it includes only a part of WISE galaxies due to limitations of the SuperCOSMOS scans of photographic plates (both in depth and in the colour space).

Until now, most of the studies dealing with WISE source classification were based on cross-matching this catalogue with other samples and using multi-band magnitudes and colours as discriminants. Stern et al. (2012), having paired-up WISE with COSMOS, proposed to use $W1 - W2 \geq 0.8$ mag (Vega) to identify WISE active galactic nuclei (AGNs). Assef et al. (2013) extended this work to a larger and deeper NOAO Deep Wide-Field Survey Boötes field and showed that this criterion is no longer optimal at fainter magnitudes. A more comprehensive effort was undertaken by Yan et al. (2013), where WISE sources were cross-matched with SDSS to derive colour cuts for object selection. The Stern et al. (2012) AGN identification was confirmed and WISE colours (especially $W1 - W2$ vs. $W2 - W3$) were shown to be sufficient for separating out star-forming galaxies from AGNs, as well as stars from some galaxies. The latter however was not the case for early-type, low redshift galaxies, which occupy practically the same region in the $W1 - W2 - W3$ colour-colour space as stars. More recently, Ferraro et al. (2015) defined their own colour cuts to identify galaxies and quasars from the WISE database, which however left position-dependent contamination visible in all-sky maps. Nikutta et al. (2014) explored WISE colours of Galactic and other nearby sources, while Mateos et al. (2012), used XMM-Newton survey to define a WISE colour-based selection of luminous AGN. The latter criterion has recently been applied to the all-sky WISE data by Secrest et al. (2015) to select a sample of 1.4 million AGN candidates. Note however that their criterion of $w1, 2, 3snr \geq 5$ (signal-to-noise ratios in $W1$, $W2$ and $W3$) eliminates about 95% of AllWISE sources from the parent catalogue, mostly due to a very low level of WISE detection in the $W3$ channel. Last but not least, Jarrett et al. (submitted) identified various source types in the G12 equa-

torial field by calibrating WISE magnitude and colour cuts on GAMA and SDSS spectroscopic data.

Some other WISE source classification studies, where additional colours from external surveys were used, include Edelson & Malkan (2012); Wu et al. (2012) for QSOs/AGNs, Tu & Wang (2013) for asymptotic giant branch stars and Kovács & Szapudi (2015) for general star-galaxy separation in a WISE – 2MASS PSC cross-match. Last but not least, Anderson et al. (2014) compiled a catalogue of Galactic H II regions from WISE, based on their mid-IR morphology.

In the present paper, we go beyond simple colour- and magnitude cuts and explore more sophisticated classification of WISE sources. Our approach is based on automated procedures of machine learning, and we use a specific algorithm – the support vector machines (SVM) – which has proven its aptitude for similar tasks within AKARI (Solarz et al. 2012) and VIPERS (Małek et al. 2013) surveys. A similar idea was explored in the independent study by Kovács & Szapudi (2015), where multi-band photometry of those WISE sources that had cross-matches with the 2MASS Point Source Catalogue (PSC) was used for SVM-based source classification. Our analysis is however more general, as we do not limit the final source selection to a cross-match with an external catalogue. In addition, for the classification we use only the two shortest WISE bands in order to retain the highest all-sky completeness possible. By training the classifier on WISE data cross-matched with the 10th spectroscopic release of the Sloan Digital Sky Survey (SDSS, Ahn et al. 2014), we make the first step towards building reliable and comprehensive WISE catalogues of stars, galaxies and AGNs/QSOs. At present, this classification is limited by the spectroscopic data that we used for training, but the methodology can be further extended with forthcoming data from different star, galaxy and quasar catalogues. To these, one should probably add also samples of Solar System bodies, as we find hints of them contaminating especially our final quasar candidate dataset.

The paper is organised as follows. Section 2 describes the data used in our analysis: the photometric sample extracted from WISE (Sec. 2.1) and the spectroscopic one from SDSS (Sec. 2.2) used by the classification algorithm. In Section 3, we present the support vector machines classifier and how to apply it to imbalanced data sets (Sec. 3.1), such as ours. Various tests of the SVM method on WISE data are shown in Section 4. Section 5 presents the application of the SVM classifier to a full-sky sample drawn from WISE data. We summarize our work in Section 6.

All the WISE magnitudes in this paper will be given in the Vega system. For transformation to AB see Jarrett et al. (2011).

2. Data selection

2.1. WISE

The WISE is a NASA-funded Medium Class Explorer mission launched in December 2009. With the use of a 40 cm space-based telescope, WISE mapped the whole sky (with a total 47x47 arcmin field of view) in four infrared bands $W1 - W4$, centred respectively at² 3.4, 4.6, 12 and 23 μm , with an angular resolution of 6.1", 6.4", 6.5" and 12.0", respectively. Its 5σ point source sensitivities exceeded 0.054, 0.071, 0.73 and 5 mJy in the four respective bands and for instance in the $W3$ channel, this is more than a hundred times better than that of IRAS at similar wavelengths. The publicly available AllWISE catalogue contains positional, photometric, quality and reliability information

² Note the recalibration of the $W4$ effective wavelength from 22 μm by Brown et al. (2014a).

and motion fit parameters for over 747 million sources (Cutri et al. 2013).

The eventual goal of the present study is to obtain a comprehensive source classification for as many WISE objects as practicable. For that reason, for object selection we have decided to rely uniquely on the parameters provided in the WISE database, as there is no other all-sky survey of a comparable depth currently available. The basic dataset we employ is the ‘AllWISE’ data release³ (Cutri et al. 2013), which was made publicly available in 2013 and combines data from the WISE cryogenic and NEOWISE (Mainzer et al. 2014) post-cryogenic survey stages. This dataset offers enhanced photometry and astrometry in comparison to the earlier WISE ‘All-Sky’ release and adds, inter alia, estimates of source apparent motions.

In our work we aim at as large sky coverage and depth as possible, we have thus chosen to be permissive in the preliminary source selection for our catalogue. As far as photometry is concerned, we use only the two shortest WISE bands, $W1$ and $W2$, and we employ additional quality parameters to ensure reliability of the sources. Our catalogue includes the sources that match the following criteria in the WISE database: $w1snr \geq 5$, $w2snr \geq 2$; $w?sat \leq 0.1$ (no more than 10% of saturated pixels in the respective bands, where ? stands for 1 or 2); $cc_flags[?] \neq ‘DPHO’$ (no severe artefacts). Such a selection ensures that the sources are detected in the two bands with reliable photometry (most of the objects have S/N much higher than the limits used for the preselection – see below). There are over 606 million such sources in WISE; however, a large number of these are concentrated in the Galactic Plane, where the WISE data suffer from severe blending and saturation due to the enhanced source density. Our ability to classify data at low Galactic latitudes will be thus very much compromised, and practically impossible within the Galactic Plane and Bulge.

An important caveat here is that the AllWISE catalogue is not complete at the very bright end ($W1 < 8$ mag and $W2 < 7$ mag) due to saturation of such sources, as well as in two strips at ecliptic longitudes of $45^\circ < \lambda < 55^\circ$ and $231^\circ < \lambda < 239^\circ$. Both these issues are related to instrumental limitations⁴, while the survey strategy causes additional patterns related to Moon avoidance manoeuvres⁵. This will be reflected in the eventual all-sky maps that we are aiming to produce.

In the present paper we will not be able to comprehensively classify all the WISE sources preselected as discussed above due to the limitations brought about by the training set from the spectroscopic SDSS DR10 that we use. Namely, that dataset cross-matched with WISE practically does not provide galaxies fainter than Vega $W1 < 16$ mag⁶. This is one magnitude brighter than the average all-sky photometric completeness of WISE, so the present analysis will need to be extended once deeper training data become available. This should be possible in the coming years thanks to the plethora of spectroscopic surveys being currently undertaken. The all-sky sample of $W1 < 16$ mag AllWISE sources preselected as above includes 314 million sources and is illustrated in the Aitoff projection in Fig. 1. Note the logarithmic

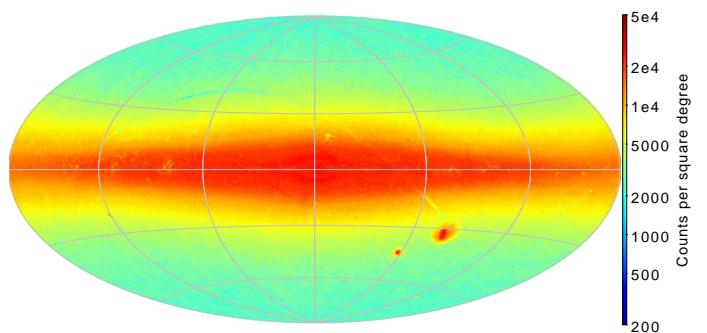


Fig. 1. Aitoff projection in Galactic coordinates of 314 million sources in the AllWISE catalogue, flux-limited to $W1 < 16$ mag.

scaling of the counts and one order of magnitude larger source density in the Galactic Plane than at high latitudes. We note that at this flux limit, most of our sources have very reliable photometry, especially in the $W1$ channel. The median signal-to-noise ratios in $W1$ and $W2$ are respectively 31.3 and 16, and more than 99% of the sources have magnitudes errors smaller than 0.08 mag for $W1$ and 0.28 mag for $W2$.

In the machine learning procedure of source classification, described later in the text, we have used the following parameters provided by WISE:

1. magnitude $w1_{\text{ipro}}$ measured with profile-fitting photometry in the $W1$ band (from now on referred to just as ‘ $W1$ ’);
2. colour $W1 - W2$ defined as the difference in the $w1_{\text{ipro}}$ and $w2_{\text{ipro}}$ profile-fitting magnitudes (the latter denoted as ‘ $W2$ ’ from now on);
3. a concentration parameter, defined as the difference of two circular aperture magnitudes in the $W1$ channel, $w1_{\text{mag_1}} - w1_{\text{mag_3}}$, measured respectively in radii 5.5” and 11” centred on the source; note that these apertures were fixed, independent of the actual sources size or shape, and not corrected for contamination or bad pixels, they cannot be thus used on their own as reliable measurements of fluxes for resolved sources.

As already mentioned, measurements in the $W1$ band in our flux-limited sample have typically very high signal-to-noise ratios; the other two parameters used for the classification are somewhat noisier. The error in the $W1 - W2$ colour is mostly driven by the less accurate $W2$ channel, and respectively 90% (99%) of the sources have $\delta(W1 - W2) < 0.16$ mag (< 0.29 mag). For the concentration parameter, the same percentiles are $\delta(w1_{\text{mag_1}} - w1_{\text{mag_3}}) < 0.17$ mag (< 0.41 mag).

We have also tested the usefulness of apparent motions for source classification. These motions, as provided in the AllWISE database (pmra and pmdec), are composed of source proper motions and those due to the parallax. One can thus expect them to be different for various source types. Note however different caveats related to their measurements by WISE, discussed by Kirkpatrick et al. (2014) and in the data description⁷. The accuracy and signal-to-noise of AllWISE proper motions are strongly correlated with source’s flux, they cannot thus be reliably used for the whole catalogue (and about 2% of WISE objects have no motion measurements at all). We consider using them as a proof of concept for future datasets and surveys that will bring much more precise and comprehensive motion measurements, such as the MaxWISE proposal (Faherty et al. 2015), Gaia (Perryman

³ Available from the NASA/IPAC Infrared Science Archive at <http://irsa.ipac.caltech.edu/>.

⁴ For details see http://wise2.ipac.caltech.edu/docs/release/allwise/expsup/sec2_2.html#cat_phot and http://wise2.ipac.caltech.edu/docs/release/allwise/expsup/sec2_2.html#w1sat.

⁵ http://wise2.ipac.caltech.edu/docs/release/allwise/expsup/sec1_2.html#survey

⁶ The situation has not improved in the final SDSS-III Data Release 12 (Alam et al. 2015).

⁷ http://wise2.ipac.caltech.edu/docs/release/allwise/expsup/sec2_6.html

et al. 2001) or LSST (Ivezić et al. 2008). The fourth parameter used in such a case was:

4. apparent motion defined as $\text{pm} = (\text{pmra}^2 + \text{pmdec}^2)^{1/2}$, where pmra and pmdec are the apparent motion in right ascension and declination, respectively.

All the tests involving the apparent motions were carried out with the imposed condition on their ‘signal-to-noise’ being larger than 1: $\text{pm} > \text{sigpm}$, where $\text{sigpm} = (\text{sigpmra}^2 + \text{sigpmdec}^2)^{1/2}$ is the motion accuracy as provided in the database. This condition introduces a selection effect as it removes mostly faint, point-like sources. This constraint will be avoided in the final classification procedure when proper motions will not be used.

One can further increase the number of parameters used for machine learning classification, it should however be first noted that every new parameter considerably extends computation time. In addition, many of the WISE database parameters are available or sufficiently reliable only for a subset of all the sources. For instance, the two longer WISE bands, $W3$ and $W4$, which are often additionally used for source classification (Ferraro et al. 2015; Kovács & Szapudi 2015), have much worse sensitivity than $W1$ and $W2$. Most of the WISE sources are not detected at the longer wavelengths or have only upper limits of $S/N < 2$. In view of eventually classifying a considerable number out of all WISE objects (selected as discussed above), we have thus decided to limit ourselves to the basic information from the $W1$ and $W2$ bands as discussed above. A possible extension employing more WISE parameters would require first determining the most optimal of them, for instance thorough a principal component analysis (Soumagnac et al. 2015). In addition, similarly to other applications of SVM in astronomy, our study will not take explicitly into account the observational errors. For the present sample this should be a good approximation, as the noise level of the parameters we use for classification is relatively low (or even very low for the $W1$ magnitude). In Section 6 we however discuss how this issue can be dealt with in future work, by using the so-called *fuzzy logic*.

2.2. Training sample: WISE \times SDSS DR10

Machine learning methods for source classification, such as the one we employ here, rely on the availability of a *training sample* that has relevant classes already identified. Ideally, such a dataset should be as representative for the whole sample as possible. At present however, such samples drawn from WISE are not available, and the only solution is to cross-match the WISE data with an external dataset which has relevant source types listed. Such an auxiliary dataset is provided by the Sloan Digital Sky Survey (SDSS, York et al. 2000) which in its third phase (SDSS III, Eisenstein et al. 2011) encompasses several dedicated star, galaxy and quasar surveys. These three classes are however available only for the *spectroscopic* part of the SDSS (the photometric classes are only ‘stars’, i.e. point-like and ‘galaxies’, i.e. resolved). For that reason we have chosen to use only those SDSS sources that have spectra (Bolton et al. 2012). Here we use the spectroscopic sample in the SDSS Data Release 10 (DR10, Ahn et al. 2014), which includes almost 3.4 million sources, of which 26% are classified by the SDSS pipeline as stars, 59% are galaxies and the remaining 15% are quasars/AGNs (class ‘QSO’ in SDSS). We have cross-matched this sample with the WISE catalogue selected as described above, using a 1” matching radius, which gave us 2.1 million sources (18% stars, 72% galaxies and 10% QSOs). However, not all of them had SDSS spectra of

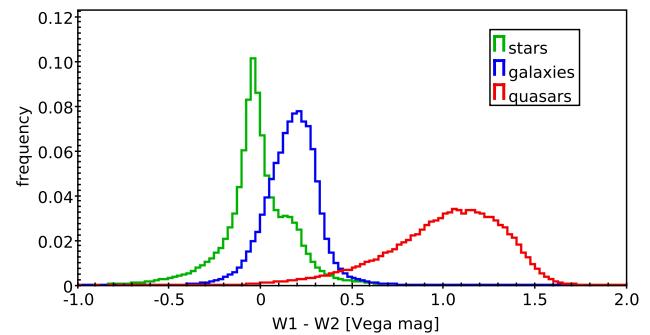


Fig. 2. Distributions of the observed $W1 - W2$ colour for stars, galaxies and quasars in the cross-matched WISExSDSS DR10 sample.

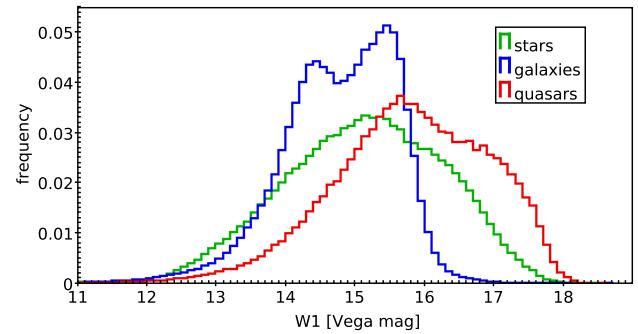


Fig. 3. Normalised apparent $W1$ magnitude counts for galaxies, quasars and stars in the WISExSDSS DR10 sample.

sufficient quality, so in order to maintain the reliability of the training sample, we have filtered the sources according to redshift (velocity) quality, keeping only those with $\text{zWarning} = 0$. Additional visual inspection of redshift and error distributions of WISExSDSS sources led us to eliminate the following outliers: $\text{zErr} > 0.001$ for stars, $\text{zErr} > 0.001$ or $\text{zErr}/\text{z} > 0.1$ for galaxies, and $\text{zErr} > 0.01$ for QSOs. This filtering left us with about 390,000 stars, 1.5 million galaxies and 190,000 quasars in our training sample (i.e. present both in SDSS and in WISE), which reduces further to 120,000 stars, 620,000 galaxies and 55,000 QSOs if the condition on the apparent motions to have $S/N > 1$ is applied (Sec 2.1).

Fig. 2 presents the $W1 - W2$ colour histogram of our training sources (as observed, i.e. without extinction- or k -corrections applied). It clearly shows that while a simple selection in this colour may allow to identify a large fraction of WISE quasars (although the $W1 - W2 > 0.8$ mag cut proposed by Stern et al. 2012 will miss some of them), it is not sufficient to reliably separate stars from galaxies. For instance, a constant $W1 - W2$ colour cut will not produce samples that are both complete and pure at the same time. Maps presented in Ferraro et al. (2015) indicate additionally that applying fixed colour cuts to WISE data may leave position-dependent contamination. In the WISExSuperCOSMOS dataset (Bilicki et al., submitted) this issue was partly alleviated by varying the star-galaxy colour separation as a function of distance from Galactic Centre. Here we move beyond this simple methodology.

As already mentioned in Sec. 2.1, using the SDSS as the training sample imposes restrictions on the depth up to which

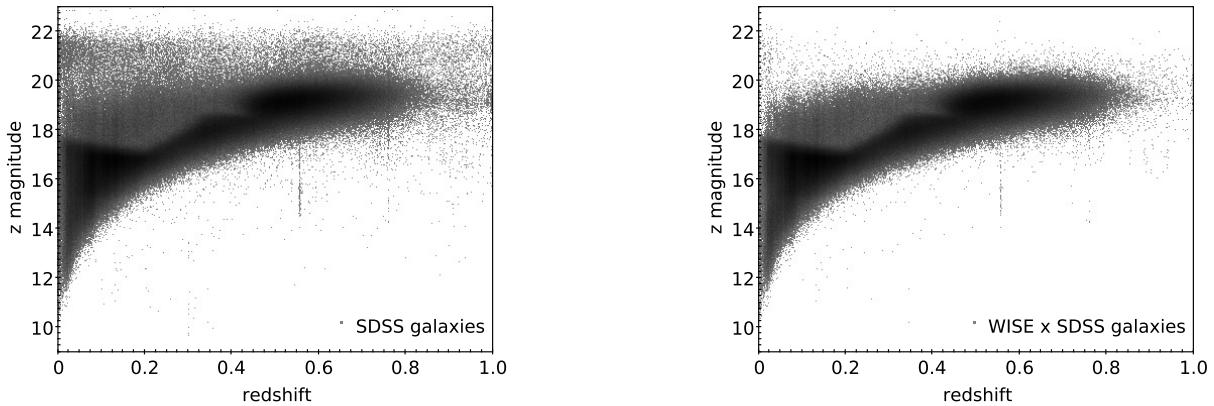


Fig. 4. Redshift – apparent magnitude diagrams for the SDSS z band: SDSS-only galaxies (left panel) and galaxies from the WISE \times SDSS cross-match (right panel).

we can classify WISE sources in the present application. As shown in Fig. 3, presenting normalised $W1$ counts for the three source types in the WISE \times SDSS cross-match, there are hardly any galaxies in there fainter than $W1 = 16$ mag, so we will not be able to create reliable training samples beyond this magnitude, although both stars and galaxies are present in the WISE \times SDSS sample at considerably fainter fluxes. The histogram for galaxies displays additionally two clear peaks. Such a shape of the galaxy counts can be attributed to the combination of two effects: 1) the first one results from the heterogeneity of the SDSS dataset itself (due to preselections of the Main, CMASS and BOSS samples) and the second one is related to the selection of the WISE \times SDSS catalogue, which is additionally based on the detections in the $W1$ filter. As demonstrated in Figs. 4 and 5, the SDSS-based galaxy selection does not sample all the regions of the redshift-magnitude space equally. In particular, as illustrated in the left panel of Fig. 4, showing the redshift–magnitude diagram for the z band (the longest wavelength measured by SDSS), in the pure SDSS data we observe a clear depletion of faint red galaxies at $z \leq 0.3$ in comparison to higher redshifts. This effect persists after adding the $W1$ -based selection of the WISE \times SDSS sample, as shown in the right panel of Fig. 4. The enhancement of this effect can be attributed to the properties of the $W1$ filter, which at low redshifts probes the part of the galaxy spectrum where strong PAH features are very prominent; 2) in Fig. 6 the convolution of the $W1$ filter with a spectrum of a typical spiral galaxy at various redshifts (taken from the Brown et al. 2014b library) demonstrates that observing in $W1$ we can expect a selection function which does not change monotonically with redshift. In particular, it has a minimum at $z \sim 0.15$, and then rises again until $z \sim 0.28$. The combination of these effects results in a relatively complex sampling of galaxies in the redshift – $W1$ magnitude space in the WISE \times SDSS data, as demonstrated in Fig. 5.

Finally, we cannot hope for reliable classification also at the very bright end. Not only the WISE $W1 < 8$ mag or $W2 < 7$ mag sources are saturated; in addition, the WISE \times SDSS sample does not include galaxies nor quasars brighter than $W1 \sim 9.5$ mag. These brightest objects will be thus removed from our final samples, which however has minor influence on the results, as the $W1 < 9.5$ mag WISE sources are concentrated mostly in the Galactic Bulge (i.e. they are stars and blends thereof).

Being aware of all these biases, we would like to reiterate that introducing them is a trade-off if we want to use the training

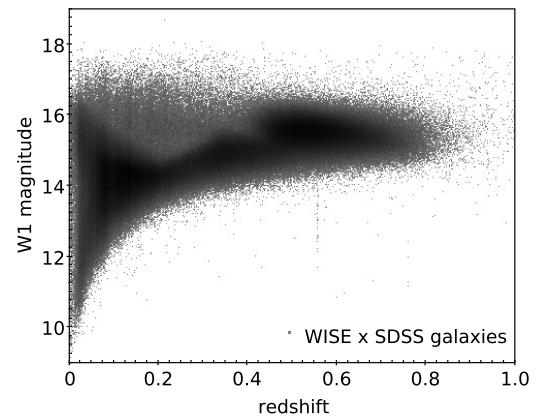


Fig. 5. Redshift – apparent magnitude diagram for the WISE $W1$ band, for the galaxies from the WISE \times SDSS cross-match.

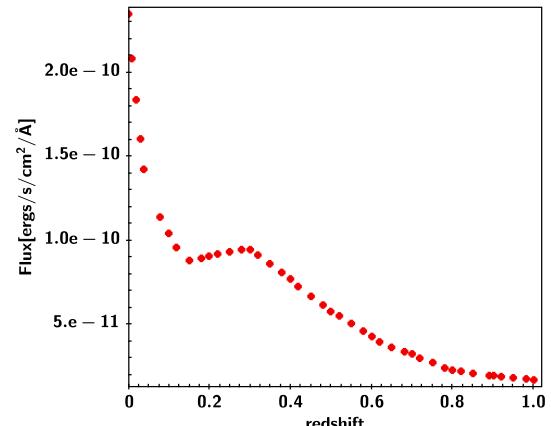


Fig. 6. Convolution of the $W1$ filter with a template spectrum of a typical spiral galaxy as a function of redshift.

sample providing star, galaxy and quasar spectral classifications, which at the depths of interest is at present available only from the SDSS. As a final caveat, the training sample applied in this study does not include Solar System bodies, which however are

present in the WISE database. Our final classification will thus ignore this fact, and as we will show, such a type of contamination affects mostly the sources flagged as quasars by our classifier.

3. Classification method: support vector machines

In general, *classification* is a process that uses pattern recognition. A *classifier* is a function which maps a feature vector of a given object characteristics into a discriminant vector containing likelihoods for the objects to belong to the different considered classes. Classification schemes rely on choosing a feature space, where different classes occupy different volumes with minimal overlapping. This approach has been utilised in developing machine learning algorithms – statistical methods which constitute a branch of artificial intelligence and are based on creating and exploiting systems which learn on data.

In this work for the task of identifying object types we adopt the support vector machines (SVM) algorithm. SVM, a supervised method based on kernel algorithms (Shawe-Taylor & Cristianini 2004), was designed to extract structures from data, and thanks to its excellent ability to deal with multi-dimensional samples, combined with providing high accuracy, has been extensively applied to many diverse astronomical problems. To name a few, SVMs have been used to solve problems like classifying different structures in the Interstellar Medium (Beaumont et al. 2011), pinpointing active galactic nucleus (AGN) candidates (Cavuoti et al. 2014), or distinguishing different subclasses of specific spectral type stars (Bu et al. 2014). Last but not least, and what is of relevance here, SVM has proven itself for selection of different classes of objects, like stars, quasars and galaxies (e.g. Saglia et al. 2012; Solarz et al. 2012; Malek et al. 2013; Kovács & Szapudi 2015).

In what follows, we draw the general outline of the nature of the algorithm; for an in-depth discussion we refer the reader to Vapnik (1999); Cristianini & Shawe-Taylor (2000); Hsu et al. (2003).

Each training object can be described by a number of quantities, N , which determine its discriminating properties. The SVM regards the values of the quantities as a position of a given object in an N -dimensional parameter space; in other words, the algorithm maps the feature vector from the input space X to a feature space H using a non-linear function $\phi : X \rightarrow H$. In the feature space H , the discriminant function, which will determine the boundary, takes a form of:

$$f(x) = \sum_{i=1}^n \alpha_i k(x, x') + b. \quad (1)$$

Here $k(x, x')$ represents the kernel function, which returns the inner product of the mapped vectors, α_i is a linear coefficient and b is a perpendicular distance called bias, which translates it into a given direction.

With a substantial amount of feature vectors representing different classes of objects, the algorithm searches for boundaries segregating those classes with the biggest possible distance from each data point (a margin). The objects laying closest to the boundary are called *support vectors*. In other words, the SVM algorithm searches for a decision boundary B that will maximise a fitness function F :

$$F = M - C \sum_i \xi_i(B, M); \quad (2)$$

where M denotes the margin of the boundary. The number of training examples violating this criterion is given by $\xi_i(B, M)$. If a position of a point i is found within a distance higher than M from B , then $\xi_i = 0$. In an opposite case, ξ_i will be equal to the distance that point i should be shifted by so that the condition be satisfied. To set a trade-off between the large margins M and misclassifications ξ_i , an adjustable cost parameter is used (more details in Beaumont et al. 2011).

The usage of kernel functions allows for shifting into a higher-dimensional parameter space, where usually the data turn out to be much more simply separated than in lower-D ones. There are many possible kernel functions to be used (such as polynomials, exponential radial basis functions, or multilayer perceptrons; Cristianini & Shawe-Taylor 2000). The choice of a proper kernel, which will be suitable for a given problem, is crucial; the usual procedure is to try several ones, beginning from the simplest cases (to avoid over-fitting and to save on parameter tuning time) and then to move towards more complex ones in order to gain accuracy. For this particular dataset we tested two kernel functions to obtain the most reliable classification outcome: the Gaussian radial basis function (GRB) and the polynomial function. The GRB is given by:

$$k(\mathbf{x}, \mathbf{x}') = \exp(-\gamma \|\mathbf{x} - \mathbf{x}'\|^2), \quad (3)$$

where \mathbf{x} and \mathbf{x}' represent feature vectors in the input space, $\|\cdot\|$ denotes the Euclidean distance and γ is the adjustable kernel width parameter, which is responsible for the curvature of the decision surface. The polynomial kernel is defined as:

$$k(\mathbf{x}, \mathbf{x}') = (\gamma(\mathbf{x} \cdot \mathbf{x}') + c_0)^d, \quad (4)$$

where \mathbf{x} and \mathbf{x}' represent feature vectors in the input space, $\mathbf{x} \cdot \mathbf{x}'$ is their inner product, d stands for the degree of the polynomial function, and c_0 is a constant coefficient.

Therefore, for the GRB kernel there are two adjustable parameters, γ and C , which will determine the separation boundary and complete the training of the SVM classifier. In case of the polynomial kernel, the amount of adjustable parameters increases to four: in addition to γ and C , the degree d and the coefficient c_0 have to be known. Then, after the most efficient kernel function is chosen, a classification of the new data points depends on their position relative to the boundary: SVM will assign a type to unknown objects based on which side of the separation hyperplane they fall.

What is more, instead of assigning discrete class labels, it is possible to find out what is the class probability for a given object. In case of binary SVM this can be done by implementing the Platt's a posteriori probabilities (Platt 1999, Lin et al. 2007): once the decision values f of the SVM classifiers are computed, a sigmoid function

$$P(i|i \text{ or } j, \mathbf{x}) = \left(1 + e^{Af+B}\right)^{-1} \quad (5)$$

is fitted (where i and j represent two classes). Then, A and B are estimated by minimizing the negative log-likelihood function. In order to extend the probabilities of classes to a three-class problem, all class probabilities from output of binary classifiers are combined (fan Wu et al. 2003). Thus calculated probabilities will be used in the final classification to eliminate sources that have low probabilities of belonging to any class (meaning that each class has $p < 0.5$, and in some cases the three probabilities are $p \sim 0.33$).

SVMs are currently available in a variety of software packages; most widely used among them is libsvm (Chang & Lin 2011)⁸, which provides robust implementation of SVM for both

⁸ <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

classification and regression. In this work we use the R ([R Development Core Team 2005](#))⁹ implementation of SVM included in the e1071 package ([Dimitriadou et al. 2005](#)), which provides an interface to `libsvm`.

3.1. SVM on imbalanced data sets: oversampling

Our training dataset is characterised by low numbers of bright objects in the QSO sample. It was then necessary to address the problem of the accuracy decreasing due to this imbalance. This feature is common in many classification schemes, especially in those which aim for the maximisation of the accuracy, like SVMs ([Akbani et al. 2004](#)). If not accounted for properly, this would result in making a simple decision, at basis of which lies the maintenance of the highest success rate: assign the most common class to the test objects. There are two ways of addressing this problem: it either can be solved through rebalancing the data set or by altering the algorithm itself. The first solution works at the level of manipulating the data, where the underrepresented population(s) can be oversampled, or the dominant class can be undersampled. In the latter case, when reducing the amount of the objects contained within the majority class, the distributional assumptions on the data must be made: some crucial information may be lost or additional noise may be introduced. On the other hand, changing the algorithm mainly relies on the cost-sensitive learning, where a higher penalty is assigned to the misclassifications, resulting in a shift of the classifiers decisions towards the minority class, improving the detection accuracy.

Since SVM decision making relies solely on the support vectors, it is working well against any noise in the data and light imbalance. Therefore, if the distribution of the training sets is very skewed, the amount of the support vectors in the majority class outweighs the ones from the minority class. In the case of WISE data we have decided to perform oversampling of the underrepresented class of QSO, i.e. additional artificial objects were created. The number of missing objects (X_{missing}) needed to be added to QSO training samples, was calculated from the following equation ([Małek et al. 2013](#)):

$$\lceil X_{\text{missing}} \rceil_{10} = NG \times 0.8 - X, \quad (6)$$

where $\lceil \cdot \rceil_{10}$ stands for rounding the value up to the nearest ten, X corresponds to the number of original QSOs in the sample, and NG is the number of galaxies. This strategy allows to obtain fully balanced training samples, which are essential to build an effective classifier.

We created mock samples of missing objects by slight changes of the real parameters. In the first step a real QSO was randomly chosen, and then its parameters were re-assigned by shifting the real ones by an amount drawn from a Gaussian distribution with specific standard deviations σ . Different values of σ were used, depending on the type of parameter: σ_{pm} for proper motions and σ_W for magnitudes. They were calculated as the median values of the proper motion and magnitude uncertainties, respectively. These values are given in Table 1, which lists the cases in which the oversampling was applied, providing numbers of objects before and after the oversampling. See the following Sec. 4 for the details on the magnitude and extinction divisions. Fig. 7 presents an example of a colour-magnitude diagram for quasars before and after oversampling. The distribution of objects after the oversampling closely mimics the real one, as designed.

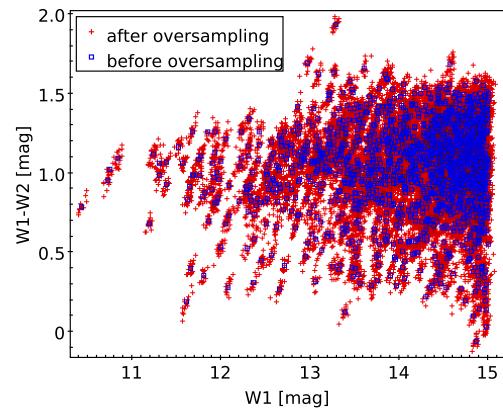


Fig. 7. Representative colour-magnitude diagram before and after oversampling for quasars.

4. Calibrating the SVM classifier for WISE data

In this section we present various tests of the SVM algorithm performed on the WISE×SDSS training data to verify and optimise the performance of the classifier. In particular, we tested algorithm's efficiency as a function of the following: i) choice of the kernel; ii) number of sources in the training samples; iii) number of parameters used for the classification; iv) Galactic extinction; v) limiting magnitude of the sample; vi) usage of source apparent motions. All this will allow us to prepare the SVM for the eventual application to the all-sky WISE dataset, presented in Section 5.

In each of the tests, we used a ten-fold cross-validation technique: we divided the training set into 10 subsets of equal size, and selected 9 subsets to train the classification model and tested it on the remaining 10th subset; this was repeated 10 times, leaving out a different subset each time. We then counted the training objects whose nature was correctly identified by SVM: TS (true star), TG (true galaxy), TQ (true QSO), as well as those which were misclassified by the algorithm: FG (false galaxy), FS (false star) and FQ (false QSO).

Then we define *completeness* c , *contamination* f and *purity* p for the three classes of objects (e.g. [Soumagnac et al. 2015](#)). For galaxies we have:

$$c_G = \frac{\text{TG}}{\text{TG} + \text{FGS} + \text{FGQ}}, \quad (7)$$

$$f_G = \frac{\text{FSG} + \text{FQG}}{\text{TG} + \text{FSG} + \text{FQG}}, \quad (8)$$

$$p_G = 1 - f_G = \frac{\text{TG}}{\text{TG} + \text{FSG} + \text{FQG}}, \quad (9)$$

where TG, FGS and FGQ stand for galaxies classified respectively as galaxies, stars and quasars, and FSG (FQG) define stars (quasars) misclassified as galaxies. Analogous definitions are used for *completeness*, *contamination* and *purity* of stellar (c_S , f_S , p_S , respectively) and quasar (c_Q , f_Q , p_Q , respectively) samples.

In each case we measure completeness, purity and contamination for two variants: a self-check and a cross-test. For the self-check we classified the same objects as were used in the

⁹ <http://www.R-project.org>

Magnitude limit	W1<14								W1<15	
Extinction I_{100}	<0;1)		<1;2)		<2;3)		<3;10)		<3;10)	
Oversampling	before	after								
Number of galaxies	36801	36801	54417	54417	22916	22916	10732	10732	33720	33720
Number of stars	6113	6113	10612	10612	4757	4757	6409	6409	13822	13822
Number of QSOs	2161	29598	2556	43998	1141	18398	525	8798	1498	27198
σ_w [mag]	0.025		0.025		0.026		0.026		0.028	
σ_{pm} [mas/yr]	84		98		99		113		165	

Table 1. Numbers of objects before and after oversampling in the bins for which oversampling was applied: for a sample with $W1 < 14$ we applied oversampling in all the extinction bins, while for $W1 < 15$ only for extinction in the range $I_{100} \in \langle 3; 10 \rangle$. Oversampling was implemented for quasars only. Parameters σ_w and σ_{pm} denote specific σ values of Gaussian distributions, respectively for magnitudes and proper motions.

SELF-CHECK						
kernel	polynomial			radial		
	Completeness	Purity	Contamination	Completeness	Purity	Contamination
Galaxy	82.9	79.4	20.6	86.3	79.9	20.1
Stars	81.0	84.1	15.9	81.0	87.1	12.9
QSO	96.9	97.8	2.2	96.9	97.9	2.1
CROSS-TEST						
kernel	polynomial			radial		
	Completeness	Purity	Contamination	Completeness	Purity	Contamination
Galaxy	81.0	77.9	22.1	84.0	75.7	24.3
Stars	78.0	83.9	16.1	77.0	86.5	13.5
QSO	97.0	94.2	5.8	96.0	96.0	4.0

Table 2. Comparison of SVM performances for two kernels, polynomial and radial, for the self-check and cross-test.

given training sample. In the cross-test we classify the objects from the subsamples different than in the current training set.

The tests described below were performed for different combinations of magnitude limits and Galactic extinction levels. Three flux limits were adopted: $W1 < 14$ mag, $W1 < 15$ mag, and $W1 < 16$ mag. In each of them data were additionally binned according to the measured Galactic dust emission. Here we chose to use the 100-micron intensity (I_{100}) sky map made from a combination of COBE/DIRBE and IRAS 100 μm measurements (Schlegel et al. 1998). We preferred this parameter over the commonly applied $E(B - V)$ as the former was directly measured from data, while the latter was derived therefrom. Our extinction bins adopted hereunder were $I_{100} < 1$, $1 \leq I_{100} < 2$, $2 \leq I_{100} < 3$ and $3 \leq I_{100} < 10$ [MJy/sr]. Above $I_{100} = 10$ MJy/sr, which constitutes about 1% of the WISE×SDSS catalogue, there are practically no galaxies nor quasars in the training set. In general, the $I_{100} \geq 10$ MJy/sr areas cover about 17% of the full sky, practically only in the Galactic Plane and regions of high dust obscuration, where our classification is not expected to be reliable.

In some cases, the above splitting of the full sample has left us with very small numbers of quasars in the relevant training sets, and the oversampling methodology had to be applied, as described in Sec. 3.1.

4.1. Kernel performance comparison

The first test that we made served to determine the optimal kernel for our application. We compared the performance of the two kernel functions described in Sec. 3, polynomial and radial (see Table 2) and analysed the so-called *univariate histograms* of projections from both self-check and cross-test of known data. A univariate histogram of projections is a graphical representation of the training data for a given binary classification (in case of a three-class classifier we have three two-class classifiers) and the decision boundary SVM provides given the data. To obtain

the best efficiency of the classification, it is a standard practice to divide the training set into two subsets: a subset which will be used for actual training and a validation subset, which will serve as a verification of the accuracy of the created hyperplane against other known, albeit not used for training, objects. In this test, the training set contains 99% of the total amount of sources with known classification, while the validation test is composed of the remaining 1% of known objects.

For non-linear SVM kernels, projected values $f(x)$ (Eq. 1) are being obtained through usage of the kernel representation in a dual space. This means that there are (in case of WISE×SDSS data) three support vector machines, each of which has its own decision function. Projection of an object x_k from a training set onto the normal direction of a nonlinear SVM boundary can be written as follows:

$$f(x_k) = \sum_{i \in s, v} \alpha_i y_i K(x_i, x_k) + b, \quad (10)$$

where x_i denotes a support vector; and classification of each example is determined by the sign of this function. The soft margin of a classifier can be written as

$$y_{f(x_k)} = \text{sign}(f(x_k))(1 - \exp^{|f(x_k)|}), \quad (11)$$

and the boundaries of the soft margin are then $y_{f(x_k)} \in [-1; 1]$. Then, $y_{f(x_k)}$ describes two aspects of each example. The first one comes from the sign: it encodes a 'hard' decision whether the example x_k belongs to a given class or not. The second one comes from its absolute value: it represents how strong the decision is. This means that the farther a given example x_k falls from the decision boundary, the more certain the decision is.

As seen from Table 2 the differences between completeness and purity of training samples and validation data sets are small for the two kernels considered. However, when we compare the histograms of projections of each point with respect to the boundary we see clear differences (Figs. 8 & 9). For the radial kernel we observe an effect of data piling on the margins, which is typical for high-dimensional data (Cherkassky & Mulier

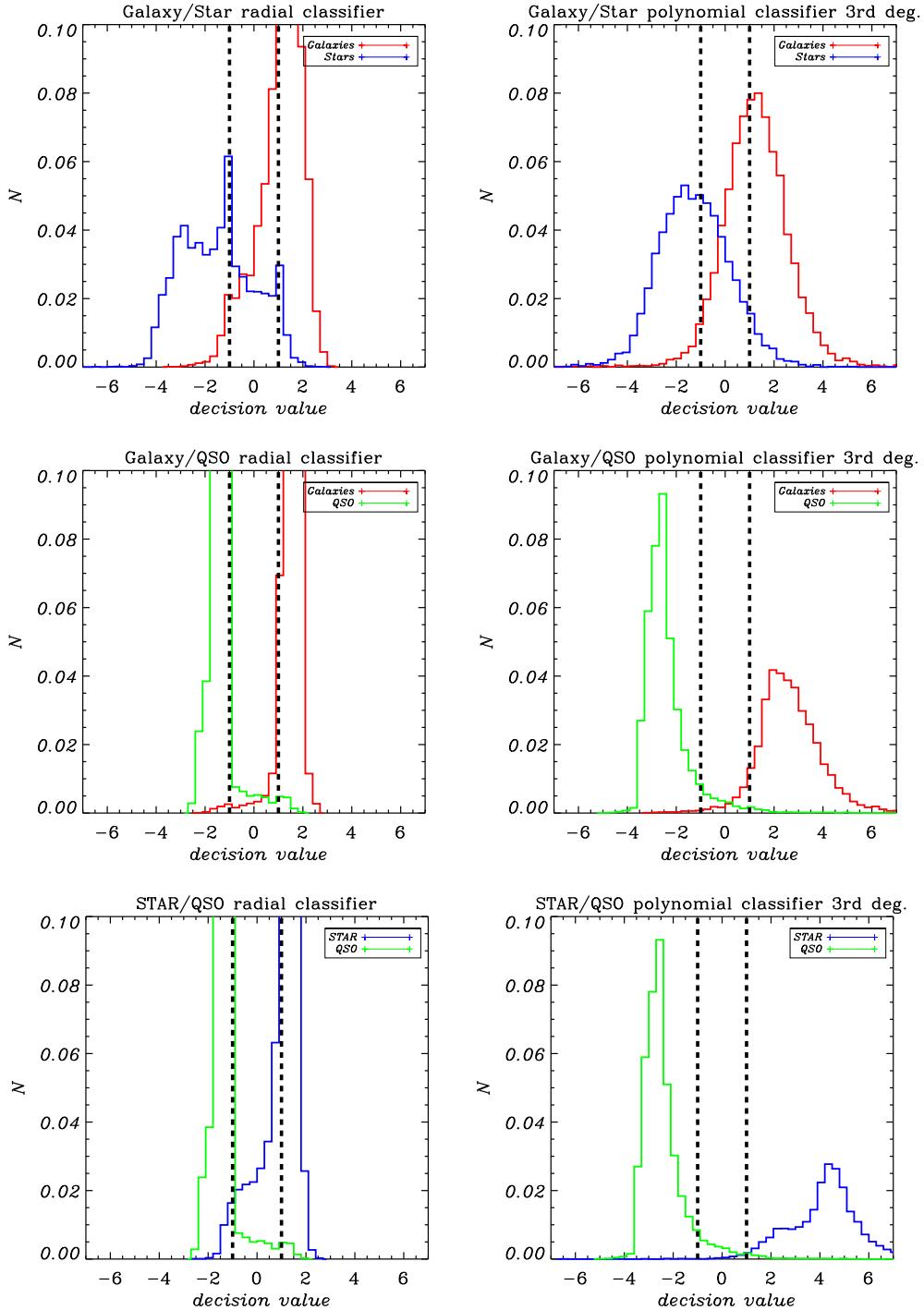


Fig. 8. Comparison of the histograms of projection values of the training data samples onto the normal direction of the SVM decision boundary for a radial (left column) and polynomial kernel (right column). Top row represents division between galaxies and stars (red and blue, respectively); middle row between galaxies and quasars (red and green); bottom row between stars and quasars. Vertical lines represent the boundaries of the decision hyperplane margins for the two classes (+1 and -1), and 0 value marks the position of the hyperplane itself.

2006). Separability of the set on which the classifier was trained does not imply that the validation or test sets will be equally well separated (as seen in Figs. 8 & 9, left columns). As the SVM optimization aims at high separability of the training data, it penalizes the data which fall into the soft margin of the separation boundary. However, the aim is to also have a good separation of the validation set (which in turn should improve the separability of the test sample). That is why a preferred model allows for data points to fall into the soft margin. Moreover, it is desirable

to have as strong decision values as possible, therefore the data piling effect should be avoided. That is why the kernel which displays a clearer division of validation – the polynomial one in this case – is more preferable, as shown in Fig. 9.

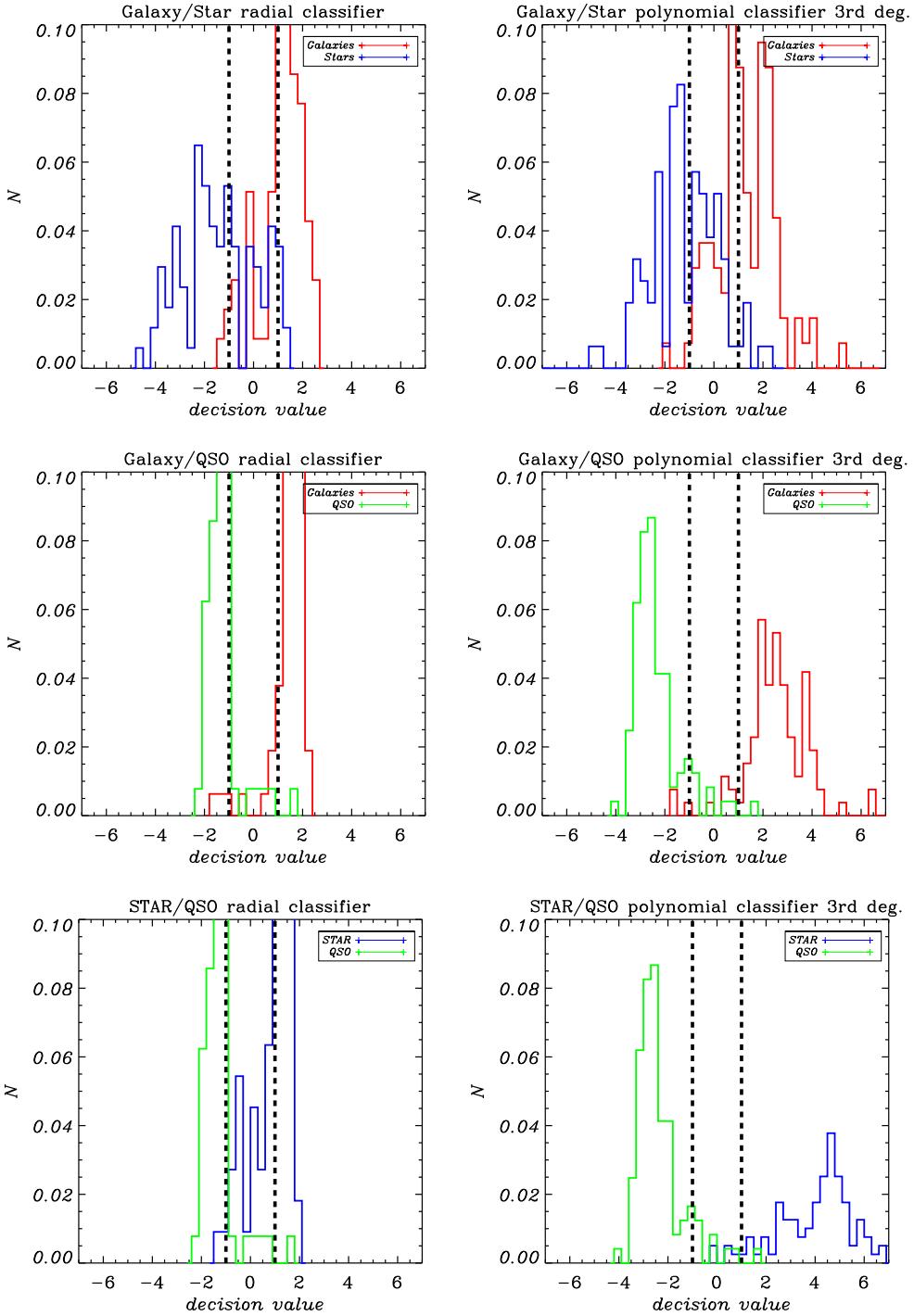


Fig. 9. Same as Fig. 8, but for the validation data-set.

4.2. Optimal number of objects in the training samples

As our WISE×SDSS training set is very large as compared to earlier SVM classification applications (e.g. in AKARI by [Solarz et al. 2012](#) or in VIPERS by [Malek et al. 2013](#)), in the first step, having decided which kernel function to use, we performed a series of tests to check whether we can calibrate our SVM method on smaller subsamples without deteriorating the results. We conducted four tests for each of the three flux-limited samples (in this case there was no extra division according to the extinction), in which we randomly chose 100, 1000, 3000 and 5000 objects for each class (i.e. 100 galaxies, 100 stars, 100 quasars, etc), and

we used such training sets to compute relevant statistics as defined above. Each test was repeated 10 times, and in all the cases the error bars provided represent the standard deviation from the mean of the 10 tests.

Fig. 10 shows, as an example, the dependence of the completeness and purity on the number of training objects for a bin $W1 < 14$ mag for the self-check and cross-test, respectively. The results for other flux limits are similar. As seen in this figure, our results stabilise for subsamples with randomly chosen 3000 objects of each class. Based on these results, the following tests were applied for these numbers of objects. This allowed us to

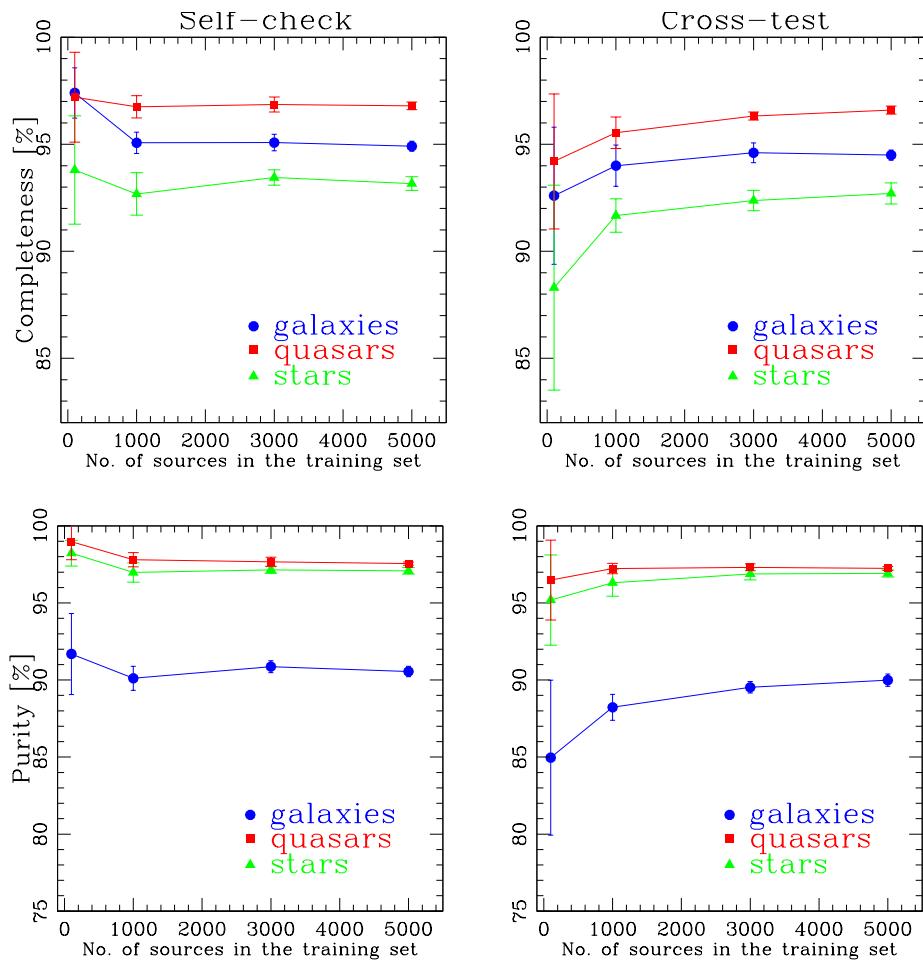


Fig. 10. Dependence of the completeness (upper panels) and purity (lower panels) on the number of objects in the training set, for a $W1 < 14$ mag flux-limited sample, for the self-check (left panels) and the cross-test (right panels) cases. Relevant contamination levels are 100%–Purity.

significantly save on computation time in the tests, as it scales highly non-linearly with the size of the training set.

4.3. Parameter space for classification

Having established the optimal size of the training samples, we performed a series of tests to determine the minimum number of parameters sufficient for optimal classification. Each additional parameter significantly extend computation time, while it does not necessarily have to improve overall accuracy, for instance in cases when it is noise-dominated and/or not related to the source type. We thus started by using just two of them, the $W1$ magnitude and the $W1 - W2$ colour, and then extended the parameter space by first adding the differential aperture magnitude $w1mag_1 - w1mag_3$, and next also the apparent motions pm (cf. Sec. 2.1 for parameter descriptions). For a proper comparison, these tests were applied on sources with ‘detected’ proper motions, i.e. having $pm > sigpm$. Additionally, they were employed in various combinations of magnitude cuts and extinction bins, as described earlier.

Fig. 11 shows how completeness and purity change when the number of implemented parameters increases. This particular example is for $W1 < 14$ mag and $I_{100} < 1$ [MJy/sr], but the results were qualitatively the same for each of the magnitude cut – extinction combinations. While quasars were very accurately classified already for only two parameters used ($W1$ and

$W1 - W2$), for stars and galaxies both completeness and purity significantly increased after the differential aperture magnitude was used as the third parameter. On the other hand, the proper motions did not bring any improvement, and sometimes even slight deterioration in accuracy was observed once they were applied. This is hardly surprising taking into account all the caveats associated with WISE apparent motion measurements (Kirkpatrick et al. 2014): they are not accurate enough for our type of analysis, and from now on we will thus focus on tests not using proper motions. We can expect, however, that longer time baselines and/or better photometric accuracy possible with the NEOWISE data (Mainzer et al. 2014), once combined into the ‘MaxWISE’ data product (Faherty et al. 2015)¹⁰, as well as with future surveys (such as the LSST) should allow for proper motions to become a useful parameter for the classification of sources, including extragalactic ones. This would also allow to identify a fourth type of source, which we have not considered here as we do not have them in the training samples, although they certainly are present in the WISE database, namely minor bodies of the Solar System. As will be shown in Sec. 5, they most likely contaminate the quasar candidate sample in the final all-sky classification.

¹⁰ See also <http://wise5.ipac.caltech.edu/posters/Eisenhardt.pdf>.

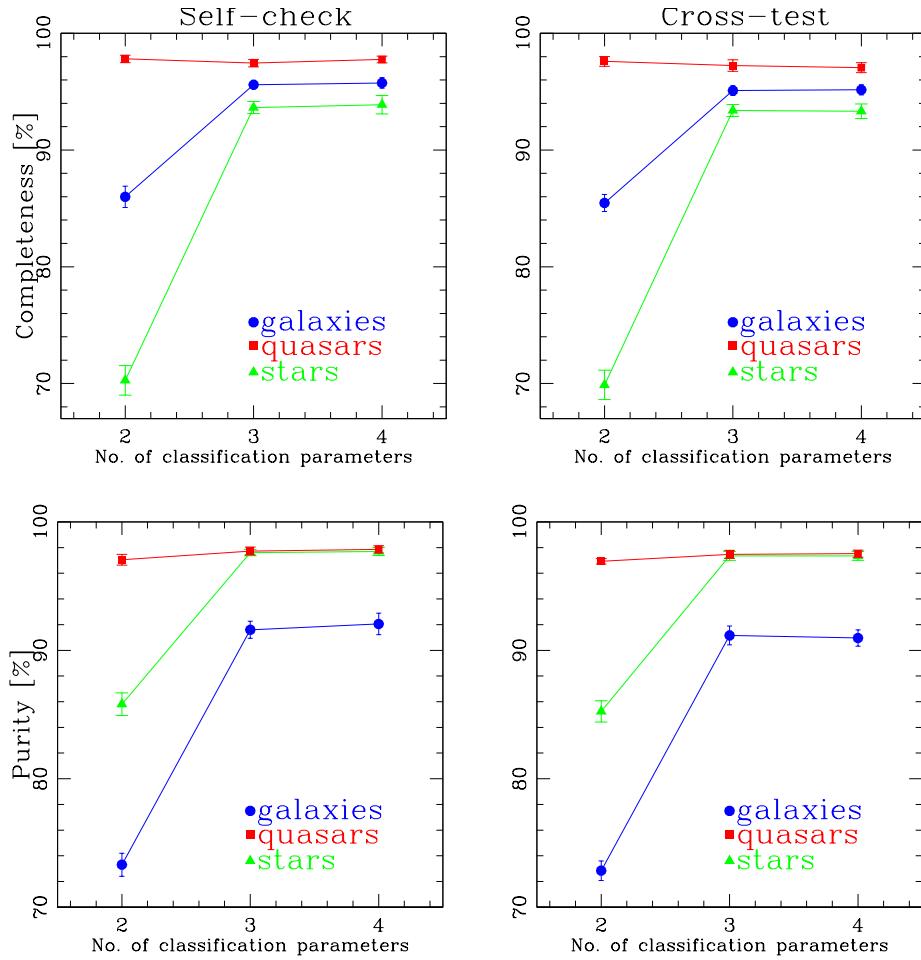


Fig. 11. Dependence of the completeness (upper panels) and purity (lower panels) on the number of parameters used for SVM training, for a sample with $W1 < 14$ mag and $I_{100} < 1$ [MJy/sr], for the self-check (left panels) and the cross-test (right panels) cases. The specific classification parameters for the training were: $W1$ (1st), $W1 - W2$ (2nd), $w1mag_1 - w1mag_3$ (3rd), and apparent motions (4th). See text for details. Relevant contamination levels are 100% – Purity.

4.4. Dependence of classification accuracy on extinction and limiting magnitude

Two final tests of the SVM algorithm applied to WISE×SDSS data were to check its performance against varying extinction levels and increased magnitude cut. Figs. 12 & 13 synthesise the results, showing how completeness and purity change with varying magnitude for four I_{100} bins, for the three classes of sources. Here we used smaller increments (by 0.5 mag) than in the other tests where it was 1 mag.

At the bright end, both completeness and purity retain very high levels of $> 90\%$ irrespectively of extinction. These numbers for stars and galaxies gradually deteriorate for fainter sources, and some dependence on the extinction starts to appear as larger magnitudes are reached. The statistics are however relatively stable and at very good levels for quasars (where a slight *increase* in purity is actually observed at the faint end). We note however that even for fainter sources, star and galaxy samples exhibit completeness of over $\sim 80\%$ and purity of over 77%. Detailed results regarding completeness, purity and contamination for all magnitude-extinction bins for the self-check and cross-test in the three dimensional parameter-space are presented in Tables A.1 and A.2, in the Appendix.

We note however that such statistics may be partly misleading, as they refer to the test sample which is statistically consis-

tent with the training set. The full-sky catalogue will however differ, and in particular may (and does) contain sources not represented in the training, such as for instance asteroids. The classifier tuned to the training set will underperform in such cases, which in particular will be reflected in low probabilities of such objects to belong to any of the three classes used in the analysis. We discuss it also in the following section.

5. Application of the SVM classifier to all-sky WISE

Having verified in various ways the performance of our classifier, we finally applied it to the all-sky WISE data limited to $W1 < 16$ mag. In this case, to tune the classifier we used the most comprehensive and general training data, by randomly selecting 10^4 galaxies, 10^4 stars and 10^4 quasars from the cross-matched WISE×SDSS dataset with $W1 < 16$ mag. Thus trained classifier flagged 70% / 27% / 3% of our WISE sources respectively as stars / galaxies / QSOs, on the full sky. These numbers are consistent with the fact that stars dominate the source counts at the bright end of WISE (Jarrett et al. 2011, Jarrett et al., in prep.), however they should not be taken at face value. At low Galactic latitudes and in other highly crowded areas (Magellanic Clouds, Galactic extended sources such as dust clouds) the classification is highly unreliable. In addition, these numbers refer to the sources for which the probability to be of a given type was

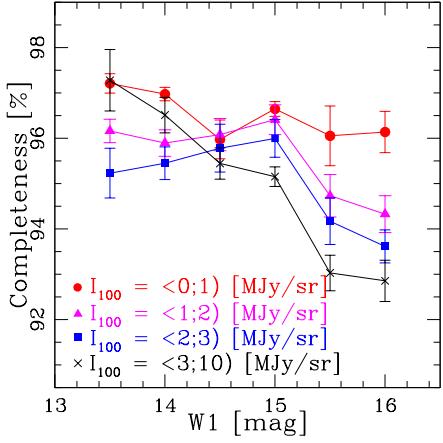
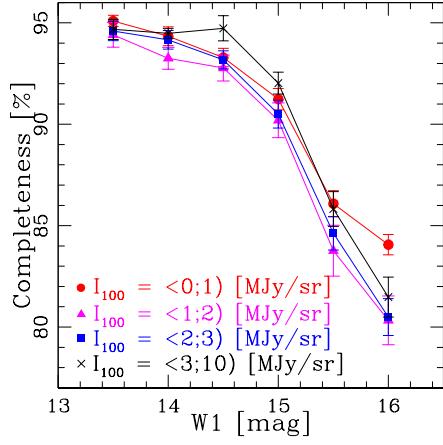
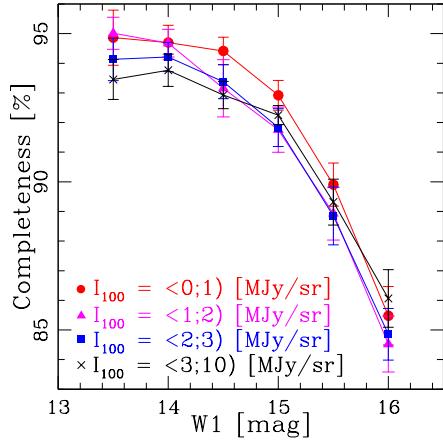


Fig. 12. Dependence of the completeness on the magnitude for four I_{100} bins when using three classification parameters (W_1 , $W_1 - W_2$ and $w1mag_1 - w1mag_3$) for galaxies (upper panel), stars (middle panel) and quasars (lower panel) from the WISE \times SDSS sample. These results are for the cross-test.

higher than of the two other ones (e.g. $p(\text{star}) > p(\text{galaxy})$ & $p(\text{star}) > p(\text{QSO})$ for stars). However, a considerable number of objects, especially in the galaxy and QSO classes, had the three probabilities comparable, which means a very low level of confidence for a class assignment. Thus, to obtain galaxy and quasar candidate catalogues based on our data, masking of problematic areas was necessary, as well as additional cleanup of low-probability sources. In case of stars this was not needed as firstly,

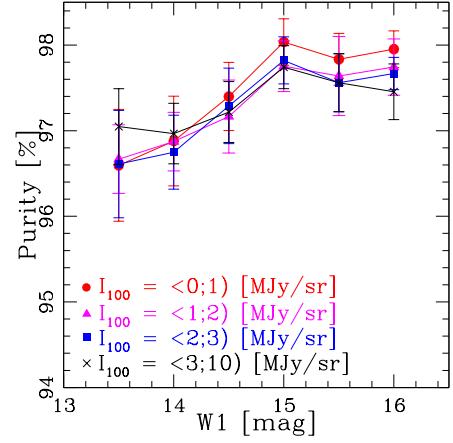
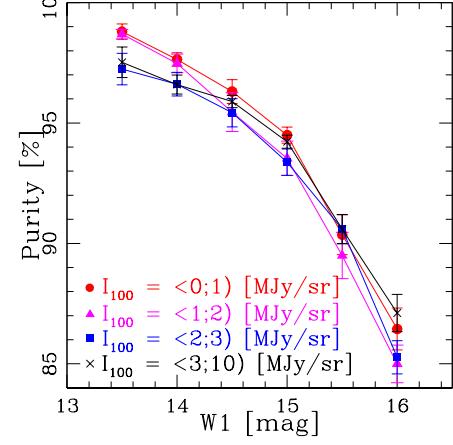
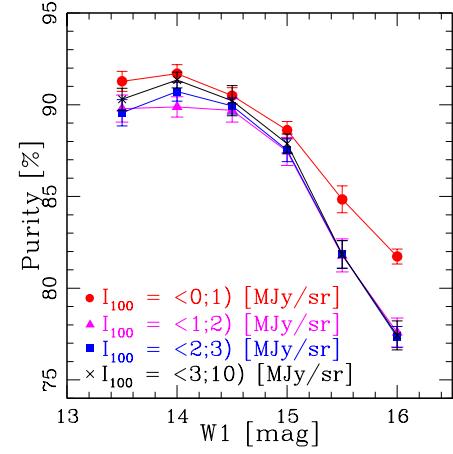


Fig. 13. Dependence of the purity on the magnitude for four I_{100} bins when using three classification parameters (W_1 , $W_1 - W_2$ and $w1mag_1 - w1mag_3$) for galaxies (upper panel), stars (middle panel) and quasars (lower panel) from the WISE \times SDSS sample. These results are for the cross-test.

we do expect them to be present in the highly crowded areas and secondly, their class assignment was the most robust: 99.4% of the sources classified as stars had $p(\text{star}) > 0.5$. A map of the 220 million star candidates is shown in Fig. 14. A decrease in counts at $b \sim 0^\circ$ is caused by saturation and blending.

The catalogue of candidate galaxies, unlike stars, needed considerable purification. First of all, we had to cut out the most confused areas of the Galactic Plane and Bulge, using

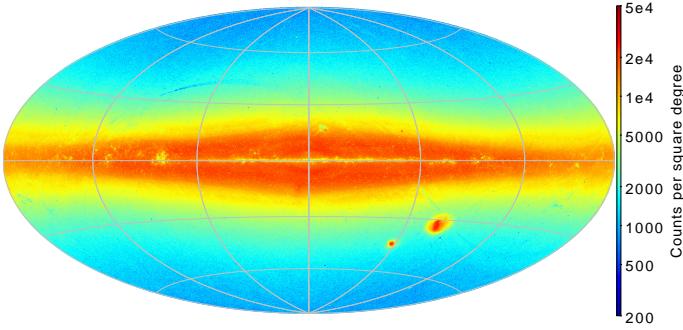


Fig. 14. All-sky map of 220 million star candidates identified by our classifier in AllWISE $W1 < 16$ mag data, in Aitoff projection in Galactic coordinates. Note the logarithmic scaling of the count colour bar.

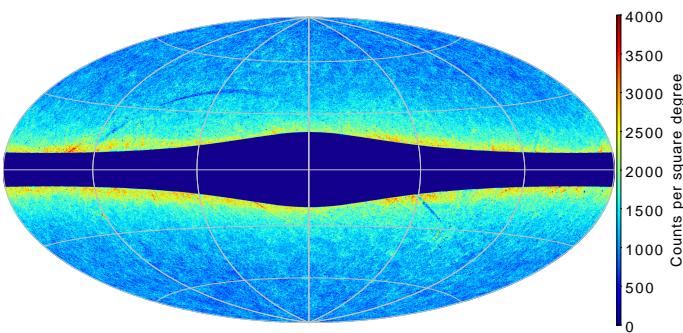


Fig. 15. All-sky map of 45 million galaxy candidates identified by our classifier in AllWISE $W1 < 16$ mag data, in Aitoff projection in Galactic coordinates. This plot shows sources over the probability threshold of $p(\text{gal}) > 0.6$, after appropriate cleanup and masking (see text for details).

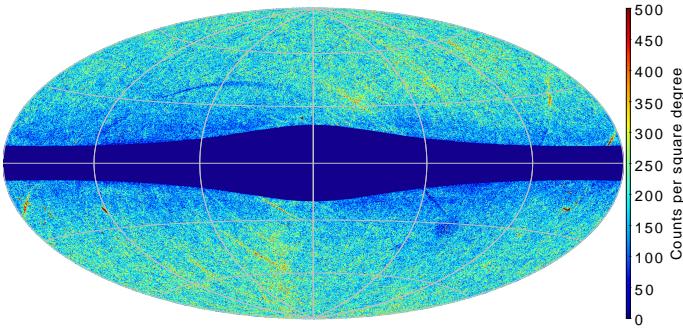


Fig. 16. Sources flagged by our classifier as quasar candidates in the WISE $W1 < 16$ mag catalogue. This sample shows 6 million objects with an SVM probability $p(\text{QSO}) > 0.5$, after appropriate cleanup and masking (see text for details).

a longitude-dependent masking of the $|b| < 6.5^\circ$ sources at $\ell = 180^\circ$ up to $|b| < 20^\circ$ by the Galactic Centre. This removed almost 30 million objects out of the 84.5×10^6 pre-assigned to the galaxy class all-sky. As seen in Fig. 15, this mask could have been wider, but we leave it in such a form to emphasise classification issues at low Galactic latitudes where blends become a significant problem for star/galaxy separation in WISE. Secondly, in addition to the bright-end cut already mentioned in Sec. 2.2 (due to saturation and lack of bright sources in the training set), which affected a very small number of the objects, we also eliminated over 400,000 outliers at the faint end in the $W2$ band, $W2 > 16.1$ mag. These were located mostly by the Ecliptic Poles where WISE coverage was the highest due to the scanning strat-

egy. This cutout also automatically removed the sources with $W1 - W2 < -0.1$ mag which are most certainly stellar (Wright et al. 2010). In the final stage of the galaxy catalogue cleanup, we used the probabilities assigned by SVM as described in Sec. 3, and examined the sky and colour distribution of the galaxy candidates after applying different thresholds in $p(\text{gal})$. More aggressive cuts in this probability lead to a more uniform distribution of the sources as a function of latitude. However, even for $p(\text{gal}) > 0.7$ or more, differences of over 50% in the source density remain between the Galactic Caps and $|b| \sim 20^\circ$. As the $W1 - W2$ colours of the objects with high galaxy probability were consistent with those of genuine galaxies, the effect of gradually increasing number counts from high to low Galactic latitudes must be related to both the stellar contamination (blending) and galaxy incompleteness going up with decreasing $|b|$.

Fig. 15 shows an example of all-sky source distribution of galaxy candidates. Included are 45 million sources obtained after the masking, bright- and faint-magnitude cutouts, and putting a threshold of $p(\text{gal}) > 0.6$. There is clear contamination at low Galactic latitudes, but the classifier seems to be working well at least for half of the sky down to $|b| = 30^\circ$. The missing data in stripes on the left above the Plane and just below it on the right are due to AllWISE instrumental artefacts (saturation at the beginning of the post-cryogenic phase), as already discussed in Sec. 2.1.

The quasar candidates underwent similar purification as the galaxy ones. The same cutout of the Galactic plane was first applied, which removed almost 30% of the 9.4 million all-sky sources flagged by SVM as QSO. In addition, more aggressive bright-end cuts than in the galaxy case were necessary to avoid dangerous extrapolation from the training sample. We removed quasar candidates with $W1 < 10.4$ mag or $W2 < 10.1$ mag, as such bright QSOs are practically non-existent in WISE×SDSS. In the SVM output, they were mostly misclassified stars or blends thereof, localised chiefly at low Galactic latitudes and in the Magellanic Clouds. A further cleanup to keep only the $p(\text{QSO}) > 0.5$ sources resulted in 6 million objects as pictured in Fig. 16. This number is most certainly an overestimate for the true WISE quasar population at $W1 < 16$ mag. Indeed, in addition to the saturation-related artefacts, we note some interesting features in the map which are not seen for the galaxy candidates. First, there is some lack of sources at low Galactic latitudes, qualitatively similar to the WISE AGN distribution presented in Ferraro et al. (2015), where sources were classified based on colour cuts. Second, various WISE scanning issues are imprinted here, the most important being overdensity stripes perpendicular to the ecliptic, resulting from Moon avoidance manoeuvres (cf. the mask applied in Ferraro et al. 2015). There is however additional spurious overdensity which seems to roughly follow the ecliptic, visible in the top-right of the map and below the Bulge, to the left. This suggests some very local contamination, such as from asteroids or maybe zodiacal light, and most likely reflects the presence of a fourth type of sources in addition to the three ones in the training set from SDSS.

Unlike in the test phase described in Sec. 4, here we do not know the ‘truth’ to which we could compare classifier’s performance. Some indirect *a posteriori* tests are however possible. The first one is the all-sky distribution, which for stars and galaxy candidates (at high latitudes) is consistent with expectations, but much less for the ‘quasars’. The second test is to verify source properties such as colours. For identified galaxies, the $W1 - W2$ colour is very much consistent with the one of the WISE×SDSS training set (cf. Fig. 2), the more the higher $p(\text{gal})$ cut is applied. In case of quasar candidates, the situation

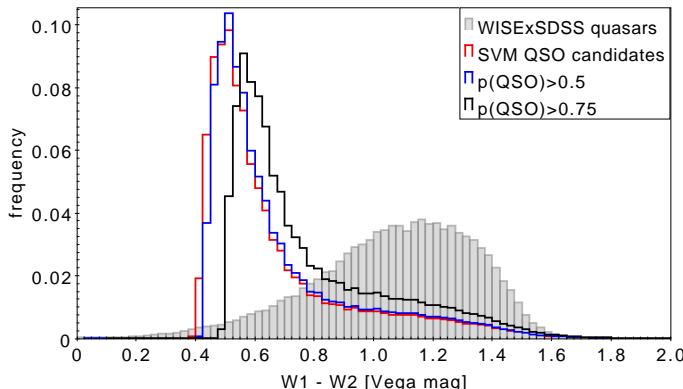


Fig. 17. Distribution of the $W_1 - W_2$ colour for WISE \times SDSS quasars (grey bars) and for SVM quasar candidates identified in all-sky WISE data, with no threshold on the SVM QSO probability (red), as well as for $p(\text{QSO}) > 0.5$ and > 0.75 (respectively blue and black).

is different. Even for very high thresholds of $p(\text{QSO})$, the peak in this colour is at $W_1 - W_2 \sim 0.6$ mag rather than ~ 1 mag as in the training set (Fig. 17). In fact, the sources of expected quasar nature with $W_1 - W_2 > 0.8$ mag (Stern et al. 2012) are only 1/5th of our QSO candidate sample. As already seen from their all-sky map, some of the contamination may come from Solar System objects. As an additional verification, we checked the location of the SVM ‘QSO’ in the $W_2 - W_3$ vs. $W_1 - W_2$ diagram, for those sources that had a W_3 detection. The bulk of the QSO candidates are located at $3 < W_2 - W_3 < 4$ [mag] and $0.4 < W_1 - W_2 < 0.7$ [mag], which does not give a clear characterisation of their nature. Indeed, following Wright et al. (2010), this is where various galaxy types overlap in this parameter space (‘normal’ spirals, Seyferts, starbursts, LIRGs...). This also indicates that even if we had used the W_3 parameter (which, as we emphasise, is robustly measured only for a small subset of our sources), this degeneracy between quasars and non-AGN galaxies would most likely remain.

6. Summary and future prospects

In this paper we presented an application of a machine learning algorithm – the support vector machines – to classify sources in an all-sky catalogue drawn from WISE. The algorithm was trained and tested on a sample of WISE objects cross-matched with SDSS spectroscopic data, where three main types of astrophysical sources – stars, galaxies and quasars – had been independently identified. To optimise the performance of SVM, we first determined that a polynomial kernel of the 3rd degree is here preferred over the traditionally used radial one. We next verified that a training sample of less than 10,000 randomly-chosen sources was sufficient to obtain stable results; in addition, the algorithm performed satisfactorily already for a 3-dimensional parameter space (W_1 magnitude; $W_1 - W_2$ colour; differential aperture mag in the W_1 channel).

Having established the optimal setup of the SVM method for our purposes, we performed several tests of its performance on WISE data. Here we focused on completeness and purity as a function of the limiting magnitude of the test sample, as well as on their dependence on Galactic extinction. For stars and galaxies both these statistics deteriorate for increasing magnitudes, but even at the faint end they hardly fall below $\sim 80\%$. On the other hand, no obvious dependence of the SVM performance on magnitude is observed for quasars. Finally, Galactic extinction does

not seem to have influence on the results, although we note that the tests were limited to regions of $EBV \lesssim 0.3$, outside of which there is practically no calibration data.

We finally applied the SVM algorithm, trained on the WISE \times SDSS sample, to the full-sky WISE data flux-limited to $W_1 < 16$ mag. About 220 million of thus preselected sources were flagged by SVM as star candidates; the remaining objects required significant cleanup to obtain galaxy- and QSO-candidate samples. This cleanup consisted in removing the brightest sources, as well as those located in the Galactic Plane and Bulge areas, for which the classification is not expected to be reliable. We also used source type probabilities provided by SVM to remove the objects of insecure classification. As a result, we obtained catalogues of 45 million galaxy candidates, as well as of 6 million ‘QSOs’. In the latter case, however, we observe significant contamination by sources consistent with dusty (non-AGN) galaxies, as well as by most likely Solar System objects.

These shortcomings of our classification are related to the limitations of the training sample, as well as to the lack of additional classification parameters that could be reliably used for the full sample together with the three basic ones employed here. While the former drawback could possibly be mitigated thanks to forthcoming spectroscopic data such as from SDSS-IV, it is not clear how much one could improve the latter, if only WISE data are to be used for the classification on the full sky, while keeping a deep and uniform sample. Measurements in WISE W_3 ($12 \mu\text{m}$) and W_4 ($23 \mu\text{m}$) channels would certainly help to break degeneracies resulting in unreliable identification of quasars in the present approach; these two bands offer however much shallower and very inhomogeneous coverage as compared to W_1 and W_2 . Some improvement in classification should also be expected if reliable proper motions were available for a much larger sample of WISE sources than presently, as they would help identify at least some of the minor bodies of the Solar System which most likely contaminate our current ‘QSO’ sample.

In general, a natural next step in the process of classification of WISE sources is to expand the current scheme to a higher number of object classes, which will allow for creation of more robust catalogues or at least for purification of the current ones. For this to be accomplished, more classification parameters will be necessary, as well as more comprehensive training sets. We plan to explore this in forthcoming studies. Last but not least, it is possible to work on improving the training scheme itself by implementing the so-called *fuzzy logic* (e.g. Klir & Yuan 1995) into the SVM algorithm. While in the classical SVM approach all training examples are treated equally, the fuzzy logic procedure allows for handling the uncertainties of the classification data by weighting the training examples (e.g. Abe & Inoue 2002; Tsujinishi & Abe 2003). Each training point may belong to no more than one class, but by weighting the training points Fuzzy-SVM (FSVM) can assure that the meaningful data points will be classified correctly, while the more noisy ones will have more freedom to be misclassified in order to ensure the maximum margin benefit. This in turn allows to expand the classification regions in the parameter space. However, this approach heavily extends the computational time due to the introduction of the additional free parameter, which, much like other SVM parameters (e.g. misclassification parameter C or kernel parameters) must be tuned for best performance. While this method could help to improve the current classification, the uncertainties of the measurements of the objects considered in this work are relatively small. In view of the largely extended computational time, the FSVM was not favourable for the purpose of the current analy-

sis, but it will be considered in our future studies of classification in more noisy WISE or other data.

Acknowledgements. We thank the referee for the helpful review. Special thanks to Mark Taylor for TOPCAT (Taylor 2005) and STILTS (Taylor 2006) software¹¹. Some of the results in this paper have been derived using the HEALPix package (Górski et al. 2005)¹². This work was supported by the Polish National Science Center under contracts # UMO-2012/07/D/ST9/02785. MB was supported by the Netherlands Organization for Scientific Research, NWO, through grant number 614.001.451, by the European Research Council through FP7 grant number 279396 and by the South African National Research Foundation (NRF). AP was partially supported by the Polish-Swiss Astro Project, co-financed by a grant from Switzerland, through the Swiss Contribution to the enlarged European Union. 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, funded by the National Aeronautics and Space Administration. Funding for SDSS-III has been 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 web site 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 Astrofísica 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.

References

- Abe, S. & Inoue, T. 2002, in in European Symposium on Artificial Neural Networks, 2002, 113–118
- Ahn, C. P., Alexandroff, R., Allende Prieto, C., et al. 2014, ApJS, 211, 17
- Akbani, R., Kwek, S., & Japkowicz, N. 2004, in In Proceedings of the 15th European Conference on Machine Learning (ECML), 39–50
- Alam, S., Albareti, F. D., Allende Prieto, C., et al. 2015, ApJS, 219, 12
- Anderson, L. D., Bania, T. M., Balser, D. S., et al. 2014, ApJS, 212, 1
- Assef, R. J., Stern, D., Kochanek, C. S., et al. 2013, ApJ, 772, 26
- Beaumont, C. N., Williams, J. P., & Goodman, A. A. 2011, ApJ, 741, 14
- 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
- Brown, M. J. I., Jarrett, T. H., & Cluver, M. E. 2014a, PASA, 31, 49
- Brown, M. J. I., Moustakas, J., Smith, J.-D. T., et al. 2014b, ApJS, 212, 18
- Bu, Y., Chen, F., & Pan, J. 2014, New A, 28, 35
- Cavuoti, S., Brescia, M., D’Abrusco, R., Longo, G., & Paolillo, M. 2014, MNRAS, 437, 968
- Chang, C.-C. & Lin, C.-J. 2011, ACM Trans. Intell. Syst. Technol., 2, 27:1
- Cherkassky, V. & Mulier, F. 2006, Learning from Data: Concepts, Theory, and Methods, Second Edition
- Cluver, M. E., Jarrett, T. H., Hopkins, A. M., et al. 2014, ApJ, 782, 90
- Cristianini, N. & Shawe-Taylor, J. 2000, An introduction to Support Vector Machines (Cambridge University Press)
- Cutri, R. M., Wright, E. L., Conrow, T., et al. 2013, Explanatory Supplement to the AllWISE Data Release Products, Tech. rep.
- Dimitriadou, E., Hornik, K., Leisch, F., Meyer, D., & Weingessel, A. 2005, e1071: Misc Functions of the Department of Statistics (e1071), TU Wien, Version 1.5-11.”
- Driver, S. P., Hill, D. T., Kelvin, L. S., et al. 2011, MNRAS, 413, 971
- Edelson, R. & Malkan, M. 2012, ApJ, 751, 52
- Eisenstein, D. J., Weinberg, D. H., Agol, E., et al. 2011, AJ, 142, 72
- Faherty, J. K., Alatalo, K., Anderson, L. D., et al. 2015, ArXiv e-prints [arXiv:1505.01923]
- fan Wu, T., Lin, C.-J., & Weng, R. C. 2003, Journal of Machine Learning Research, 5, 975
- Ferraro, S., Sherwin, B. D., & Spergel, D. N. 2015, Phys. Rev. D, 91, 083533
- Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759
- Hambly, N. C., MacGillivray, H. T., Read, M. A., et al. 2001, MNRAS, 326, 1279
- Hsu, C.-W., Chang, C.-C., & Lin, C.-J. 2003, Bioinformatics, 1, 1
- Ivezic, Ž., Monet, D. G., Bond, N., et al. 2008, in IAU Symposium, Vol. 248, IAU Symposium, ed. W. J. Jin, I. Platais, & M. A. C. Perryman, 537–543
- 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
- Kirkpatrick, J. D., Schneider, A., Fajardo-Acosta, S., et al. 2014, ApJ, 783, 122
- Klir, G. J. & Yuan, B. 1995, Fuzzy Sets and Fuzzy Logic: Theory and Applications (Upper Saddle River, NJ, USA: Prentice-Hall, Inc.)
- Kovács, A. & Szapudi, I. 2015, MNRAS, 448, 1305
- Lin, H.-T., Lin, C.-J., & Weng, R. C. 2007, Mach. Learn., 68, 267
- Mainzer, A., Bauer, J., Cutri, R. M., et al. 2014, ApJ, 792, 30
- Małek, K., Solarz, A., Pollo, A., et al. 2013, A&A, 557, A16
- Mateos, S., Alonso-Herrero, A., Carrera, F. J., et al. 2012, MNRAS, 426, 3271
- Murakami, H., Baba, H., Barthel, P., et al. 2007, PASJ, 59, 369
- Neugebauer, G., Habing, H. J., van Duinen, R., et al. 1984, ApJ, 278, L1
- Nikutta, R., Hunt-Walker, N., Nenkova, M., Ivezic, Ž., & Elitzur, M. 2014, MNRAS, 442, 3361
- Perryman, M. A. C., de Boer, K. S., Gilmore, G., et al. 2001, A&A, 369, 339
- Platt, J. C. 1999, in ADVANCES IN LARGE MARGIN CLASSIFIERS (MIT Press), 61–74
- R Development Core Team. 2005, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0
- Saglia, R. P., Tonry, J. L., Bender, R., et al. 2012, ApJ, 746, 128
- Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525
- Secrest, N. J., Dudik, R. P., Dorland, B. N., et al. 2015, ApJS, 221, 12
- Shawe-Taylor, S. & Cristianini, N. 2004, Kernel Methods for Pattern Analysis (Cambridge, UK: Cambridge, UP)
- Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163
- Solarz, A., Pollo, A., Takeuchi, T. T., et al. 2012, A&A, 541, A50
- Soumagnac, M. T., Abdalla, F. B., Lahav, O., et al. 2015, MNRAS, 450, 666
- Stern, D., Assef, R. J., Benford, D. J., et al. 2012, ApJ, 753, 30
- Taylor, M. B. 2005, in Astronomical Society of the Pacific Conference Series, Vol. 347, Astronomical Data Analysis Software and Systems XIV, ed. P. Shopbell, M. Britton, & R. Ebert, 29
- Taylor, M. B. 2006, in Astronomical Society of the Pacific Conference Series, Vol. 351, Astronomical Data Analysis Software and Systems XV, ed. C. Gabriel, C. Arviset, D. Ponz, & S. Enrique, 666
- Tsujinishi, D. & Abe, S. 2003, Neural Networks, 16, 785 , advances in Neural Networks Research: [IJCNN] ’03
- Tu, X. & Wang, Z.-X. 2013, Research in Astronomy and Astrophysics, 13, 323
- Vapnik, V. 1999, IEEE Transactions on Neural Networks, 10, 988
- Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, AJ, 140, 1868
- Wu, X.-B., Hao, G., Jia, Z., Zhang, Y., & Peng, N. 2012, AJ, 144, 49
- Yan, L., Donoso, E., Tsai, C.-W., et al. 2013, AJ, 145, 55
- York, D. G., Adelman, J., Anderson, Jr., J. E., et al. 2000, AJ, 120, 1579

Appendix A: Tables with detailed results of the tests

In this appendix we provide tables with detailed results of the tests described in Section 4. Tables A.1 and A.2 summarize the statistics of the completeness, purity and contamination for various combinations of extinction bins and flux limits in the test sets, for the self-check and cross-test cases.

¹¹ <http://www.star.bris.ac.uk/~mbt/>

¹² <http://healpix.jpl.nasa.gov/>

SELF-CHECK						
Magnitude limit	$W1 < 14 \text{ mag}$					
Extinction [MJy/sr]	$\langle 0; 1 \rangle$			$\langle 1; 2 \rangle$		
Galaxy	95.0	91.8	8.2	90.1	89.9	10.1
Stars	94.5	97.8	2.2	97.5	97.5	2.5
QSO	96.9	97.1	2.9	97.0	96.9	3.1
Extinction [MJy/sr]	$\langle 2; 3 \rangle$			$\langle 3; 10 \rangle$		
Galaxy	Completeness	Purity	Contamination	Completeness	Purity	Contamination
Galaxy	94.6	91.3	8.7	94.5	92.2	7.8
Stars	94.6	96.9	3.1	94.9	97.1	2.9
QSO	95.8	97.0	3.0	97.0	97.3	2.7
Magnitude limit	$W1 < 15 \text{ mag}$					
Extinction [MJy/sr]	$\langle 0; 1 \rangle$			$\langle 1; 2 \rangle$		
Galaxy	Completeness	Purity	Contamination	Completeness	Purity	Contamination
Galaxy	93.3	89.1	10.9	92.2	87.8	12.2
Stars	91.5	94.9	5.1	90.5	93.8	6.2
QSO	97.0	98.2	1.8	96.6	98.0	2.0
Extinction [MJy/sr]	$\langle 2; 3 \rangle$			$\langle 3; 10 \rangle$		
Galaxy	Completeness	Purity	Contamination	Completeness	Purity	Contamination
Galaxy	92.3	88.4	11.6	92.5	88.5	11.5
Stars	91.4	93.9	6.1	92.6	94.2	5.8
QSO	96.1	97.9	2.1	95.3	98.1	1.9
Magnitude limit	$W1 < 16 \text{ mag}$					
Extinction [MJy/sr]	$\langle 0; 1 \rangle$			$\langle 1; 2 \rangle$		
Galaxy	Completeness	Purity	Contamination	Completeness	Purity	Contamination
Galaxy	87.7	82.7	17.3	87.2	79.3	20.7
Stars	84.4	88.0	12.0	81.4	87.1	12.9
QSO	96.6	98.5	1.5	95.0	98.4	1.6
Extinction [MJy/sr]	$\langle 2; 3 \rangle$			$\langle 3; 10 \rangle$		
Galaxy	Completeness	Purity	Contamination	Completeness	Purity	Contamination
Galaxy	86.6	78.5	21.5	87.6	79.0	21.0
Stars	81.2	86.6	13.4	82.9	88.1	11.9
QSO	94.1	98.1	1.9	93.3	98.1	1.9

Table A.1. Overall classification statistics (in %) for various combinations of extinction bins and flux limits, for the self-check case (classified objects were the same as in the training sample).

CROSS-TEST						
Magnitude limit	$W1 < 14 \text{ mag}$					
Extinction [MJy/sr]	$\langle 0; 1 \rangle$			$\langle 1; 2 \rangle$		
Galaxy	94.7	91.7	8.3	94.7	89.9	10.1
Stars	94.3	97.6	2.4	93.3	97.5	2.5
QSO	97.0	96.9	3.1	95.9	96.9	3.1
Extinction [MJy/sr]	$\langle 2; 3 \rangle$			$\langle 3; 10 \rangle$		
Galaxy	94.2	90.7	9.3	93.8	91.3	8.7
Stars	94.2	96.6	3.4	94.5	96.6	3.4
QSO	95.4	96.8	3.2	96.5	97.0	3.0
Magnitude limit	$W1 < 15 \text{ mag}$					
Extinction [MJy/sr]	$\langle 0; 1 \rangle$			$\langle 1; 2 \rangle$		
Galaxy	92.9	88.6	11.4	91.8	87.5	12.5
Stars	91.2	94.5	5.5	90.2	93.5	6.5
QSO	96.6	98.0	2.0	96.4	97.8	2.2
Extinction [MJy/sr]	$\langle 2; 3 \rangle$			$\langle 3; 10 \rangle$		
Galaxy	91.8	87.5	12.5	92.2	87.9	12.1
Stars	90.5	93.4	6.6	92.0	94.2	5.8
QSO	96.0	97.8	2.2	95.2	97.7	2.3
Magnitude limit	$W1 < 16 \text{ mag}$					
Extinction [MJy/sr]	$\langle 0; 1 \rangle$			$\langle 1; 2 \rangle$		
Galaxy	85.5	81.7	18.3	84.5	77.6	22.4
Stars	84.0	86.4	13.6	80.3	85.0	15.0
QSO	96.1	98	2.0	94.3	97.7	2.3
Extinction [MJy/sr]	$\langle 2; 3 \rangle$			$\langle 3; 10 \rangle$		
Galaxy	84.9	77.3	22.7	86.0	77.4	22.6
Stars	80.5	85.3	14.7	81.5	87.1	12.9
QSO	93.6	97.7	2.3	92.9	97.5	2.5

Table A.2. Overall classification statistics (in %) for various combinations of extinction bins and flux limits, for the cross-test case (classified objects were different from those in the training sample).