



# Quasar lenses and galactic streams: outlier selection and *Gaia* multiplet detection

Adriano Agnello<sup>★</sup>

European Southern Observatory, Karl-Schwarzschild-Strasse 2, D-85748 Garching bei München, Germany

Accepted 2017 June 28. Received 2017 June 26; in original form 2017 April 21

## ABSTRACT

I describe two novel techniques originally devised to select strongly lensed quasar candidates in wide-field surveys. The first relies on outlier selection in optical and mid-infrared magnitude space; the second combines mid-infrared colour selection with *Gaia* spatial resolution, to identify multiplets of objects with quasar-like colours. Both methods have already been applied successfully to the Sloan Digital Sky Survey, ATLAS and Dark Energy Survey footprints: besides recovering known lenses from previous searches, they have led to new discoveries, including quadruply lensed quasars, which are rare within the rare-object class of quasar lenses. As a serendipitous by-product, at least four candidate Galactic streams in the South have been identified among foreground contaminants. There is considerable scope for tailoring the *WISE*–*Gaia* multiplet search to stellar-like objects, instead of quasar-like, and to automatically detect Galactic streams.

**Key words:** gravitational lensing: strong – methods: statistical – surveys – Galaxy: structure.

## 1 INTRODUCTION

The many virtues of strongly lensed quasars, both for astrophysics and cosmography, are hampered by the relative paucity of systems with sufficient ancillary data. On the one hand, lensed quasars are valuable probes of cosmological distances and stellar and dark matter in  $z \approx 1$  galaxies (Courbin, Saha & Schechter 2002): the delays between light curves of different quasar images can be translated in one-step measurements of the Hubble constant (Refsdal 1964), making a low-redshift complement to cosmic microwave background experiments (Suyu et al. 2017); lens modelling and source-reconstruction enables super-resolved studies of quasar host galaxies at  $z \approx 2$  (Ding et al. 2017); and micro-lensing yields a detailed view of the source central engine (Bate et al. 2011; Sluse et al. 2012; Hutsemékers et al. 2015) and luminous and dark mass in the deflector (Oguri, Rusu & Falco 2014; Schechter et al. 2014). On the other hand, a few tens of these systems are currently known and few of them are suitable for time-delay cosmography or detailed lens modelling, whence the need for larger samples. With one every  $\mathcal{O}(10^4)$  quasars being strongly lensed (Oguri & Marshall 2010), and  $\approx 0.2$  quasar lenses per square degree, these are a class of rare objects to be mined in wide-field surveys. In particular, predicted quasar lens samples are predominantly doubles, with only  $\approx 14$  per cent being highly valuable quads.

Different searches, tailored to different data sets and surveys, have been developed to find new lensed quasars. The Cosmic Lens All Sky Survey (CLASS; Myers et al. 2003) and a parallel JVLA search (King et al. 1999) targeted radio-loud objects that could be

resolved in multiple components by follow-up observations; the Sloan Quasar Lens Search (SQLS; Oguri et al. 2006) and its BOSS extension (BQLS; More et al. 2016) targeted objects that were identified as quasars from previous fibre spectroscopy. Both CLASS and SQLS/BQLS uncovered the bright end of the lensed quasar population, and were focused on either radio or UV excess pre-selection (for the challenges in observing radio-quiet lensed quasars; see e.g. Jackson et al. 2015). In order to expand this effort to fainter systems and higher redshift, a variety of techniques (Morgan et al. 2004; Agnello et al. 2015a; Ostrovski et al. 2017; Williams, Agnello & Treu 2017a) has been tailored to wide-field photometric and morphological searches.

Here, I illustrate two novel search methods that I have developed and used to discover new quasar lenses and whose performance has been complementary to other searches applied over the last year. The first, described in Section 2, relies on outlier selection in the optical and mid-infrared, selecting lensed quasar targets among objects that do not ‘obviously’ belong to classes/clusters of more common contaminants. The second, described in Section 3, combines the good spatial resolution and depth of the *Gaia* mission (Gaia Collaboration 2016; Lindegren et al. 2016; van Leeuwen et al. 2017) with a mid-infrared colour selection of quasar-like objects. Four new Galactic streams, found as a by-product of this method, are also briefly described. Given the setup of this search (tailored on quasar lenses), this discovery may be considered serendipitous, so I will outline how a similar search can be tailored on Galactic substructure. Concluding remarks are summarized in Section 4. Discoveries of quasar lenses from both methods, involving different collaborations, are reported elsewhere.

In what follows, some nomenclature will be consistently used. *Objects* are selected at query level from wide-field surveys, *targets*

\* E-mail: [aagnello@eso.org](mailto:aagnello@eso.org)

are a sub-sample of objects selected based on their catalogue properties, and *candidates* are a sub-sample of targets further selected based on their images either via visual inspection or cutout modelling. For the sake of brevity, I designate as quasar ‘pairs’ both physically associated quasars and chance alignments of quasars at different redshifts. The same holds for line-of-sight (LOS) quasar-star ‘pairs’.

Wide-field surveys are abbreviated as follows: SDSS is the Sloan Digital Sky Survey (Abazajian et al. 2009); ATLAS is the VLT Survey Telescope ATLAS survey (Shanks et al. 2015); *WISE* is the Wide Infrared Survey Imager (Wright et al. 2010); DES is the Dark Energy Survey (Sánchez & DES Collaboration 2010); iPTF is the intermediate Palomar Transient Factory (Law et al. 2009) DR3 and PS1 is the first Pan-STARRS (Kaiser et al. 2010; Chambers et al. 2016) telescope data release. Throughout this paper, *griz* magnitudes are in the AB system, and mid-infrared *W1*, *W2*, *W3* magnitudes from *WISE* are in the Vega system.

## 2 OUTLIER SELECTION

Depending on the survey image quality and depth, and lens configuration and image separation, lensed quasars have colours intermediate between those of the source quasars and those of the deflector galaxies. Likewise, they can result in groups of point-like or extended sources, or as extended objects due to blending by image processing pipelines. Due to their rarity and intermediate colours and morphologies, lensed quasar candidate samples suffer from significant contamination by more common classes of objects. This problem can be mitigated by selecting objects that do *not* have typical colours of more common contaminants.

As a *training* set for the outlier-selection method, I use the 10 quasar lenses and 40 false-positives in the morphologically selected sample of Inada et al. (2012). This is because most quasar lenses are marginally deblended by the pipelines of ground-based surveys, and those with larger separation in previous searches have colours that are more typical of nearby quasars, due to the UVx and spectroscopic pre-selections that were applied there. The 10+40 morphological candidates of Inada et al. (2012) should then be a good guidance to a homogeneous sample of quasar-like objects with nearby companions, with different kinds of contaminants, and small enough that data mining techniques trained on it do not overspecialize, thus remaining complementary to previous searches.

As a *test* set, to evaluate the method performance on a wider sample, I use a list of 149 known lenses compiled from the CASTLES<sup>1</sup> data base and SQLS full sample from SDSS-DR7 given by Inada et al. (2012). In particular, I will consider the 132 known systems with Dec.  $> -20$  for tests on the SDSS footprint. These are the same used by Williams et al. (2017a) to evaluate the performance of population-mixture methods on quasar lenses in different ranges of image separation. Some of the outlier-selected targets were later found to be already known quasar lenses and pairs in the BQLS sample of More et al. (2016).

### 2.1 Setup, clusters and object classes

In order to describe different populations in colour–magnitude space, and similarly to previous work (Williams et al. 2017a), I consider  $g - r$ ,  $g - i$ ,  $r - z$ ,  $i - W1$ ,  $W2$ ,  $W1 - W2$  and  $W2 - W3$ ,

<sup>1</sup> <https://www.cfa.harvard.edu/castles/>

compressing the catalogue information to a seven-dimensional *feature* space. I do not use UV excess information, since it is not always available in current wide-field surveys (e.g. DES), and it is less efficient at identifying quasars at redshifts  $z_s \gtrsim 2.5$ , where Ly  $\alpha$  emission exits the *u* band.

Object pre-selection is based on their extended morphology and some loose colour–magnitude requirements. As a morphological pre-selection, I concentrate on objects that have  $psf\_r\_mod\_r > 0.075$  and  $psf\_i\_mod\_i > 0.075$  (referred below as *magnitude* criterion), or  $\log_{10} \mathcal{L}_{star,i} < -11$  (resp. *stellarity* criterion).

The *psf* and *model* magnitudes, as well as the *i*-band stellarity likelihood  $\log \mathcal{L}_{star,i} = \ln LStar_i$ , are taken directly from the SDSS catalogue.<sup>2</sup> Their definition is, in fact, somewhat survey specific and different thresholds must be explored for ATLAS, DES and PS1, separately for each survey. Since this search is tailored on lensed quasars, I further select objects satisfying

$$\begin{aligned} W1 - W2 &> 0.55, \quad 2.2 < W2 - W3 < 3.8, \\ W1 &< 17.0, \quad W2 < 15.4, \quad W3 < 11.6, \\ \delta W1 &< 0.25, \quad \delta W2 < 0.3, \quad \delta W3 < 0.35, \\ 2.2 < i - W1 &< 5.75, \quad i - W3 < 8.9, \\ g - i &< \max(0.65; 1.2(i - W1) - 2.4), \\ g - i &< 2.55, \quad r - z < 1.5, \quad 15.0 < i < 20.5, \end{aligned} \quad (1)$$

which eliminates most stellar contaminants, blue galaxies and low-redshift quasars. Here, *griz* magnitudes are SDSS *model* magnitudes, *WISE* *W*X magnitudes are *wXmpo* and  $\delta W^X$  are the uncertainties *wXsigmpo* on the corresponding magnitude ( $X = 1, 2, 3$ ). Magnitude-selected extended objects will also be split into ‘c0’ objects, having

$$W2 - W3 < \max[2.7; 3.15 + 1.5(W1 - W2 - 1.075)], \quad (2)$$

and ‘c1’ objects, occupying the remaining wedge in *WISE* colours. This colour distinction delimits the locus where all SQLS training-set objects lie, and roughly traces the distinction between quasar-dominated and galaxy-dominated objects (e.g. Wright et al. 2010).

Main contaminant classes are identified using Gaussian population mixture as a clustering algorithm, which has been validated as a means of object classification across different surveys (e.g. Bovy et al. 2011; DiPompeo et al. 2015; Chehade et al. 2016; Tie et al. 2017; Williams et al. 2017a, where hybrid cuts on  $g - i$  and  $i - W1$  similar to the above have been tested). A recurring theme of semisupervised clustering and classification methods is whether class parameters should be initialized: (i) based on where we expect different objects to lie a priori or (ii) identifying clusters independently and labelling them a posteriori. Here, I choose the second option, even though some clusters can be easily interpreted in terms of known object populations, as will be done below.

Each class is then described by a mean  $\mu$  and a covariance matrix  $\Sigma$ . The relative class abundances will not be used at classification stage. To each object, with feature vector  $\mathbf{f}_i$ , a pseudo-distance to the  $k$ th class is defined as

$$d_{i,k} = \langle (\mathbf{f}_i - \mu_k), \Sigma_k^{-1}(\mathbf{f}_i - \mu_k) \rangle / 2. \quad (3)$$

<sup>2</sup> A description can be found in the SDSS Schema Browser, e.g. at <http://skyserver.sdss.org/dr7/en/help/browser/browser.asp>, Table ‘PhotoObjAll’.

**Table 1.** Means  $\mu_{k,j}$  and widths  $\sqrt{\Sigma_{k,ii}}$  of the object clusters computed as in Section 2.1; the association between clusters and classes is done a posteriori, and described in the text.

$k$	$g - r$	$g - i$	$r - z$	$i - W1$	$W1 - W2$	$W2$	$W2 - W3$
1	0.15	0.44	0.48	4.34	0.89	–	2.80
	0.31	0.37	0.21	0.32	0.14	–	0.32
2	0.32	0.52	0.25	4.45	0.55	16.30	3.90
	0.20	0.32	0.29	1.07	0.61	0.35	0.50
3	0.16	0.14	0.00	3.60	1.24	14.90	2.94
	0.10	0.13	0.17	0.15	0.11	0.47	0.15
4	0.95	1.40	0.78	3.51	0.55	–	3.5
	0.24	0.30	0.16	0.41	0.40	–	0.55
5	0.35	0.61	0.61	4.17	0.82	14.19	3.12
	0.16	0.18	0.16	0.48	0.13	0.32	0.20
6	0.48	0.88	0.84	4.16	0.92	13.90	2.35
	0.25	0.25	0.12	0.47	0.21	0.48	0.36
7	0.22	0.25	0.44	4.50	0.79	14.20	3.10
	0.22	0.38	0.21	0.47	0.11	0.30	0.26
8	0.17	0.25	0.60	4.56	1.03	13.18	2.49
	0.24	0.39	0.23	0.46	0.12	0.53	0.28
9	0.15	0.25	0.32	3.05	1.00	14.12	3.20
	0.11	0.12	0.10	0.20	0.13	0.33	0.28
10	0.13	0.31	0.34	3.04	1.26	13.63	3.30
	0.09	0.15	0.10	0.25	0.09	0.37	0.23

The mean and covariance of each class are computed iteratively using Expectation–Maximization. In order to ensure convergence, I borrow from the strategy of adaptive second moments (Bernstein & Jarvis 2002): at each step in the computations of  $\mu_k$  and  $\Sigma_k$ , each object is additionally weighed with  $e^{-d_{i,k}}$ , multiplying the covariances by 2. This avoids iteration instabilities due to neighbouring classes or ‘large’ widths in the classes. Some of the mean features are kept fixed, as they have values outside those of the queried sample (e.g.  $W1 - W2$  of galaxies) or are unstable to Expectation–Maximization.

The first four classes are common to objects selected via magnitude or stellarity morphological criteria. Most of the magnitude-selected sample is described by six clusters, whereas eight clusters are needed to encompass the majority of the stellarity-selected sample. Table 1 lists the averages  $\mu_k$  and for the sake of brevity the widths  $\sqrt{\Sigma_{k,jj}}$  ( $j = 1, \dots, 7$ ) of the classes. By comparing the colours of different classes with those from spectroscopic subsamples (see e.g. Williams et al. 2017a), one can roughly associate the first four classes with: (i) isolated quasars at  $z_s \lesssim 0.35$ , which have bright  $W2$ , high  $i - W1$  and low  $g - i$  but can extend to redder colours due to contribution from their host galaxies; (ii) isolated quasars at redshift  $z_s \approx 2$ , with fainter  $W2$ ,  $g - i \lesssim 0.6$  and lower  $i - W1$ ; (iii) narrow-line galaxies at  $z \approx 0.2 - 0.3$ , with  $W2$  and  $i - W1$  comparable to  $z_s \approx 2$  quasars but higher  $g - i$  and (iv) fainter galaxies with  $W1 - W2$  spanning a wide range. In fact, the empirical *WISE* colour–magnitude cuts of Assef et al. (2013) were designed to minimize contamination from the fourth class leaking into *WISE*-selected quasar samples.

## 2.2 Adapting on SQLS training set

Once the classes/clusters are defined, targets must be selected based on how far they lie from different classes. In order to do so, I use linear combinations of the pseudo-distances  $d_k$ . The first set of

pseudo-distance cuts, using  $d_1, \dots, d_4$  as

$$\begin{aligned} d_1 &\geq 4.5, d_2 \geq 15, d_4 \geq 4, \\ 12d_2 - 17 &\geq d_3 \geq 225 - 3d_2, \\ d_3 &\geq 175 - 10d_1, d_3 \geq 6d_1 - 20, \\ 125 - 4d_4 &< d_3 \leq 100d_4 \end{aligned} \quad (4)$$

is displayed in Fig. 1. Some of the cuts can be made more or less restrictive, with the aim of retaining as many quasar-like systems as possible, while rejecting galaxies and SQLS false-positives without penalizing the SQLS true-positives.

With this combination, 9 out of 10 training-set lenses are retained, and 20 out of 40 false-positives are rejected. At face value, this would mean that outlier selection is 90 per cent complete with respect to the SQLS and has half as many contaminants, but matters are more complicated. One the one hand, the SQLS and BQLS relied on spectroscopic information, so contamination by narrow-line galaxies is completely suppressed there, while it is non-negligible in photometric searches. On the other, the same spectroscopic pre-selection limits the SQLS and BQLS to objects for which spectra are available, and has in fact precluded the discovery of some noteworthy lenses.

When the first cuts of equations (1) and (4) are applied to objects in the SDSS-DR12 footprint, 8543 (resp. 15839) objects are still left in the magnitude-selected (resp. stellarity-selected) sample.<sup>3</sup> In the colour–colour diagrams of remaining objects, there are some overdensities corresponding to (apparently) isolated quasars with  $z_s = 0.5 \pm 0.1$ , based on the available fibre spectra. Their  $\mu$  values are different between magnitude-selected (two clusters,  $k = 5, 6$ ) and stellarity-selected (four clusters,  $k = 7, \dots, 10$ ) objects. The lower panels of Fig. 1 show where objects surviving the cuts of equation (4) lie in terms of the distances from additional clusters ( $d_5, \dots, d_{10}$ ).

To further guide the selection, I visually inspected the first-pass targets and retained 82 magnitude-selected and 157 stellarity-selected candidates, which are not ‘obvious’ kinds of contaminants such as blue galaxies or nearby quasars and Seyfert galaxies. Their pseudo-distance distribution, occupying mostly the outskirts of the object clusters (see Fig. 1), suggests the following cuts

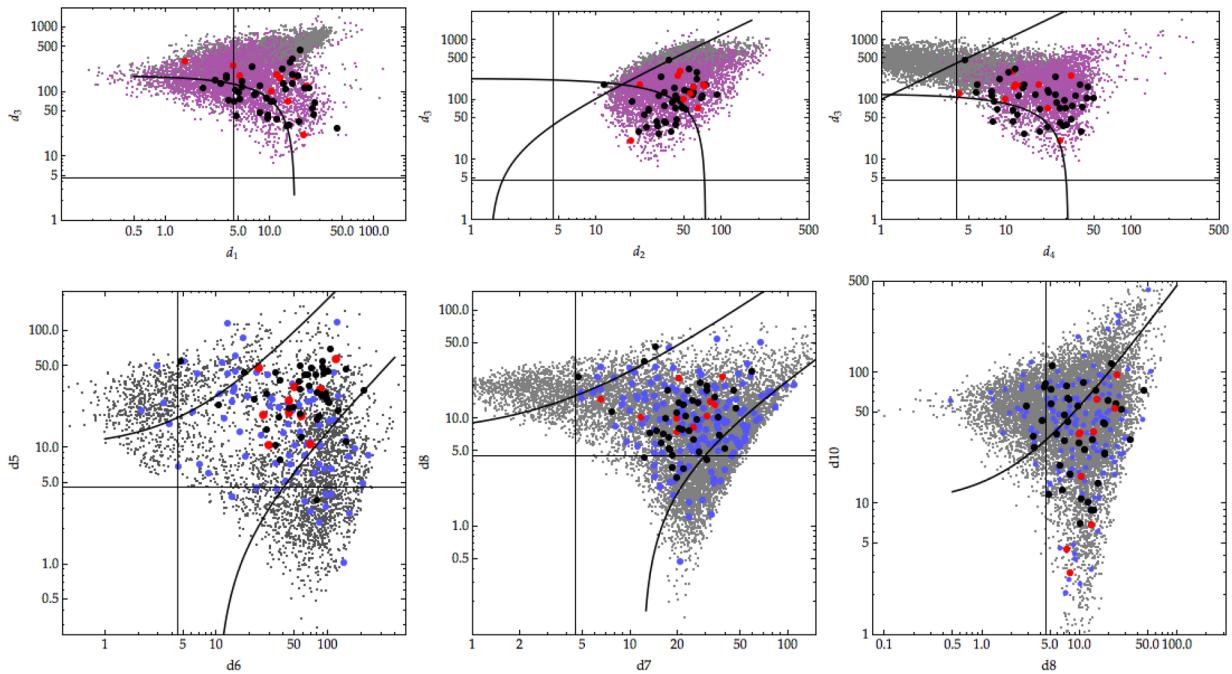
$$0.15d_6 - 1.5 \leq d_5 \leq 1.75d_6 + 10 \quad (5)$$

for magnitude-selected objects, and

$$\begin{aligned} 2.0d_7 + 7 &\geq d_8 \geq 0.25d_7 - 3, \\ d_{10} &\leq 4.5d_8 + 10, d_9 \leq 10d_8 + 10 \end{aligned} \quad (6)$$

for stellarity-selected objects. Once applied to the first-pass candidates, they result in 3728 magnitude-selected and 4712 stellarity-selected targets, 60 per cent of which are non-repeated catalogue entries. This has reduced the initial  $\approx 5 \times 10^5$  queried objects to a manageable sample for visual inspection; in fact, the cuts in  $d_1, \dots, d_4$  were already enough to obtain a reasonable reduction in objects to be inspected, whence the 82 + 157 candidates were obtained.

<sup>3</sup> These are *non-unique* identifications, as some of the objects are identified and retained multiple times, through the whole selection process from query to final candidate sample. Non-repeated objects are  $\approx 60$  per cent.



**Figure 1.** Cuts in pseudo-distances from main clusters/classes. Red (resp.) black bullets mark SQLS lenses (resp. false positives) in the training set. *Top:* First set of cuts on  $d_1, \dots, d_4$ ; grey (resp. purple) swarms trace magnitude-selected (resp. stellarity-selected) objects. *Bottom:* Cuts in  $d_5, \dots, d_{10}$  depending on the extended-morphology criterion; light-blue bullets mark the candidates selected after visual inspection of the first-pass targets, and c0 (resp. c1) targets correspond to high  $d_6$  (resp.  $d_5$ ). The axes are shown mostly as a guidance to the eye.

### 2.3 Blind test on SDSS

Using the CASTLES+SQLS *test* set introduced above, we can quantify how many known lenses are lost at each stage and why. At pre-selection level, 57 of the 132 test objects are retained, due primarily to the extendedness criteria as noticed already by Williams et al. (2017a), and secondarily to *WISE* colour selection. Of these, 35 remain after the first cuts (in equation 4); 36 satisfy equations (5) and (6) and only 22 satisfy the cuts in all pseudo-distances. Most of the rejected lenses lie close to the selection boundaries, which could be re-adjusted post hoc to increase the completeness. However, I preferred to perform a blind test of this method, trained solely on 10 lenses and 40 non-lenses.

Fig. 2 shows the distribution of queried objects (grey swarm), candidates selected after  $d_1, \dots, d_4$  cuts (blue bullets) and known lenses (red star symbols). One every 40 queried objects is shown, for convenience. Inhomogeneous spatial coverage is a direct consequence of the SDSS scanning, and affects the distribution of targets, candidates and known lenses. Candidates and known lenses have compatible distributions in *WISE* colours (not used by previous campaigns) and morphological parameters.

This outlier-selection seems to retain known quads and doubles alike. Some lenses are present multiple times in the queried sample, the most popular being: J2343–0050 (78 matches, m.), J1001+5553 (20 m.), J0145–0945 (7 m.), J1206+4332 (6 m.), J0806+2006 (5 m.), J1304+2001 (4 m.). Besides these, 10 lenses are flagged 3 times by the object query, and 20 are flagged twice. Some BQLS objects, not included in the test set, have been rediscovered as well. When applied to the DES catalogue, with suitably translated magnitudes (as discussed by Agnello et al. 2017a), it recognized both large-separation lenses like DES0408 (Lin et al. 2017) and small-separation blends like DES0115 (Agnello et al. 2015b). The main reason is that, while all these lenses have markedly quasar-

like *WISE* colours, their hybrid colours are not typical of unlensed and low-redshift quasars.

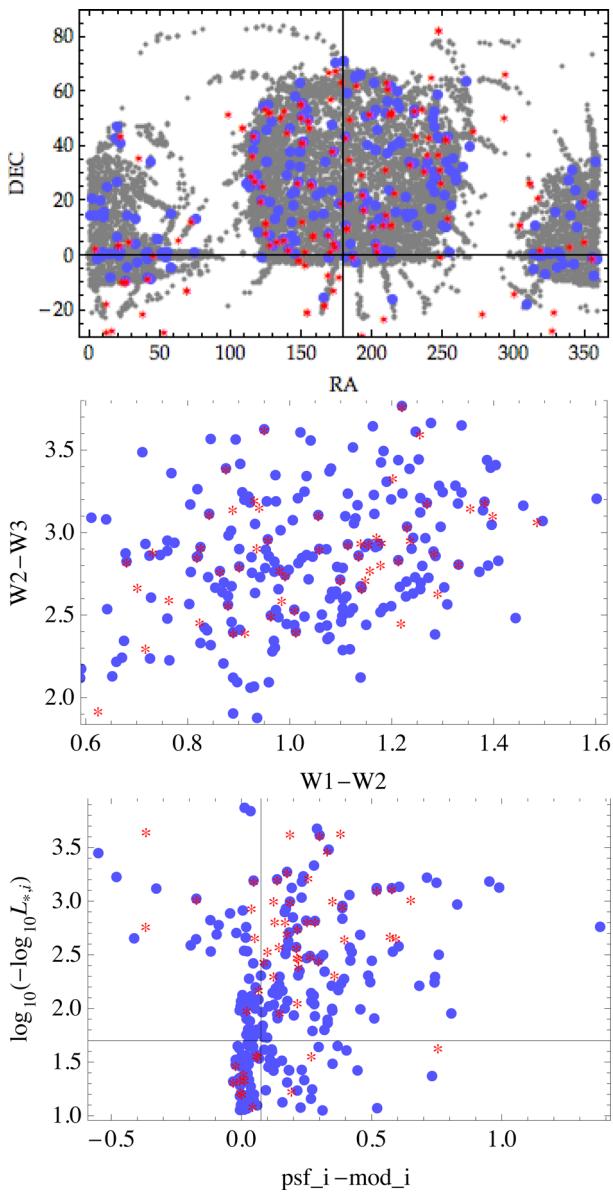
Whether quasar ‘pairs’, which are interesting for other applications, are eliminated or retained depends chiefly on how much their overall colours resemble those of low-redshift quasars. Based on the results of different spectroscopic follow-up campaigns,<sup>4</sup> quasar pairs with sources at  $z_s \approx 2$  are present among quasar lens candidates. Some quasar pairs with nearly identical spectra are also present, but whether they are lenses or physically associated quasars will require deeper follow-up.

### 3 GAIA MULTIPLET DETECTION

By using different all-sky surveys, one can combine their separate advantages and exploit the largest possible footprint. In particular: *WISE* enables a mid-IR selection of quasar-like objects, gathering the light from different quasar images within the same  $\approx 6$  arcsec beam; while *Gaia* offers higher spatial resolution, enabling a distinction between isolated quasars and quasars with companions within the same *WISE* beam. From this viewpoint, combining *WISE* and *Gaia* for quasar lens searches is reminiscent of the CLASS/JVLA strategy of radio-loud detection and higher resolution follow-up.

Here, I illustrate the general properties of objects selected with *WISE* W1, W2, W3 and *Gaia* DR1 data, using the CASTLES+SQLS known systems as guidance. With the addition of SQLS quasar pairs, the test set amounts to 197 objects. Visual inspection is needed to obtain a candidate sample, so I will explore the performance of this combined search over three footprints: one covering the SDSS-DR12; one covering the ATLAS-DR3; and one covering slightly more than the DES-Y3. For SDSS and ATLAS, I will also rely on

<sup>4</sup> Papers by Williams et al. (2017b), Agnello et al. (in preparation) and by the STRIDES collaboration (strides.astro.ucla.edu).



**Figure 2.** Distribution of query-selected objects (grey), candidates (light-blue bullets) and known lenses (red stars). The SDSS coverage is not homogeneous, as reflected at all stages of query, target- and candidate-selection and by the known systems, which have consistent distributions in *WISE* colours and stellarity/psf-mod distribution. The axes are shown mostly as a guidance to the eye.

public survey catalogues and images, whereas for the DES footprint I will delineate some general properties based solely on *WISE* and *Gaia* data. To quantify how many lens candidates can be expected with this method, I select objects with  $W1, W2, W3$  as in Section 2 (over the three considered footprints) and match them to *Gaia* using a 6 arcsec search radius. In the SDSS and ATLAS, when *griz* catalogue magnitudes are available, I also require that objects satisfy the colour-magnitude requirements of equations (1) and (2). As a separate footprint, to examine the role of hybrid optical-IR colours, I will study objects with Dec.  $>-30$  deg that are detected also in the iPTF catalogue. These have the advantage that they can also be visually inspected using PS1 images, and that quasar-like objects can be further selected based on their variability signatures (Schmidt et al. 2012).

As an illustration of the method, and to minimize contamination by star-quasar pairs, the *WISE-Gaia* matches will be performed on objects with  $G > 15.0$  and  $|b| > 20.0$  and visually inspected for  $|b| > 30.0$  when possible. This is also why the recent quad candidate of Berghea et al. (2017) is discarded at pre-selection, while (at least) four known quads are retained: HE0435 (Wisotzki et al. 2002), RXJ1131 (Sluse et al. 2003), PG1115 and SDSS J1433+60.

### 3.1 General behaviour

Most known lenses and pairs correspond to one detection per system in *WISE*, whereas 55 of them are recognized as separate objects in *Gaia*-DR1. Their overall properties are summarized in Fig. 3: the *WISE-Gaia* match can be performed down to  $G = 20.7$ ; most lenses with a *Gaia* multiplet counterpart have  $G < 20$ ; and the flux-ratios of different components are not extreme, mostly  $|2.5 \log_{10}(f_1/f_2)| \lesssim 2$ .

The grey points in Fig. 3 display general properties of the *WISE-Gaia* multiplets, for a random sub-set, showing the same behaviour as for the known lenses in red. This first match results in 1868 multiplets in SDSS-DR12, 648 in ATLAS-DR3, and 2679 in the approximate DES footprint. All of these are manageable numbers for a quick visual inspection, and have led to new lenses being discovered with this method, which will be reported elsewhere.

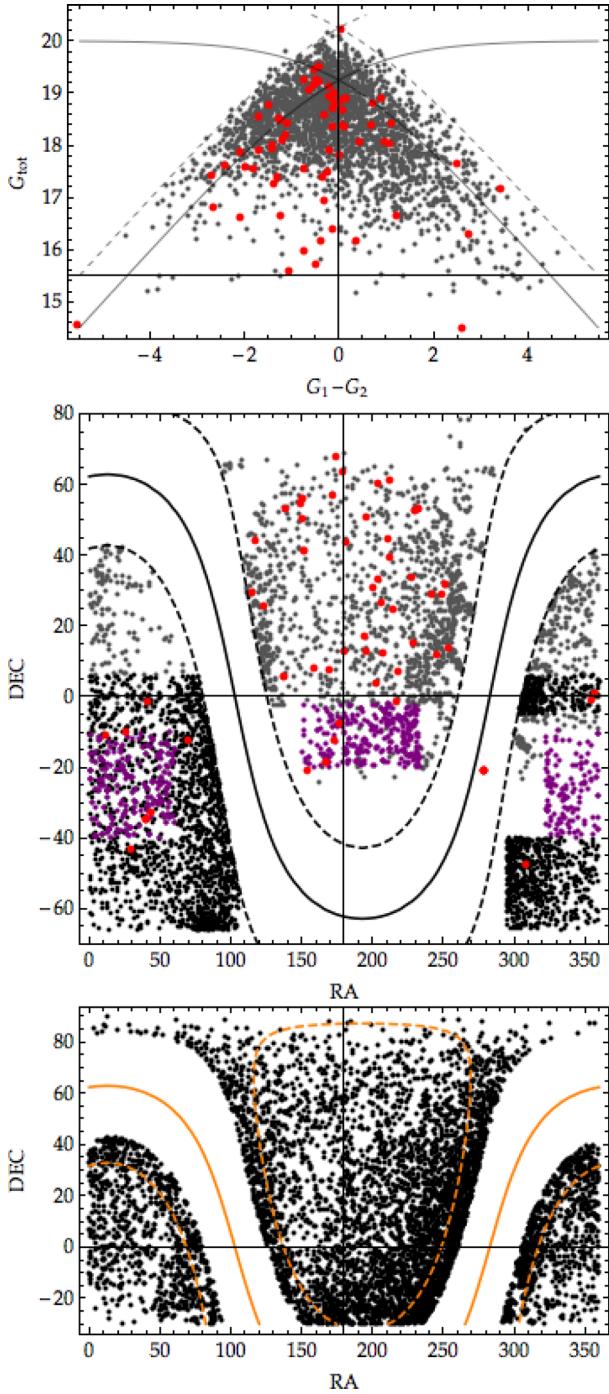
Exploiting the homogeneous (sometimes spectroscopic) coverage of SDSS, we can further quantify the fraction of promising multiplets. A quick visual inspection reveals different contaminant classes: galaxy groups and mergers; stars, especially at low galactic latitudes  $|b| < 33$  deg; quasar-star and quasar-galaxy LOS alignments. Those that are not ‘obvious’ contaminants amount to 127 candidates, some more convincing than others. By construction, they include known lenses with different configurations (quads and doubles) and image separations.

The situation in other footprints is complicated by the patchy (ATLAS) or limited (DES) coverage of public catalogues. Visual inspection of ATLAS-DR3 targets with at least one valid magnitude results in 28 unique candidates. The known quad RXJ1131 is flagged four times by this search (corresponding to its four quasar images), even though it is covered just in  $i$  band in the public footprint. Visual inspection produces one candidate every 20–30 targets, so one can expect 60–100 candidates to be selected in the DES footprint with this method; the depth, image-quality and *grizY* coverage of DES also help to further select the most promising systems for follow-up.

Since the multiplet search targets objects with quasar-like *WISE* colours and nearby neighbours, it also retains LOS alignments of quasars and stars. This aspect is further discussed below.

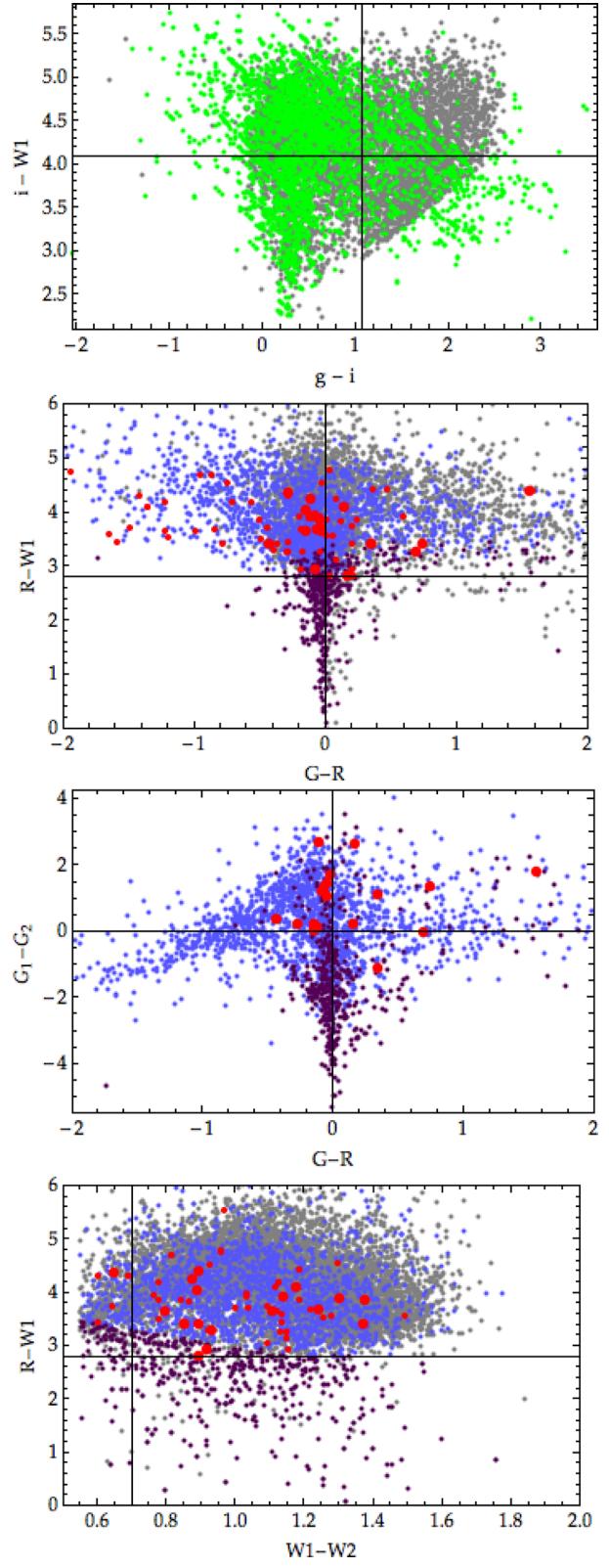
### 3.2 Hybrid colours

Since hybrid colours help separate quasar-like objects from galaxies and stars, an ideal search would incorporate the multiplet information from *Gaia* with optical/IR colours to reduce contamination. The top panel in Fig. 4, displaying hybrid colours with *Gaia*, SDSS and *WISE*, suggests that these may still be used to aid the *WISE-Gaia* multiplet detection in the case of partial (e.g. only  $i$  band) ground-based coverage. There, I am using  $0.175 + G - i$  (green points) instead of  $g - i$  (grey swarm) to quantify the red excess of SDSS objects corresponding to *Gaia* multiplets. Indeed, most quasar-like objects have compatible colours across SDSS and *Gaia*. Some objects have  $G - i$  versus  $i - W1$  falling beyond the pre-selection boundaries, and may be LOS pairs of quasars and stars that are resolved by *Gaia* but not by the SDSS.



**Figure 3.** *Top:* Total magnitude versus magnitude difference of individually resolved components, for *WISE* objects split by *Gaia*. The lines delimit systems with a faintest image with  $G = 20$  (full) or  $G = 21$  (dashed). *Middle:* Spatial distribution of *WISE*–*Gaia* multiplets. Grey (resp. purple, black) points correspond to the SDSS (resp. ATLAS-DR3, pseudo-DES) footprint. In both panels, red bullets indicate known quasar lenses/pairs. *Bottom:* *WISE*–*Gaia* multiplets with Dec.  $> -30$ ; the dashed contours correspond to  $b = \pm 30$  deg.

A similar scenario is shown by matching *WISE*–*Gaia* objects (singlets and multiplets) with the iPTF. The lower panels in Fig. 4 shows their hybrid colours across the three surveys, with  $R$  band from the iPTF when available. Similarly to the pseudo-DES footprint, objects were selected based only on *WISE* colours, so a large



**Figure 4.** Using hybrid colours across three surveys, for patchy ground-based coverage. *Top:* Grey (resp. green) points show  $g_{\text{sdss}} - i_{\text{sdss}}$  (resp.  $0.175 + G - i_{\text{sdss}}$ ) versus  $i_{\text{sdss}} - W_1$  for SDSS queried objects (resp. *Gaia* multiplet counterparts). *Lower panels:* Hybrid colours of *WISE*–*Gaia* multiplets that also fall in the iPTF footprint, with purple (resp. blue) indicating objects with lower (resp. higher)  $R - W_1$ ; red bullets indicate known lenses.

contamination by LOS quasar-star pairs is present, as seen in the spatial distribution of multiplets: many abound at  $|b| < 30$  deg, besides clumps corresponding to M31 and M33. In terms of colours, flux ratios  $G_1 - G_2 = 2.5 \log_{10}(f_1/f_2)$  and separations, different object classes can be recognized: (i) quasar-like objects, with  $R - W1 \gtrsim 2.8$ ,  $R - W2 \gtrsim 3.9$ , themselves divided in three main clusters at  $W1 - W2 = (0.750 \pm 0.125), (0.94 \pm 0.07), (1.225 \pm 0.175)$ ; (ii) quasar-star pairs in which the star dominates in the optical, with low iPTF-WISE colours and  $G - R \approx -0.02 \pm 0.05$ , with object separations distributed smoothly and regardless of colours; (iii) ‘bluer’ objects, with  $G - R \lesssim -0.2$ ,  $|G_1 - G_2| \lesssim 2.0$ , and typical separations  $\gtrsim 5.0$  arcsec.

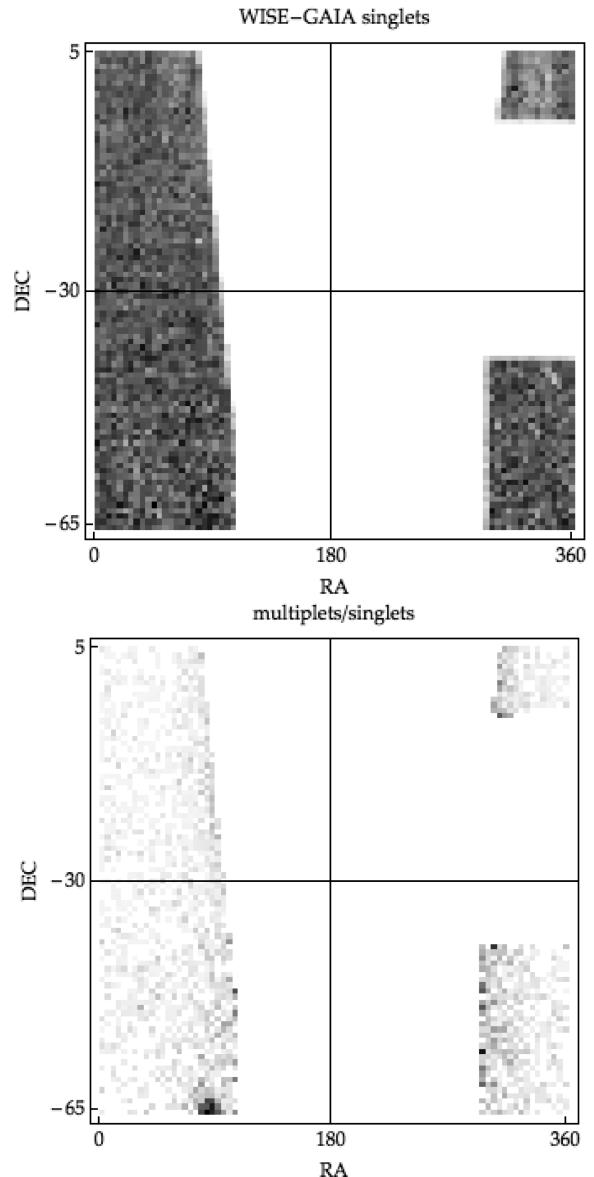
Different object classes are all intersecting, at least in projection, and again the known quasar lenses and pairs seem to occupy the outskirts of different populations. Quasar-star objects with low  $R - W1$  seem just the visible ‘tip of the contaminant iceberg’, since multiplets with  $|b| < 30$  deg (perhaps mostly quasar-star pairs) form a tight sequence with  $G - R \approx -0.02 \pm 0.05$  but extend to higher values of  $R - W1$ , depending on the relative contribution of quasar and star. Since quasars dominate in the infrared, the *WISE*  $W1 - W2$ ,  $W2 - W3$  colours or quasar-star pairs are indistinguishable from those of isolated and lensed quasars.

The analysis of hybrid colours suggests that an integrated approach, incorporating both the multiplet detection and outlier selection, may yield cleaner lens candidate samples. This has not been used for this work yet, and is shown as a possible extension for future investigation. It may be replaced altogether by *Gaia* colours, based on  $G_{BP}$  and  $G_{RP}$ , if these are available in future releases.

### 3.3 Foregrounds: Galactic substructure

Where the stellar density is higher, e.g. at low Galactic latitudes, the *WISE+Gaia* multiplet search triggers more often. While this is a nuisance for quasar lens searches, it also means that it can be used to locate stellar overdensities such as satellite galaxies, globular clusters and streams. Fig. 3, plotting multiplets among non-duplicate *Gaia* detections, shows some of these. Besides the Large Magellanic Cloud and the Orphan Stream, some Southern candidate streams appear. I designate them as WG1, with RA, Dec.  $\approx (64.16, -23.2)$  to  $(97.2, -55.5)$ ; WG2, with RA, Dec.  $\approx (70.5, -51.4)$  to  $(95.7, -47.1)$ ; WG3, with RA, Dec.  $\approx (304.1, -47.7)$  to  $(351.3, -44.8)$  and WG4, with RA, Dec.  $\approx (300.0, -49.4)$  to  $(333.5, -66.0)$ . Of these, two (WG3, WG4) may have counterparts independently identified<sup>5‡</sup> in DES data (Shipp & DES Collaboration, in preparation). The sharpest stream is WG1, WG2 is weaker and barely visible in the DES search,<sup>5</sup> and the thin stream WG3 seems to cross the Galactic plane, whereas WG4 is considerably thicker and may be a superposition of two, almost-parallel streams. Their quoted endpoints should be solely a guidance to the eye: they are not the primary aim of the *WISE-Gaia* multiplet search and so they are roughly traced by the detected quasar+star alignments. Besides WG1, ..., WG4, the thin stream of Balbinot et al. (2016) can be seen crossing the ATLAS and DES footprints.

Other candidate streams can be seen at low grazing angles from the Galactic plane, or in the ATLAS footprint, but are less sharp and may be given by patchy footprint coverage, as traced by the distribution of *WISE-Gaia* quasar-like singlets. An example of spurious overdensity, in the Northern Galactic Hemisphere, is given by two wide regions with low *WISE-Gaia* singlet density separated by a



**Figure 5.** Distribution of (non-duplicate) *WISE-Gaia* singlets and multiplets over the pseudo-DES footprint. *Top:* density of single matches; *bottom:* ratio of multiplets to singlets, showing the same inhomogeneities identified from Fig. 2. The edge of the MW disc is visible in the increased multiplet counts at low  $|b|$ .

thin strip that has more complete coverage, which could otherwise be confused for a stream extending from RA, Dec.  $\approx (160, 15)$  deg to RA, Dec.  $\approx (150, 45)$  deg. The four candidates WG1, ..., WG4 seem more robust than others because they do not seemingly coincide with inhomogeneities in the *WISE-Gaia* coverage: as shown by Fig. 5 in the ratio of multiplets to singlets, the same overdensities corresponding to WG1, ..., WG4 remain once their distribution is normalized to that of singlets.

Given the primary scope of this paper, i.e. lensed quasar searches, the identification of WG1, ..., WG4 may be regarded as an unintended by-product. Searches for Galactic substructure via stellar overdensities have been applied extensively in the past to the SDSS (Belokurov et al. 2006; Zucker et al. 2006; Belokurov et al. 2007; Irwin et al. 2007; Koposov et al. 2007; Belokurov et al. 2009),

<sup>5‡</sup>E. Balbinot, private communication.

ATLAS (Belokurov et al. 2014) and DES (Bechtol et al. 2015; Koposov et al. 2015; Balbinot et al. 2016; Li et al. 2016) footprints, and have also led to the discovery of two star clusters using *Gaia* data alone (Koposov et al. 2017). Northern streams, dwarf galaxies and globular clusters have also been found as overdensities in the PS1 footprint (e.g. Laevens et al. 2015; Bernard et al. 2016, and references therein). The search outlined here differs from those as it targets neighbours within the same 6–10 arcsec radius, as opposed to overdensities on arc-minute scales from binned star counts.

Two steps forward may be envisioned, enabling an all-sky *WISE+Gaia* search of Galactic substructure down to the  $G = 20.7$  detection limit. First, it may be tailored directly on stellar objects ( $W_1 - W_2 \lesssim 0.3$ , instead of  $W_1 - W_2 > 0.55$ ), and may also be augmented with 2MASS (Skrutskie et al. 2006) information, such as  $J - W_1 < 1.2$  (Vickers, Röser & Grebel 2016). The quasar-like colour selection and match to *Gaia* are, in fact, one reason for the circular gaps in the distribution of multiplets over the sky, due to crowding or higher extinction. Second, while WG1, ..., WG4 were identified solely based on visual inspection, a more complete search may automatically select associations of multiplets that lie on the same line over  $\approx 5\text{--}10$  deg-long distances. This would also help identify more diffuse streams from more massive progenitors, which (Balbinot & Gieles 2017) can otherwise be confused with random fluctuations in the distribution of halo stars.

Even fainter magnitudes may be reached by trading the resolution of *Gaia* with deeper *grizY* magnitudes from DES in the South or PS1 in the North. In this case, the *WISE* pre-selection might be dropped altogether, by-passing completeness limits at  $i \gtrsim 21$ ; however, a quasar-like *WISE-Gaia* singlet selection is still useful to characterize the survey coverage, and hence evaluate spurious substructure detections.

#### 4 CONCLUSIONS AND FUTURE PROSPECTS

I have introduced two new methods to find quasar lenses, and illustrated their performance on publicly accessible footprints, using training- and test sets of known lenses and pairs in the CASTLES and SQLS data bases. The first concentrates on objects that are not well described as isolated quasars or blue galaxies; the second joins *WISE* and *Gaia* to detect multiplets of point-sources with overall quasar mid-IR colours. Besides recovering known lenses (with varying degree of success), both searches have led to new discoveries, reported elsewhere. A by-product of the *WISE+Gaia* search has been the identification of Galactic stream candidates that (as I later learned) are being found within the DES collaboration using different techniques.

A common feature of many searches, including the outlier-selection and multiplet-detection, is the trade-off between completeness and manageable size of candidate samples to be visually inspected and followed-up at the telescope. Searches concentrating on ‘lens’ colours (e.g. this outlier-selection and the population-mixture classification by Williams et al. 2017a) discard a high fraction of lenses with colours akin to nearby unlensed quasars. This is a price to pay in order to find systems with sources at higher redshift, or that could not be pre-selected for SDSS fibre spectroscopy due to their unusual colours. On the other hand, searches tailored on ‘quasar’ colours, amending the loss of lenses with quasar-like colours, result in large candidate samples to be examined: according to Ostrovski et al. (2017), a double was found after eyeballing

5000 objects brighter than  $i = 19$  in a  $1500 \text{ deg}^2$  footprint. By comparison, the outlier selection resulted in 5064 SDSS targets brighter than  $i = 20$ .

The pseudo-distance cuts reduce the number of objects considerably, from  $5 \times 10^5$  to  $\approx 8600$  targets, of which  $\approx 60$  per cent non-repeated objects, down to 239 final candidates, to the price of a harsh selection on known lenses. Even though few systems are rejected by each pair-wise combination of pseudo-distances, and most of the rejected lenses lie close to the cut boundaries, all cuts combined reduce the overall completeness to 30 per cent, or 50 per cent when only cuts in  $d_1, \dots, d_4$  are used. Since the number of targets for visual inspection is manageable, one could loosen the cuts in  $d_5, \dots, d_{10}$ ; this would mean that the whole test set would now be a training-set for the method. Despite the 50 per cent completeness with respect to already-known lenses, previously unknown lenses have been discovered with this technique (e.g. the new quad J1433+60; Agnello et al. 2017b), which suggests that outlier-selection is indeed complementary to previous searches.

Comparable numbers, from query to final candidates, are found when applying the *WISE-Gaia* search to the SDSS footprint. The resolving power of *Gaia* enables multiplet detection for objects that are otherwise blended by the pipelines of ground-based surveys. New lenses, not found by other techniques, have been discovered with this method and will be presented elsewhere. Its performance, however, is not obvious to characterize: while it recognizes the two quasar images of the close double DES J0115 (Agnello et al. 2015b), and all quasar images of the fold PG1115 (Weymann et al. 1980), it detects only two images in the DES J0408 quad (Lin et al. 2017), and two the WFI2033 quad (Morgan et al. 2004), whereas the WFI2026 quad is seemingly unresolved. Due to  $G = 20.7$  depth and pipeline-specific choices in resolving close pairs with high flux-ratios, roughly one every three known lenses/pairs is resolved by *Gaia*.

Hybrid colours may be used as additional information, either using (limited) ground-based information or via  $G_{\text{RP}}$ ,  $G_{\text{BP}}$  photometry, if available in future releases. This will be particularly interesting in view of the EUCLID mission, which should provide a wide-field, deep and sharp counterpart to the slitless spectroscopic Hamburg-ESO survey (Wisotzki et al. 2000), from which some remarkable quasar lenses have been identified.

Classification via clustering and outlier-selection has been implemented with different methods, for other purposes, over the last 2 yr. Clustering in colour-space has been used to evaluate photometric redshifts (Rahman et al. 2016b; Rahman, Ménard & Scranton 2016a), select LRGs (Rozo et al. 2016) and identify galaxy clusters from improved red-sequence finders (Rykoff et al. 2014). Outlier-selection, implemented via random-forest classification on SDSS fibre spectra of galaxies, has been used to select peculiar galaxies of various types (Baron & Poznanski 2017). While the search in this paper was tailored on (lensed) quasars, and others were concentrated on galaxies, there is significant scope for integrating these into a comprehensive photometric classification across wide-field surveys and selection of the most peculiar objects.

A by-product of the *WISE-Gaia* multiplet search is the detection of (candidate) Milky Way streams, as coherent structures traced by stars in front of *WISE* quasar candidates. Multi-band follow-up is needed, however, to confirm them and characterize their stellar populations. A *WISE-Gaia* search can be tailored on streams, using a different *WISE* selection and possibly augmenting with near-IR colours to separate extragalactic objects.

## ACKNOWLEDGEMENTS

I wish to thank Eduardo Balbinot, Nora Shipp, Alex Drlica-Wagner for the comparison between my MW stream candidates and theirs; and Elizabeth Buckley-Geer, Huan Lin, Paul Schechter and Tommaso Treu for discussions on quasar lens searches. Careful comments by the anonymous referee have led to clarifications and improvements of this paper. I gratefully acknowledge hospitality by the ITC-Harvard, where this work was completed.

This research has made use of the NASA/ IPAC Infrared Science Archive, which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

This work has made use of data from the European Space Agency (ESA) mission *Gaia* (<https://www.cosmos.esa.int/gaia>), processed by the Gaia Data Processing and Analysis Consortium (DPAC, <https://www.cosmos.esa.int/web/gaia/dpac/consortium>). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the Gaia Multilateral Agreement.

## REFERENCES

- Abazajian K. N. et al., 2009, *ApJS*, 182, 543  
 Agnello A., Kelly B. C., Treu T., Marshall P. J., 2015a, *MNRAS*, 448, 1446  
 Agnello A. et al., 2015b, *MNRAS*, 454, 1260  
 Agnello A. et al., 2017a, *MNRAS*, preprint ([arXiv:1702.00406](https://arxiv.org/abs/1702.00406))  
 Agnello A., Grillo C., Jones T., Treu T., Bonamigo M., Suyu S. H., 2017b, *MNRAS*, preprint ([arXiv:1702.03942](https://arxiv.org/abs/1702.03942))  
 Assef R. J. et al., 2013, *ApJ*, 772, 26  
 Balbinot E., Gieles M., 2017, *MNRAS*, preprint ([arXiv:1702.02543](https://arxiv.org/abs/1702.02543))  
 Balbinot E. et al., 2016, *ApJ*, 820, 58  
 Baron D., Poznanski D., 2017, *MNRAS*, 465, 4530  
 Bate N. F., Floyd D. J. E., Webster R. L., Wyithe J. S. B., 2011, *ApJ*, 731, 71  
 Bechtol K. et al., 2015, *ApJ*, 807, 50  
 Belokurov V. et al., 2006, *ApJ*, 647, L111  
 Belokurov V. et al., 2007, *ApJ*, 654, 897  
 Belokurov V. et al., 2009, *MNRAS*, 397, 1748  
 Belokurov V., Irwin M. J., Koposov S. E., Evans N. W., Gonzalez-Solares E., Metcalfe N., Shanks T., 2014, *MNRAS*, 441, 2124  
 Berghea C. T., Nelson G. J., Rusu C. E., Keeton C. R., Dudik R. P., 2017, *ApJ*, preprint ([arXiv:1705.08359](https://arxiv.org/abs/1705.08359))  
 Bernard E. J. et al., 2016, *MNRAS*, 463, 1759  
 Bernstein G. M., Jarvis M., 2002, *AJ*, 123, 583  
 Bovy J. et al., 2011, *ApJ*, 729, 141  
 Chambers K. C. et al., 2016, preprint ([arXiv:1612.05560](https://arxiv.org/abs/1612.05560))  
 Chan J. H. H., Suyu S. H., Chiueh T., More A., Marshall P. J., Coupon J., Oguri M., Price P., 2015, *ApJ*, 807, 138  
 Chehade B. et al., 2016, *MNRAS*, 459, 1179  
 Courbin F., Saha P., Schechter P. L., 2002, in Courbin F., Minniti D., eds, *Gravitational Lensing: An Astrophysical Tool*, Lecture Notes in Physics, Vol. 608. Springer-Verlag, Berlin, p. 1  
 Ding X., 2017, *MNRAS*, 465, 4634  
 DiPompeo M. A., Bovy J., Myers A. D., Lang D., 2015, *MNRAS*, 452, 3124  
 Gaia Collaboration, 2016, *AA*, 595, A1  
 Hutsemékers D., Sluse D., Braibant L., Anguita T., 2015, *AA*, 584, A61  
 Inada N. et al., 2012, *AJ*, 143, 119  
 Irwin M. J. et al., 2007, *ApJ*, 656, L13  
 Jackson N., Tagore A. S., Roberts C., Sluse D., Stacey H., Vives-Arias H., Wucknitz O., Volino F., 2015, *MNRAS*, 454, 287  
 Kaiser N. et al., 2010, in Stepp L. M., Gilmozzi R., Hall H. J., eds, *Proc. SPIE Vol. 7733, Ground-based and Airborne Telescopes III*. SPIE, Bellingham, p. 77330E  
 King L. J., Browne I. W. A., Marlow D. R., Patnaik A. R., Wilkinson P. N., 1999, *MNRAS*, 307, 225  
 Koposov S. et al., 2007, *ApJ*, 669, 337  
 Koposov S. E., Belokurov V., Torrealba G., Evans N. W., 2015, *ApJ*, 805, 130  
 Koposov S. E., Belokurov V., Torrealba G., 2017, *MNRAS*, 470, 2702  
 Laeens B. P. M. et al., 2015, *ApJ*, 813, 44  
 Law N. M. et al., 2009, *PASP*, 121, 1395  
 Li T. S. et al., 2016, *ApJ*, 817, 135  
 Lin H. et al., 2017, *ApJ*, 838, L15  
 Lindegren L. et al., 2016, *AA*, 595, A4  
 More A. et al., 2016, *MNRAS*, 456, 1595  
 Morgan N. D., Caldwell J. A. R., Schechter P. L., Dressler A., Egami E., Rix H.-W., 2004, *AJ*, 127, 2617  
 Myers S. T. et al., 2003, *MNRAS*, 341, 1  
 Oguri M., Marshall P. J., 2010, *MNRAS*, 405, 2579  
 Oguri M. et al., 2006, *AJ*, 132, 999  
 Oguri M., Rusu C. E., Falco E. E., 2014, *MNRAS*, 439, 2494  
 Ostrovski F. et al., 2017, *MNRAS*, 465, 4325  
 Rahman M., Ménard B., Scranton R., 2016a, *MNRAS*, 457, 3912  
 Rahman M., Mendez A. J., Ménard B., Scranton R., Schmidt S. J., Morrison C. B., Budavári T., 2016b, *MNRAS*, 460, 163  
 Refsdal S., 1964, *MNRAS*, 128, 307  
 Rozo E. et al., 2016, *MNRAS*, 461, 1431  
 Rykoff E. S. et al., 2014, *ApJ*, 785, 104  
 Sánchez E., DES Collaboration, 2010, *J. Phys. Conf. Ser.*, 259, 012080  
 Schechter P. L., Poole D., Blackburne J. A., Wambsganss J., 2014, *ApJ*, 793, 96  
 Schmidt K. B., Rix H.-W., Shields J. C., Knecht M., Hogg D. W., Maoz D., Bovy J., 2012, *ApJ*, 744, 147  
 Shanks T. et al., 2015, *MNRAS*, 451, 4238  
 Skrutskie M. F. et al., 2006, *AJ*, 131, 1163  
 Sluse D., Surdej J., Claeekens J.-F., Hutsemékers D., Jean C., Courbin F., Nakos T., Billeres M., Khmil S. V., 2003, *AA*, 406, L43  
 Sluse D., Hutsemékers D., Courbin F., Meylan G., Wambsganss J., 2012, *AA*, 544, A62  
 Suyu S. H. et al., 2017, *MNRAS*, 468, 2590  
 Tie S. S. et al., 2017, *AJ*, 153, 107  
 van Leeuwen F. et al., 2017, *AA*, 599, A32  
 Vickers J. J., Röser S., Grebel E. K., 2016, *AJ*, 151, 99  
 Weymann R. J. et al., 1980, *Nature*, 285, 641  
 Williams P., Agnello A., Treu T., 2017a, *MNRAS*, 466, 3088  
 Williams P. R. et al., 2017b, *MNRAS*, preprint ([arXiv:1706.01506](https://arxiv.org/abs/1706.01506))  
 Wisotzki L., Christlieb N., Bade N., Beckmann V., Köhler T., Vanelle C., Reimers D., 2000, *AA*, 358, 77  
 Wisotzki L., Schechter P. L., Bradt H. V., Heinmüller J., Reimers D., 2002, *AA*, 395, 17  
 Wright E. L. et al., 2010, *AJ*, 140, 1868  
 Zucker D. B. et al., 2006, *ApJ*, 643, L103

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