

# Identifying quasars with astrometric and mid-infrared methods from APOP and ALLWISE★

Sufen Guo (郭素芬)<sup>1,2</sup>, Zhaoxiang Qi (齐朝祥)<sup>1,2</sup>, Shilong Liao (廖石龙)<sup>1,2</sup>, Zihuang Cao (曹子皇)<sup>2,4</sup>, Mario G. Lattanzi<sup>3</sup>, Beatrice Bucciarelli<sup>3</sup>, Zhenghong Tang (唐正宏)<sup>1,2</sup>, and Qing-Zeng Yan (闫庆增)<sup>1,2</sup>

<sup>1</sup> Shanghai Astronomical Observatory, Chinese Academy of Sciences, 80 Nandan Rd, 200030 Shanghai, PR China  
e-mail: kevin@shao.ac.cn

<sup>2</sup> School of Astronomy and Space Science, University of Chinese Academy of Sciences, Beijing 100049, PR China

<sup>3</sup> INAF, Astrophysical Observatory of Torino, Via Osservatorio 20, 10025 Pino Torinese (Torino), Italy

<sup>4</sup> National Astronomical Observatory, Chinese Academy of Sciences, PR China

Received 29 March 2018 / Accepted 10 June 2018

## ABSTRACT

**Context.** Quasars are spatially stationary, and they are essential objects in astrometry when defining reference frames. However, the census of quasars is far from complete. Mid-infrared colors can be used to find quasar candidates because active galactic nuclei (AGNs) show a peculiar appearance in mid-infrared color, but these methods are incapable of separating quasars from AGNs.

**Aims.** The aim of our study is to use astrometric and mid-infrared methods to select quasars and get a reliable quasar candidates catalog.

**Methods.** We used a near-zero proper motion criterion in conjunction with WISE [W1–W2] color to select quasar candidates. The [W1–W2] color criterion is defined by the linear boundary of two samples: LAMOST DR5 quasars, which serve as the quasar sample, and LAMOST DR5 stars and galaxies, which serve as the non-quasar sample. The contamination and completeness are evaluated.

**Results.** We present a catalog of 662 753 quasar candidates, with a completeness of about 75% and a reliability of 77.2%.

**Key words.** catalogs – quasars: general – proper motions – galaxies: active – infrared: galaxies

## 1. Introduction

Quasars, one type of active galactic nuclei (AGNs), are point-like and generally at high redshifts. They are vital objects to construct the International Celestial Reference Frame (ICRF), which is the realization of the International Celestial Reference System (ICRS). This is based on the fact that the proper motions of quasars caused by secular aberration drift (Titov 2010) is non-detectable now, and therefore they are regarded as zero, which can keep the reference direction fixed. Consequently, quasars are ideal objects with which to construct a non-rotating reference frame. The International Astronomical Union (IAU) stipulated the precise coordinates of extragalactic radio sources as the ICRF (Arias et al. 1995). The first (ICRF1, Ma et al. 1998) and the second realization (ICRF2, Fey et al. 2015) of ICRF are defined by the positions of compact radio sources (mostly quasars) obtained with VLBI. With the improvement of precisions and the spatial and frequency coverages at radio band, ICRF3 (Jacobs et al. 2014; Malkin et al. 2015) will be released in 2018, containing a larger number of radio-loud quasars.

Radio-loud quasars that have optical counterparts can be used to tie ICRF to optical frames (Bourda et al. 2011). However, this alignment requires extensive overlapping sources, which possess highly accurate radio as well as optical positions and a uniform sky coverage. European Space Agency's

(ESA) Gaia mission (Lindegren et al. 2016; Mignard et al. 2016) is realizing a new optical reference frame. This mission is dedicated to measure three-dimensional positions and velocities with unprecedented precision for more than 1 billion objects within five years. Observations will contain ~500 000 quasars, 20 000 of which are brighter than visual magnitude 18. The first (Gaia DR1, Lindegren et al. 2016) and second (Gaia DR2, Gaia Collaboration 2018) datasets have been released, and the final data release is expected to be published by 2020. Unlike the preceding HIPPARCOS mission (Perryman et al. 1997), Gaia realizes the extragalactic celestial reference frame directly at optical wavelengths, using the highest accurate positions of quasars (Lindegren et al. 2016, 2018; Mignard et al. 2016). It is crucial to guarantee the consistency between this optical reference frame and the radio reference frame. To achieve this goal, a large number of quasars are needed to produce a uniformly distributed sample and to place severe constraints on the overall geometry.

Many surveys have been performed to complete the census of quasars, and the number of quasars now is approximately 700 000. The main surveys include the Sloan Digital Sky Survey (SDSS, York et al. 2000), the catalog from the large Sky Area Multi-Object Fiber Spectroscopic Telescope (hereafter LAMOST, Cui et al. 2012), the Large Bright Quasar Survey (Hewett et al. 1995), the Hamburg Quasar Survey (Hagen et al. 1995), the Canada-France High-z Quasar Survey (Willott et al. 2010), the INT Wide Angle Survey (Sharp et al. 2001), and the 2dF Quasar Survey (Boyle et al. 2000). The SDSS data release series (Pâris et al. 2012, 2014, 2017, 2018; Ross et al. 2012; Myers et al. 2015) contribute much of the effort.

\* The catalog of AGN candidates is only available in electronic form at the CDS via anonymous ftp to [cdsarc.u-strasbg.fr](ftp://cdsarc.u-strasbg.fr) (130.79.128.5) or via <http://cdsarc.u-strasbg.fr/viz-bin/qcat?J/A+A/618/A144>

Although a large number of quasars have been confirmed by their spectra, many quasars have not yet been identified. Recently, several selection methods for AGN candidates using photometric observations have been developed, and among them, the mid-infrared color criteria, which are based on the fact that AGNs occupy a specific locus in the mid-infrared color space (Lacy et al. 2004; Richards et al. 2006), are proved to be very efficient. These methods can efficiently select AGNs, but not quasars.

Mid-infrared selection methods are relatively insensitive to extinction caused by interstellar dust which plagues AGN surveys toward shorter wavelengths. Notably, with the release of the mid-infrared photometric data of all-sky Wide-field Infrared Survey Explorer (WISE, Wright et al. 2010), some selection criteria using WISE colors have been developed. For instance, Stern et al. (2012) using  $[W1-W2]$  color identified 78% AGNs from AGN candidates with 95% reliability; Mateos et al. (2012) defined a criterion with  $[W1-W2]$  and  $[W2-W3]$  colors. However, the reason why AGNs are located in particular locus in mid-infrared color space is not yet fully understood. Stern et al. (2005, 2012) assumed that in the mid-infrared band, the spectral energy distribution (SED) of AGNs are dominated by a power law, which is different from blackbody spectrum objects like stars and normal galaxies. Alternatively, Nikutta et al. (2014) claims that the difference in mid-infrared results from the peculiar characteristics of the dust shells surrounding different type of objects.

With mid-infrared methods, Secrest et al. (2015) selected 1.4 million AGN candidates (hereafter, MIRAGN). This large sample of AGNs together with a tiny fraction of VLBI sources was used to define the reference frame of Gaia DR2 (Mignard et al. 2018). Due to the deeper WISE coverage, the objects' density in MIRAGN near the ecliptic poles is much larger than that in other regions (see detail in Fig. 1 of Secrest et al. 2015). Furthermore, in MIRAGN, about 70 000 candidates have proper motions larger than  $10 \text{ mas yr}^{-1}$  either in right ascension or in declination, suggesting that they are probably stars or extended galaxies. This may lead to the instability of Gaia reference frame, and it is therefore necessary to test this reference frame with a uniformly distributed quasar sample.

Non-detectable proper motion is a basic characteristic of quasars and a constraint on this can purify quasar candidates selected in mid-infrared color maps. In fact, because most proper motion catalogs are derived from optical observations, the astrometric criterion can eliminate most dwarfs that are ultracool but have high mid-infrared colors (see Fig. 1 in Kirkpatrick et al. 2011). However, the proper motion has not been widely used as a selection criterion because of lacking precise proper motions. The catalog of the Absolute Proper motions Outside the galactic Plane (APOP, Qi et al. 2015), which contains precise proper motions, makes it feasible to identify quasars in such a way. APOP covers  $\sim 12\,000$  square degrees of the southern sky and the limiting magnitude  $R_F$  is 20.8, so it includes many dark sources beyond the regime of SDSS.

In order to derive a reliable sample of quasars, we present a quasar selection method by combining mid-infrared criterion with a near-zero proper motion criterion. This method is the first attempt to use both astrometric and mid-infrared data, and can be applied to subsequent catalogs that contain precise proper motions.

This paper is organized as follows. We describe the data used in Sect. 2 and the detailed quasar selection criteria in Sect. 3. In Sect. 4 we present our quasar candidate catalog. We describe the catalog in Sect. 5. Finally, in Sect. 6 we summarize our results.

## 2. Data

### 2.1. The APOP catalog

APOP is derived from historical digitized Schmidt survey plates, which covers 22 525 square degrees of the sky where  $|b| \geq 27$  degrees. Using stars and galaxies, APOP removed the systematic errors in proper motions related to positions, magnitudes and colors. Additional  $R_F$ ,  $B_J$ ,  $I_V$  and  $V$  magnitudes extracted from GSC 2.3 (Lasker et al. 2008) and JHK magnitudes from 2MASS (Skrutskie et al. 2006) are listed. The reference frame is aligned with the ICRF2 within  $\pm 0.2$  mas at J2000.0. More than a hundred million sources are cataloged and the magnitude reaches  $R_F \sim 20.8$ . The internal accuracy of proper motion is better than  $4.0 \text{ mas yr}^{-1}$  for stellar objects brighter than  $R_F = 18.5$ , and  $9.0 \text{ mas yr}^{-1}$  for stellar objects with magnitudes  $18.5 < R_F < 20.0$ . The nominal global zero point error estimated by quasars is better than  $0.6 \text{ mas yr}^{-1}$ .

### 2.2. The ALLWISE catalog

WISE is a satellite with a 40 cm aperture launched by NASA in 2009 December. It scanned the whole sky at four mid-infrared bands, centering at 3.4, 4.6, 12 and  $22 \mu\text{m}$  (hereafter referred to as W1, W2, W3, and W4, respectively) with image resolutions (FWHM) of 6.1, 6.4, 6.5, and  $12.0 \text{ arcsec}$ . The achieved sensitivities in the preliminary data release are about 0.008, 0.11, 1 and 6 mJy ( $5\sigma$ ) for the four bands, respectively. Began in 2010 January, WISE finished the first all-sky cryogenic survey in August 2010. After that, an additional 30% of the sky was mapped. The final post-cryogenic survey (NEOWISE, Mainzer et al. 2011a,b) started on September 2010 and finished on February 2011, covering 70% of the sky.

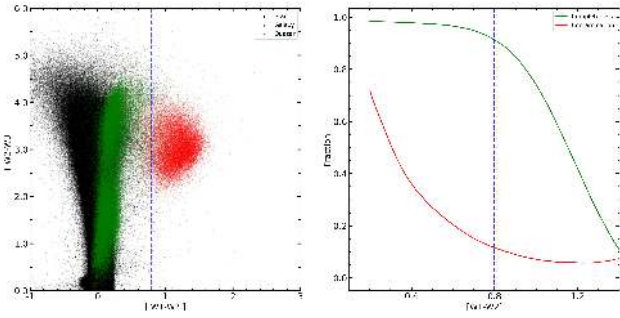
The ALLWISE catalog (Cutri et al. 2013) is a compilation of the data from the cryogenic and post-cryogenic survey phases of the WISE mission. It contains more than seven hundred million celestial objects with improved photometric and astrometric accuracy compared with the WISE All-Sky Release Catalog (Cutri et al. 2012). ALLWISE catalog lists positions, magnitudes, astrometric and photometric uncertainties, flags, and counterparts in the 2MASS catalog.

### 2.3. The LAMOST catalog

LAMOST is a reflecting Schmidt telescope located at Xinglong Observing station in China. With a wide field of view (FOV) of  $5^\circ$  that can allocate 4000 optical fibers, it can obtain 4000 spectra simultaneously. A large aperture with diameter of about 4 m is designed to observe deep objects. It is dedicated to observe the spectrum of celestial objects over the northern sky with resolution  $R = \lambda/\Delta\lambda = 1800$ . The limiting magnitude in the  $i$  band is 20.5. LAMOST began its first spectroscopic survey in September 2012, and it has released the fifth data release (LAMOST DR 5) with more than 1.3 million spectra, containing 1.2 million stars, 36093 galaxies and 14,782 quasars. Before DR5, LAMOST used a method similar to SDSS to identify quasars (Luo et al. 2015), while for DR5, quasars were confirmed by human eyes, making identifications more reliable.

## 3. Quasar selection

We used mid-infrared colors in conjunction with a near-zero proper motion criterion to select quasars. In this section, we discuss in detail our selection criteria of quasar candidates.

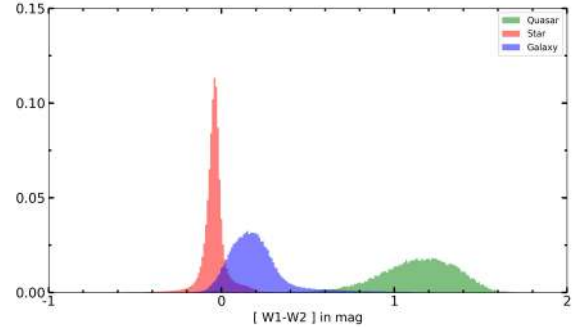


**Fig. 1.** *Left panel:* quasar locus in  $[W1-W2]$  and  $[W2-W3]$  space. *Right panel:* quasar completeness (blue curve) and the contamination (red curve) by stars and galaxies as a function of a  $[W1-W2]$  color cut. The dotted lines in the two panels cut at  $[W1-W2] = 0.8$ .

### 3.1. Mid-infrared color-selection criterion

With the rise of mid-infrared surveys, researchers discovered that AGNs primarily reside in a restricted locus in mid-infrared color space, which allows them to separate AGNs from stars and normal galaxies in mid-infrared color maps. For the first time, Lacy et al. (2004) and Stern et al. (2005) investigated AGNs with mid-infrared color using the Spitzer Space Telescope First Look Survey, and since then, the development of mid-infrared method has profited significantly from the successful WISE mission. After the WISE mission was completed, researchers (Stern et al. 2012; Secrest et al. 2015; Assef et al. 2011) found that WISE colors are efficient to select AGNs. Mainly there are two methods that rely *only* on colors from WISE imaging: one uses the  $[W1-W2]$  color only (Stern et al. 2012); the other was developed by Mateos et al. (2012): using two WISE colors ( $[W1-W2]$  and  $[W2-W3]$ ) to define the boundaries of the region where AGNs are located. All these criteria generally agree with each other. The majority of the spectroscopically confirmed or X-ray confirmed AGNs were used to test the reliability of these methods.

Building upon these valuable works, we carried out an independent investigation of mid-infrared selection criteria using new data. Because the newly released LAMOST DR5 has a highly reliable sample of quasars, we used it to define our selection criteria. 8 169 852 common sources are found in ALLWISE and LAMOST DR5, including quasars, galaxies and stars. We used SVMs (support vector machines) to determine the color boundaries. The SVM algorithm is a widely used technique capable of classifying objects into different groups in a supervised machine-learning environment; in particular, linear SVM classifiers are suitable in this context, as previous works (Stern et al. 2012; Mateos et al. 2012) showed that AGNs are linearly separable from other objects. The SVM method is performed with the *Python* package *sklearn*, in which the penalty value  $C$  is a tunable parameter: a larger  $C$  means higher penalties for misclassification whereas a smaller  $C$  allows more misclassification. Because we want our quasar sample to be reliable, we chose an intermediate value of  $C = 10$ . We investigated two-color selection criteria, i.e.,  $[W1-W2]$  and  $[W2-W3]$ , and single-color selection criteria, i.e.,  $[W1-W2]$ . We found that single-color criterion work better on classification. As illustrated in the left panel of Fig. 1, quasars can be efficiently isolated from stars and galaxies by a simple  $[W1-W2]$  cut. The right panel of Fig. 1 shows that the contamination declines at the cost of losing completeness, and  $[W1-W2] \geq 0.8$  is a balance between low contamination and high completeness. The histogram of quasars, stars and



**Fig. 2.** Normalized histogram of  $[W1-W2]$  color of quasars (green), stars (red) and galaxies (blue).

galaxies in Fig. 2 clearly shows this separation. Accordingly, we defined the mid-infrared criterion as:  $[W1-W2] \geq 0.8$ , in agreement with Stern et al. (2012). With this criterion, we find that the completeness of LAMOST quasars is 91.4% and the contamination by sources that are not quasars is 11.1%. Additionally, in order to obtain a reliable sample of mid-infrared photometry, we rejected sources from ALLWISE having signal-to-noise ratios (S/N) lower than five in the W1 or W2 bands.

### 3.2. Proper motion criteria

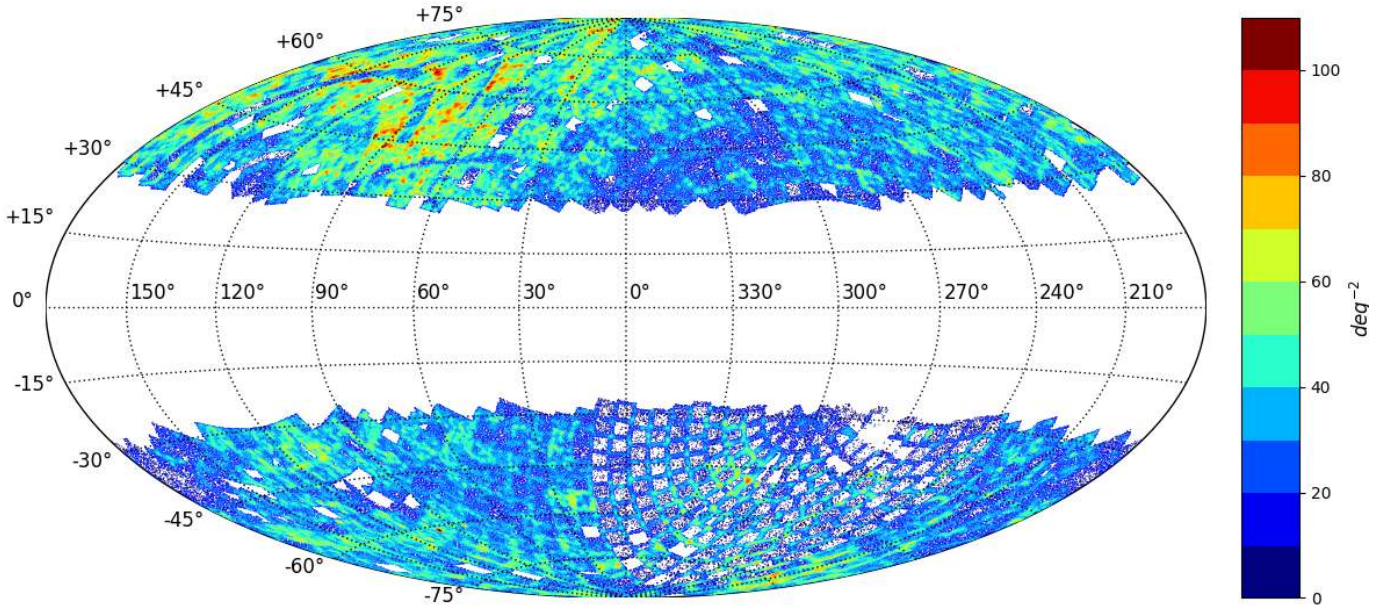
Due to their distant extragalactic distances, the proper motion of quasars can be considered to be zero. On the other hand, ground-based proper motion catalogs suffer from low-precision data and/or systematic error sources and, consequently, are affected by zero-point error. Therefore, if we use the APOP proper motions to build a criterion for quasar selection, we need to consider the zero-point error of APOP first, and then we can define a suitable proper motion criterion.

The APOP catalog estimated the global zero-point using the quasars confirmed by the *Gaia* Initial QSO Catalog (GIQC), which contains about 1.2 million quasar candidates from LQAC2 (Souhay et al. 2012), SDSS DR7 (Schneider et al. 2010), 2dF, BOSS, VLBI observation and data from Richards et al. (2009). GIQC potentially included some spurious sources, resulting in an inaccurate zero-point.

To calculate a global zero-point purely based on confirmed quasars, we cross-matched APOP with LAMOST DR5, obtaining 42 140 quasars. We rejected the sources with proper motion errors larger than  $10 \text{ mas yr}^{-1}$  either in right ascension or declination. With 2% outlier rejection, we obtained mean proper motions of  $0.0 \pm 4.6$  and  $0.5 \pm 4.4 \text{ mas yr}^{-1}$  in right ascension and declination, respectively.

The newly released third Large Quasars Astrometric Catalog (hereafter LQAC3, Souhay et al. 2015) includes most of the confirmed quasars. We cross-matched APOP with LQAC3 and estimated the zero-point as a comparison. There are 185 197 sources in common. With the same procedure, we got mean proper motions of  $0.3 \pm 4.4$  and  $0.6 \pm 4.0 \text{ mas yr}^{-1}$  in right ascension and declination, respectively.

We adopted the zero-point error derived from LAMOST DR5. Although LQAC3 shows a similar result, quasars in LAMOST DR5 are more reliable, because they have been confirmed by visual inspection. Finally, in order to take into account the astrometric signature of quasars, we removed objects whose proper motions in right ascension or declination deviate from the mean APOP zero-point error by more than  $3\sigma$ ; this is not a strict criterion but one deemed necessary to exclude most of



**Fig. 3.** Sky density plot for the selected quasar candidates of this work, Aitoff projection in galactic coordinates. The sources are outside the Galactic plane by virtue of the quasar proper motion criteria which used APOP catalog.

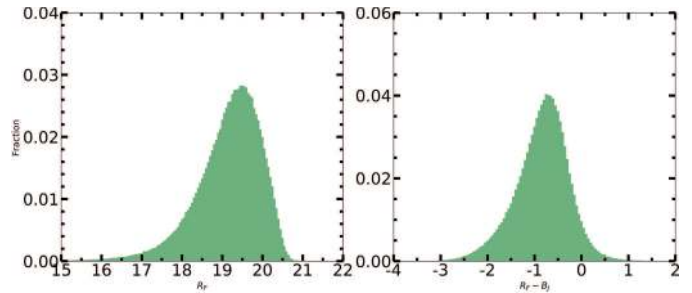
the contamination of stars inside the Milky Way and extended AGNs. Furthermore, sources with APOP proper motion errors in right ascension or declination larger than  $10 \text{ mas yr}^{-1}$  were excluded.

#### 4. Results

We obtain a catalog of 662 753 quasar candidates outside the Galactic plane, and 488 130 are new quasars that were not identified either by LAMOST, SDSS, or X-ray observations. The mean source density is about  $30 \text{ deg}^{-2}$ , covering 22 525 square degrees. Figure 3 shows the density plot of our candidates. Except for the region near the ecliptic poles, it is relatively uniformly distributed outside the Galactic plane. The reason causing the non-uniformity near the ecliptic poles is the deeper WISE coverage toward the vicinity of this region. This problem is also present in MIRAGN, but our results have been improved because we imposed extra restrictions on proper motions. In Table 2 we detail the content of our catalog. The histogram of  $R_F$  magnitudes is shown in the left panel of Fig. 4 and the right panel of Fig. 4 is the histogram of  $R_F - B_J$  color. We excluded the candidates that lack  $R_F$  or  $B_J$  magnitude values, and in total, about 66 000 candidates remain in this plot.

##### 4.1. Completeness

We used the compiled quasar catalog LQAC3 (Souhay et al. 2015) to estimate the completeness of our sample. In total, 185 197 common sources are found in all three catalogs: LQAC3, APOP and ALLWISE. Of these, 148,138 sources have both clear ALLWISE detections ( $S/N$  in W1 and W2  $> 5$ ) and well-qualified proper motions (error  $< 10 \text{ mas yr}^{-1}$ ). With this robust sample, we calculated the completeness of the sources going back to the raw sources of LQAC3 to check how many quasars are found. The result is shown in Table 1. Because we added a proper motion criterion to the mid-infrared method, as expected, the completeness is lower than the catalogs which use mid-infrared selection method purely. All the completeness is

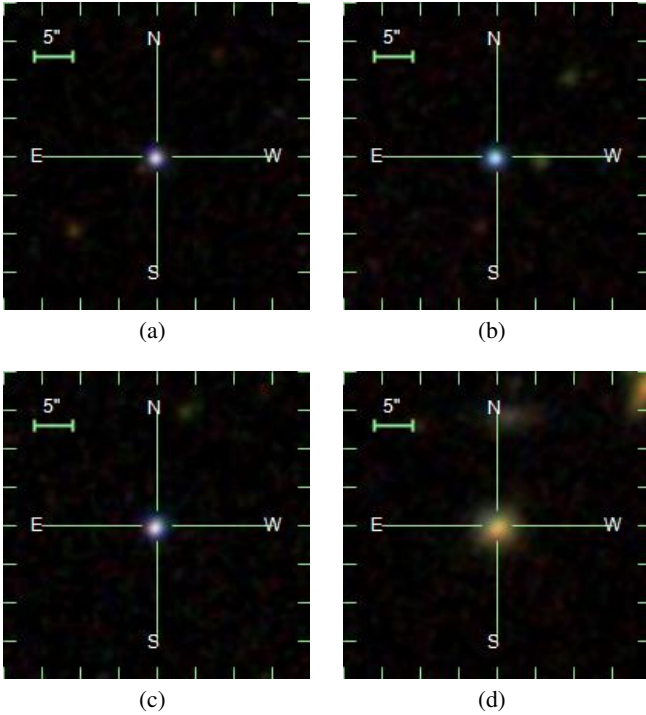


**Fig. 4.** Normalized histograms of  $R_F$  magnitudes and  $R_F - B_J$  color. The left panel is the histogram of the  $R_F$  magnitudes of sources in our catalog, which shows a peak at about  $R_F = 19.5$ . The right panel shows the  $R_F - B_J$  color.

**Table 1.** Completeness compared with LQAC3.

Resource	Completeness	Quasars found/All quasars
ICRF2	84.6%	1057/1250
VLBA	70.8%	1731/2446
VLA	87.0%	613/705
JVAS	73.3%	631/861
FIRST	92.9%	769/828
2QZ	79.5%	12279/15439
SDSS DR7-DR10	81.0%	96014/118521
2df-SDSS LRG	69.5%	1477/2125
GSC2.3	84.0%	86312/102806
B1.0	84.2%	86280/102519
2MASS	77.3%	14071/18207
HB	84.9%	4043/4762
V&V	32.2%	3765/11701

around 75% but for the result from Véron-Cetty & Véron (2010; hereafter V&V) catalog. Among the 11 701 common sources from APOP and V&V, 7936 sources in V&V catalog are absent in our final catalog. We found that the  $[W1 - W2]$  value of most of the sources (7648) in V&V catalog is smaller than 0.8



**Fig. 5.** Four sources taken from our catalog matching with SDSS DR14: *panels a–c*: point-like sources and *panel d*: an extended source.

(rejected by our mid-infrared selection criterion), the rest 288 sources have proper motion errors larger than  $10 \text{ mas yr}^{-1}$  (rejected by our proper motion criterion).

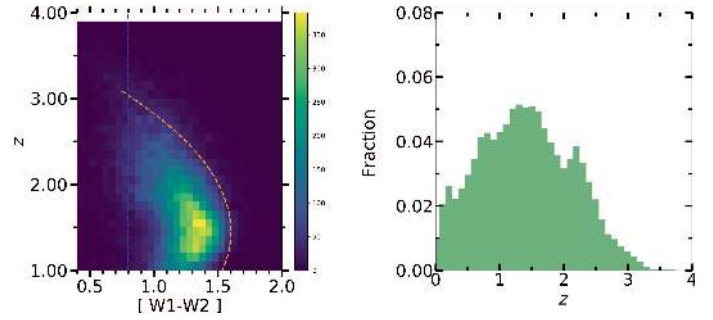
#### 4.2. Contamination

To evaluate the contamination by stars and galaxies, we randomly chose a  $30^\circ \times 30^\circ$  testing region. Within this region, we compiled all sources with confirmed classification (star, galaxy, quasars) from SDSS DR14, LAMOST DR5 and LQAC3 and take it as a reference catalog. Such catalog lists 8043 sources in common with our quasar candidate catalog, of which 7778 are confirmed quasars (96.7%), 69 are confirmed stars (0.9%) and 196 are confirmed galaxies (2.4%). This suggests that only  $\sim 3\%$  of our sources are not AGN-like. We note, however, that a reasonable fraction of the confirmed “quasars” in our sample are low-luminosity sources at low redshift. Such sources may have a visible host galaxy, and so be extended in imaging, limiting their utility for astrometric calibration.

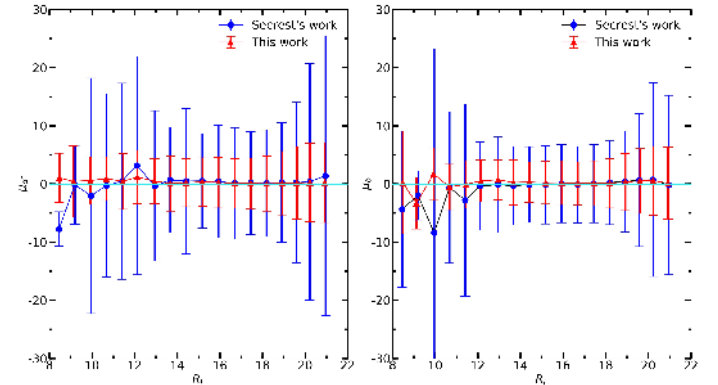
Our purpose is to identify as many quasar candidates as possible, which means the main contents in our final catalog should be point-like sources. We investigated the photometric nature of our catalog by matching it to the SDSS imaging database to evaluate the extended sources contamination. We randomly chose 1000 sources and inspected their morphology using the SDSS tools. 632 sources are found in the SDSS imaging database, of which 488 are point-like (77.2%), 144 are extended (22.8%). Fig. 5 shows the SDSS thumbnails<sup>1</sup> of four candidates in our catalog.

The contamination by other type of AGNs is difficult to remove because of the inevitable proper motions errors in APOP.

<sup>1</sup> <http://skyserver.sdss.org/dr14/en/tools/chart/navi.aspx>



**Fig. 6.** *Left panel*: redshifts of quasars against  $[W1-W2]$  color. 48 282 quasars with valid redshift from LAMOST DR5 are plotted. The dashed blue line is the  $[W1-W2] \geq 0.8$  cut and dashed orange line is the fitting curve of the edge. *Right panel*: normalized histogram of the redshift of quasar candidates in this work. Redshift value is extracted from SDSS DR14 (145 205 sources) and LAMOST DR5 (14 265 sources). In total, 159 470 valid redshifts of quasars are plotted.



**Fig. 7.** Proper motions of quasars against  $R_F$  magnitude. The red triangles indicate the mean proper motions of our catalog of  $\alpha^*$  and  $\delta$  in that magnitude bin and the error bar shows their standard deviations in  $\text{mas yr}^{-1}$ . The blue circles indicate the mean proper motions of MIRAGNs in each 0.75-magnitude bin and the error bar indicates their standard deviations.

It is possible to exclude many more spurious sources, with the astrometry technique improved in the future.

#### 4.3. Redshifts

At high redshift ( $z > 3.5$ ), when the  $H\alpha$  line is redshifted to the W1 band or the  $\sim 1 \mu\text{m}$  minimum enters the W2 band (see detail in Stern et al. 2012), the  $[W1-W2]$  color of AGNs may be bluer (i.e., smaller) than 0.8. The left panel of Fig. 6 confirms this behavior. Therefore, our catalog loses some high-redshift quasars as all mid-infrared selection methods do. The right panel of Fig. 6 is the histogram of the quasars which have valid redshifts from our catalog. Most of the redshifts are less than three and no significant high-redshift ( $z > 3.5$ ) quasars are identified. It should be noted that since the left panel of Fig. 6 contains only the quasars from LAMOST, this figure may be biased by the observation strategy of LAMOST.

#### 4.4. Comparison with MIRAGN

Figure 7 shows the proper motions of the MIRAGNs (Secrest et al. 2015) and our quasar sample against  $R_F$  magnitude and indicates that they are consistent with the zero-point

**Table 2.** Description of our catalog.

Label	Type	Units	Detail
ID1	Int	–	ID1 of the unique union id (ID1, ID2)
ID2	Int	–	ID2 of the unique union id (ID1, ID2)
RAdeg	Double	degree	Right Ascension in J2000.0
DEdeg	Double	degree	Declination in J2000.0
eRAmas	Double	mas	Error of right ascension
eDecmas	Double	mas	Error of declination
pmRA	Double	mas yr <sup>-1</sup>	$\mu_\alpha \cos \delta$
pmDec	Double	mas yr <sup>-1</sup>	$\mu_\delta$
epmRA	Double	mas yr <sup>-1</sup>	Error of $\mu_\alpha \cos \delta$
epmDec	Double	mas yr <sup>-1</sup>	Error of $\mu_\delta$
Rmag	Float	mag	$R_F$ photographic magnitude
Bmag	Float	mag	$B_J$ photographic magnitude
Vmag	Float	mag	$V$ photographic magnitude
Imag	Float	mag	$I_N$ photographic magnitude
Jmag	Float	mag	2MASS J magnitude
Hmag	Float	mag	2MASS H magnitude
Kmag	Float	mag	2MASS K magnitude
W1mag	Float	mag	ALLWISE W1 magnitude
W2mag	Float	mag	ALLWISE W2 magnitude
W3mag	Float	mag	ALLWISE W3 magnitude
W4mag	Float	mag	ALLWISE W4 magnitude
QSOflag	Char	–	Flag on the presence of SDSS or LAMOST
z	Float	–	Redshift
zflag	Char	–	Flag on the sources of redshift
Morflag	Char	–	Flag on the morphology

**Notes.** The catalog is available at the CDS.

error estimated by APOP. The red lines are the proper motions from our catalog and the blue lines from MIRAGN. The mean and dispersion of our work are closer to zero and shows less trend along  $R_F$  magnitude.

## 5. Description of the catalog

The quasar candidate catalog presented in this paper contains 662 753 objects. The catalog is available at the CDS; the description of the catalog data is detailed in Table 2 and we describe the items for each line of the catalog as follows.

- Columns 1 and 2 give the unique union identifier (ID1, ID2)
- Columns 3 and 4 are the  $\alpha$ ,  $\delta$  coordinates given by the APOP catalog in epoch J2000.0 and their errors are in Col. 5 and 6, respectively.
- Columns 7 and 8 give the proper motions  $\mu_\alpha \cos \delta$ ,  $\mu_\delta$  and in Cols. 9 and 10 are their errors, respectively.
- Columns 11, 12, 13, and 14 provide the  $R_F$ ,  $B_J$ ,  $V$ ,  $I_N$  magnitudes directly extracted from APOP.
- Columns 15, 16, and 17 give the near-infrared J ( $1.25 \mu\text{m}$ ), H ( $1.65 \mu\text{m}$ ) and K ( $2.16 \mu\text{m}$ ) magnitudes directly extracted from 2MASS catalog.
- Columns 18, 19, 20, and 21 are the mid-infrared W1 ( $3.4 \mu\text{m}$ ), W2 ( $4.6 \mu\text{m}$ ), W3 ( $8 \mu\text{m}$ ), W4 ( $22 \mu\text{m}$ ) magnitudes directly extracted from ALLWISE.
- Column 22 gives the flags indicating the presence of the quasar in SDSS (flag: “s”) or LAMOST (flag: “g”).
- Column 23 gives the redshift value obtained from SDSS if QSOflag (Col. 22) = “s”; from LAMOST if QSOflag (Col. 22) = “g”; else  $z = -99$ .

- Column 24 gives the flags indicating the sources of redshift from SDSS (zflag = “s”) or from LAMOST (zflag = “g”);  $z = “o”$  for no valid redshift.
- Column 25 is the morphological flag indicating whether these objects are point-like or extend sources. “p” = point-like sources; “e” = extend sources.

## 6. Conclusions

We have presented a catalog of quasar candidates selected using mid-infrared and proper motion methods; 96.7% of our sources are confirmed to be quasar-like, but not all of these quasars are point-like in imaging. We estimate that  $\sim 22.8\%$  may be extended in imaging, limiting their utility for establishing an astrometric reference frame.

Newly released LAMOST data allowed us to investigate the  $[W1-W2] \geq 0.8$  criterion to discriminate quasars from stars and galaxies, confirming previous studies based on SDSS data (Stern et al. 2012). Our sample lies outside the Galactic plane, allowing us to avoid many strongly polluted WISE images near the plane. Also, having used astrometric constraints, the contamination of brown dwarfs with detectable proper motions are also small. A tiny fraction of normal galaxies may be included in our catalog because of mid-infrared emission from star bursts or strong star formation regions in them.

Finally, although about a quarter of the sources in our catalog are extended in imaging, the near-zero proper motion criterion ensures the effectiveness of an external test of the reference frame of Gaia DR2 which was established mainly by MIRAGNs (Mignard et al. 2018). Our quasar candidates can be used to construct a uniformly distributed quasar catalog together

with the confirmed quasars to analyze the parallax zero-point (Michalik & Lindegren 2016) of the Gaia DR2 and the link of the optical reference frame to the ICRS (Mignard et al. 2016).

*Acknowledgements.* We thank the referee for constructive scientific comments. This work is supported by grants from the National Science Foundation of China (NSFC No. 11573054, 11503042, 11703065). 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. Guoshoujing Telescope (the Large Sky Area Multi-Object Fiber Spectroscopic Telescope LAMOST) is a National Major Scientific Project built by the Chinese Academy of Sciences. Funding for the project has been provided by the National Development and Reform Commission. LAMOST is operated and managed by the National Astronomical Observatories, Chinese Academy of Sciences. Funding for the Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions. SDSS-IV acknowledges support and resources from the Center for High-Performance Computing at the University of Utah. The SDSS web site is [www.sdss.org](http://www.sdss.org). SDSS-IV is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, the Chilean Participation Group, the French Participation Group, Harvard-Smithsonian Center for Astrophysics, Instituto de Astrofísica de Canarias, The Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU)/University of Tokyo, Lawrence Berkeley National Laboratory, Leibniz Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), National Astronomical Observatories of China, New Mexico State University, New York University, University of Notre Dame, Observatório Nacional/MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University.

## References

- Arias, E., Charlot, P., Feissel, M., & Lestrade, J.-F. 1995, *A&A*, 303, 604  
 Assef, R., Kochanek, C., Ashby, M., et al. 2011, *ApJ*, 728, 56  
 Bourda, G., Collioud, A., Charlot, P., Porcas, R., & Garrington, S. 2011, *A&A*, 526, A102  
 Boyle, B. J., Shanks, T., Croom, S., et al. 2000, *MNRAS*, 317, 1014  
 Cui, X.-Q., Zhao, Y.-H., Chu, Y.-Q., et al. 2012, *RAA*, 12, 9  
 Cutri, R. M. & et al. 2012, *VizieR Online Data Catalog: II/311*  
 Cutri, R. M. & et al. 2013, *VizieR Online Data Catalog: II/328*  
 Fey, A. L., Gordon, D., Jacobs, C. S., et al. 2015, *AJ*, 150, 58  
 Gaia Collaboration (Brown, A. G. A., et al.) 2018, *A&A*, 616, A1  
 Hagen, H.-J., Groote, D., Engels, D., & Reimers, D. 1995, *A&AS*, 111, 195  
 Hewett, P. C., Foltz, C. B., & Chaffee, F. H. 1995, *AJ*, 109, 1498  
 Jacobs, C. S., Bourda, G., Charlot, P., et al. 2014, in *Astronomía Dinámica en Latino-América (ADeLA)*, 1  
 Kirkpatrick, J. D., Cushing, M. C., Gelino, C. R., et al. 2011, *ApJ*, 197, 19  
 Lacy, M., Storrie-Lombardi, L., Sajina, A., et al. 2004, *ApJ*, 154, 166  
 Lasker, B. M., Lattanzi, M. G., McLean, B. J., et al. 2008, *AJ*, 136, 735  
 Lindegren, L., Lammers, U., Bastian, U., et al. 2016, *A&A*, 595, A4  
 Lindegren, L., Hernandez, J., Bombrun, A., et al. 2018, *A&A*, 616, A2  
 Luo, A.-L., Zhao, Y.-H., Zhao, G., et al. 2015, *Res. A&A*, 15, 1095  
 Ma, C., Arias, E., Eubanks, T., et al. 1998, *AJ*, 116, 516  
 Mainzer, A., Bauer, J., Grav, T., et al. 2011a, *AJ*, 731, 53  
 Mainzer, A., Bauer, J., Grav, T., et al. 2011b, *AJ*, 731, 53  
 Malkin, Z., Jacobs, C., Arias, F., et al. 2015, in *Proc. of the Journées 2014 "Systèmes de référence spatio-temporels": Recent developments and prospects in ground-based and space astrometry*, eds. Z. Malkin, & N. Capitaine, 3  
 Mateos, S., Alonso-Herrero, A., Carrera, F. J., et al. 2012, *MNRAS*, 426, 3271  
 Michalik, D., & Lindegren, L. 2016, *A&A*, 586, A26  
 Mignard, F., Klioner, S., Lindegren, L., et al. 2016, *A&A*, 595, A5  
 Mignard, F., Klioner, S., Lindegren, L., et al. 2018, *A&A*, 616, A14  
 Myers, A. D., Palanque-Delabrouille, N., Prakash, A., et al. 2015, *ApJ*, 221, 27  
 Nikutta, R., Hunt-Walker, N., Nenkova, M., Ivezić, Ž., & Elitzur, M. 2014, *MNRAS*, 442, 3361  
 Pâris, I., Petitjean, P., Aubourg, É., et al. 2012, *A&A*, 548, A66  
 Pâris, I., Petitjean, P., Aubourg, É., et al. 2014, *A&A*, 563, A54  
 Pâris, I., Petitjean, P., Ross, N. P., et al. 2017, *A&A*, 597, A79  
 Pâris, I., Petitjean, P., Aubourg, E., et al. 2018, *A&A* 613, A51  
 Perryman, M. A., Lindegren, L., Kovalevsky, J., et al. 1997, *A&A*, 323, L49  
 Qi, Z., Yu, Y., Bucciarelli, B., et al. 2015, *AJ*, 150, 137  
 Richards, G. T., Lacy, M., Storrie-Lombardi, L. J., et al. 2006, *ApJ*, 166, 470  
 Richards, G. T., Deo, R. P., Lacy, M., et al. 2009, *AJ*, 137, 3884  
 Ross, N. P., Myers, A. D., Sheldon, E. S., et al. 2012, *ApJ*, 199, 3  
 Schneider, D. P., Richards, G. T., Hall, P. B., et al. 2010, *AJ*, 139, 2360  
 Secrest, N., Dudik, R., Dorland, B., et al. 2015, *ApJ*, 221, 12  
 Sharp, R., McMahon, R., Irwin, M., & Hodgkin, S. 2001, *MNRAS*, 326, L45  
 Skrutskie, M., Cutri, R., Stiening, R., et al. 2006, 2001, *AJ*, 131, 1163  
 Souchay, J., Andrei, A., Barache, C., et al. 2012, *A&A*, 537, A99  
 Souchay, J., Andrei, A., Barache, C., et al. 2015, *A&A*, 583, A75  
 Stern, D., Eisenhardt, P., Gorjian, V., et al. 2005, *AJ*, 631, 163  
 Stern, D., Assef, R. J., Benford, D. J., et al. 2012, *AJ*, 753, 30  
 Titov, O. 2010, *MNRAS*, 407  
 Véron-Cetty, M.-P., & Véron, P. 2010, *A&A*, 518, A10  
 Willott, C. J., Delorme, P., Reylé, C., et al. 2010, *AJ*, 139, 906  
 Wright, E. L., Eisenhardt, P. R., Mainzer, A. K., et al. 2010, *AJ*, 140, 1868  
 York, D. G., Adelman, J., Anderson Jr., J. E., et al. 2000, *AJ*, 120, 1579