Reaching completness of the radio loud AGN sample

1 Background

1.1 Complete VLBI samples of extragalactic radio sources

About 10% of active galactic nuclea (AGNs) exhibit parsec scale radio emission. Emission from the core region characterizes the processes that occur in the vicinity of a supermassive black hole. Understanding these processes can be achieved with two ways: 1) selecting a specific source and making intensive observations of the object using different instruments; 2) observing a sample of sources. In the first case we aim to draw causative conclusions, in the second case we aim to draw statistical inference. The statistical inference can be generalized to the entire population if the sample is complete. If the sample is incomplete and biased, there is a risk of propagating this bias to conclusions.

Among full samples, the all-sky samples are special. First, if we found no object in the specific search area, we can say more than just "we did not find": we can say *"there did not exist an object brighter than some limit there on the epoch of observation"*. Second, if we cross-match a trial population with poor position localization against the complete sample, we can draw an inference about their interrelationship using statistical methods even if we cannot reliably associate each given object.

1.2 Current status of VLBI surveys

Starting with the VLBA Calibrator Survey–1 (Beasley et al. 2002), a systematic study of the VLBI sky at the declination range $[-40°, +90°]$ commenced. Observations of 44 observing campaigns were used to derive the Radio Fundamental Catalogue[1] that is updated on a quarterly basis since 2009. The catalogue contains positions, images and estimates of median flux densities at projected baselines shorter 900 km and longer 5,000 km at 8 GHz and other frequencies. Parsec-scale emission at 8 GHz originates mainly from the core. Therefore, this sample, hereafter called the RFC sample, senses synchrotron emission in the jet base. By 2021.01.15 the RFC has 18,764 objects (17,437 with declinations $> -40°$). However, its completeness is poorly understood. Classical logN–logS test shows a deviation from the straight line below 260 mJy. Figure 1 highlights the area of parsec scale flux densities 100–260 mJy by yellow color. We interpret this deviation as a lack of completeness. The origin of the incompleteness is the incompleteness of parent samples used in prior VLBI surveys and the selection bias.

1.3 The radio blazar – neutrino breakthrough

VLBI-selected samples of extragalactic sources are used in many applications including astrometry and space navigation, identification of γ-ray sources, VLBI-*Gaia* differencing, scattering and core-

[1] Available at http://astrogeo.org/rfc

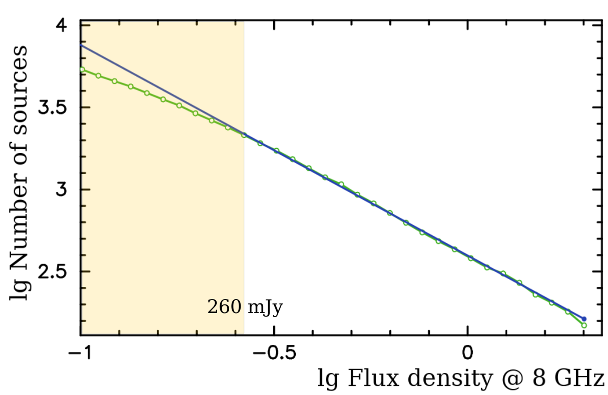

Fig. 1: logn–logs dependence of radio loud AGNS from the RFC catalogue at $\delta > -40°$. The median flux density at 8 GHz at projected VLBA baseline length less than 900 km is used for computation.

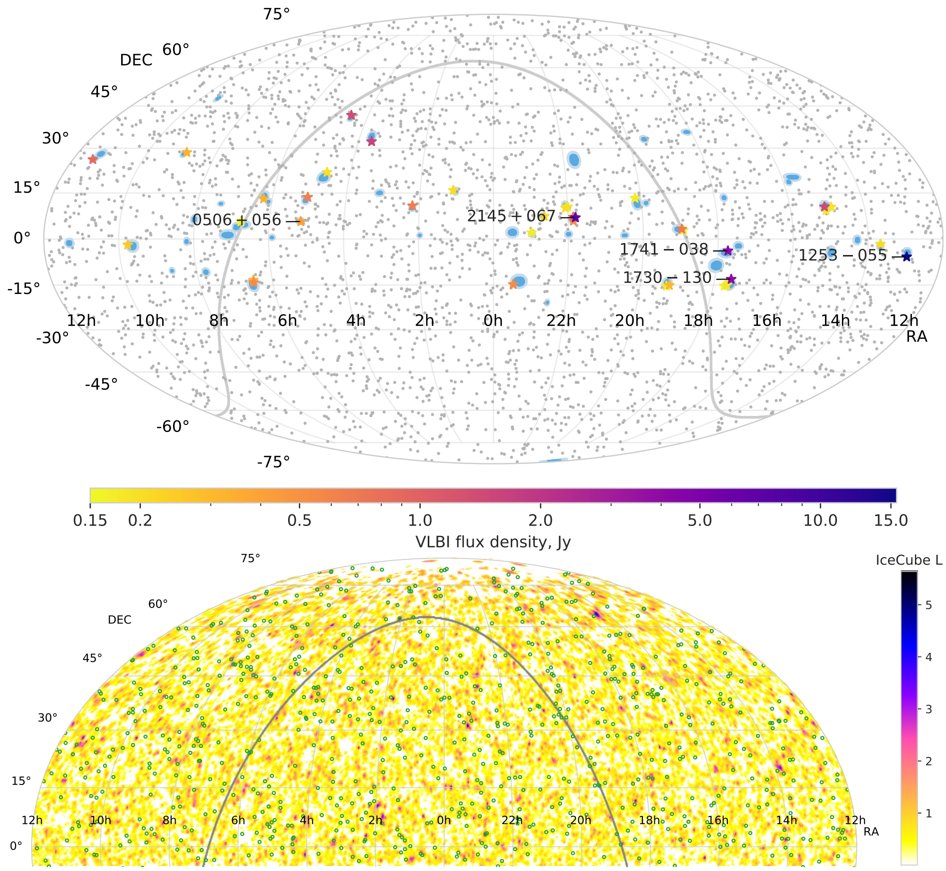

Fig. 2: *Top:* error ellipses of high energy IceCube neutrino events (blue) analyzed against an RFC sub-sample ($S > 0.33$ Jy) (grey dots) from Plavin et al. (2020). *Bottom:* all-sky IceCube astrophysical neutrino probability (color) compared with the same sub-sample (green dots) from Plavin et al. (2021).

shift studies. Recently, exciting new results were presented from an analysis against IceCube neutrino events. The large complete sample of VLBI-selected blazars turned out to be the key to high energy neutrino associations and delivered highly significant result from this analysis (see for details Plavin et al. 2020, 2021). Future associations of new high energy IceCube alerts and analyses of stacked multi-year data will allow scientist to check these results and answer important question regarding neutrino production and protons acceleration in blazars. This requires a deep complete VLBI-selected sample with no significant biases which will serve both, the AGN and neutrino communities in the exciting multi-messenger era.

2 Sample size justification

There is an intrinsic connection between parsec-scale radio emission and the contemporary processes that occur in the vicinity of a supermassive black holes. There is a correction between **parsec-scale** flux density and γ-ray, TeV fluxes, and neutrino events. Due to a limited sensitivity of high energy instruments, we see emission from bright compact radio sources. Small samples like MOJAVE sample that is complete to 1.5 Jy is very useful for a study of strong flaring events. We often can associate these events with individual sources because bright events usually have small localization errors, and high count rate allows to discriminate them from the background. However, small samples are often too shallow to study high energy emission for sources that do not flare or flare not so strongly. That is why there is a need for deeper samples. In that case association is made statistically. For instance, if to use only the sources brighter 1.5 Jy, the p-value of the correlation between neutrinos and blazars is 50% (i.e. insignificant), but if to use a sample of all compact sources brighter 0.33 Jy at 8 GHz, the p-value drops to 0.04% (highly significant). With a planned increase of the sensitivity of neutrino and high energy telescopes, the shallowness of the sample of compact radio sources will become bigger and bigger problem. Figure 3 shows an excess of missed source the zone [13, 22] hours of the right ascension due to a bias in prior observations.

The RFC subsample of sources with parsec-scale flux density brighter 100 mJy at 8 GHz will be a factor of 15 deeper and will have one and half orders of magnitude more sources ($\sim 6,000$ versus 180) than the complete MOJAVE sample. Its average spacial density will be around 0.2 sources per square degree. The specific cutoff 100 mJy reflects the trade-off: above that limit we mainly eliminate the selection bias in prior surveys and below that we tap the population of weaker sources that was not heavily targeted in the past.

3 Proposed observations

In order to reach completeness, we cross-matched the list of sources observed with VLBI that were used for deriving the RFC against a sample that we consider a priori complete within a $60''$ search radius. We take single dish catalogues GB6 (Gregory et al. 1996) and PMN (Griffth et al. 1996) at 4.85 GHz as the parent sample. These catalogues cover the declination range $[-90°, +75°]$ and have a completeness over 99% at a 100 mJy level.

We propose to observe all remaining sources from GB6 and PMN brighter 100 mJy at $\delta > -40°$ that have not been observed before with VLBI. In total, there are 5543 such sources. GB6 and PMN have a considerable fraction of extended sources, especially within 5° of the Galactic plane. To filter out extended sources, we used a preliminary VLASS release (Lacy et al. 2020). We cross-matched the 5543 identified GB6 and PMN sources against VLASS within a $3'$ search radius. If VLASS showed no source with the total flux density over 50 mJy computed over the area of a $2.5''$ radius of the centroid, we discarded it from our target list. The diameter of the total flux density integration area is twice greater than the VLASS beam FWHM, and therefore, provides the upper limit of the unresolved flux density at 3 GHz. A large VLASS cross-match search radius was selected to provide an allowance for position error of GB6 and PMN. Rather low VLASS flux density limit 50 mJy was selected to avoid throwing our a source with rising spectra. (VLASS reference frequency is 3 GHz). We have discarded 1213 sources, mainly in the Galactic plane.

GB6 does not cover the area with $\delta > +75°$. We have selected 100 sources from VLASS with the total flux density greater 100 mJy integrated within a $2.5''$ radius of the centroid position to cover this area. Thus, in total, our parent sample has 4430 objects.

We propose to observe sources using the same approaches as we successfully applied for the Wide-Filed VLBA Calibrator Survey (projects BP192/BC191/BC196/BC201/BP171/BP177 and similar BP242/BP245/BR235) that also used GB6 and PMN as the parent sample among other catalogues. A detailed description of the methodology and results of that survey can be found in Petrov (2021). We have developed automatic scheduling software that has a web interface. In order to generate a schedule, the array operator specifies in the Web-form the start and stop dates during a gap between normal priority projects and subsequently retrieves the key schedule file after 2–3 minutes. This method is very flexible, since only short blocks of time (> 3.5 hours) and networks as small as 7 antennas are required. This approach makes the VLBA scheduling truly dynamic, thereby driving the cost of the project down. Since we request **priority C scheduling mode**, this project does not compete or displace other projects.

Each source will be observed at two wings of C-band at 4.3 and 7.6 GHz in one scan of 60 s long at 2 Gbps. Due to slewing overheads, single frequency or the 4 Gbps mode provide only a

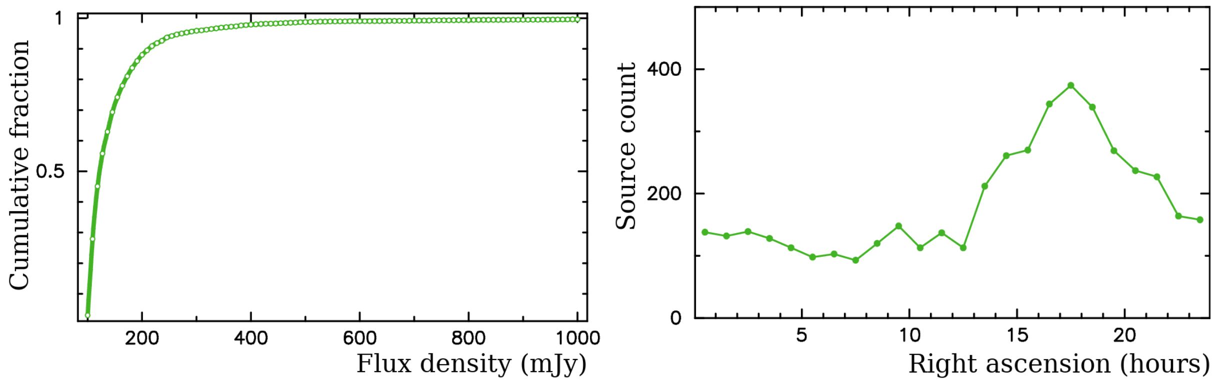

Fig. 3: *Left:* cumulative distribution of target sources over flux density; *Right:* distribution of target sources over right ascensions.

small decrease of observing time. Since all known RFC sources brighter 100 mJy were observed in the dual-band mode, the continuity of dual-band observations has a merit because this provides important information about of radio core spectral indices for the entire sample. Every hour a burst of 4 atmospheric calibrators, aka geodetic blocks will be observed to improve estimates of the atmospheric path delay. Our experience shows that accounting for slewing and atmospheric calibrators, on average, 23 target sources per hour can be scheduled. Thus, 192 hours are required. Since the distribution of time slots suitable for filler schedules over siderial time is not uniform, some zones over right ascensions may be under-observed. Therefore, we request 15% time to allocate in a non-filler mode to catch up sources missed in the filler part of the project. Therefore, we request 168 hours of filler time and 6 blocks of 4 hours of dynamic (i.e. non-filler time). We request the project to span over two semesters.

4 Data release plan

The deliverables of the project are: a) a catalogue of detected sources with positions at milliarcsecond accuracies; b) images at 4.3 and 7.6 GHz. Combining with prior observation, we will deliver a catalogue of compact radio sources that is complete at a 100 mJy level at 8 GHz.

We waive the proprietary period. Images and source positions will be available from the project web site immediately upon processing, typically within 4–8 weeks of observations. Upon project completion, the source positions will augment the RFC, and the images will be submitted to the public Astrogeo VLBI FITS image database. Well designed surveys have considerable "staying power", as illustrated by the citation statistics for two surveys conducted using the Very Large Array (see Figure 1 in Lacy et al. (2020)). The complete unbiased all-VLBA-sky sample of compact sources will be a legacy of the instrument. This sample will be used for a study of the radio-loud AGN population in general and in addition, as a source of phase calibrators, target sources for geodesy and space navigation, matching with other AGNs catalogue, etc.

References

Beasley, A. J., Gordon, D., Peck, A. B., et al. 2002, ApJS, 141, 13.

Gregory, P. C., et al., 1996, ApJS, 103, 427

Griffth, M. R., & Wright, A. E. 1993, AJ, 105, 1666.

Lacy, M., et al. 2020, PASP, 132, 035001

Petrov, L., 2021, AJ, 161, 15

Plavin, A., Kovalev, Y. Y., Kovalev, Y. A., Troitsky, S. V. 2020, ApJ, 894, 101.

Plavin, A. V., Kovalev, Y. Y., Kovalev, Y. A., Troitsky, S. V. 2021, ApJ in press; arXiv:2009.08914