Identifying active galactic nuclei for the SPHEREx space observatory

Supakorn (Nikko) Juengsophonvitavas[1]

Supervisor: Juan V. Hernández Santisteban

[1] *School of Physics and Astronomy, University of St Andrews, UK*

ABSTRACT

We present a catalogue of active galactic nuclei (AGN) candidates identified in the north celestial cap (NCC), a region within a 10° radii around the north celestial pole. The North Celestial Cap Catalogue (NCC-CAT) consists of 914 sources brighter than 19 magnitude with astrometric and photometric characteristics of AGN, and matches in external AGN catalogues. Data used to generate this catalogue consists of Gaia Data Release 3, Gaia Celestial Reference Frame 3, Panoramic Survey Telescope and Rapid Response System Data Release, and Million Optical/Radio/X-Ray Associations Catalogue. The Las Cumbres Observatory will further verify these candidates in this catalogue for reverberation mapping study by the SPHEREx space observatory

Keywords: Active Galactic Nuclei — Catalogue — NCC — NCC-CAT — Gaia DR3 — Gaia CRF3 — Pan-STARRS1 — SPHEREx

1. INTRODUCTION

Active galactic nuclei (AGN) play a significant role in understanding extragalactic astronomy and the Universe's evolution. Although we cannot directly observe the supermassive black hole at the centre, from our understanding of the Unified Model of AGN (Antonucci 1993), we could determine the mass of the black hole by studying the constituent parts using reverberation mapping (Peterson 1993). However, due to the distance of these AGN, it is challenging to observe them using ground-based observatories due to the air mass and long observing time (Horne et al. 2004). In 2025, NASA's SPHEREx mission (Doré et al. 2014), a sky survey satellite, will enable us to study the various structures of the AGN in detail by having multiple passes over the north celestial cap (NCC) region during its mission lifetime with its multiple bandpass filters (Doré et al. 2018). As such, this will overcome the limitation presented above. Although SPHEREx can only survey the sky four times during its two-year mission period, its orbit allows it to monitor the celestial cap up to 300 times around the celestial pole, as seen in Fig. 1. The frequency of observation will allow us to study the variability of the AGN in the NCC. Therefore, it is crucial that we create and validate AGN candidates within $80° \leq \delta \leq 90°$ of the NCC prior to the mission as it is a poorly covered region of the sky by many recent surveys. Especially by those with the capability of allowing us to easily determine whether an object is an AGN, such as SDSS (Gunn et al. 2006), GALEX (Martin et al. 2005) and XMM-Newton (Jansen et al. 2001).

The recent release of Gaia Data Release 3 (Gaia DR3) (Gaia Collaboration et al. 2022b, 2016) come with the Gaia celestial reference frame 3 (Gaia-CRF3) (Gaia Collaboration et al. 2022c) with never seen before completeness (High percentage of AGN) of potential AGN candidates in all parts of the sky. However,

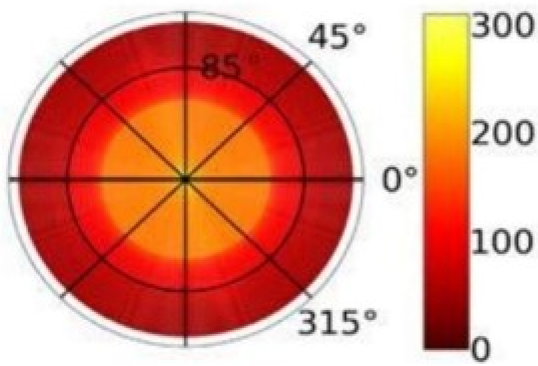

Figure 1. SPHEREx coverage in the celestial cap region. This plot shows the number of times the SPHEREx will pass by the area within its two-year mission life span, with increasing observations towards the pole. Plot taken from the "preparing for the 21-cm cosmology revolution 2015" conference by Asantha Cooray

the completeness comes at the cost of purity (not all sources flagged as AGN are correct). (Gaia Collaboration et al. 2022a) In this work, we use data from Gaia DR3, Panoramic Survey Telescope and Rapid Response System Data Release 2 (PS1 DR2) (Flewelling et al. 2020), Gaia CRF3, and the Million Optical/Radio/X-Ray Associations Catalogue (Flesch 2016) to produce an AGN catalogue with higher purity for the SPHEREx mission.

This paper present the North Celestial Cap Catalogue (NCC-CAT). The candidates presented in this paper are selected based on their astrometric, photometric and other properties indicative of an AGN. Candidates with high luminosity in the short wavelength of the visible light (as these will most likely present "lag" or "echoing") will be observed using optical spectroscopy to verify their nature using the Las Cumbres Observatory (LCO) in the next observing cycle. Section 2 presents all the databases used in this paper which are publicly available with the aim of repeatability by others. A python notebook was also developed to assist with querying PS1 DR2 data where there were querying constraints with the MAST API. Section 3 explores the database cuts performed to create the catalogue. Section 4 discuss the property of the catalogue presented.

2. DATABASES

2.1. *Gaia Data Release 3*

The Gaia satellite (Gaia Collaboration et al. 2016) is an astrometric all-sky survey satellite, which aims to take precise position and proper motion (movement) measurements of all objects up to 20 magnitudes dim in the sky. The smaller the magnitude, the brighter the object. Gaia DR3 provides us with the most accurate astrometric measurements to date and additional information such as low-resolution spectroscopy and light curves on some extragalactic sources. However, data on the extragalactic properties, such as redshift measurement (a unit used to measure the distance of extragalactic sources), is currently not as reliable as SDSS data. As such, Gaia DR3 data is used primarily to determine the proper motion of the sources to locate extragalactic sources. Furthermore, Gaia photometric data will better represent what the SPHEREx satellite will observe as both satellites do not observe through an atmosphere.

2.2. *Pan-STARRS1 Data Release 2*

Pan-STARRS1 (Tonry et al. 2012) is a ground-based digital sky survey which aims to take photometric measurements (Object's brightness) in 5 different photometric bands and astrometric measurements of objects primarily in the north celestial sphere. PS1 DR2 provides us with the most up-to-date photometric data in the NCC compared to past surveys, which allows us to analyse the brightness of these sources in different bands accurately. A known caveat of PS1 DR2 is that the astrometric data in the pole region is inaccurate up to 1"; this error was taken into consideration when performing the cross-matching between PS1 DR2 and Gaia DR3. Additionally, PS1 DR2 photometric data is used to gauge the sources' magnitude when observing through an atmosphere as we aim to verify these sources with spectroscopy using the LCO.

2.3. *Gaia Celestial Reference Frame 3*

Gaia Celestial Reference Frame 3 (Gaia Collaboration et al. 2022c) is the third iteration of Gaia's reference frame system; it is used to keep track and measure how objects move from our point of view changes relative to the background. The background sources would consist of extragalactic sources as they do not appear to move. The Gaia consortium included a cross-matched with 17 additional external AGN and Quasi Stellar Object (QSO) catalogues in this release. QSO are extremely bright AGN.

2.4. *Million Optical/Radio/X-Ray Associations Catalog*

The Million/Optical/Radio/X-ray Associations Catalogue (MORX) (Flesch 2016) is a catalogue of published optical,radio , and x-ray sources in the sky. AGN are known to be very luminous in the X-ray (Elvis et al. 1978). For our purposes, we chose to use this catalogue to perform x-ray matches with our catalogue as Gaia-CRF3 does not have any dedicated x-ray catalogue matches. This catalogue match will help compliment Gaia-CRF3 external catalogue matches.

3. AGN SELECTION METHOD

The definitive way of determining if a source is an AGN is to look at its spectral energy distribution (SED), as AGN has a unique SED, as shown in Elvis et al. (1994). However, this is a time-consuming and costly process; as such, we choose to look for specific characteristics that AGN exhibit using currently available data to determine sources with the highest probability before observing their spectrum. Fig. 2 shows a visual summary of the various cuts performed on the parent sample, and the decisions for these cuts are explained further below.

3.1. *Parent Sample*

As mentioned in 2.1, we used Gaia DR3 database as our parent sample due to its astrometric precision and accuracy. We treated Gaia Database as the leading source, whereas other smaller databases are matched against the leading table. This allows for a better one-to-one matching where a single Gaia source is matched with a single source from the external catalogue instead of a one-to-many match. Furthermore, with the release of Gaia-CRF3, the majority of external AGN catalogues had already been cross-matched with Gaia DR3 data which significantly helped streamline our selection process.

3.2. *Photometry*

Photometry is the study of objects' brightness. AGN are known to be visibly blue in colour (Sandage 1971). Using photometric data from PS1 (how bright the object is in each wavelength), we can deduce the "colour" by subtracting the object's brightness in different bandpasses. Using the conditions from Finlator et al. (2000), we then plot a colour-colour diagram to locate all sources with colours associated with low redshift AGN, as shown in Fig.3. This method is particularly effective against locating low redshift AGN, as the further away the AGN are, the redder it appears. As such, they would move on the colour-colour diagram. However, this is not an issue for us as we aim to identify bright AGN and, therefore, usually closer to us.

Furthermore, we can use photometry to determine if the source is bright enough to observe using specific equipment, i.e. SPHEREx satellite and LCO. In this catalogue, we chose our brightness limit to be 19 magnitude in PS1 DR2 G mean aperture magnitude. This ensures that all sources in this optical cut will be observable by the SPHEREx satellite, which has a brightness limit of 18 magnitude in the i optical band.

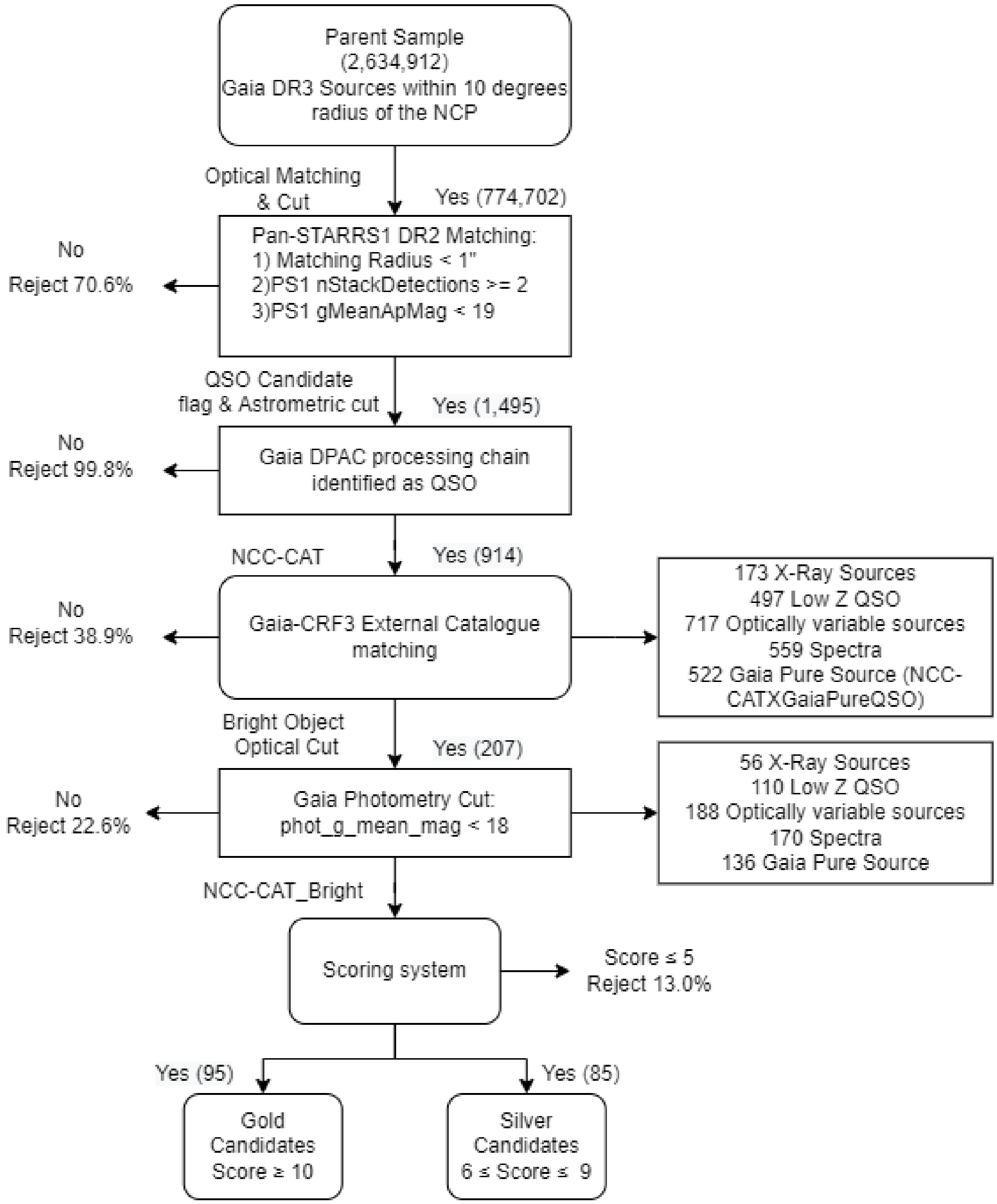

Figure 2. Decision tree showing the various cut made on the Gaia DR3 parent sample to arrive at the catalogue.

A second photometric cut is performed once in the NCC-CAT to locate sources brighter than 18 magnitude in the Gaia optical G band; this cut is performed to locate the brighter sources, which will be verified using the LCO.

At this stage, we are able to find all Gaia sources with PS1 DR2 equivalent that is expected to be observable by SPHEREx; however, these sources will also include bright stars in the Milky Way galaxy and other bright galaxies, not just AGN. As such, we require additional cuts to remove these unwanted objects.

3.3. *Astrometry*

Astrometry is the study of the position and movement of objects in space. AGN are extragalactic objects, meaning they exist outside our galaxy, the Milky Way. Unlike nearby stars in

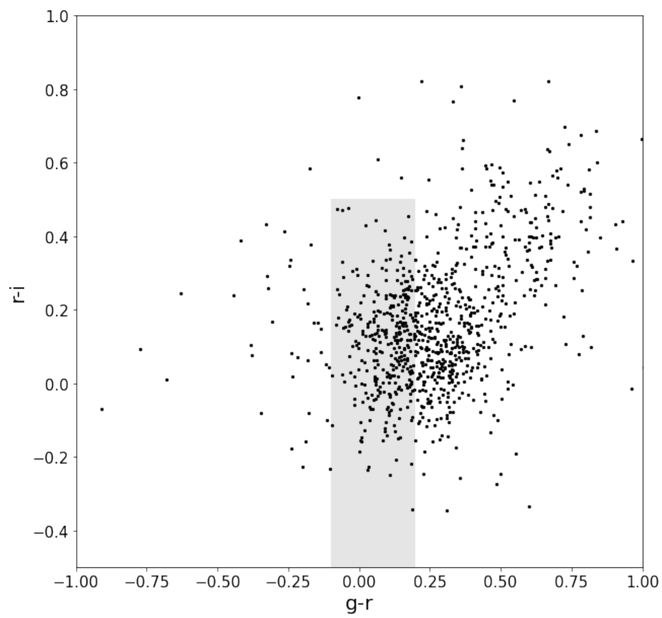

Figure 3. Colour-colour diagram of NCC-CAT candidates. Sources located in the grey region of the diagram are considered low redshift AGN according to Finlator et al. (2000) and are flagged accordingly in our catalogue.

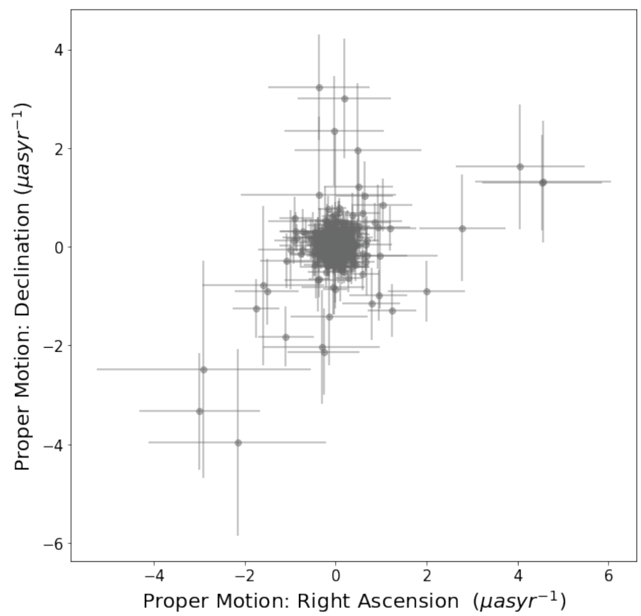

Figure 4. NCC-CAT candidates' proper motion. We can clearly see that the majority of sources are clustered around the centre of the plot with small errors consistent with what we expect of extragalactic sources. Few sources exhibit large proper motion and error, which is consistent with known caveats from Gaia Collaboration et al. (2022a); these are resolved sources.

our galaxy, we would not expect to observe any proper motion from extragalactic objects. We use this property of AGN to perform the second data cut on our parent sample to remove the majority of stars present in our data set. Objects with proper motion equal to zero within statistical uncertainty as provided by Gaia Collaboration et al. (2022c) are kept and further analysed as shown in Fig.4. The astrometry cut was highly effective as it removed over 98.9% of our sources, which are stars, from the photometric cut. However, this method presents some issues as resolved AGN, objects which can be visibly verified, exhibit spurious optical data, which leads to errors in the astrometric data (Gaia Collaboration et al. 2022a).

Despite this, there are workarounds, as stated by Lindegren, L. et al. (2021). Fortunately, GAIA DR3 internal AGN classifier has taken advantage of these workarounds and considered the spurious astrometric data as potential AGN candidates. Using the classifier method has shown to include more sources, those with spurious data while matching with all sources from the astrometric cut alone from our testing; as such, we believe this does not introduce any biases from Gaia. This also has the added benefit of removing normal galaxies from our sample; we now have a sample of likely AGN candidates.

3.4. *External Catalogue and Boolean Credit Scoring*

There are currently still too many candidates that need to be verified. We look to external catalogues that observe these candidates in other spectrums. As mentioned in section 2.3, Gaia-CRF3 had already performed cross-matches with 17 other catalogues; this is incredibly helpful as the more each candidate is flagged as an AGN candidate, the more likely it is going to be an AGN. Additionally, the MORX

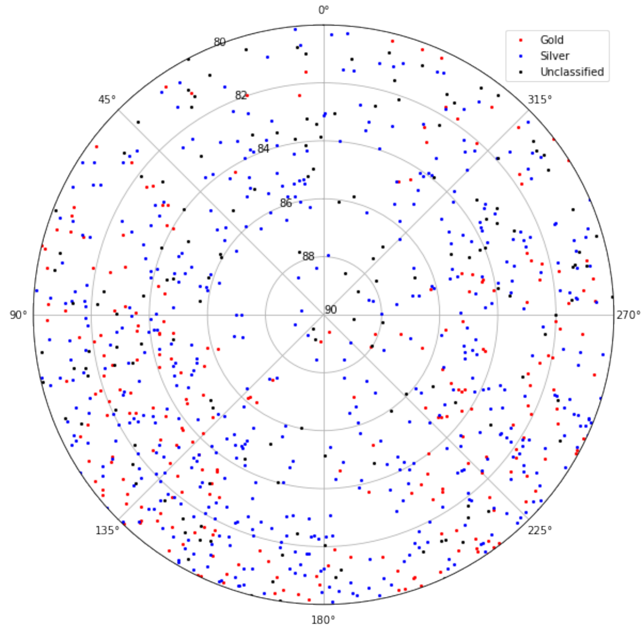

Figure 5. All 914 sources plotted on the NCC region of the sky with colour coding showing the different classification as described in section 3.4. Red points are Gold, blue points are Silver, and black points are unclassified.

catalogue is used to check for x-ray emission from these candidates.

To further assist in selecting the candidates for further verification, we chose to use a Boolean credit scoring system to rank the candidates. To implement this, we assigned additional columns for each property, e.g. x-ray sources, Low Z QSO, individual external catalogue matches, etc, with Boolean value where True equals to 1. Subsequently, we can then sum up the Boolean of each source. As such, we ended up with two classifications of sources, Gold and Silver, with the following conditions. $Gold : Score \geq 10, Silver : 9 \leq score \leq 6$

4. CATALOGUE

The NCC-CAT consists of 914 sources with PS1 DR2 G band magnitude lower than 19 with properties consistent with AGN. Two CSV files are available with this paper which consist of *NCC-CAT.csv* and *NCC-CAT-master.csv*. The former is intended for general viewing of the candidates with all the necessary information used in this paper. The latter contains all the additional information on the candidates, such as the external catalogue name, astrometric and photometric errors, and Gaia's data on the extragalactic nature of these sources from Gaia Collaboration et al. (2022a). The catalogue is generated mainly using TOPCAT (Taylor 2005) for crossing matching source_id and coordinates from different external databases. Table.1 shows the header for NCC-CAT.csv.

4.1. *Completeness and Purity*

This catalogue focuses on the completeness as we want to be able to identify as many AGN as possible throughout the NCC. Despite this, we were able to get 78.4% match with all sources and 100% for sources brighter than 18 G band magnitude when cross-matching with Gaia extragalactic variable classifier (vari_best_class_name = "AGN") (Gaia Collaboration et al. 2022b); this is different from the DPAC classifier or the internal classifier used during the astrometric cut as this take in further consideration of both variability of the source and external catalogue matches. As such, a small fraction of candidates are unlikely to be AGN; however, they exhibit interesting features during the external catalogue matching and would be interesting for further evaluation if possible.

Table 1. *NCC-CAT.csv* Header explanation.

Column	Header	Data Type	Unit	Notes
1	NCC_ID	String	Dimensionless	Unique catalogue's identifier
2	source_id	Long	Degrees	Gaia DR3 unique catalogue's identifier
3	solution_id	Long	Degrees	Gaia DR3 Solution ID
4	RA	Float	Degrees	RA coordinates from Gaia DR3 (J2016)
5	DEC	Float	Degrees	DEC coordinates from Gaia DR3 (J2016)
6	pole_score	Double	Dimensionless	A measure of how close the source is to the pole
7	parallax	Double	Degrees	Parallax from Gaia DR3
8	pm	Double	$\mu as\ yr^{-1}$	Parallax from Gaia DR3
9	astrometric_selection_flag	Boolean	Dimensionless	Candidate's astrometric data consistent with extragalactic sources
10	phot_g_mean_mag	Double	Mag	G band magnitude from Gaia DR3
11	gMeanApMag	Double	Mag	G band aperture magnitude from PS1
12	redshift_qsoc	Double	z	Redshift estimate from (Gaia Collaboration et al. 2022a)
13	classprob_dsc_combmod_quasar	Double	Dimensionless	Probability of candidate being a quasar by Gaia DR3
14	Gold_Candidates	Boolean	Dimensionless	Gold Classification Candidate
15	Silver_Candidates	Boolean	Dimensionless	Silver Classification Candidate
16	Boolean_score	Short	Dimensionless	Boolean credit score as mentioned in 3.4
17	Variable	Boolean	Dimensionless	Candidate has variability as detected by Gaia DR3
18	in_qso_candidates	Boolean	Dimensionless	Identified as QSO candidate by Gaia DR3
19	has_xp_continuous	Boolean	Dimensionless	Candidate has BP/RP spectrum in continuous representation from Gaia DR3
20	has_epoch_photometry	Boolean	Dimensionless	Candidate has light curve data from Gaia DR3
21	LowZQSO	Boolean	Dimensionless	Candidate lies within the low redshift region of the color-color diagram as mentioned in 3.2
22	phot_g_mean_mag¡18	Boolean	Dimensionless	Candidate is brighter than 18 magnitude in Gaia's G band
23	Pure_sample	Boolean	Dimensionless	Candidate has a probability of greater than 50% of being an AGN by Gaia classifier
24 - 42	External Catalogue Matches	Boolean	Dimensionless	Candidates with a match in a specific external catalogue is given a TRUE in the corresponding column. Column (42) is a match with MORX catalogue

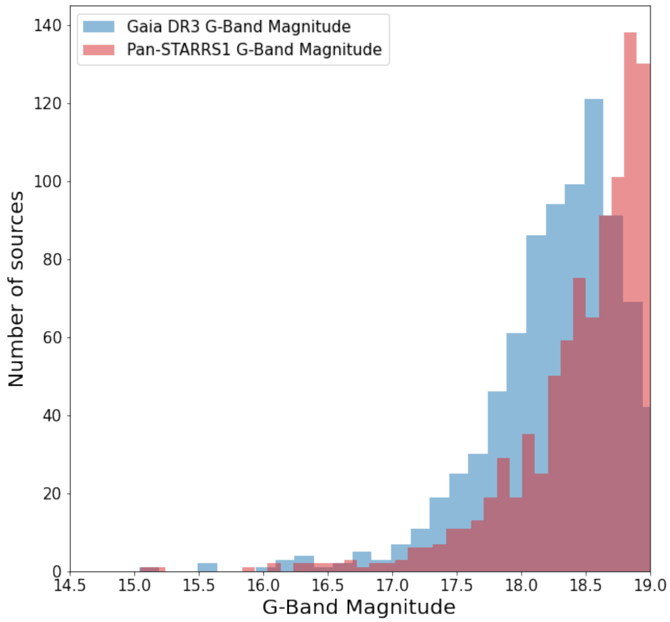

Figure 6. NCC-CAT candidates' magnitude histogram shows the magnitude distribution from Pan-STARRS1 and Gaia. We observe that magnitude measured by Gaia is brighter; this is due to Gaia not having to observe through an atmosphere.

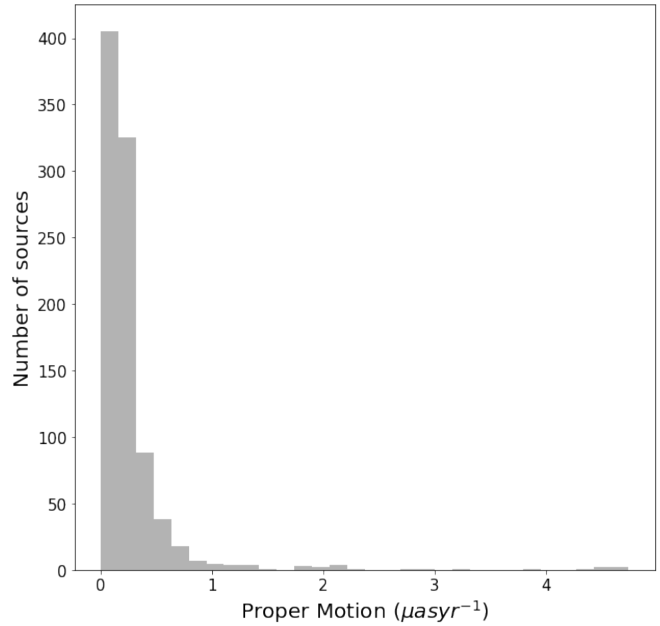

Figure 7. NCC-CAT candidates' proper motion histogram.

4.2. *Candidate Distribution and Pole Score*

Fig.5 shows all the candidates plotted on the NCC region, with the classification label in the legend. As shown in Fig.1, the closer to the celestial pole, the higher the frequency the SPHEREx satellite will observed. As such, candidates closer to the pole will be observe at a higher frequency which is ideal for measuring the echoing from the AGN. The following calculation was made to determine how close to the poles these candidates are, $pole\_score = 1 - \frac{90-dec}{10}$. A score of 0 is given to candidates located at 80° declination and increases linearly to a score of 1 for candidates at 90° declination. Unfortunately, the density of candidates 5 °around the pole is only 1.77 Sources per deg^2 in contrast to the area between 80°and 85°where the density nearly doubles at 3.29 Sources per deg^2.

4.3. *Photometric and Astrometric properties*

Fig.6 shows the magnitude distribution of the candidates in both Gaia DDR3 and PS1 DR2 G band. As expected, we observe that the photometric data from Gaia is brighter than PS1 DR2 due to PS1 DR2 observing through the atmosphere. 497 candidates lie within the low redshift QSO region of the colour-colour diagram as indicated by Finlator et al. (2000), of which 76 candidates are brighter than 18 magnitudes in PS1 DR2.

Fig.7 shows the proper motion distribution of the candidates in this catalogue, as alluded to in section 3.3, the majority of the candidates have small proper motion consistent with extragalactic sources.

4.4. *Scorecard*

A scorecard system is used to ease accessing and viewing the individual candidates. The scorecards can be viewed in *NCC-CAT-Scorecard.pdf.* Each pdf page of the score card file shows the information of individual candidates on the page as well as Digitized Sky Survey (DSS-2) (Lasker 1995) in different bands, Bp/Rp spectra, and photometry where available. The candidates are ordered by the NCC_ID. If additional information is needed, either the NCC_ID or source_id could be used to search for the candidate in *NCC-CAT-master.csv* using any CSV reader of your choice. An example of a candidate scorecard is shown in Fig.8

5. CONCLUSION

We presented a catalogue of potential AGN candidates within 10° radii of the north celestial pole using Pan-STARRS1 DR2, Gaia DR3/CRF3, and MORX catalogue for the SPHEREx space observatory. The catalogue contains 914 candidates with a magnitude less than 19 in Pan-STARRS1 G Band; each ranked into different tiers depending on their likelihood. We find that our methodology cuts were able to remove the majority of stellar and galaxies contamination as indicated by Gaia extragalactic data (Gaia Collaboration et al. 2022a). This catalogue should help validate extragalactic sources in Gaia DR3 data and be ready for further validation using the Las Cumbres Observatory. Three files are included alongside this paper upon request which includes *NCC-CAT.csv*, *NCC-CAT-master.csv* and *NCC-CAT-Scorecard.pdf*.

ACKNOWLEDGMENTS

This work would not have been possible without the generous support and opportunity from my supervisor Dr Juan V. Hernández Santisteban, Lord Laidlaw, and the Laidlaw Foundation. I would like to further thank my fellow Laidlaw scholars, friends, and the PhD students for supporting me throughout the research, as well as Mr.Rothwell and Mr.Mitchell for supporting me on this journey.

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.

The Pan-STARRS1 Surveys (PS1) and the PS1 public science archive have been made possible through contributions by the Institute for Astronomy, the University of Hawaii, the Pan-STARRS Project Office, the Max-Planck Society and its participating institutes, the Max Planck Institute for Astronomy, Heidelberg and the Max Planck Institute for Extraterrestrial Physics, Garching, The Johns Hopkins University, Durham University, the University of Edinburgh, the Queen's University Belfast, the Harvard-Smithsonian Center for Astrophysics, the Las Cumbres Observatory Global Telescope Network Incorporated, the National Central University of Taiwan, the Space Telescope Science Institute, the National Aeronautics and Space Administration under Grant No. NNX08AR22G issued through the Planetary Science Division of the NASA Science Mission Directorate, the National Science Foundation Grant No. AST-1238877, the University of Maryland, Eotvos Lorand University (ELTE), the Los Alamos National Laboratory, and the Gordon and Betty Moore Foundation.

The Digitized Sky Survey was produced at the Space Telescope Science Institute under U.S. Government grant NAG W-2166. The images of these surveys are based on photographic data obtained using the Oschin Schmidt Telescope on Palomar Mountain and the UK Schmidt Telescope. The plates were processed into the present compressed digital form with the permission of these institutions.

DATA AVAILABILITY

The NCC-CAT files and other python notebooks used in this paper is available in a OneDrive folder upon request by contacting the author at sj91@st-andrews.ac.uk.

APPENDIX

A. SCORECARD EXAMPLE

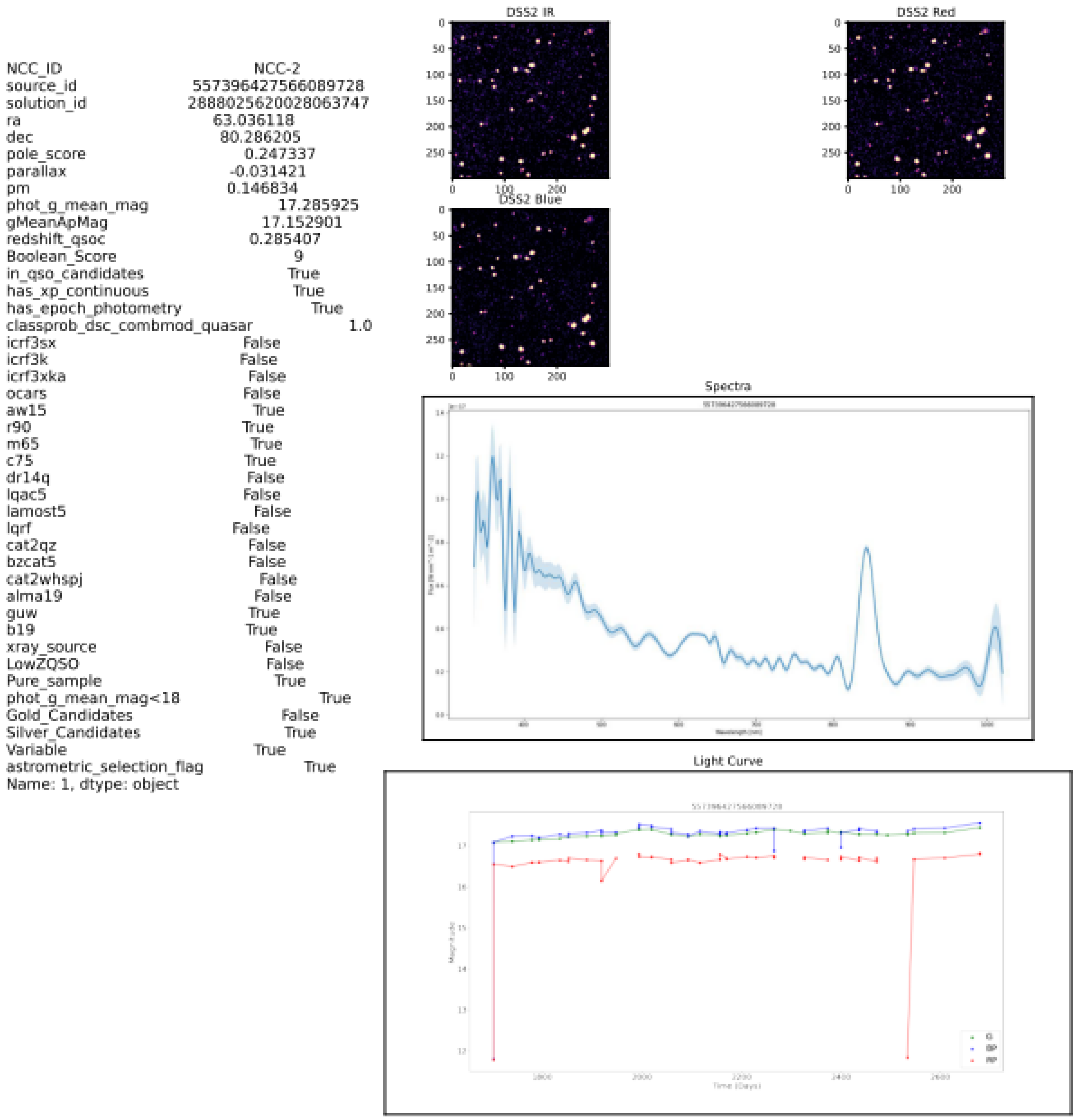

Figure 8. Example of a candidate scorecard

REFERENCES

Antonucci, R. 1993, ARA&A, 31, 473, doi: 10.1146/annurev.aa.31.090193.002353

Doré, O., Bock, J., Ashby, M., et al. 2014, arXiv e-prints, arXiv:1412.4872. https://arxiv.org/abs/1412.4872

Doré, O., Werner, M. W., Ashby, M. L. N., et al. 2018, Science Impacts of the SPHEREx All-Sky Optical to Near-Infrared Spectral Survey II: Report of a Community Workshop on the Scientific Synergies Between the SPHEREx Survey and Other Astronomy Observatories, arXiv, doi: 10.48550/ARXIV.1805.05489

Elvis, M., Maccacaro, T., Wilson, A. S., et al. 1978, MNRAS, 183, 129, doi: 10.1093/mnras/183.2.129

Elvis, M., Wilkes, B. J., McDowell, J. C., et al. 1994, ApJS, 95, 1, doi: 10.1086/192093

Finlator, K., Ivezić, Ž., Fan, X., et al. 2000, AJ, 120, 2615, doi: 10.1086/316824

Flesch, E. W. 2016, PASA, 33, e052, doi: 10.1017/pasa.2016.44

Flewelling, H. A., Magnier, E. A., Chambers, K. C., et al. 2020, The Astrophysical Journal Supplement Series, 251, 7, doi: 10.3847/1538-4365/abb82d

Gaia Collaboration, Bailer-Jones, Coryn A. L., & al. 2022a, A&A, doi: 10.1051/0004-6361/202243232

Gaia Collaboration, Vallenari, A., Brown, A.G.A., Prusti, T., & et al. 2022b, A&A, doi: 10.1051/0004-6361/202243940

Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, A&A, 595, A1, doi: 10.1051/0004-6361/201629272

Gaia Collaboration, Klioner, S. A., Lindegren, L., et al. 2022c, arXiv e-prints, arXiv:2204.12574. https://arxiv.org/abs/2204.12574

Gunn, J. E., Siegmund, W. A., Mannery, E. J., et al. 2006, AJ, 131, 2332, doi: 10.1086/500975

Horne, K., Peterson, B. M., Collier, S. J., & Netzer, H. 2004, PASP, 116, 465, doi: 10.1086/420755

Jansen, F., Lumb, D., Altieri, B., et al. 2001, A&A, 365, L1, doi: 10.1051/0004-6361:20000036

Lasker, B. M. 1995, PASP, 107, 763, doi: 10.1086/133620

Lindegren, L., Klioner, S. A., Hernández, J., et al. 2021, A&A, 649, A2, doi: 10.1051/0004-6361/202039709

Martin, D. C., Fanson, J., Schiminovich, D., et al. 2005, ApJL, 619, L1, doi: 10.1086/426387

Peterson, B. M. 1993, PASP, 105, 247, doi: 10.1086/133140

Sandage, A. R. 1971, in Study Week on Nuclei of Galaxies, ed. D. J. K. O'Connell, 271

Taylor, M. B. 2005, in Astronomical Society of the Pacific Conference Series, Vol. 347, Astronomical Data Analysis Software and Systems XIV, ed. P. Shopbell, M. Britton, & R. Ebert, 29

Tonry, J. L., Stubbs, C. W., Lykke, K. R., et al. 2012, The Astrophysical Journal, 750, 99, doi: 10.1088/0004-637x/750/2/99