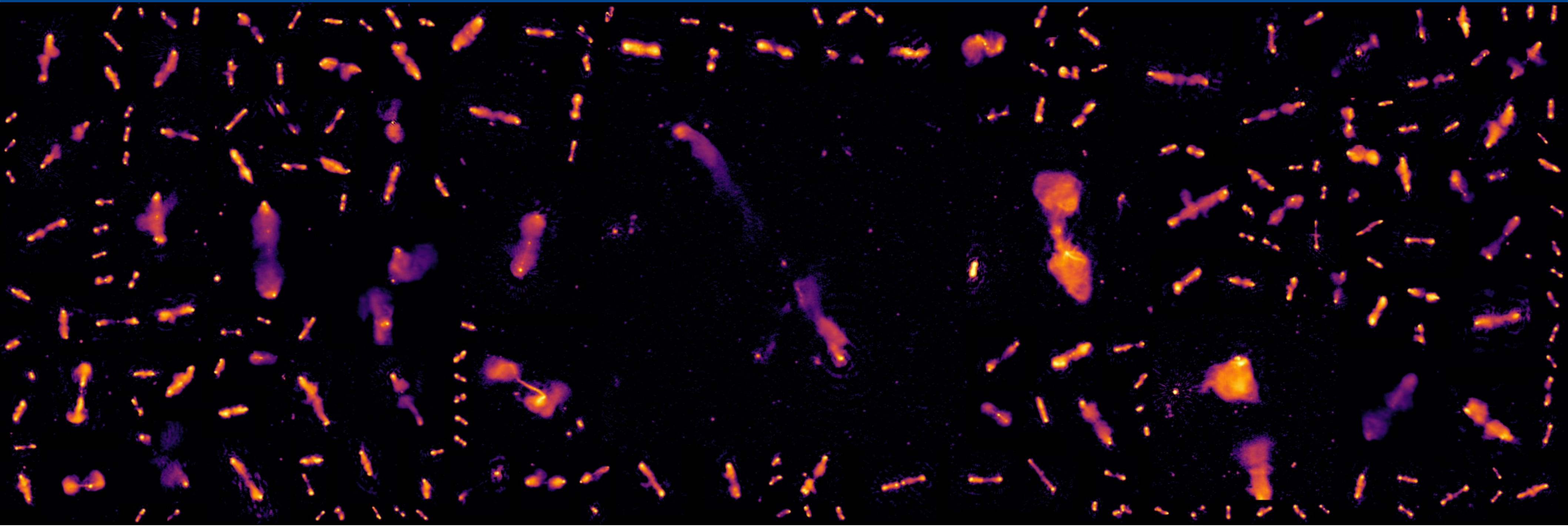
# Optical identifications in LoTSS DR2: from citizens to science



Martin Hardcastle

LOFAR family meeting, Leiden, June 2024

Thanks to: Maya Horton, Wendy Williams, Tim Shimwell, Ken Duncan ++

University of Hertfordshire UH

# Optical IDs for DR2

- Summarizing published paper
- See arXiv:2309.00102
  for many more details

A&A, 678, A151 (2023)
https://doi.org/10.1051/0004-6361/202347333


Astronomy & Astrophysics

## The LOFAR Two-Metre Sky Survey

### VI. Optical identifications for the second data release★

M. J. Hardcastle[1], M. A. Horton[1,2], W. L. Williams[3], K. J. Duncan[4], L. Alegre[4], B. Barkus[5], J. H. Croston[5], H. Dickinson[5], E. Osinga[6], H. J. A. Röttgering[6], J. Sabater[4], T. W. Shimwell[7], D. J. B. Smith[1], P. N. Best[4], A. Botteon[16], M. Brüggen[17], A. Drabent[10], F. de Gasperin[16,17], G. Gürkan[1,10,20], M. Hajduk[15], C. L. Hale[4], M. Hoeft[10], M. Jamrozy[8], M. Kunert-Bajraszewska[14], R. Kondapally[4], M. Magliocchetti[12], V. H. Mahatma[10], R. I. J. Mostert[6,7], S. P. O'Sullivan[21], U. Pajdosz-Śmierciak[8], J. Petley[13], J. C. S. Pierce[1], I. Prandoni[16], D. J. Schwarz[11], A. Shulewski[7], T. M. Siewert[11], J. P. Stott[19], H. Tang[22], M. Vaccari[23,24,16], X. Zheng[6,18], T. Bailey[25], S. Desbled[25], A. Goyal[7], V. Gonano[25], M. Hanset[25], W. Kurtz[25], S. M. Lim[25], L. Mielle[25], C. S. Molloy[25], R. Roth[25], I. A. Terentev[25], and M. Torres[9]

*(Affiliations can be found after the references)*

Received 1 July 2023 / Accepted 29 August 2023

**ABSTRACT**

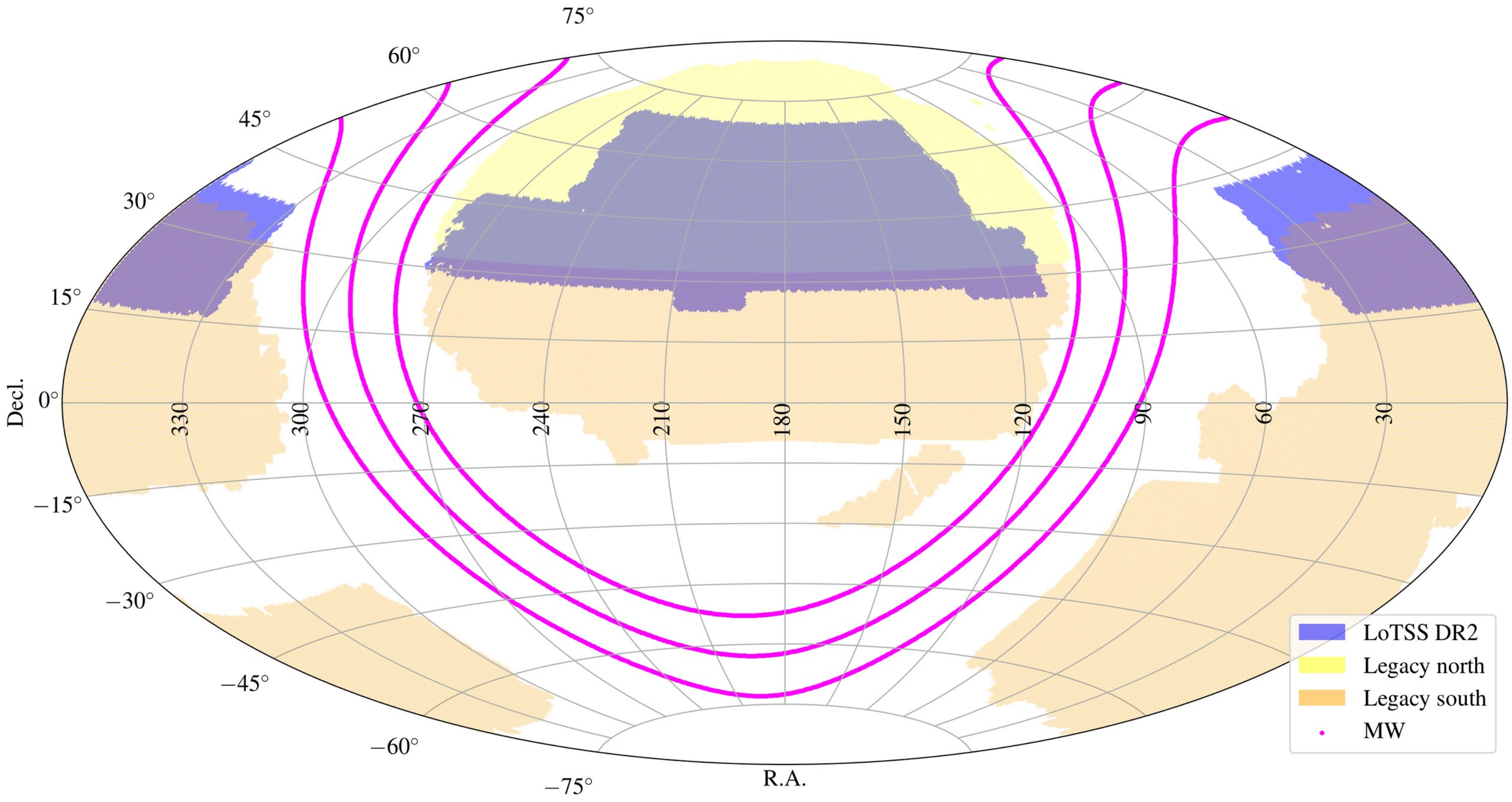
The second data release of the LOFAR Two-Metre Sky Survey (LoTSS) covers 27% of the northern sky, with a total area of ~5700 deg$^2$. The high angular resolution of LOFAR with Dutch baselines (6 arcsec) allows us to carry out optical identifications of a large fraction of the detected radio sources without further radio followup; however, the process is made more challenging by the many extended radio sources found in LOFAR images as a result of its excellent sensitivity to extended structure. In this paper we present source associations and identifications for sources in the second data release based on optical and near-infrared data, using a combination of a likelihood-ratio cross-match method developed for our first data release, our citizen science project Radio Galaxy Zoo: LOFAR, and new approaches to algorithmic optical identification, together with extensive visual inspection by astronomers. We also present spectroscopic or photometric redshifts for a large fraction of the optical identifications. In total 4 116 934 radio sources lie in the area with good optical data, of which 85% have an optical or infrared identification and 58% have a good redshift estimate. We demonstrate the quality of the dataset by comparing it with earlier optically identified radio surveys. This is by far the largest ever optically identified radio catalogue, and will permit robust statistical studies of star-forming and radio-loud active galaxies.

**Key words.** catalogs – radio continuum: galaxies

# Optical ids for DR2

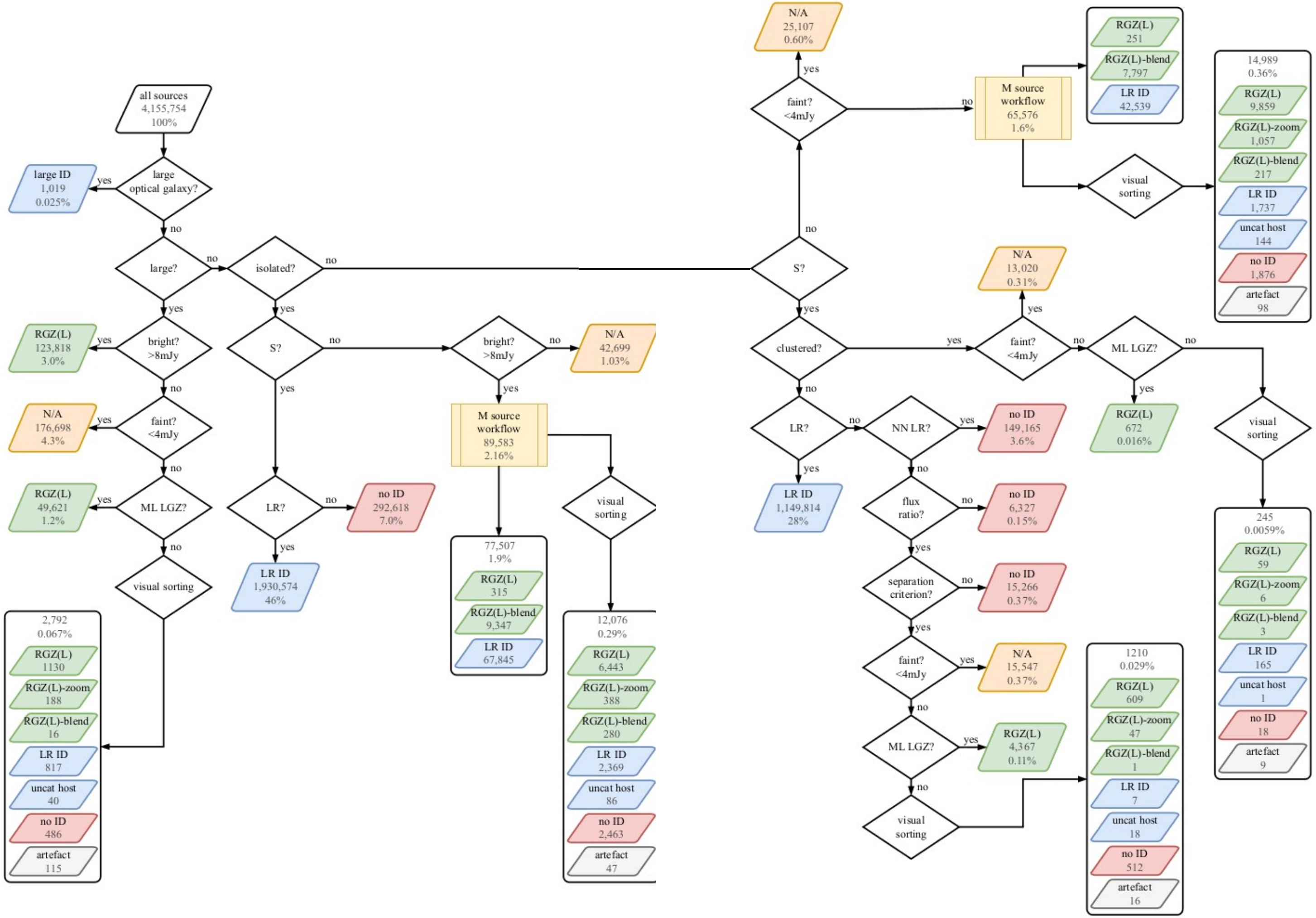
- ‘Nothing so useless as a radio source’
- To understand & model the radio source population we need to find the optical sources corresponding to each radio object
- Radio structures can be physically much larger than host galaxy in the case of radio AGN → can’t just use positional crossmatching
- Start from DESI Legacy Survey and WISE...

# DESI Legacy Survey coverage



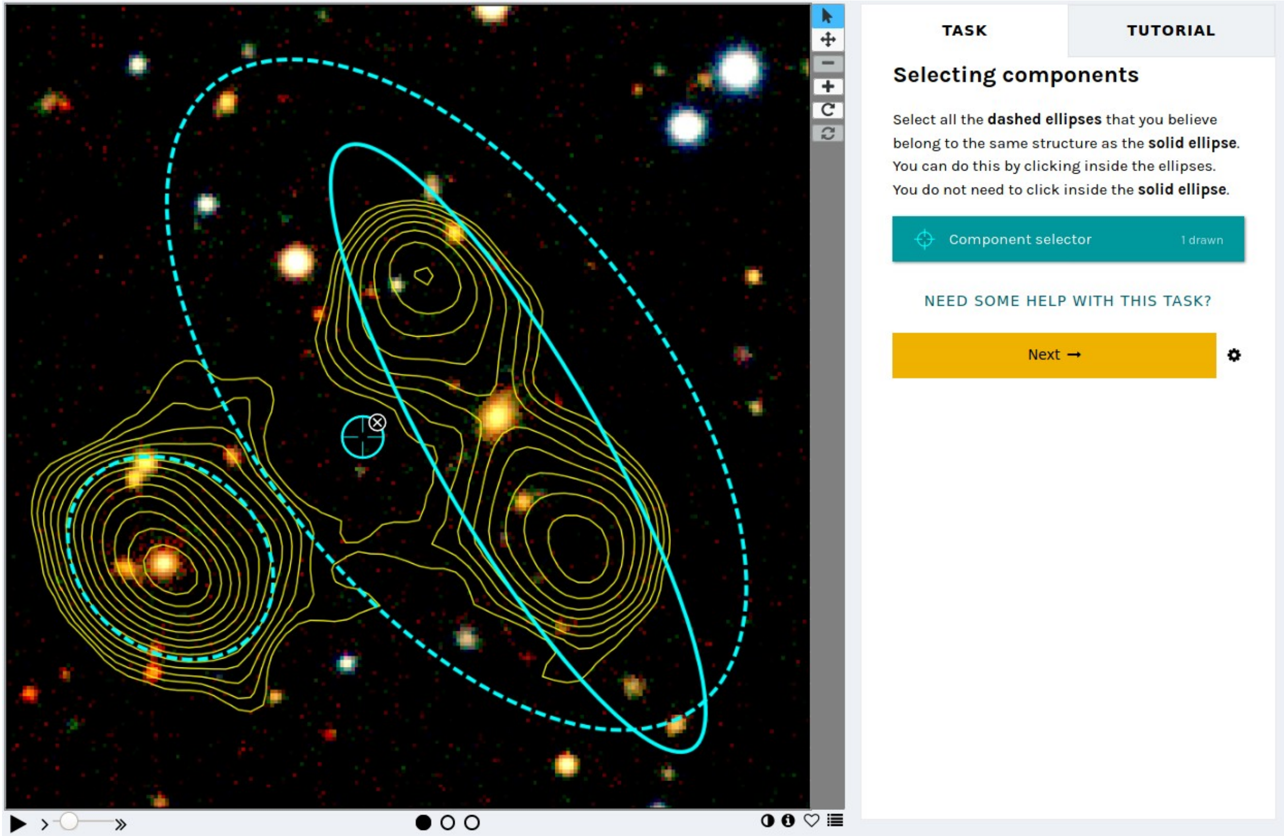
# ‘Decision tree’ for optical IDs

LR/auto
RGZ
No id
Faint



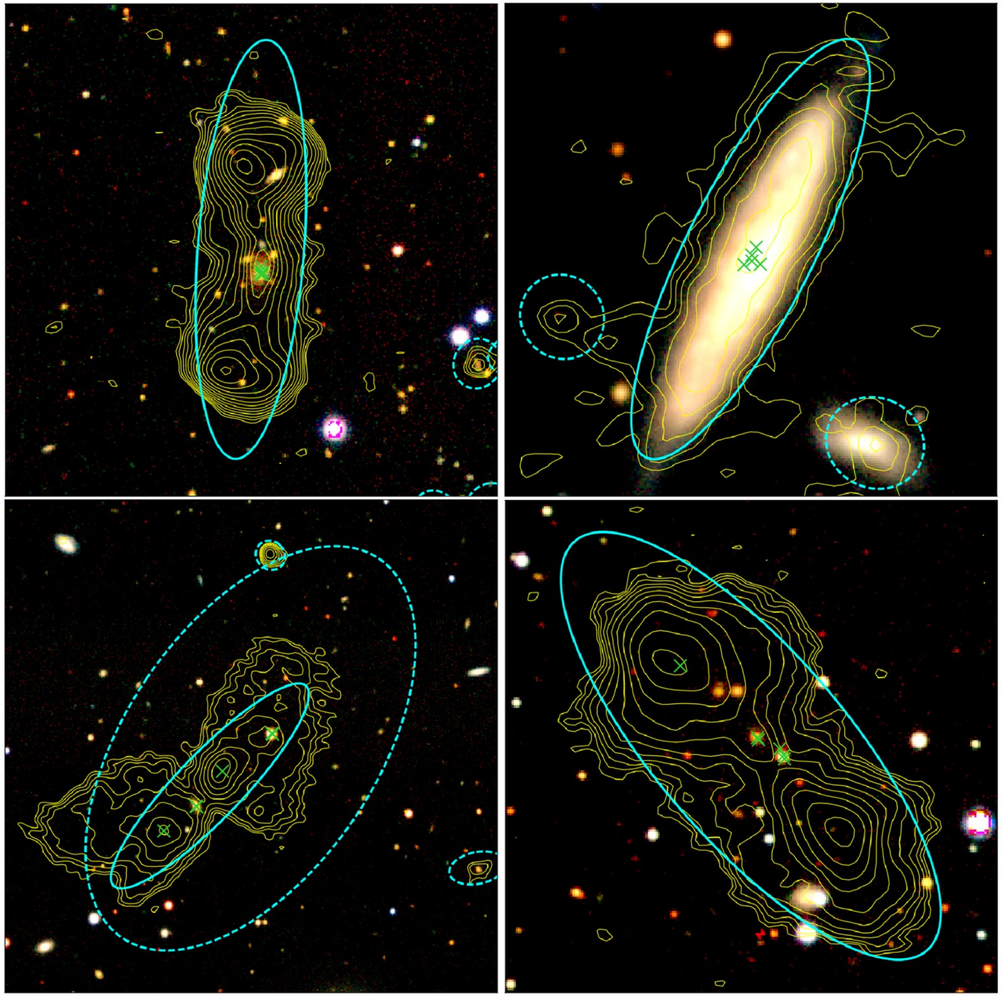
# Citizen science for DR2

- Following on from successful Radio Galaxy Zoo project using FIRST and WISE images
- Volunteers simultaneously ‘associate’ radio sources – say which ones go together – and ‘identify’ them – what is the optical counterpart to a given possibly composite radio source
- 189,375 sources sent to classification by citizen scientists (at 5 views each)
- 957,374 classifications
- 13,711 distinct volunteers
- About 3 years to run!!!



# Citizen science outputs

- Association usually fairly good
- Optical identification could be poor even on ‘obvious’ sources…
- Examples to the right – crosses show positions of citizen science ID clicks. Top two successfully IDed, bottom two failed even though host is clear



# Volunteer population stats

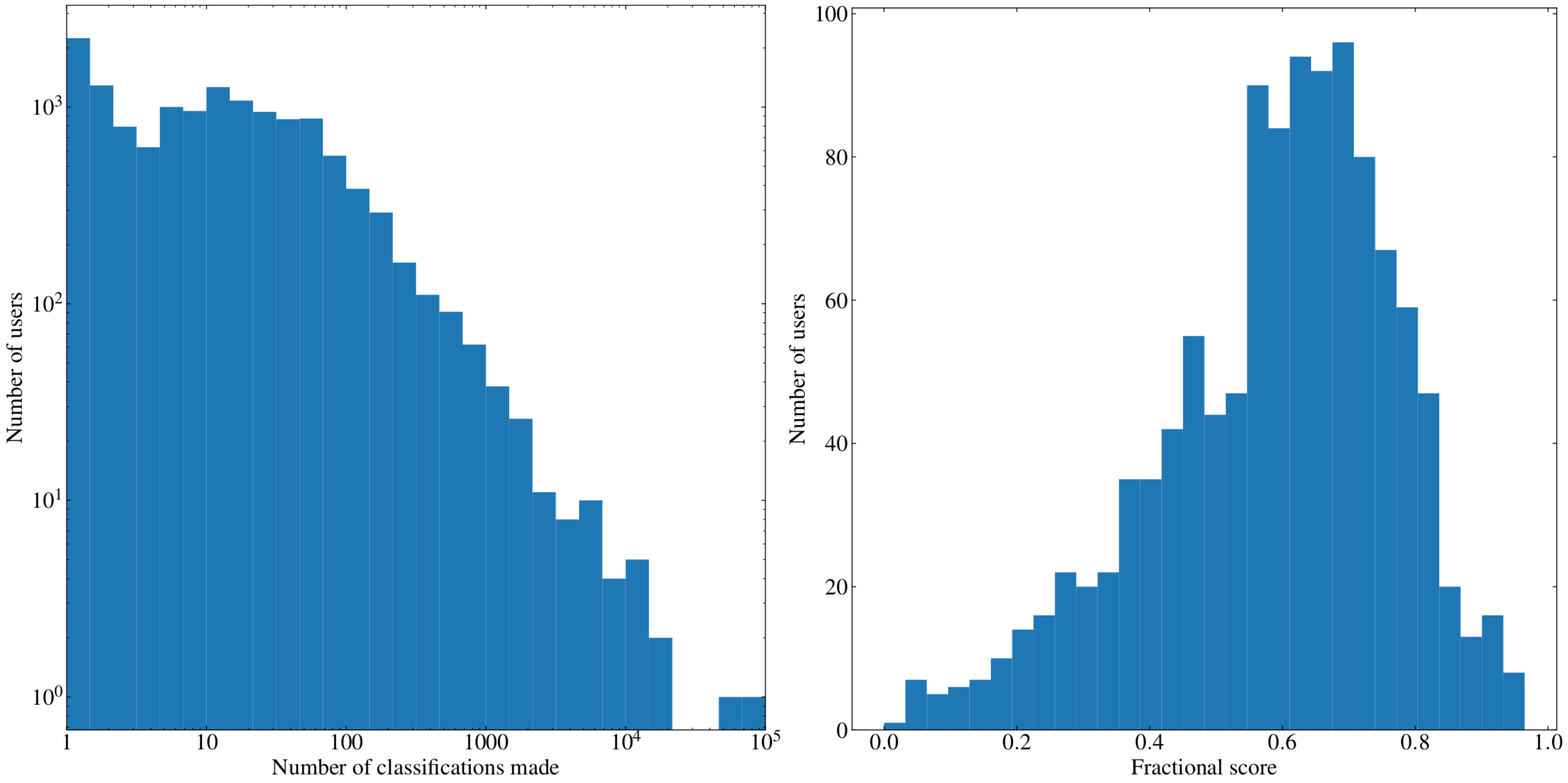
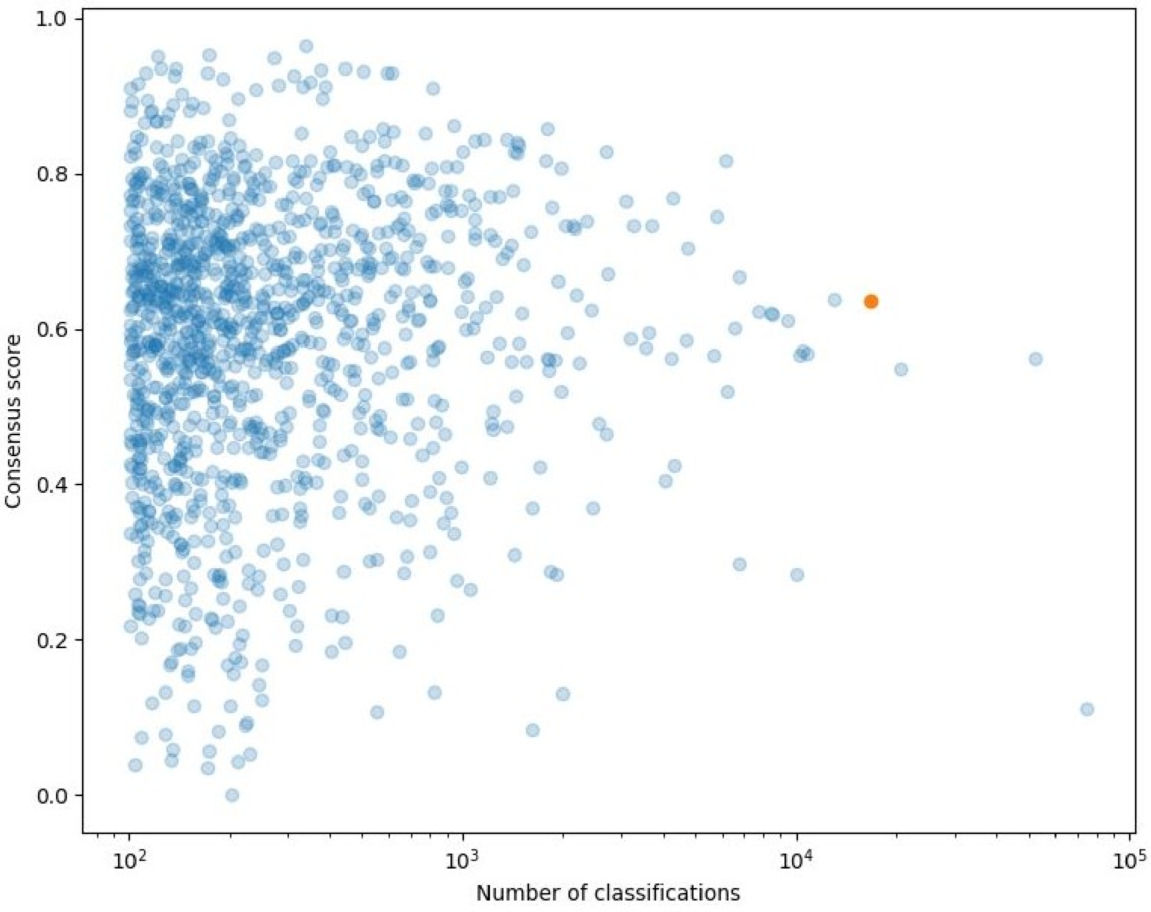



**Fig. 5.** Statistics of the Zooniverse volunteer population. Left: histogram showing the numbers of Zooniverse volunteers who made a certain number of classifications. On a log scale the rough power-law distribution of classification numbers is apparent, with a slope $\approx -1$. Right: histogram of the distribution of optical ID consensus scores for volunteers with more than 100 classifications.

# Consensus vs classification count

- Most optical IDs are majority decisions so hard to achieve a high consensus score even if you are right all the time – unless you refuse to classify anything remotely doubtful
- ‘Experts’ do not have clearly higher consensus scores than citizens
- Some enthusiastic participants were not seeing the same things as everyone else…
- Final ID voting was weighted by a user’s consensus score



# Expert inspection and ridge line code

- Inspect sources that:
  - Have duplicate optical IDs
  - Have ‘too zoomed in’ or blend flags in RGZ or pre-filter
  - Are larger than 1 arcmin
- Around 150,000 inspections by a small team of ‘experts’
- Several different interactive workflows to allow the editing of the PyBDSF source decomposition and optical IDs
- Ridge line code of Barkus+22 then applied to search for optical counterparts missed in visual inspection for sources > 15 arcsec and brighter than 10 mJy

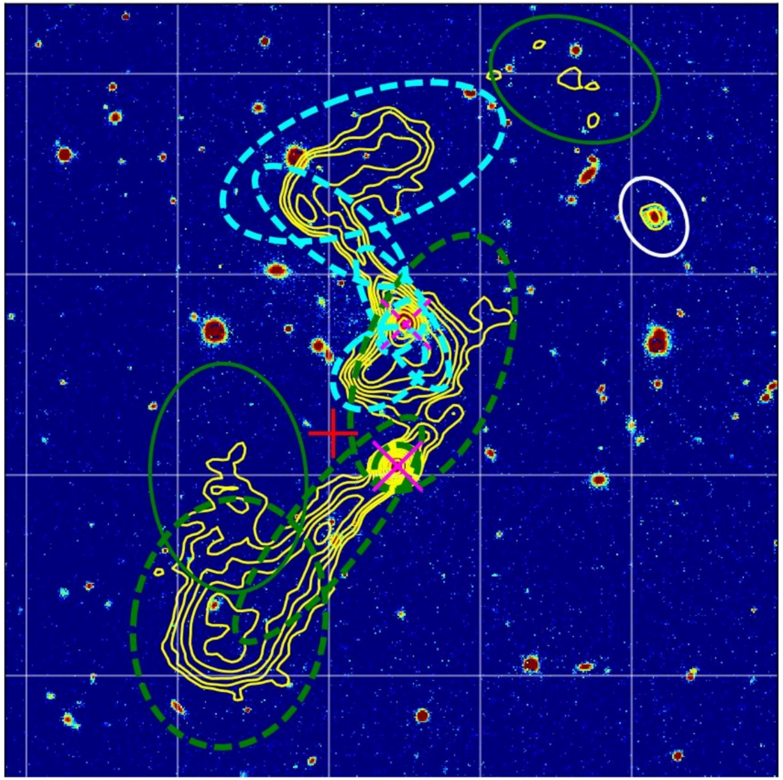
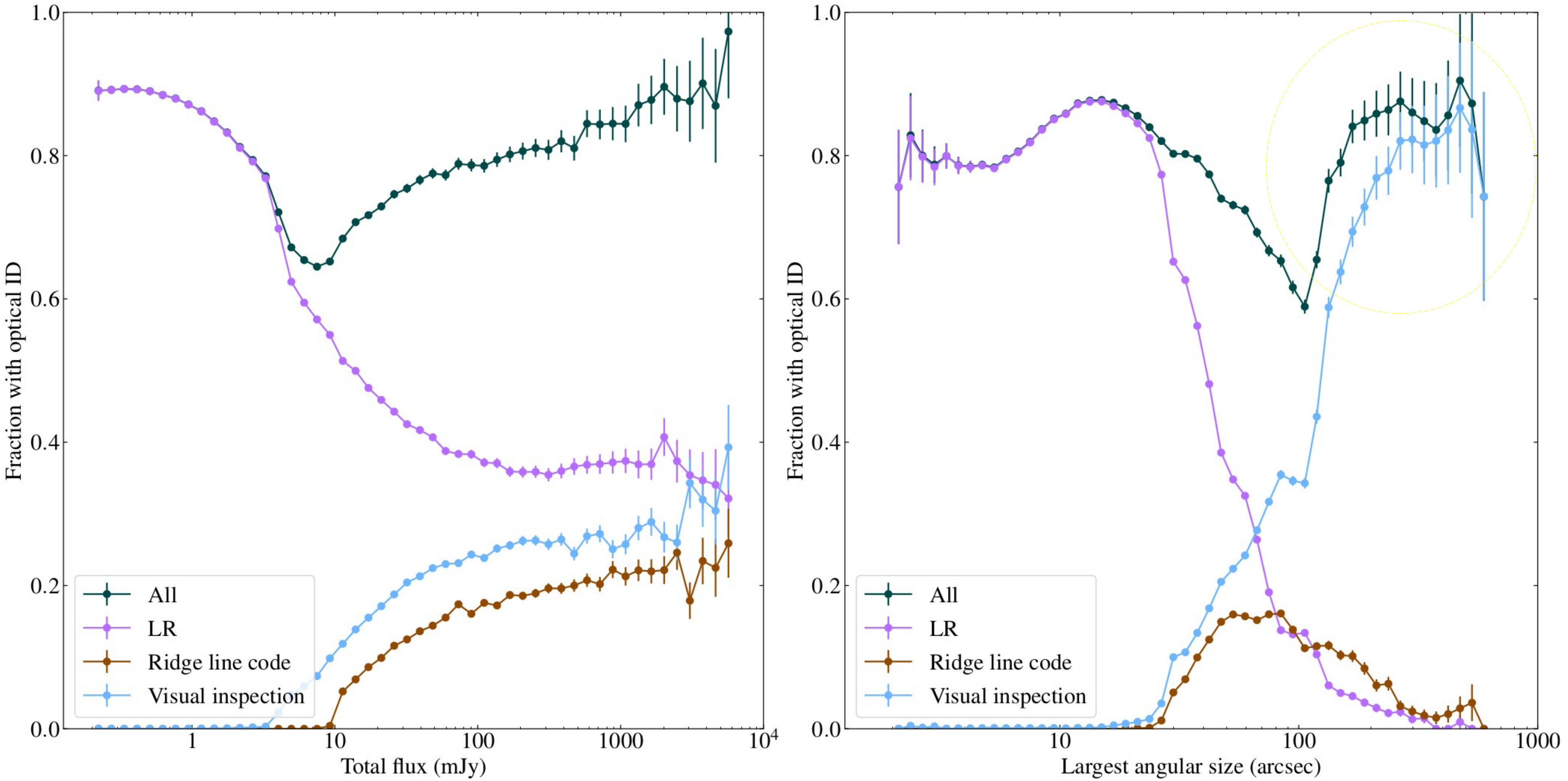
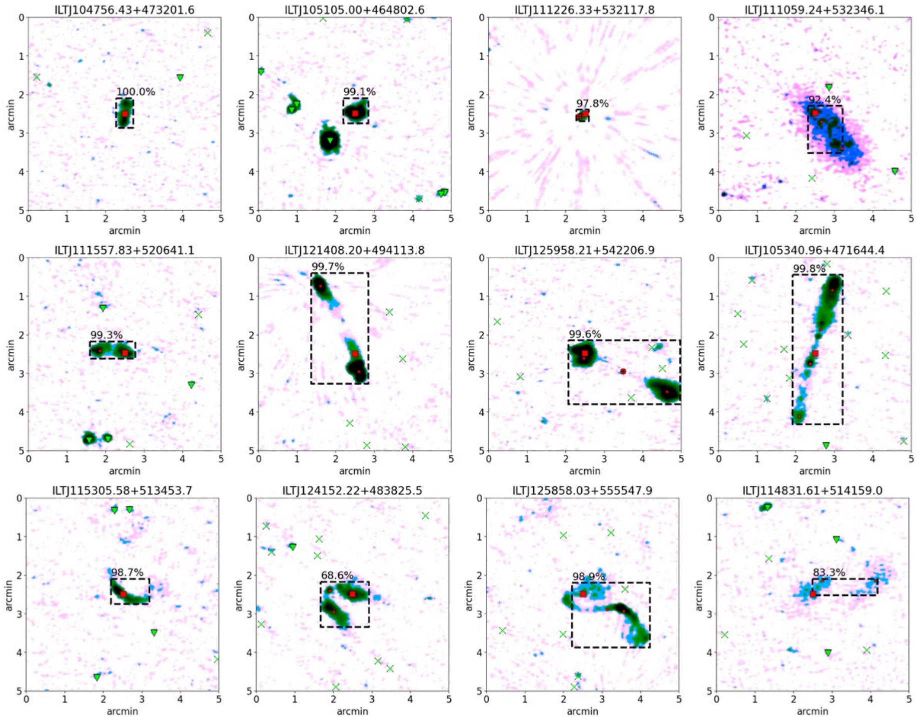


**Fig. 7.** Example user interface for the ‘Second deblend’ workflow. In an interactive Matplotlib window the expert classifier has separated the emission from two extended sources that had been combined in PYBDSF, seen in green and cyan, and has selected optical IDs for both. An unrelated source marked in white has been left unchanged. The new source is a mixture of PYBDSF components (solid lines) and Gaussians (dashed lines).

# DR2 optical ID fraction

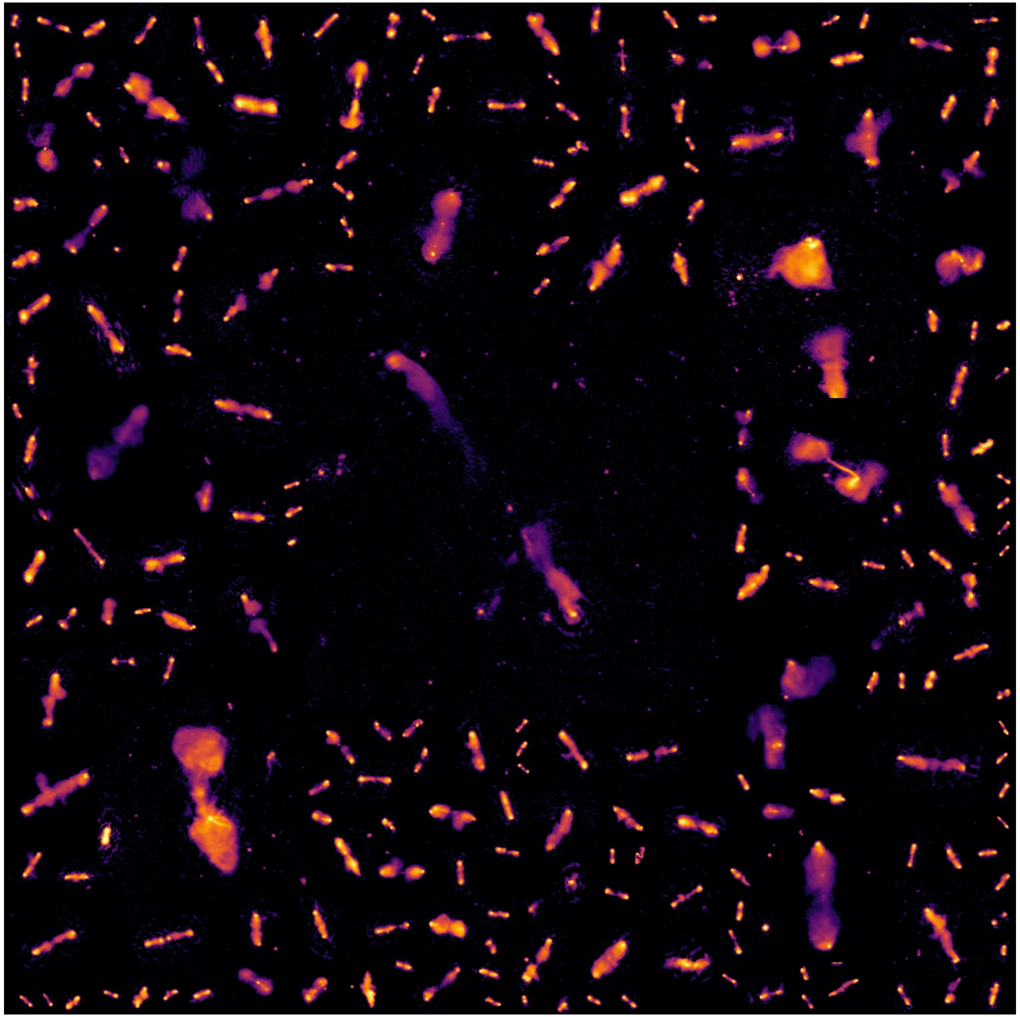


- Task is hard!
- Citizens have very varying degrees of (a) engagement and (b) correctness. These two may not correlate…
- Slow to complete because of (relatively) small engaged volunteer pool and large number of sources
- Optical IDs a particular challenge – choice of Legacy rather than WISE as optical image didn't help
- For DR3 need to **vastly reduce number of images sent to visual inspection**
- Can machine learning save us? Maybe – see Mostert+22, Mostert, Oei+24

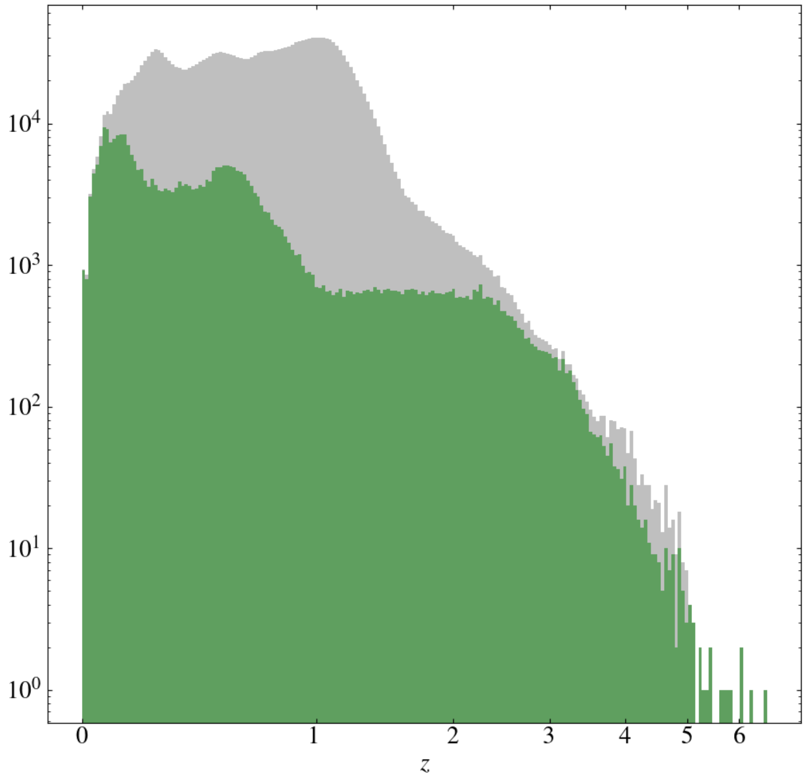


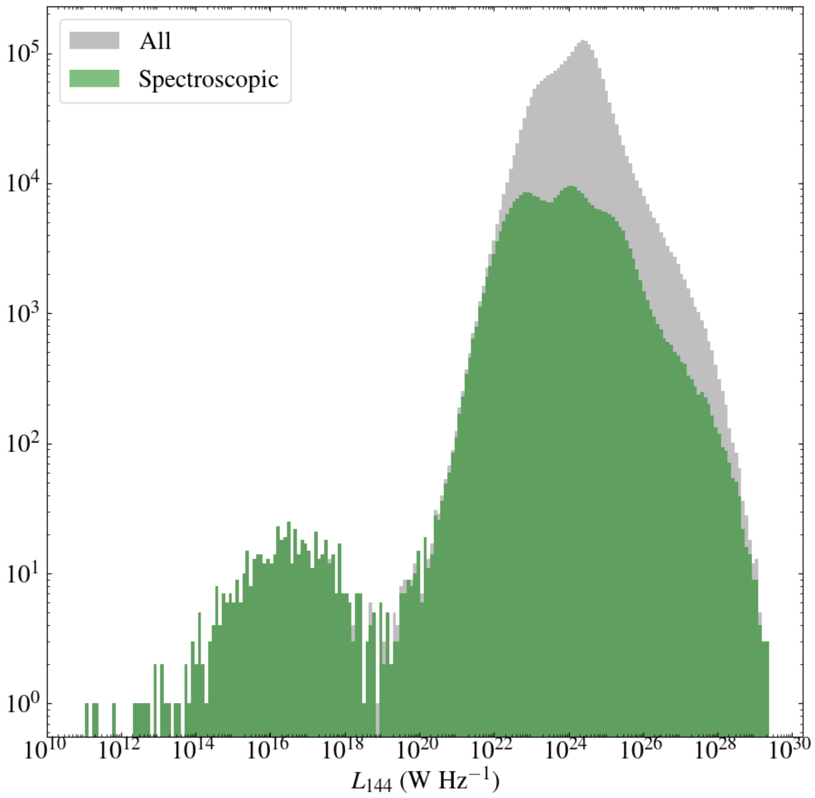
# Output DR2 catalogue

- 4,116,934 sources in DR2/Legacy overlap
- 85% optical/WISE ID fraction
- 58% good redshift fraction (84% of sources with IDs in Legacy catalogue)
- 305,000 spec-z from SDSS and DESI – lots more to come from future DESI releases (2025+) and WEAVE-LOFAR (same??).
- 1,856,041 good mass estimates
- By far the largest optically identified catalogue ever (order of magnitude improvement on DR1)
- Right, some > 2 Jy FRIIs from DR2



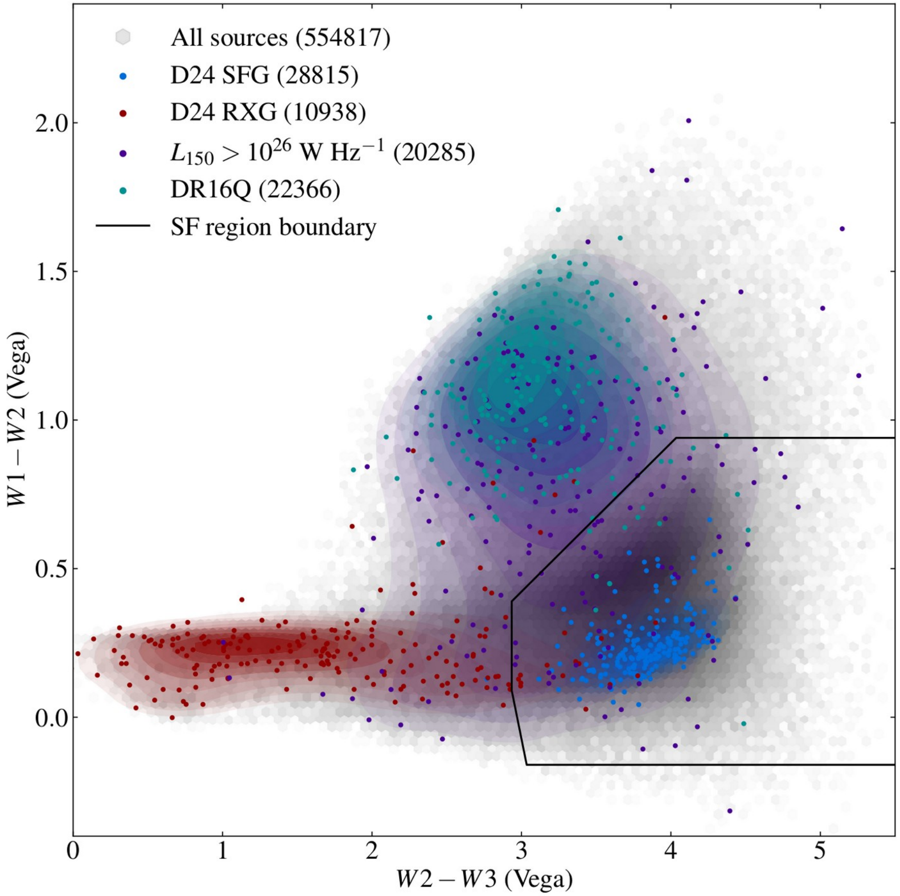
# DR2 $z$ and luminosity distribution



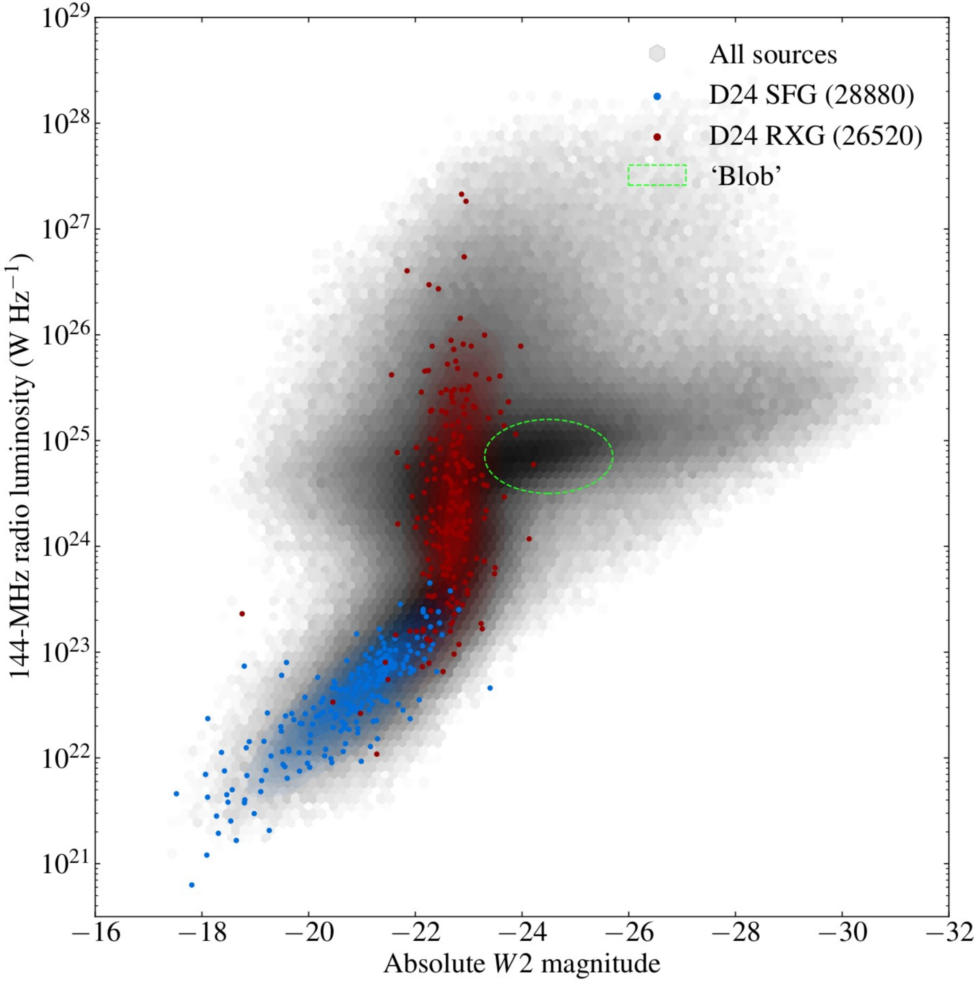



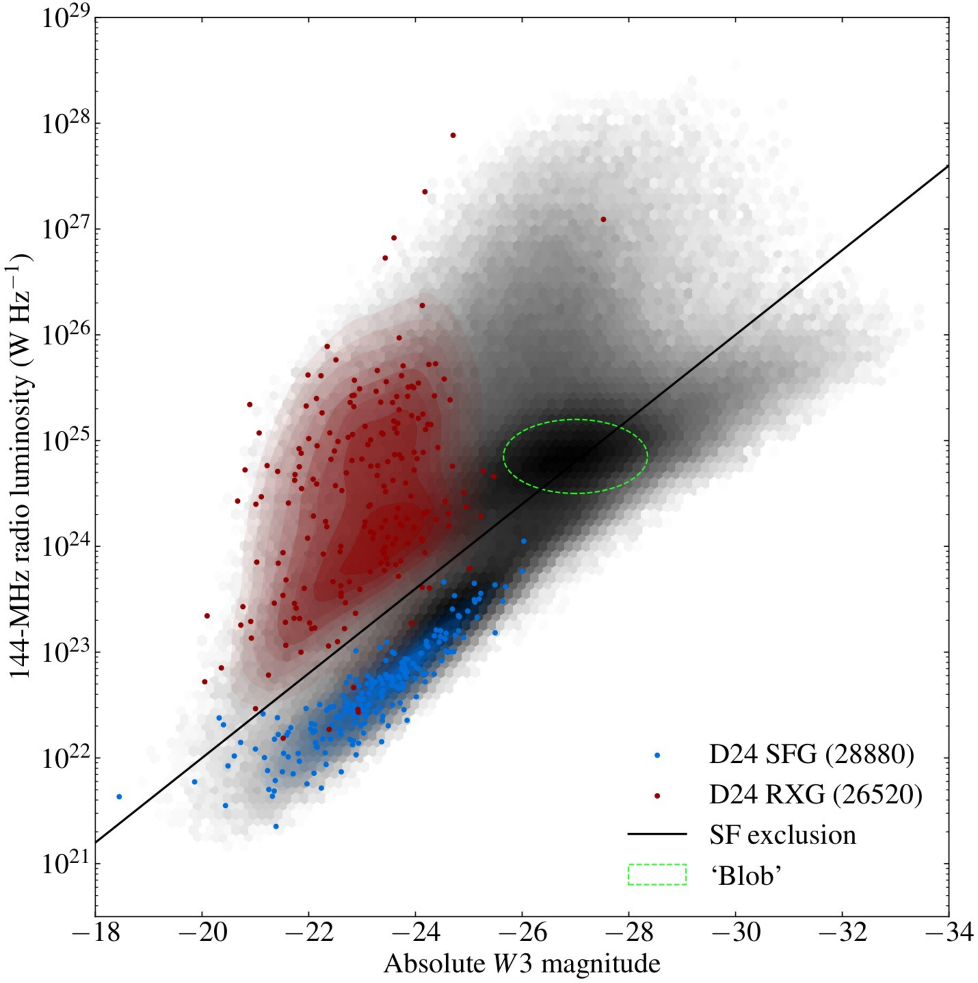
# Emission-line classifications (1)

- Drake+ (in prep: D24) are using BPT-type emission-line diagnostics to classify 150k DR2 sources with SDSS spectra as radio-excess (radio luminosity exceeds expectation from Halpha) and HERG/LERG.
- Classification is consistent with expectations from WISE colour-colour plots
- Bright AGN and HERGs lie in the same location as SFG
- Not safe to select AGN as in Hardcastle+19 with cuts in WISE colour space!
- But W2/W3 vs radio luminosity is more interesting...

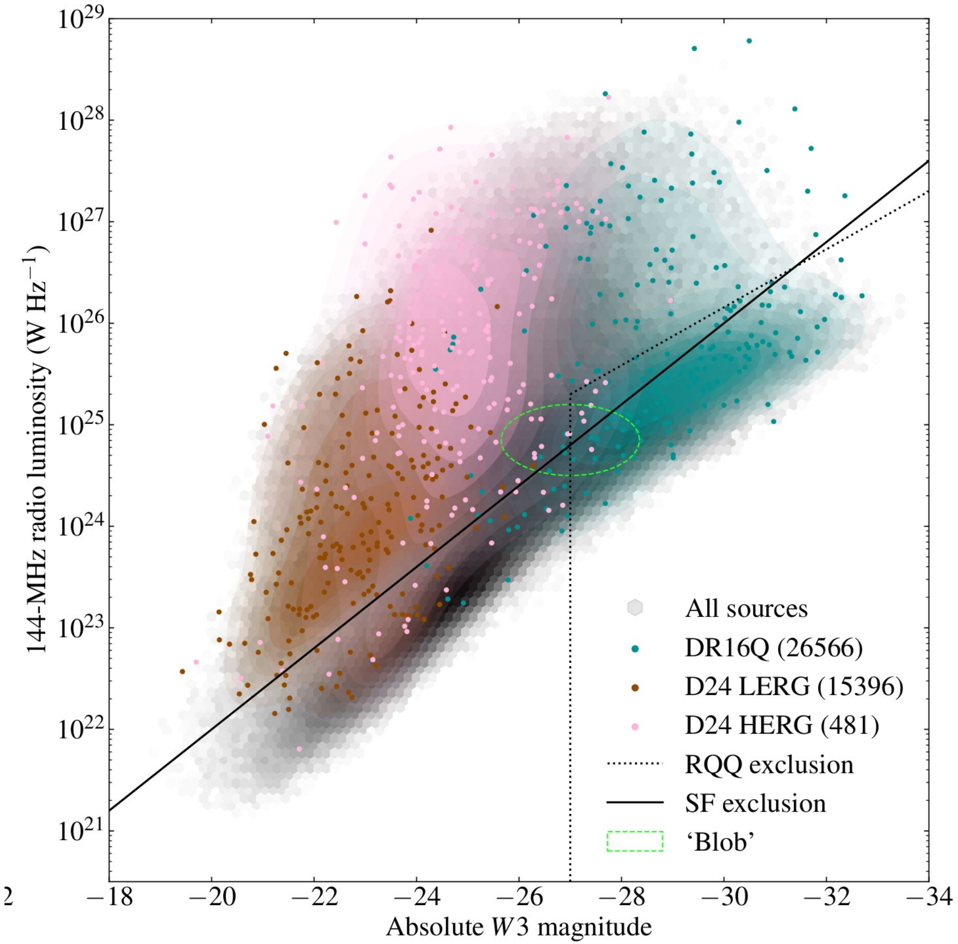


# Emission-line classifications (2)



# AGN selection




- Start from flux-complete sample with S > 1.1 mJy, good z > 0.01 (963,764 objects)
- W3-radio diagram allows us to select different classes of objects (although note that many objects are undetected in W3)
- Quasars and HERGs lie to the right of this plot (W3 luminosity from the torus – cf Gurkan et al 2014)
- SDSS DR16 quasars populate the far right including a clear 'RQQ' branch
- Not clear whether these should or should not be excluded! But we can remove them for consistency with Hardcastle+19
- Net result is a plot with two cuts in this space.
- Can decide whether a W3 detection is required or use limits – overall around 600,000 AGN.

# Masses: AGN host galaxy mass relation to luminosity and size

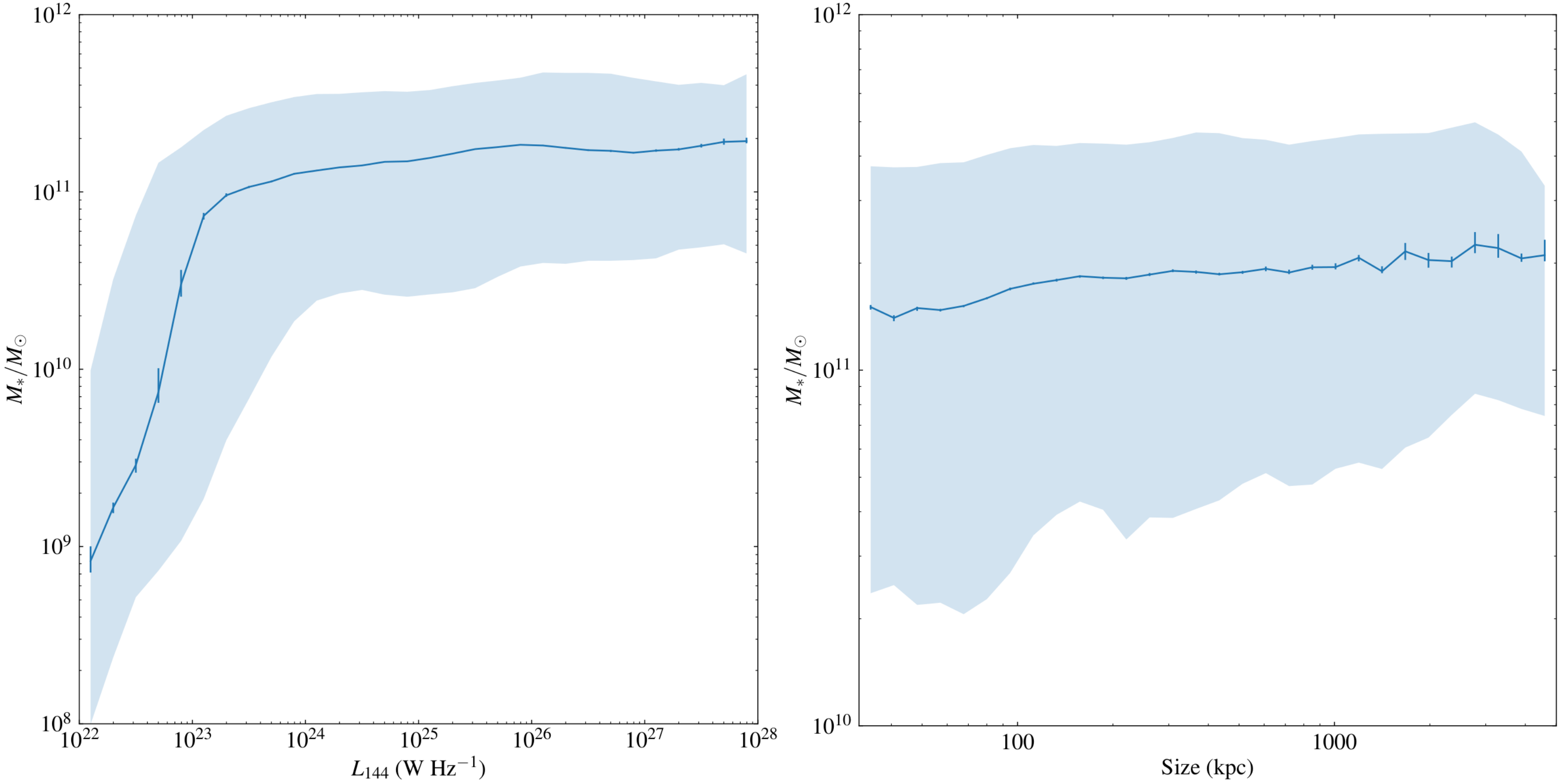
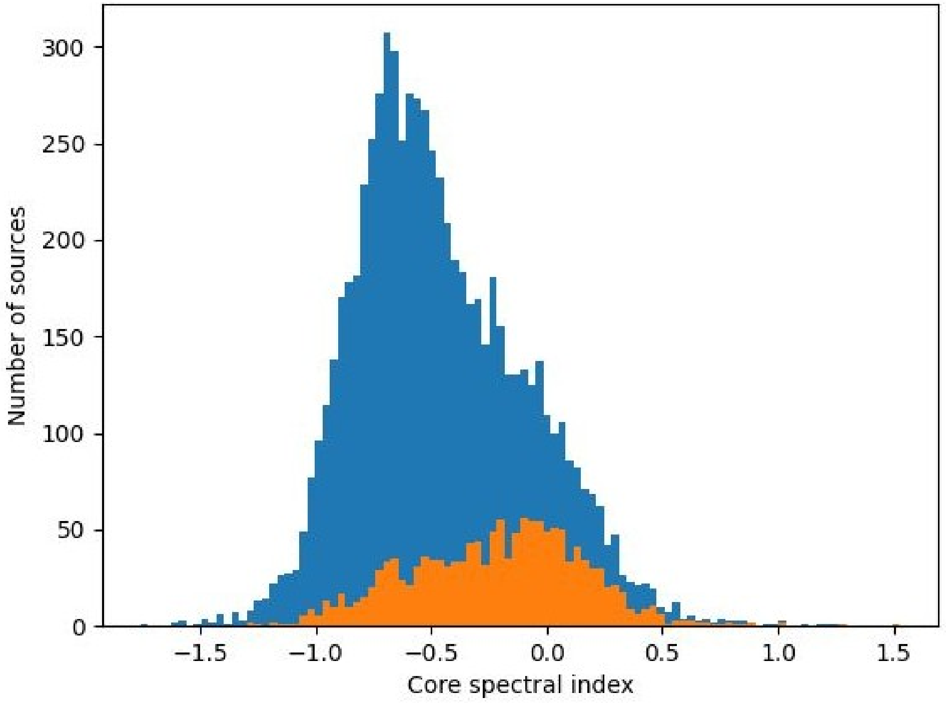



**Figure 8.** Host galaxy stellar mass estimates as a function of radio luminosity (left) and source linear size (right). Only sources with `flag_mass==True` are plotted in both figures, and a redshift cut $z < 1.2$ is imposed. Both figures show binned median masses together with their $1\sigma$ bootstrap uncertainty (line and error bars) together with the 5-95 percentile range of the mass estimates to give a sense of the breadth of the distribution (shaded area). In the right-hand figure only sources with $L_{144} > 10^{25}$ W Hz$^{-1}$ are shown. Note the different scales on the $y$-axes of the two plots.

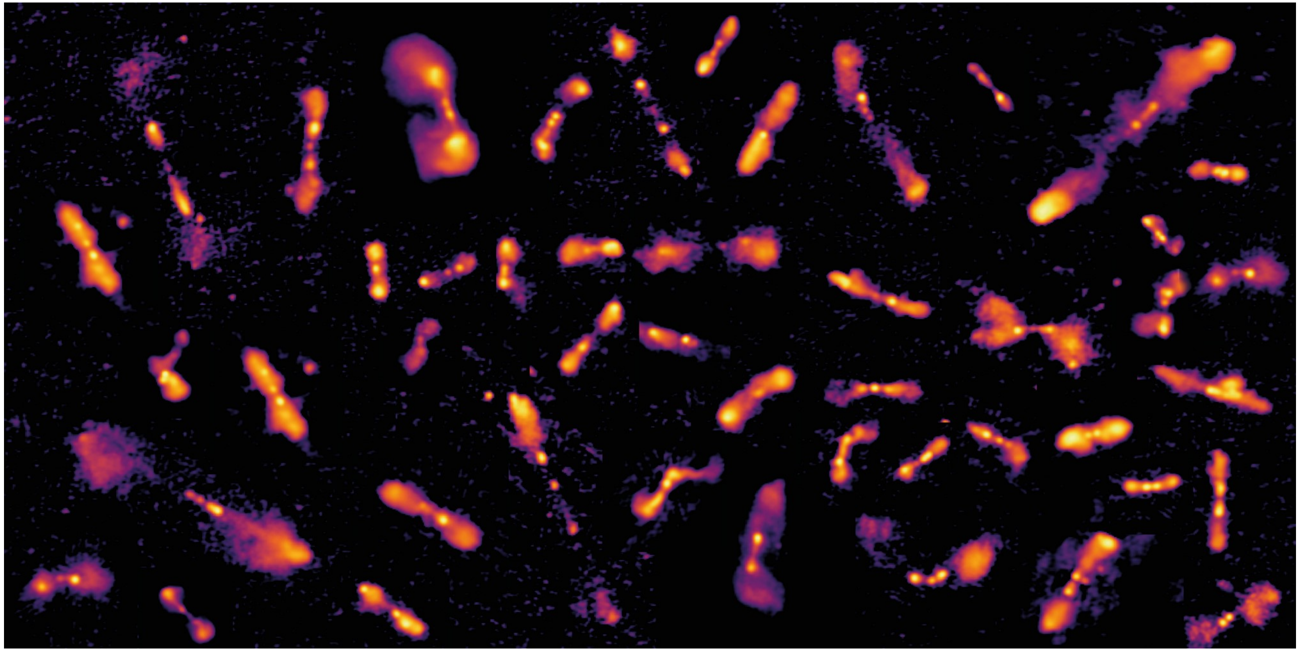
# Cores: match with VLASS

- Can start to look at the radio core properties of extended AGN
- Some interesting results here with samples of tens of thousands of resolved AGN
- Care needed as heavily biased to steep-spectrum cores
- Can use VLASS to investigate variability of central AGN in extended sources



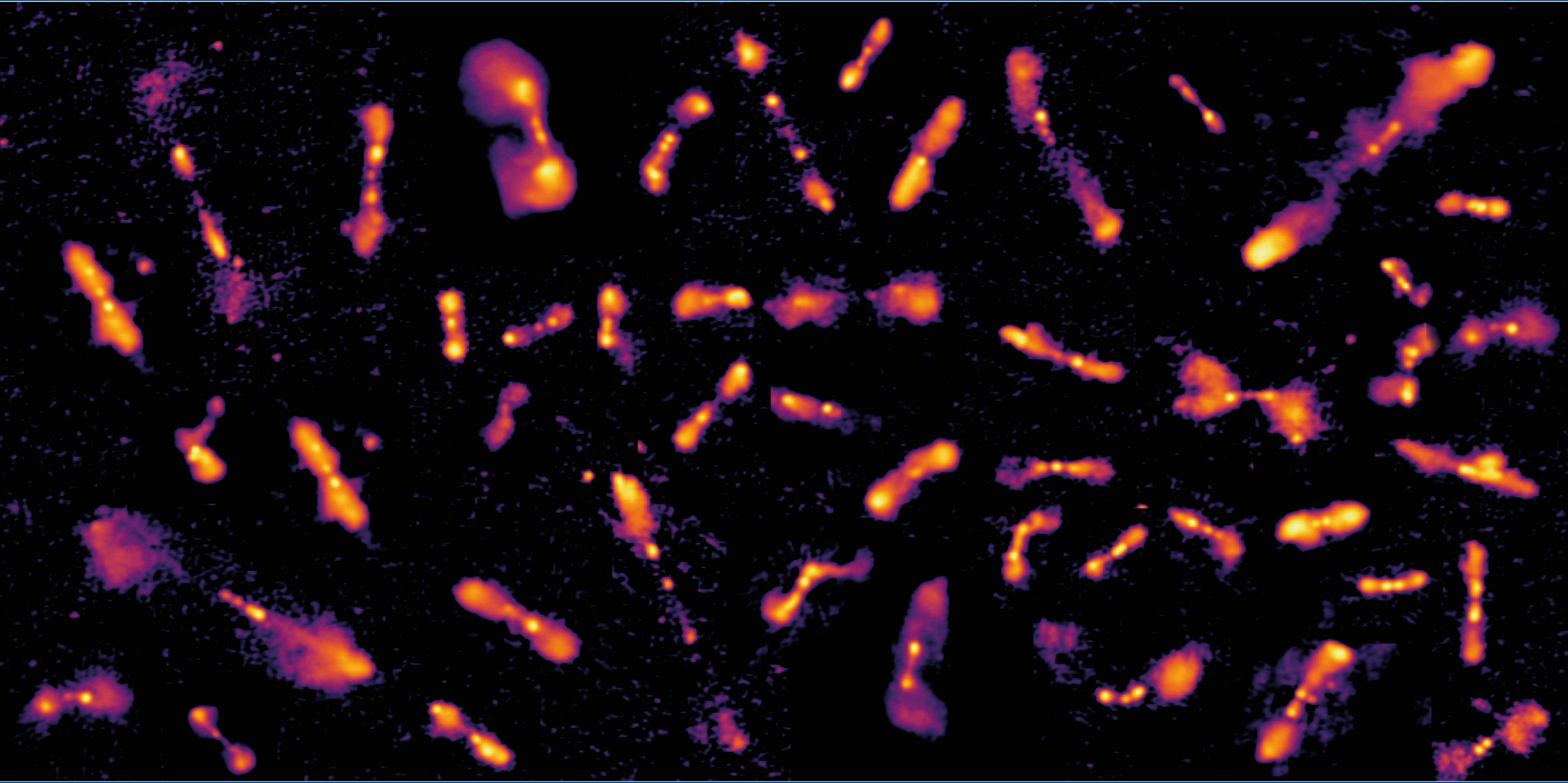
# Restarting radio galaxies

- If jets switch on and off again then we will see pairs of inner and outer lobes ('double-double' radio sources)
- No evidence that sources that do this are different from the general population (Mahatma+19)
- Implies that brief (few Myr) interruptions to jet energy supply *with return to a similar level* are normal
- Important selection effects
- Fraction is low (few %) because double-double phase can only be seen for a short time.

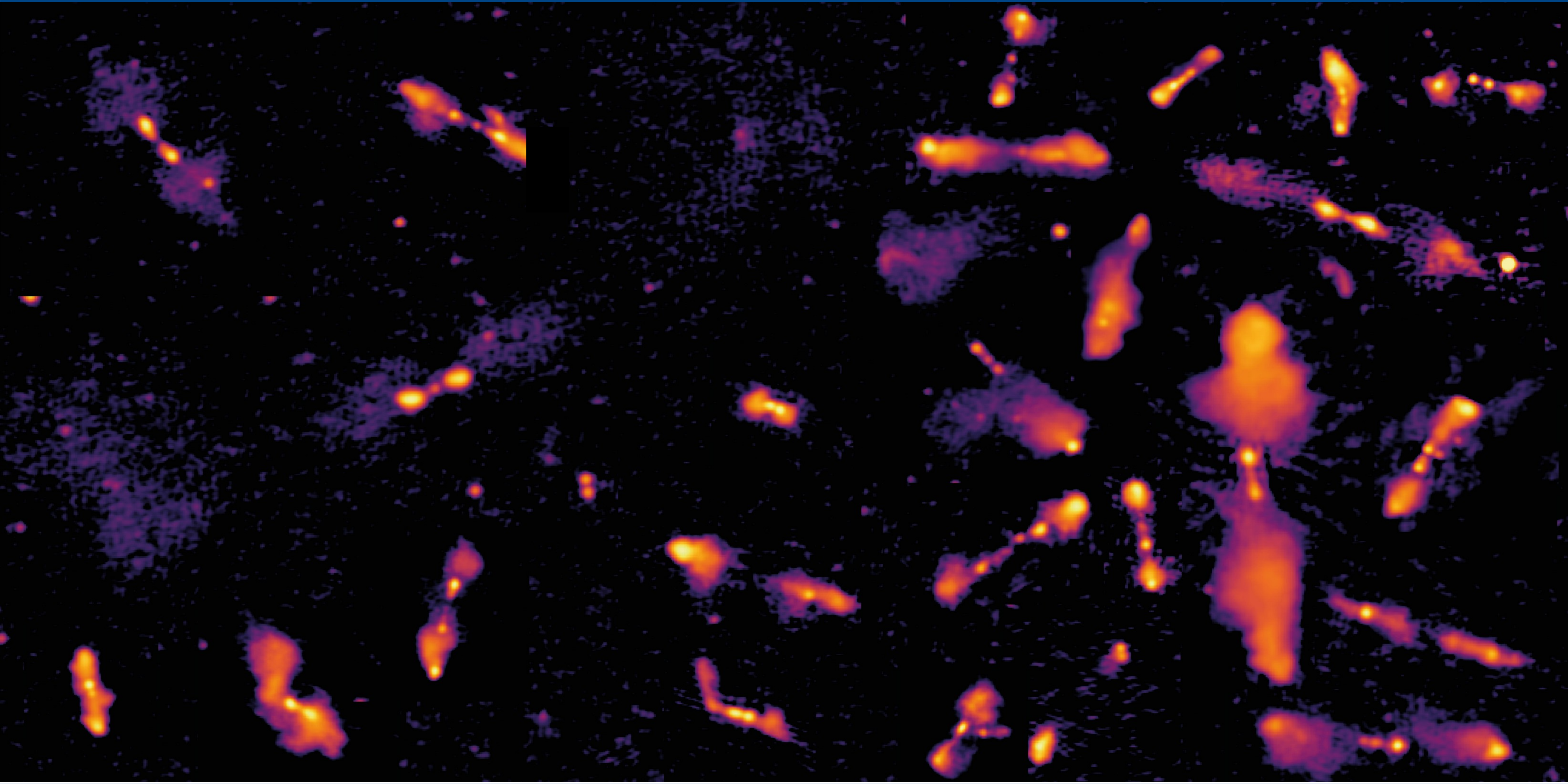


Right: candidate restarting sources from Horton+ in prep.

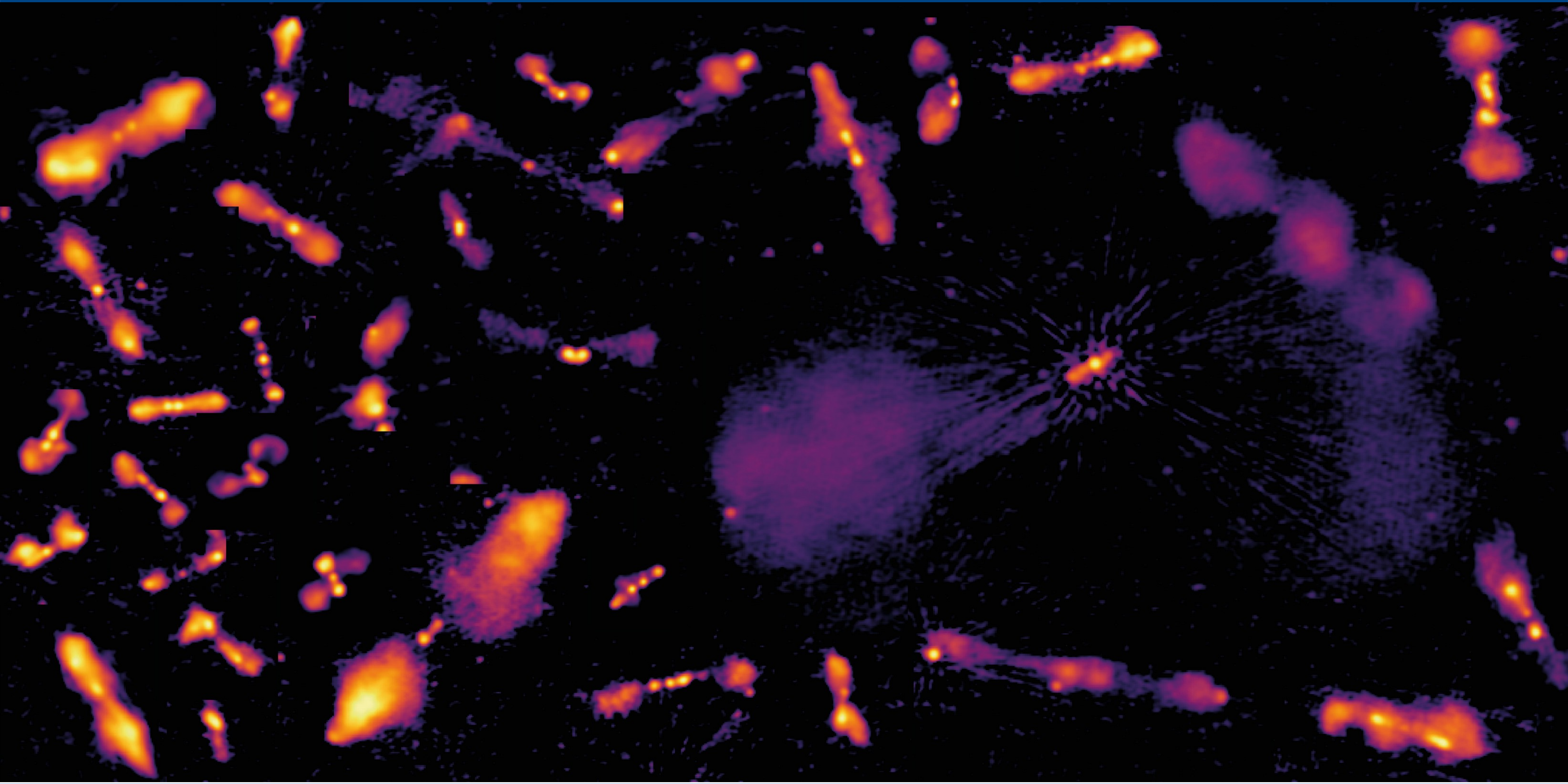
# Restarting radio galaxies



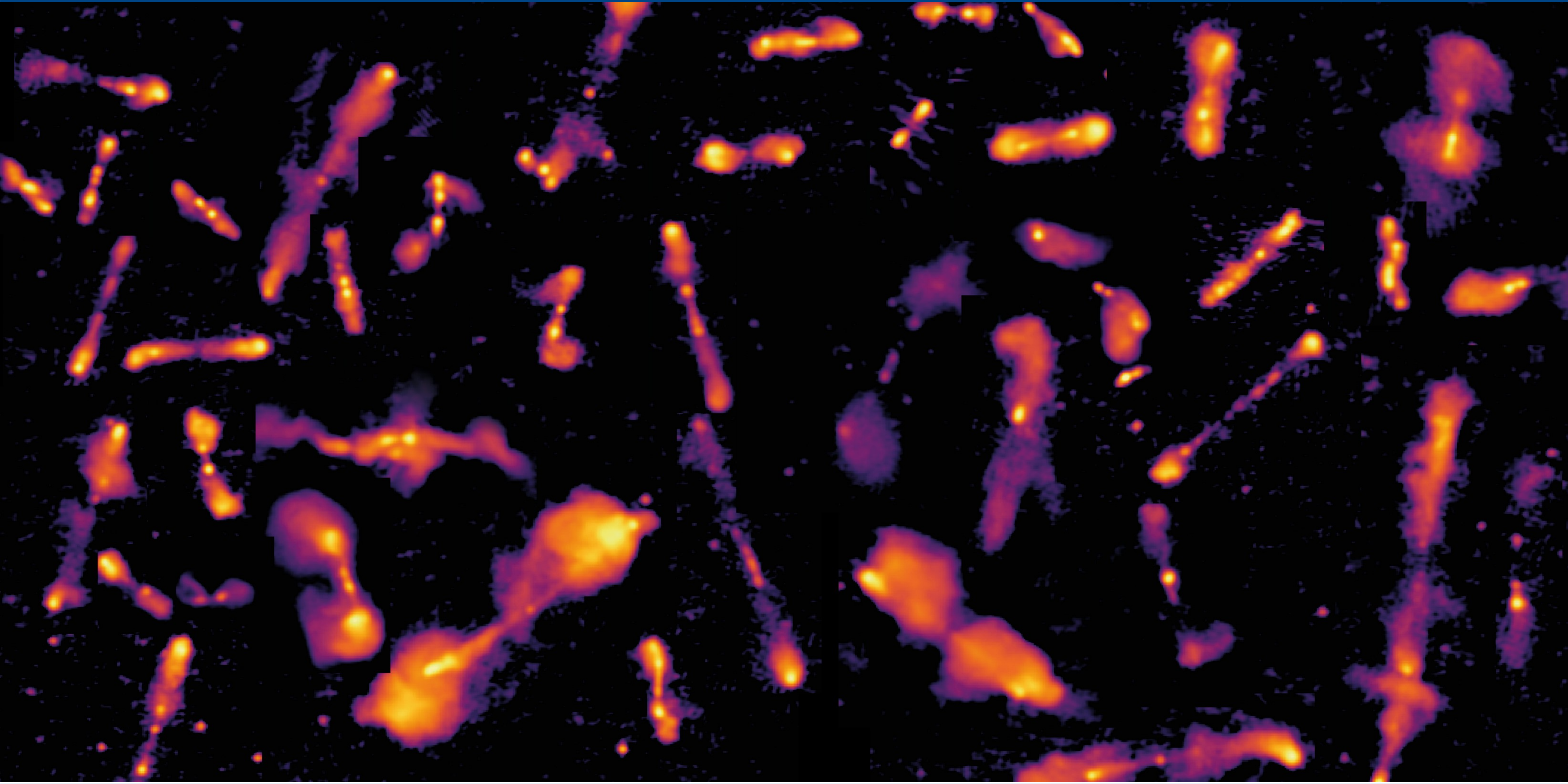
# Restarting radio galaxies



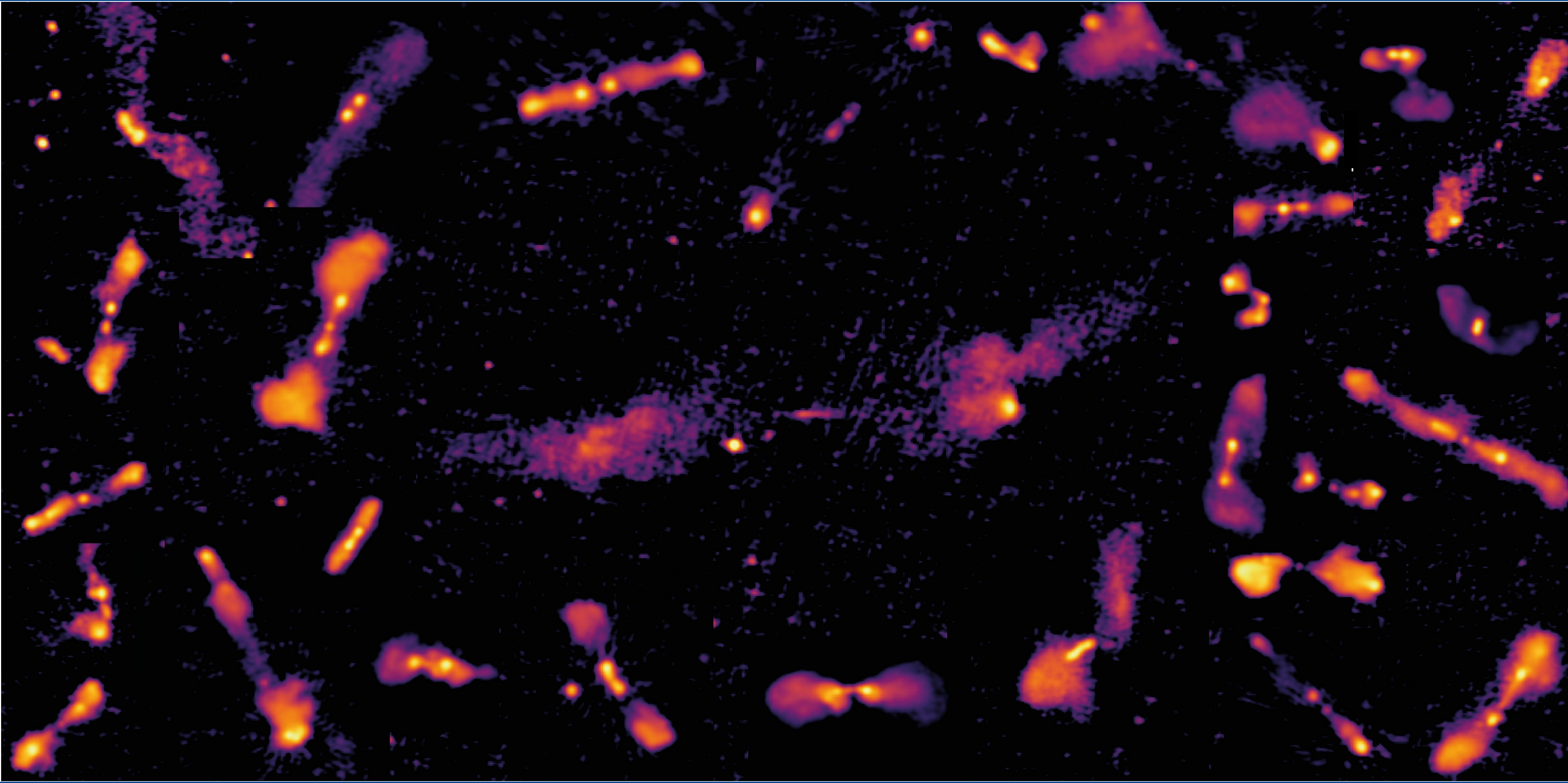
# Restarting radio galaxies



# Restarting radio galaxies



# Restarting radio galaxies



# Summary

- DR2 optical ID catalogue exists: see Hardcastle+23 and https://lofar-surveys.org/
- Probably reasonably complete for bright radio AGN below z=1 and massive SFG with z<0.3
- We plan to keep updating the DR2 catalogue through and past the DR3 LoTSS release
- Improved method of AGN selection currently using WISE data – WEAVE will enable much more by providing emission lines
- ~ 600,000 sources should be enough for a lot of robust statistical analysis!
- Work to follow on:
  - Morphology and relation to emission-line class
  - Jet power inference as a function of environment