

From Radio Sources to Optical Counterparts: LoTSS DR3 and Euclid DR1

LOFAR Annual Meeting June 2026

Bonny Barkus, Lara Alegre, Martin Hardcastle, Johannes Allotey

Background

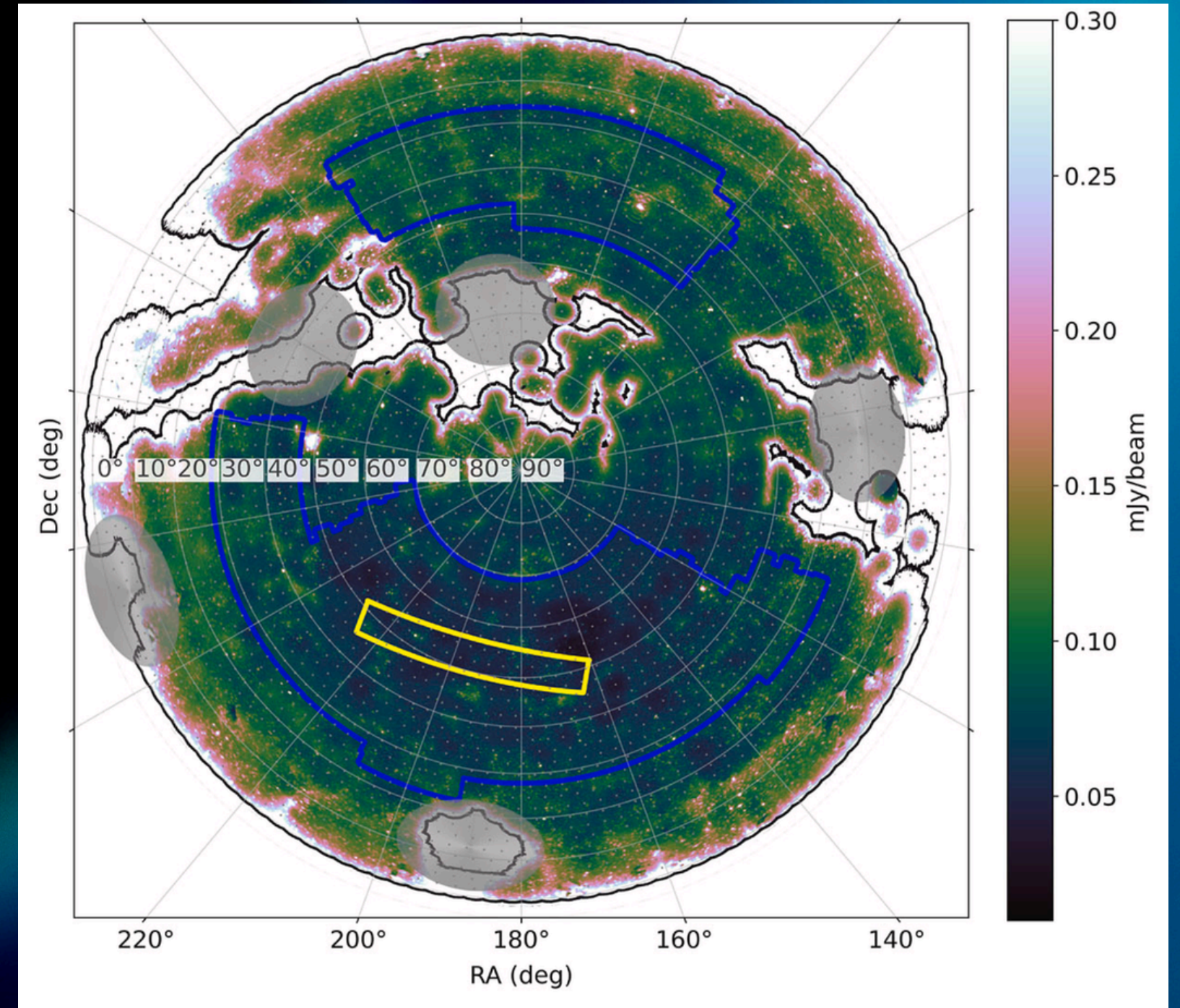
- One of the **biggest challenges** large radio surveys like LoTSS
- **Cross-matching** is matching **infrared(IR)/optical** sources to their **radio counterparts**
- Important:
 - Properties - **redshift, size, luminosity, environment**
- Why? Better **science**, comprehensive catalogues, overall view of the Universe
- Challenging data-wise because **huge surveys** and **large datasets**



Image generated by ChatGPT for fun!

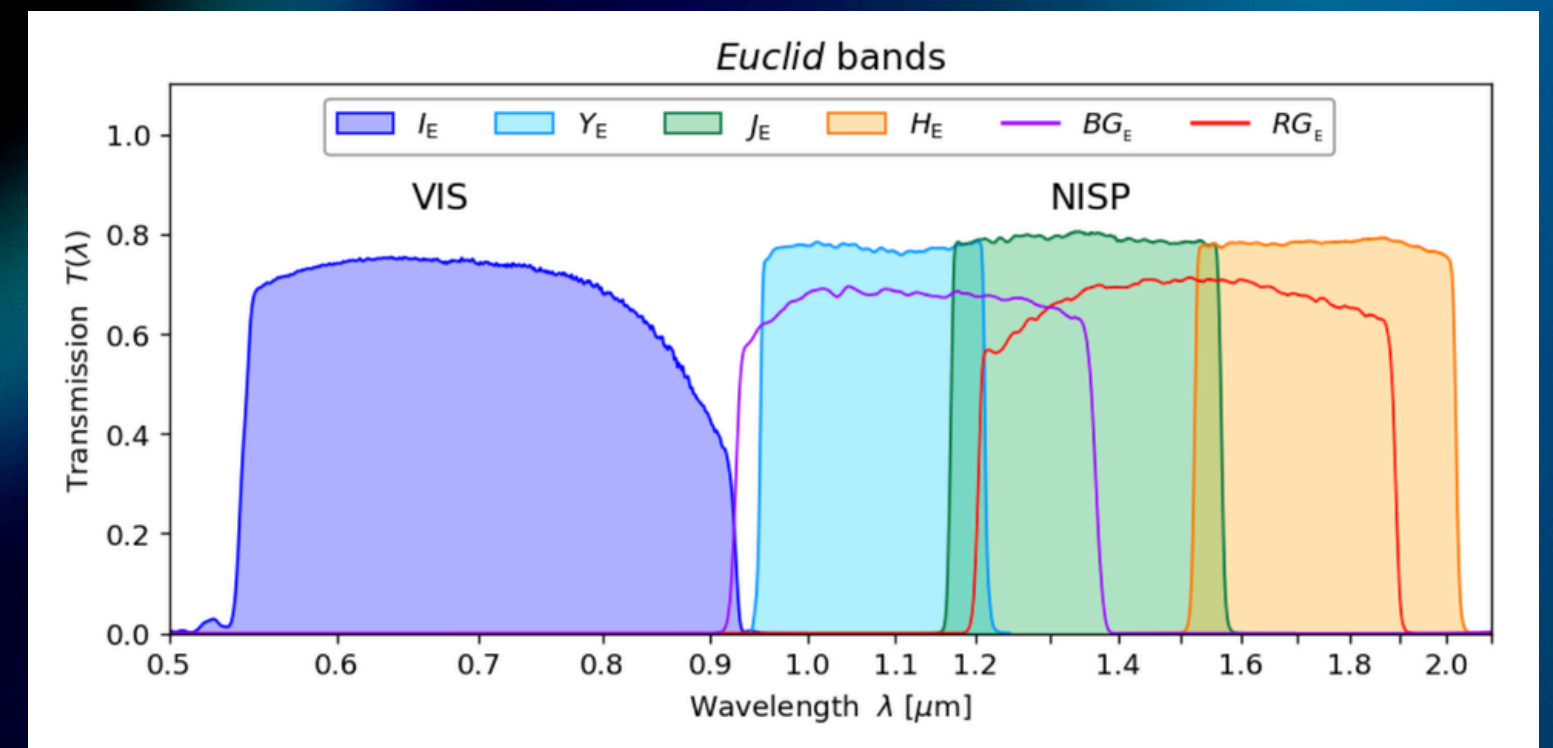
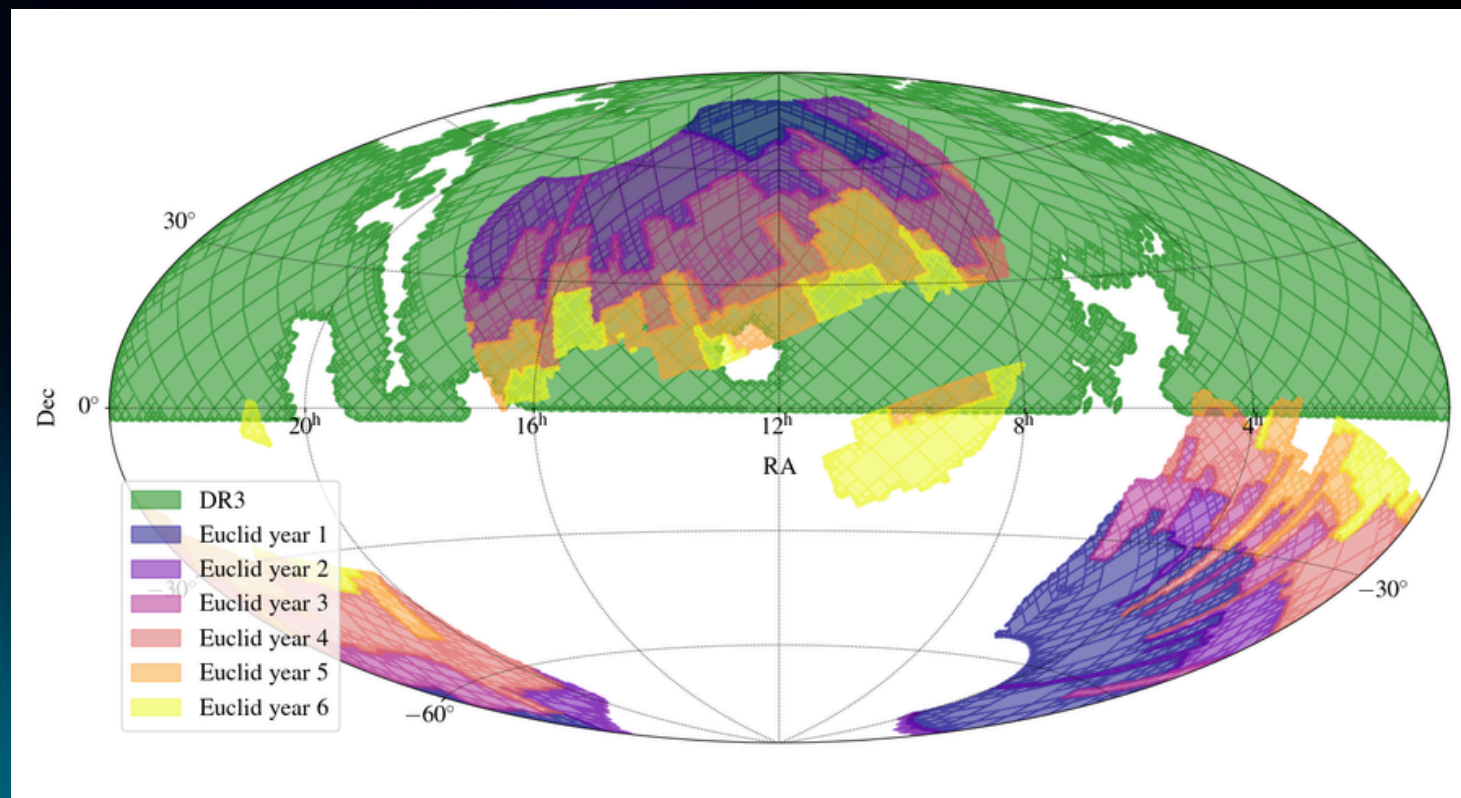
LOFAR LoTSS DR3

- See Martin Hardcastle's talk for all the details!
- Wide area continuum survey centred at **144 MHz**, with a resolution of **6 arcseconds**
- DR3 covers **88%** of the **northern sky** and has over **13.6 million** radio sources in it
- Also had to do all the cross matching for this already so we have methods we can adapt



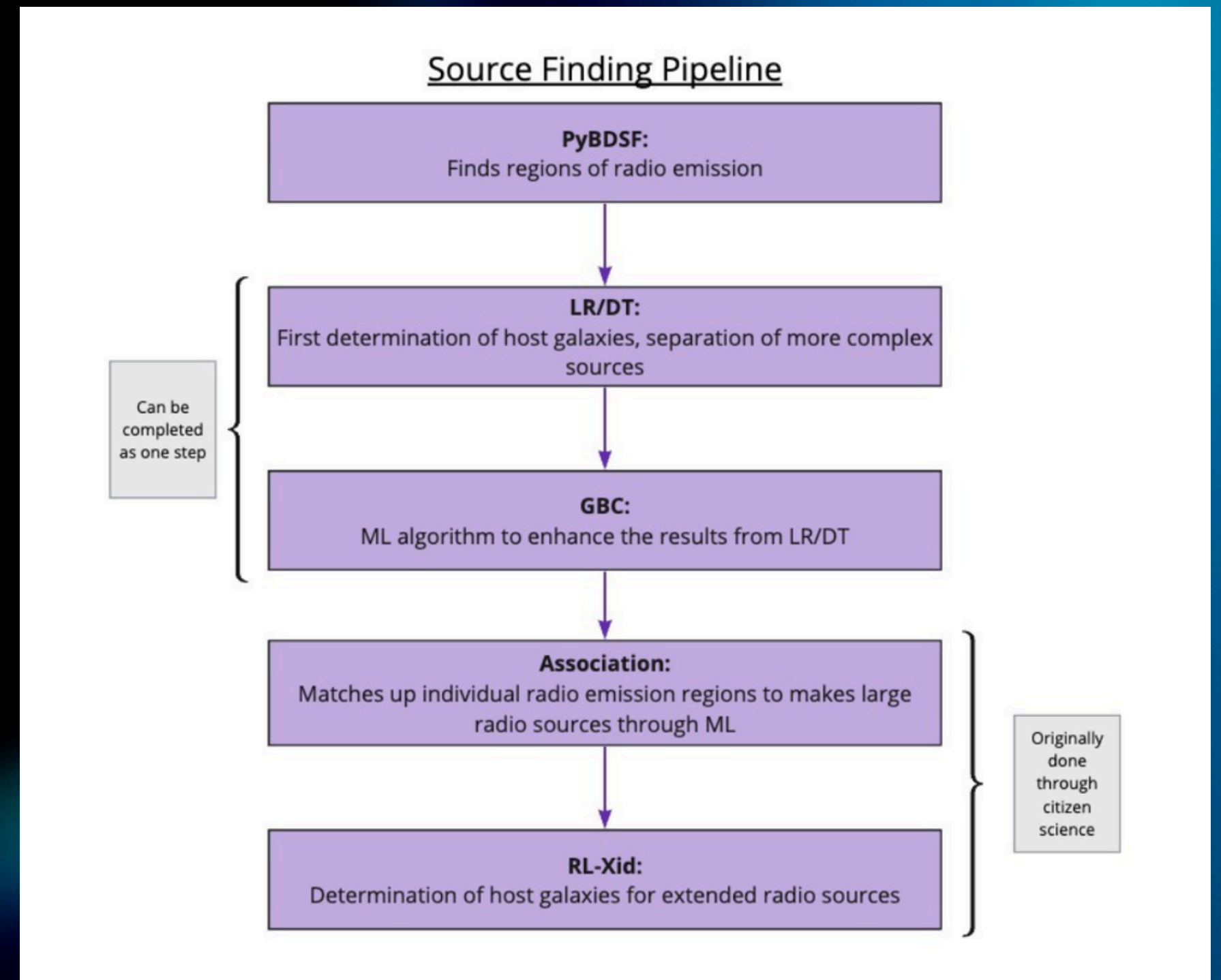
Euclid Wide Survey

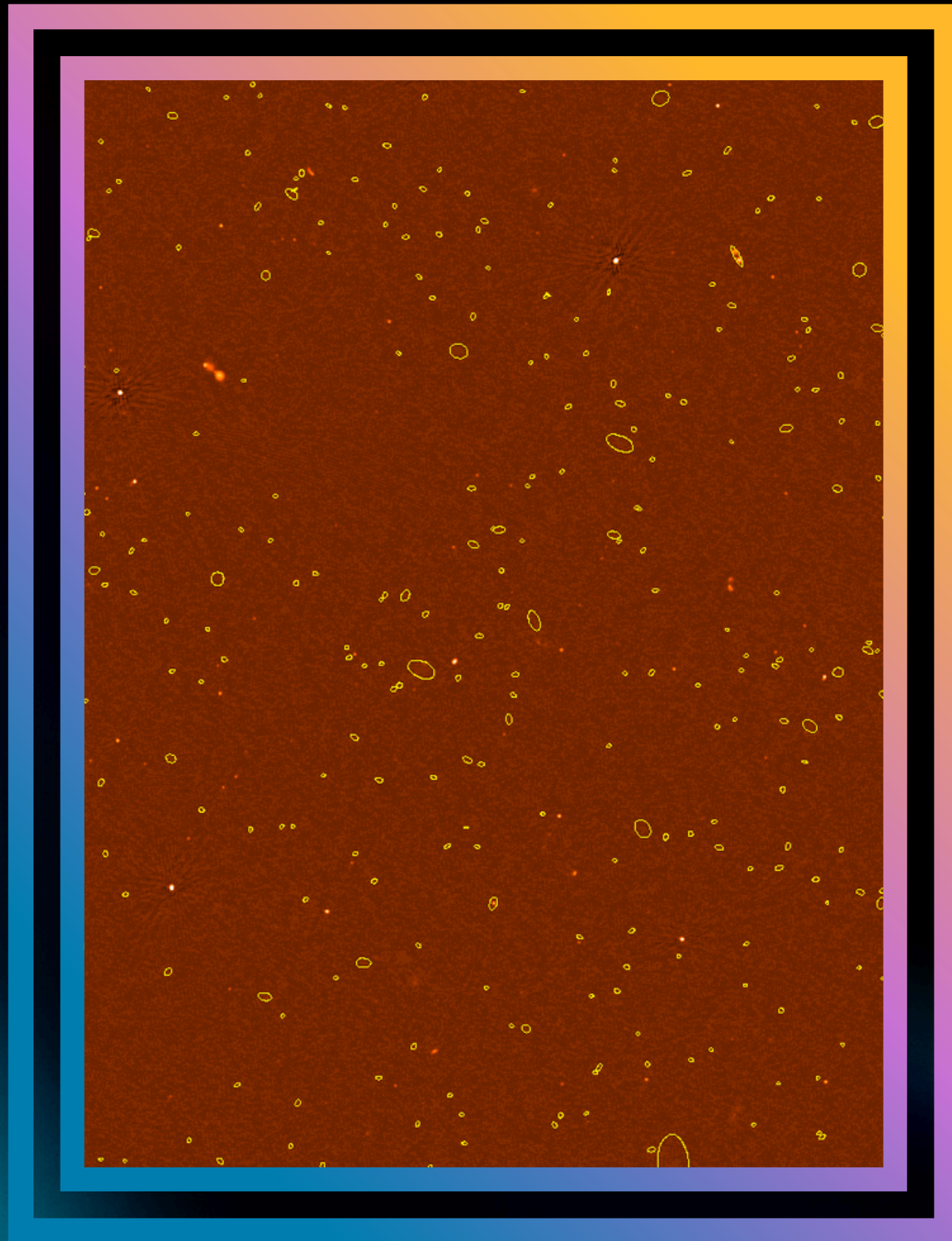
- ESA's Euclid mission: to better understand the nature of dark energy through observations of galaxy clustering and weak gravitational lensing
- **Euclid Wide Survey** (EWS) is the all sky survey
- EWS will cover approximately **13,000 square degrees** of the sky, over 6 years
- We are using DR1 - Year 1 data - northern hemisphere
- **VIS** instrument for covering the visible band (**I**) over 550 to 900 nm
- NISP instrument for the near infrared (**NIR**) bands (**Y, J, and H**) centred at 1080.9 nm, 1367.3 nm, and 1771.4 nm (respectively)



Step by Step

- Previous methods: **Statistical/ML** and **citizen science**
- Automation!
- Five main steps:
 - **Source finding**
 - Likelihood ratio (**LR**)
 - Gradient Boosting Classifier (**GBC**)
 - Association (**citizen science**)
 - Extended host identification (**RL-Xid** + **citizen science**)
- **Flowchart** refines the sources sent to each step of the process





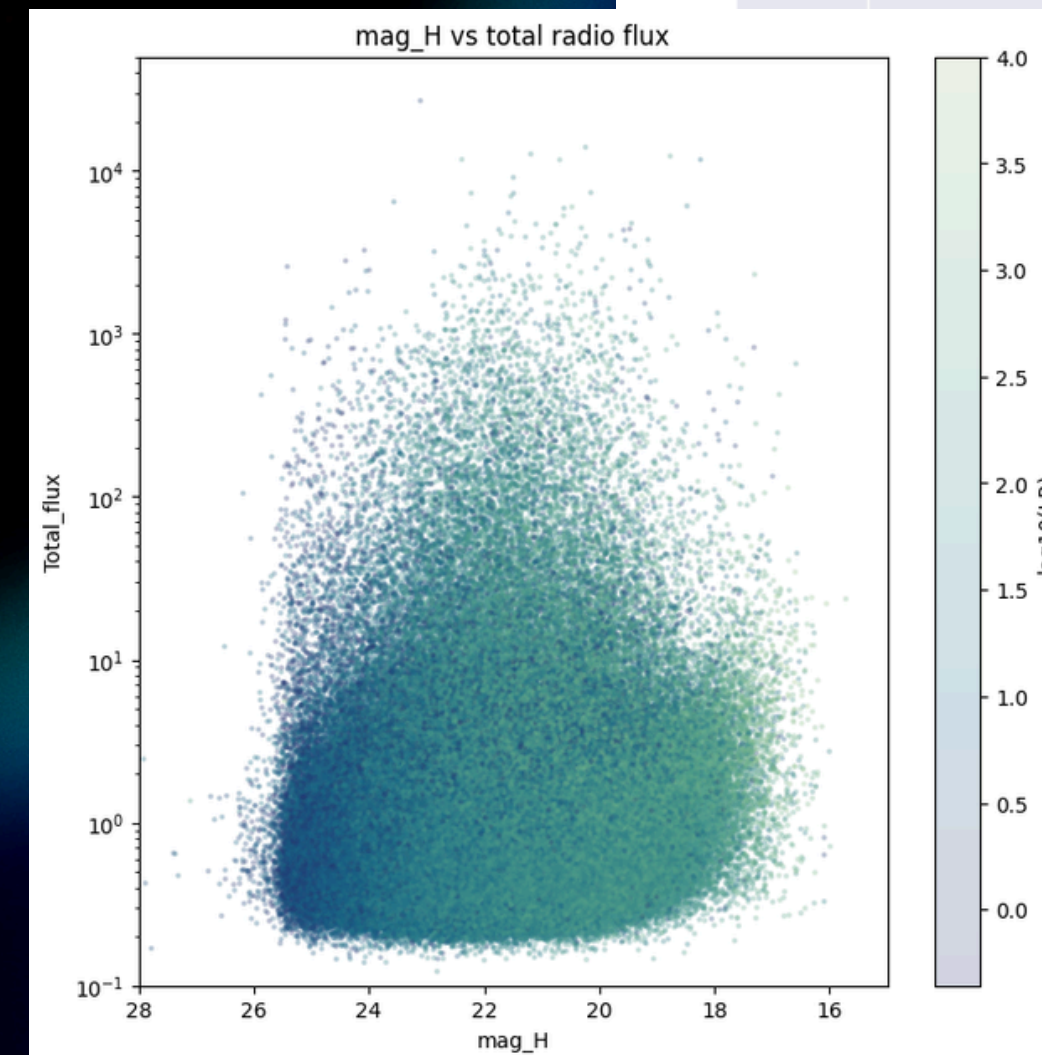
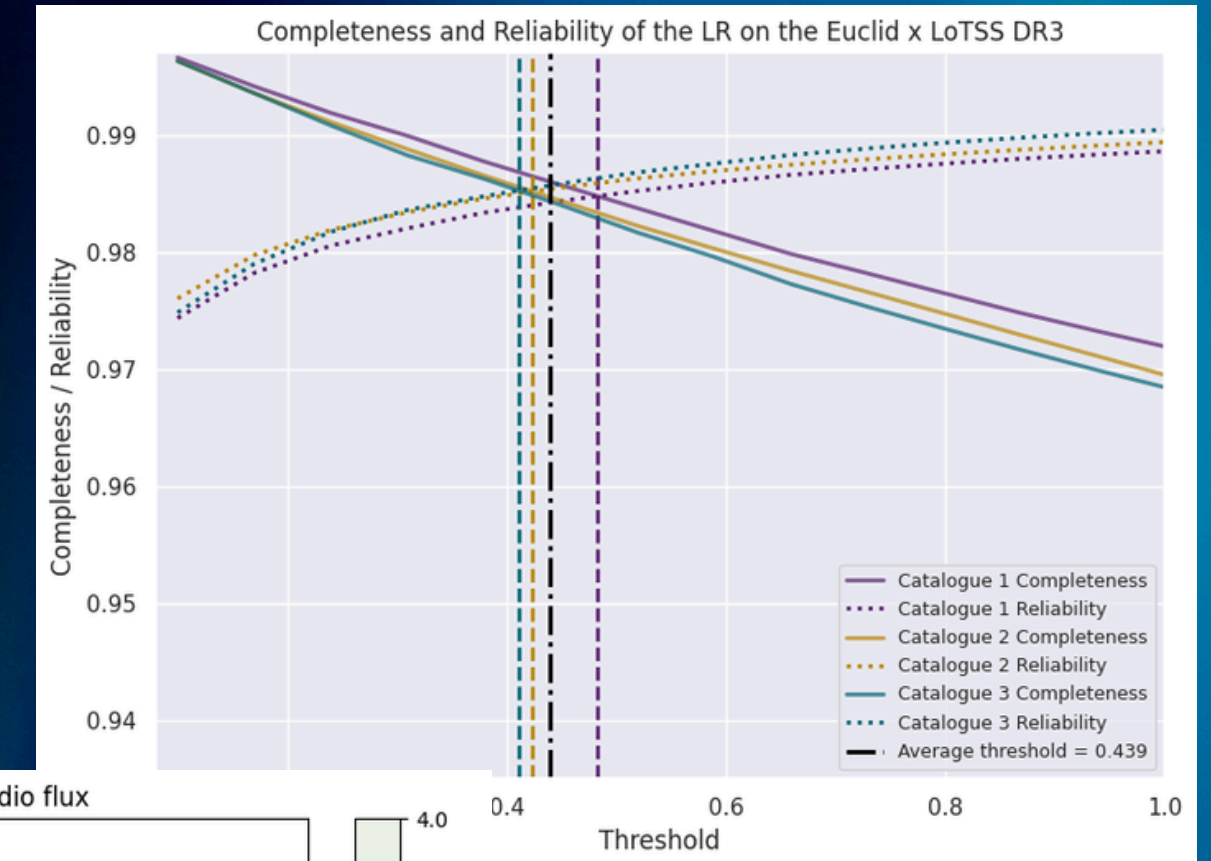
Source Finding

PyBDSF

- PyBDSF finds the radio sources in the radio images
- Need to know where the **radio** is to **match** to the **multiwave** data
- DR3: Batch run on healpix areas
 - **resource** management
 - **reproducibility** of outcomes
- Sky area batching carried through to following steps

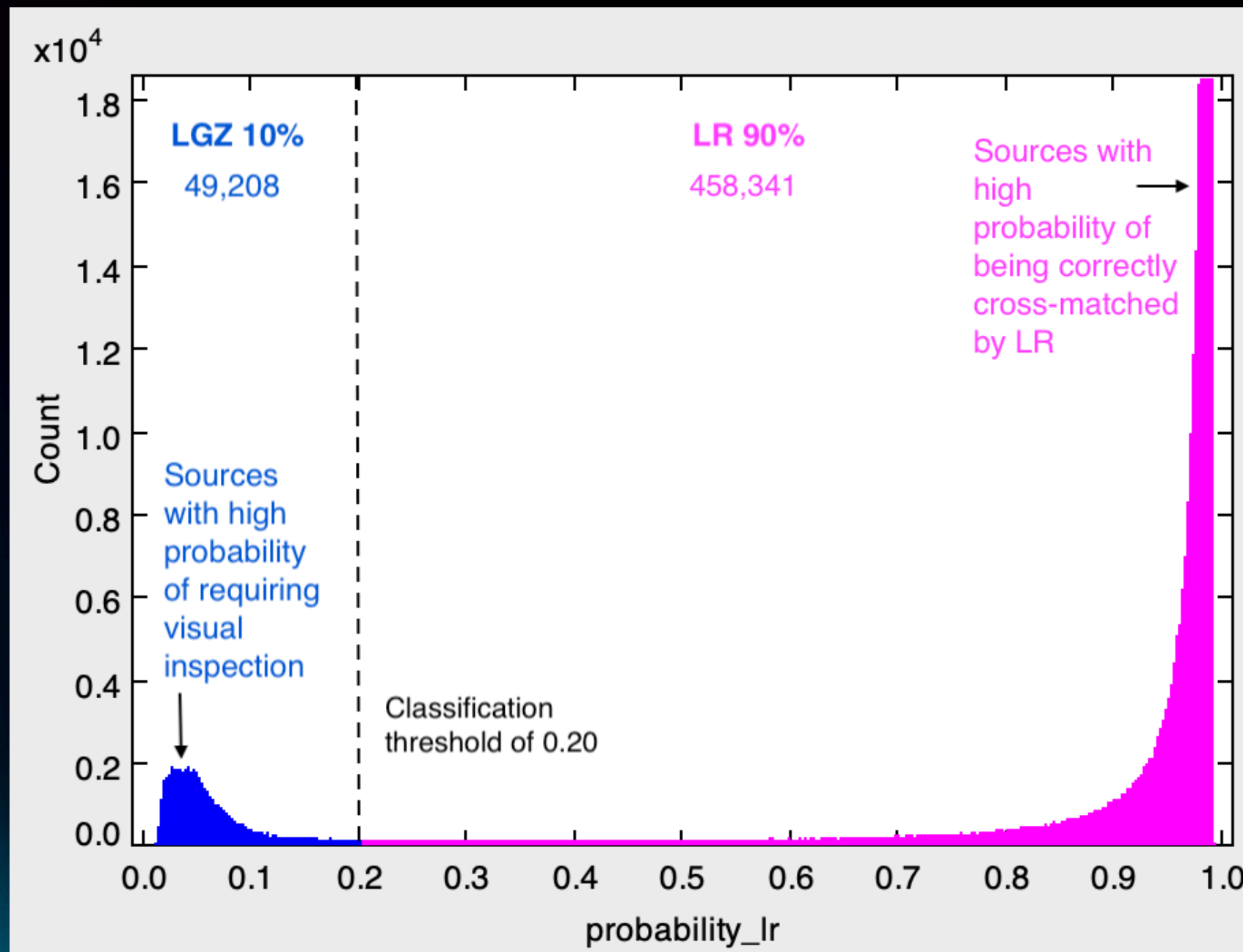
Likelihood Ratio

- **Statistical** method to determine a **probability** that an **optical/IR** source is a **counterpart** to a **radio** source
- Works really well for point sources, not so much for larger more complex sources
- Parameters: **distance**, **magnitude**, **colour**
- **Automated** and **batched**
- Euclid split into three areas
- Able to use both **VIS** and **NIR** to create the **colour** component
- Completeness and reliability **~98.5%**
- Threshold then passed into the following steps



GBC

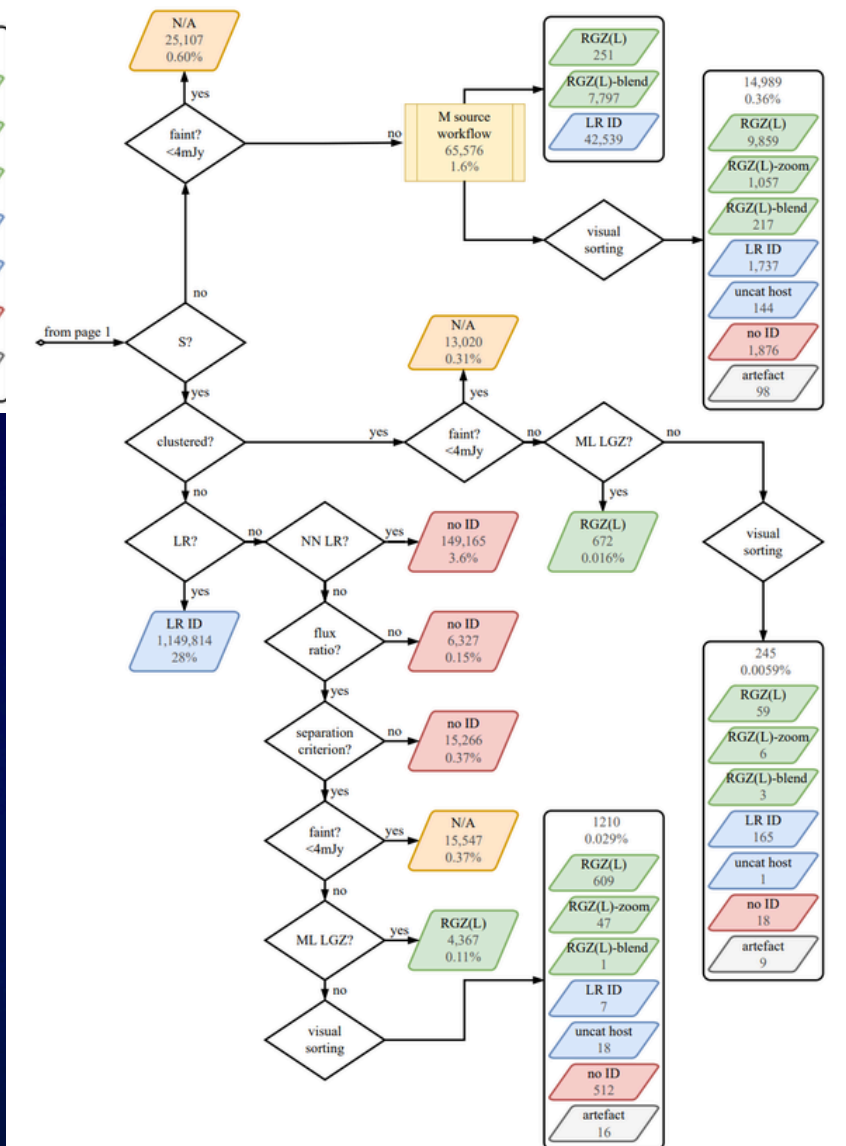
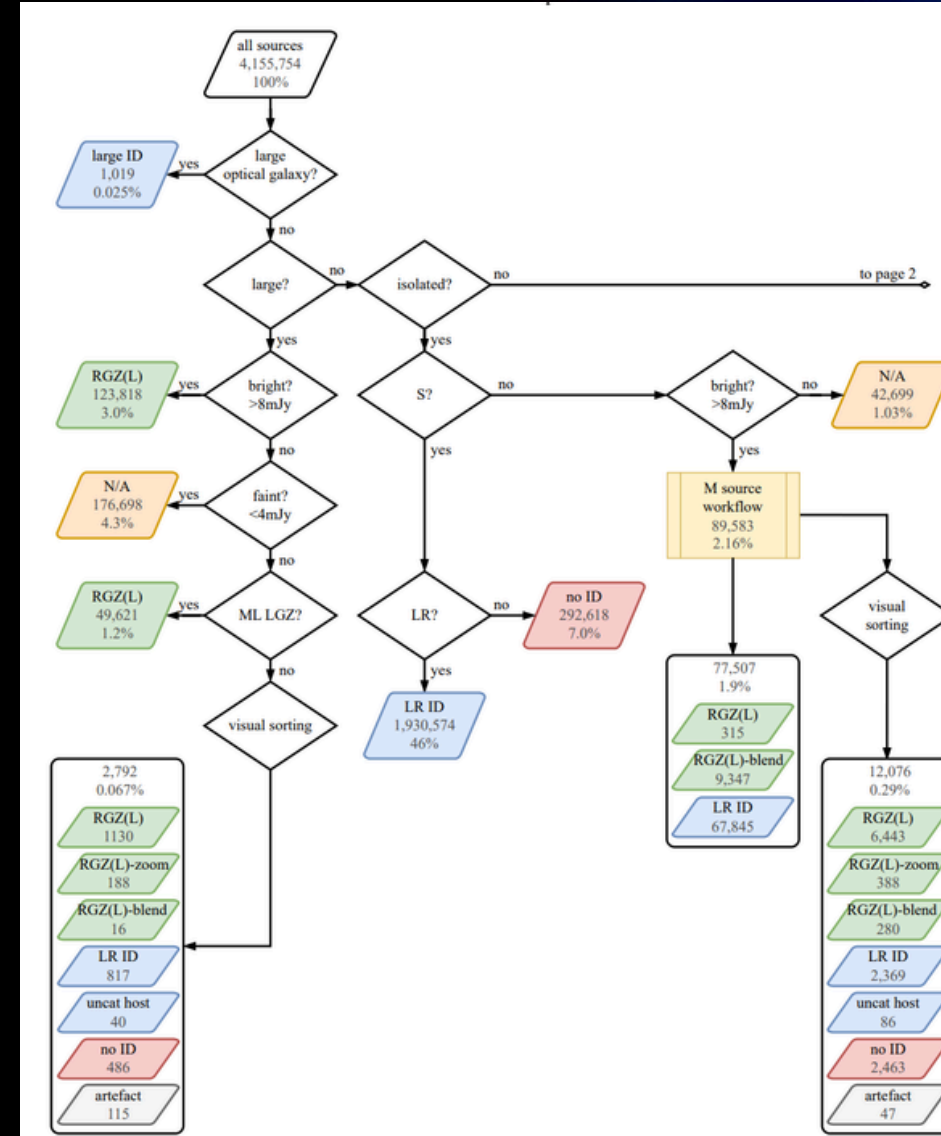
(gradient boosting classifier)



- ML model which uses: PyBDSF source and gaussian sizes, fluxes, and calculated LR values
- Outputs: probability that a PyBDSF source can be cross-matched with the LR or needs further inspection
- We get 90% of sources can be cross-matched with the LR
- Many sources over 0.95 - high probability of a LR match
- Increase of sources under 0.1 - high probability of visual inspection

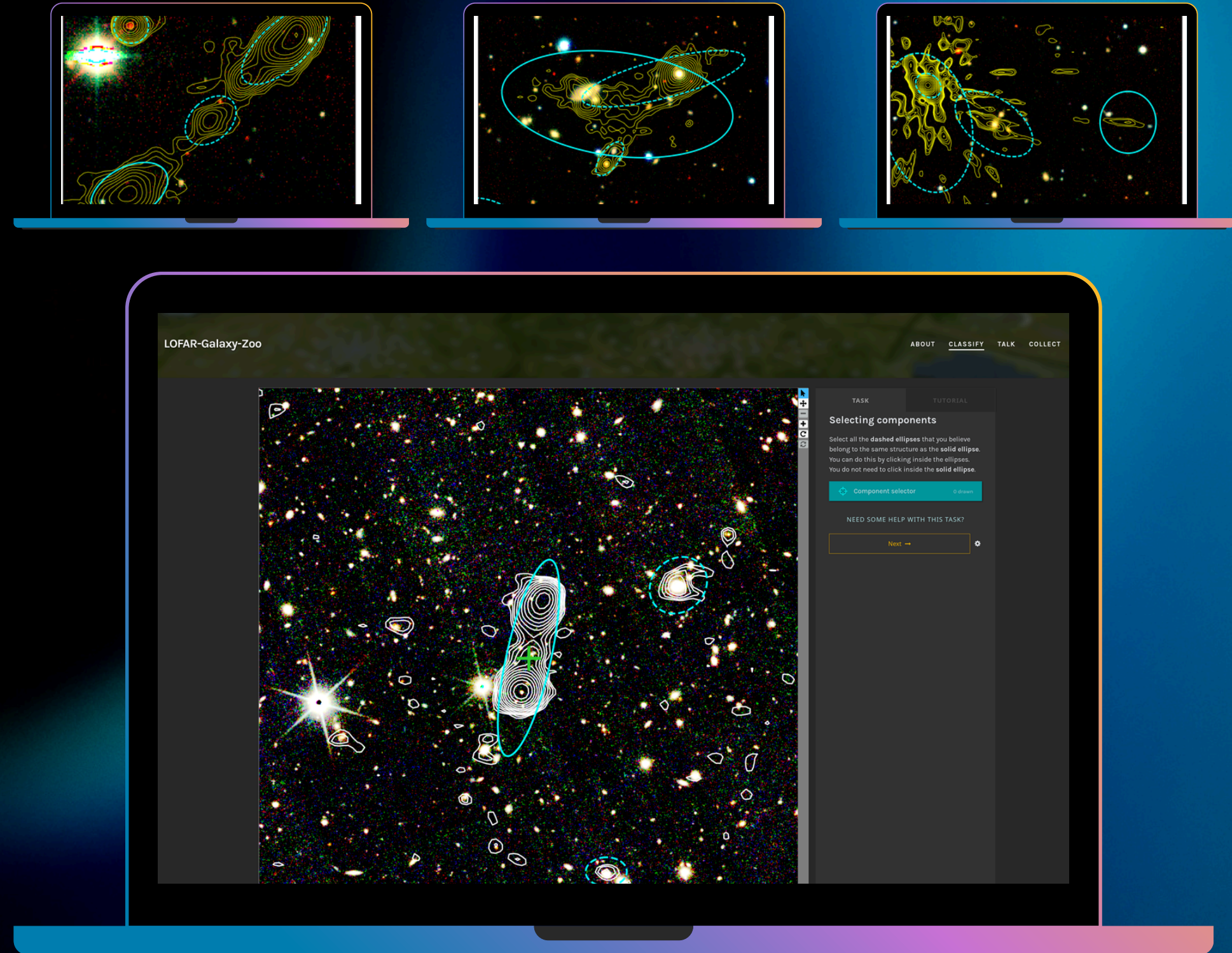
FlowChart

- LR and GBC feed into the flowchart, runs alongside the process
- Flags sources which can be identified now or need further investigation
- Series of scripts which need to be manually run and configured
- Currently we are automating the process with a single config file and a simple command process
- Removes a large proportion of sources
- All combined - initial point source catalogue



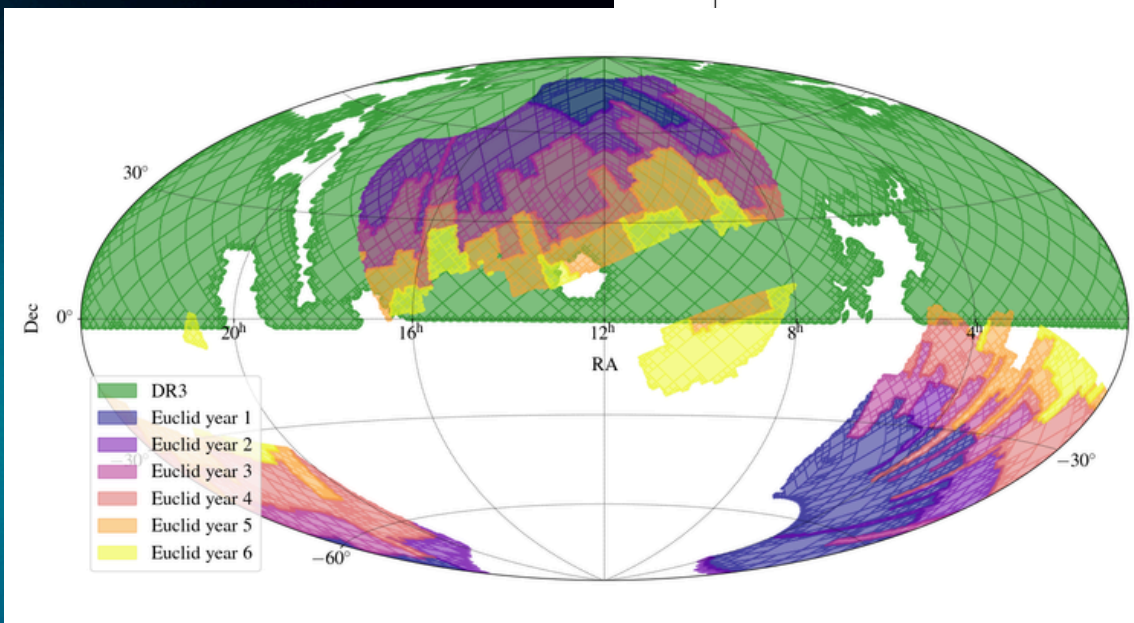
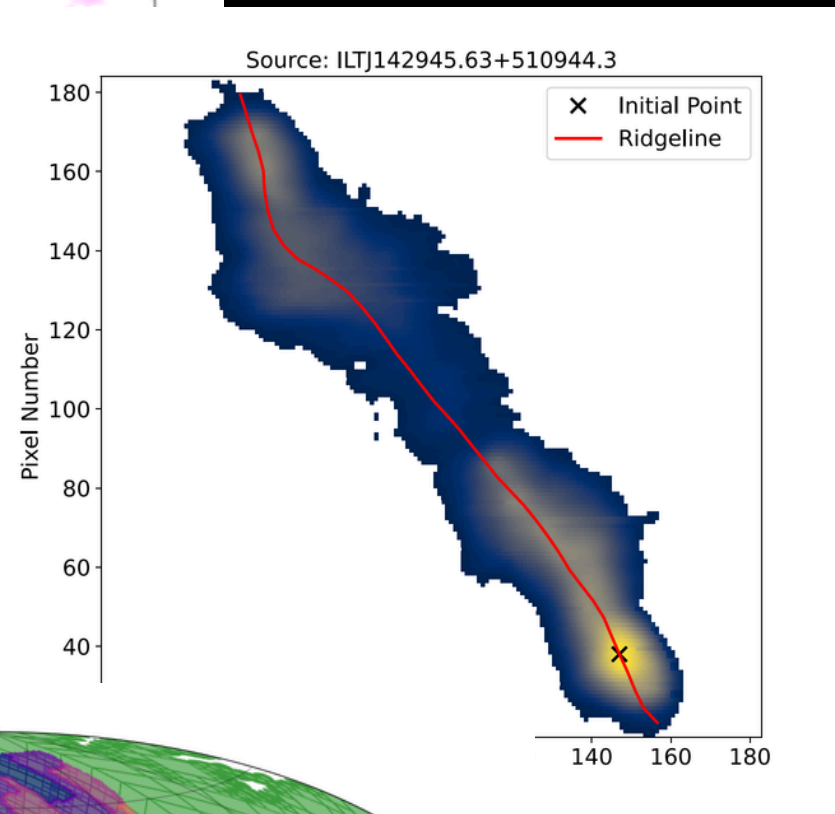
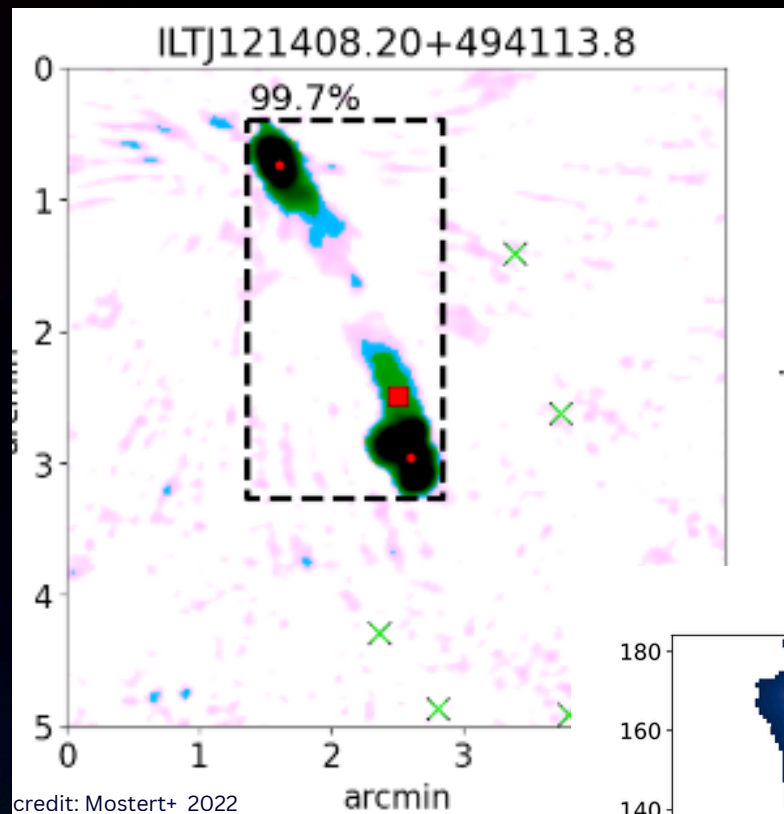
LOFAR Galaxy Zoo

- Regions of **higher complexity** flagged out such as: **association, blends, artefacts**
- More **complex sources** need their **hosts identified**
- Euclid DR1 we have gone with **visual inspection** - LOFAR galaxy zoo; not many need further investigation
- Up and running soon...
- Come and **talk to us** if you want to join the **exclusive sign-up** list



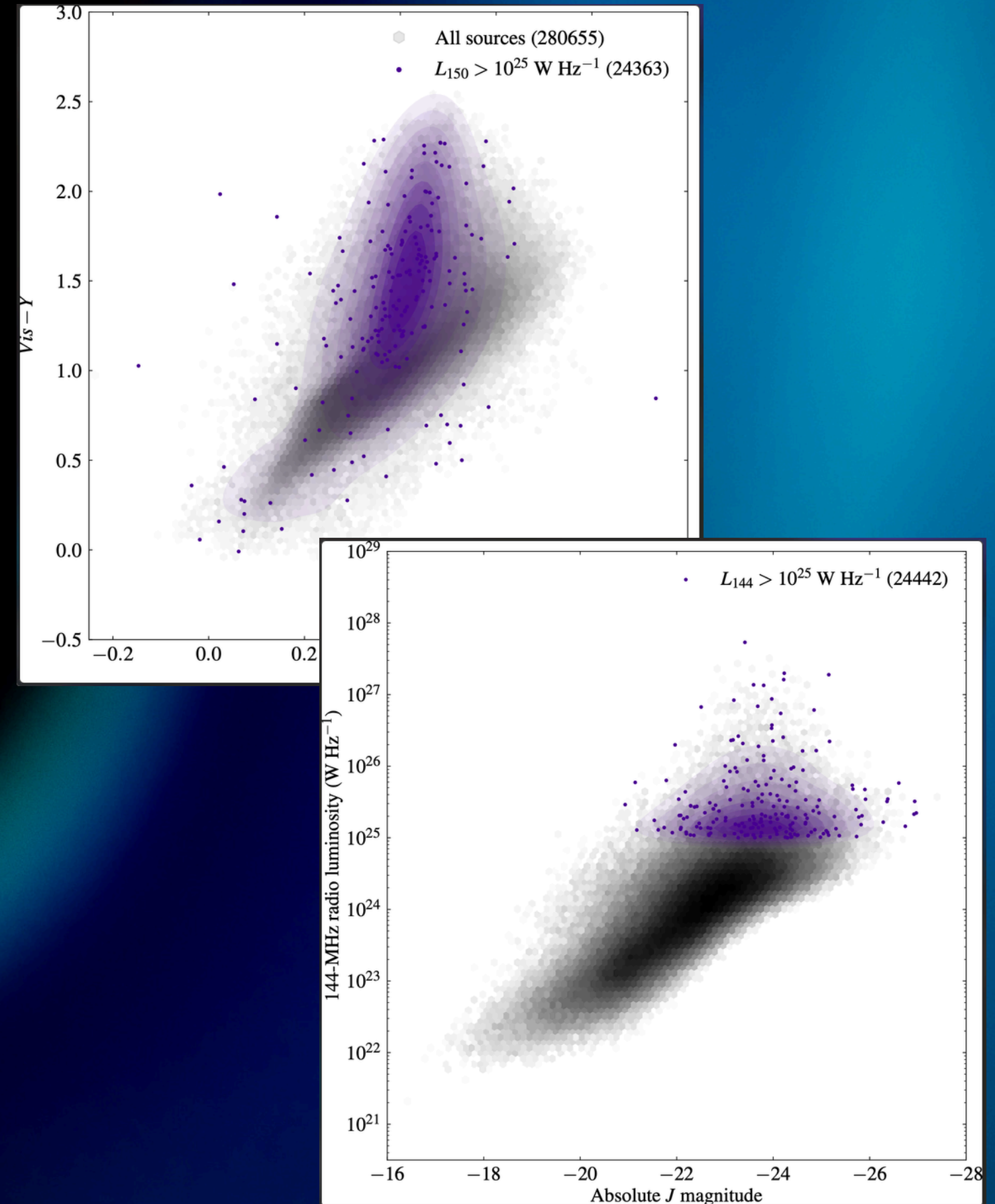
Future Steps...

- Looking to **automate** the visual inspection part in the future
 - ML techniques for the **association** (Mostert+ 2022)
 - **RL-Xid** for the host identification (Barkus+2022)
- Moving through the future releases of the EWS
- More of DR3 but with other optical/IR catalogues?



Science!

- Using the **initial point source** catalogue from the LR matched sources
- **~60%** have redshifts using the same method as DR2
- Demonstrate that we can make use of the Euclid photometry
- **Correlation** between galaxy magnitude and radio luminosity **up to a point** - main sequence of star formation gets messy once the AGN start kicking in.
- Luminous/AGN - lie on a line on the colour-colour plot associated with quiescent galaxies (Bisigello+ 2020)
- **It all makes sense!!** 🤯



Summary

- One of the **biggest challenges** faced by large radio sky surveys
- Introduced the **different stages** of the LoTSS DR3 and Euclid pipeline: source finding, LR, GBC, and the flowchart
- Which has given us the start of a catalogue with **~90% of our sources matched**
- **LOFAR galaxy zoo** for the more complex sources - you can help!
- Looking to **automate** the visual inspection. We can cover the **remaining Euclid** areas
- **~60%** have redshifts in the initial LR catalogue

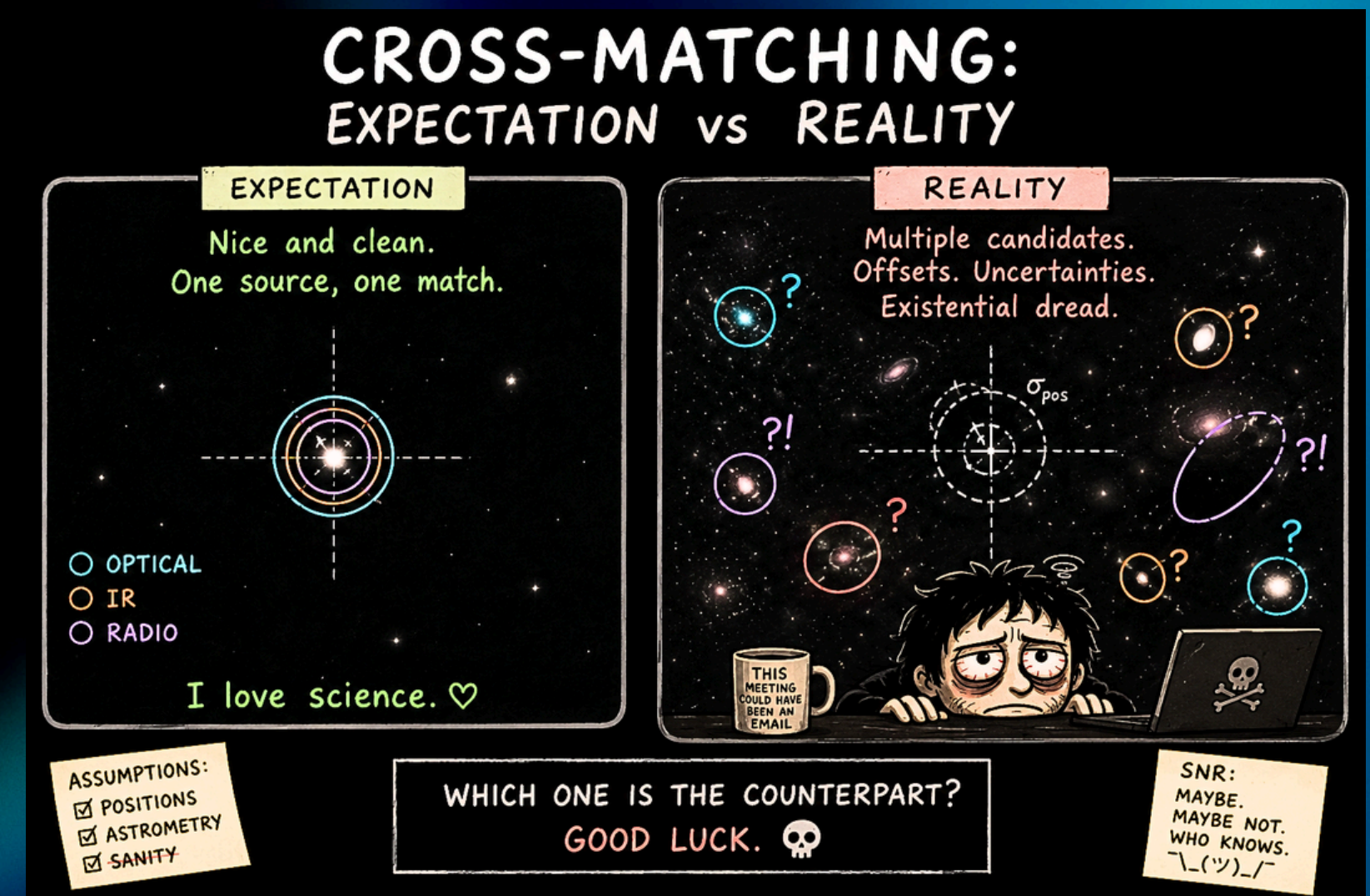


Image generated by ChatGPT for fun!

Thank you!

Questions?