



Ciências
ULisboa



Towards Better Selection and Characterisation Criteria for High-Redshift Radio Galaxies Using Machine-Assisted Pattern Recognition

RODRIGO CARVAJAL

SUPERVISED BY
DR J. AFONSO
DR I. MATUTE
DR H. MESSIAS



Ciências
ULisboa



Towards Better Selection and Characterisation Criteria for High-Redshift Radio Galaxies Using Machine-Assisted Pattern Recognition

THIS WORK WAS SUPPORTED BY THE FUNDAÇÃO PARA A CIÊNCIA E A TECNOLOGIA (FCT) THROUGH THE GRANT UIDP/04434/2020, UIDB/04434/2020, AND THE PHD FELLOWSHIP PD/BD/150455/2019 (PHD::SPACE DOCTORAL NETWORK PD/00040/2012) AND POCH/FSE (EC).



FCT
PROGRAMMES

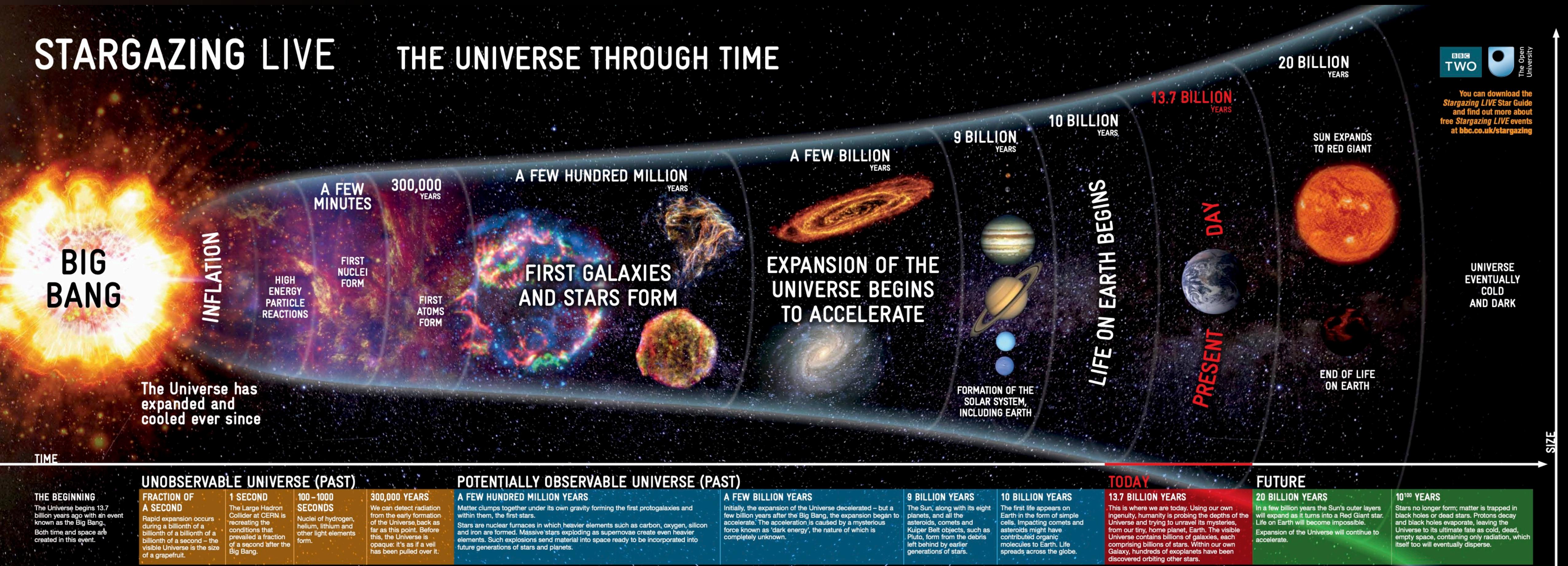


COFINANCIAMENTO / COFINANCING



STARGAZING LIVE

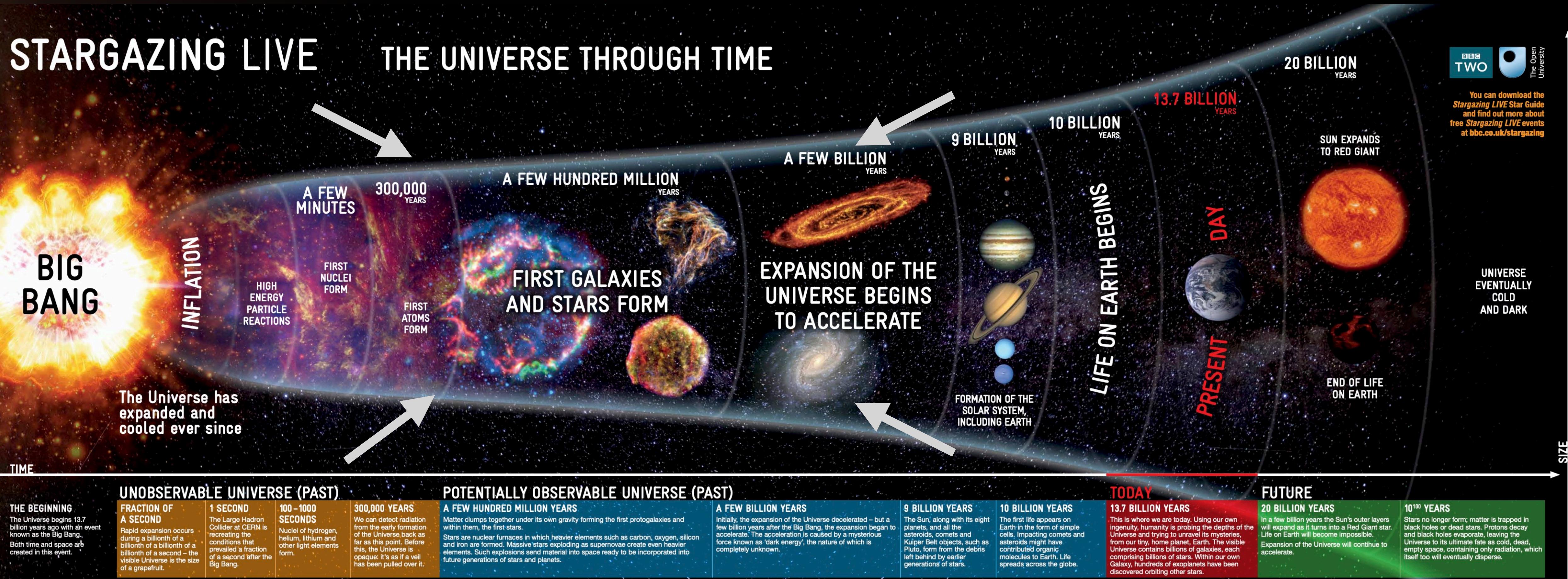
THE UNIVERSE THROUGH TIME



Stargazing LIVE is a BBC and Open University co-production. Credit: Photography sourced from NASA.

STARGAZING LIVE

THE UNIVERSE THROUGH TIME



Stargazing LIVE is a BBC and Open University co-production. Credit: Photography sourced from NASA.

STARGAZING LIVE

THE UNIVERSE THROUGH TIME

Need to understand role of SMBH in galaxy evolution!

The Universe has expanded and cooled ever since

TIME



The Open University

You can download the
Stargazing LIVE Star Guide
and find out more about
free Stargazing LIVE events
at bbc.co.uk/stargazing

UNOBSERVABLE UNIVERSE (PAST)

THE BEGINNING

The Universe begins 13.7 billion years ago with an event known as the Big Bang.

Both time and space are created in this event.

FRACTION OF A SECOND

The Large Hadron Collider at CERN is recreating the conditions that prevailed a fraction of a second after the Big Bang.

1 SECOND

Rapid expansion occurs during a billionth of a billionth of a billionth of a second – the visible Universe is the size of a grapefruit.

100 - 1000 SECONDS

Nuclei of hydrogen, helium, lithium and other light elements form.

We can detect radiation from the early formation of the Universe back as far as this point. Before this, the Universe is opaque: it's as if a veil has been pulled over it.

A FEW HUNDRED MILLION YEARS

Matter clumps together under its own gravity forming the first protogalaxies and within them, the first stars. Stars are nuclear furnaces in which heavier elements such as carbon, oxygen, silicon and iron are formed. Massive stars exploding as supernovae create even heavier elements. Such explosions send material into space ready to be incorporated into future generations of stars and planets.

POTENTIALLY OBSERVABLE UNIVERSE (PAST)

A FEW BILLION YEARS

Initially, the expansion of the Universe decelerated – but a few billion years after the Big Bang, the expansion began to accelerate. The acceleration is caused by a mysterious force known as 'dark energy', the nature of which is completely unknown.

9 BILLION YEARS

The Sun, along with its eight planets, and all the asteroids, comets and Kuiper Belt objects, such as Pluto, form from the debris left behind by earlier generations of stars.

10 BILLION YEARS

The first life appears on Earth in the form of simple cells. Impacting comets and asteroids might have contributed organic molecules to Earth. Life spreads across the globe.

TODAY

13.7 BILLION YEARS

This is where we are today. Using our own ingenuity, humanity is probing the depths of the Universe and trying to unravel its mysteries. From our tiny, home planet, Earth, the visible Universe contains billions of galaxies, each comprising billions of stars. Within our own Galaxy, hundreds of exoplanets have been discovered orbiting other stars.

FUTURE

20 BILLION YEARS

In a few billion years the Sun's outer layers will expand as it turns into a Red Giant star. Life on Earth will become impossible. Expansion of the Universe will continue to accelerate.

10¹⁰⁰ YEARS

Stars no longer form; matter is trapped in black holes or dead stars. Protons decay and black holes evaporate, leaving the Universe to its ultimate fate as cold, dead, empty space, containing only radiation, which itself too will eventually disperse.

STARGAZING LIVE

THE UNIVERSE THROUGH TIME

Need to understand role of SMBH in galaxy evolution!

The Universe has expanded and cooled ever since

AGN + SF

TIME



The Open University

You can download the
Stargazing LIVE Star Guide
and find out more about
free Stargazing LIVE events
at bbc.co.uk/stargazing

↑
SIZE
↓

UNOBSERVABLE UNIVERSE (PAST)

THE BEGINNING

The Universe begins 13.7 billion years ago with an event known as the Big Bang.

Both time and space are created in this event.

FRACTION OF A SECOND

The Large Hadron Collider at CERN is recreating the conditions that prevailed a fraction of a second after the Big Bang.

1 SECOND

Rapid expansion occurs during a billionth of a billionth of a billionth of a second – the visible Universe is the size of a grapefruit.

100 - 1000 SECONDS

Nuclei of hydrogen, helium, lithium and other light elements form.

300,000 YEARS

We can detect radiation from the early formation of the Universe back as far as this point. Before this, the Universe is opaque: it's as if a veil has been pulled over it.

POTENTIALLY OBSERVABLE UNIVERSE (PAST)

A FEW HUNDRED MILLION YEARS

A FEW BILLION YEARS

Matter clumps together under its own gravity forming the first protogalaxies and within them, the first stars.

Stars are nuclear furnaces in which heavier elements such as carbon, oxygen, silicon and iron are formed. Massive stars exploding as supernovae create even heavier elements. Such explosions send material into space ready to be incorporated into future generations of stars and planets.

TODAY

9 BILLION YEARS

The Sun, along with its eight planets, and all the asteroids, comets and Kuiper Belt objects, such as Pluto, form from the debris left behind by earlier generations of stars.

FUTURE

10 BILLION YEARS

The first life appears on Earth in the form of simple cells. Impacting comets and asteroids might have contributed organic molecules to Earth. Life spreads across the globe.

13.7 BILLION YEARS

This is where we are today. Using our own ingenuity, humanity is probing the depths of the Universe and trying to unravel its mysteries. From our tiny, home planet, Earth, the visible Universe contains billions of galaxies, each comprising billions of stars. Within our Galaxy, hundreds of exoplanets have been discovered orbiting other stars.

20 BILLION YEARS

In a few billion years the Sun's outer layers will expand as it turns into a Red Giant star. Life on Earth will become impossible. Expansion of the Universe will continue to accelerate.

Stars no longer form; matter is trapped in black holes or dead stars. Protons decay and black holes evaporate, leaving the Universe to its ultimate fate as cold, dead, empty space, containing only radiation, which itself too will eventually disperse.

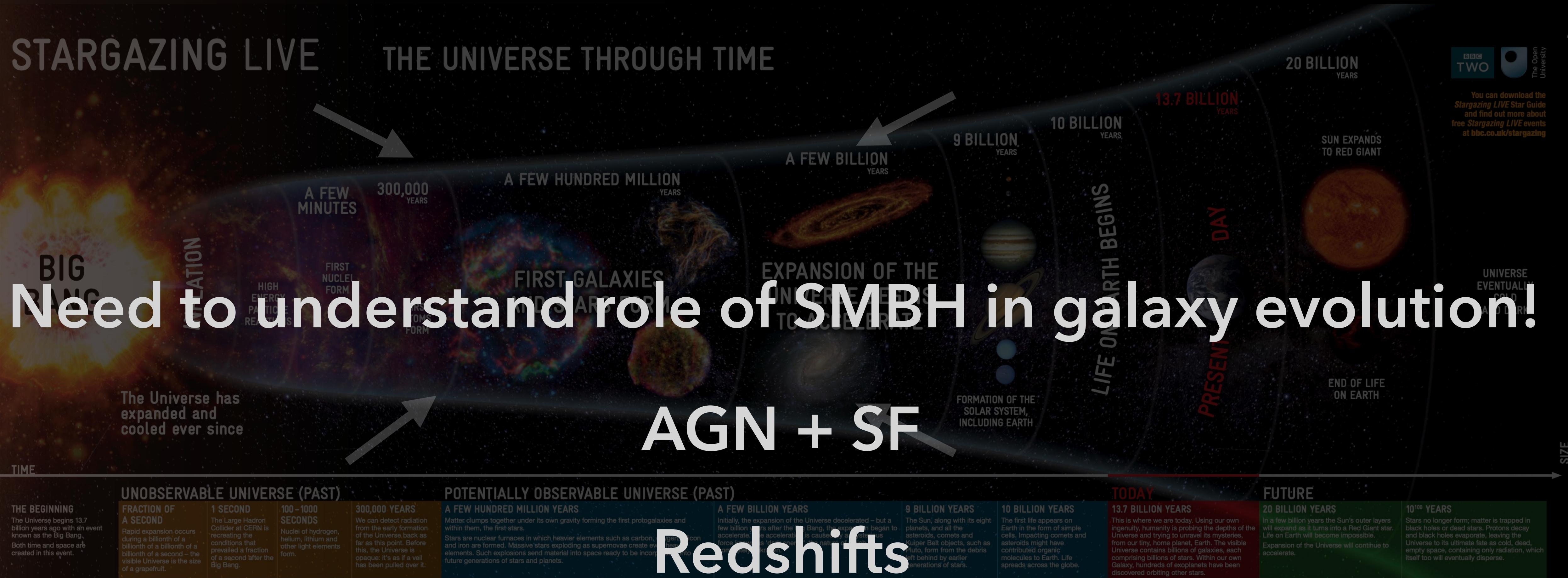
STARGAZING LIVE

THE UNIVERSE THROUGH TIME

Need to understand role of SMBH in galaxy evolution!

AGN + SF

Redshifts



OUTLINE

Separating AGN from SFGs

Our approach

Results

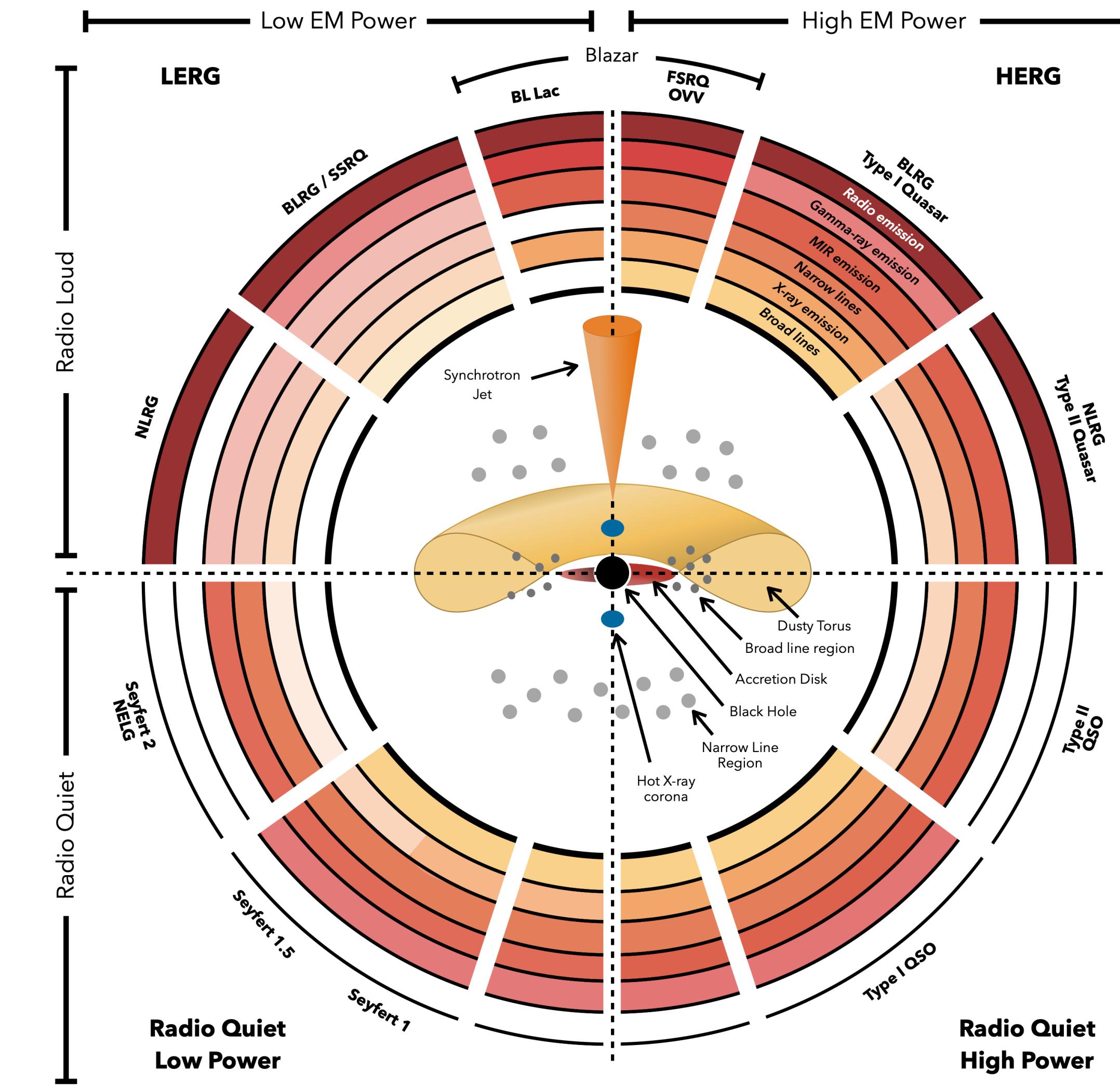
Analysing our tool

Conclusions and future work

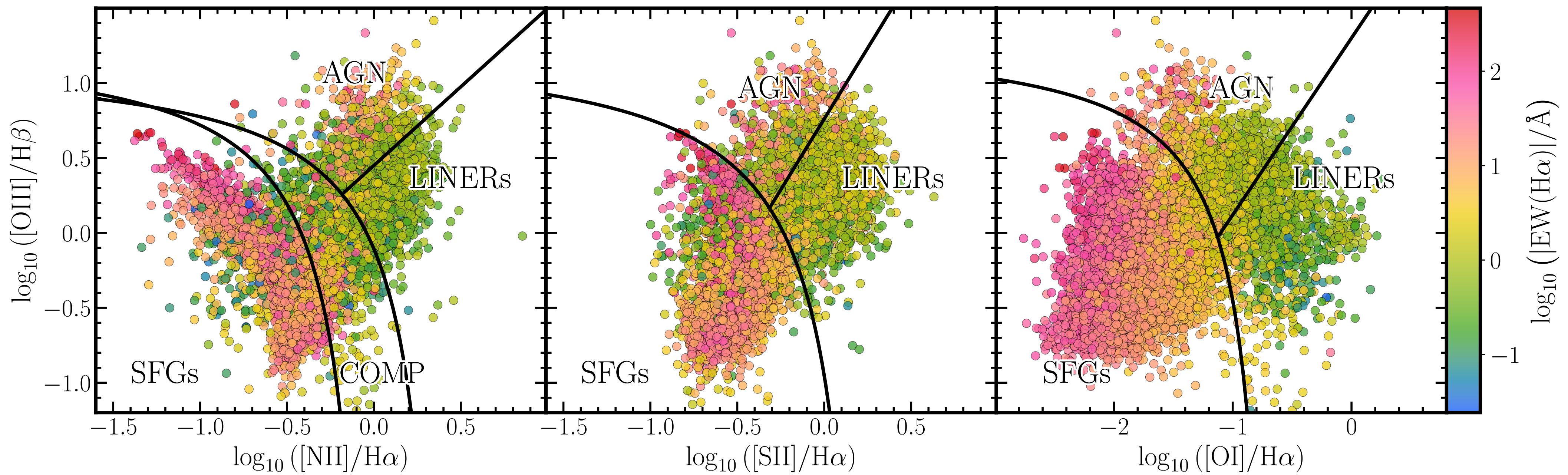
SEPARATING AGN FROM SFG (NON-AGN)

DETECT AGN THROUGH DIFFERENT WAVELENGTHS

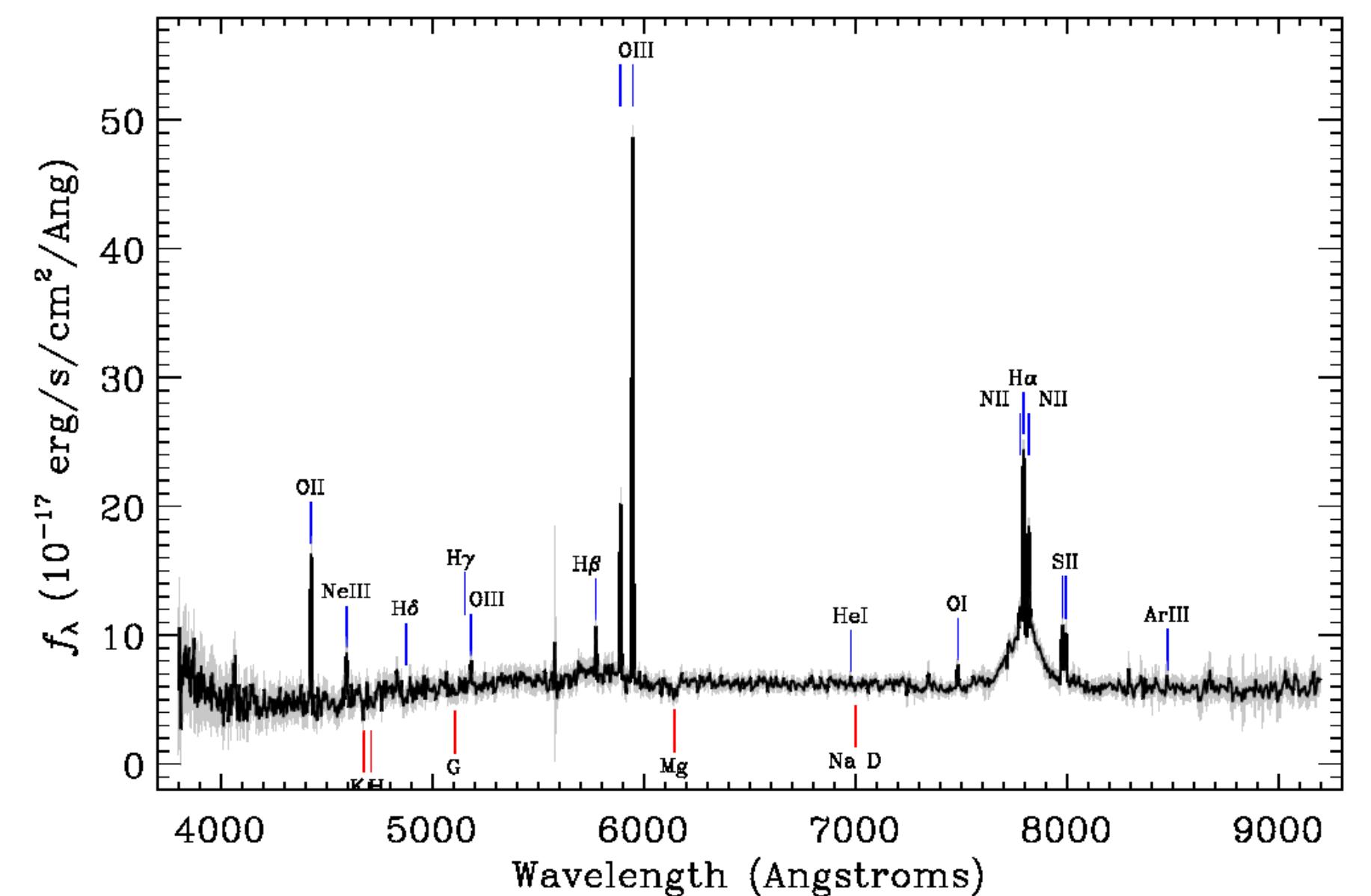
ACCESSING VARIOUS PHYSICAL
PROCESSES AND PROPERTIES



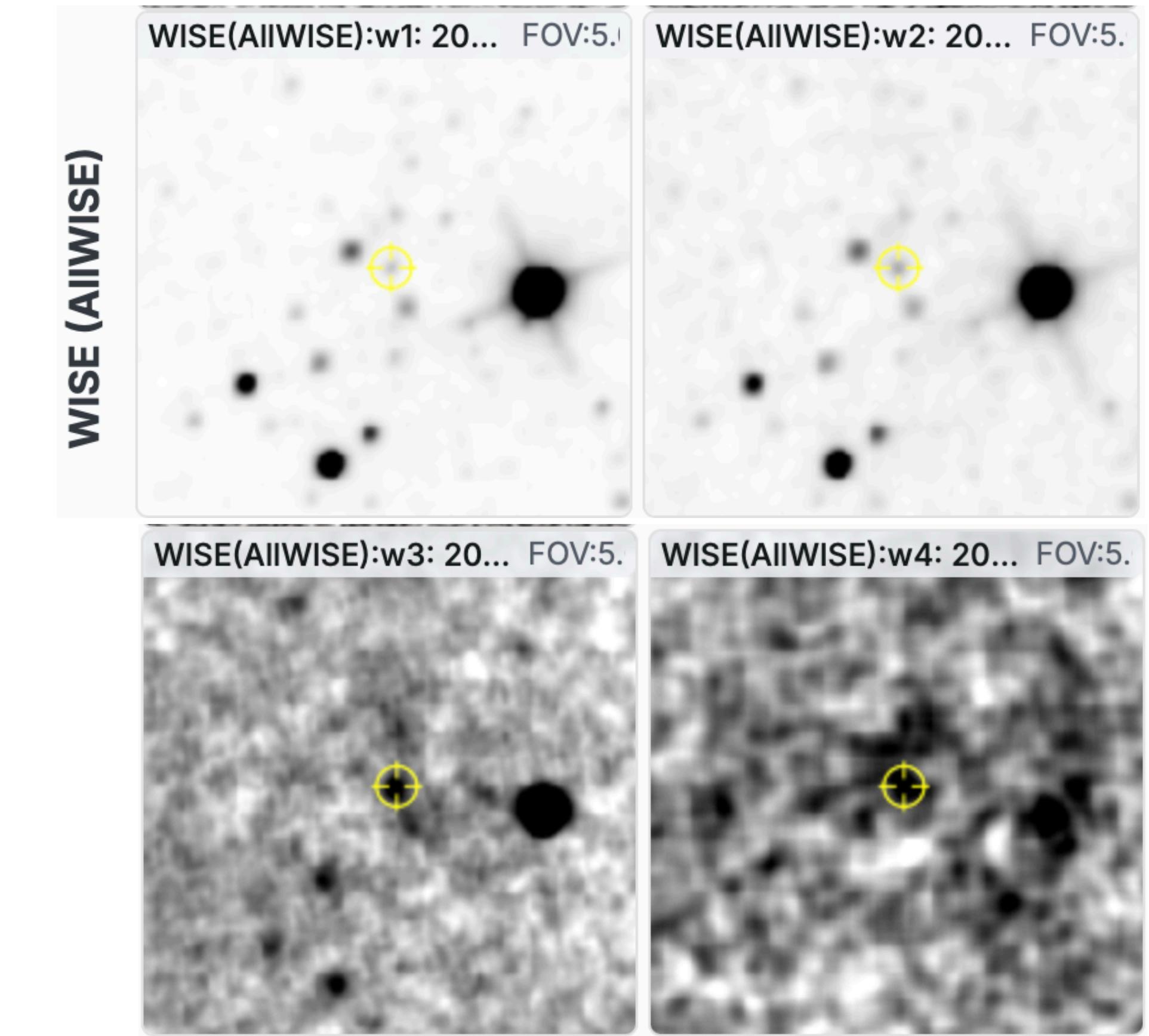
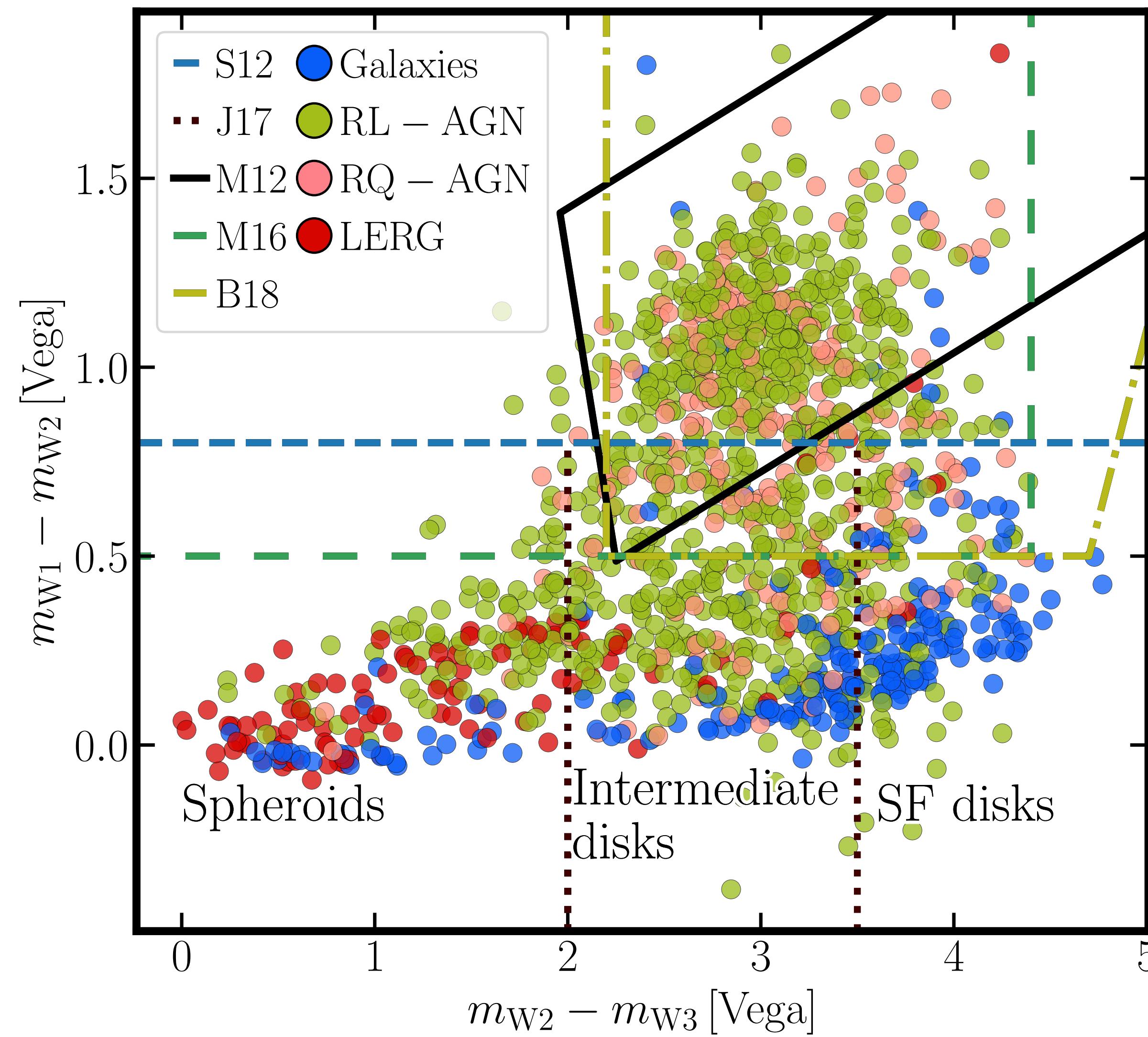
J. E. Thorne



BPT Diagram SPECTROSCOPY



SDSS DR18 Quick look



PHOTOMETRY

For ex. WISE Colours

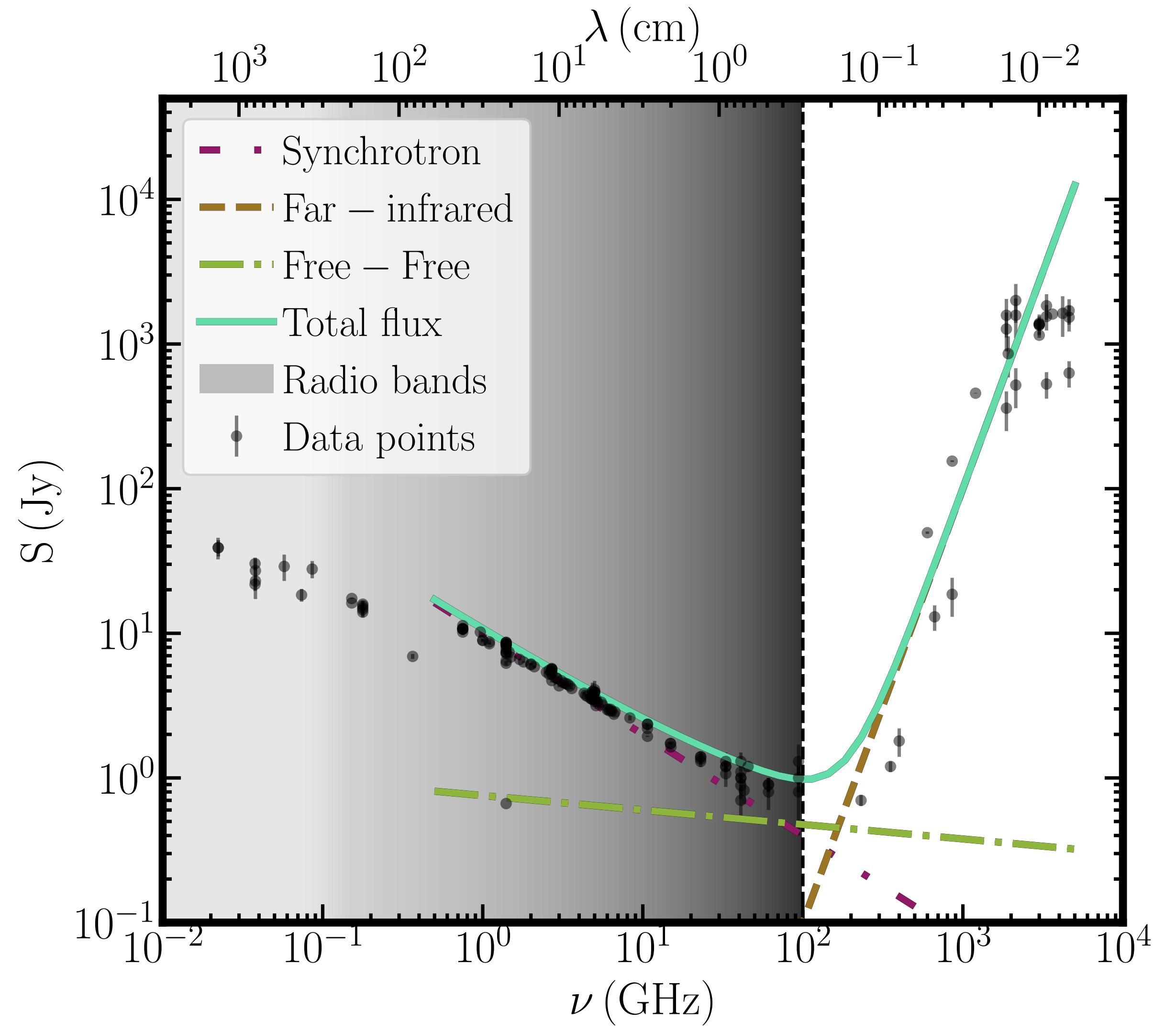
IRSA Finder Chart

ALMOST DIRECT OBSERVATION OF AGN

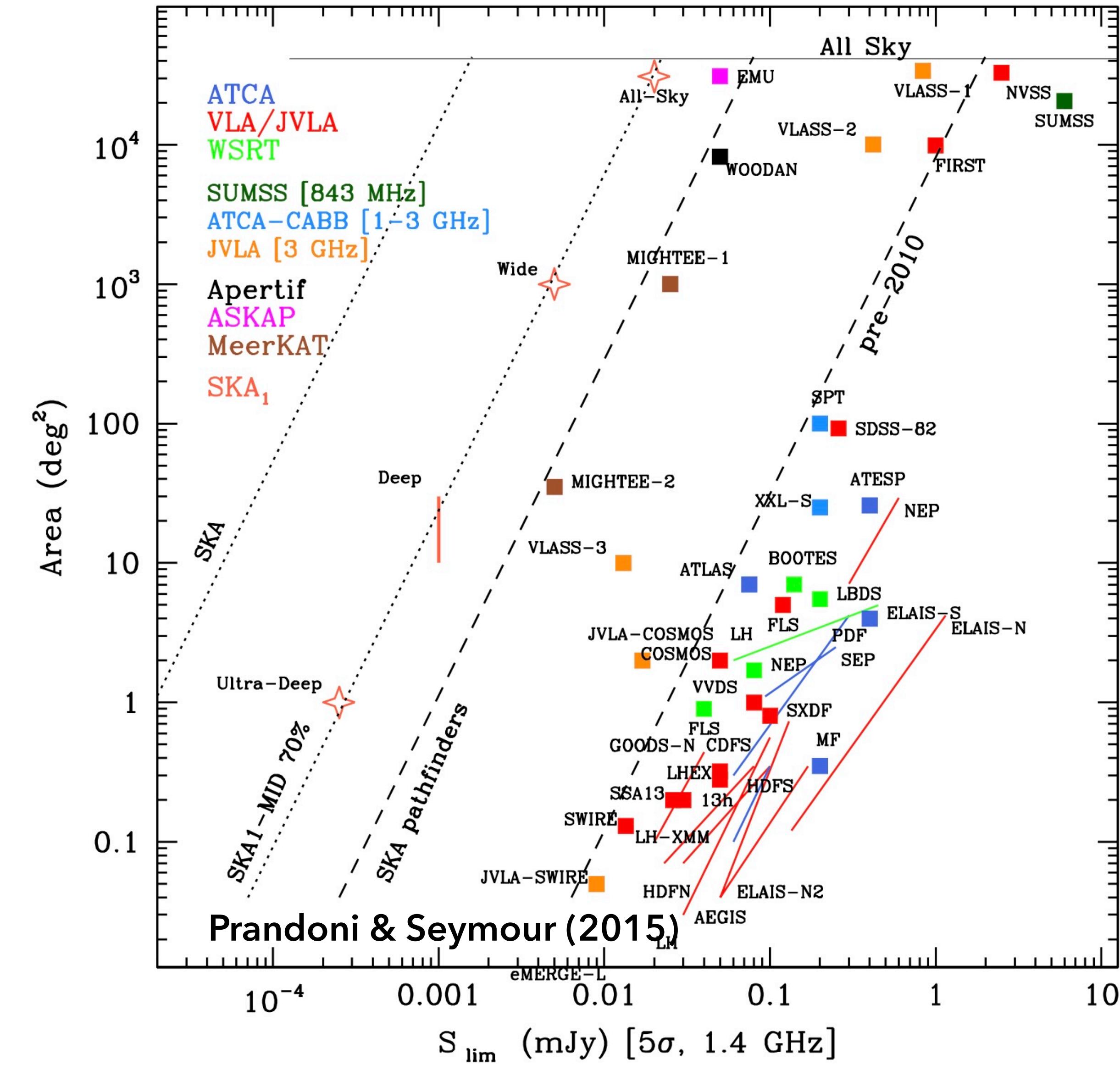
SKA + PRECURSORS

LARGE DATASETS

RADIO BANDS



BIG-DATA ERA

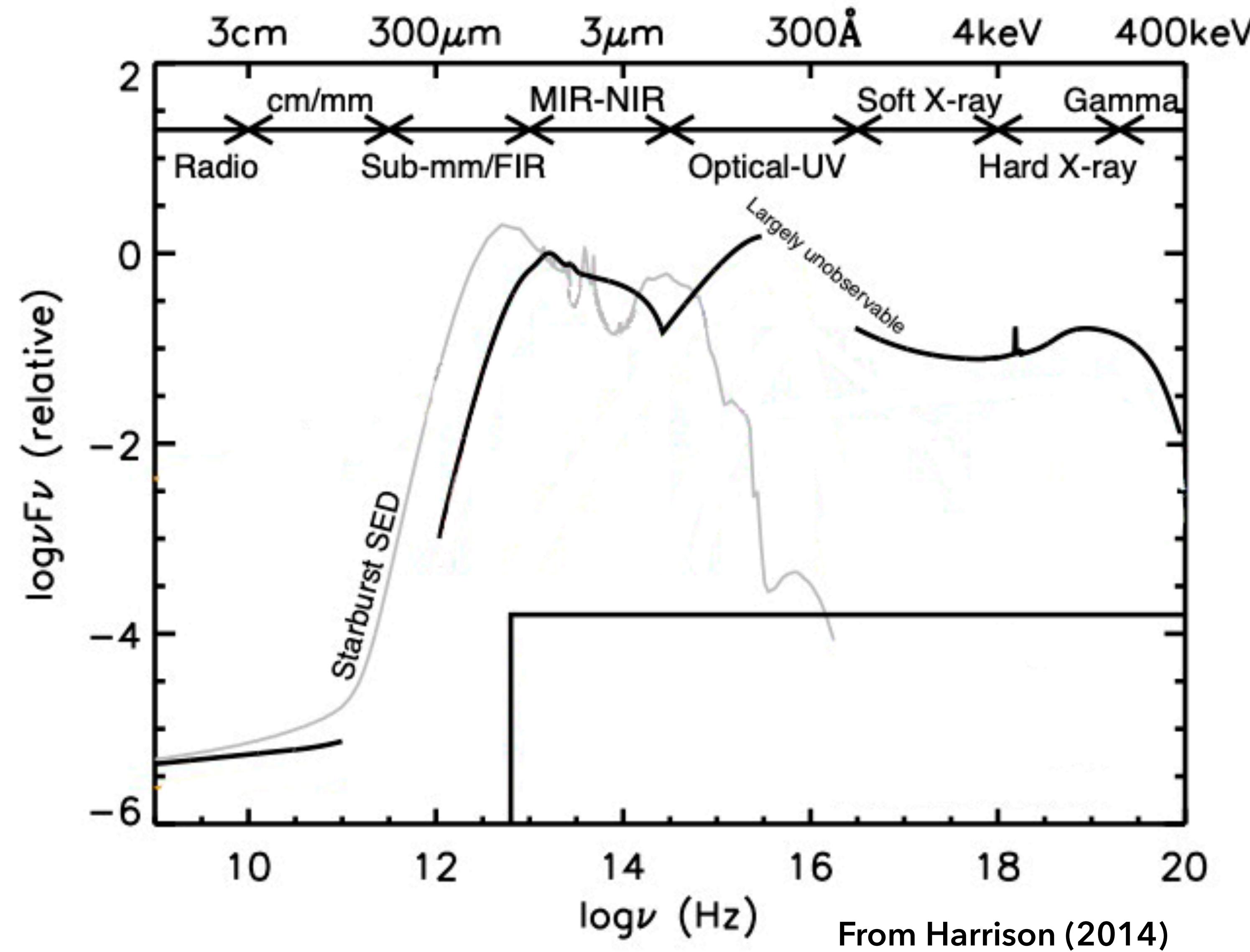


**WE NEED A MULTI-WAVELENGTH
APPROACH!**

**IS IT POSSIBLE TO INCORPORATE SEVERAL
INDICATORS INTO ONE TOOL?**

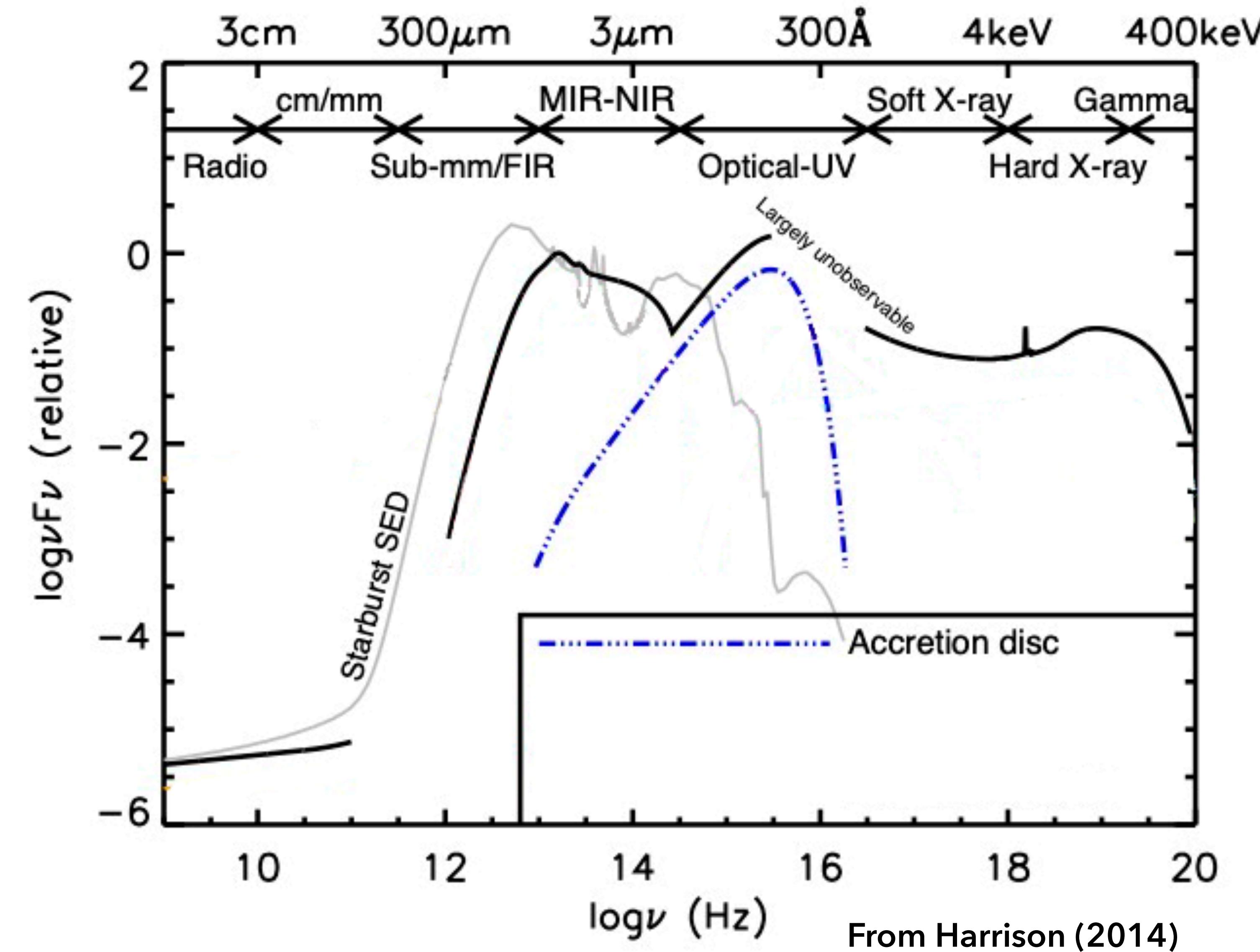
SED FITTING

ONE WAY TO DO
MULTI-WAVELENGTH
ANALYSIS



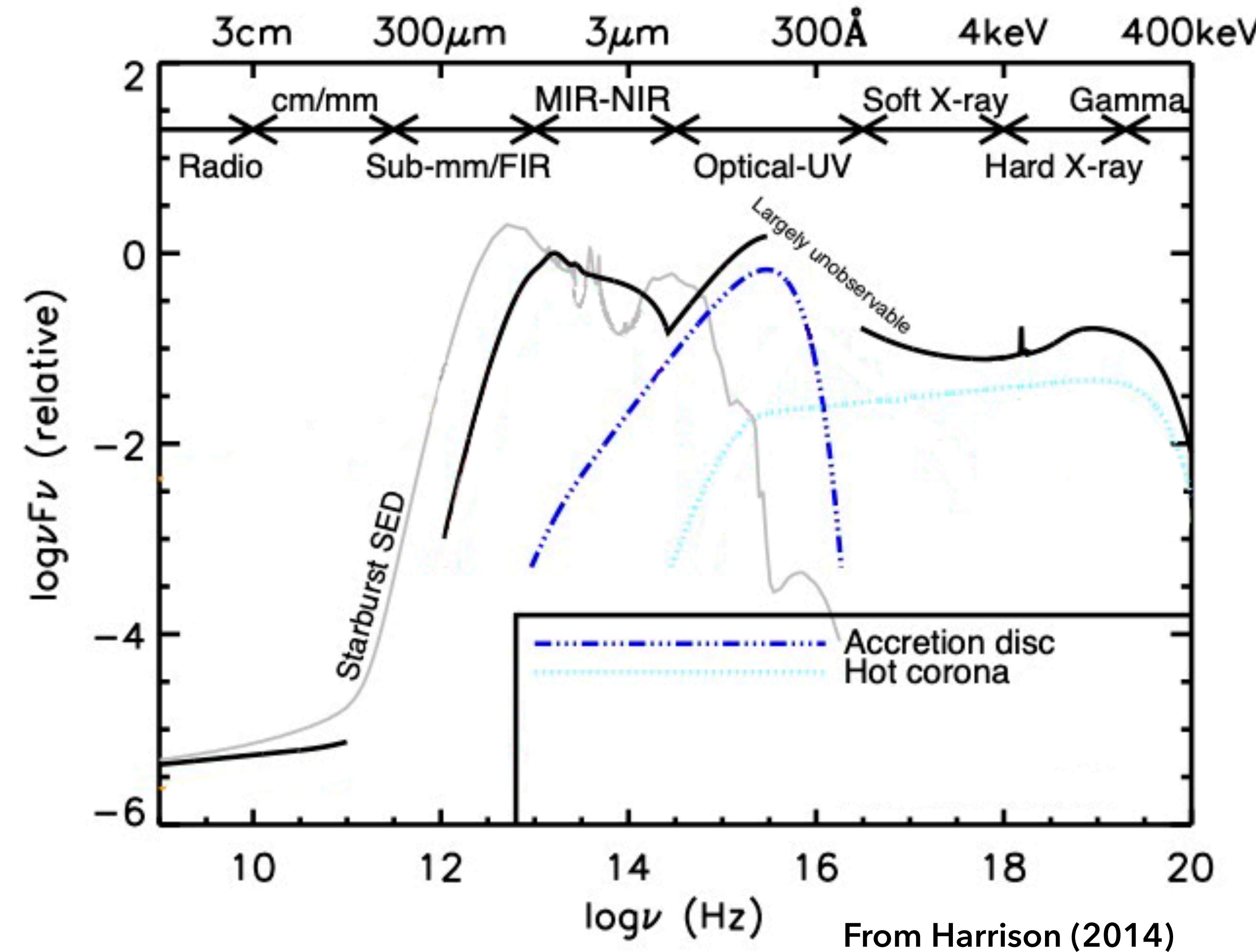
SED FITTING

ONE WAY TO DO
MULTI-WAVELENGTH
ANALYSIS



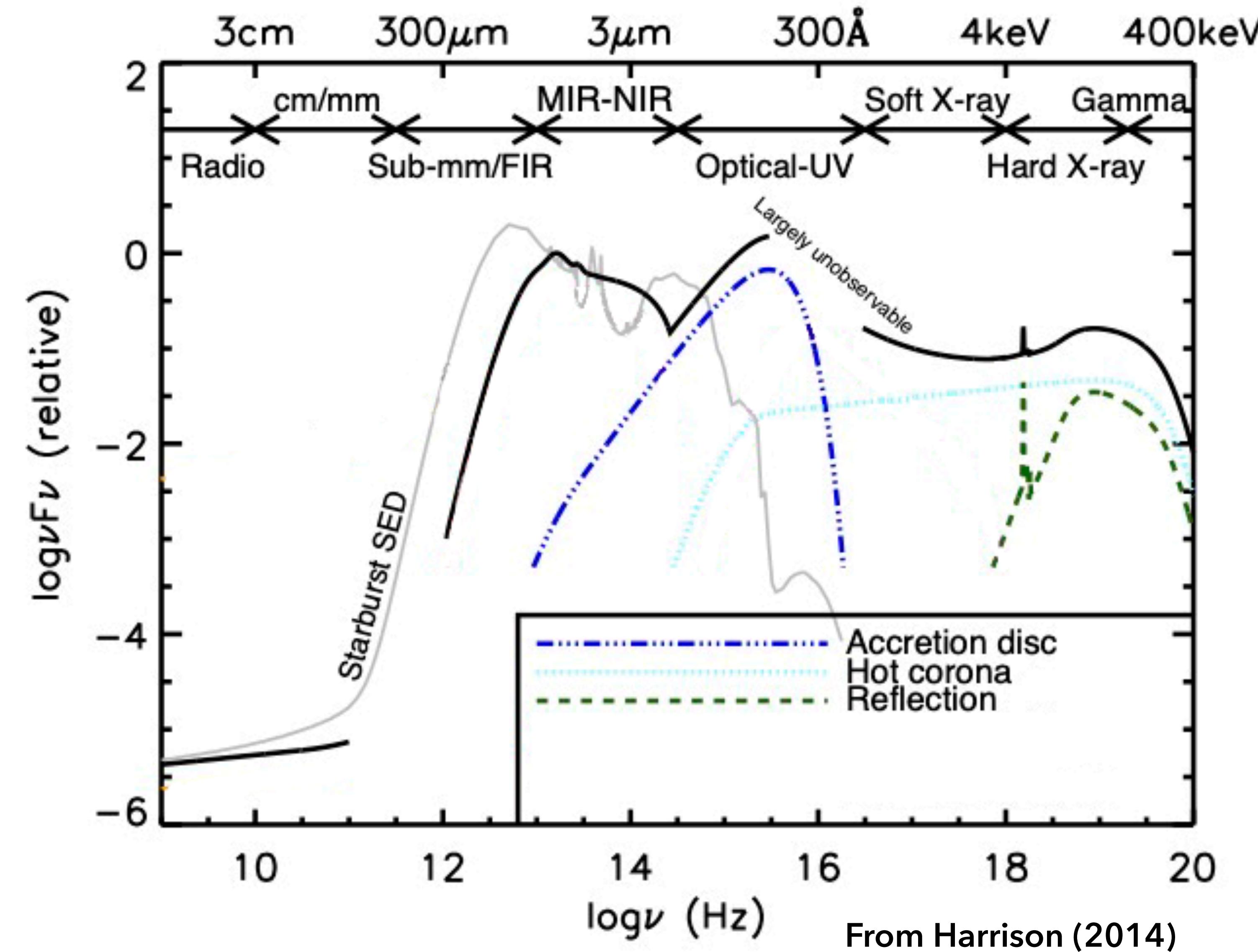
SED FITTING

ONE WAY TO DO
MULTI-WAVELENGTH
ANALYSIS



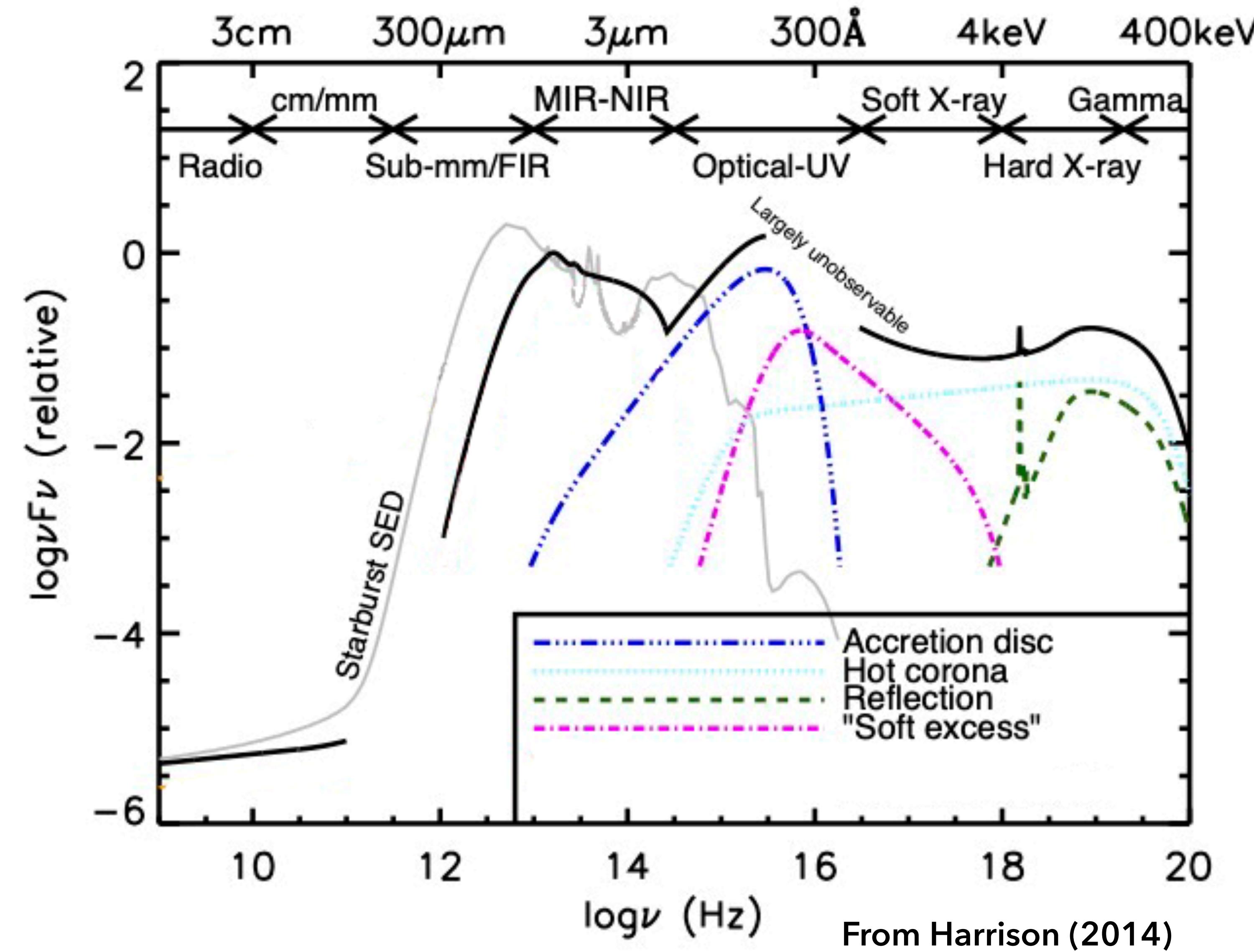
SED FITTING

ONE WAY TO DO
MULTI-WAVELENGTH
ANALYSIS



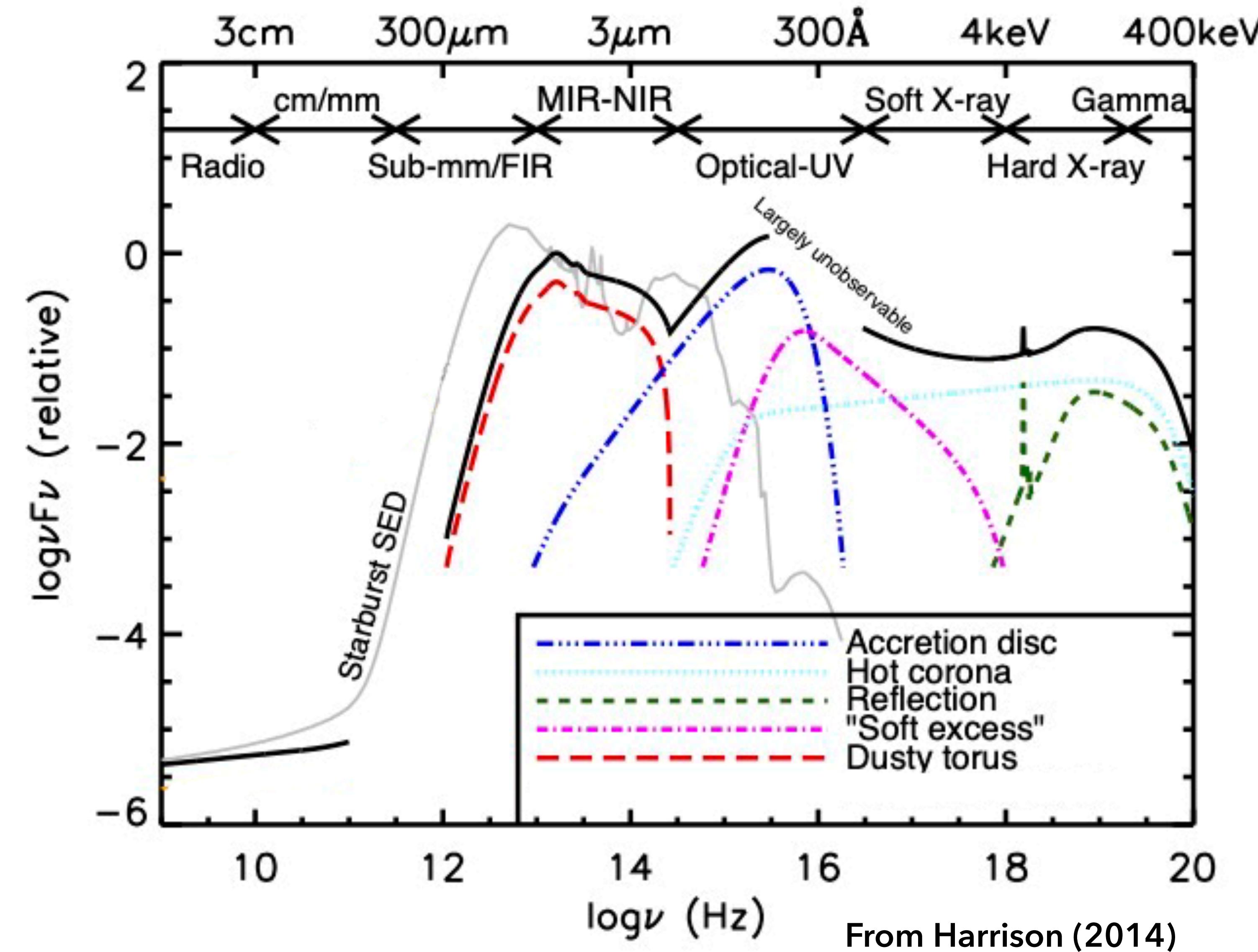
SED FITTING

ONE WAY TO DO
MULTI-WAVELENGTH
ANALYSIS



SED FITTING

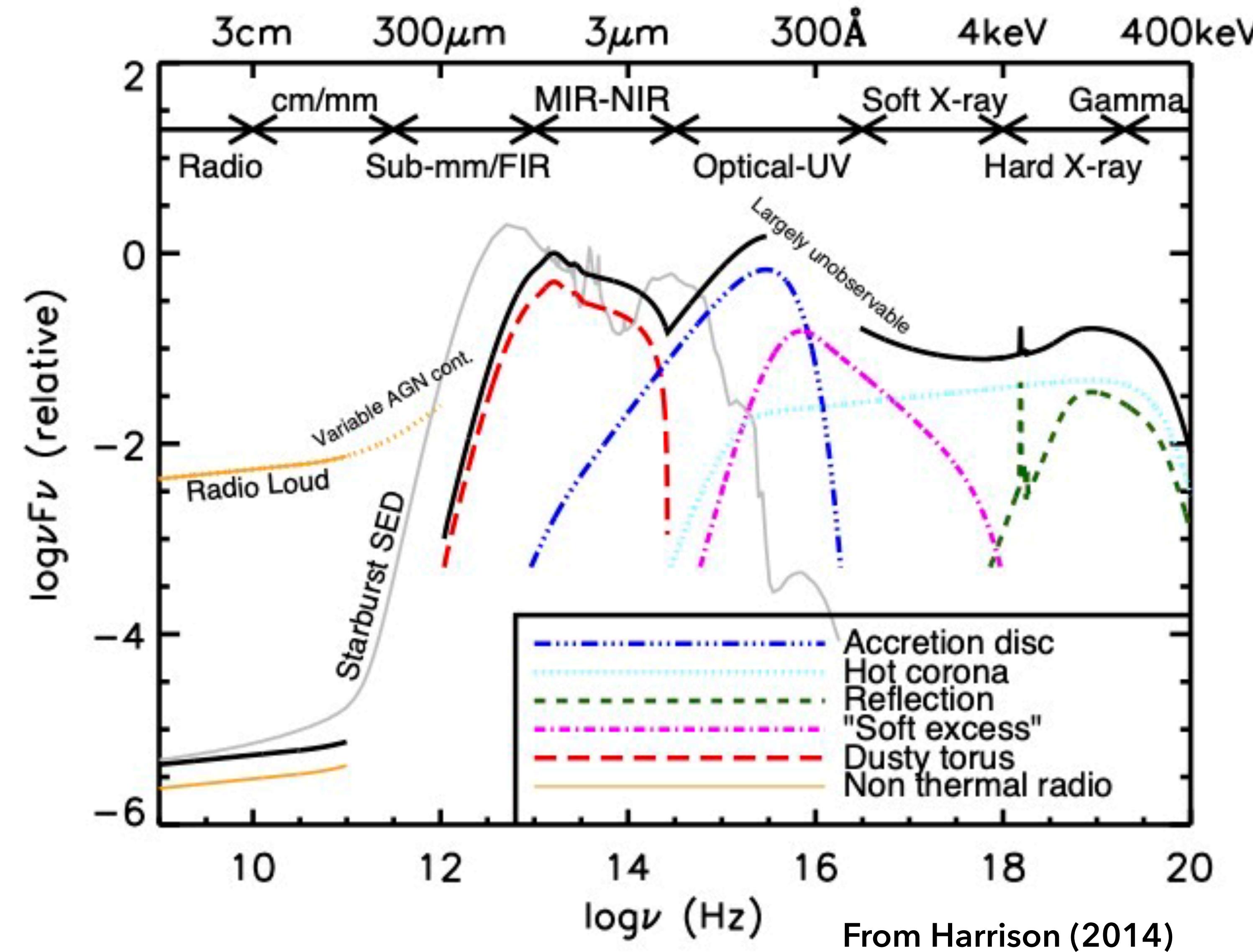
ONE WAY TO DO
MULTI-WAVELENGTH
ANALYSIS



From Harrison (2014)

SED FITTING

ONE WAY TO DO
MULTI-WAVELENGTH
ANALYSIS



MACHINE LEARNING CAN HELP!

MACHINE LEARNING

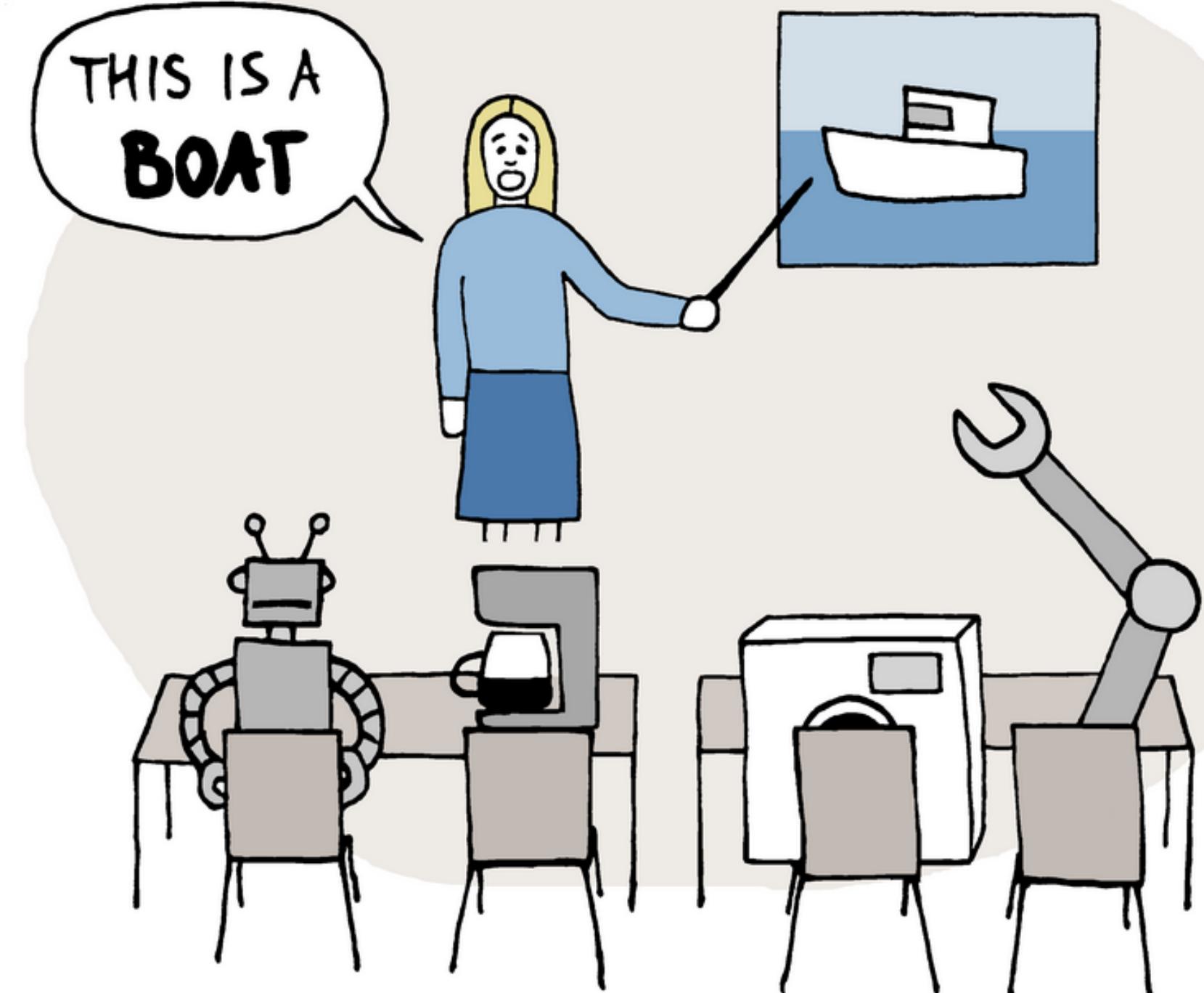
It can take advantage of very high dimensional datasets

It can determine patterns and trends within data and apply them to new measurements

We can examine predictions to further understand physical processes

It can guide us towards otherwise hidden research paths

MACHINE LEARNING



OUR GOAL

**Establish a method to select and characterise
radio-detected sources from large datasets**

PREDICTION PIPELINE

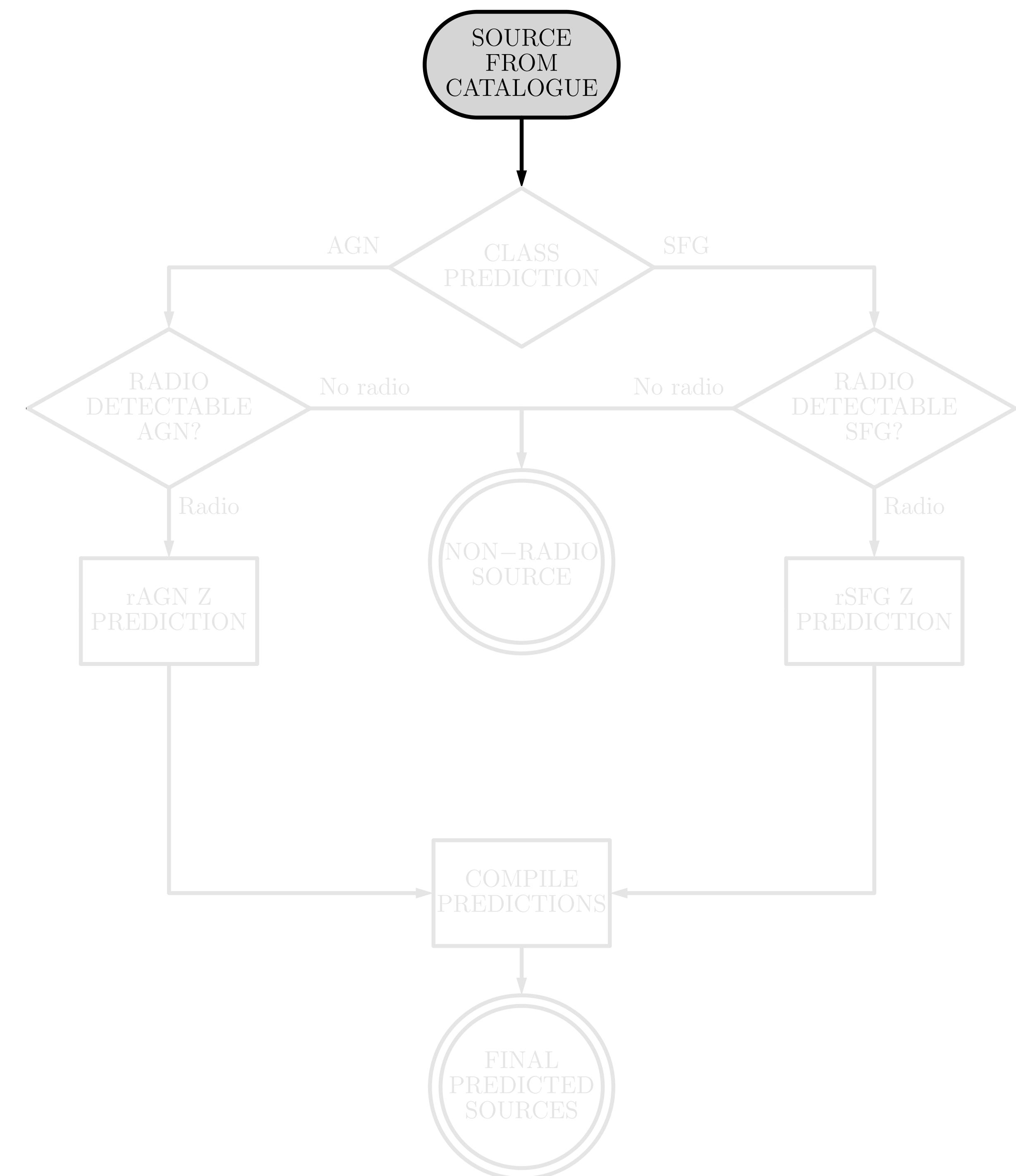
OUR APPROACH TO SELECTING AGN AND SFG WITH ML

Three different levels

Classify as AGN or SFG

Select radio-detectable sources

Estimate redshift for radio-detectable sources



PREDICTION PIPELINE

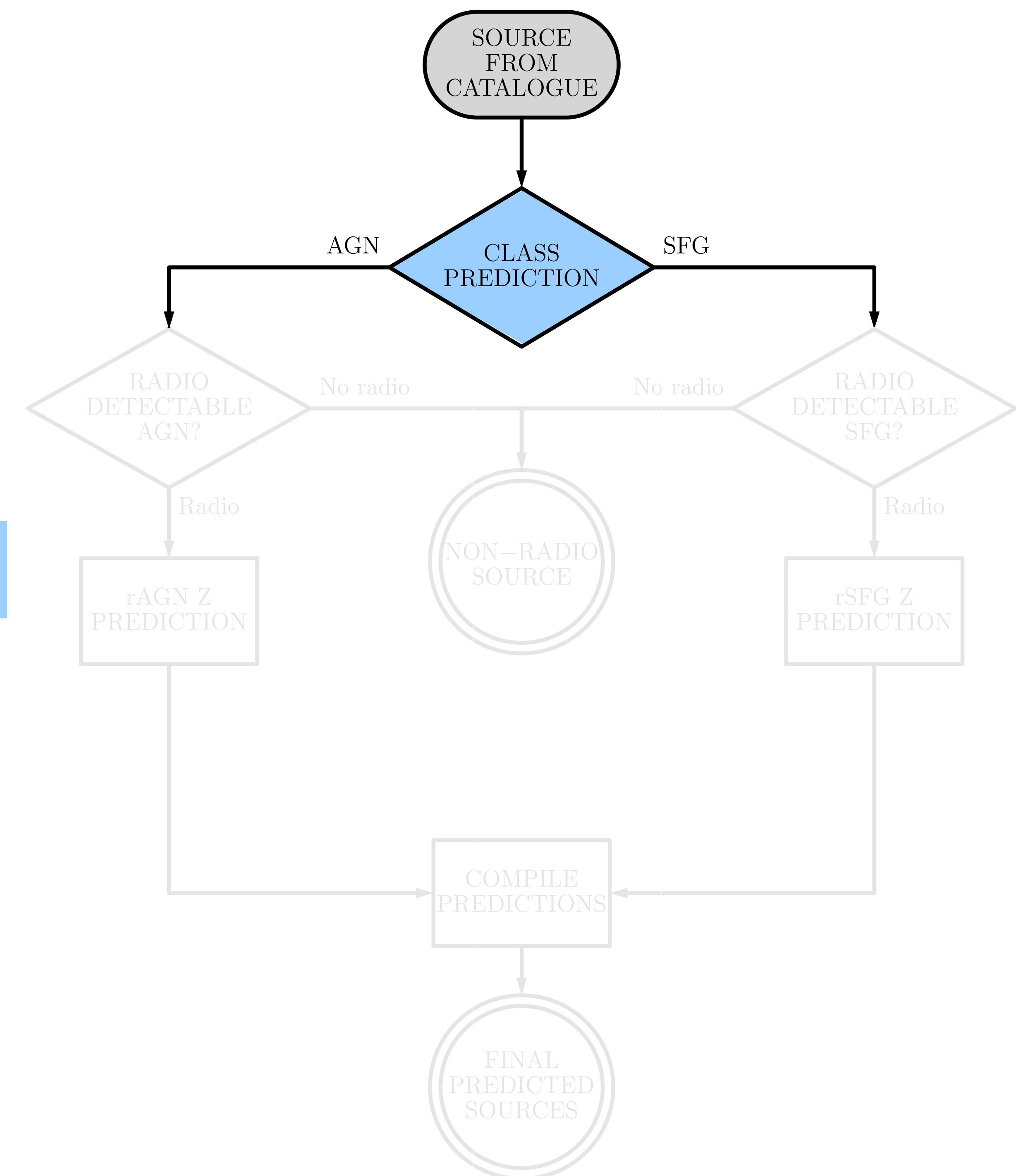
OUR APPROACH TO SELECTING AGN AND SFG WITH ML

Three different levels

Classify as AGN or SFG

Select radio-detectable sources

Estimate redshift for radio-detectable sources



PREDICTION PIPELINE

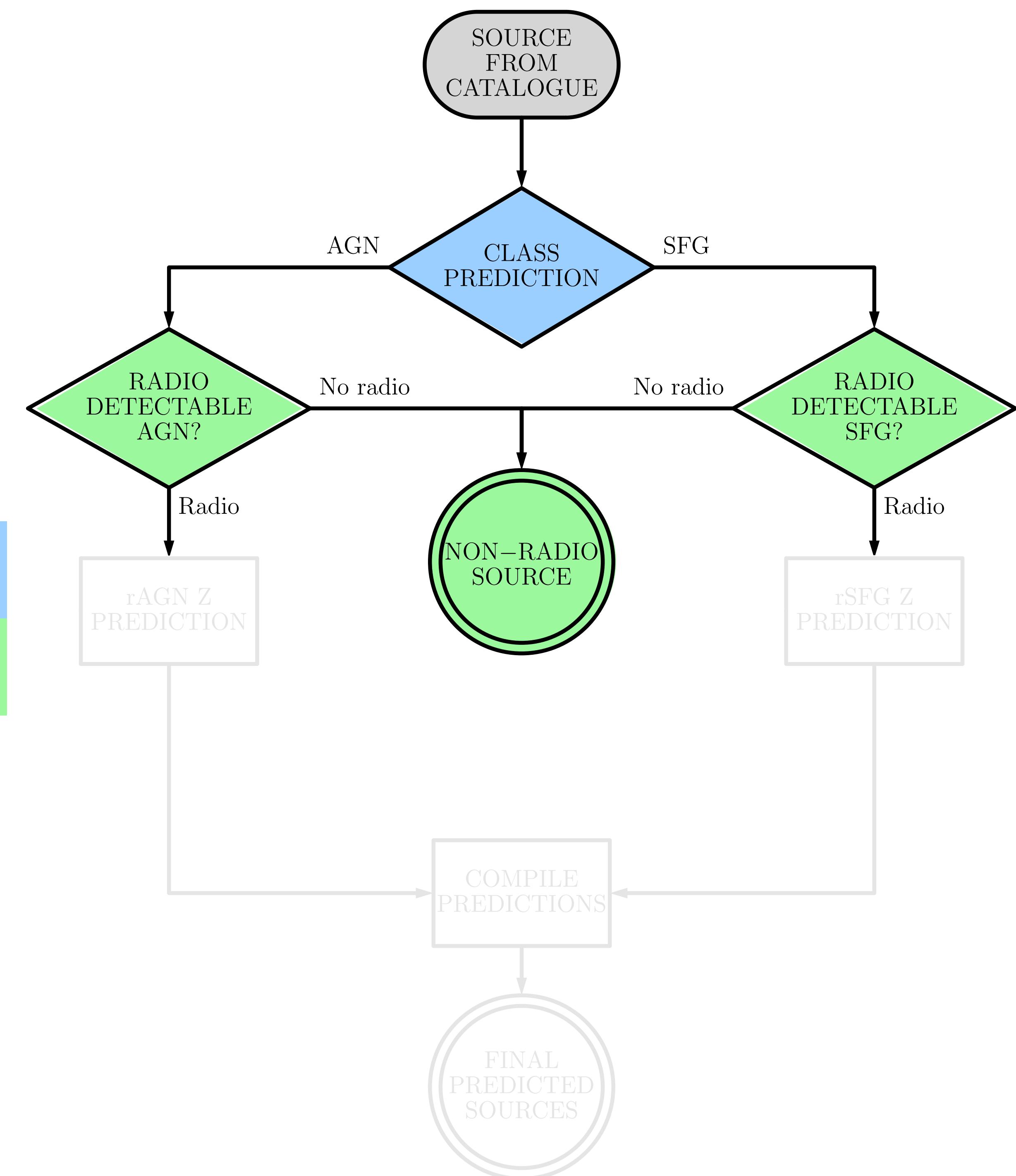
OUR APPROACH TO SELECTING AGN AND SFG WITH ML

Three different levels

Classify as AGN or SFG

Select radio-detectable sources

Estimate redshift for radio-detectable sources



PREDICTION PIPELINE

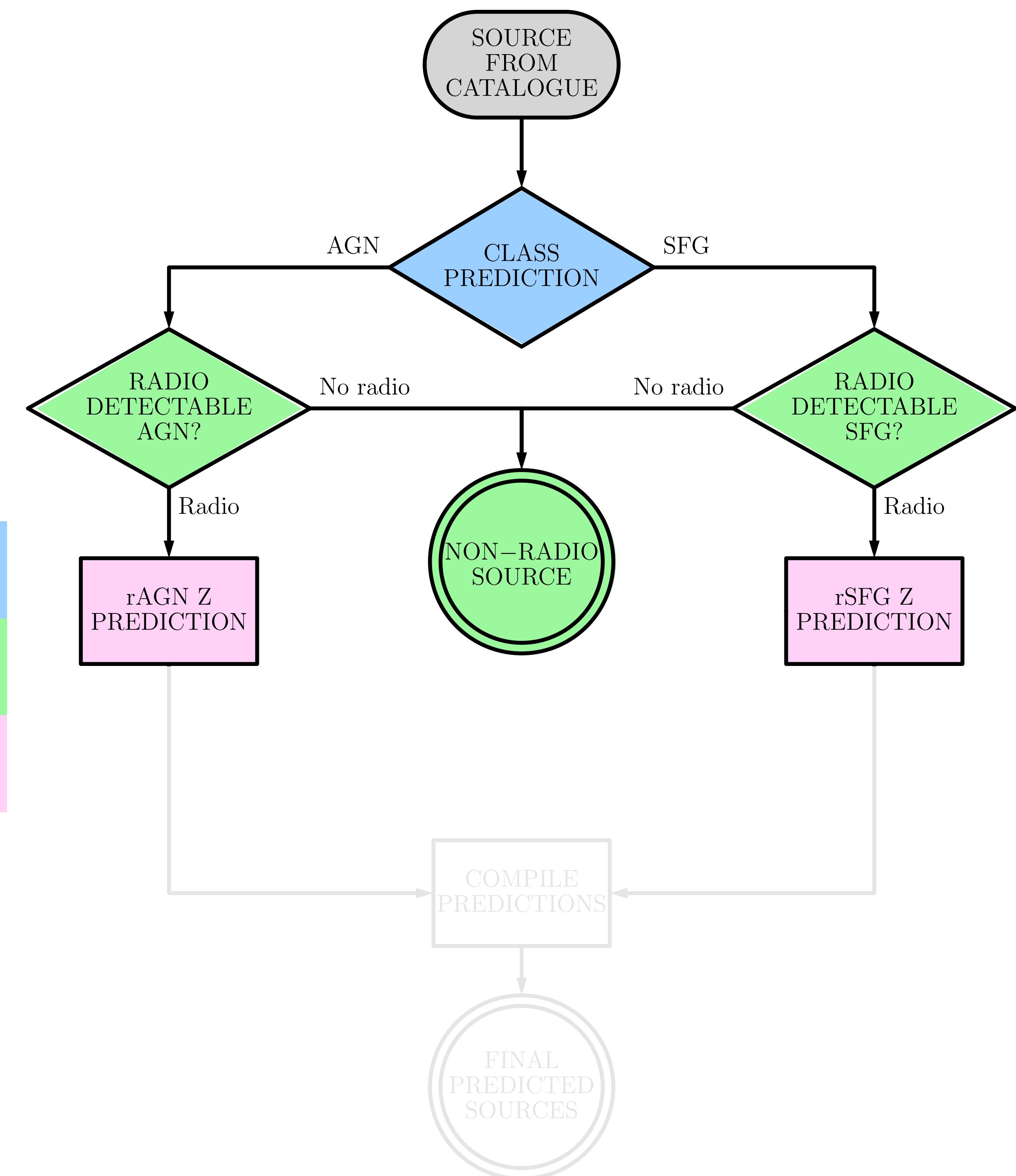
OUR APPROACH TO SELECTING AGN AND SFG WITH ML

Three different levels

Classify as AGN or SFG

Select radio-detectable sources

Estimate redshift for radio-detectable sources



PREDICTION PIPELINE

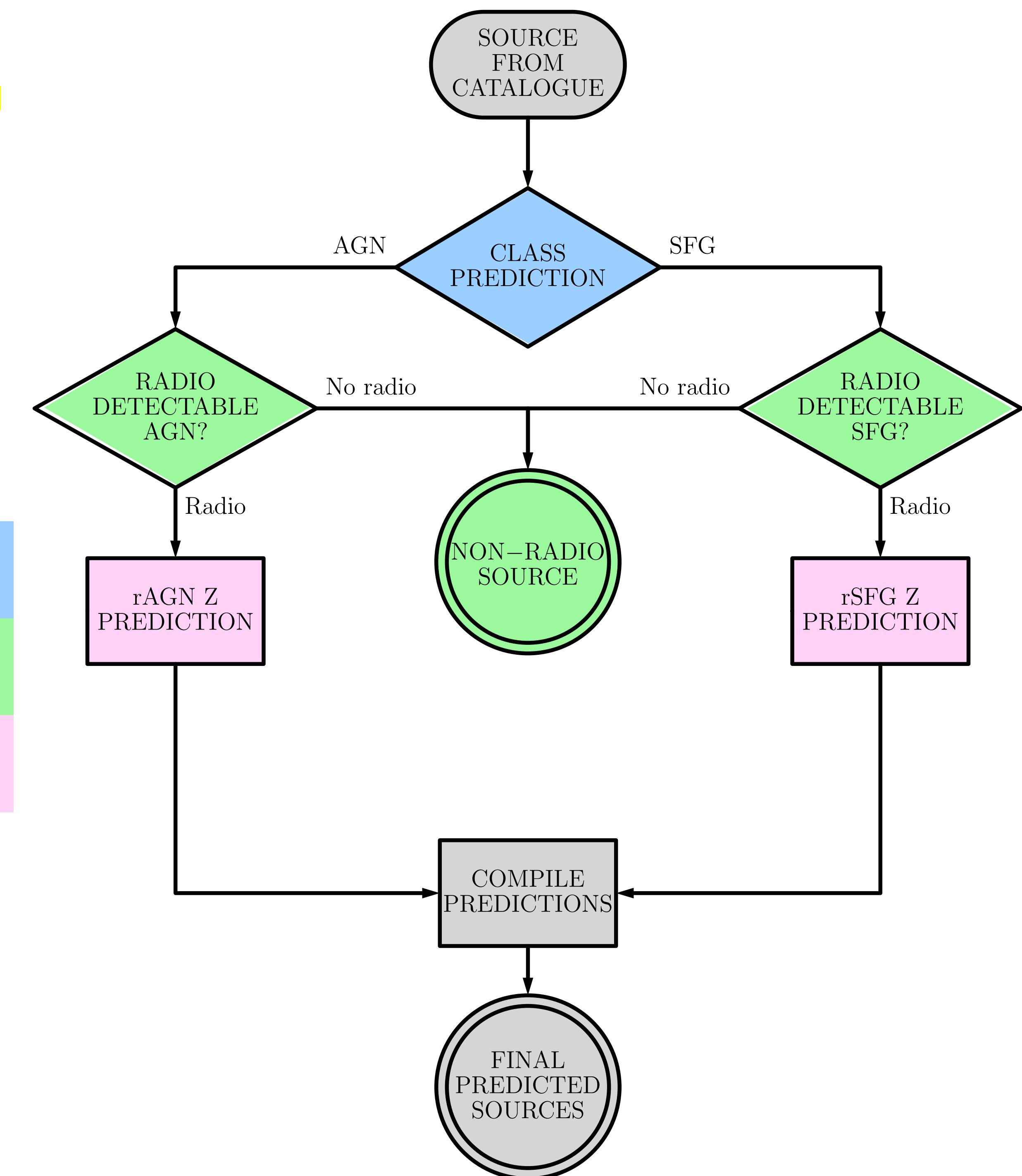
OUR APPROACH TO SELECTING AGN AND SFG WITH ML

Three different levels

Classify as AGN or SFG

Select radio-detectable sources

Estimate redshift for radio-detectable sources



PREDICTION PIPELINE

OUR APPROACH TO SELECTING AGN AND SFG WITH ML

Three different levels

Classify as AGN or SFG

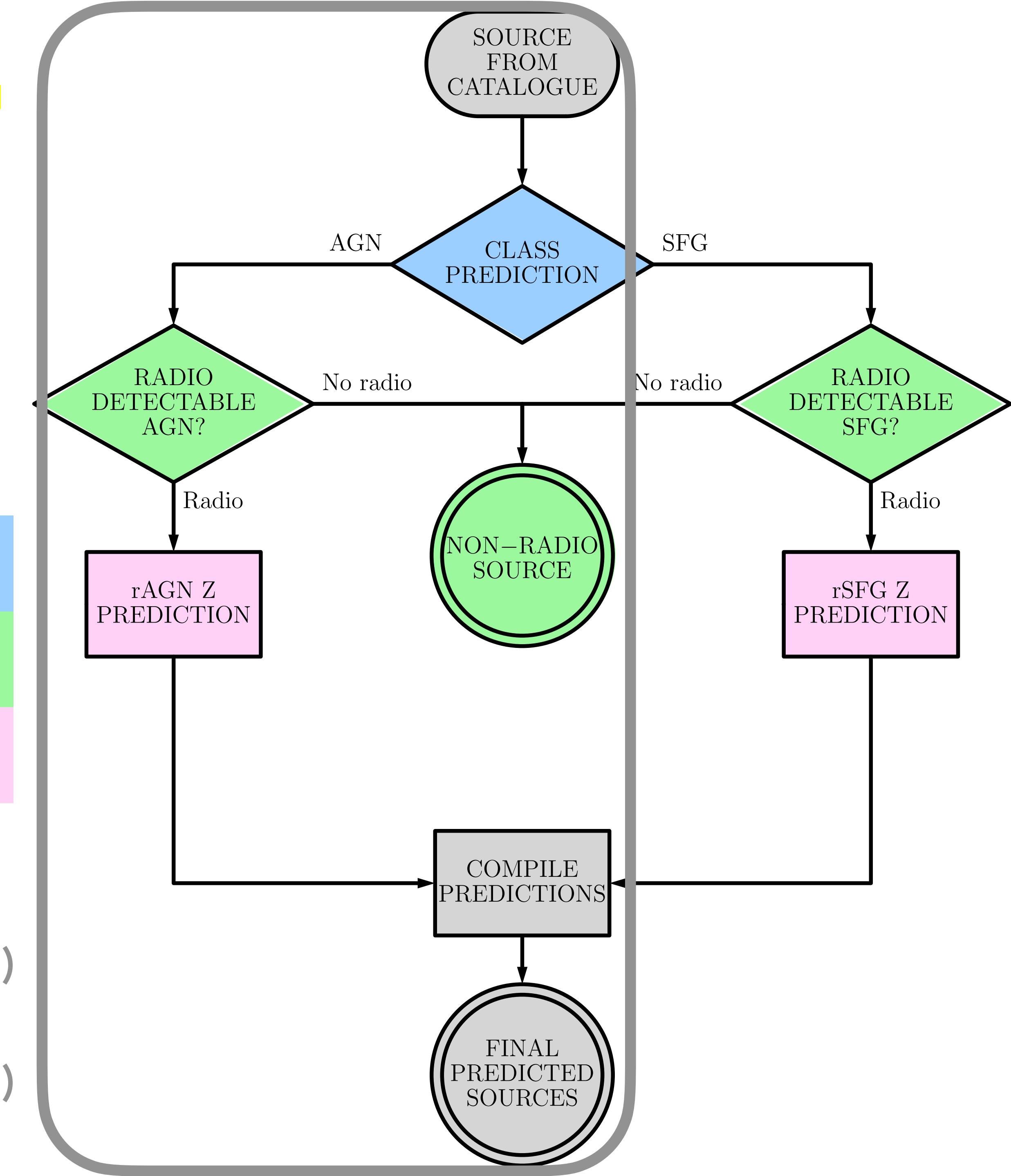
Select radio-detectable sources

Estimate redshift for radio-detectable sources

Carvajal et al. (2021)

+

Carvajal et al. (2023)



PREDICTION PIPELINE

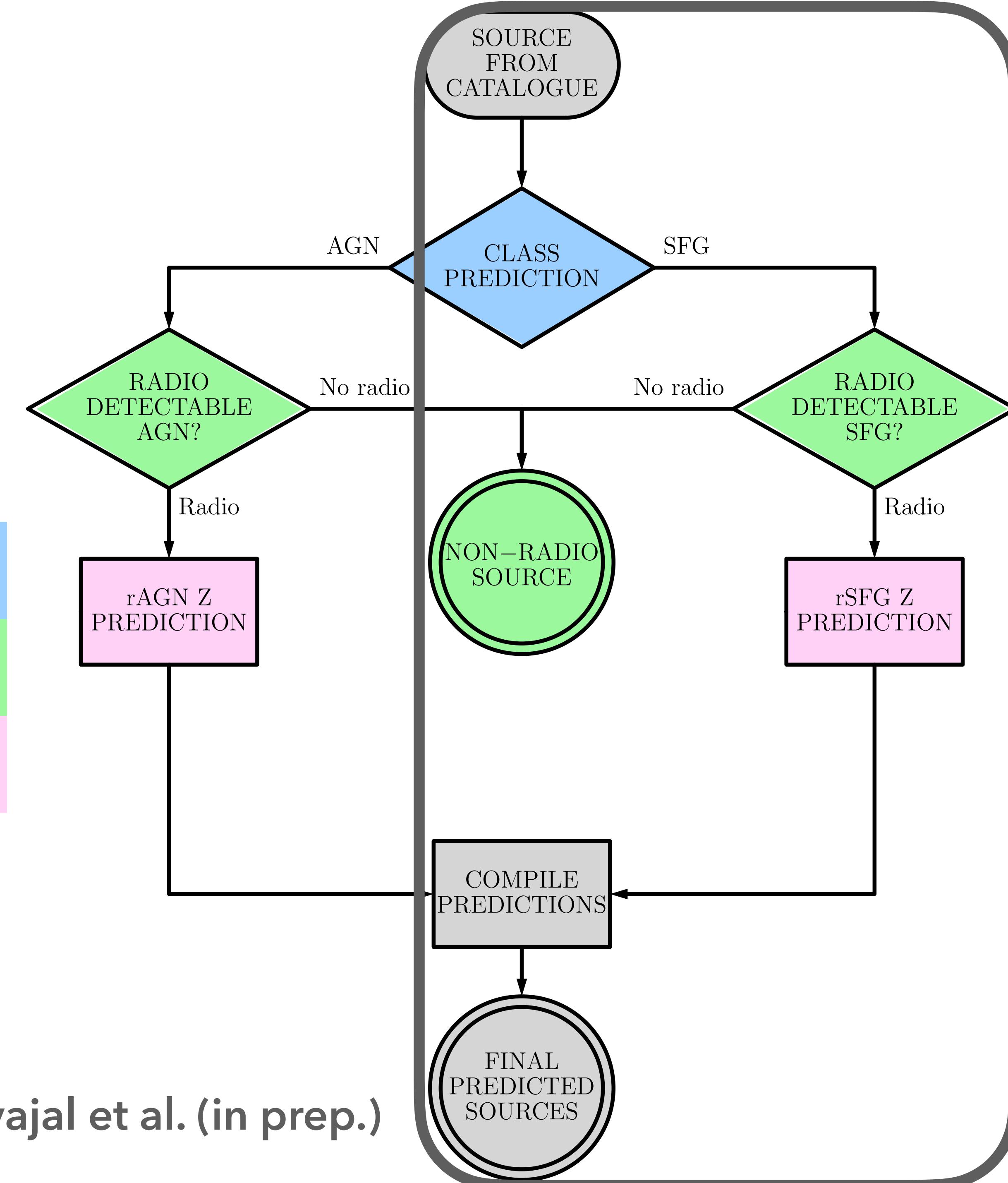
OUR APPROACH TO SELECTING AGN AND SFG WITH ML

Three different levels

Classify as AGN or SFG

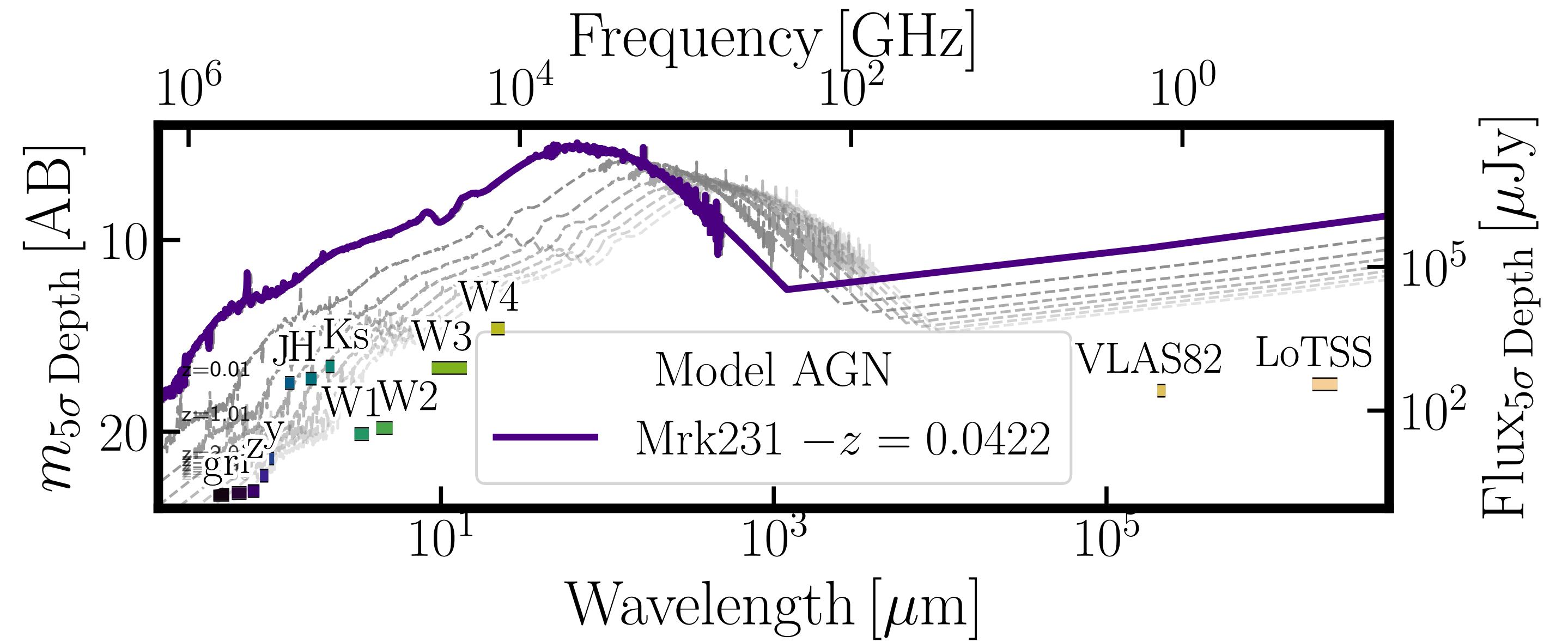
Select radio-detectable sources

Estimate redshift for radio-detectable sources



OUR DATA

DATASET – PHOTOMETRY



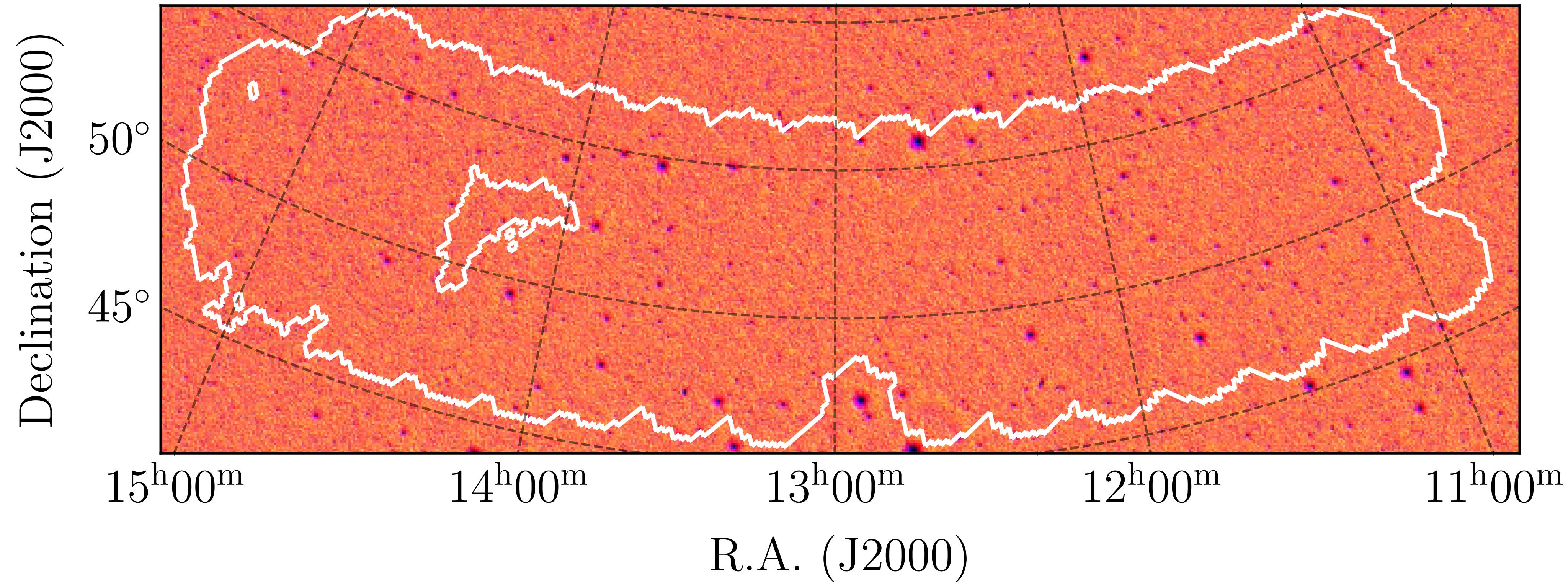
Carvajal et al. (2023)

Base catalogue: CatWISE2020 (Marocco et al. 2021, W1, W2)

Counterparts from: Pan-STARRS DR1, 2MASS, AllWISE

Colours from all bands (selected subset per model)

Target labels: class (AGN or SFG), radio detection, redshift



HETDEX SPRING FIELD

424 deg² covered by LoTSS-DR1 @ 144 MHz, 71 μJy, 6'' resolution

~15 million CatWISE2020 detections (~190k with LoTSS counterpart, 1%)

~50k spec. confirmed AGN (~6.4k radio, 13%) + ~70k spec. confirmed SFGs (~6.6k radio, 9%)

MODELS' RESULTS

**AGN SELECTION
COMPLETENESS: 96 %**
INCREASE FROM BASELINE AGN FRACTION OF 43%

**SFG SELECTION
COMPLETENESS: 96 %**
INCREASE FROM BASELINE SFG FRACTION OF 57%

True Classes	SFG	AGN	
SFG	13 072	567	13 639
AGN	383	9725	9768
Predicted Classes			

RADIO SELECTION IN AGN COMPLETENESS: 52 %

INCREASE FROM BASELINE RADIO FRACTION IN AGN OF 13%

True Classes
Radio No-Radio

7568	1242
621	677

No-Radio Radio
Predicted Classes

RADIO SELECTION IN SFG COMPLETENESS: 46 %

INCREASE FROM BASELINE RADIO FRACTION IN SFGs OF 9%

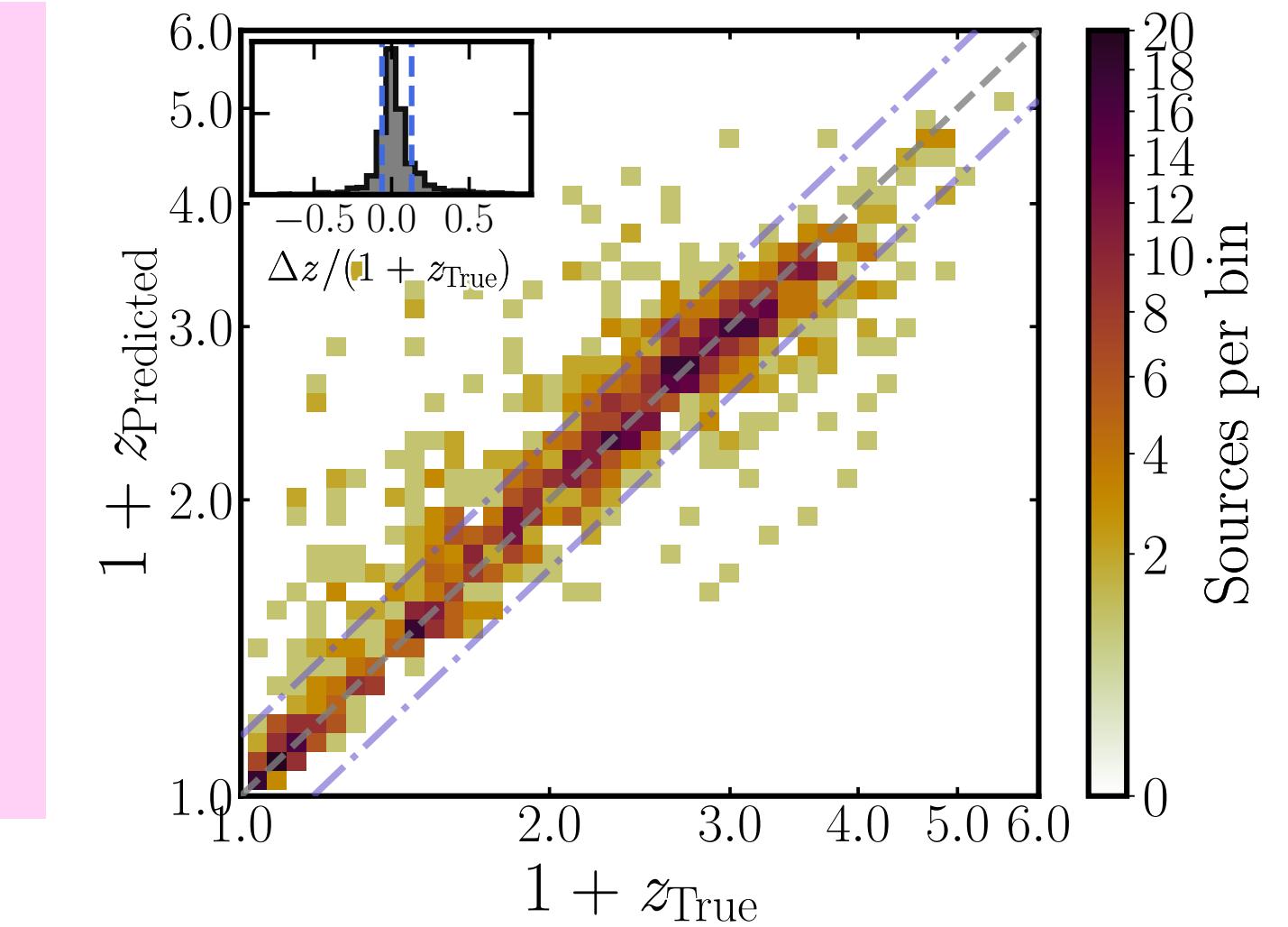
True Classes
Radio No-Radio

10566	1069
1089	915

No-Radio Radio
Predicted Classes

REDSHIFT IN RAGN: 81 % ACCURATE

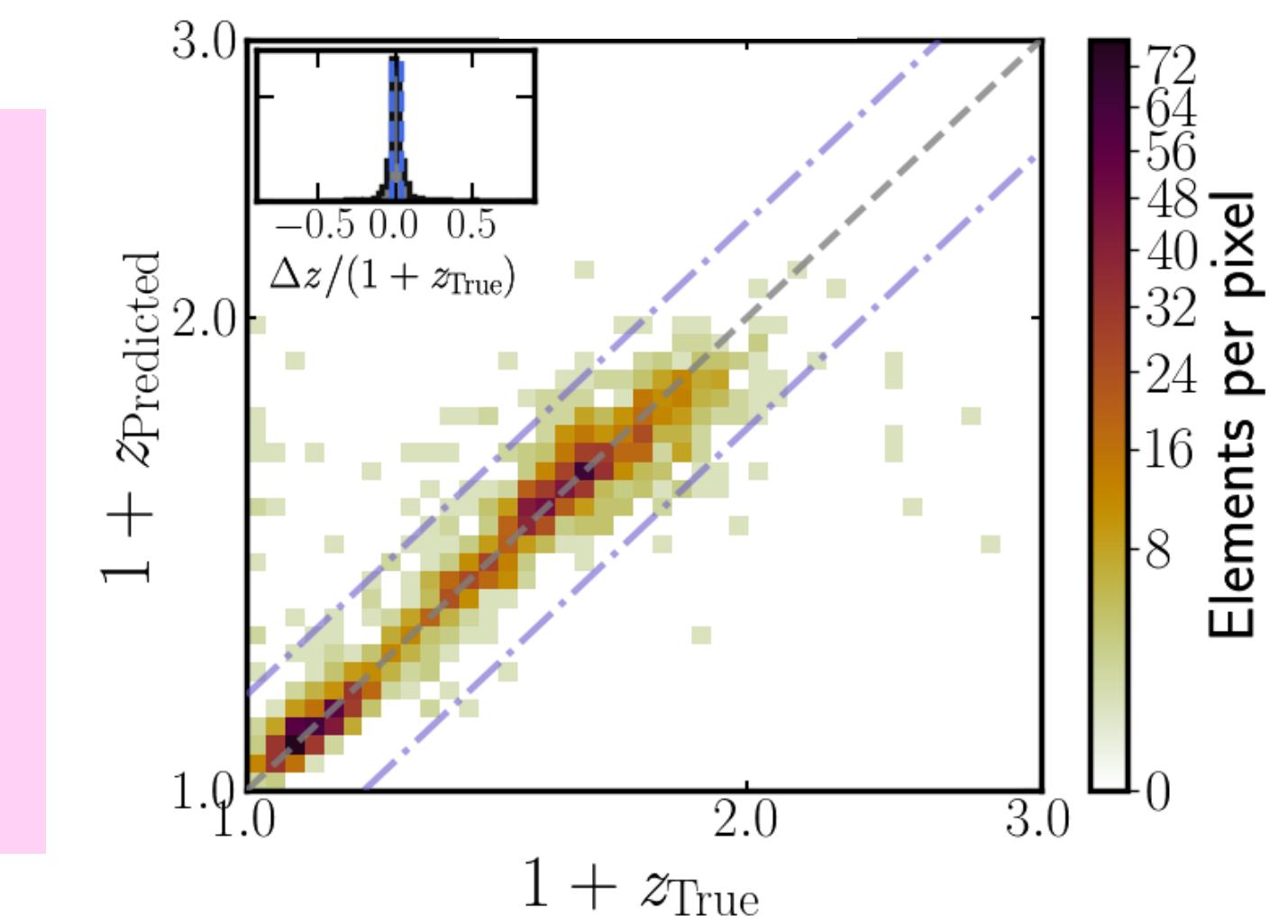
OUTLIER FRACTION OF 19%



Carvajal et al. (2023)

REDSHIFT IN RSFG: 97 % ACCURATE

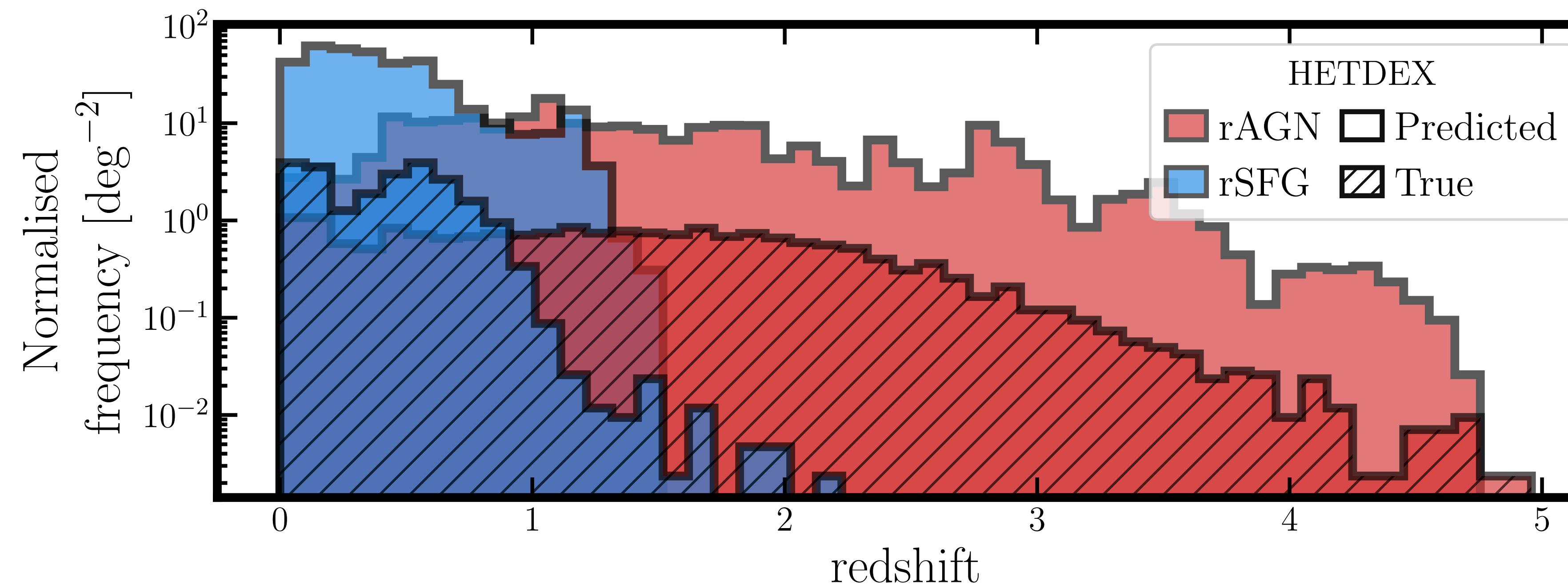
OUTLIER FRACTION OF 3%



APPLYING PIPELINE TO FULL HETDEX DATASET

NEW SAMPLE OF 68K RAGN CANDIDATES 115K RSFG CANDIDATES

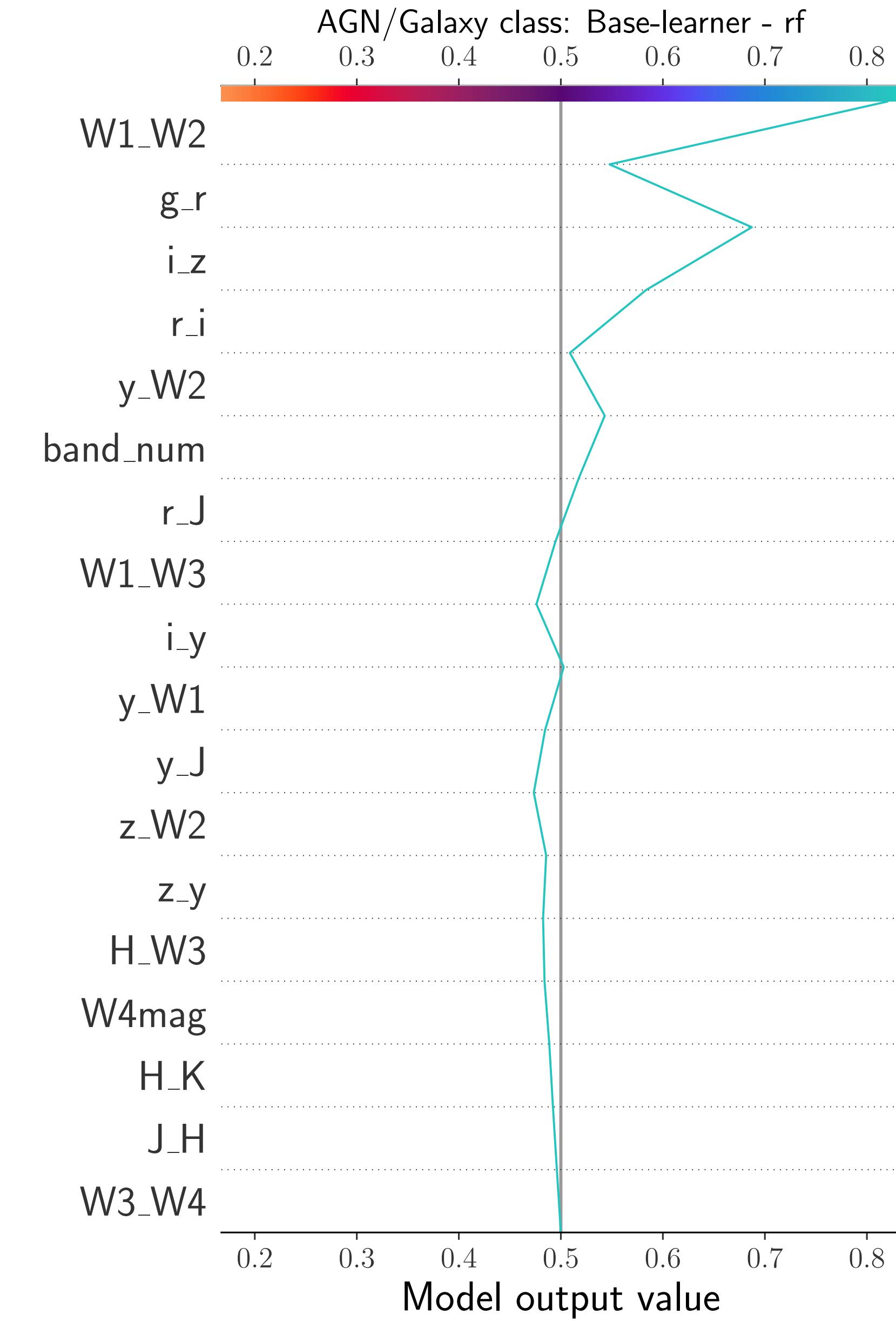
MORE THAN 10x ORIGINAL SAMPLE (6.4K RAGN AND 6.6K RSFG)



UNDERSTANDING OUR PREDICTIONS

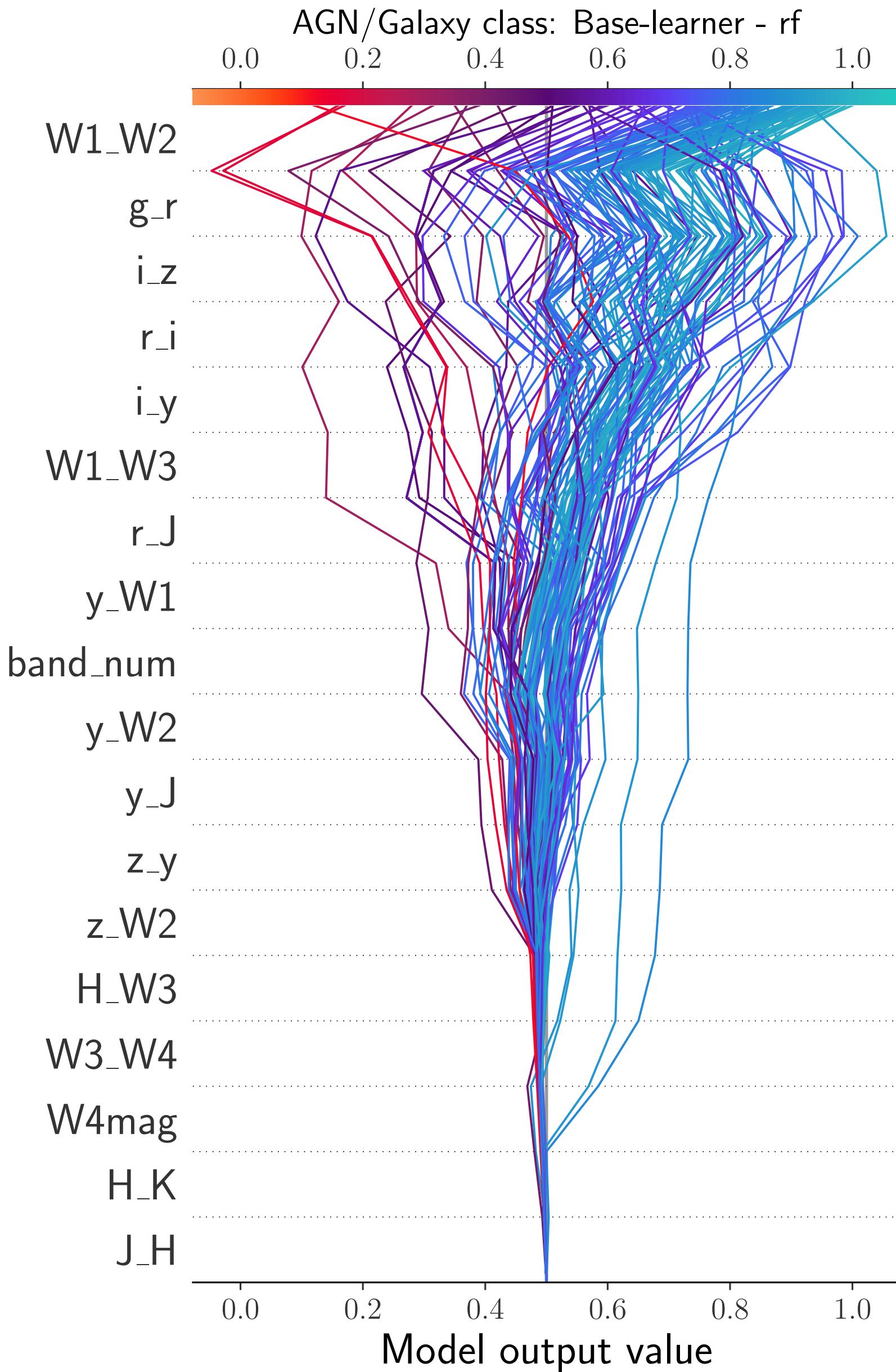
FEATURE IMPORTANCE

UNDERSTAND WHICH FEATURES DRIVE
PREDICTIONS MORE STRONGLY



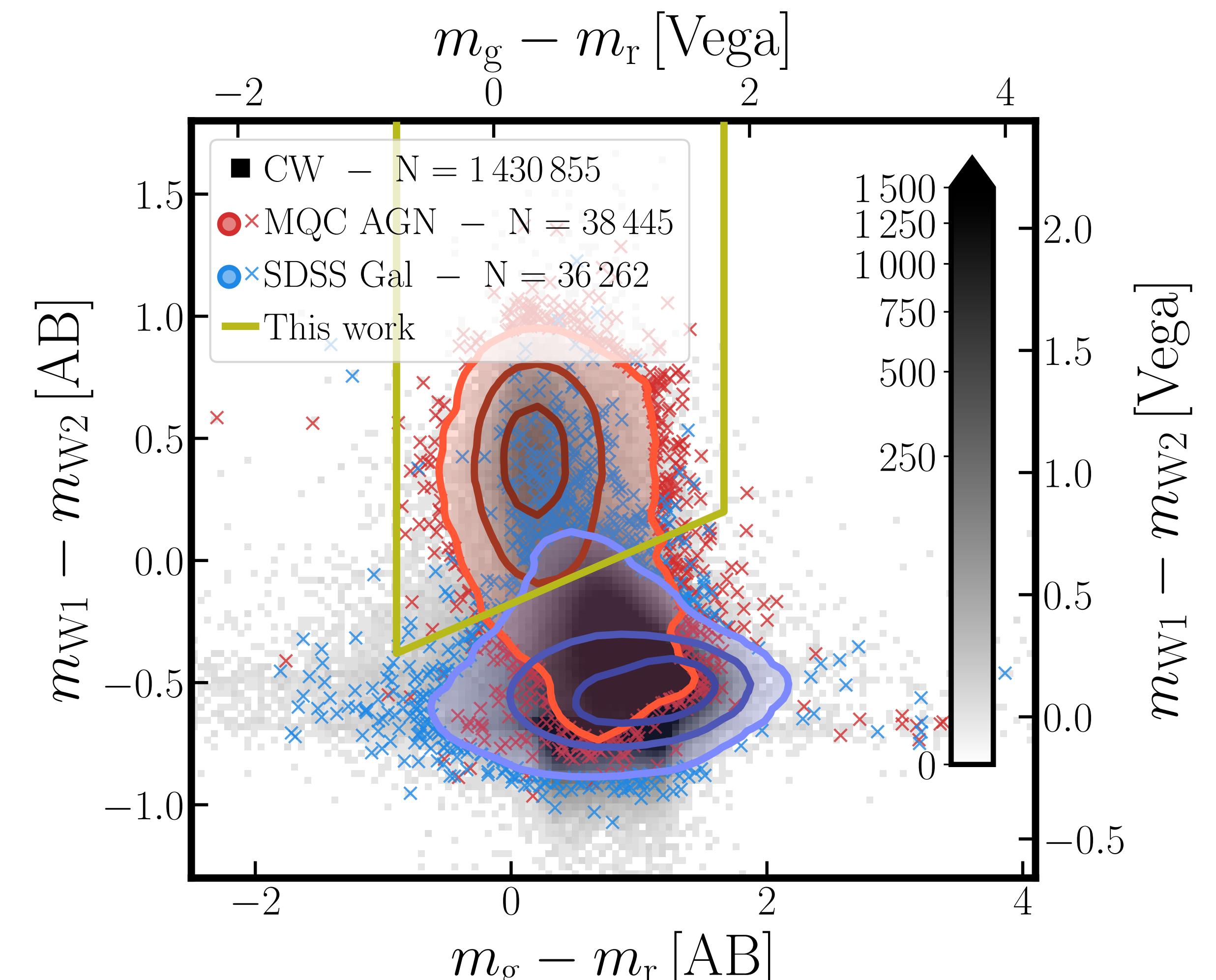
FEATURE IMPORTANCE

UNDERSTAND WHICH FEATURES DRIVE
PREDICTIONS MORE STRONGLY



ML-BASED COLOUR-COLOUR DIAGRAM

g-r – W1-W2 DIAGRAM

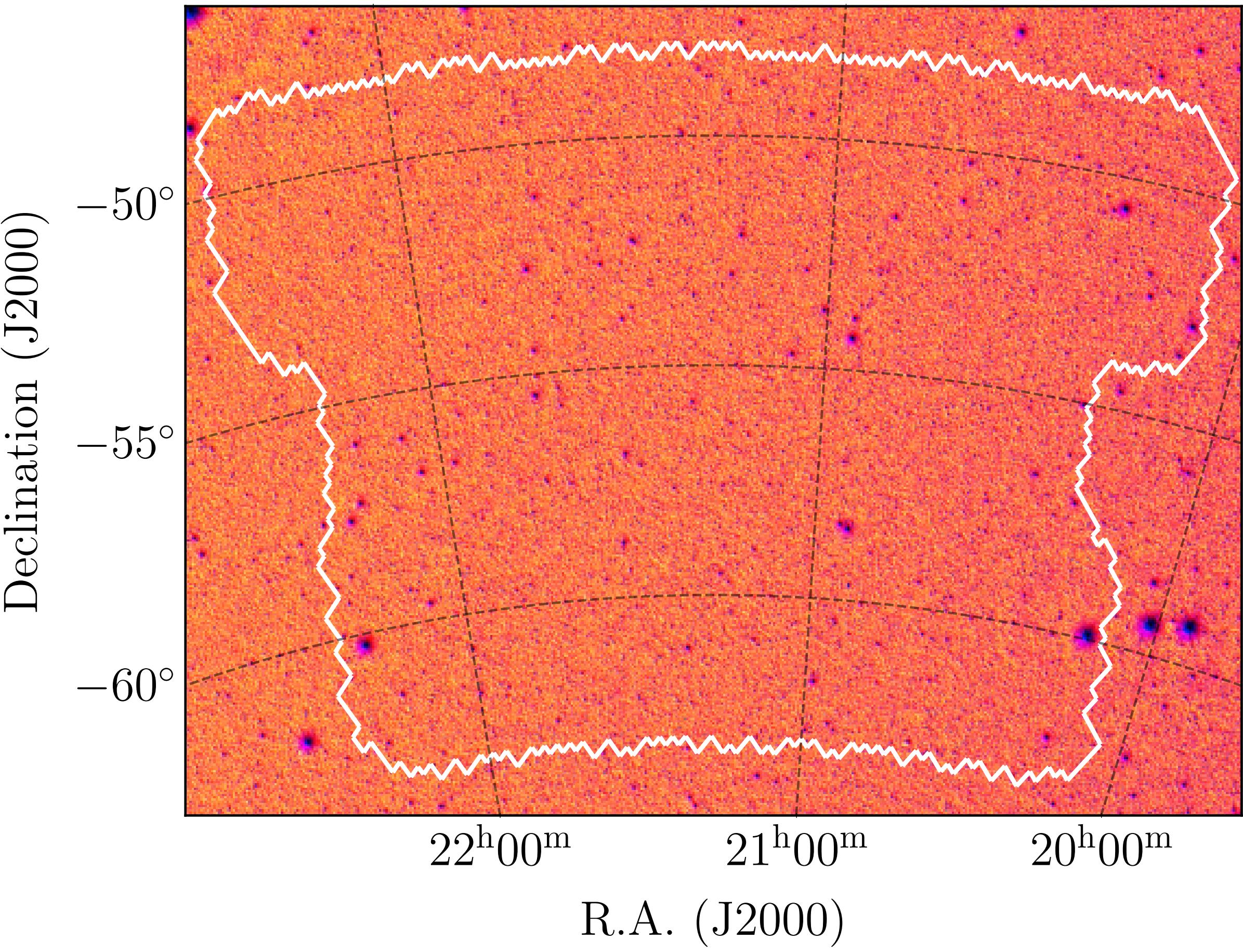


Carvajal et al. (2023)

APPLYING PIPELINE ELSEWHERE?

APPLY PIPELINE IN THE SOUTHERN SKY

EMU Pilot Survey (EMU-PS) – 270 deg²
– 18 arcsec resolution @ 944 MHz –
25-30 μ Jy rms
SKA-like conditions



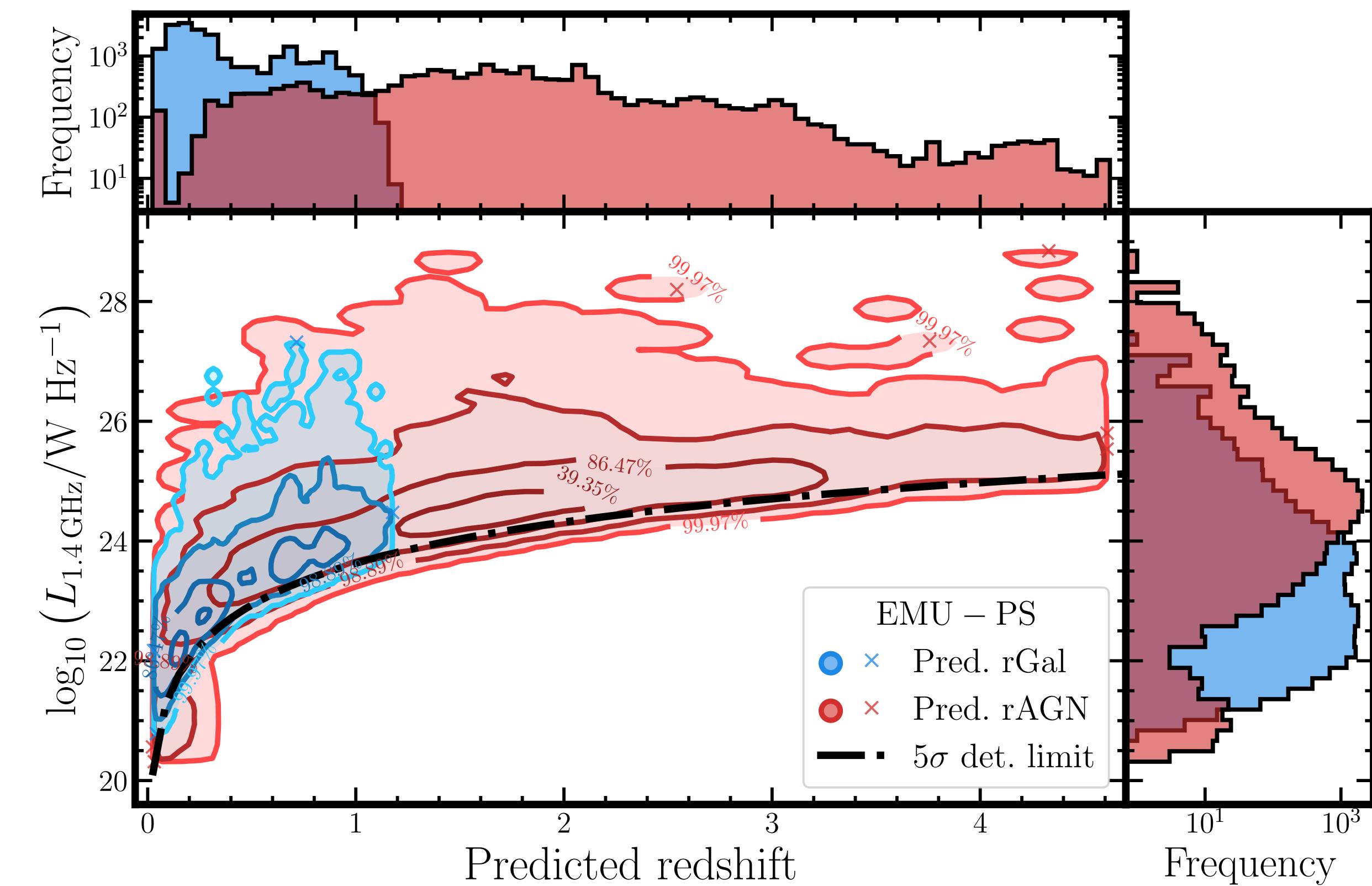
EMU-PS

More than 10x new rAGN and rSFG candidates

rAGN: 92113 (originally 2367)

rSFG: 128249 (originally 870)

Such numbers allow for population studies



RADIO LUMINOSITY FUNCTION

RADIO LUMINOSITY FUNCTION

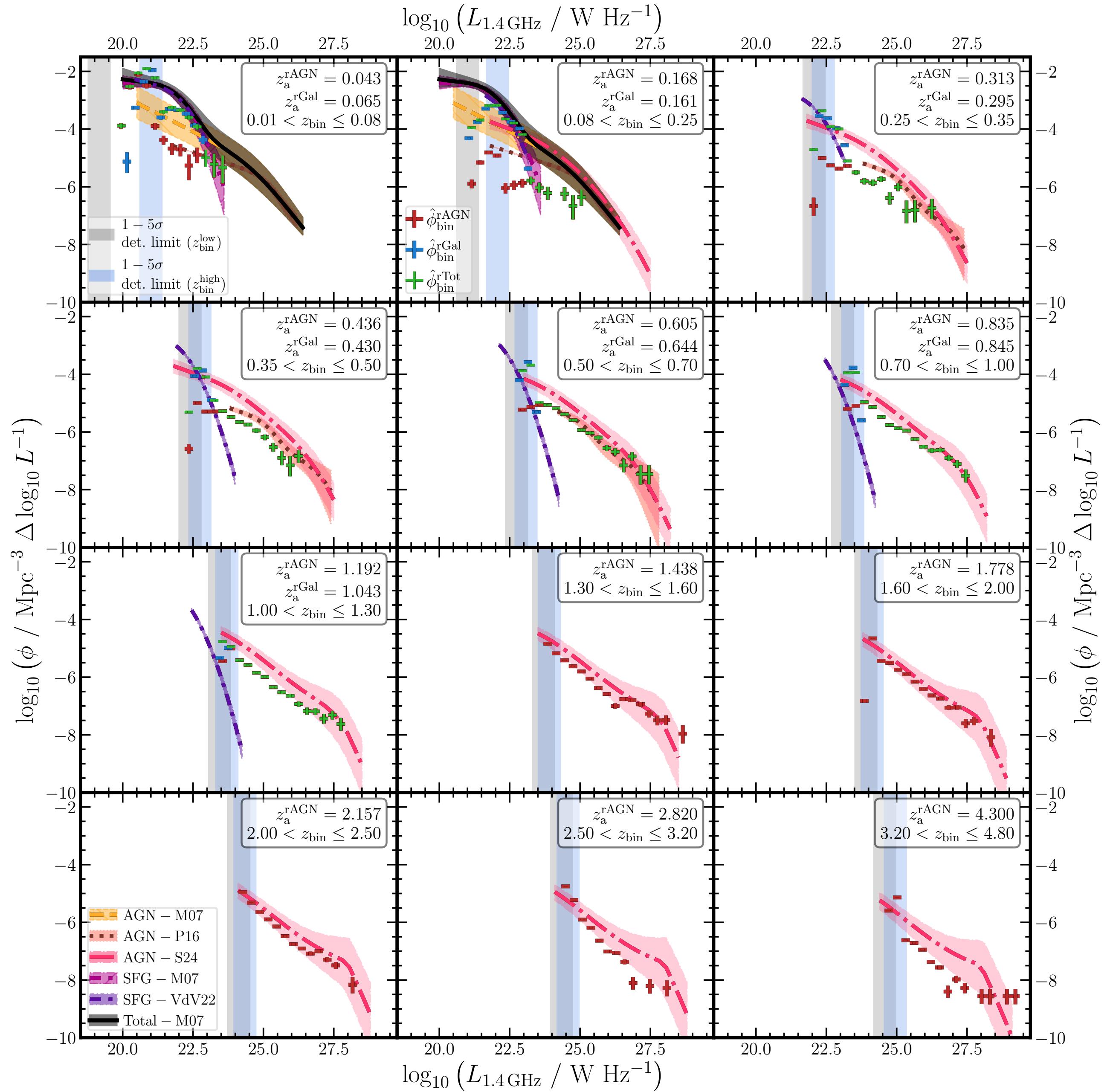
Number density candidate SFGs and AGN from pipeline in EMU-PS

Binned luminosity function ($1/V_{\max}$ approach, Page & Carrera, 2000)

Calculated over twelve redshift bins [0.1, 4.8]

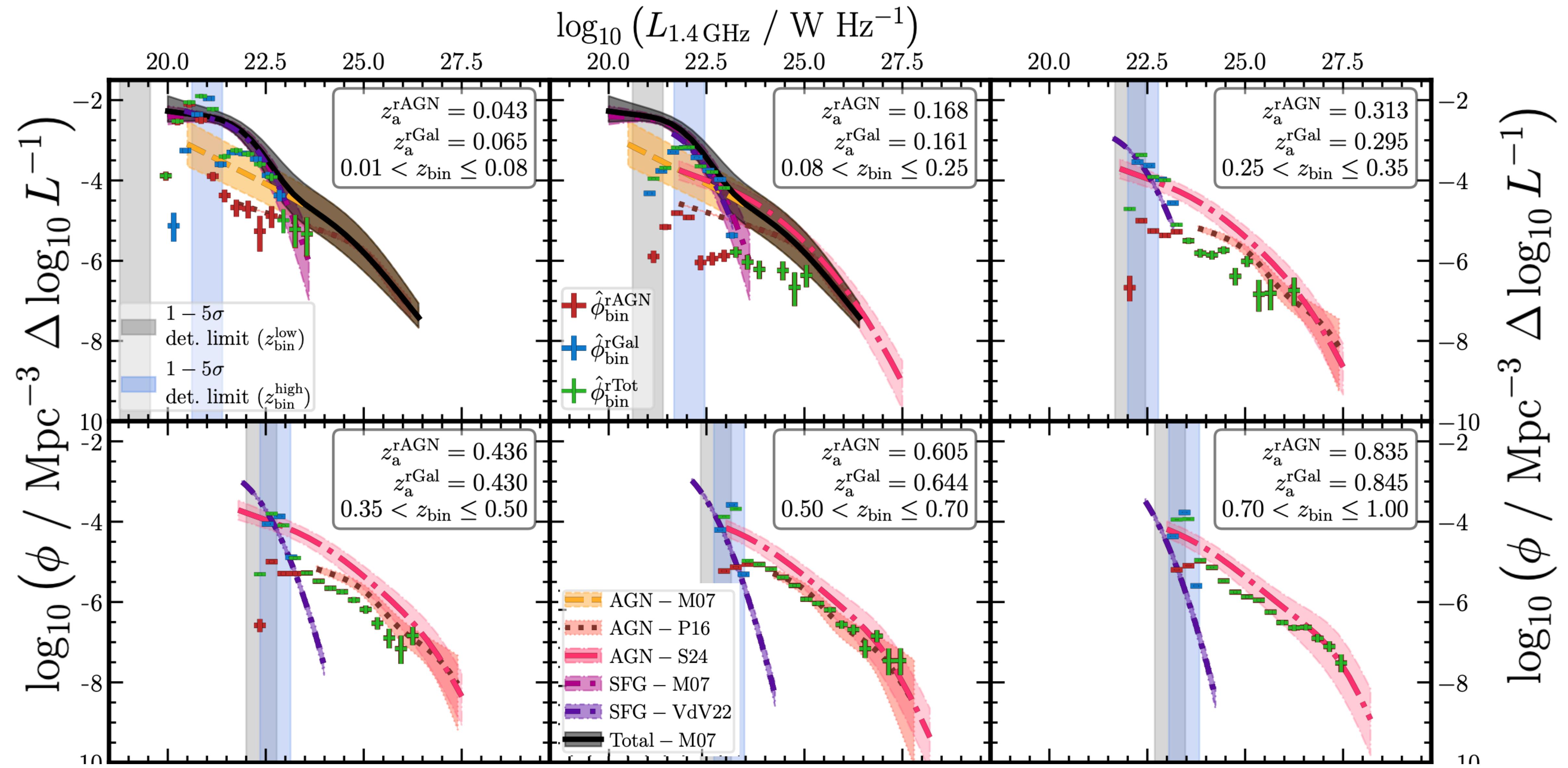
Correct individual counts by completeness and purity from pipeline

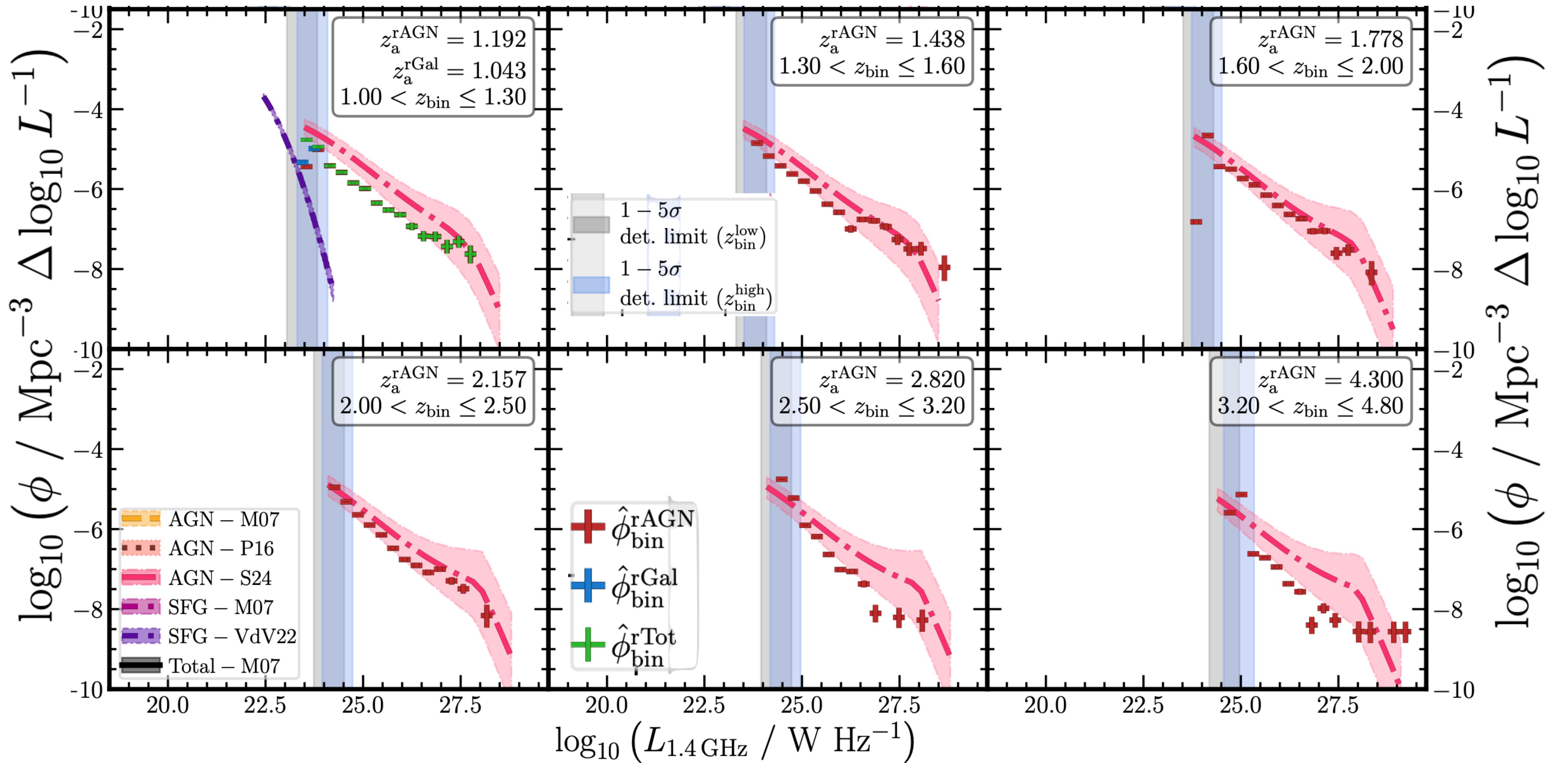
Compared with previous parametric functions



**Close to previous work only
from ML corrections**

**Small error bars from very
sample (at all z)**





IN CONCLUSION

CONCLUDING REMARKS

Want to determine contribution of AGN into galaxy evolution

Clear characterisation of sources is needed: class + redshift

Need to exploit large radio surveys for AGN and SFG studies

Machine-assisted pipeline to select rAGN + rSFGs candidates and redshift

Understand inner works of algorithms to extract physical insight

Increase sample sizes improving statistics

WHERE TO GO FROM HERE?

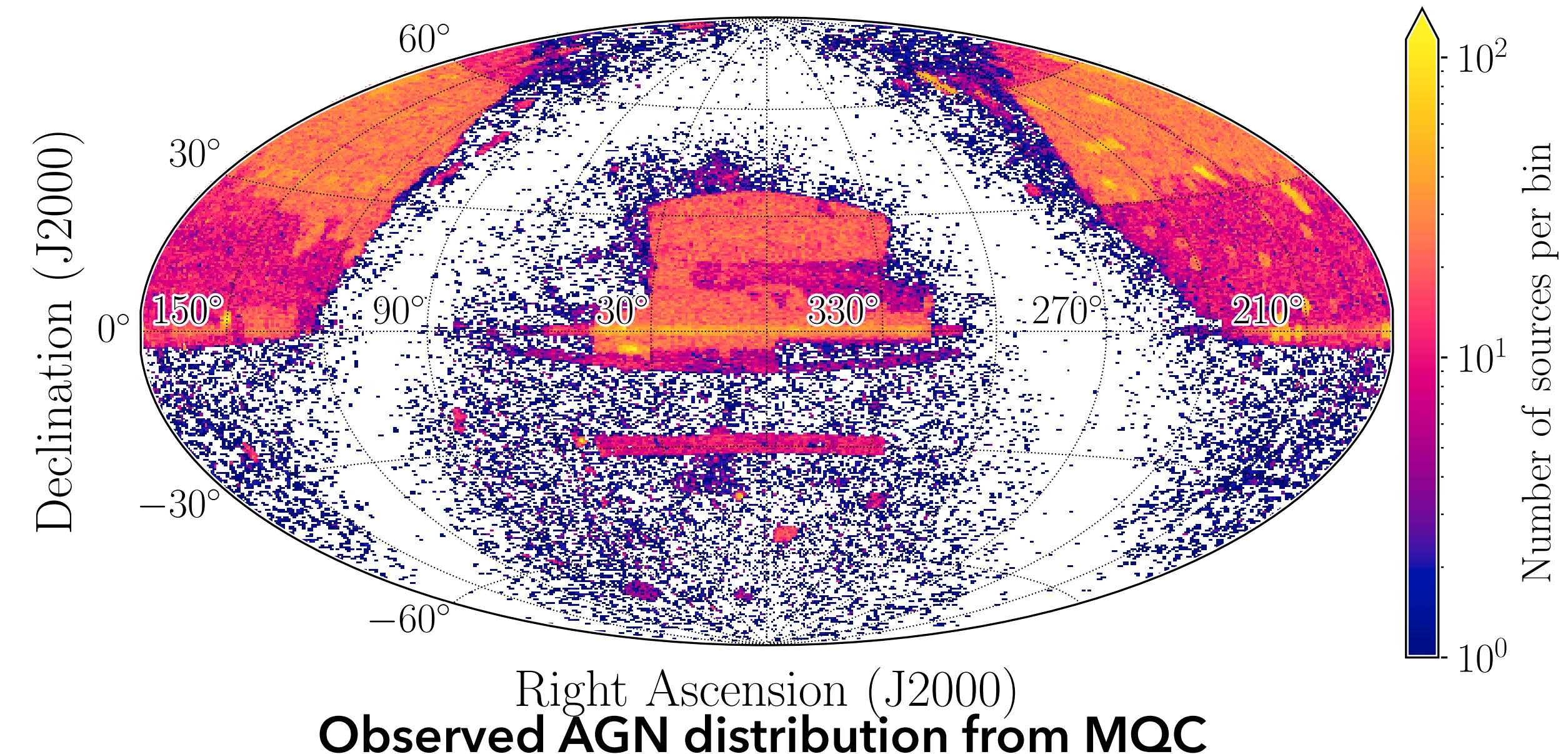
Need to go to the southern hemisphere!

AGN distribution strongly skewed to the north

SKA (and precursors: EMU, MIGTHEE) observations

Deeper surveys: VHS, KIDS-S, DES

Code/pipeline improvements





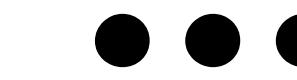
Ciências
ULisboa

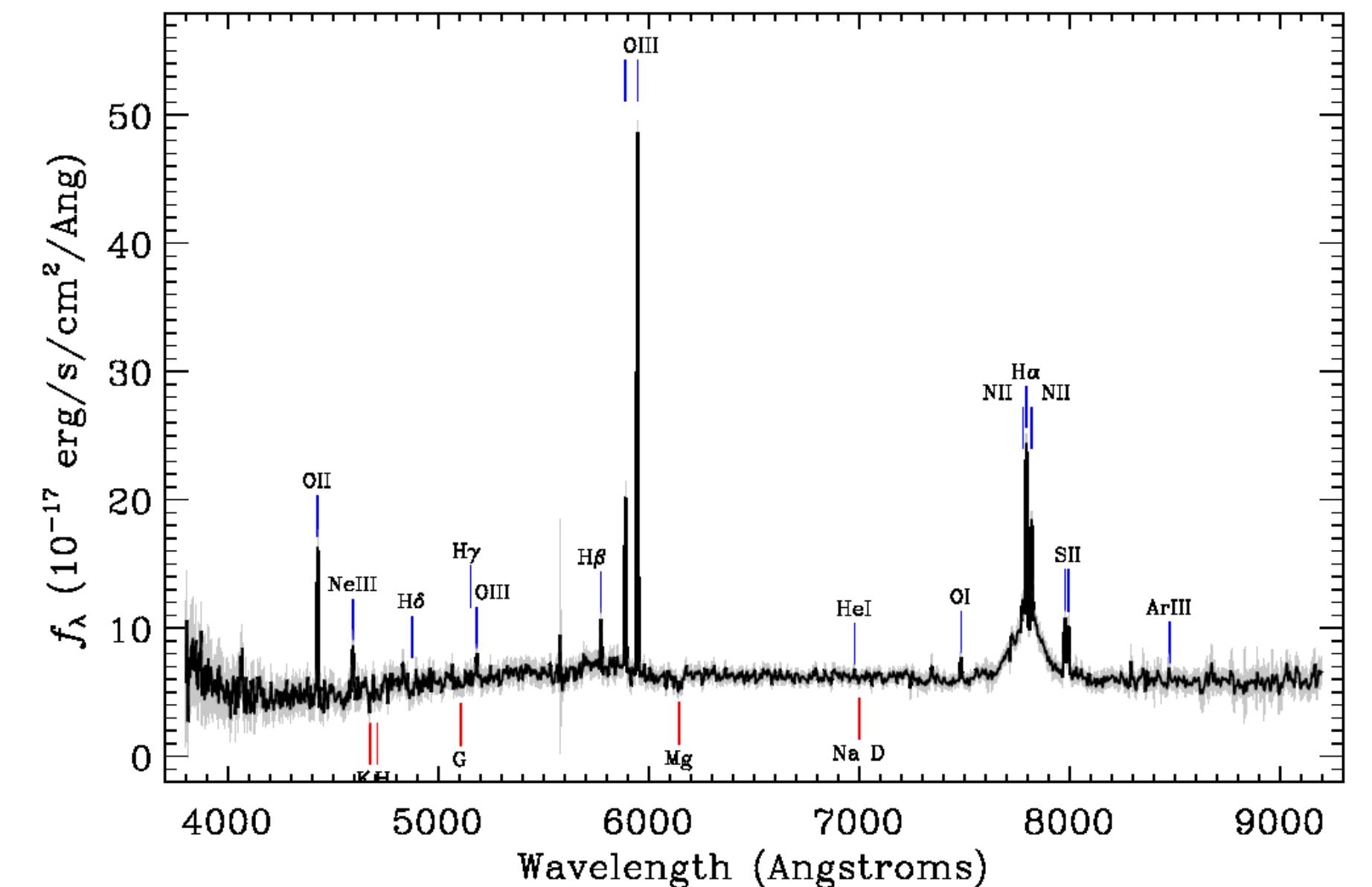
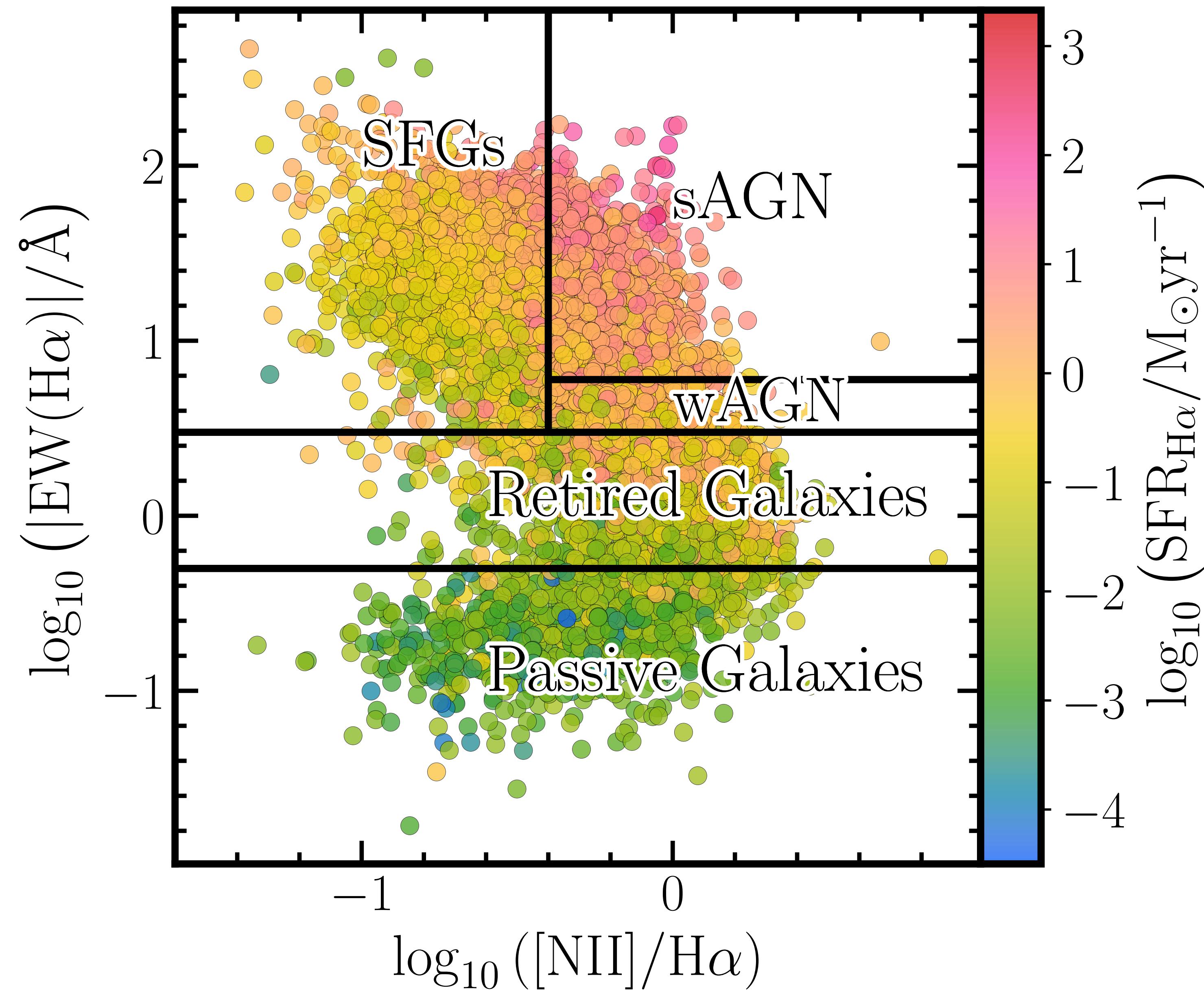


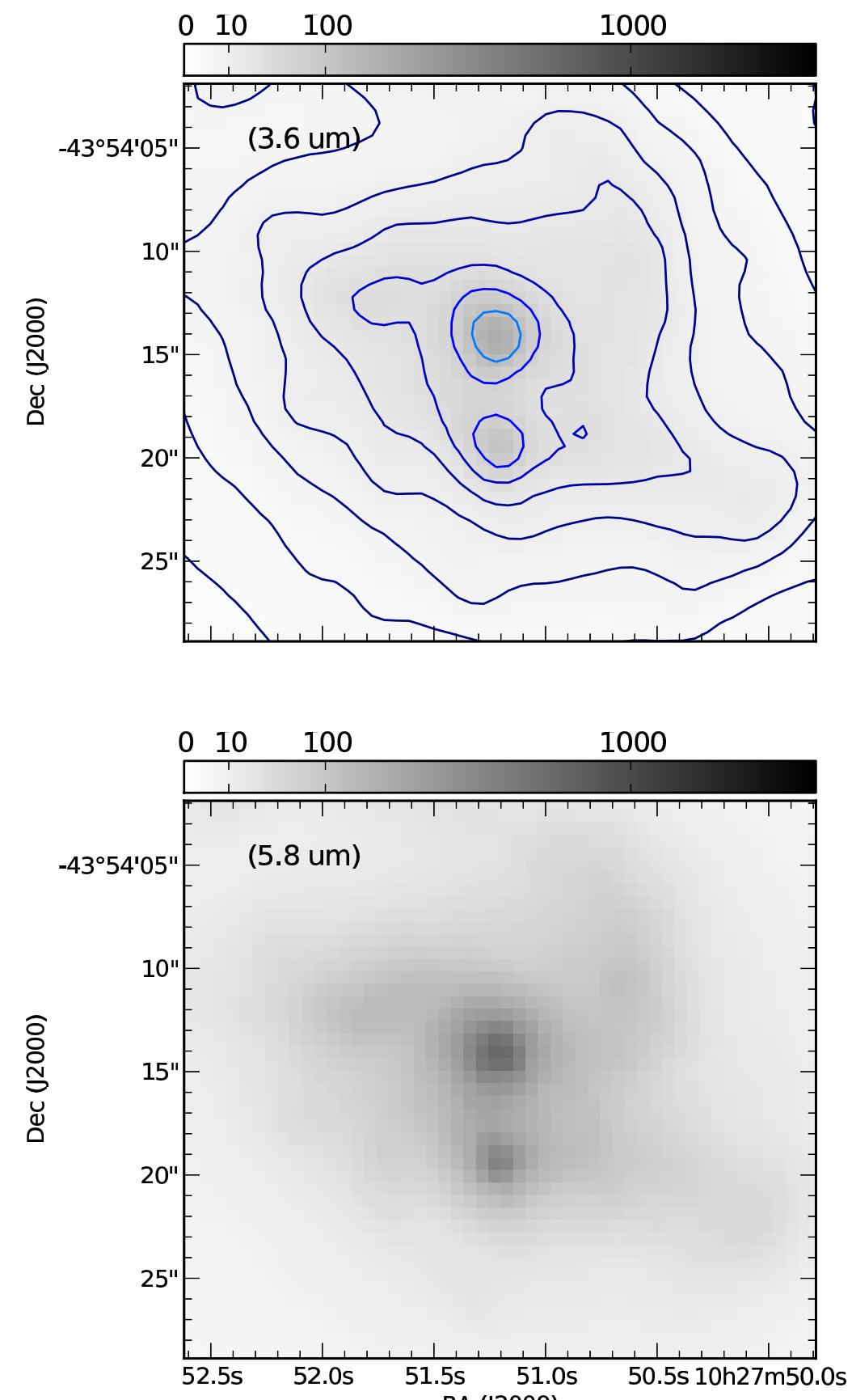
Towards Better Selection and Characterisation Criteria for High-Redshift Radio Galaxies Using Machine-Assisted Pattern Recognition

RODRIGO CARVAJAL

SUPERVISED BY
DR J. AFONSO
DR I. MATUTE
DR H. MESSIAS

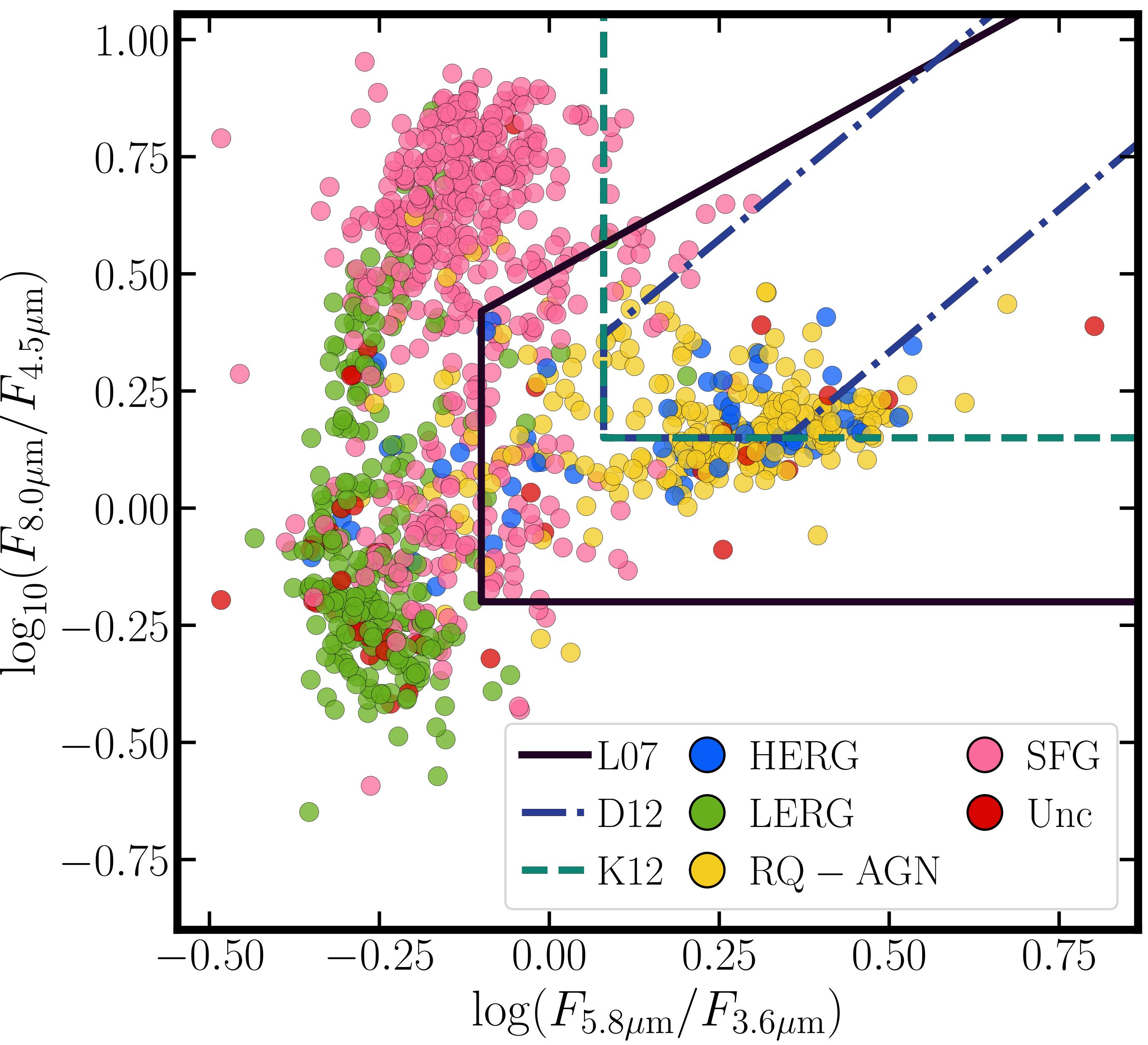
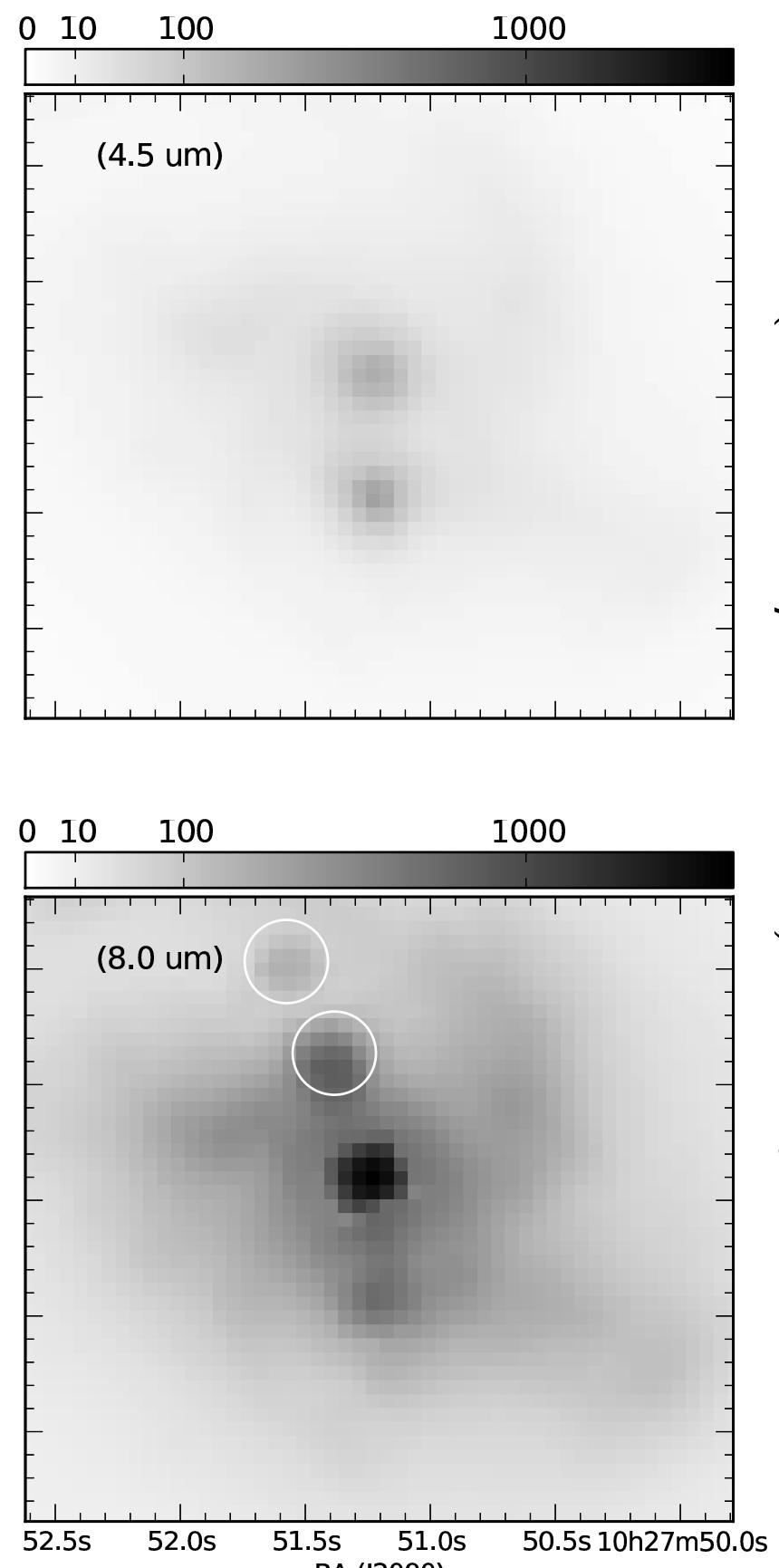


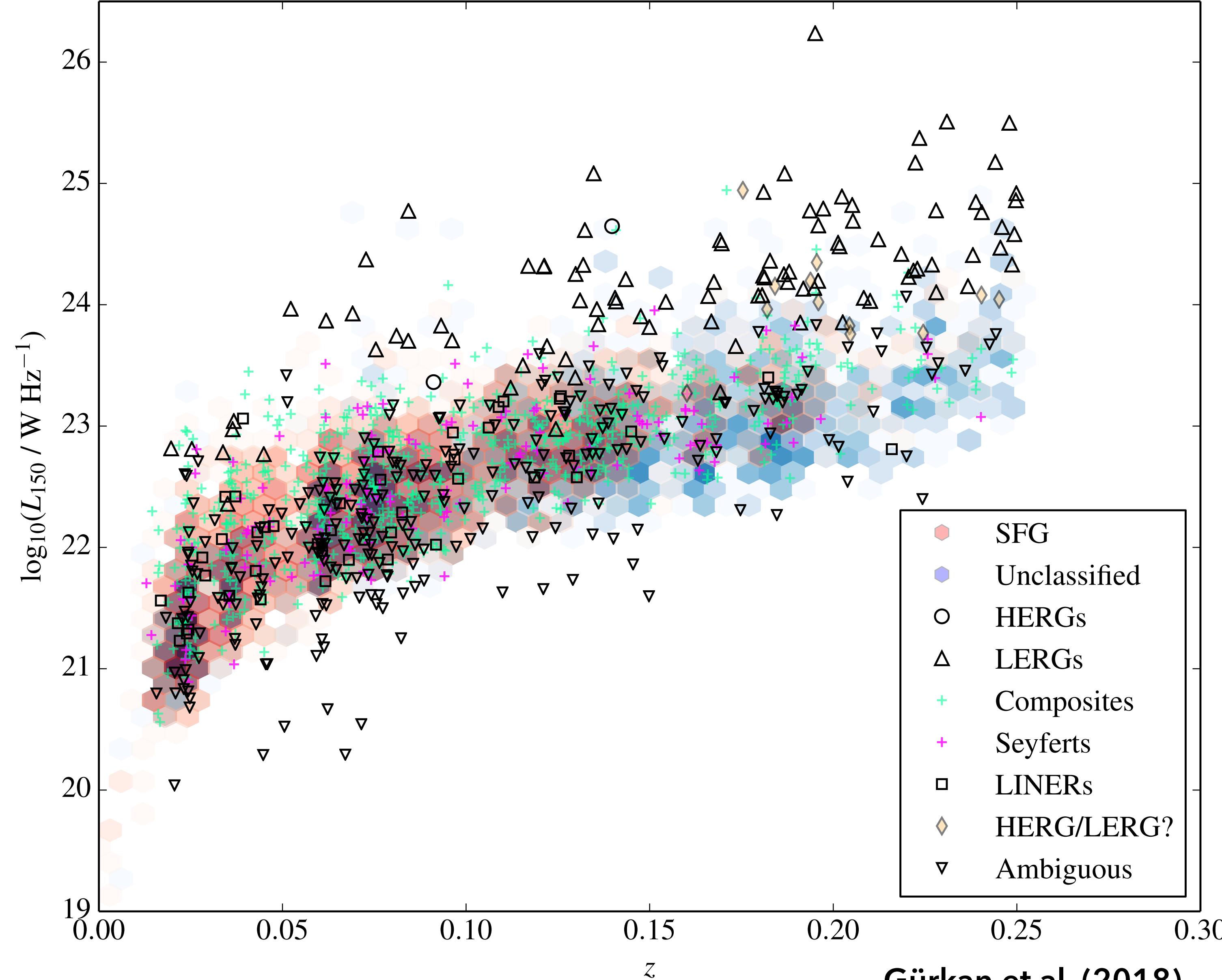




Ohyama et al. (2015)

Spitzer Colours PHOTOMETRY



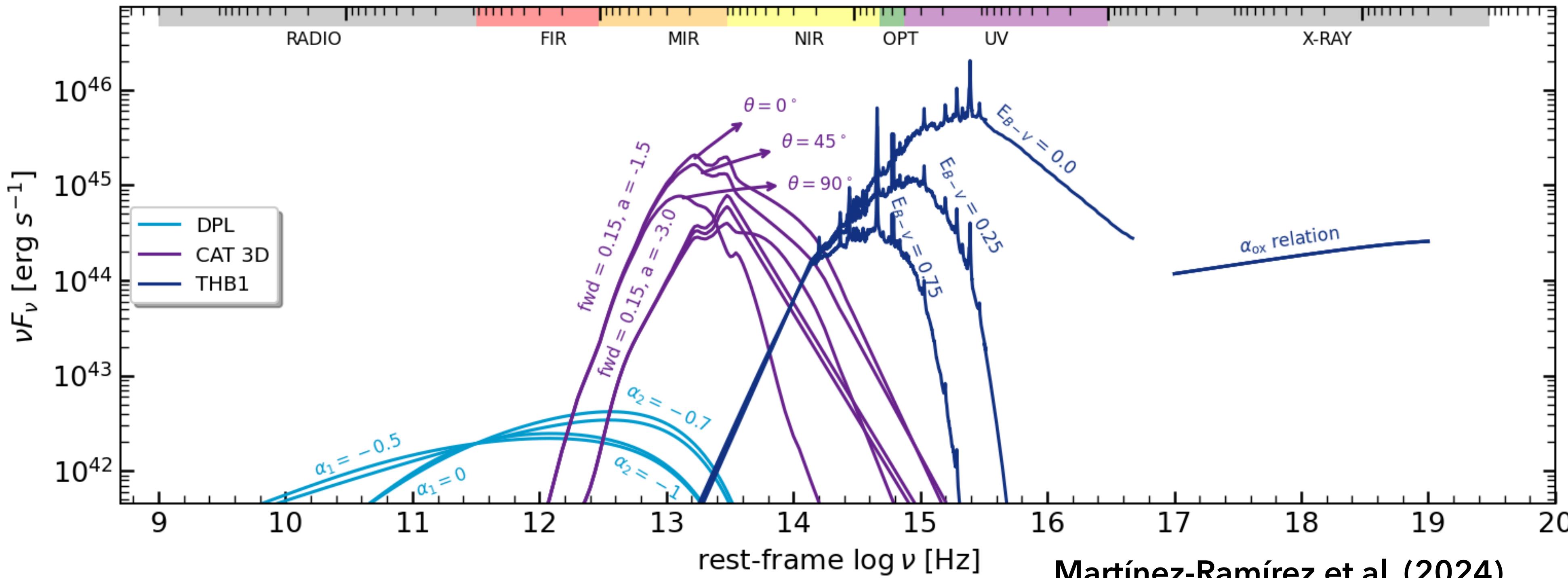
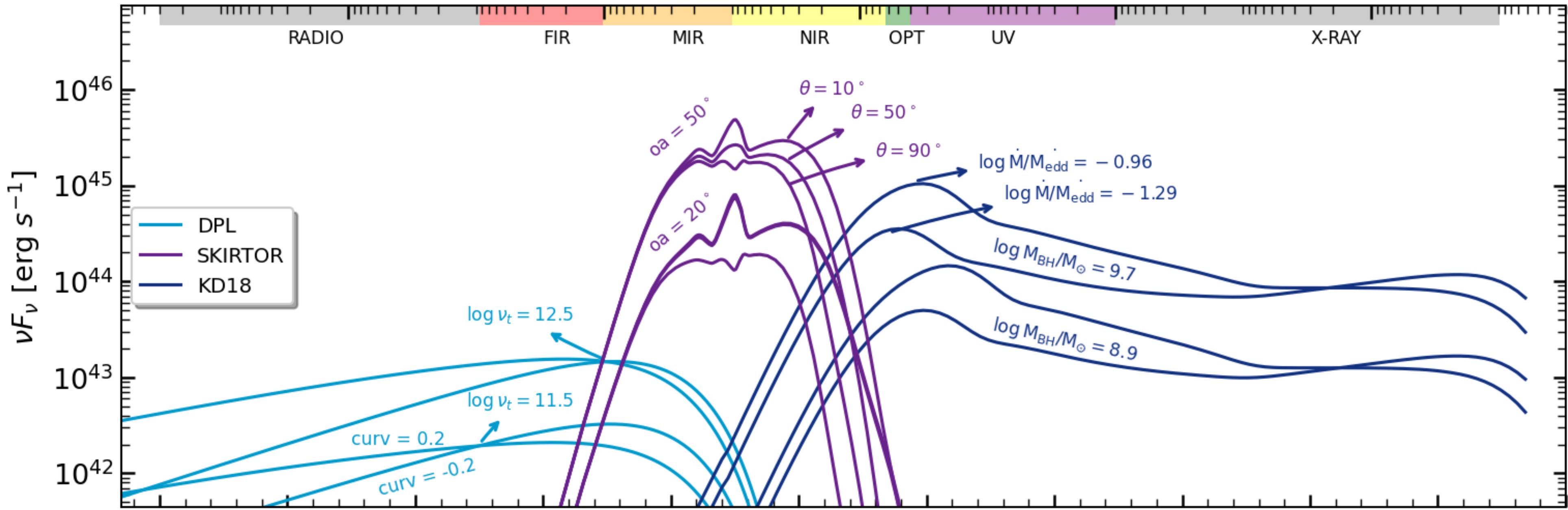


RADIO POWER

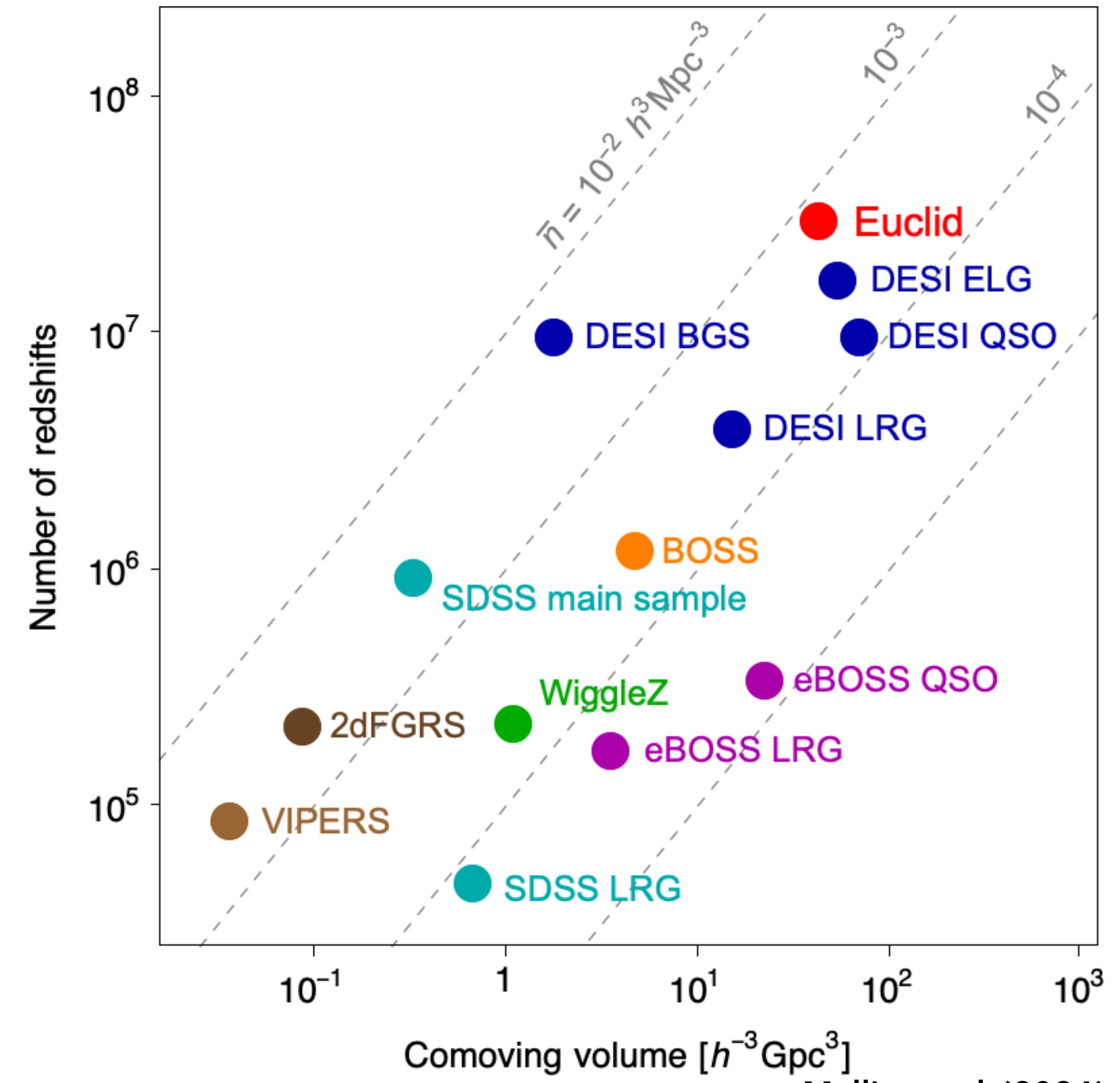
Gürkan et al. (2018)

SED FITTING

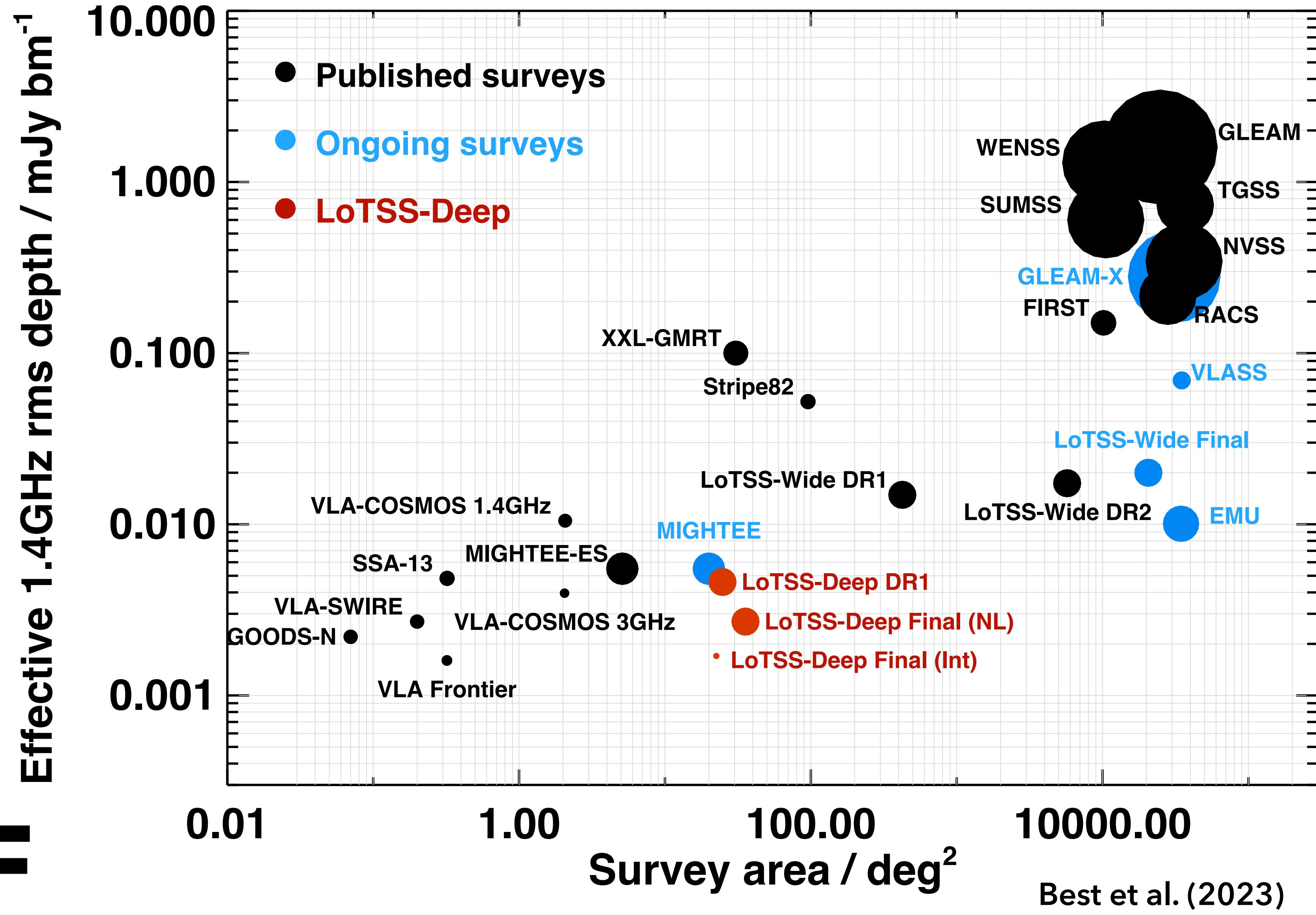
ONE WAY TO DO
MULTI-WAVELENGTH
ANALYSIS



PLENTY OF DATA



PLENTY OF DATA





SOME ISSUES ARISE

TOO MUCH DATA FOR TRADITIONAL METHODS

Data Volume (TB)

10^6

10^5

10^4

10^3

10^2

10^1

1995

2000

2005

2010

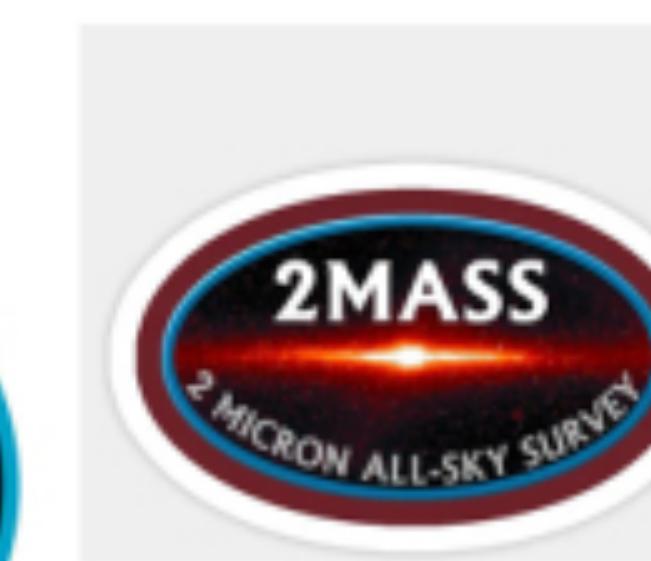
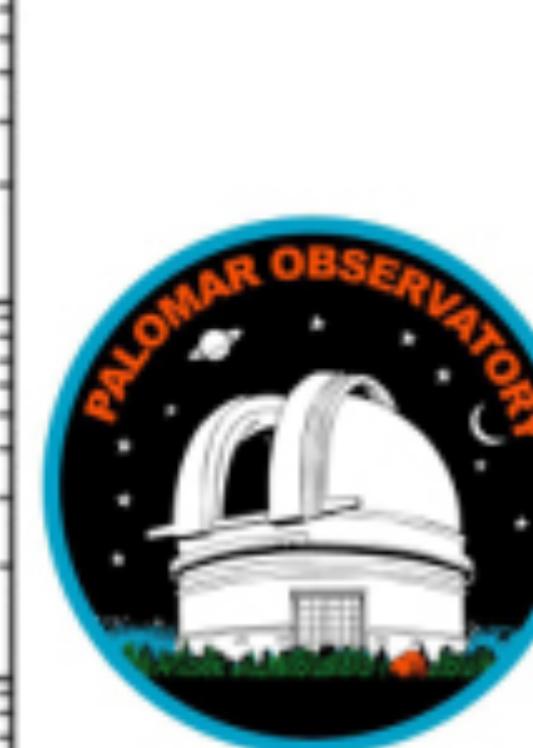
2015

2020

2025

2030

Year



DATA FOR ML

Training data – Initial fit of model parameters – Bulk of data

Validation data – Fit hyperparameters of model

Calibration data – Calibrate probabilities

Testing data – Final assessment of predictions

Prediction data – Measurements without labels

DATA FLOW

Start with CatWISE2020 sources

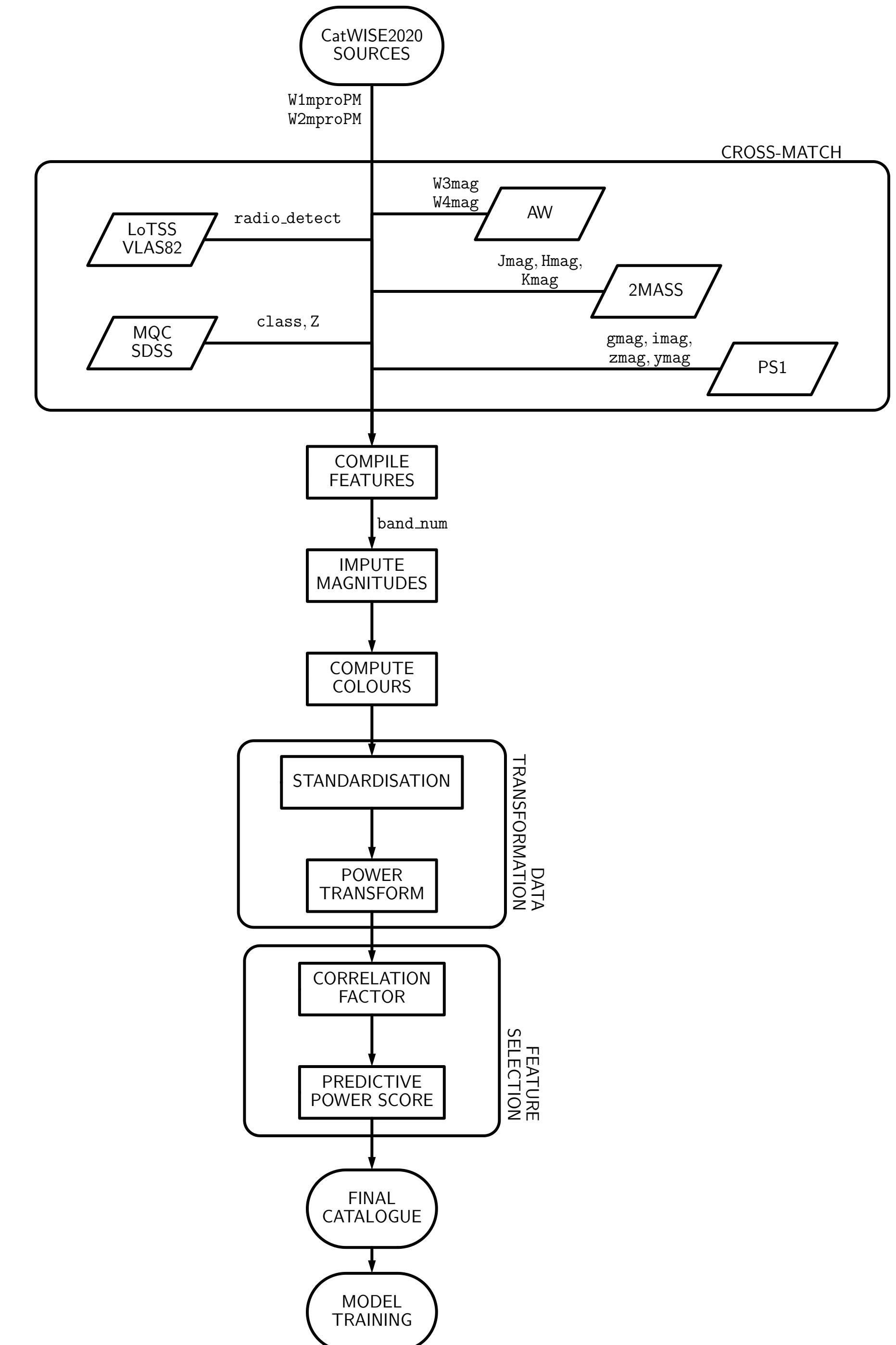
Add Optical, NIR, MIR counterparts

Impute missing values

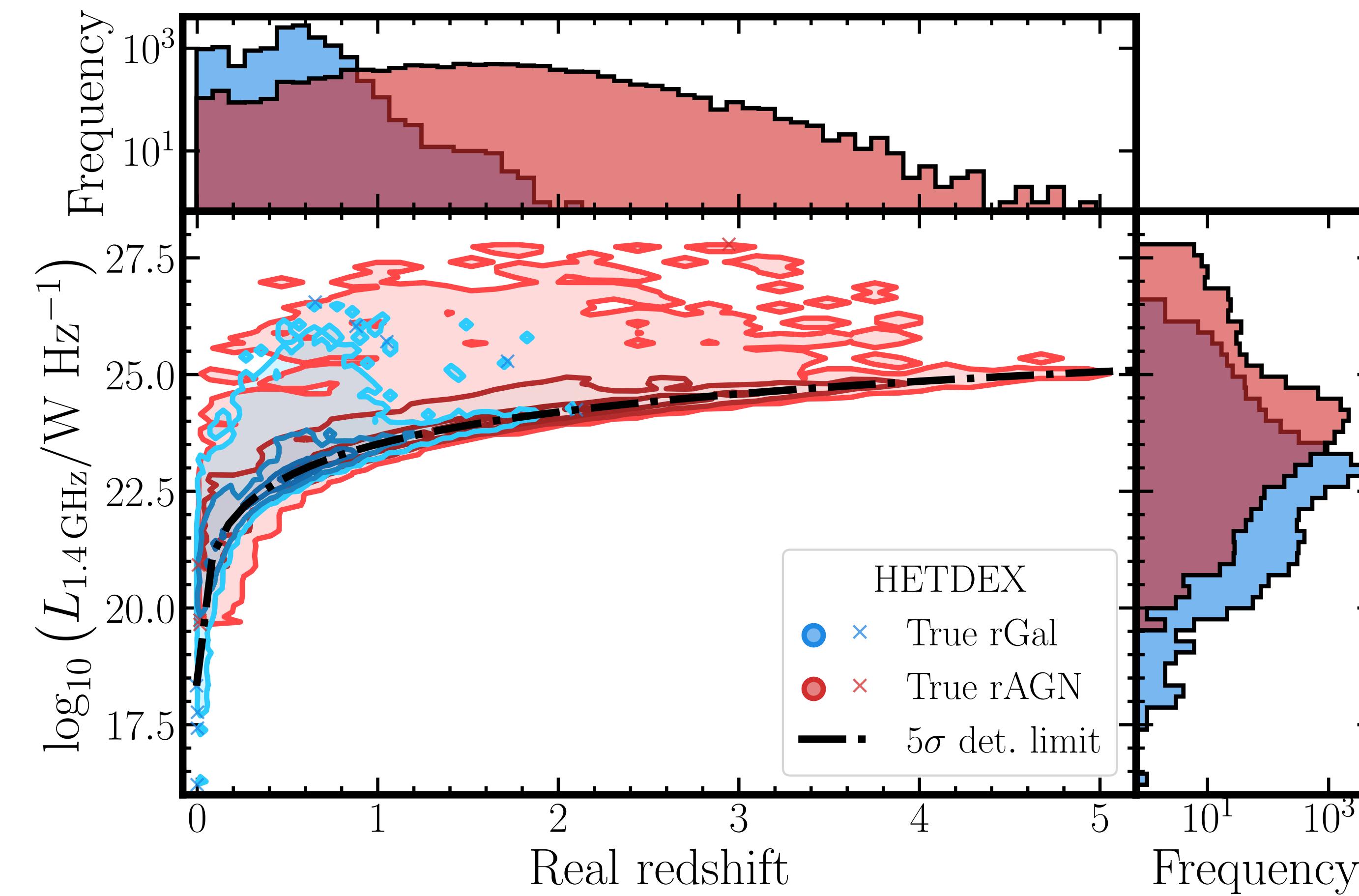
Include colours

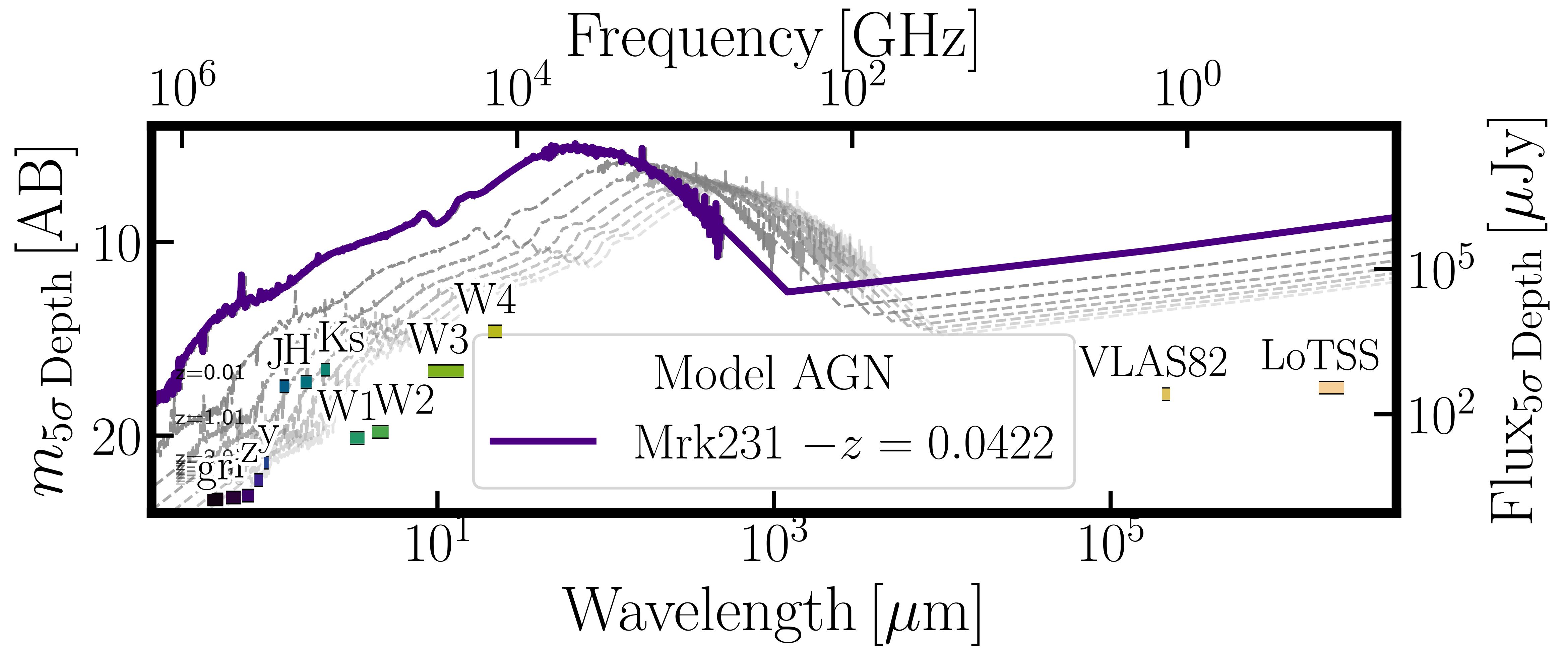
Add labels for training: Class, radio-detection, redshift

For each step, determine most informative features

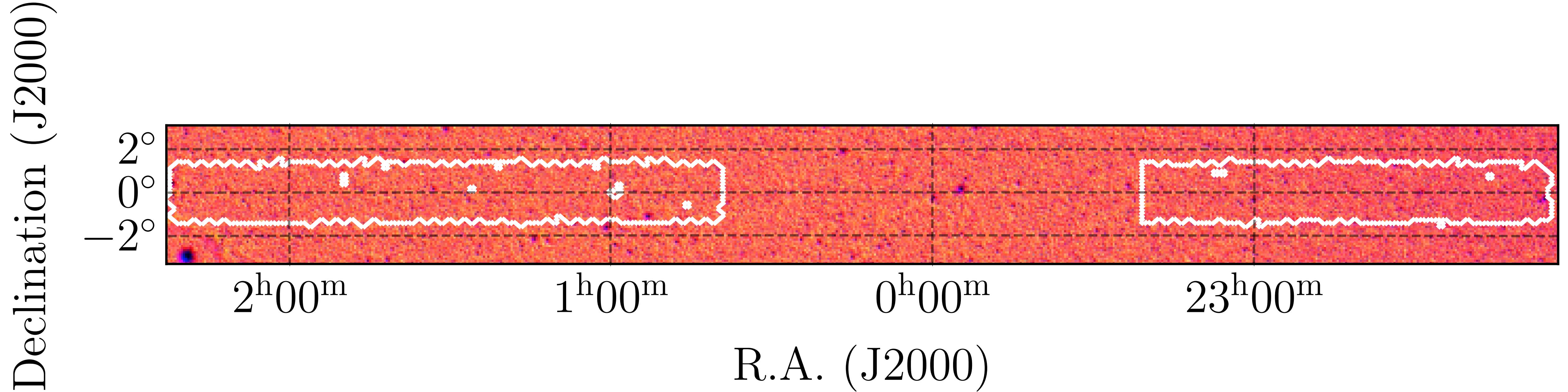


SOURCE DISTRIBUTION HETDEX





HETDEX PHOTOMETRY



STRIPE 82 (S82) FIELD

Only for final testing

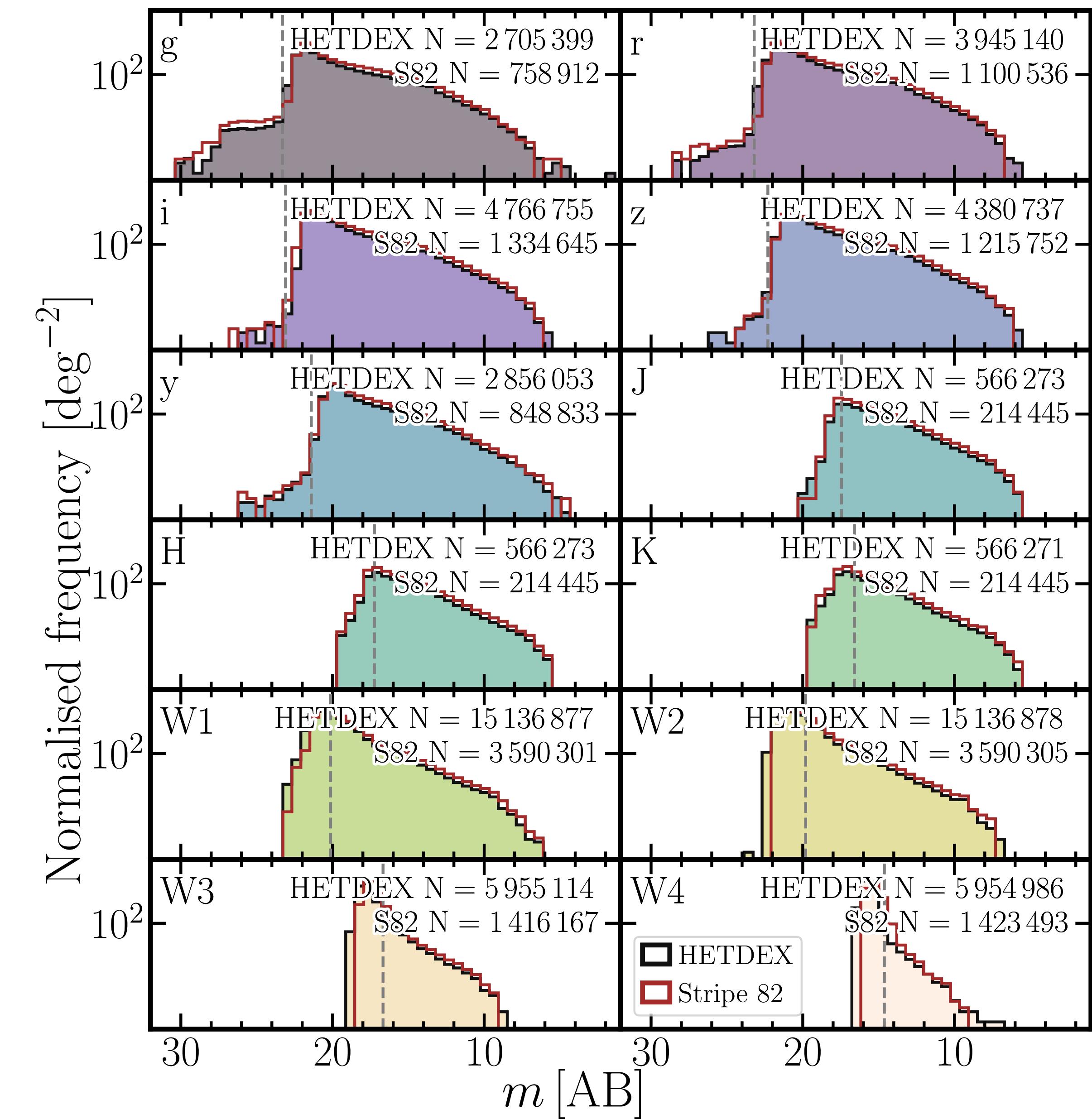
92 deg² covered by VLA @ 1.4 GHz, 52μJy, 1.8'' resolution

~3.5 million CatWISE2020 detections

~18k spectroscopically-confirmed AGN + ~4k spectroscopically-confirmed SFGs

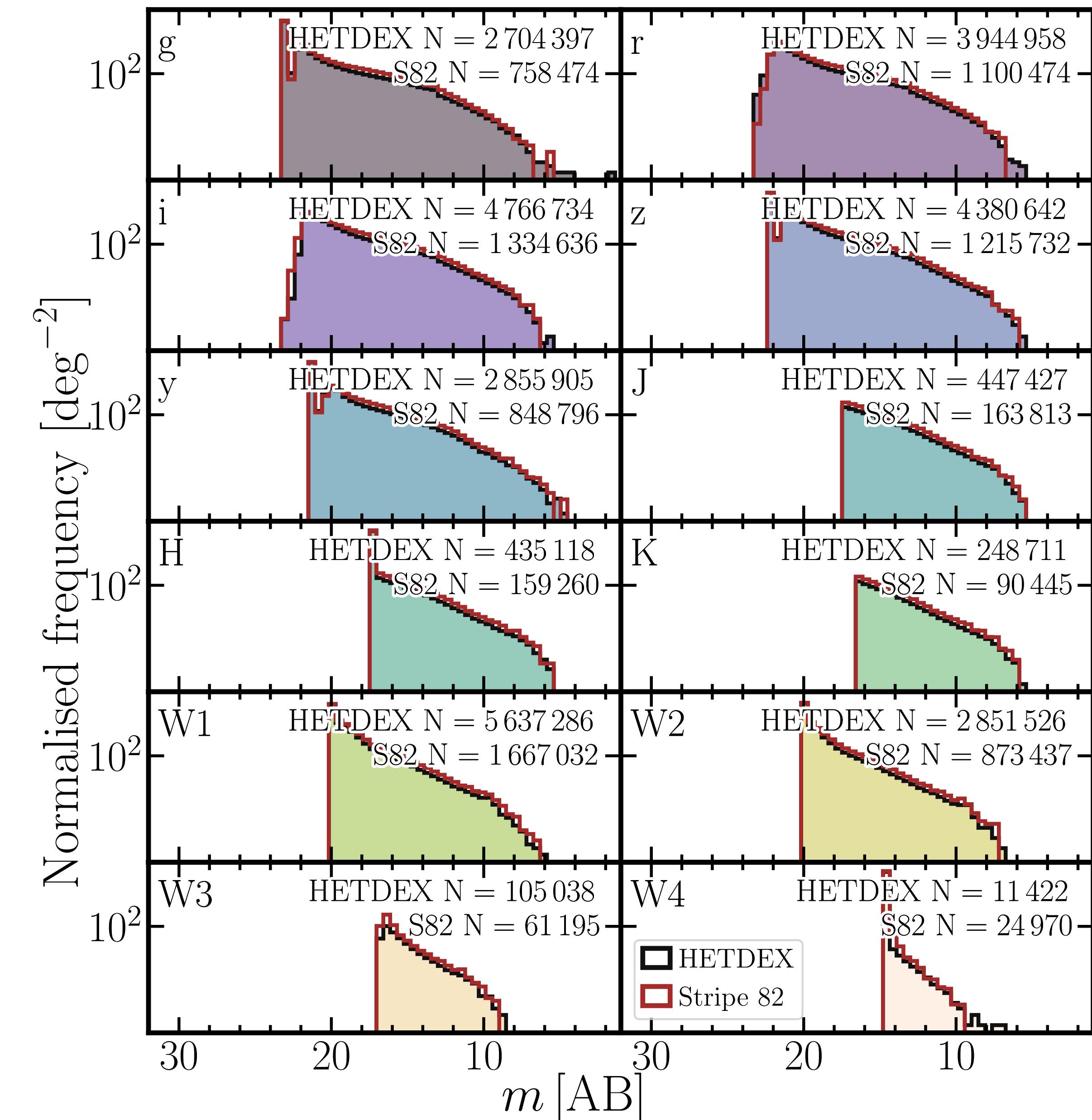
FULL PHOTOMETRY

SIMILAR ACROSS FIELDS



IMPUTED PHOTOMETRY

SIMILAR ACROSS FIELDS



BASELINE SELECTION

Based only on the fraction of sources in the sample:

Selection of AGN: Probability 43 % - Selection of SFG: Probability 57 %

**Selection of radio in AGN: Probability 13 % - Selection of radio in SFG:
Probability 13 %**

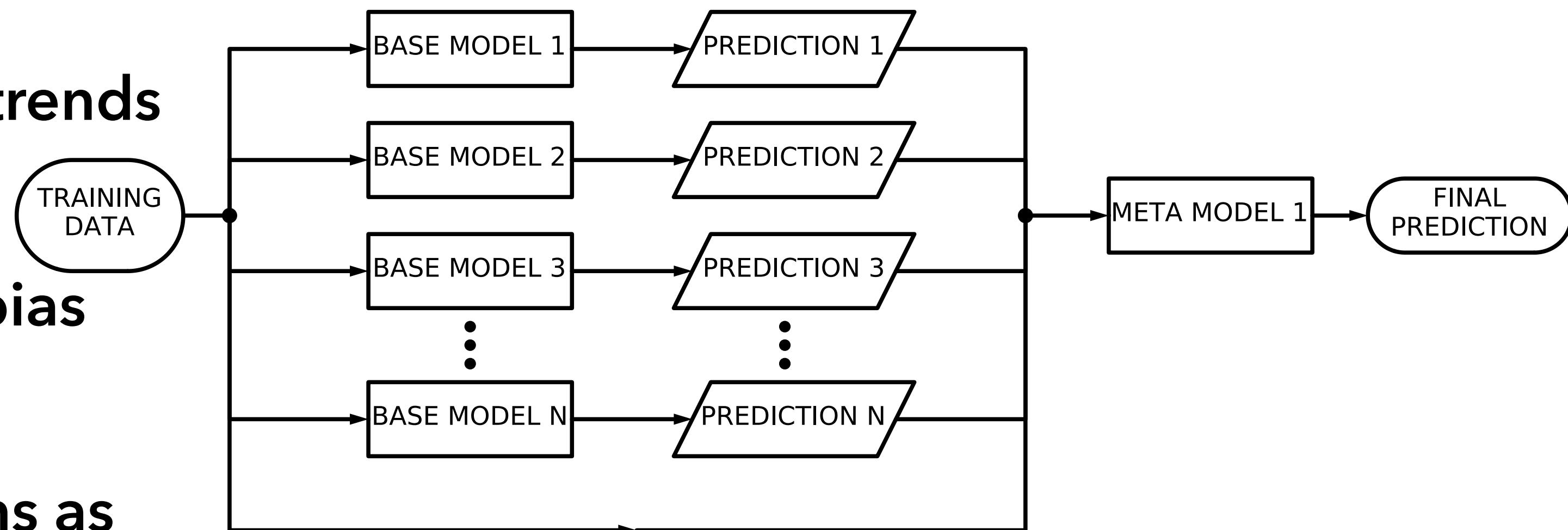
MODEL STACKING

For each step of pipeline, combine several algorithms

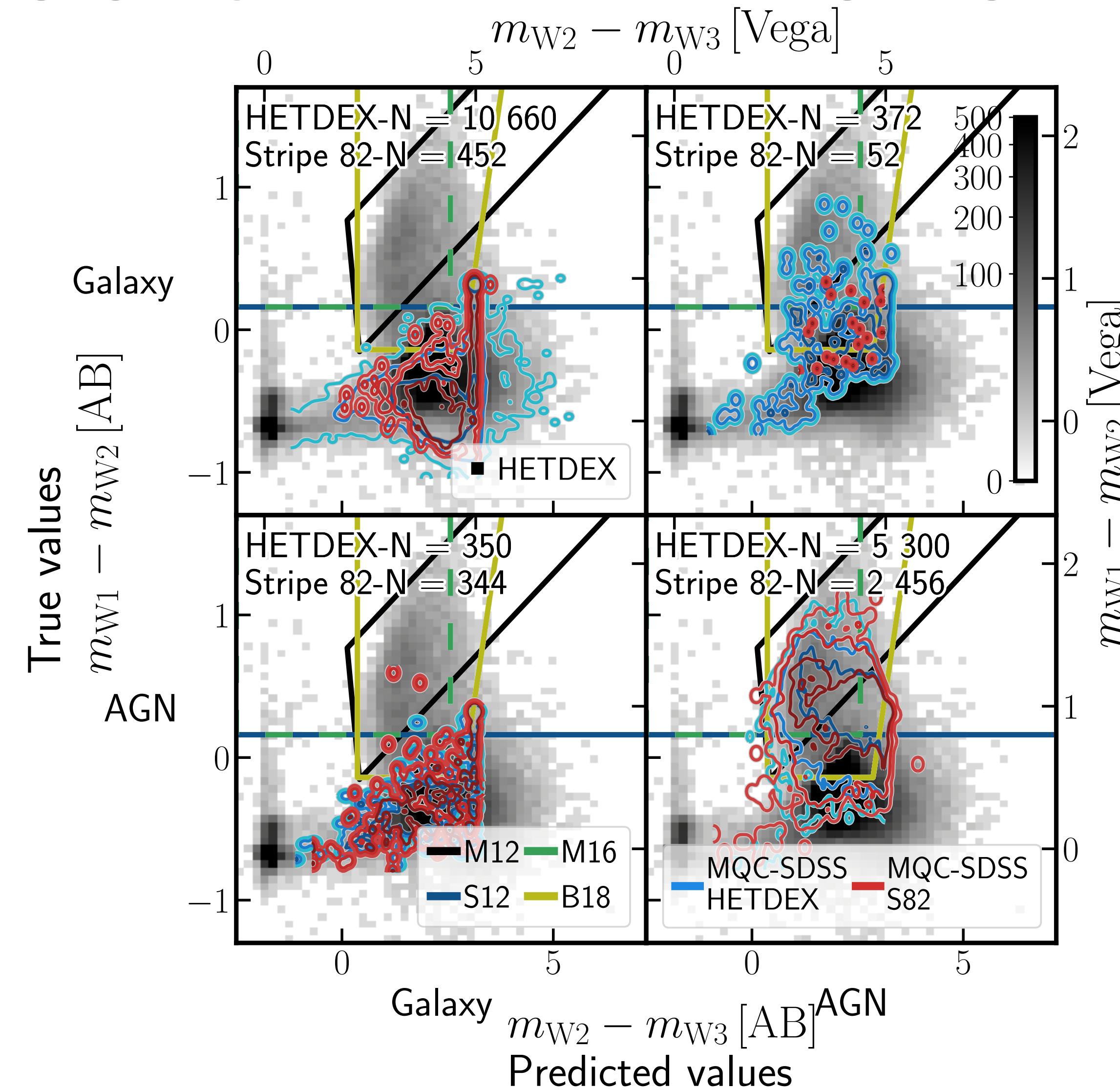
Each model learns slightly different trends

Aims at improving results reducing bias

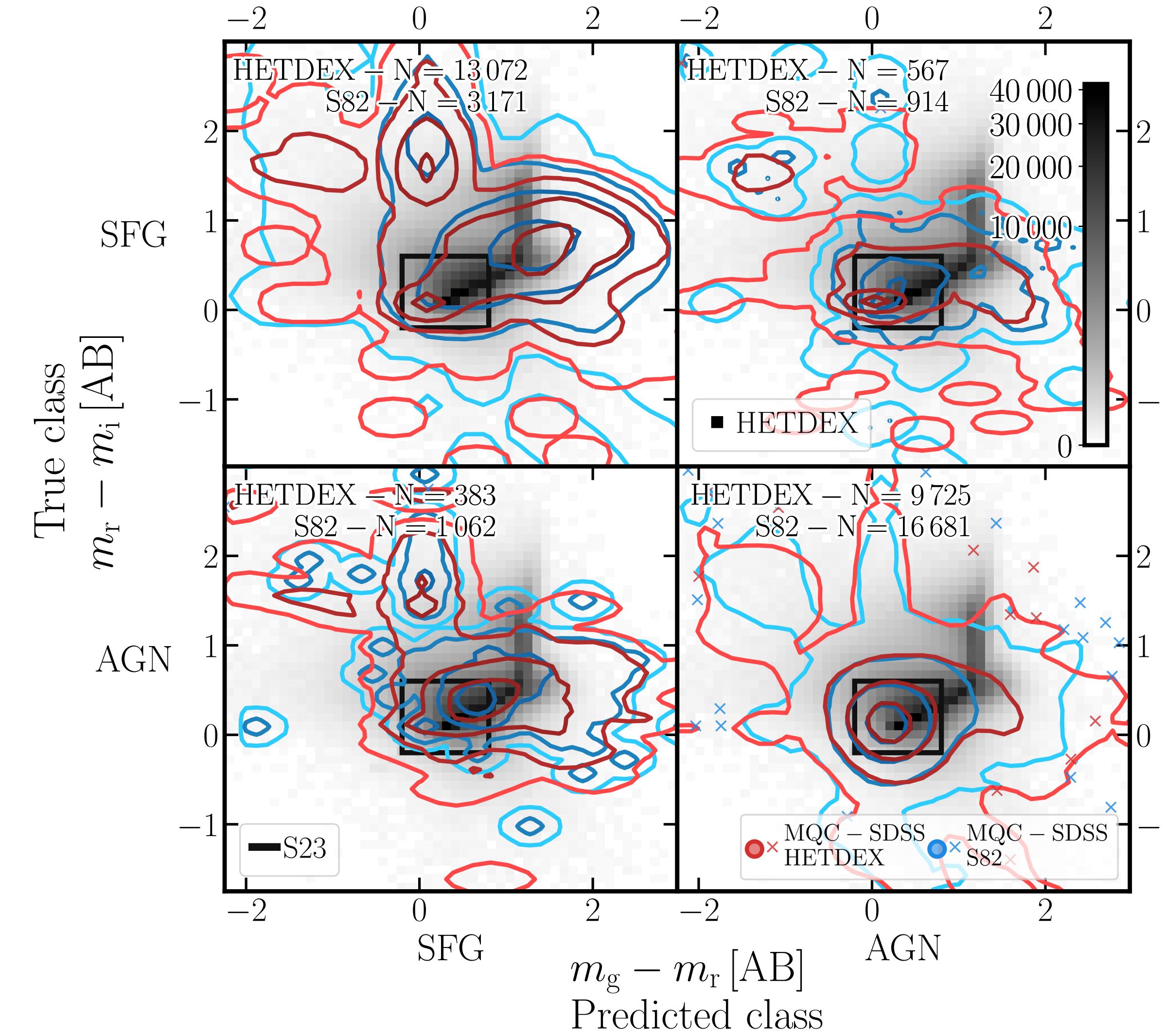
Final model uses previous predictions as input



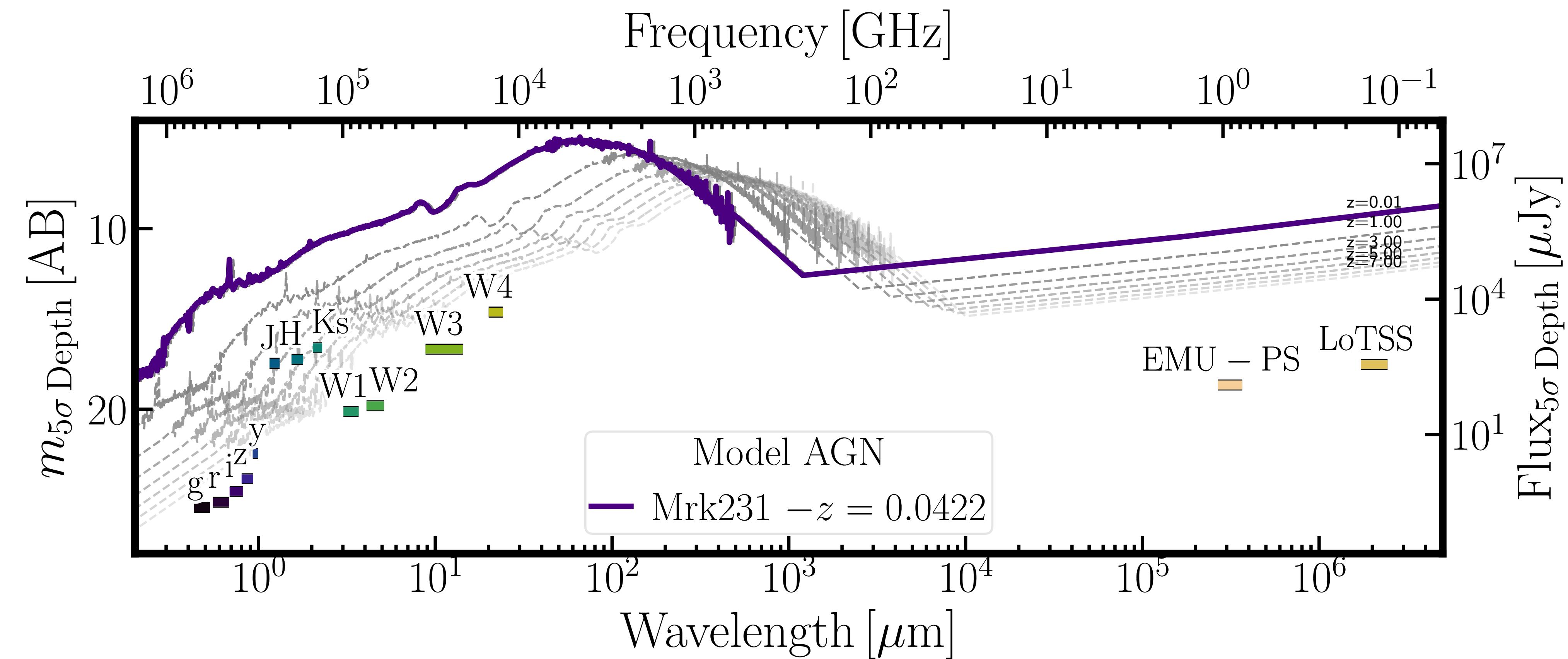
WISE COLOURS AND PREDICTIONS



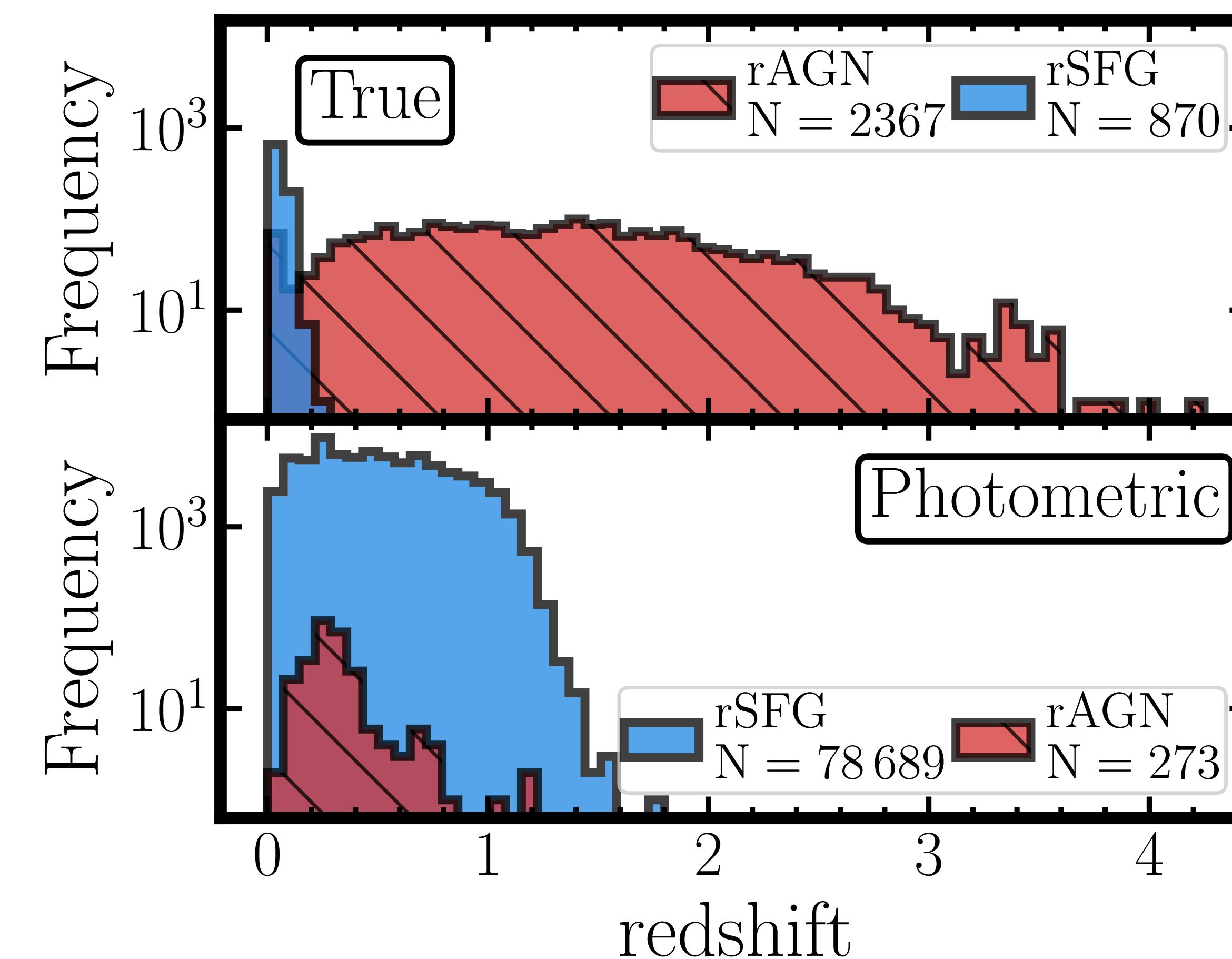
WISE + PS1 COLOURS AND PREDICTIONS



EMU PHOTOMETRY BANDS



PREDICTION IN EMU-PS



RADIO LUMINOSITY FUNCTION CORRECTIONS

