Ciências ULisboa
ia instituto de astrofísica e ciências do espaço

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

PD F FCT PhD PROGRAMMES
COFINANCIAMENTO / COFINANCING
fct Fundação para a Ciência e a Tecnologia
COMPETE 2020
PORTUGAL 2020
UNIÃO EUROPEIA Fundo Europeu de Desenvolvimento Regional

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 PhD PROGRAMMES
COFINANCIAMENTO / COFINANCING
Fundação para a Ciência e a Tecnologia
COMPETE 2020
PORTUGAL 2020
UNIÃO EUROPEIA
Fundo Europeu de Desenvolvimento Regional

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

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

Need to understand role of SMBH in galaxy evolution!

Need to understand role of SMBH in galaxy evolution!

AGN + SF

Need to understand role of SMBH in galaxy evolution!

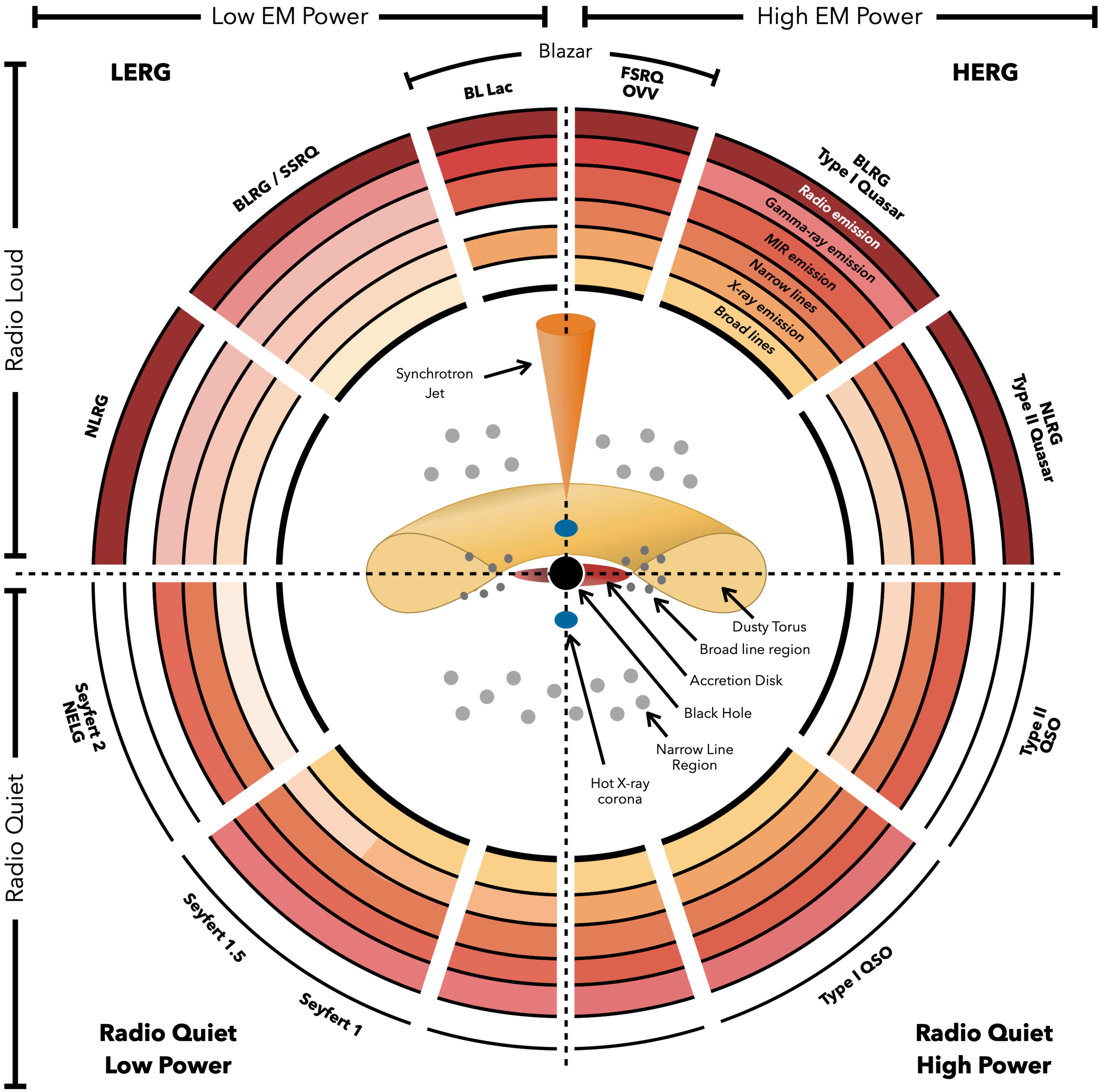
AGN + SF

Redshifts

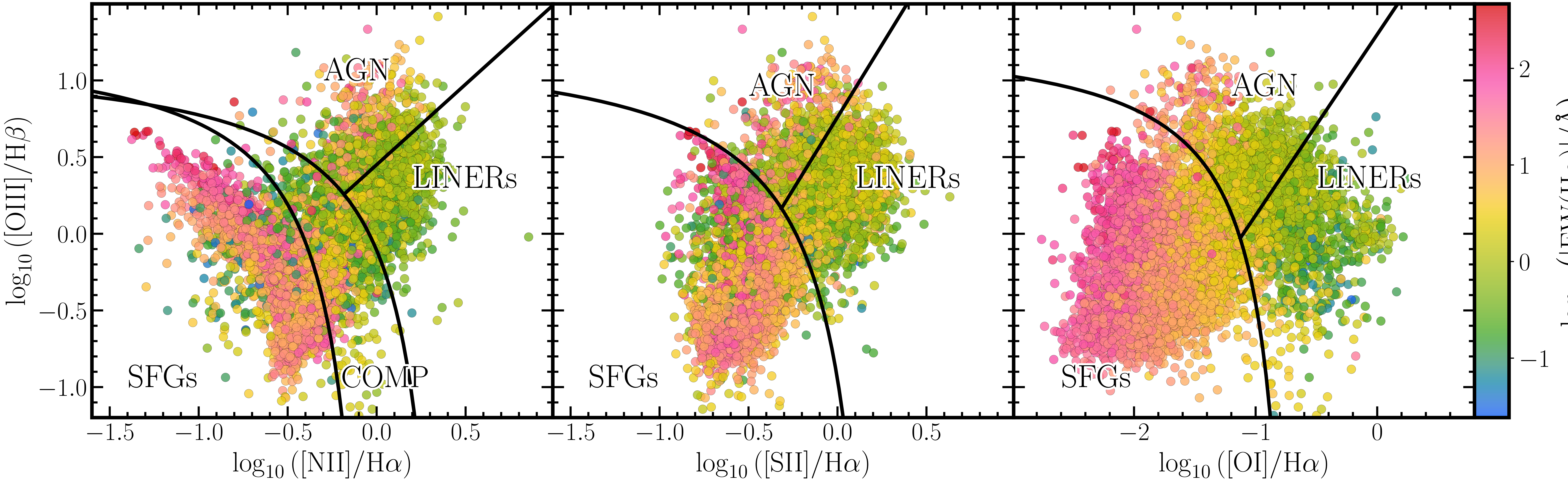
Separating AGN from SFG (non-AGN)

DETECT AGN THROUGH DIFFERENT WAVELENGTHS

ACCESSING VARIOUS PHYSICAL PROCESSES AND PROPERTIES

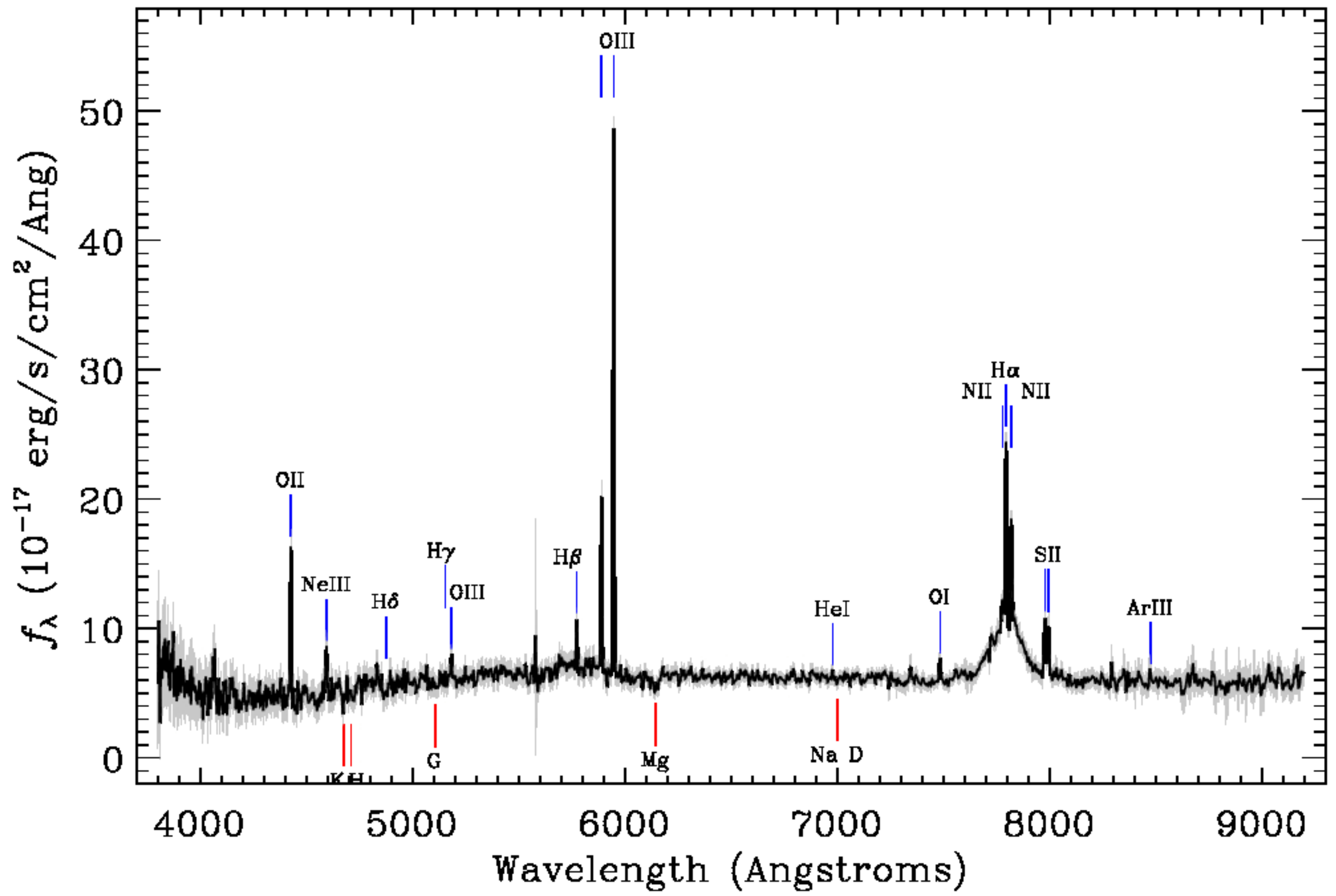


J. E. Thorne

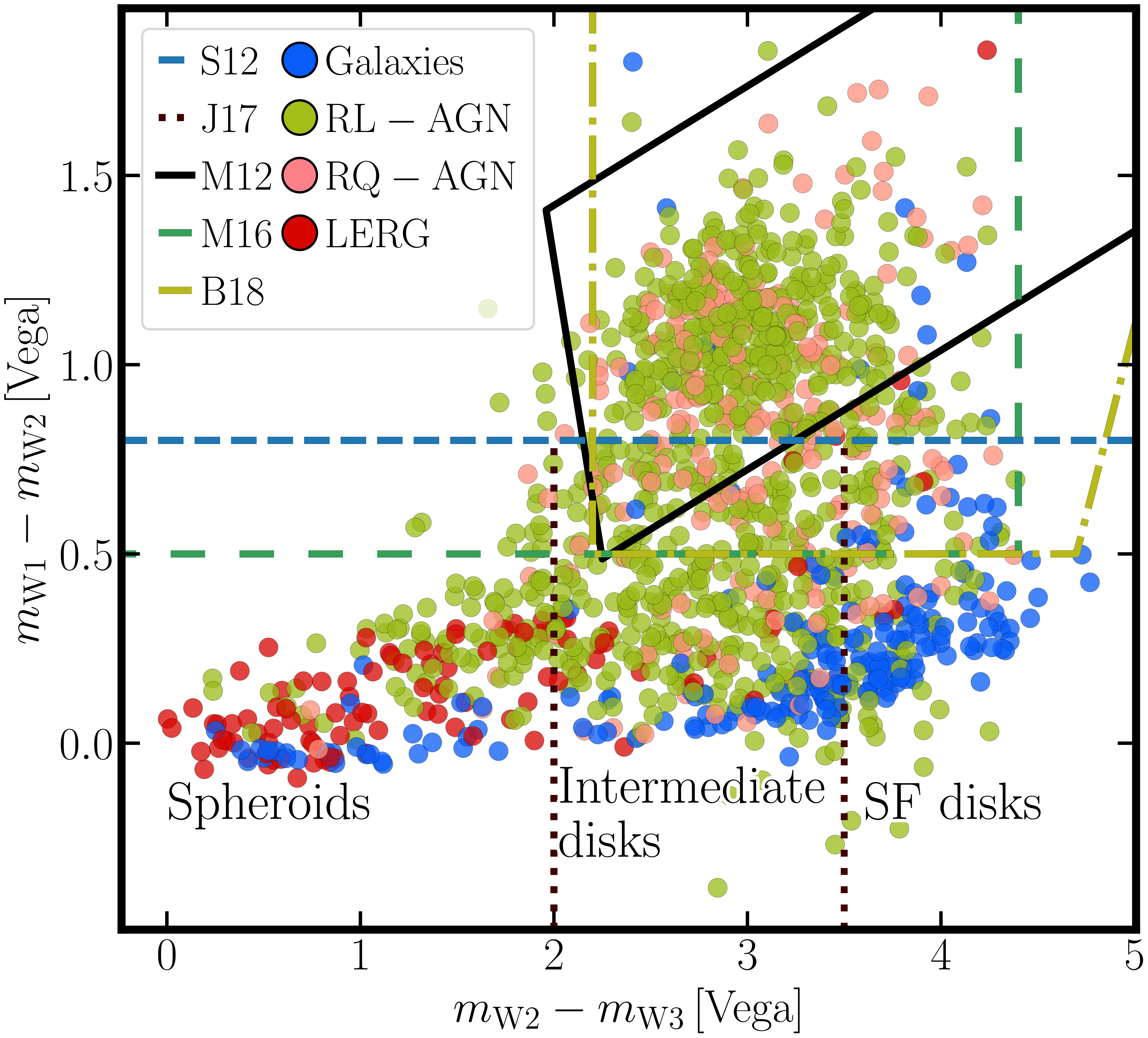


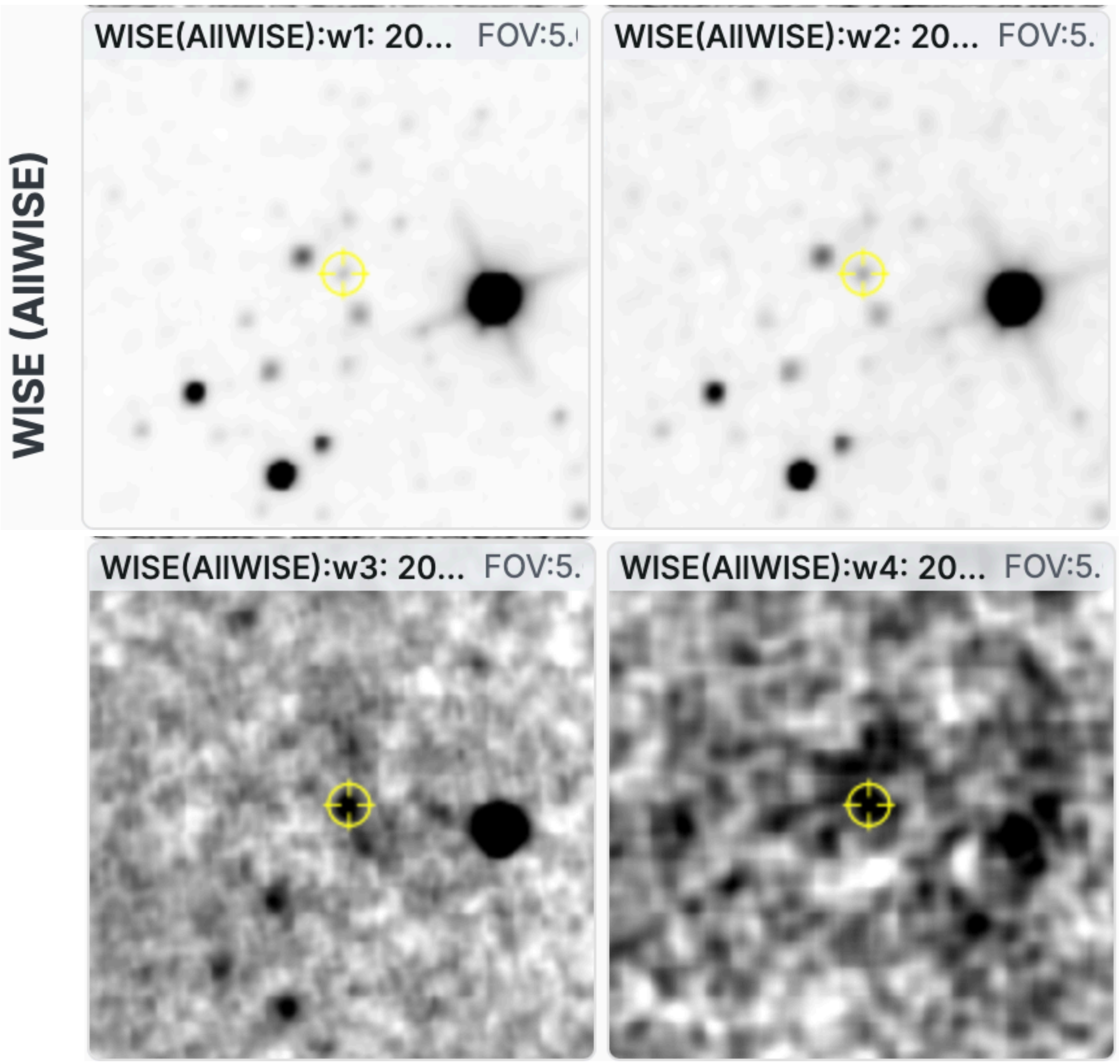
BPT Diagram

SPECTROSCOPY



SDSS





PHOTOMETRY

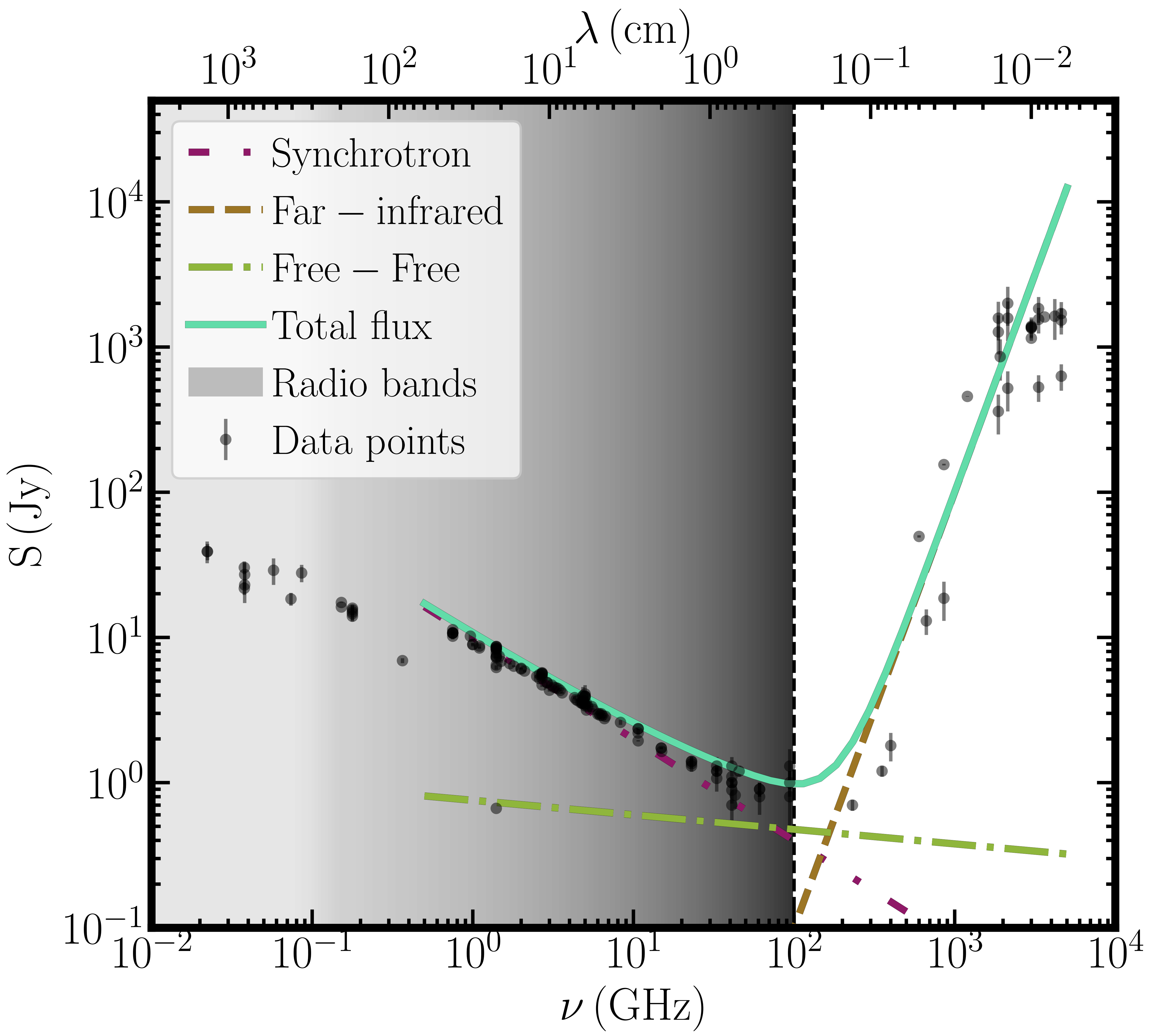
WISE Colours

Radio bands

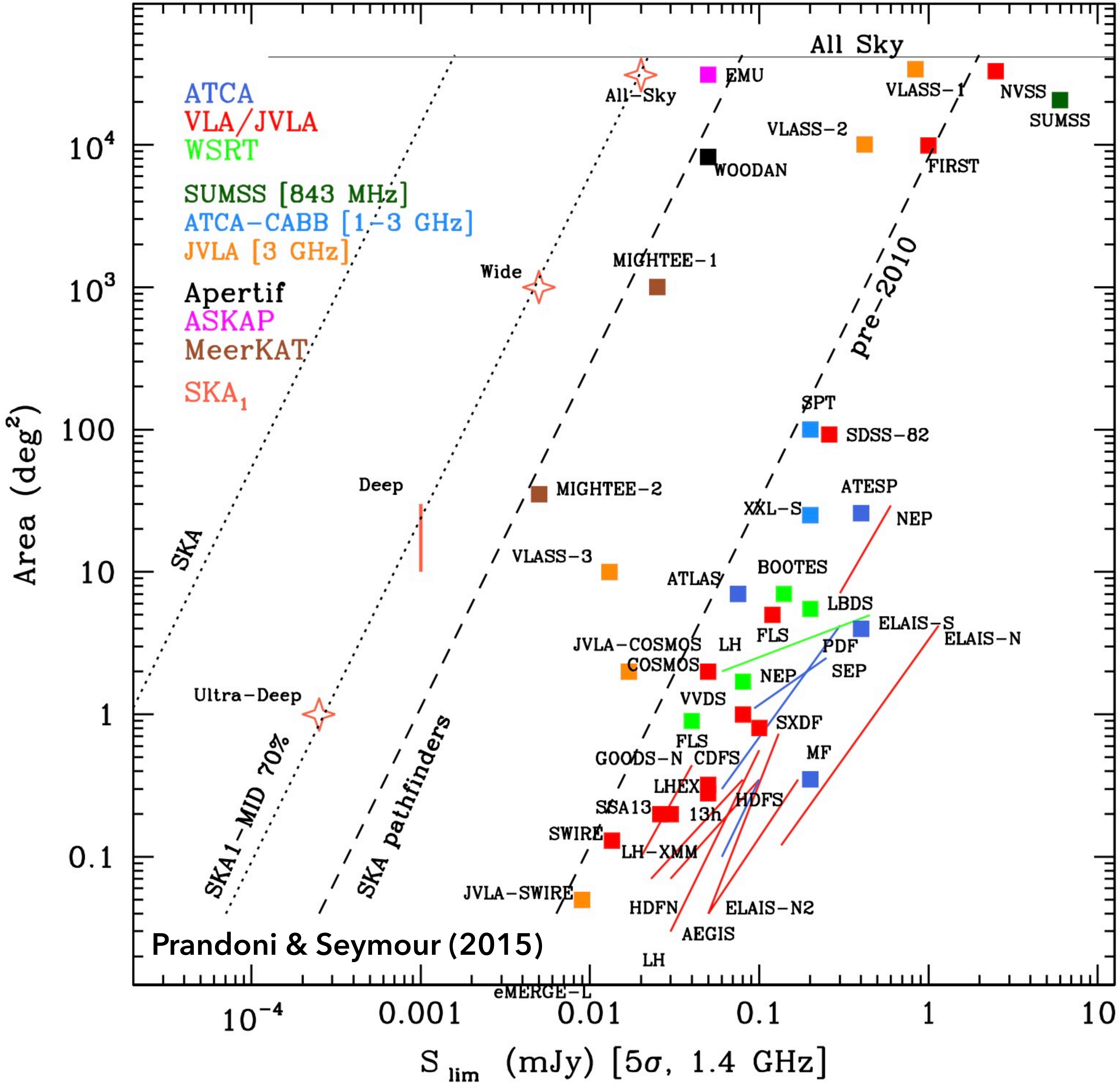
Almost direct observation of AGN

SKA + Precursors

Large datasets



PLENTY OF DATA

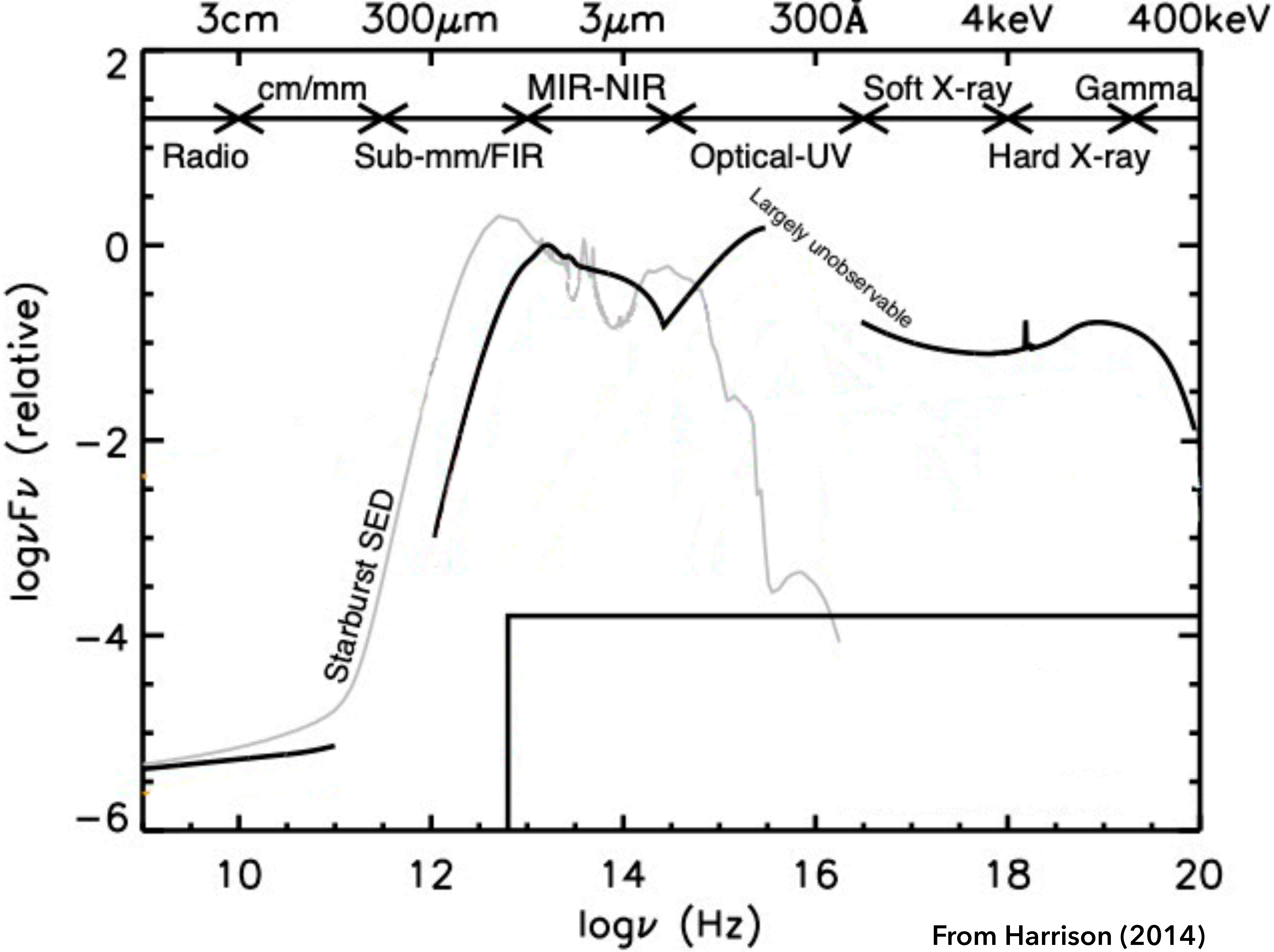


Is it possible to incorporate several indicators into one tool?

We need a multi-wavelength approach!

SED FITTING

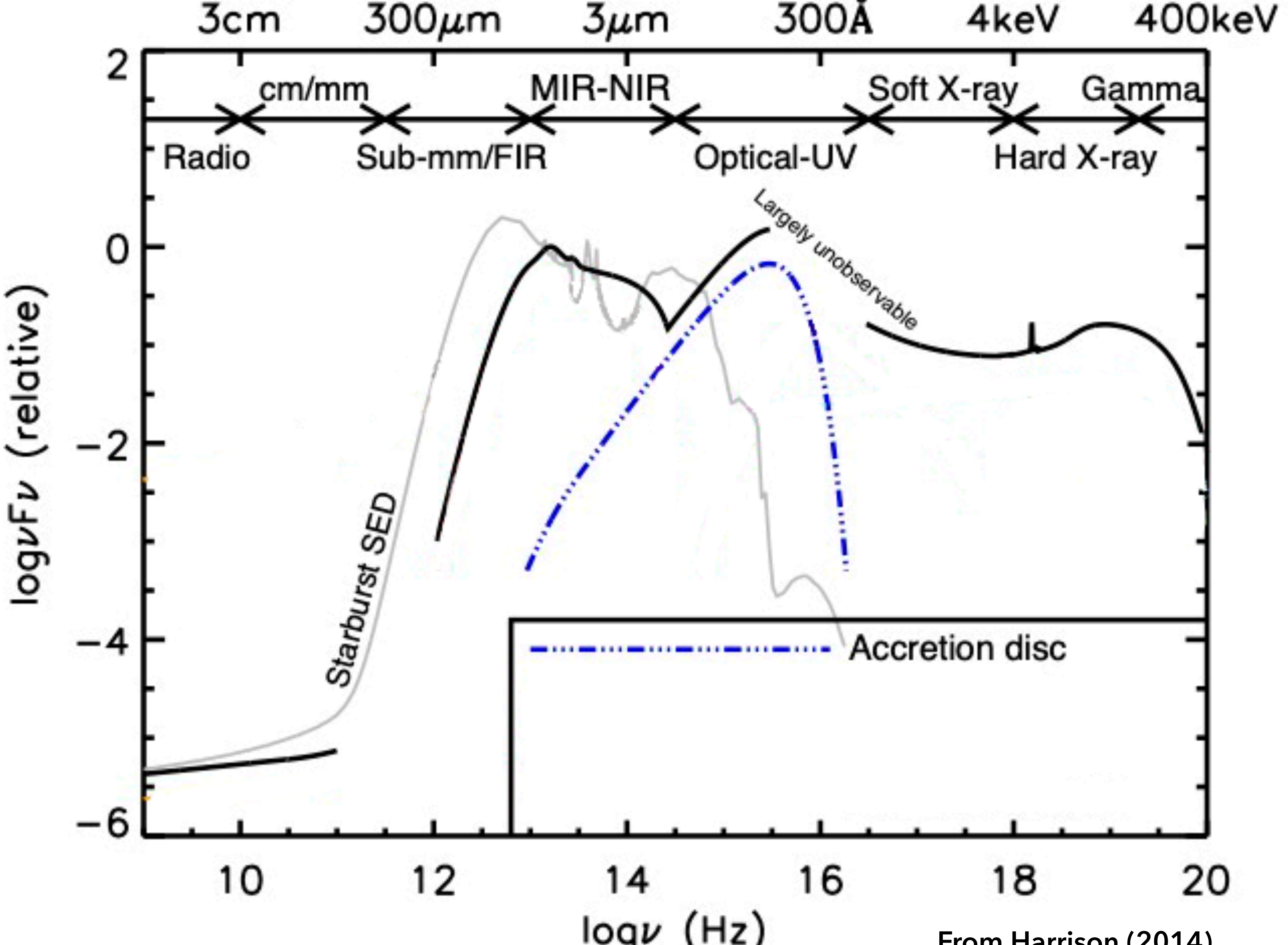
ONE WAY TO DO MULTI-WAVELENGTH ANALYSIS



From Harrison (2014)

SED FITTING

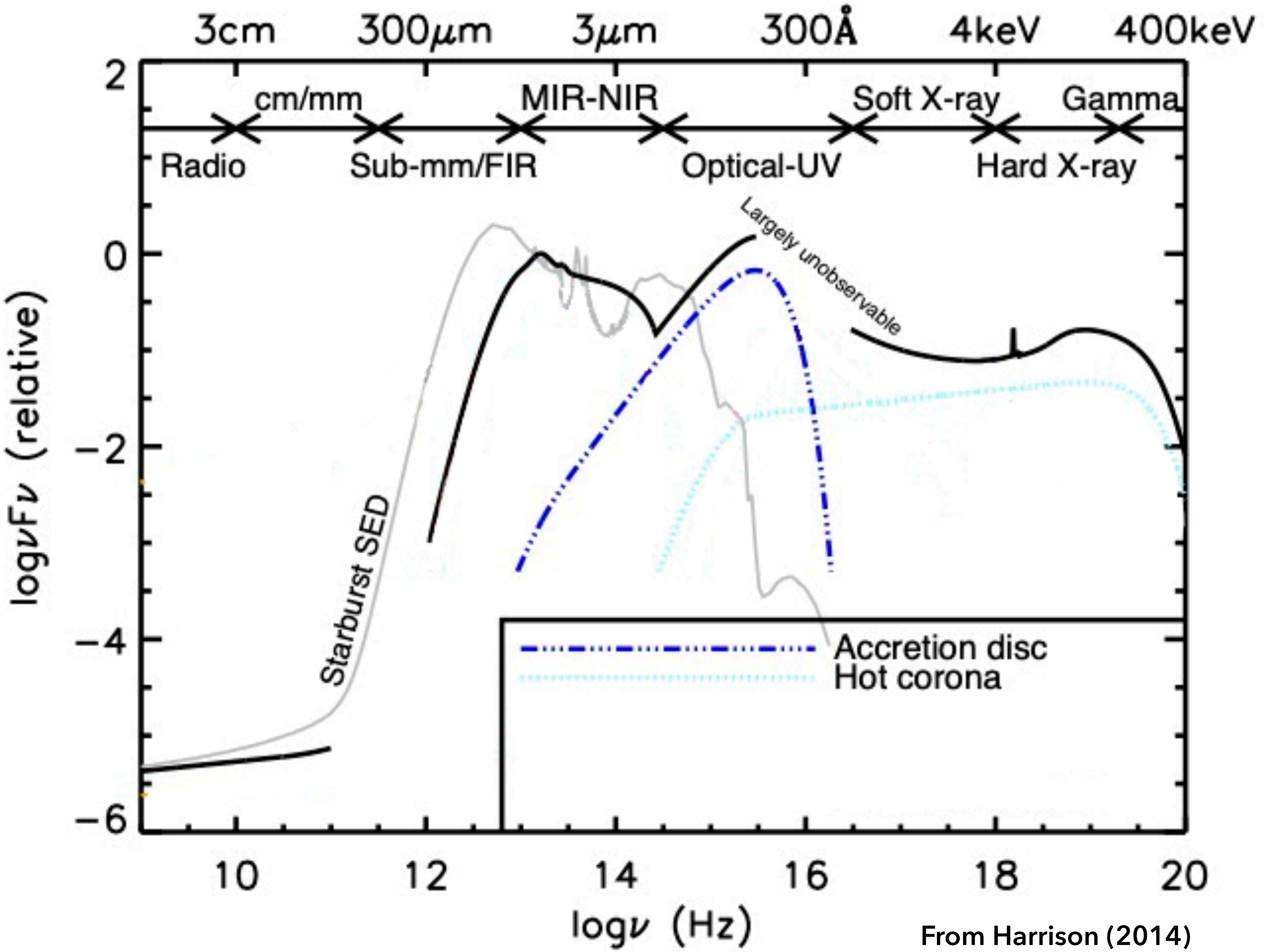
ONE WAY TO DO MULTI-WAVELENGTH ANALYSIS



From Harrison (2014)

SED Fitting

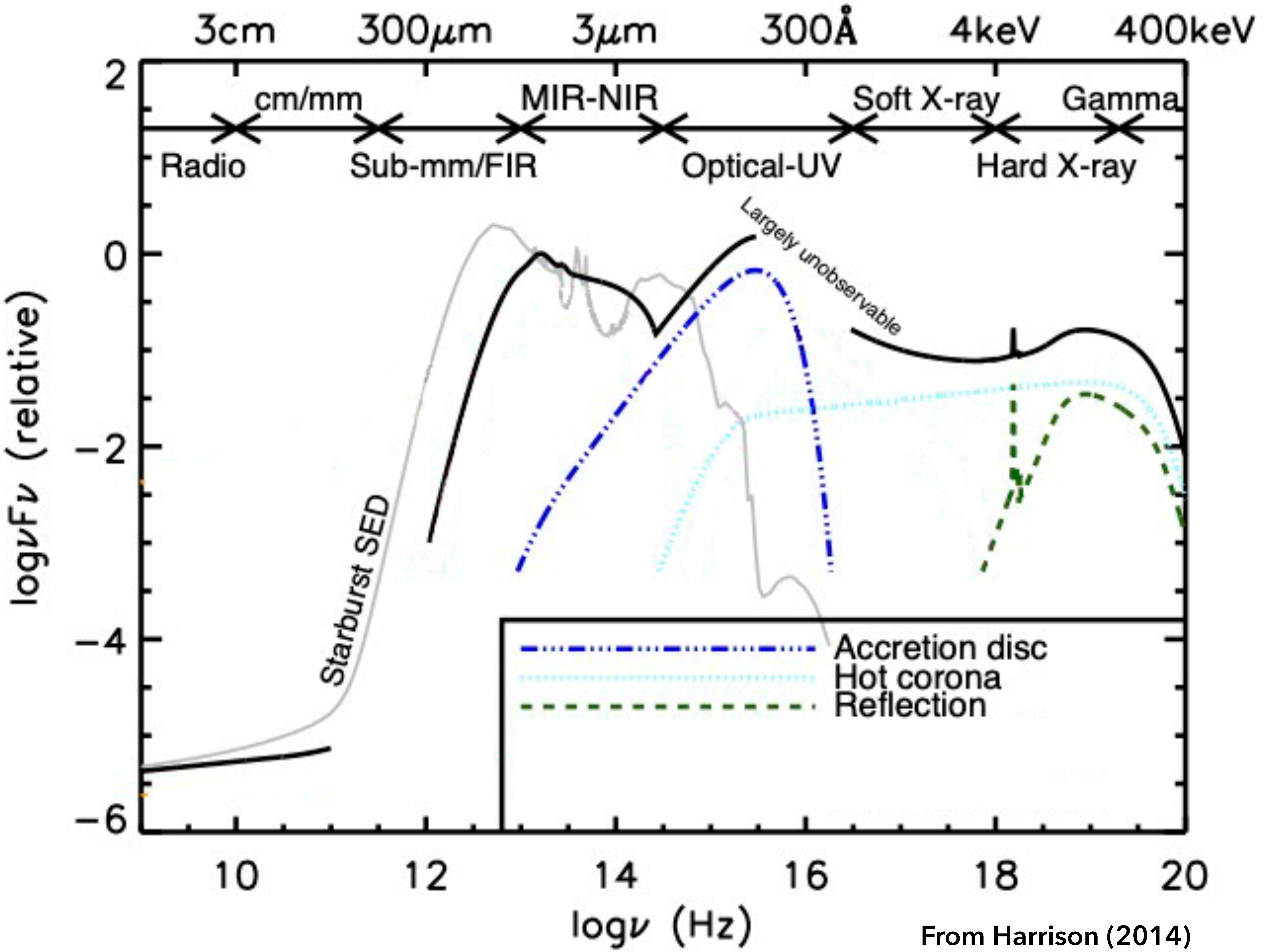
One way to do multi-wavelength analysis



From Harrison (2014)

SED FITTING

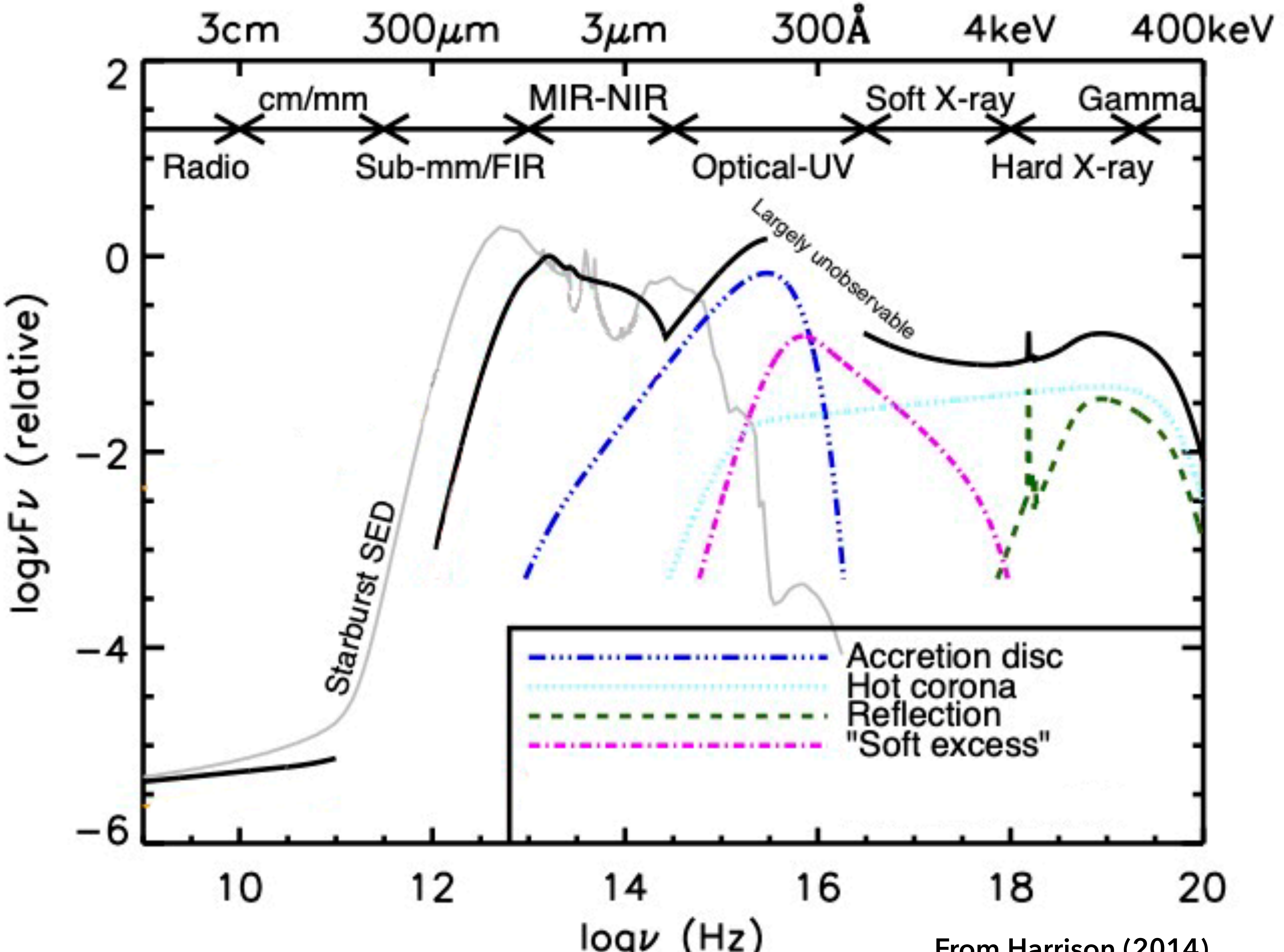
ONE WAY TO DO MULTI-WAVELENGTH ANALYSIS



From Harrison (2014)

SED FITTING

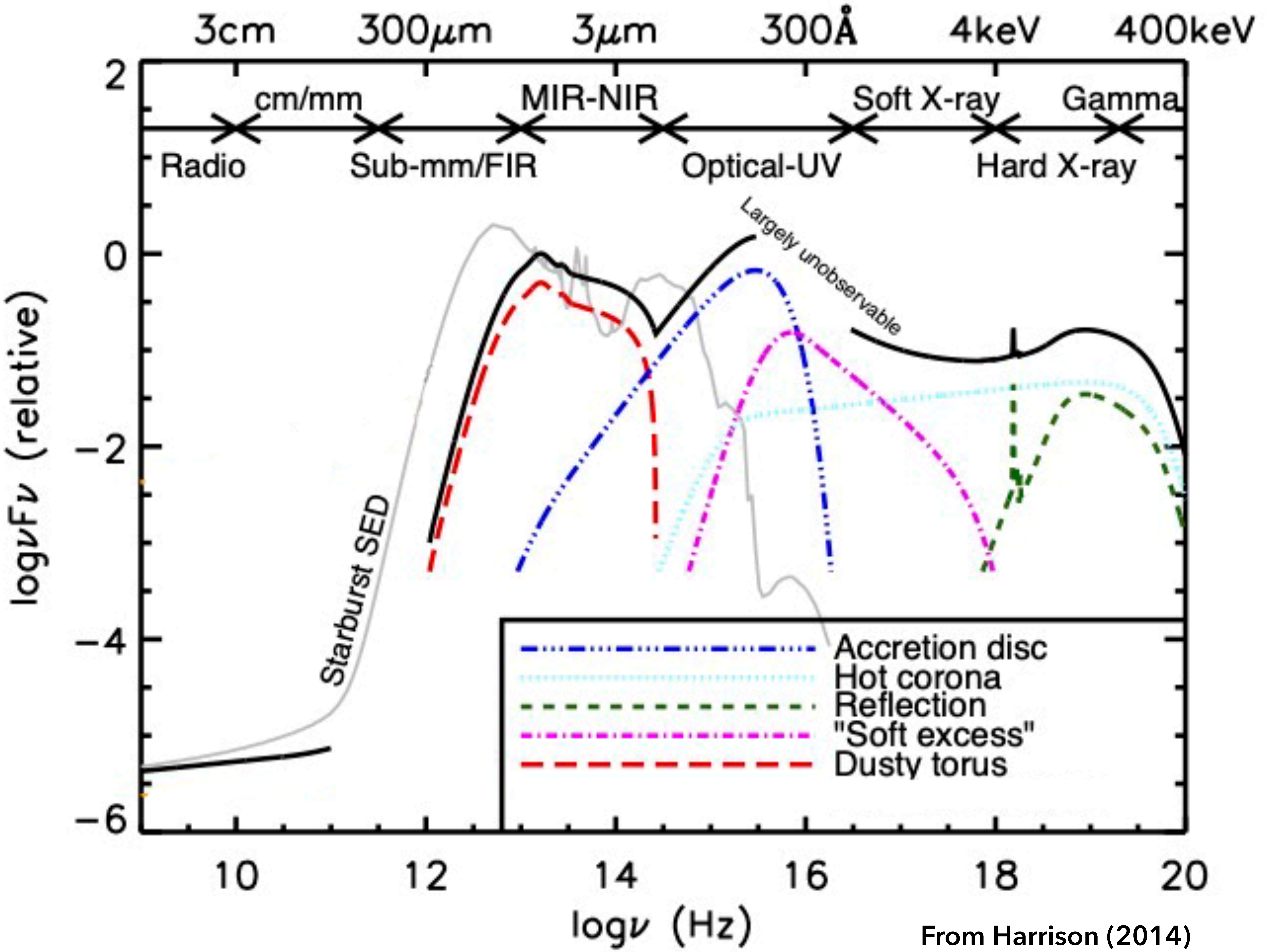
ONE WAY TO DO MULTI-WAVELENGTH ANALYSIS



From Harrison (2014)

SED FITTING

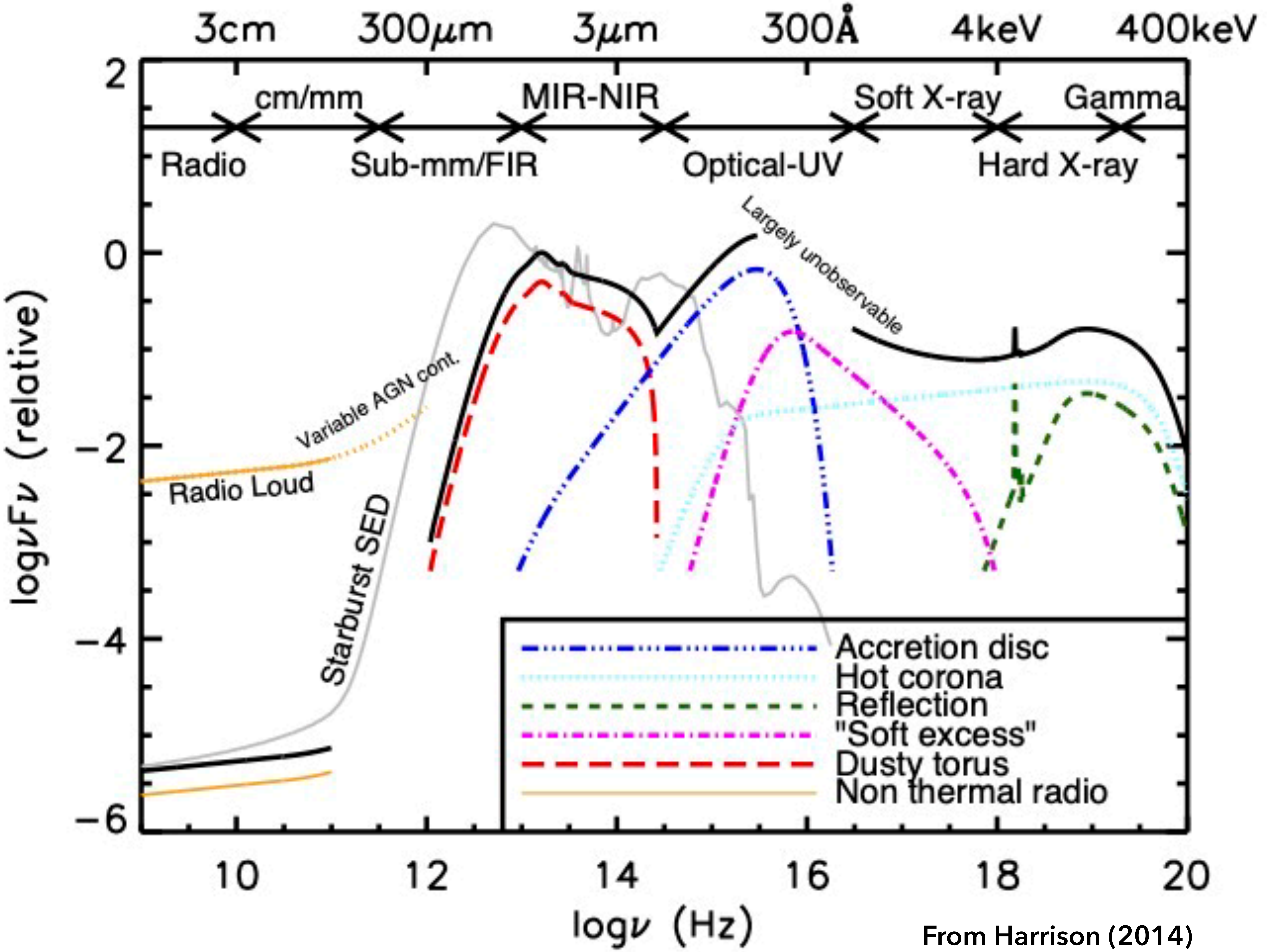
ONE WAY TO DO MULTI-WAVELENGTH ANALYSIS



From Harrison (2014)

SED FITTING

ONE WAY TO DO MULTI-WAVELENGTH ANALYSIS



From Harrison (2014)

Machine Learning can help!

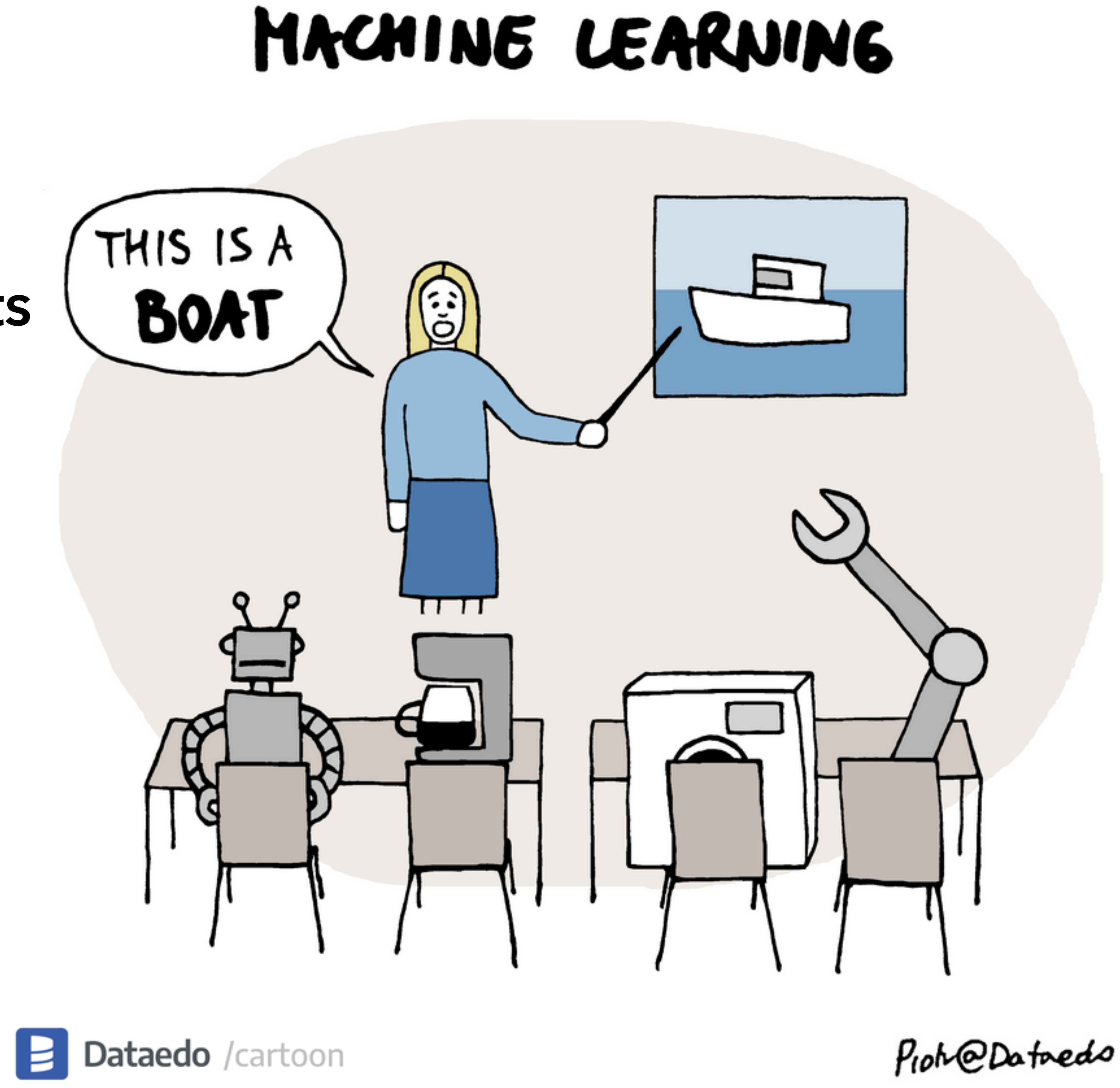
MACHINE LEARNING

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



Our goal

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

PREDICTION PIPELINE

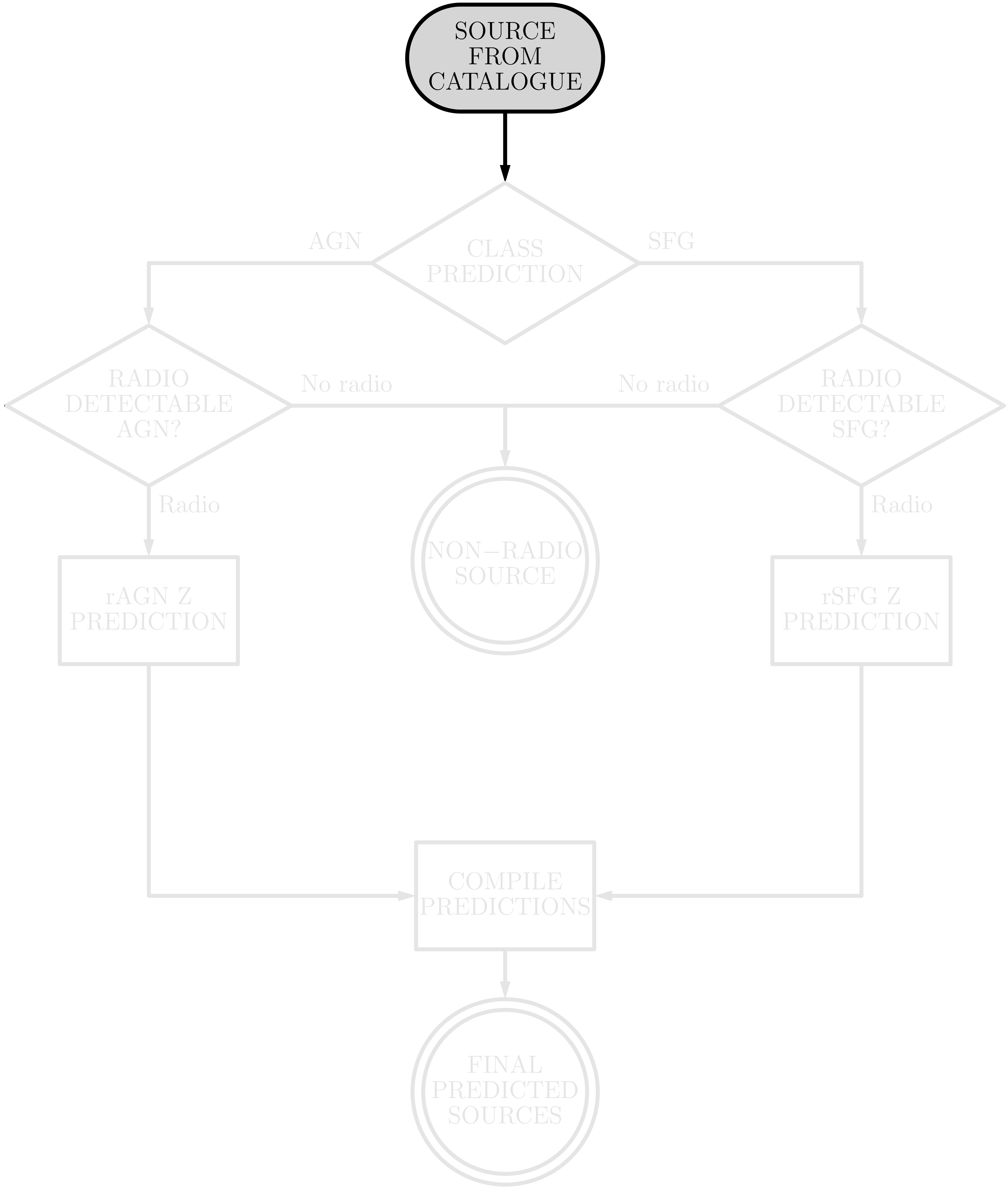
OUR APPROACH TO SELECTING AGN AND SFG WITH ML

Three different levels

Classify as AGN or SFG

Select radio-detectable sources

Estimate redshift for radio-detectable sources



PREDICTION PIPELINE

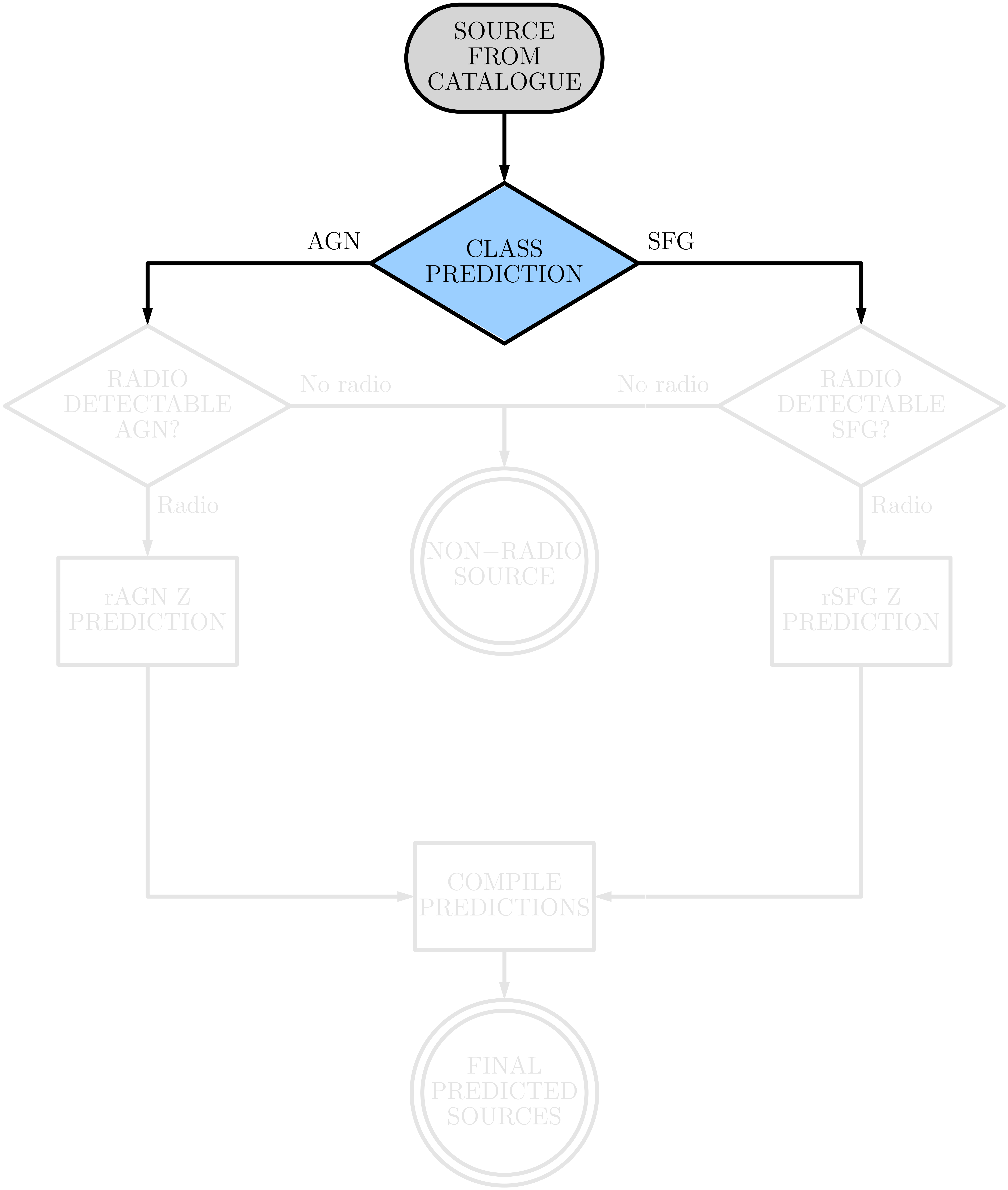
OUR APPROACH TO SELECTING AGN AND SFG WITH ML

Three different levels

Classify as AGN or SFG

Select radio-detectable sources

Estimate redshift for radio-detectable sources



PREDICTION PIPELINE

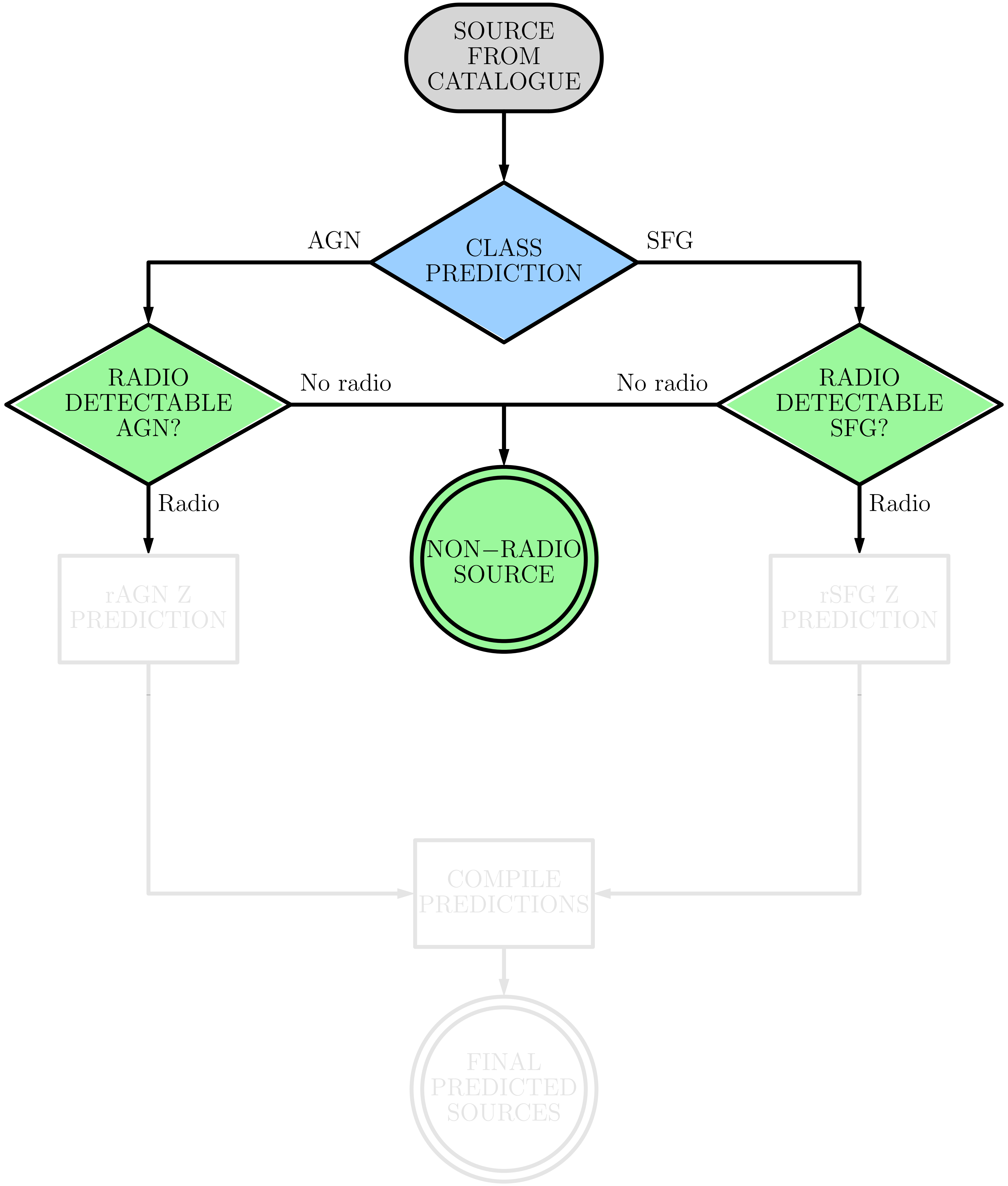
OUR APPROACH TO SELECTING AGN AND SFG WITH ML

Three different levels

Classify as AGN or SFG

Select radio-detectable sources

Estimate redshift for radio-detectable sources



PREDICTION PIPELINE

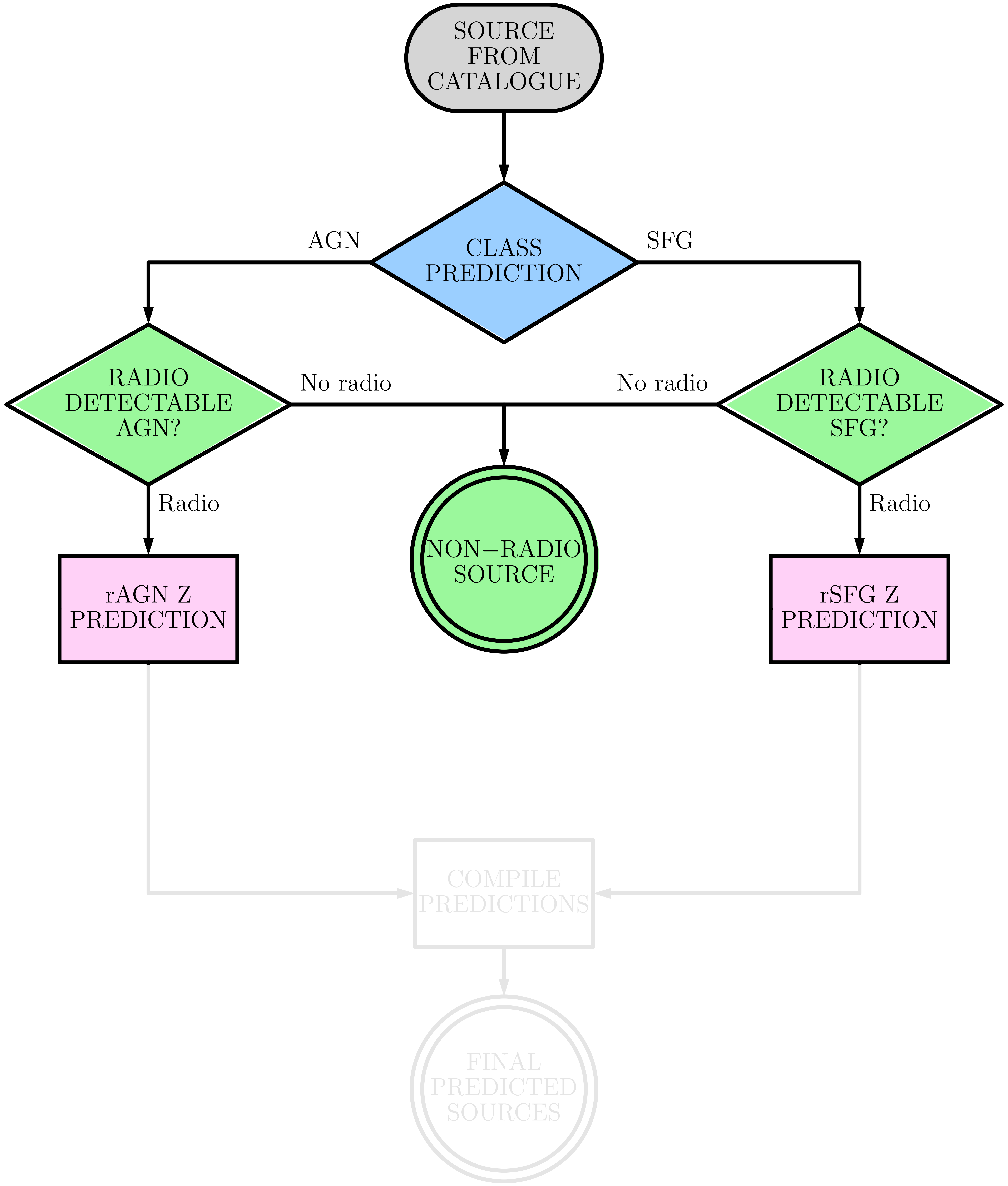
OUR APPROACH TO SELECTING AGN AND SFG WITH ML

Three different levels

Classify as AGN or SFG

Select radio-detectable sources

Estimate redshift for radio-detectable sources



PREDICTION PIPELINE

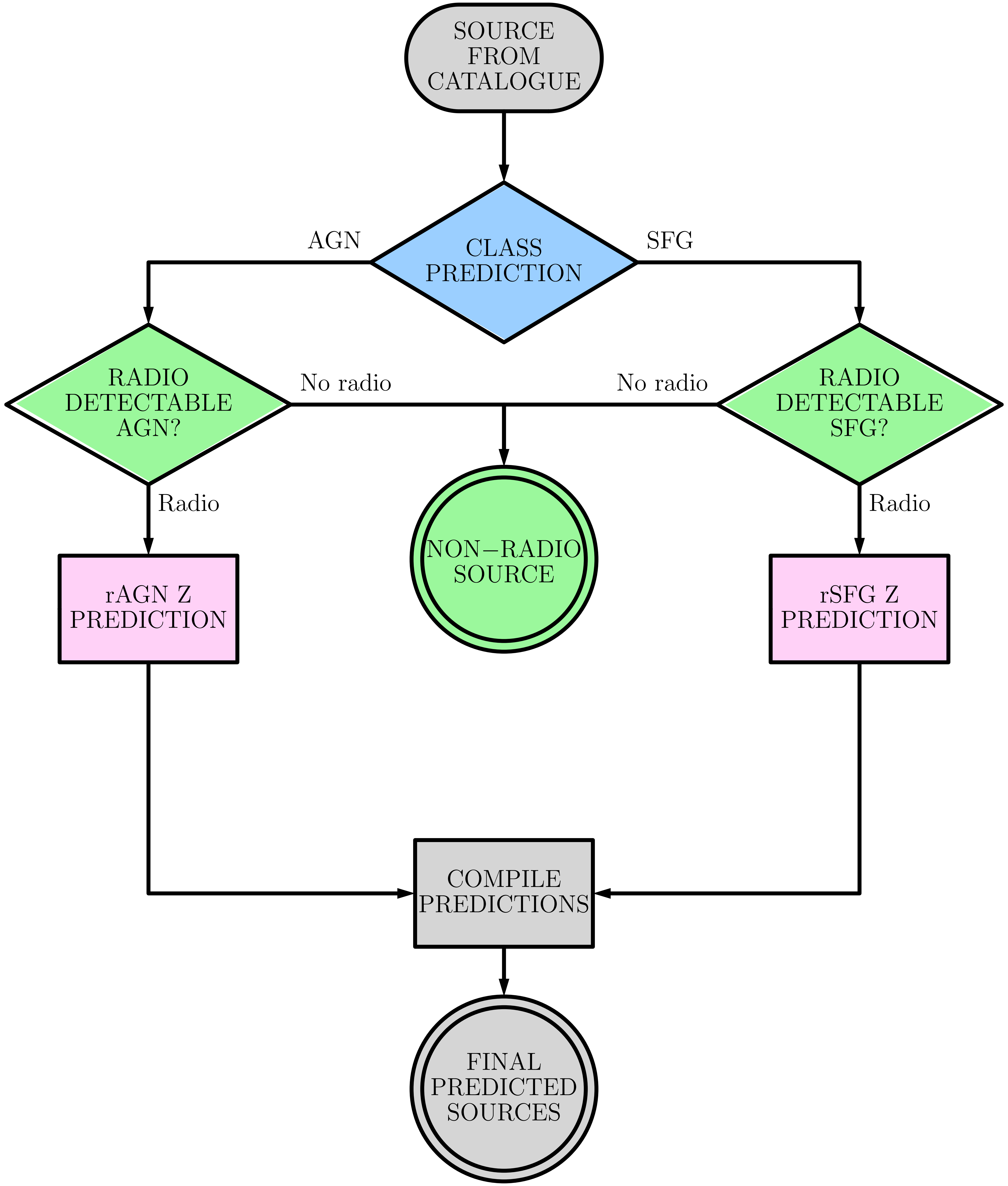
OUR APPROACH TO SELECTING AGN AND SFG WITH ML

Three different levels

Classify as AGN or SFG

Select radio-detectable sources

Estimate redshift for radio-detectable sources

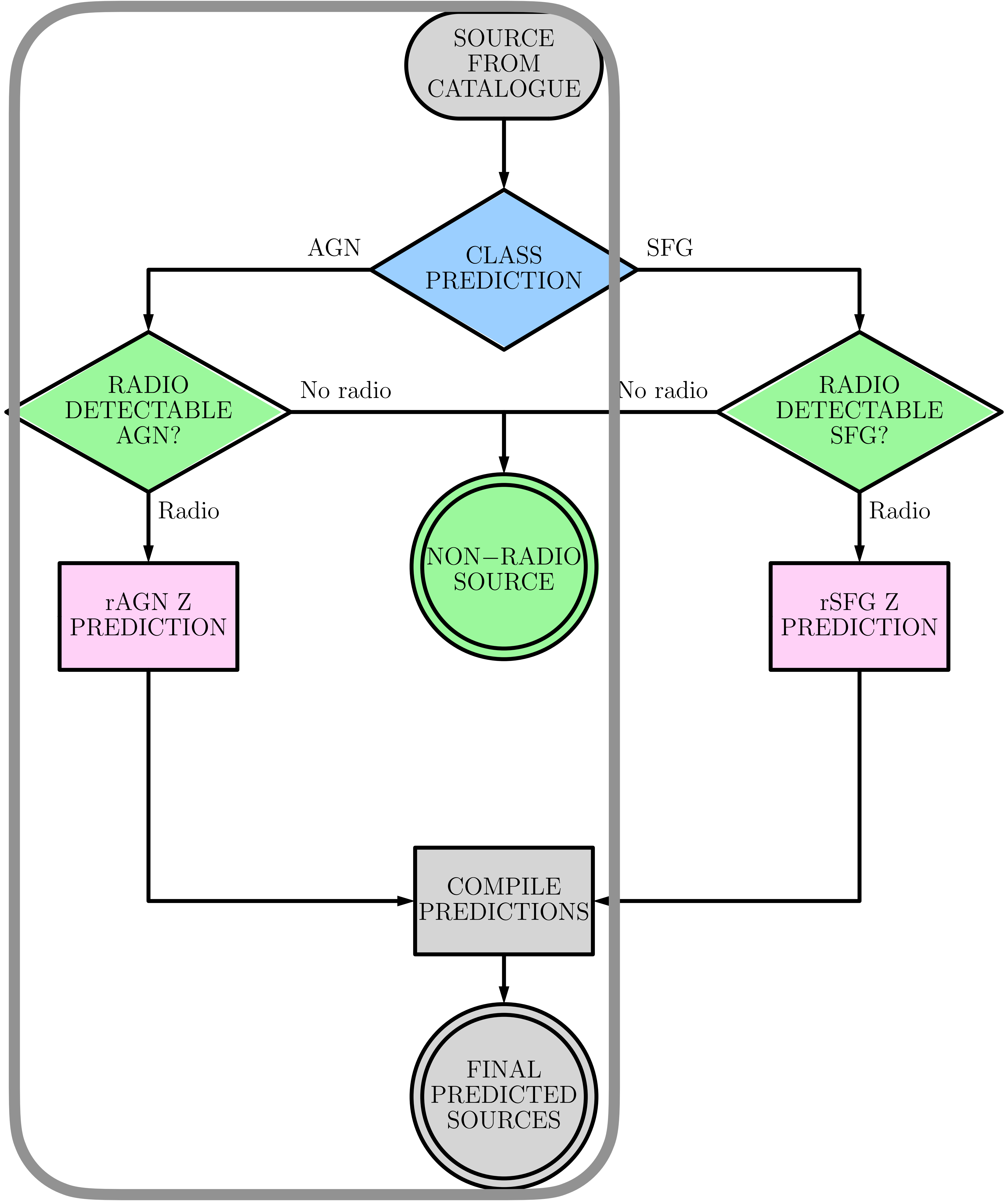


Prediction Pipeline

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



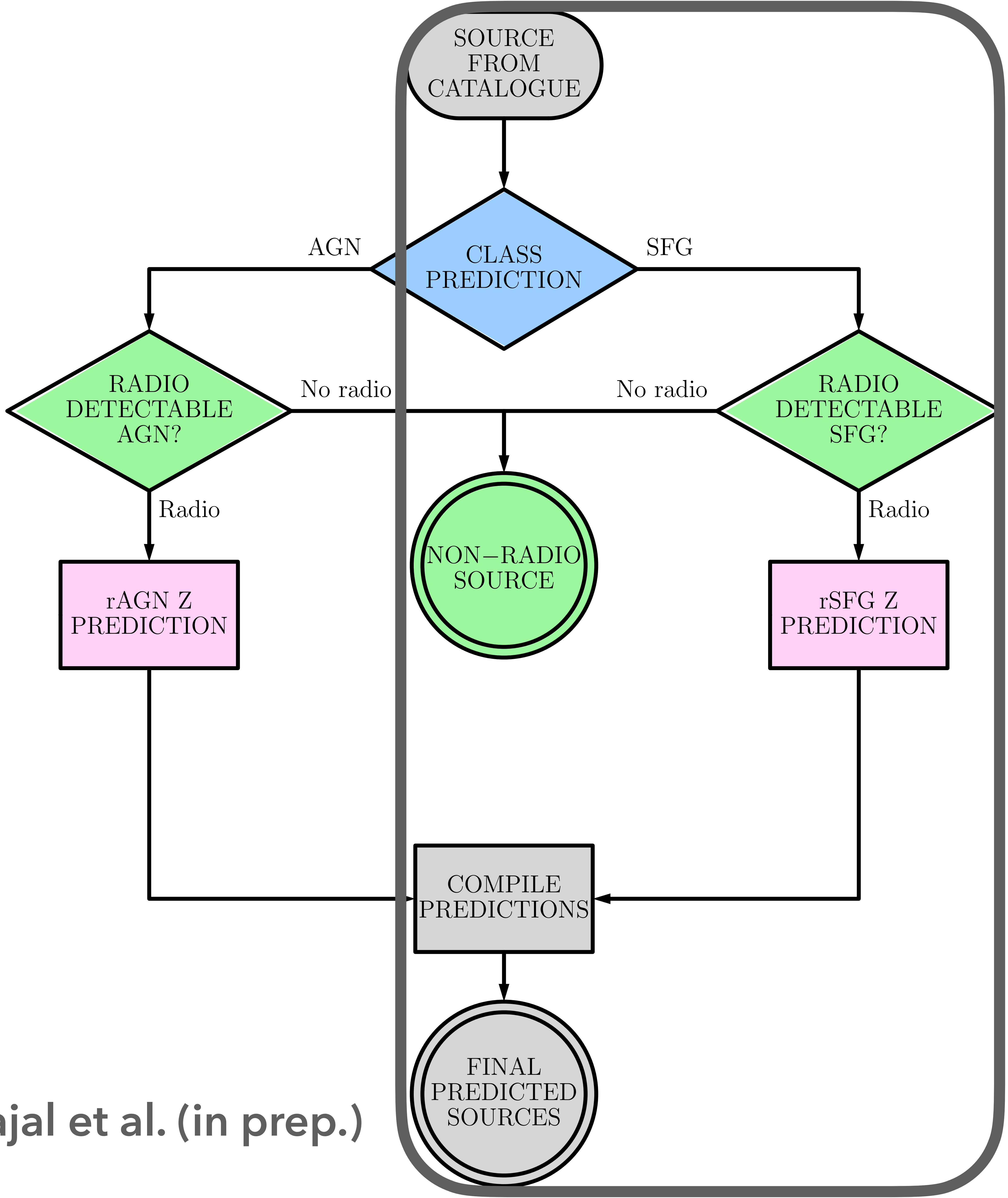
Carvajal et al. (2021)
+
Carvajal et al. (2023)

Prediction Pipeline

Our approach to selecting AGN and SFG with ML

Three different levels

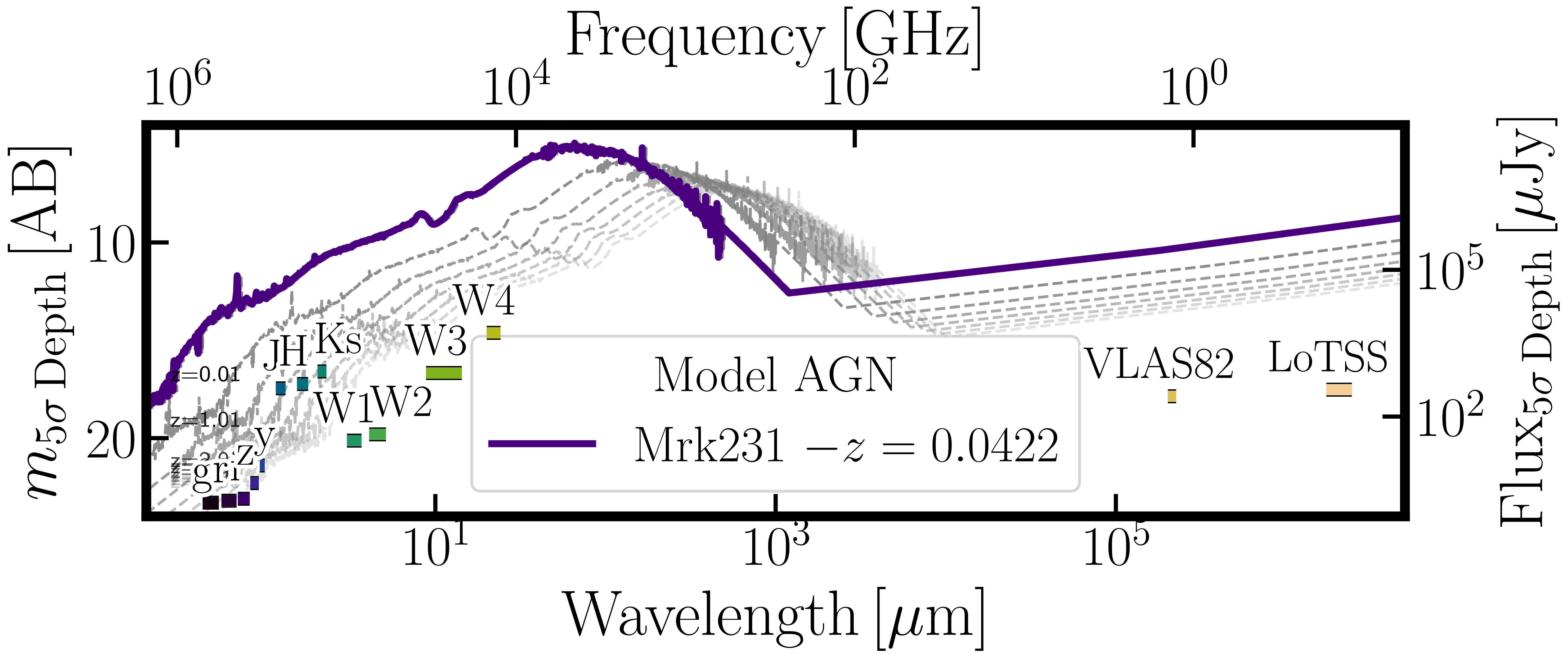
- Classify as AGN or SFG
- Select radio-detectable sources
- Estimate redshift for radio-detectable sources



Carvajal et al. (in prep.)

OUR DATA

DATASET – PHOTOMETRY



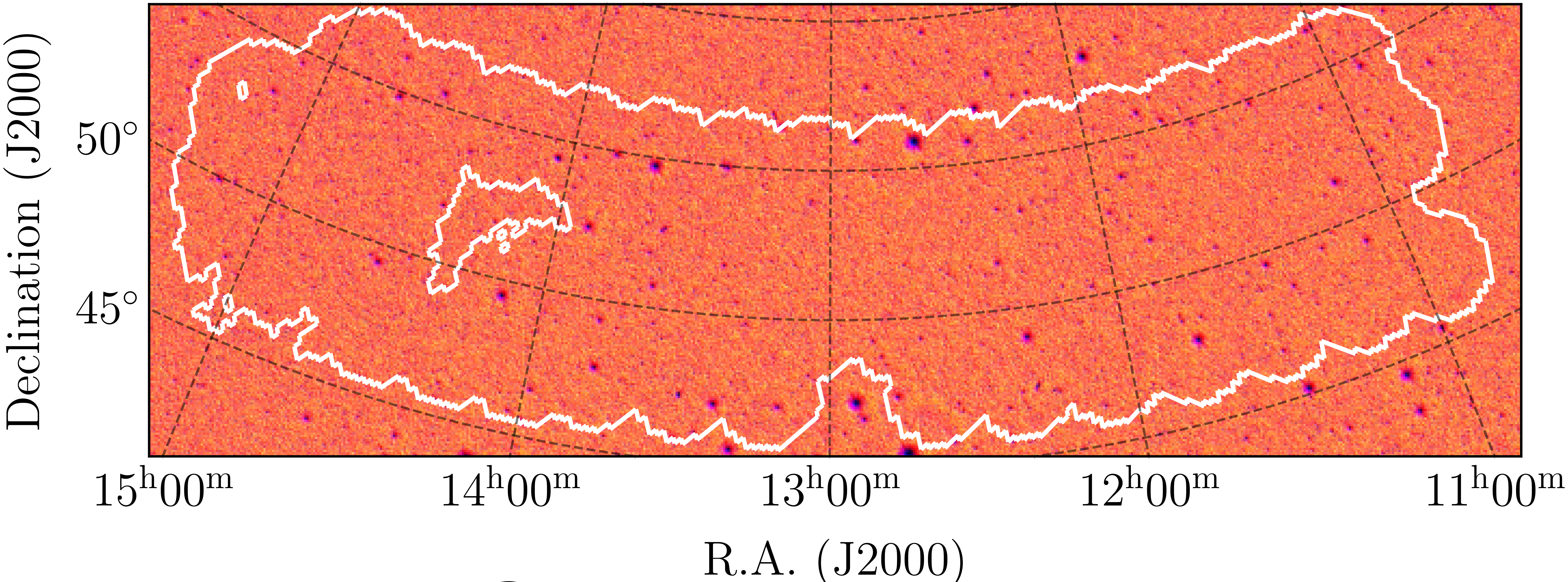
Carvajal et al. (2023)

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, 71μJy, 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%

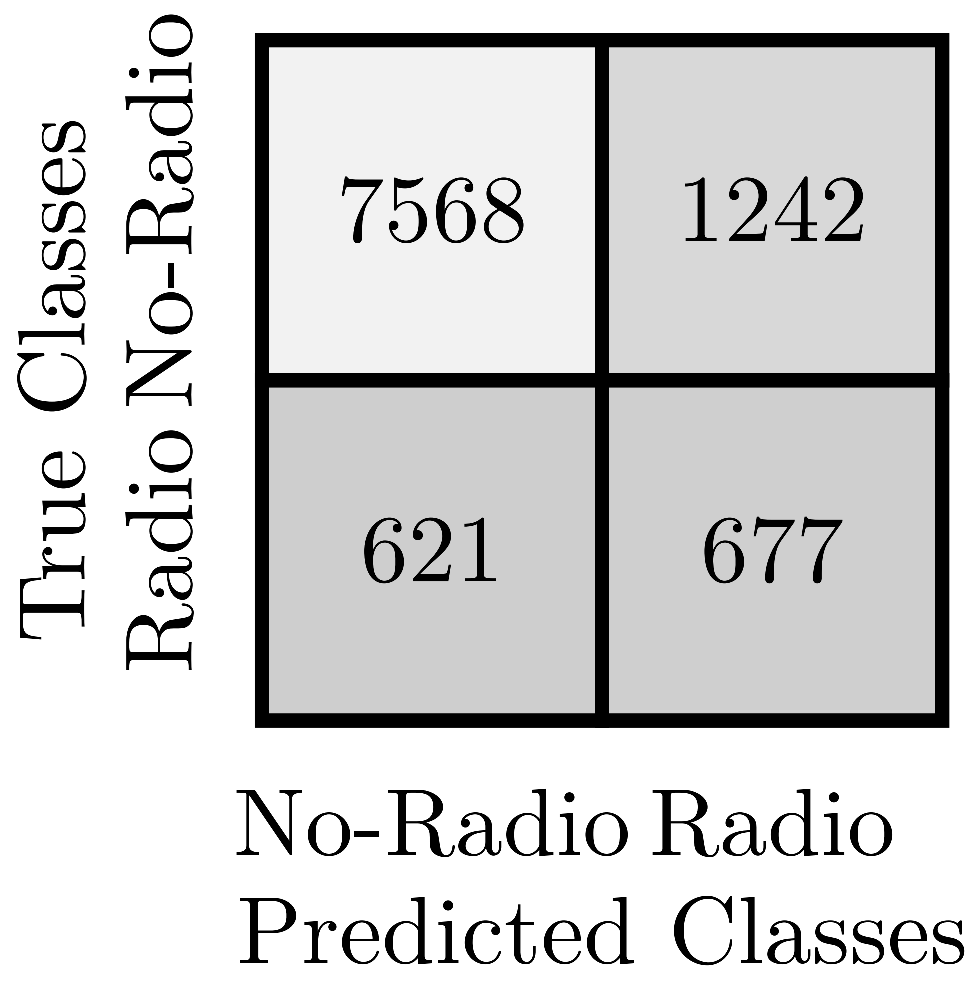
SFG Selection Completeness: 96 %

Increase from Baseline SFG fraction of 57%

True Classes \ Predicted Classes	SFG	AGN
SFG	13 072	567
AGN	383	9725

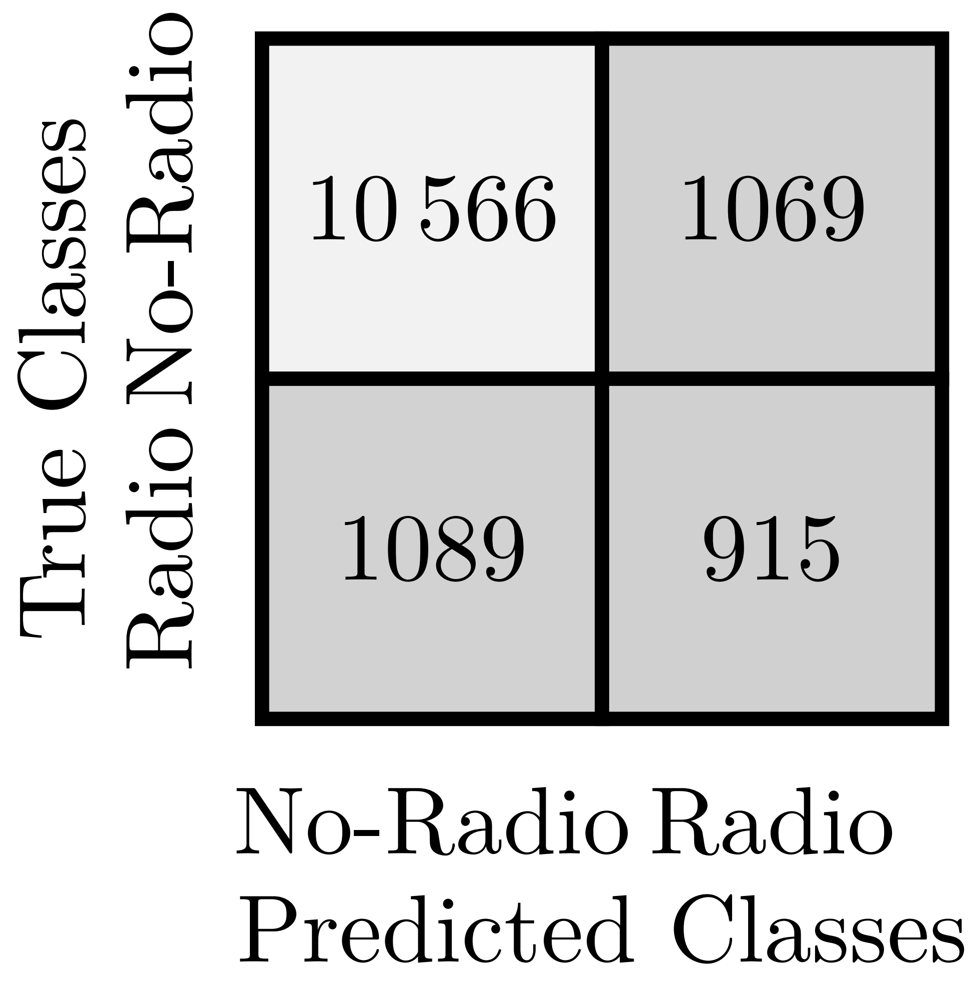
Radio Selection in AGN
Completeness: 52 %

Increase from Baseline radio fraction in AGN of 13%



Radio Selection in SFG
Completeness: 46 %

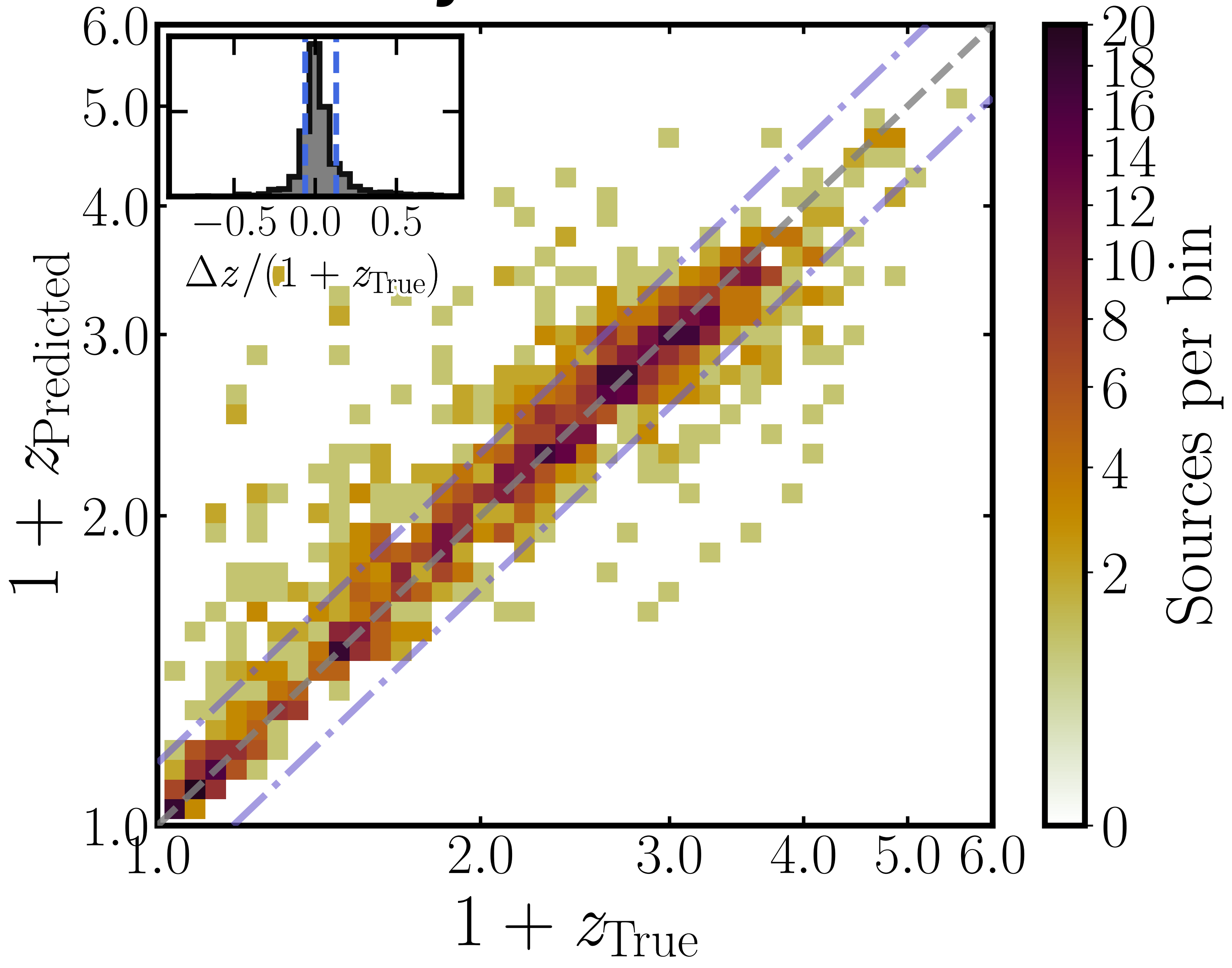
Increase from Baseline radio fraction in SFGs of 9%



Redshift in rAGN: 81 % Accurate

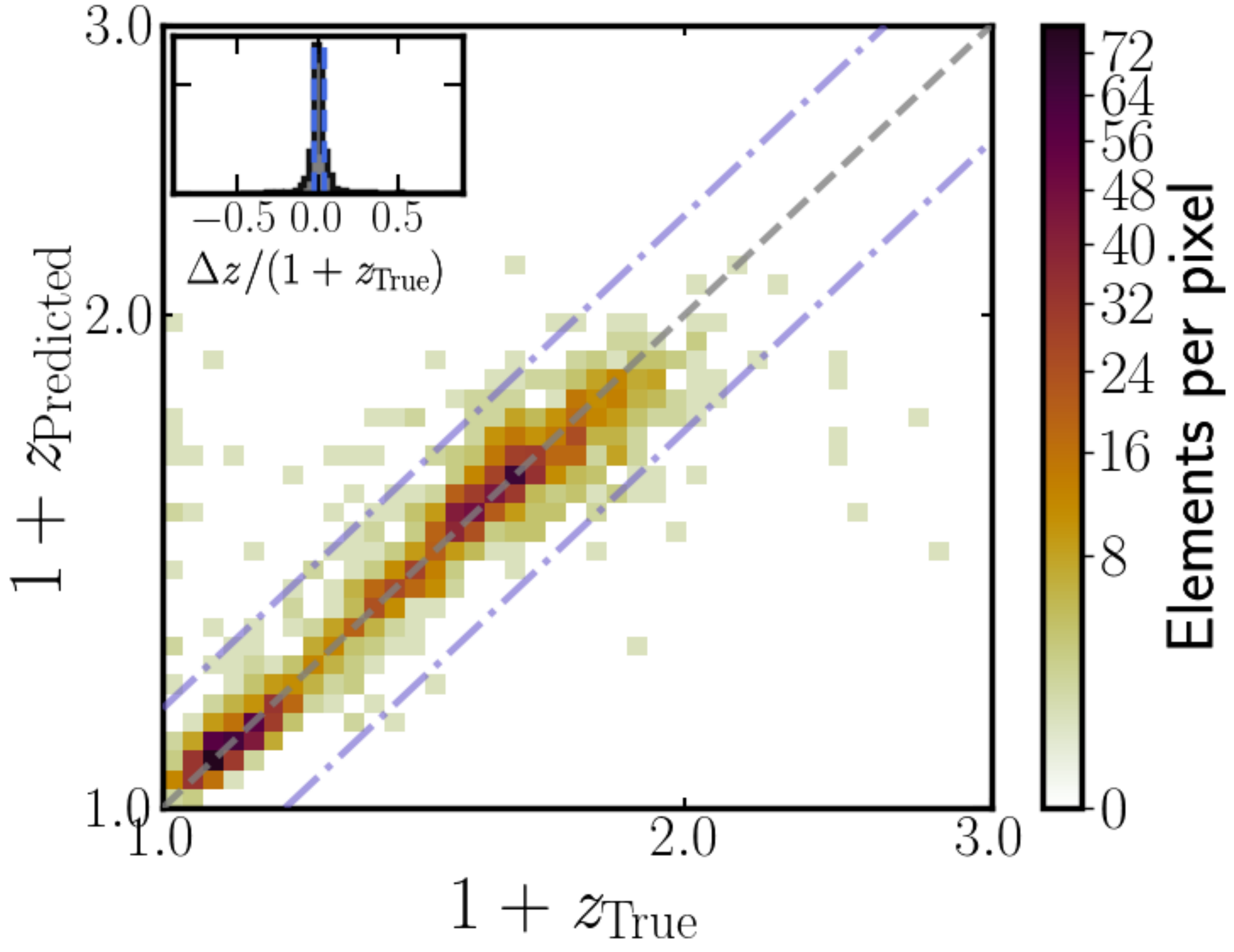
Outlier fraction of 19%

Carvajal et al. (2023)



Redshift in rSFG: 97 % Accurate

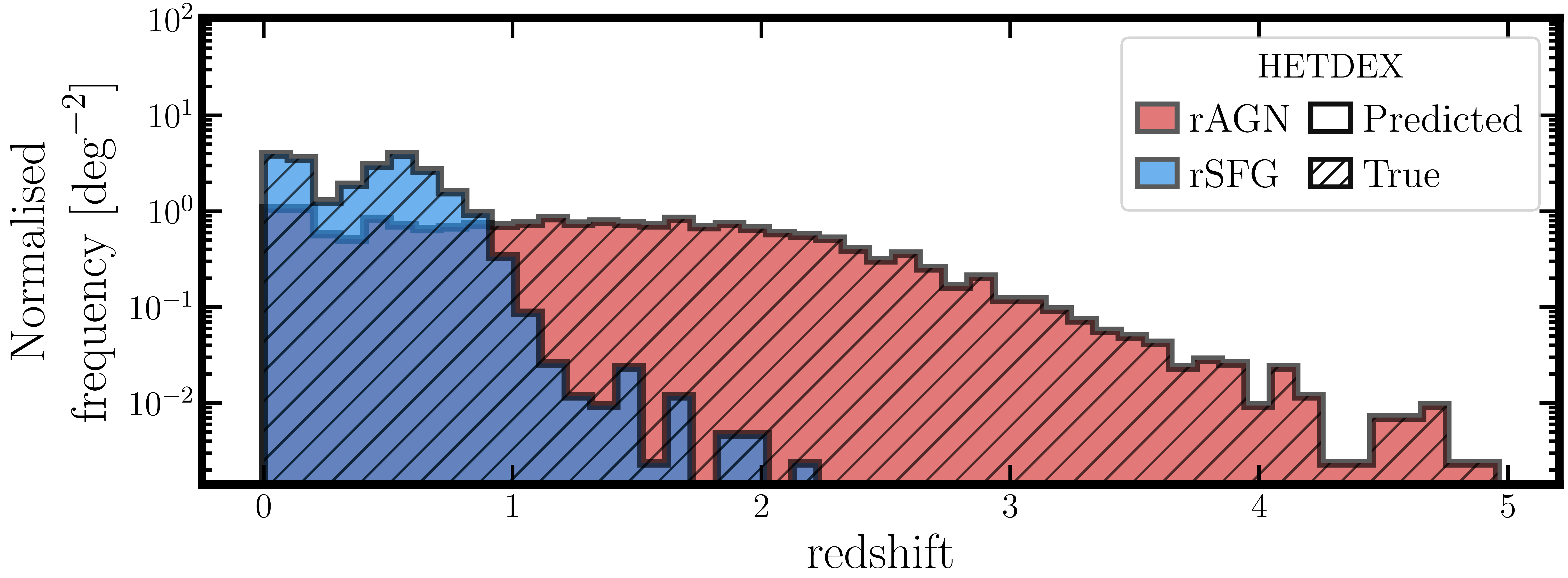
Outlier fraction of 3%



Applying pipeline to full HETDEX dataset

New sample of 68k rAGN candidates 115k rSFG candidates

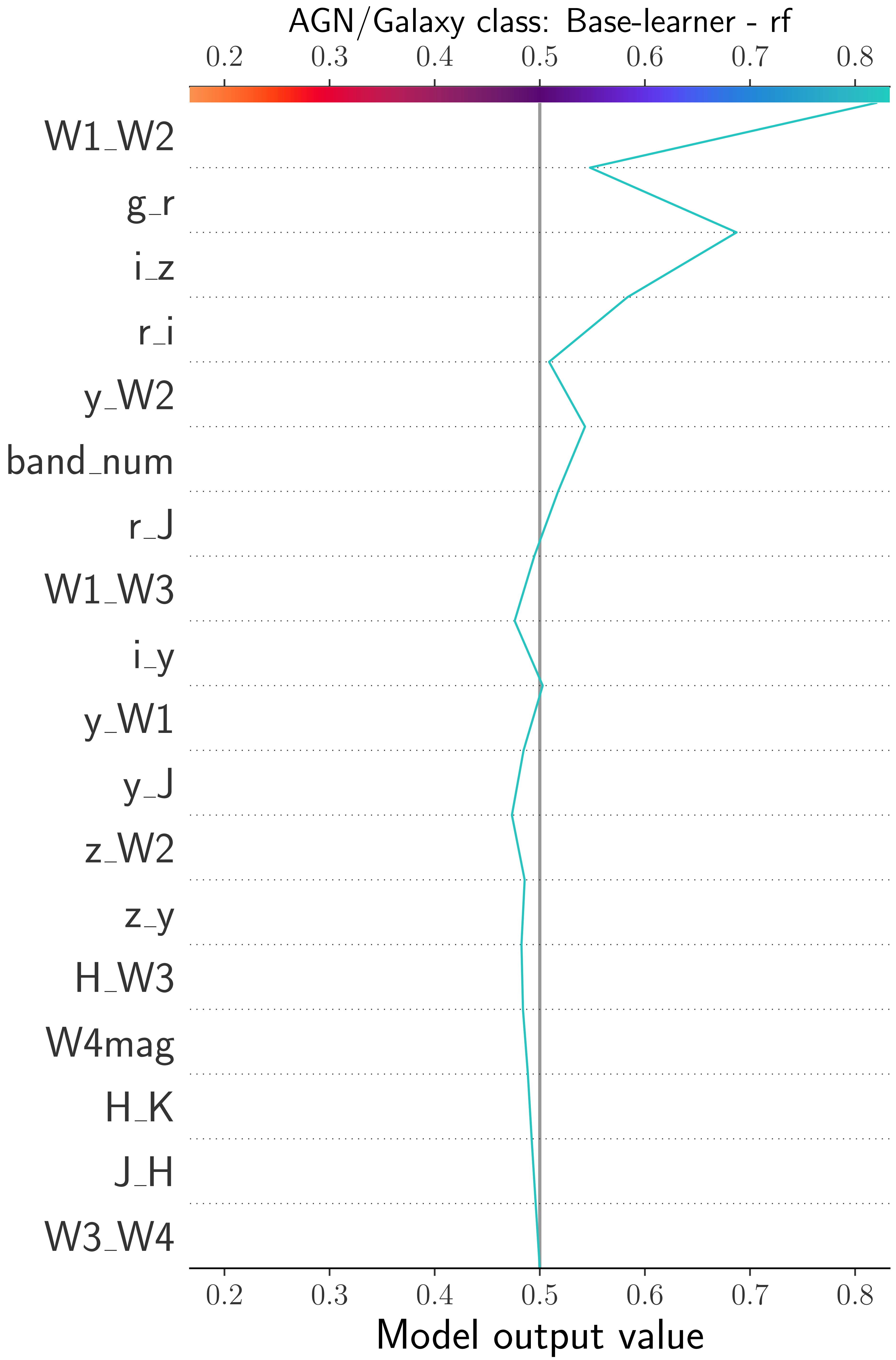
More than 10x original sample (6.4k rAGN and 6.6k rSFG)



Understanding our predictions

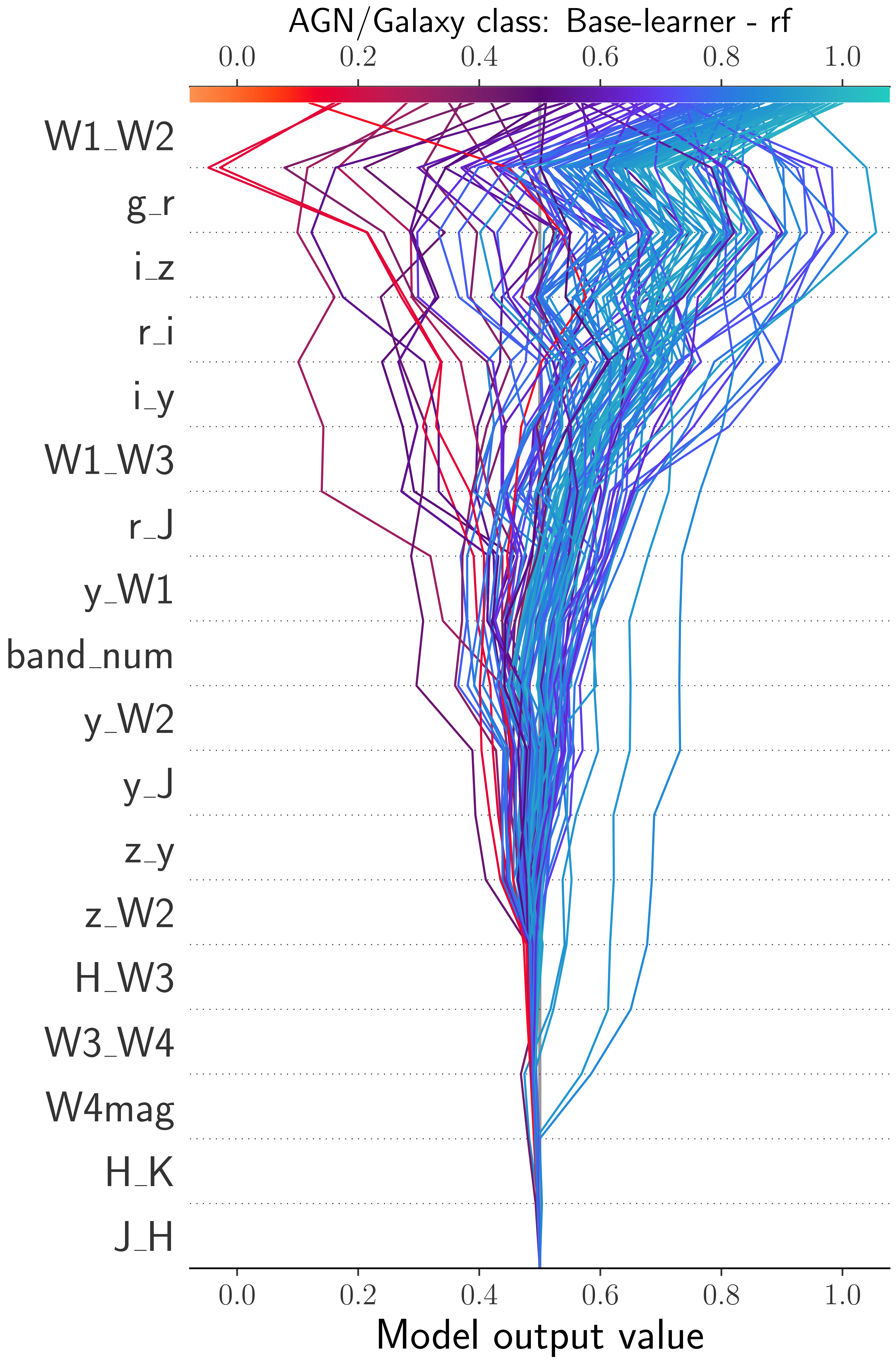
Feature Importance

Understand which features drive predictions more strongly



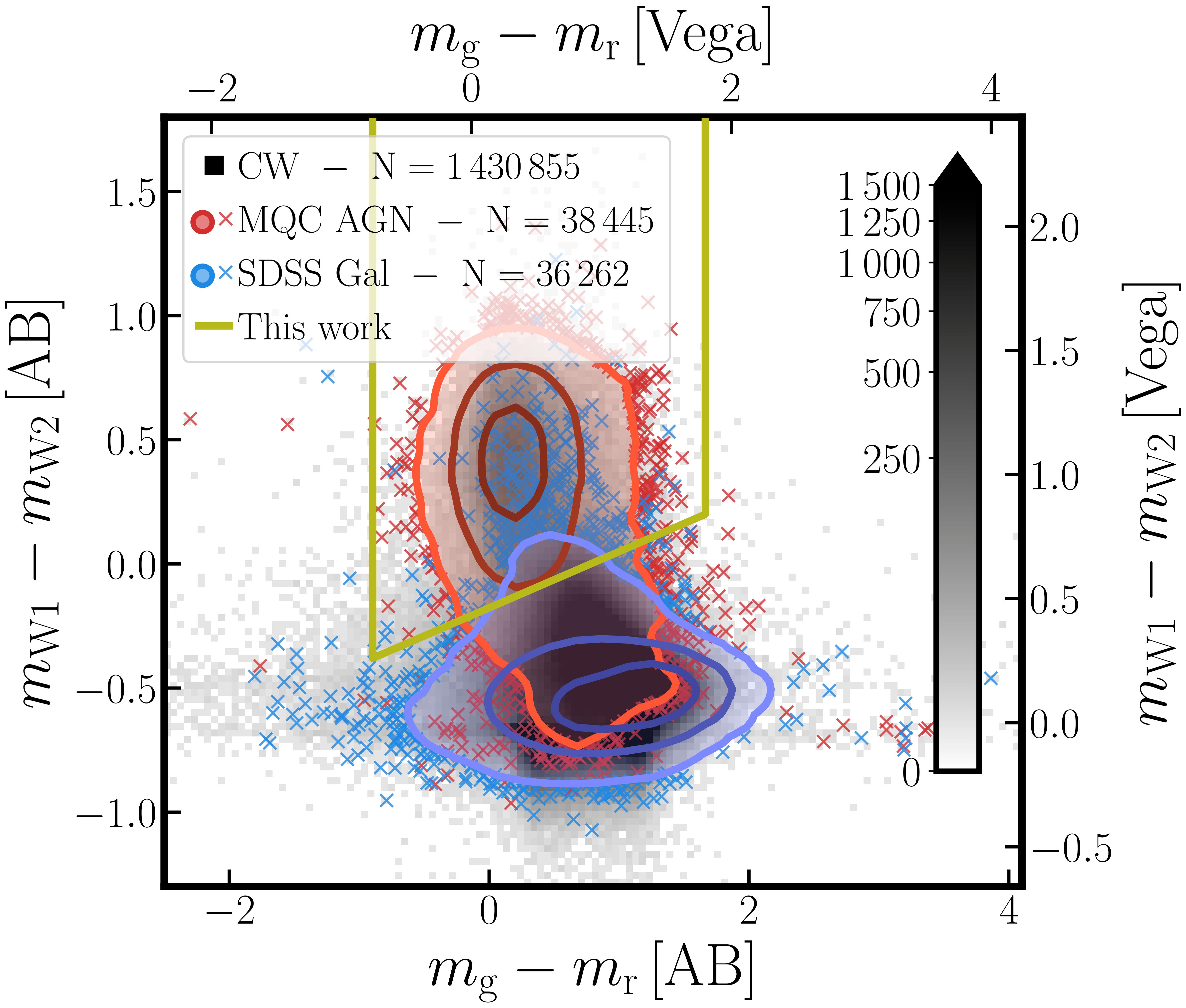
Feature Importance

Understand which features drive predictions more strongly



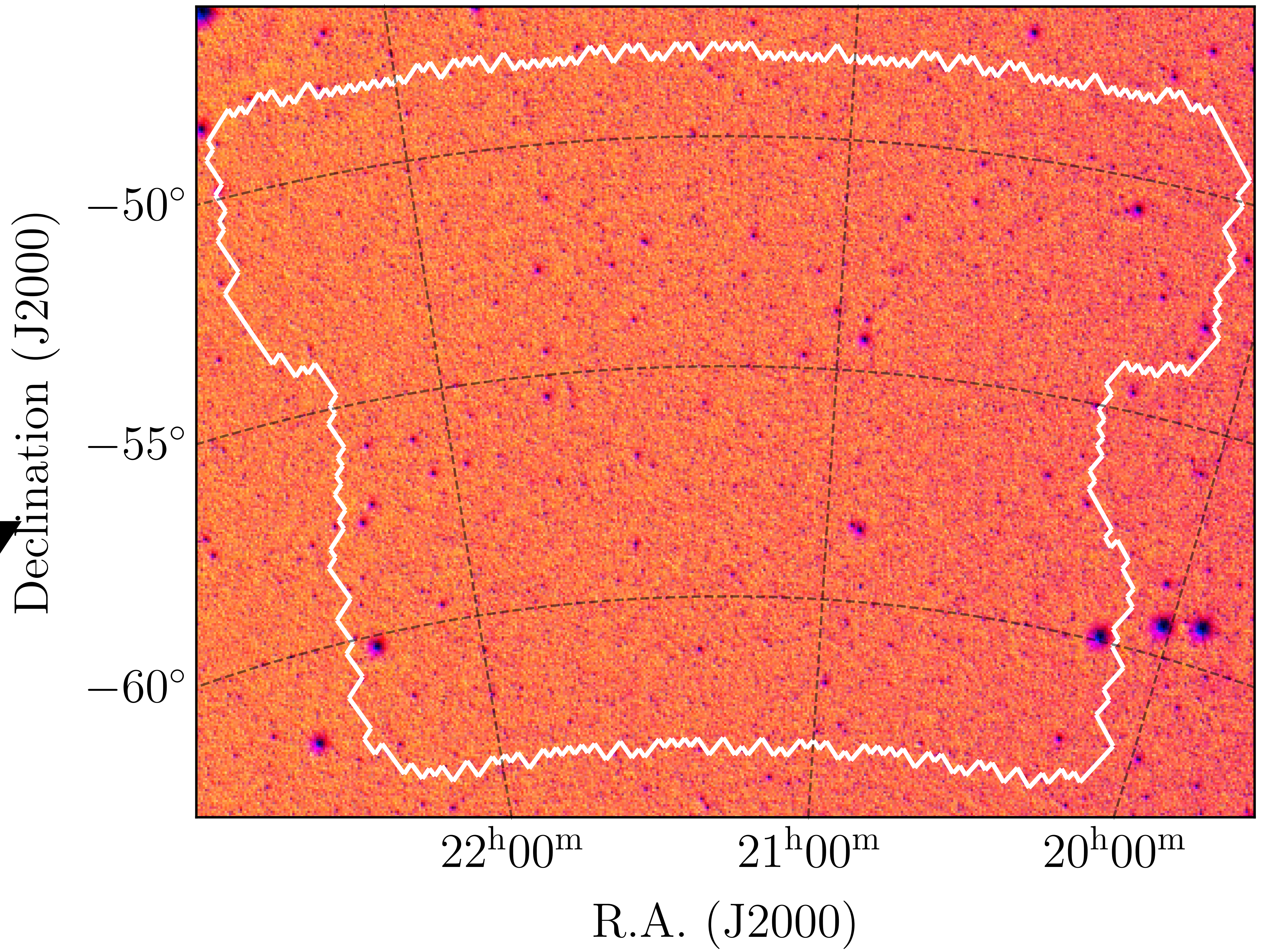
ML-BASED COLOUR-COLOUR DIAGRAM

g-r – W1-W2 DIAGRAM



Carvajal et al. (2023)

APPLY PIPELINE IN THE SOUTHERN SKY



EMU Pilot Survey (EMU-PS) – 270 deg² –
18 arcsec resolution – 25-30 µJy rms
SKA-like conditions

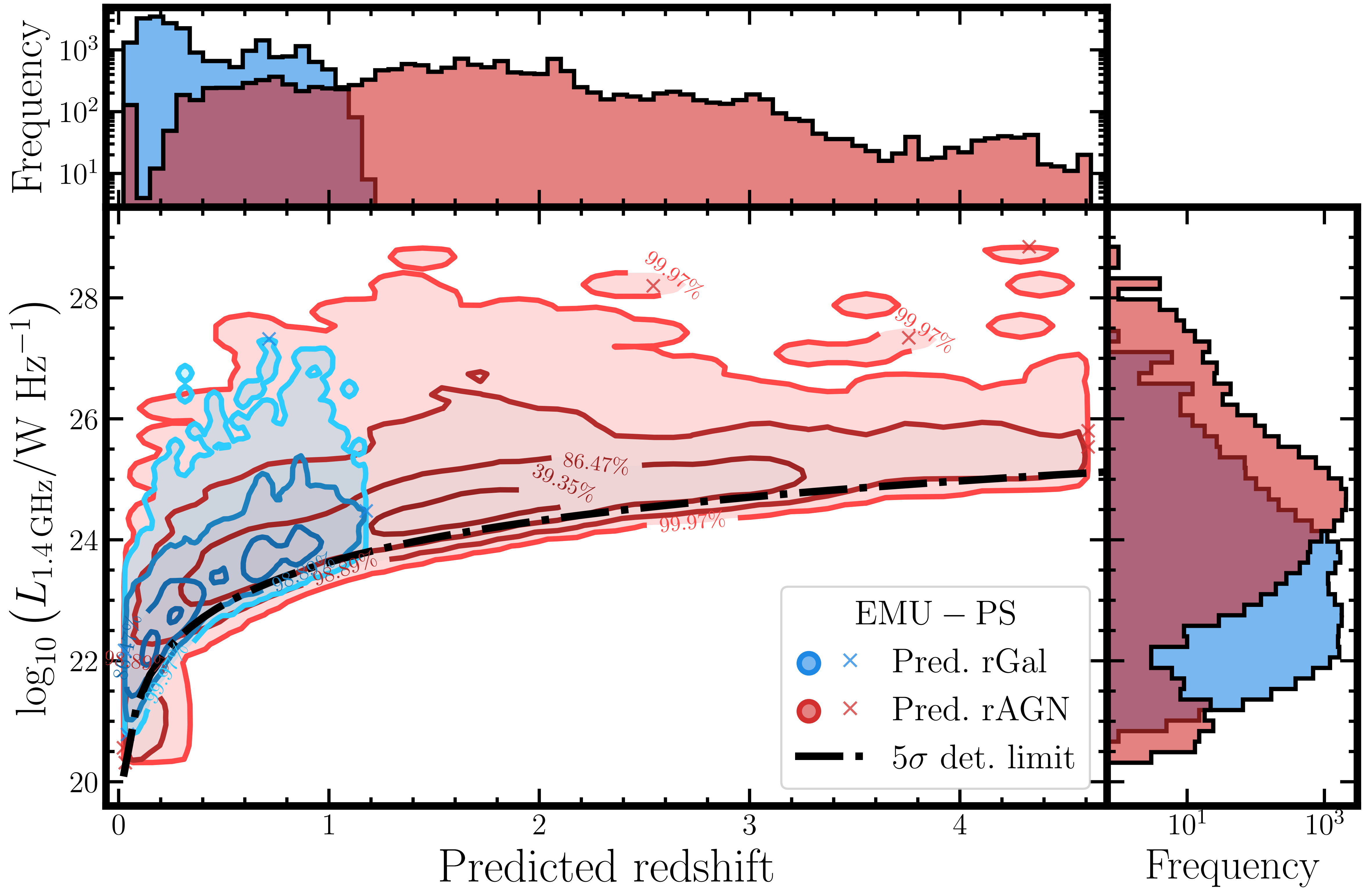
EMU-PS

More than 10x new rAGN and rSFG candidates

rAGN: 92 113 (originally 2367)

rSFG: 128 249 (originally 870)

Such numbers allow for population studies



Radio Luminosity Function

Radio Luminosity Function

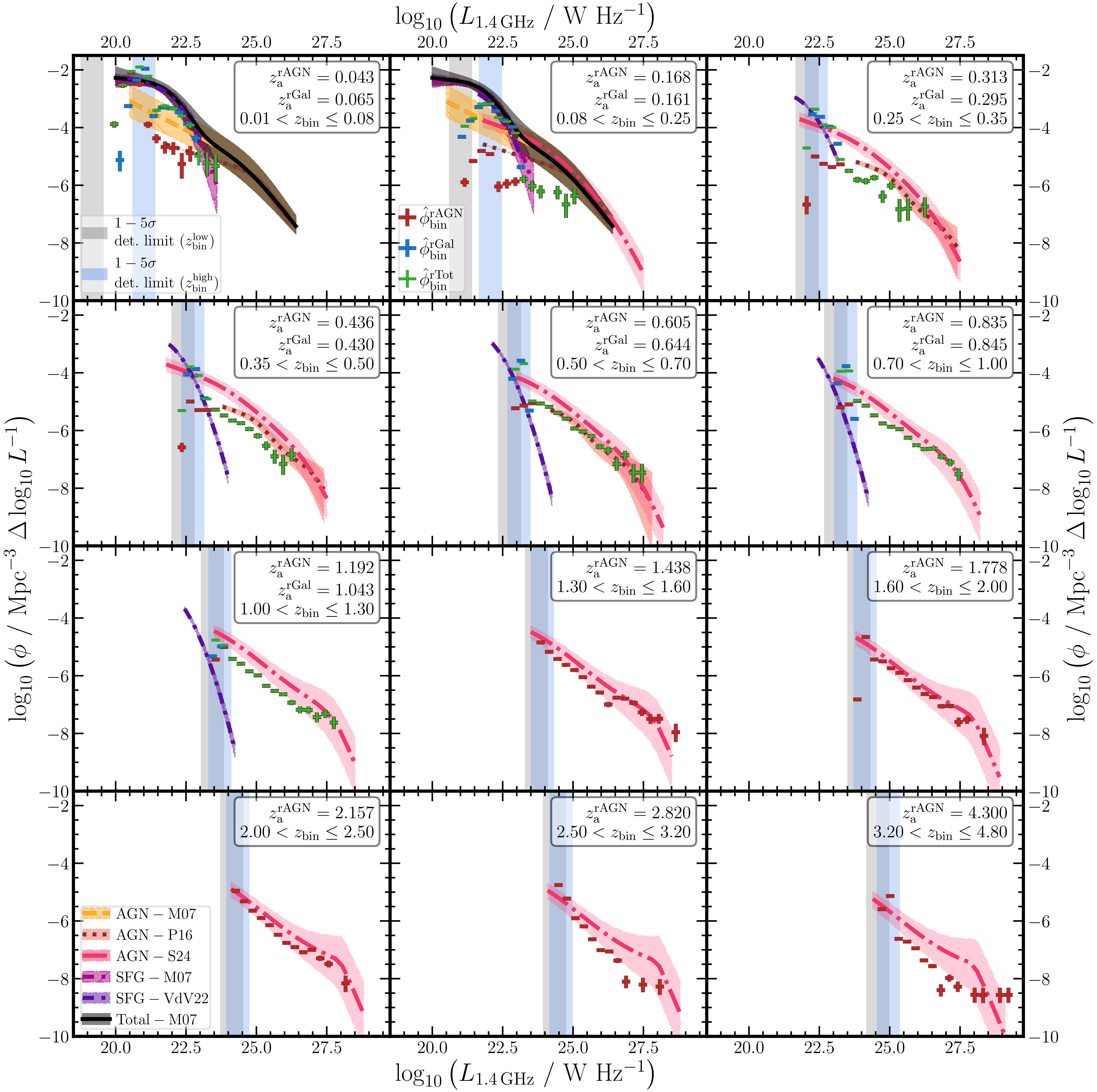
Candidate SFGs and AGN from pipeline in EMU-PS

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



$\log_{10}(L_{1.4\,\mathrm{GHz}}$ / W Hz$^{-1})$
$\log_{10}(\phi$ / Mpc^{-3} $\Delta\log_{10} L^{-1})$
$z_a^{\mathrm{rAGN}} = 0.043$
$z_a^{\mathrm{rGal}} = 0.065$
$0.01 < z_{\mathrm{bin}} \leq 0.08$
$z_a^{\mathrm{rAGN}} = 0.168$
$z_a^{\mathrm{rGal}} = 0.161$
$0.08 < z_{\mathrm{bin}} \leq 0.25$
$z_a^{\mathrm{rAGN}} = 0.313$
$z_a^{\mathrm{rGal}} = 0.295$
$0.25 < z_{\mathrm{bin}} \leq 0.35$
$z_a^{\mathrm{rAGN}} = 0.436$
$z_a^{\mathrm{rGal}} = 0.430$
$0.35 < z_{\mathrm{bin}} \leq 0.50$
$z_a^{\mathrm{rAGN}} = 0.605$
$z_a^{\mathrm{rGal}} = 0.644$
$0.50 < z_{\mathrm{bin}} \leq 0.70$
$z_a^{\mathrm{rAGN}} = 0.835$
$z_a^{\mathrm{rGal}} = 0.845$
$0.70 < z_{\mathrm{bin}} \leq 1.00$
$z_a^{\mathrm{rAGN}} = 1.192$
$z_a^{\mathrm{rGal}} = 1.043$
$1.00 < z_{\mathrm{bin}} \leq 1.30$
$z_a^{\mathrm{rAGN}} = 1.438$
$1.30 < z_{\mathrm{bin}} \leq 1.60$
$z_a^{\mathrm{rAGN}} = 1.778$
$1.60 < z_{\mathrm{bin}} \leq 2.00$
$z_a^{\mathrm{rAGN}} = 2.157$
$2.00 < z_{\mathrm{bin}} \leq 2.50$
$z_a^{\mathrm{rAGN}} = 2.820$
$2.50 < z_{\mathrm{bin}} \leq 3.20$
$z_a^{\mathrm{rAGN}} = 4.300$
$3.20 < z_{\mathrm{bin}} \leq 4.80$
$1-5\sigma$ det. limit ($z_{\mathrm{bin}}^{\mathrm{low}}$)
$1-5\sigma$ det. limit ($z_{\mathrm{bin}}^{\mathrm{high}}$)
$\hat{\phi}_{\mathrm{bin}}^{\mathrm{rAGN}}$
$\hat{\phi}_{\mathrm{bin}}^{\mathrm{rGal}}$
$\hat{\phi}_{\mathrm{bin}}^{\mathrm{rTot}}$
AGN – M07
AGN – P16
AGN – S24
SFG – M07
SFG – VdV22
Total – M07

$\log_{10}\left(L_{1.4\,\mathrm{GHz}}\ /\ \mathrm{W\ Hz}^{-1}\right)$
$\log_{10}\left(\phi\ /\ \mathrm{Mpc}^{-3}\ \Delta\log_{10}L^{-1}\right)$
$z_a^{\mathrm{rAGN}} = 0.043$ $z_a^{\mathrm{rGal}} = 0.065$ $0.01 < z_{\mathrm{bin}} \leq 0.08$
$z_a^{\mathrm{rAGN}} = 0.168$ $z_a^{\mathrm{rGal}} = 0.161$ $0.08 < z_{\mathrm{bin}} \leq 0.25$
$z_a^{\mathrm{rAGN}} = 0.313$ $z_a^{\mathrm{rGal}} = 0.295$ $0.25 < z_{\mathrm{bin}} \leq 0.35$
$z_a^{\mathrm{rAGN}} = 0.436$ $z_a^{\mathrm{rGal}} = 0.430$ $0.35 < z_{\mathrm{bin}} \leq 0.50$
$z_a^{\mathrm{rAGN}} = 0.605$ $z_a^{\mathrm{rGal}} = 0.644$ $0.50 < z_{\mathrm{bin}} \leq 0.70$
$z_a^{\mathrm{rAGN}} = 0.835$ $z_a^{\mathrm{rGal}} = 0.845$ $0.70 < z_{\mathrm{bin}} \leq 1.00$
$1-5\sigma$ det. limit ($z_{\mathrm{bin}}^{\mathrm{low}}$)
$1-5\sigma$ det. limit ($z_{\mathrm{bin}}^{\mathrm{high}}$)
$\hat{\phi}_{\mathrm{bin}}^{\mathrm{rAGN}}$
$\hat{\phi}_{\mathrm{bin}}^{\mathrm{rGal}}$
$\hat{\phi}_{\mathrm{bin}}^{\mathrm{rTot}}$
AGN – M07
AGN – P16
AGN – S24
SFG – M07
SFG – VdV22
Total – M07

$z_a^{rAGN} = 1.192$
$z_a^{rGal} = 1.043$
$1.00 < z_{bin} \leq 1.30$
$z_a^{rAGN} = 1.438$
$1.30 < z_{bin} \leq 1.60$
$z_a^{rAGN} = 1.778$
$1.60 < z_{bin} \leq 2.00$
$z_a^{rAGN} = 2.157$
$2.00 < z_{bin} \leq 2.50$
$z_a^{rAGN} = 2.820$
$2.50 < z_{bin} \leq 3.20$
$z_a^{rAGN} = 4.300$
$3.20 < z_{bin} \leq 4.80$
$1-5\sigma$ det. limit (z_{bin}^{low})
$1-5\sigma$ det. limit (z_{bin}^{high})
AGN – M07
AGN – P16
AGN – S24
SFG – M07
SFG – VdV22
Total – M07
$\hat{\phi}_{bin}^{rAGN}$
$\hat{\phi}_{bin}^{rGal}$
$\hat{\phi}_{bin}^{rTot}$
$\log_{10}(\phi$ / Mpc^{-3} $\Delta \log_{10} L^{-1})$
$\log_{10}(L_{1.4\,GHz}$ / W Hz$^{-1})$

In conclusion

CONCLUDING REMARKS

We 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

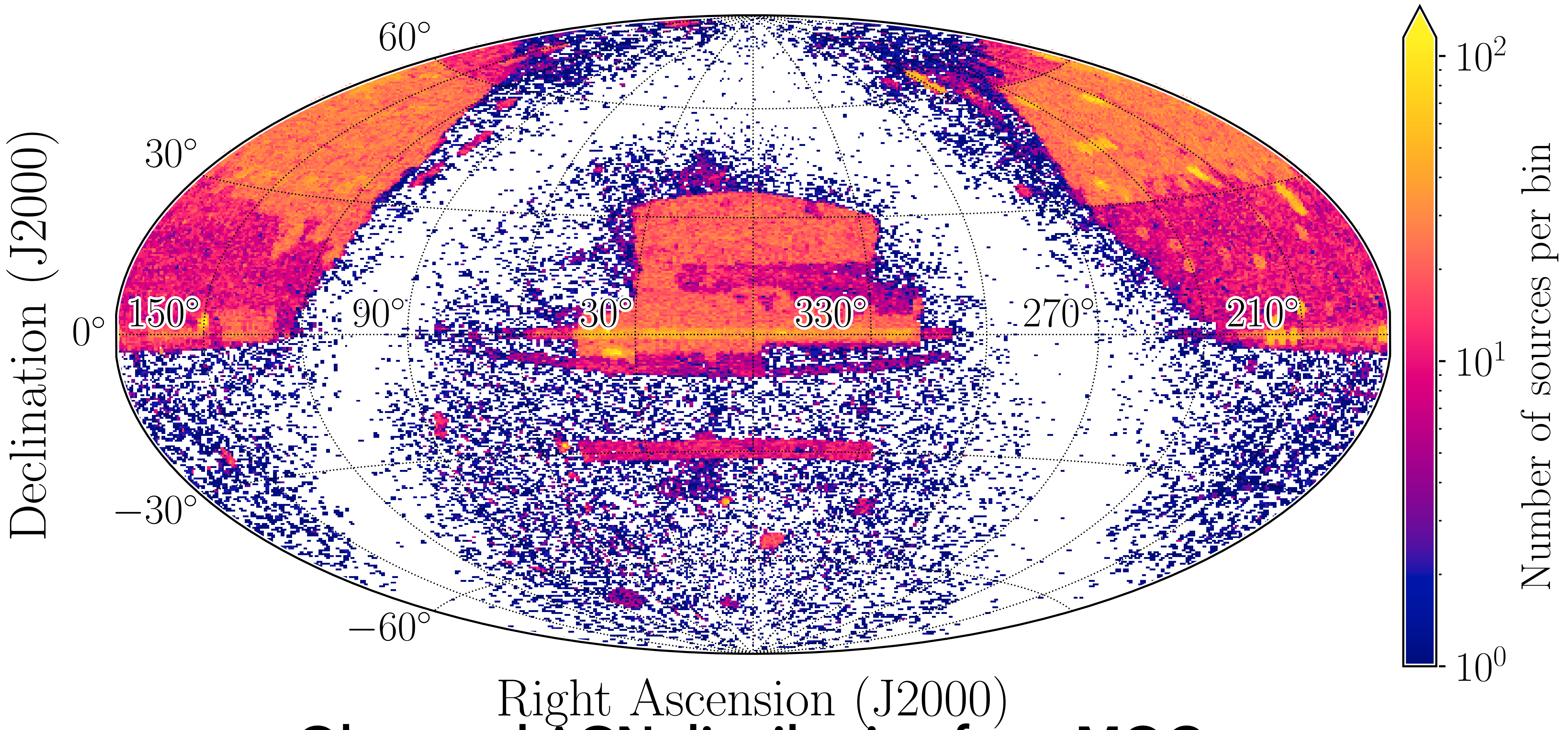
Where to go from here?

We 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



Observed AGN distribution from MQC

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

Cofinanciamento / Cofinancing