Search for neutrinos from AGN using a data-driven source selection

ERLANGEN CENTRE FOR ASTROPARTICLE PHYSICS

Sebastian Schindler
25th June 2024
MMS Annual Meeting 2024

SPONSORED BY THE
Federal Ministry of Education and Research

ICECUBE
SOUTH POLE NEUTRINO OBSERVATORY

ERLANGEN CENTRE FOR ASTROPARTICLE PHYSICS

FAU
Friedrich-Alexander-Universität
Naturwissenschaftliche Fakultät

Neutrinos from Active Galactic Nuclei

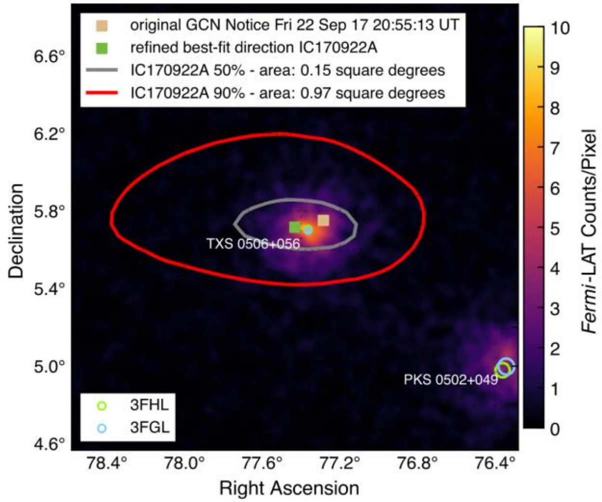
TXS 0506+056

significance: ~ 3 σ

type: ***Blazar***

Multimessenger observations of a flaring blazar coincident with high-energy neutrino IceCube-170922A

The IceCube Collaboration, *Fermi*-LAT, MAGIC, *AGILE*, ASAS-SN, HAWC, H.E.S.S., *INTEGRAL*, Kanata, Kiso, Kapteyn, Liverpool Telescope, Subaru, *Swift*/*NuSTAR*, VERITAS, and VLA/17B-403 teams*†



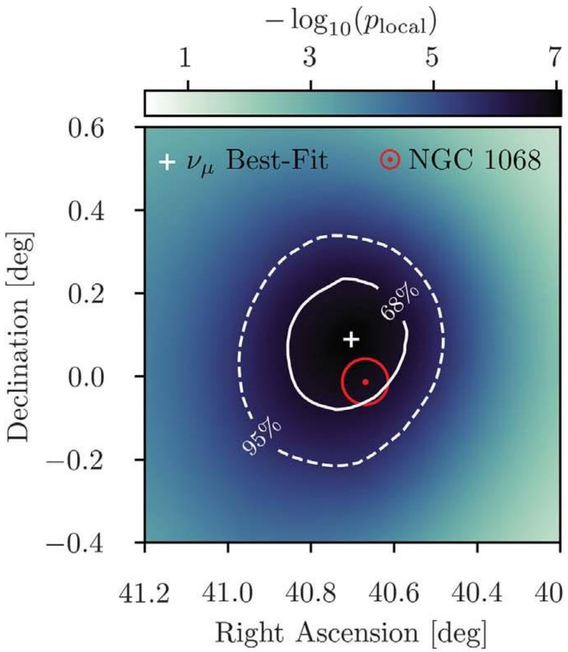
NGC 1068

significance: 4.2 σ

type: ***Seyfert 2 galaxy***

Evidence for neutrino emission from the nearby active galaxy NGC 1068

IceCube Collaboration*†

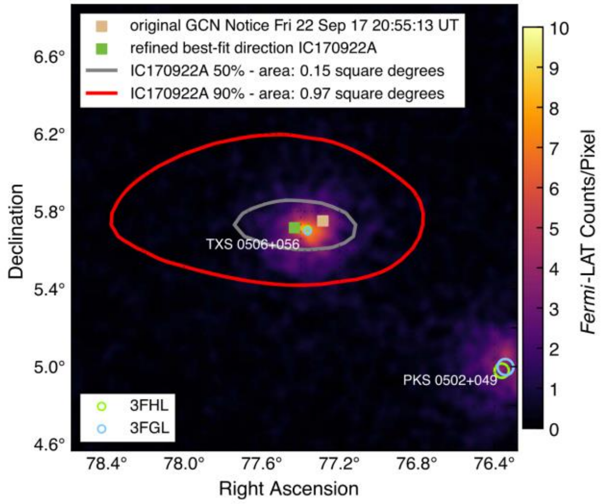


Neutrinos from Active Galactic Nuclei

TXS 0506+056
significance: ~ 3 σ
type: ***Blazar***

Multimessenger observations of a flaring blazar coincident with high-energy neutrino IceCube-170922A

The IceCube Collaboration, *Fermi*-LAT, MAGIC, *AGILE*, ASAS-SN, HAWC, H.E.S.S., *INTEGRAL*, Kanata, Kiso, Kapteyn, Liverpool Telescope, Subaru, *Swift/NuSTAR*, VERITAS, and VLA/17B-403 teams*†

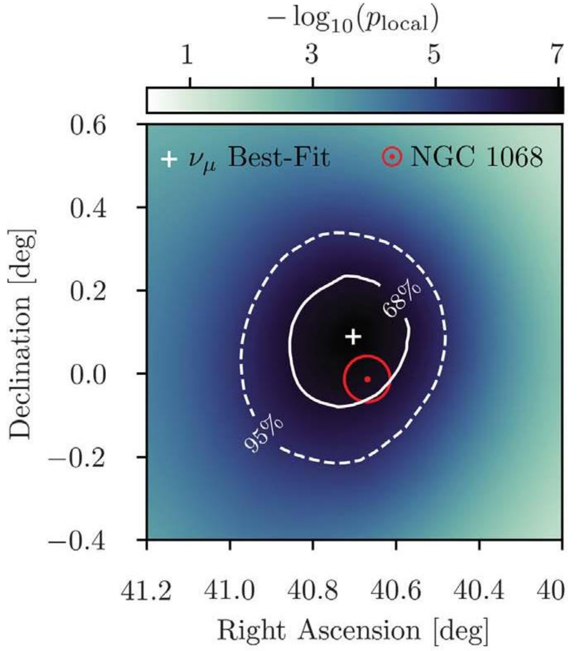


Active Galactic Nuclei

NGC 1068
significance: 4.2 σ
type: ***Seyfert 2 galaxy***

Evidence for neutrino emission from the nearby active galaxy NGC 1068

IceCube Collaboration*†



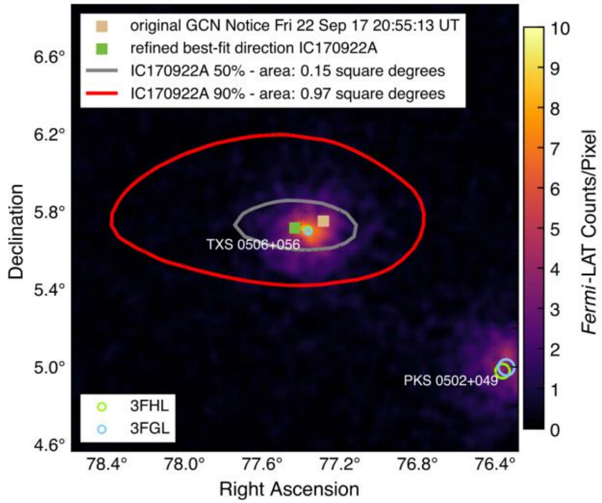
→ so… stacking search with AGN ???

Neutrinos from Active Galactic Nuclei

TXS 0506+056
significance: ~ 3 σ
type: ***Blazar***

Multimessenger observations of a flaring blazar coincident with high-energy neutrino IceCube-170922A

The IceCube Collaboration, *Fermi*-LAT, MAGIC, *AGILE*, ASAS-SN, HAWC, H.E.S.S., *INTEGRAL*, Kanata, Kiso, Kapteyn, Liverpool Telescope, Subaru, *Swift*/*NuSTAR*, VERITAS, and VLA/17B-403 teams*†

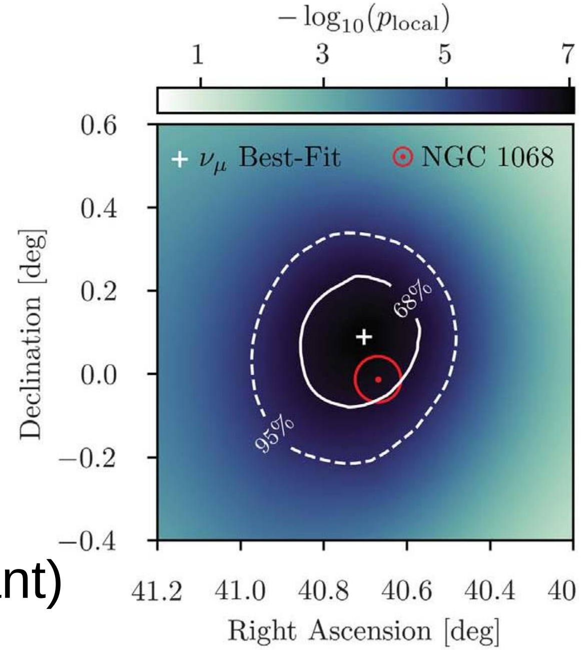


Active Galactic Nuclei

NGC 1068
significance: 4.2 σ
type: ***Seyfert 2 galaxy***

Evidence for neutrino emission from the nearby active galaxy NGC 1068

IceCube Collaboration*†

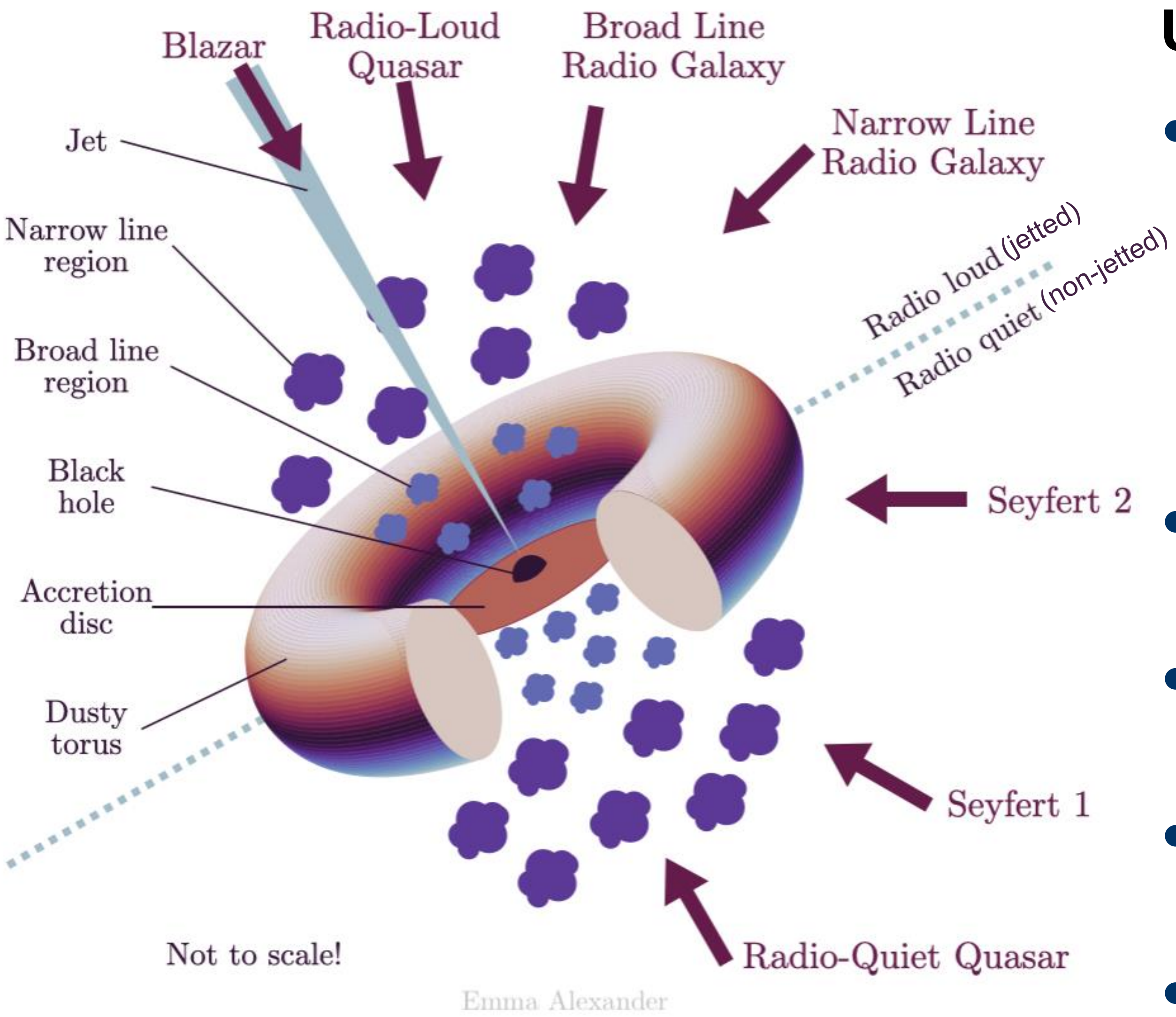


but which subset?

→ so… stacking search with AGN ???

(stack signal from (not-too-large) number of sources to become significant)

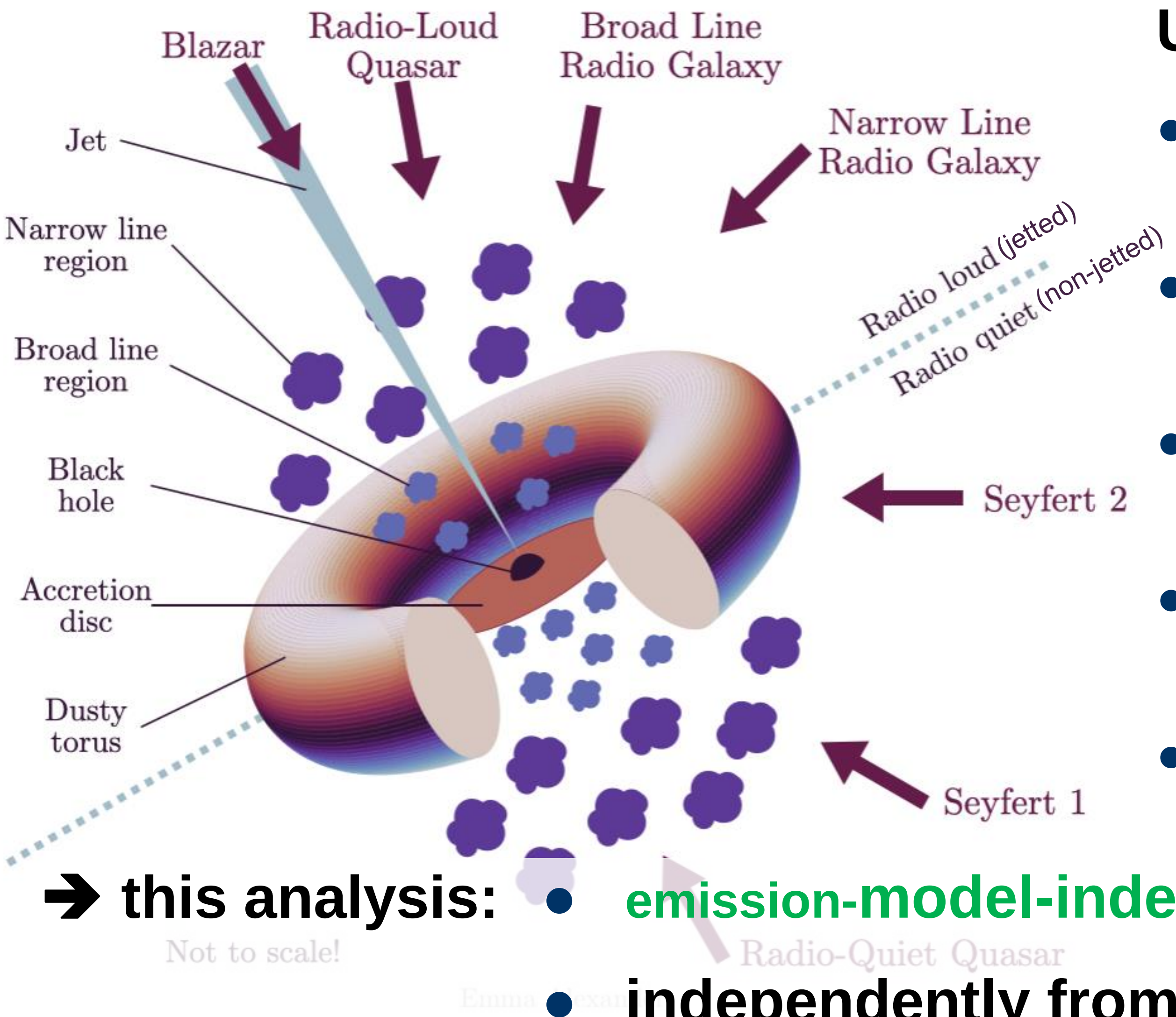
Classification of AGN as a problem



Unified Model of AGN (Urry, Padovani 1995)

- many observationally diverse extragalactic objects
  - e.g. small/large radio flux
  - spectral lines large/small or broad/narrow
  - variability in time yes/no etc. pp. …
- explained as one type of object, appearing differently due to orientation to observer
- Type 1 *vs* Type 2 = **black hole** in center is obscured *vs* unobscured by a **dusty torus**
- **sub-classification** remains quite historical
  → not necessarily optimal for neutrino search!
- **neutrino emission** via several processes

Classification of AGN as a problem

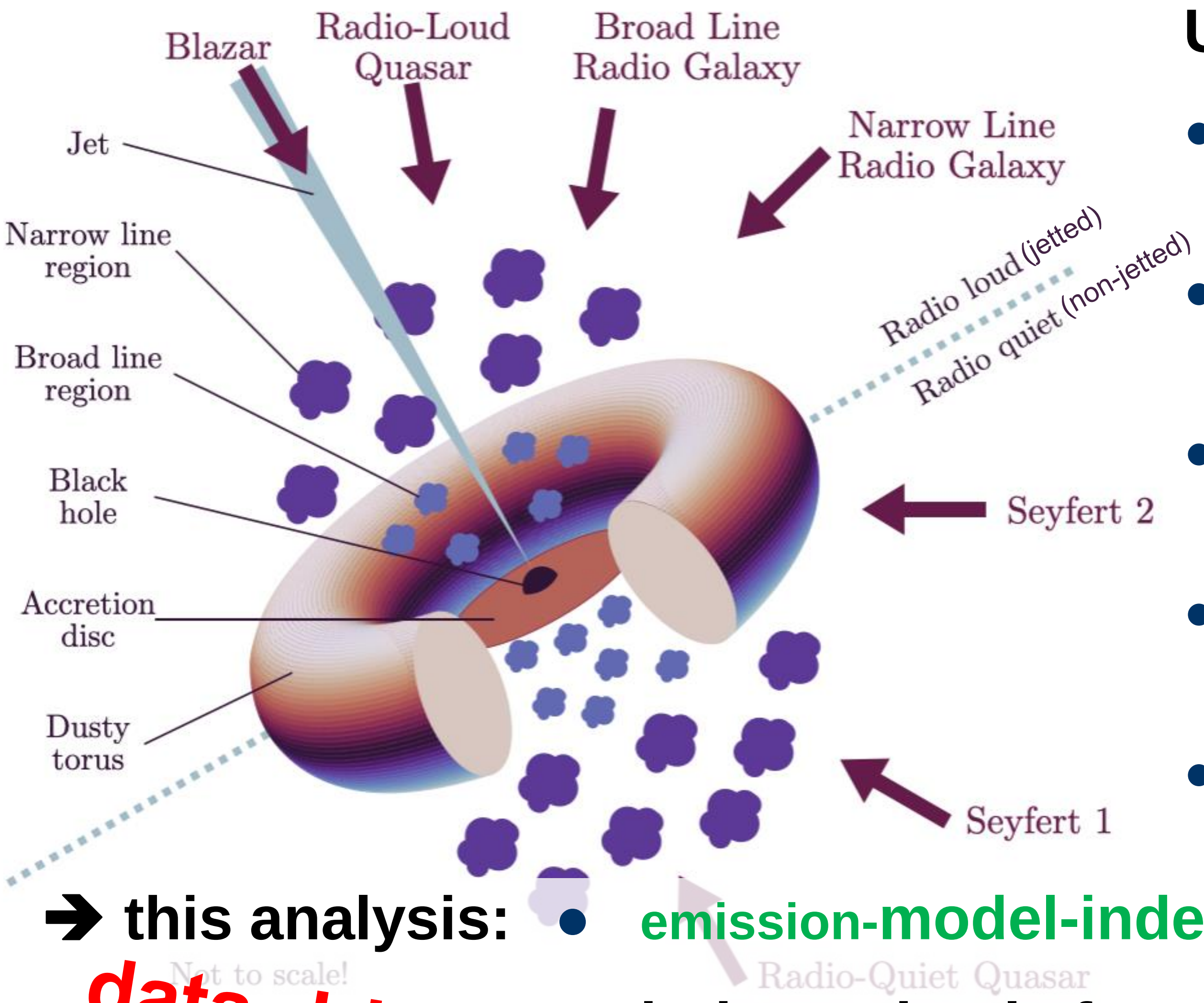


Unified Model of AGN (Urry, Padovani 1995)

- many observationally diverse extragalactic objects
- explained as one type of object, appearing differently due to orientation to observer
- Type 1 *vs* Type 2 = **black hole** in center is obscured *vs* unobscured by a **dusty torus**
- **sub-classification** remains quite historical
  → not necessarily optimal for neutrino search!
- **neutrino emission** via several processes

➔ **this analysis:**
- **emission-model-independent search**
- **independently from existing AGN classification scheme**

Classification of AGN as a problem



Unified Model of AGN (Urry, Padovani 1995)

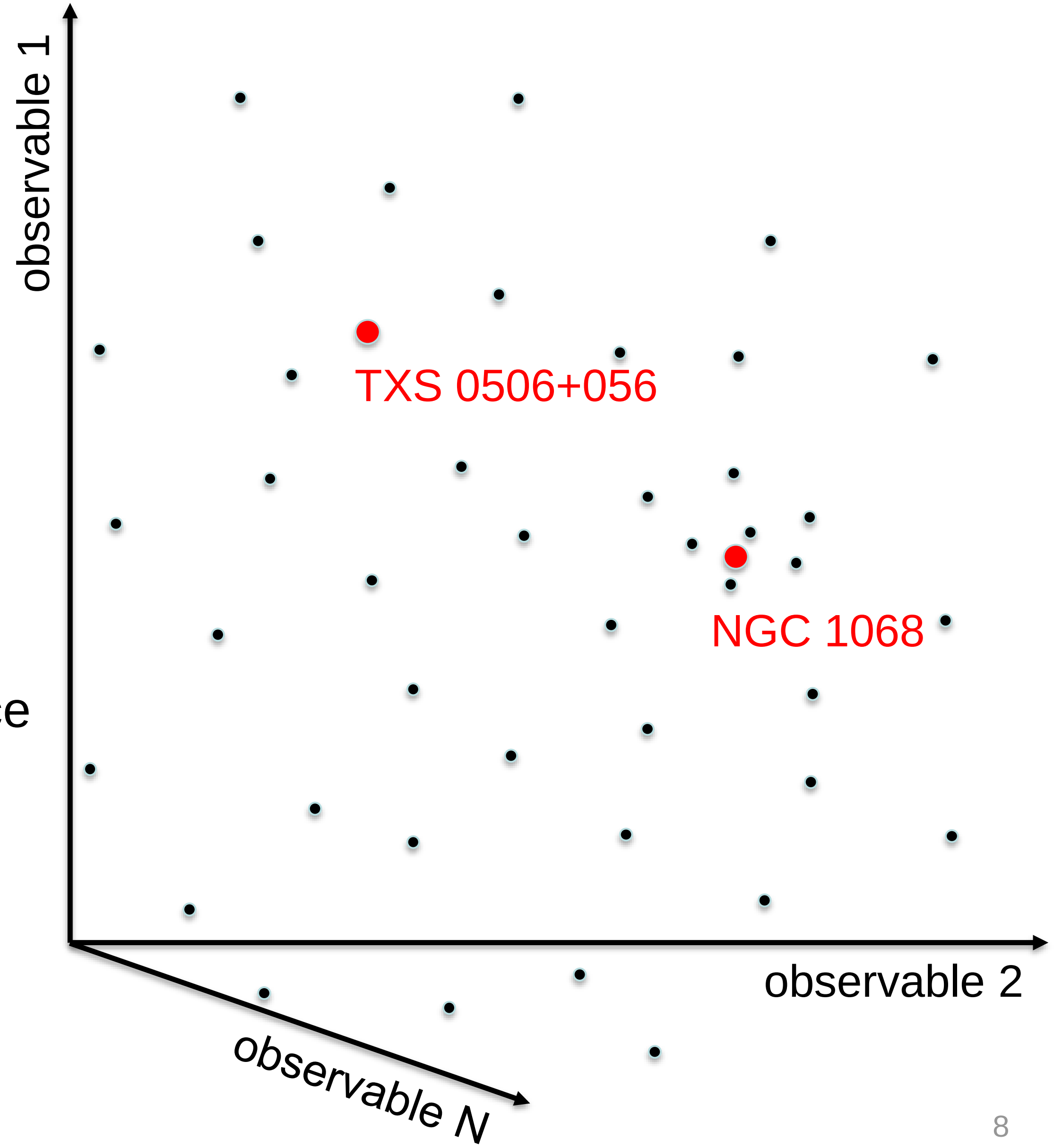
- many observationally diverse extragalactic objects
- explained as one type of object, appearing differently due to orientation to observer
- Type 1 *vs* Type 2 = **black hole** in center is obscured *vs* unobscured by a **dusty torus**
- **sub-classification** remains quite historical
  → not necessarily optimal for neutrino search!
- **neutrino emission** via several processes

➔ this analysis:
- **emission-model-independent search**
- **independently from existing AGN classification scheme**

data-driven!

Phase space of AGN observables

- **many different observable quantities** from AGN:
  - flux in some waveband: radio, x-ray, γ-ray etc.
  - strength, broadness, shape etc.
    of spectral line X, Y, Z etc.
  - polarization etc.
- **each observable is a continuum** (axis in space)
  → many observables span a high-dimensional space
- **populate space** with many observed AGN
  → probably not homogenously distributed



Phase space of AGN observables

distance in phase space

↕

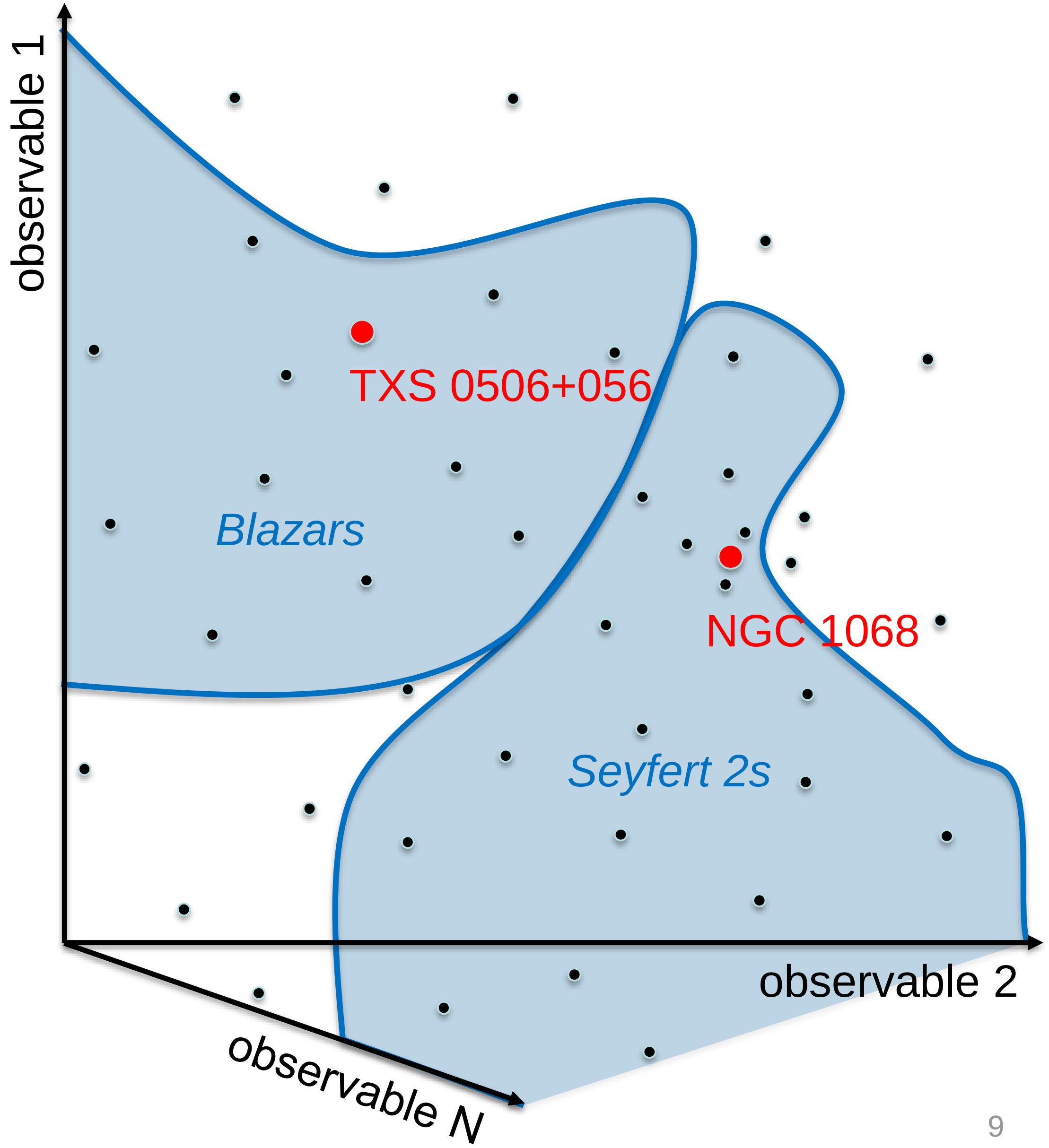
measure for similarity of AGN

clusterings / distinguishable groups

↕

common features,

indication for sub-classes?



Phase space of AGN observables

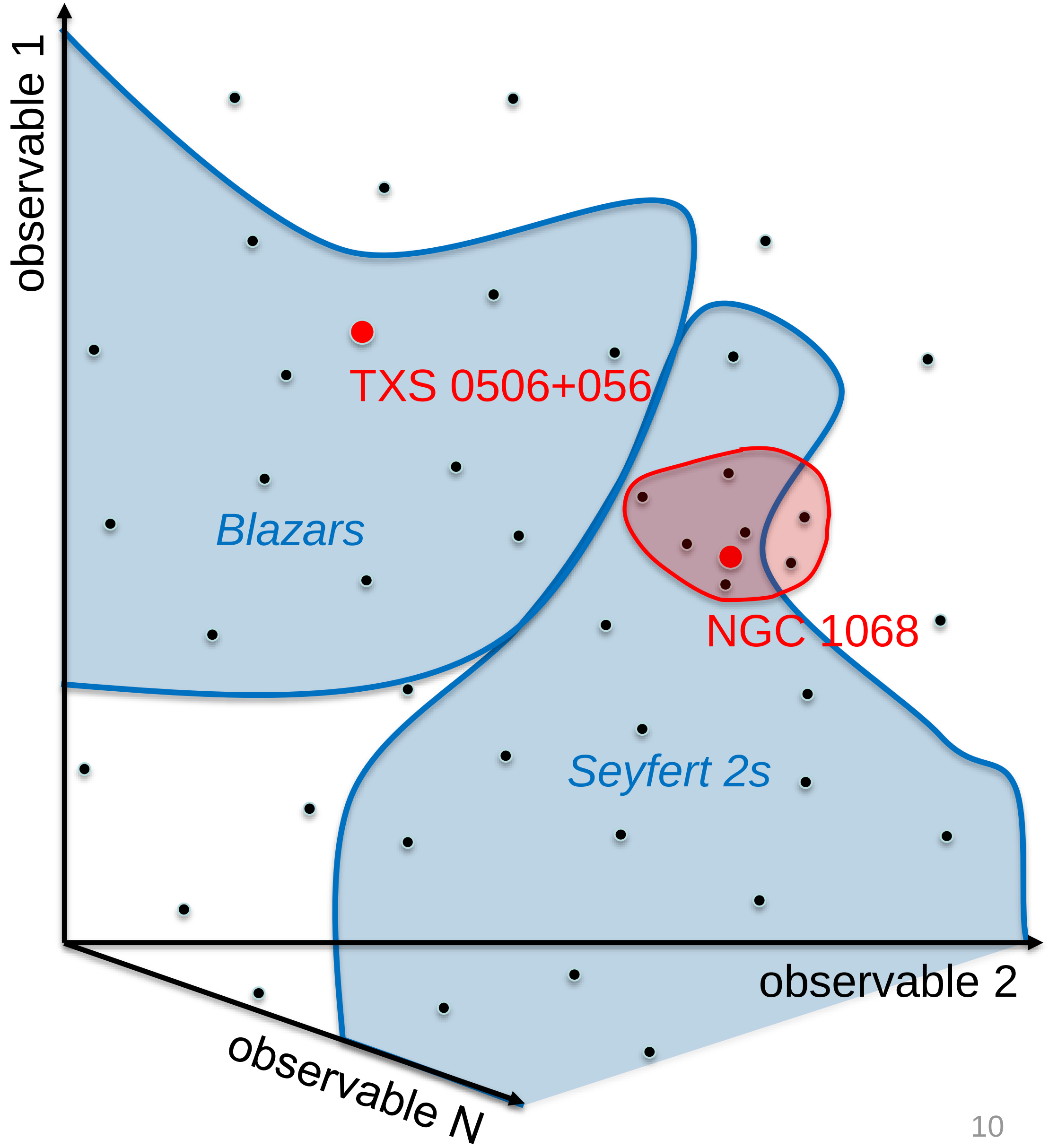
- use **interesting clusters of AGN** as source list for stacking search
- interesting clusters?

  e.g. inclusion of **existing candidate** in cluster

  → "which smallest cluster includes NGC 1068?"

→ stack sources that are **intrinsically similar to existing candidates**

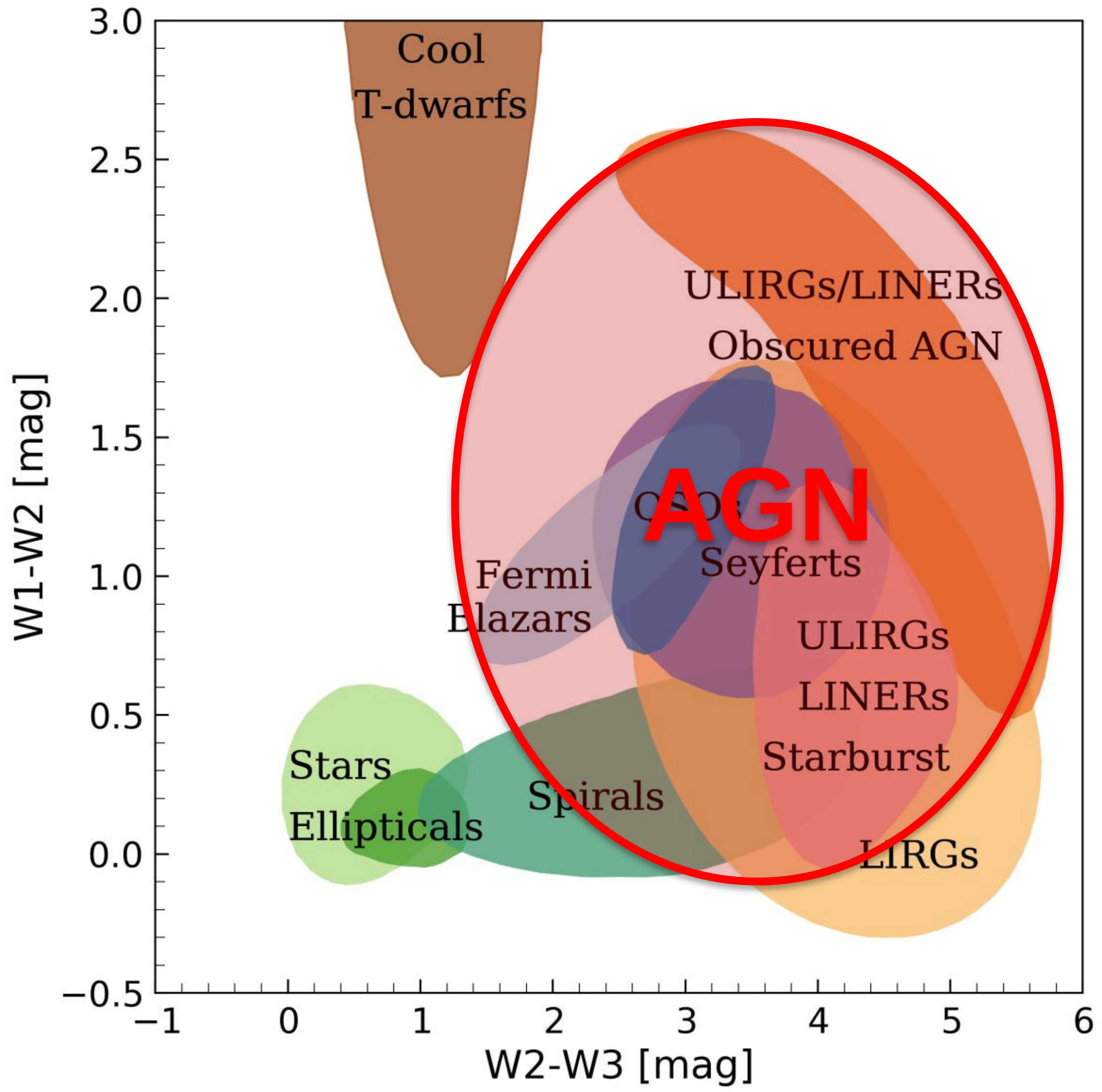
(e.g. TXS 0506+056 & NGC 1068)



IR color-color diagrams

AllWISE: magnitude in 4 infrared bands (W1, …, W4)

- differences of bands („colors“) = slope of spectrum
- color-color diagrams: shown to have good discrimination power between classes of objects
  Nikutta et al. https://doi.org/10.1093/mnras/stu1087
- e.g. selection of AGN possible with cuts in this 2D space

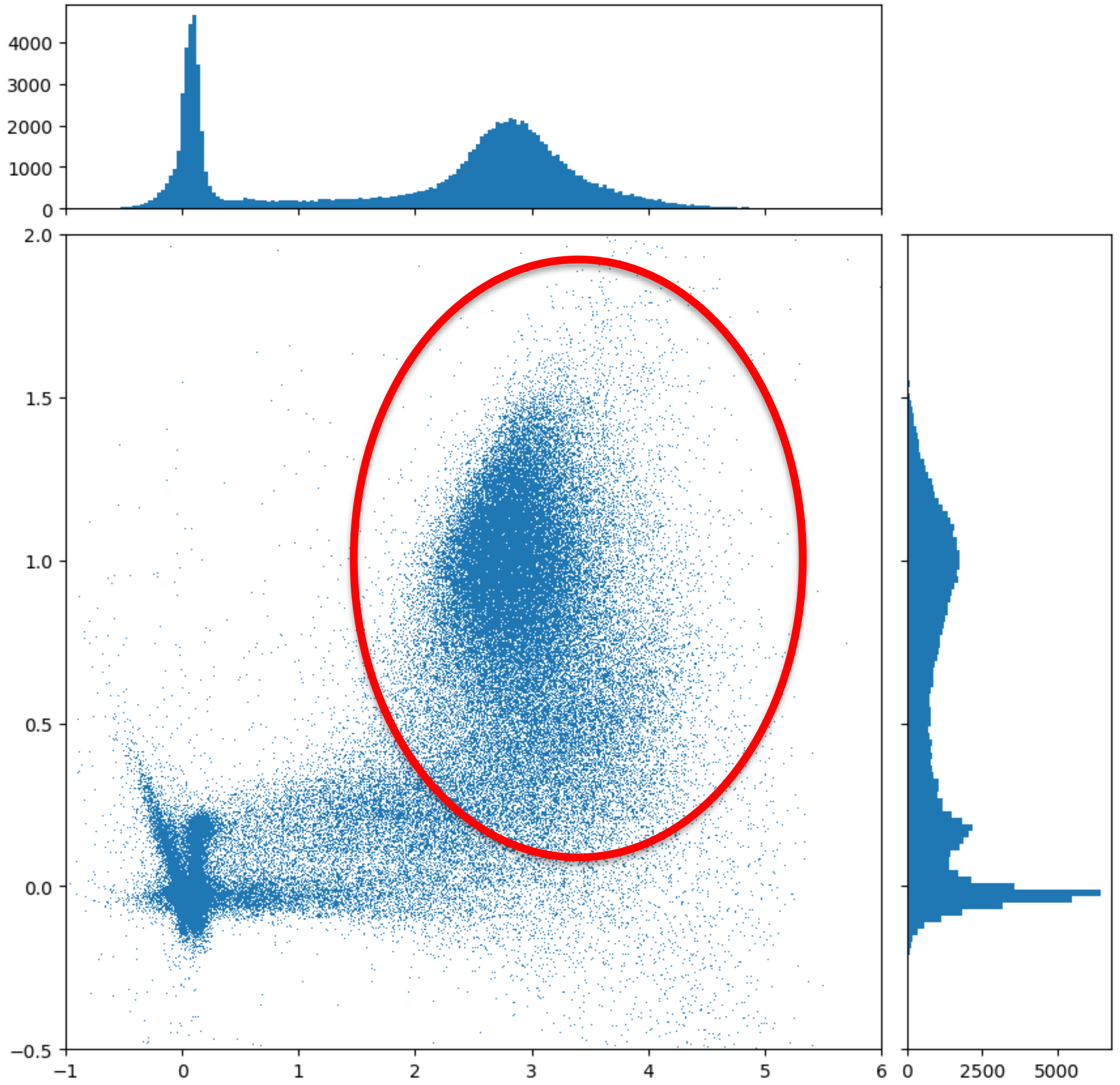


IR color-color diagrams

AllWISE: magnitude in 4 infrared bands (W1, …, W4)

- differences of bands („colors“) = slope of spectrum
- color-color diagrams: shown to have good discrimination power between classes of objects

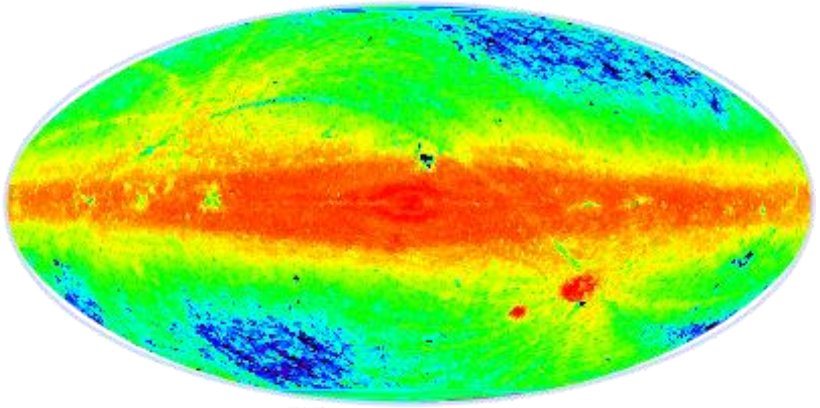
  Nikutta et al. https://doi.org/10.1093/mnras/stu1087
- e.g. selection of AGN possible with cuts in this 2D space

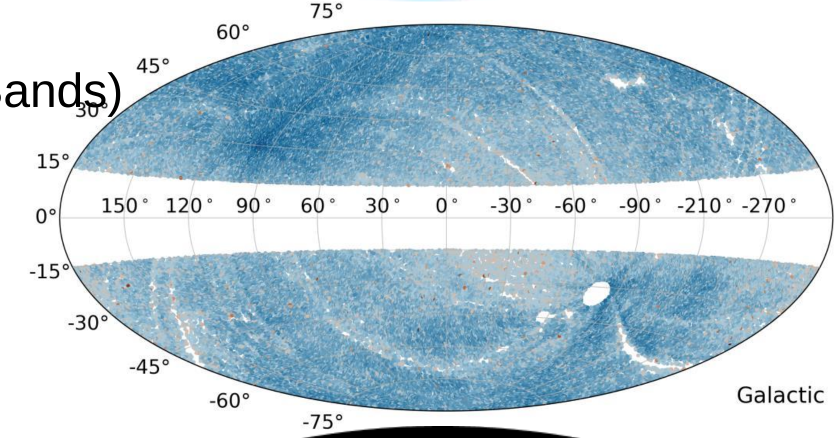


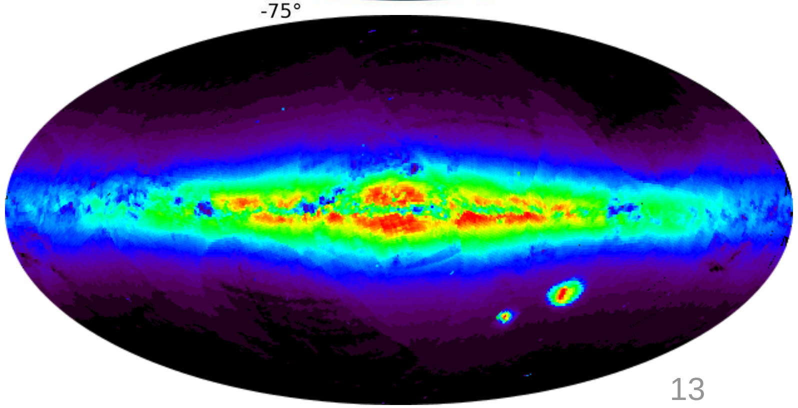
Dataset

add more observables:

- infrared: **AllWISE** (*Wide-field Infrared Survey Explorer*, 2010, 4 bands at a few μm)
- x-ray: **2RXS** (*ROSAT All-sky Survey*, 1990, 0.1 – 2.4 keV)
  + XMMSL2 (*XMM-Newton Slew 2 Survey*, 2017, 0.2 – 12 keV)
- optical: **Gaia** + **SDSS** (*Sloan Digital Sky Survey*, mean flux + u g r i z Bands)



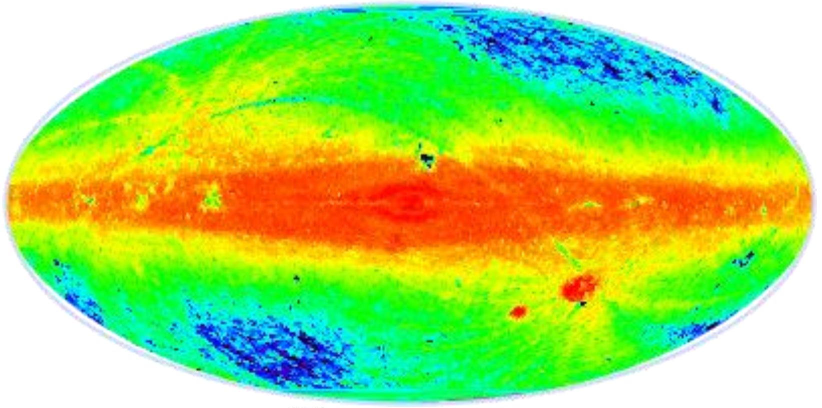


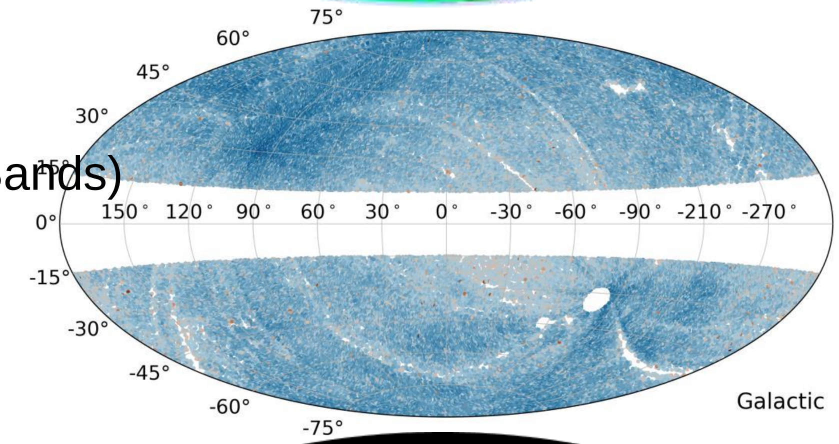


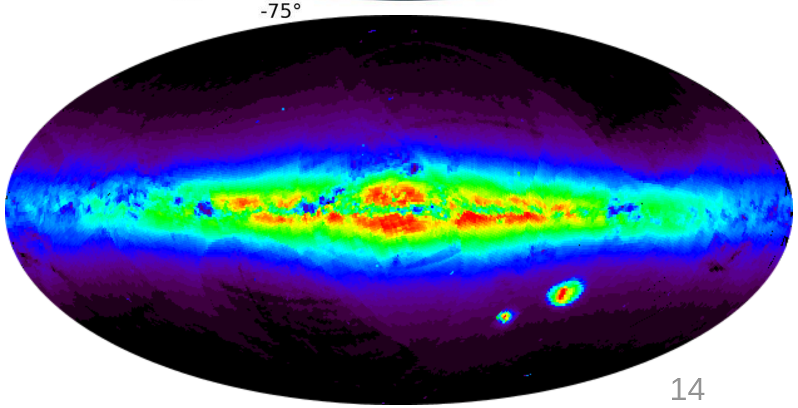
Dataset

add more observables:

- infrared: **AllWISE** (*Wide-field Infrared Survey Explorer*, 2010,
  https://doi.org/10.1088/0004-6256/140/6/1868 4 bands at a few μm)
  → all-sky, millions of AGN candidates
- x-ray: **2RXS** (*ROSAT All-sky Survey*, 1990, 0.1 – 2.4 keV)
  + XMMSL2 (*XMM-Newton Slew 2 Survey*, 2017, 0.2 – 12 keV)
- optical: **Gaia** + **SDSS** (*Sloan Digital Sky Survey*, mean flux + u g r i z Bands)



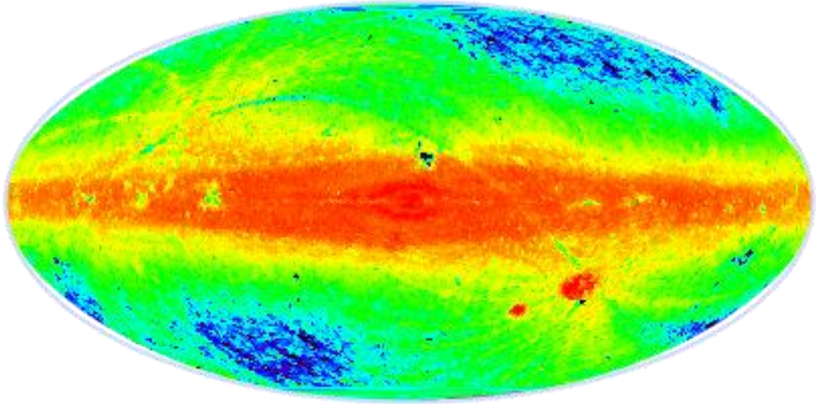


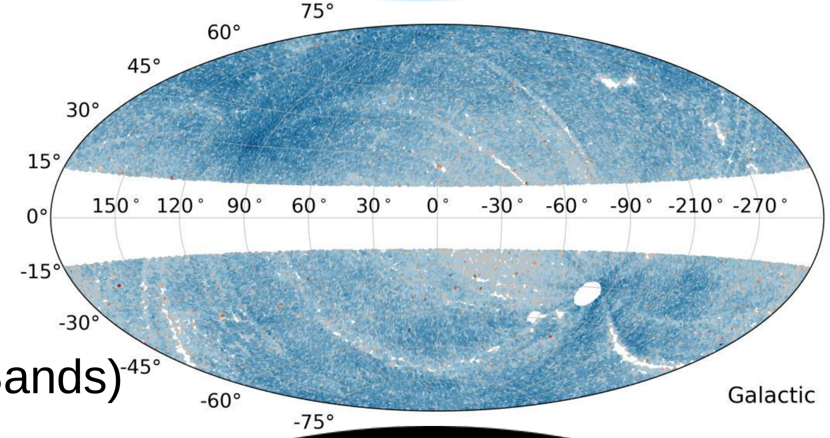


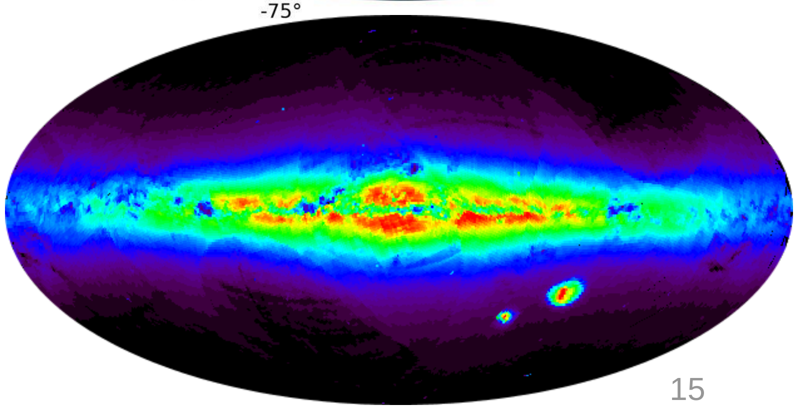
Dataset

add more observables:

- infrared: **AllWISE** (*Wide-field Infrared Survey Explorer*, 2010, https://doi.org/10.1088/0004-6256/140/6/1868 4 bands at a few µm)
  → all-sky, millions of AGN candidates
- x-ray: **2RXS** (*ROSAT All-sky Survey*, 1990, 0.1 – 2.4 keV)
  → all-sky, ~ 100,000 sources https://www.mpe.mpg.de/ROSAT/2RXS
  correlated catalog of 2RXS and AllWISE by Salvato et al. 2017
  https://doi.org/10.1093/mnras/stx2651
  + XMMSL2 (*XMM-Newton Slew 2 Survey*, 2017, 0.2 – 12 keV)
- optical: **Gaia** + **SDSS** (*Sloan Digital Sky Survey*, mean flux + u g r i z Bands)



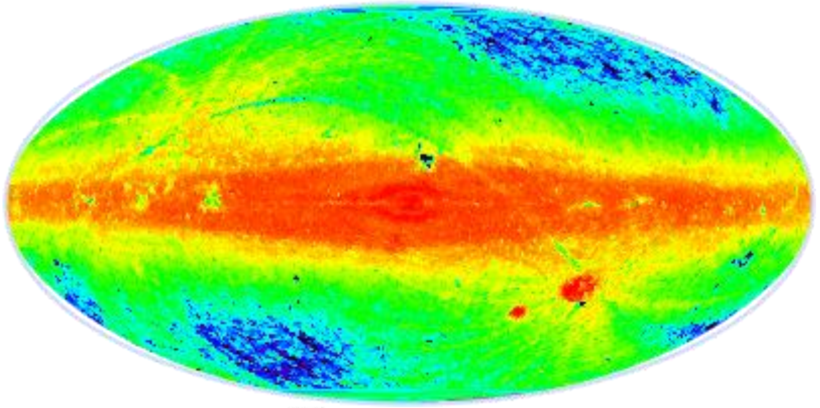


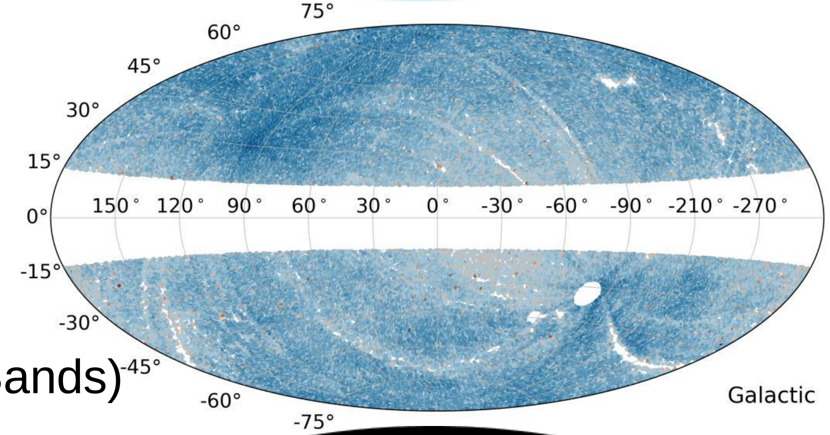


Dataset

add more observables:

- infrared: **AllWISE** (*Wide-field Infrared Survey Explorer*, 2010, https://doi.org/10.1088/0004-6256/140/6/1868 4 bands at a few µm)
  → all-sky, millions of AGN candidates
- x-ray: **2RXS** (*ROSAT All-sky Survey*, 1990, 0.1 – 2.4 keV)
  → all-sky, ~ 100,000 sources https://www.mpe.mpg.de/ROSAT/2RXS
  correlated catalog of 2RXS and AllWISE by Salvato et al. 2017
  https://doi.org/10.1093/mnras/stx2651
  + XMMSL2 (*XMM-Newton Slew 2 Survey*, 2017, 0.2 – 12 keV)
- optical: **Gaia** + **SDSS** (*Sloan Digital Sky Survey*, mean flux + u g r i z Bands)

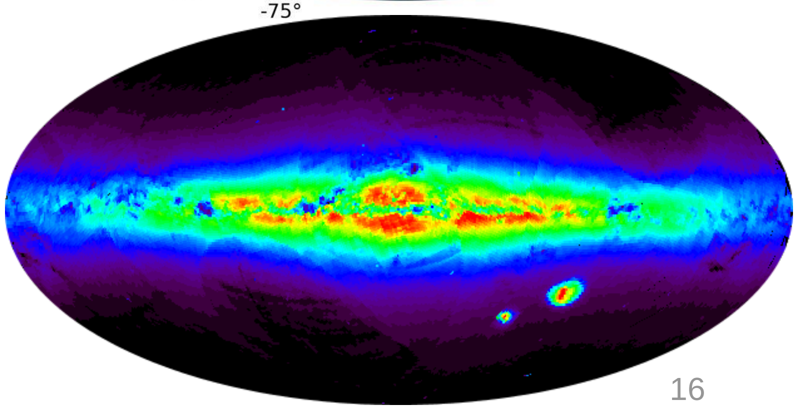




→ photometric data: 18 observables

- 4 infrared bands
- x-ray flux
- optical mean flux + 5 optical bands
- + 7 differences of fluxes (e.g. W1-W2, u-g)

combined catalogs from Mechbal et al. (pre-print)
arXiv:2303.18076



18-dim. AGN data

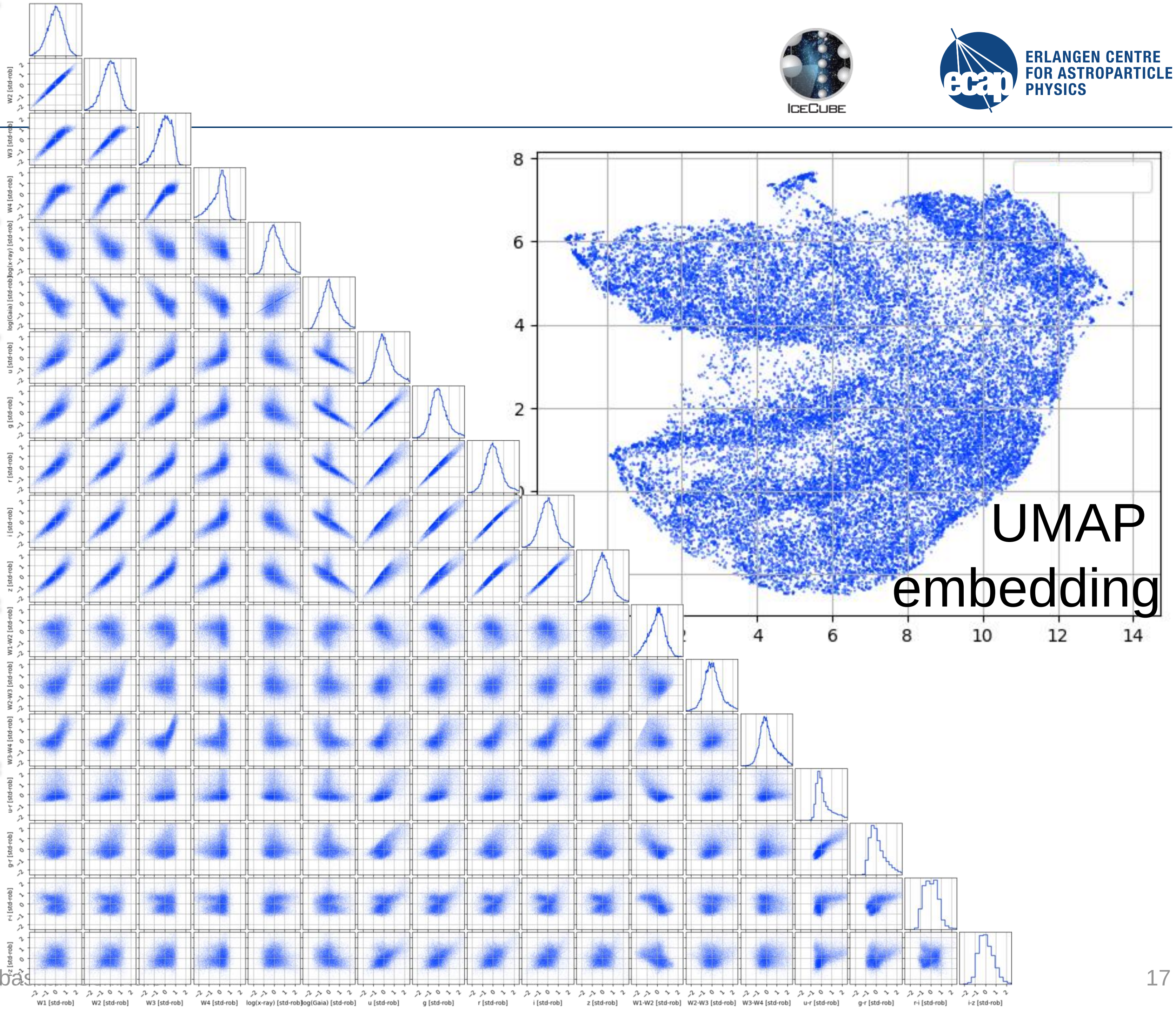
infrared (W1, …, W4)

log(x-ray)
log(optical mean)

optical bands (u, g, r, i, z)

infrared differences

optical differences

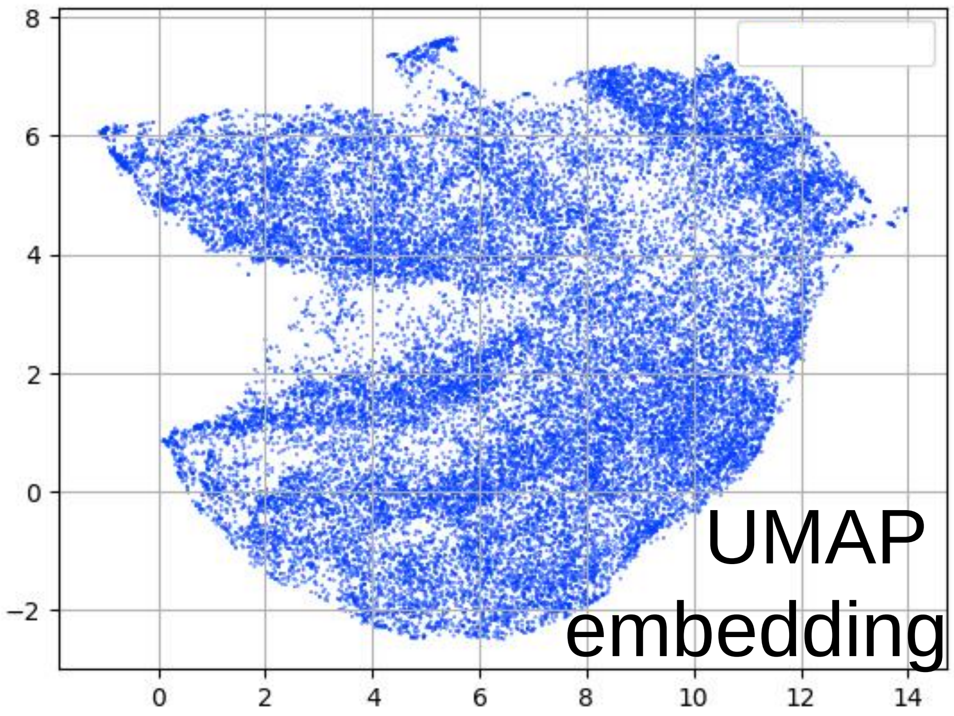


18-dim. AGN data

- rather homogeneous, connected distribution in most observables
  → no to-the-eye obvious distinguishable classes

- but high-dimensional = cannot see all hidden details even in corner plot
  illustrative example:
  sphere inside a spherical shell
  → 2D projections cannot depict problem entirely

- UMAP embedding:
  - difficult to interpret, however:
  - some structure: lower & upper part, upper-right part? → investigate!

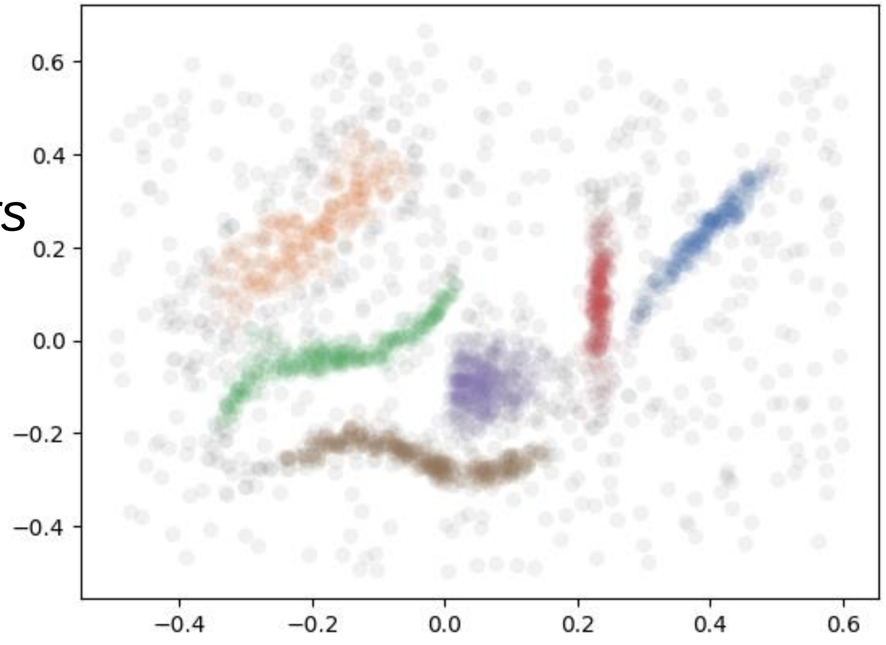


Cluster algorithm

HDBSCAN: state-of-the-art algorithm for searching for clusters in arbitrary data

(*Hierarchical Density-Based Spatial Clustering of Applications with Noise*
Campello, Moulavi, Sander https://doi.org/10.1007/978-3-642-37456-2_14)

- high dimensionality of data possible
- explorative: no required knowledge of number of clusters, arbitrary cluster shape
  → *but tends to produce some small additional clusters that seem like noise*
- fast (few seconds for 30k points in 18 dimensions)
- two main parameters to tune
  → *but not very easy to interpret these hyperparameters*
- has notion of unclustered data
  → *but often just leaves points around a central cluster unclustered, which is not that helpful*



Clustering applied

(one particular choice of hyperparameters)

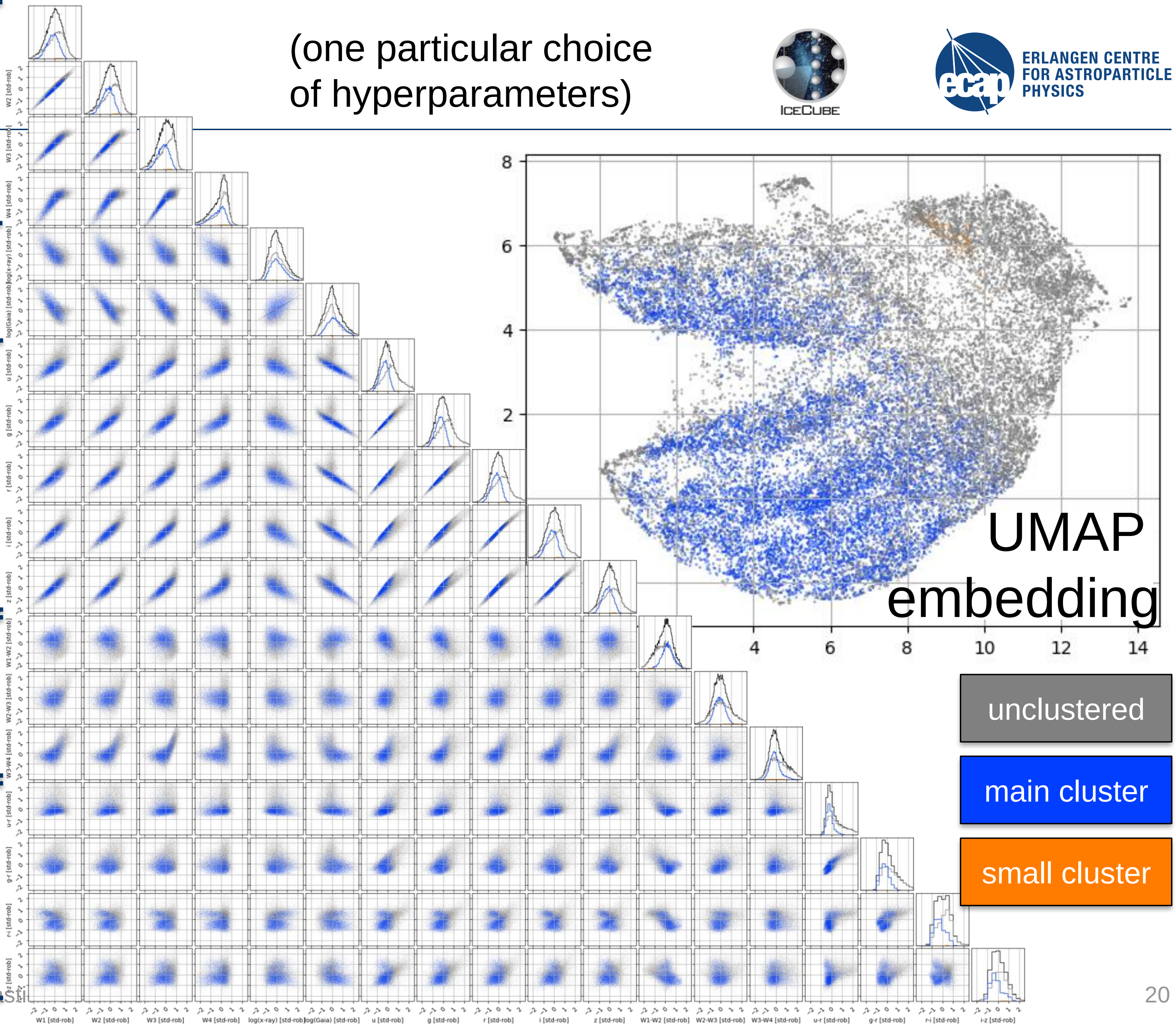
infrared (W1, …, W4)

log(x-ray)
log(optical mean)

optical bands (u, g, r, i, z)

infrared differences

optical differences



unclustered

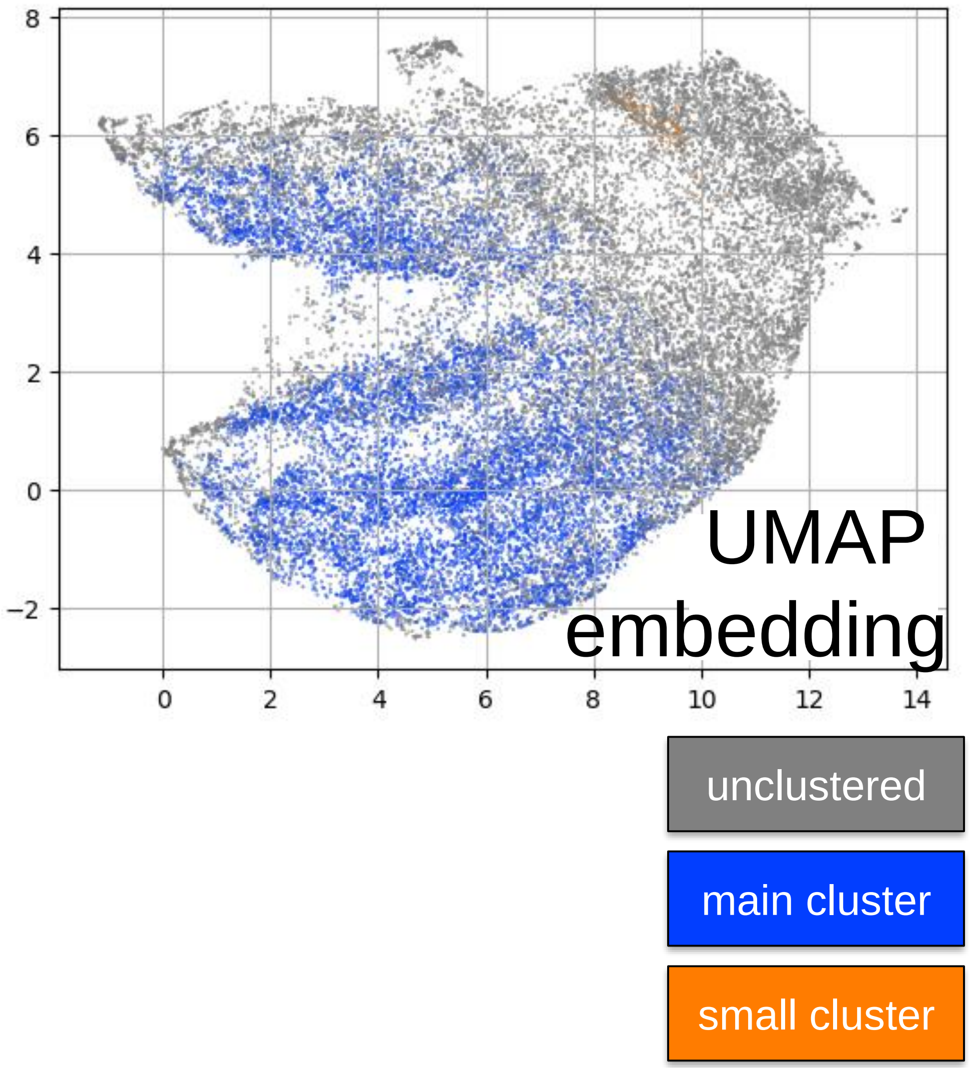
main cluster

small cluster

Clustering applied

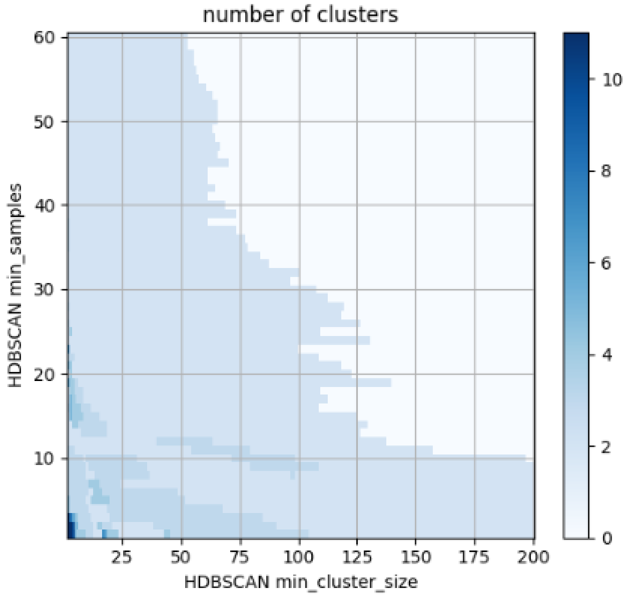
(one particular choice of hyperparameters)

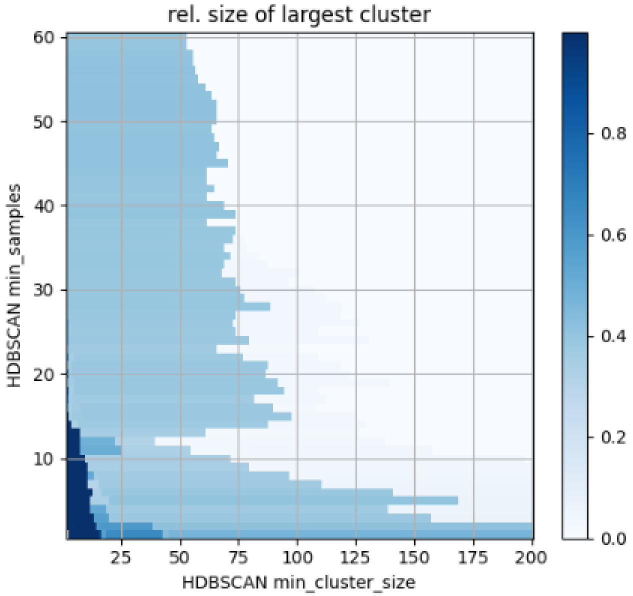
- one continuous, regularly shaped main cluster in most dense area
- many points around it remain unclustered (few micro clusters)
- hyperparameter choice does not change much
- UMAP embedding: clustering does not follow the visible structure…

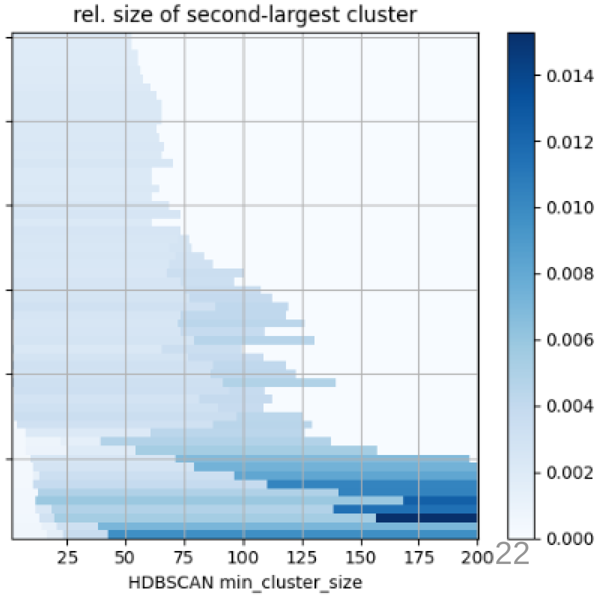


Hyperparameter scan

- hyperparameter choice for cluster algorithm is not obvious
  → scan hyperparameter space (i.e. run HDBSCAN for all combinations)
- look at summary quantities
  - number of clusters found
  - relative size of largest & second-largest cluster (fraction of points in this cluster)
- what choice is sensible?
  → introduce artificial clusters to test hyperparameter choice
  - introduce obvious clusters into the data
  - only hyperparameters that can identify those should be used







What to make of this?

naïve and hopeful idea does not really work:

- not that much structure in the phase space
- algorithm does not seem to find interesting non-obvious clusters

→ fault of HDBSCAN? use different cluster algorithm?

other Ansatz:

- compile list of interesting sources and mark them in the phase space
- (should not be a problem with blindness, as no neutrino data yet involved)

→ can we see something around the sources by eye?

fall-back option:

- use all sources in a pre-defined radius around very interesting sources (like NGC 1068)
- maybe radius such that ~ 100 sources are contained

Summary

Summary

- look for source list for stacking search based on similarity to known candidate sources
- high-dimensional dataset: about 30k AGN in 18 observables
- search for structures/clusters using HDBSCAN and UMAP
- not successfull currently → possibly pursue different Ansatz

Different idea

- instead of many different observables, use wealth of data in **optical only**
- optical spectra: treat each bin as one dimension (~ 1000 dimensions)
- apply UMAP (or other dimensionality reduction), then search for clusters again

Thank you for your attention

Questions?

ERLANGEN CENTRE
FOR ASTROPARTICLE
PHYSICS

SPONSORED BY THE
Federal Ministry
of Education
and Research

ICECUBE
SOUTH POLE NEUTRINO OBSERVATORY

ERLANGEN CENTRE
FOR ASTROPARTICLE
PHYSICS

FAU
Friedrich-Alexander-Universität
Naturwissenschaftliche Fakultät

Backup

Analysis plan

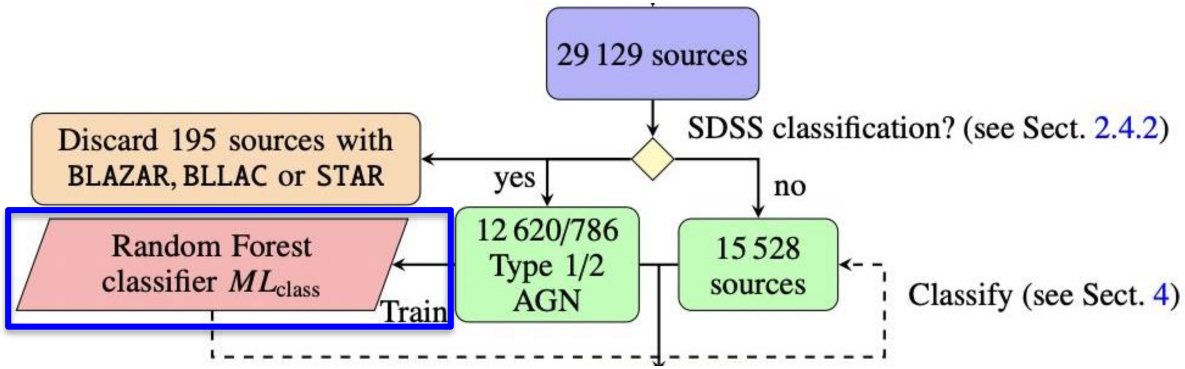
1. **build extensive catalog** of AGN with many different observables
   a) decide on observables to use (flux in certain waveband, spectral lines etc.)
   b) combine astronomical catalogs
   c) populate phase space
2. **cluster search** in observable space to select AGN closest to candidates
   a) run cluster search algorithm
   b) select interesting clusters based on location of candidate sources
3. perform **stacking analysis** with AGN within selected clusters

Testing with a supervised-learning classifier

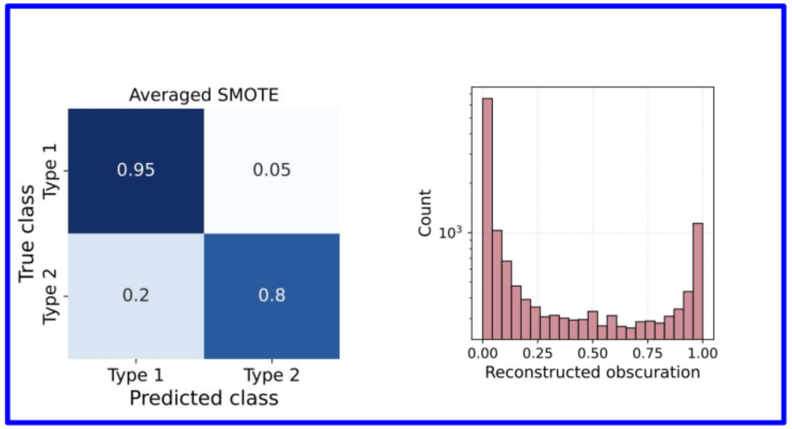
- can classify AGN as…
  - Type 1: unobscured disk, broad lines
  - Type 2: obscured disk, narrow lines
- Mechbal et al. pre-print arXiv:2303.18076:

  classification exists for ~ ½ of AGNs in dataset
  (from SDSS catalog by visual inspection)

  → **train machine-learning classifier**
  - photometric data of AGNs as input
  - supervised learning with labels from existing classification
  - apply to other ½ of AGNs



Classification task



→ also possible with **unsupervised clustering?**

- classification problem is binary (Type 1 or 2)
- … so there should be two clusters in the data

Classifier (Mechbal et al.)

infrared (W1, …, W4)

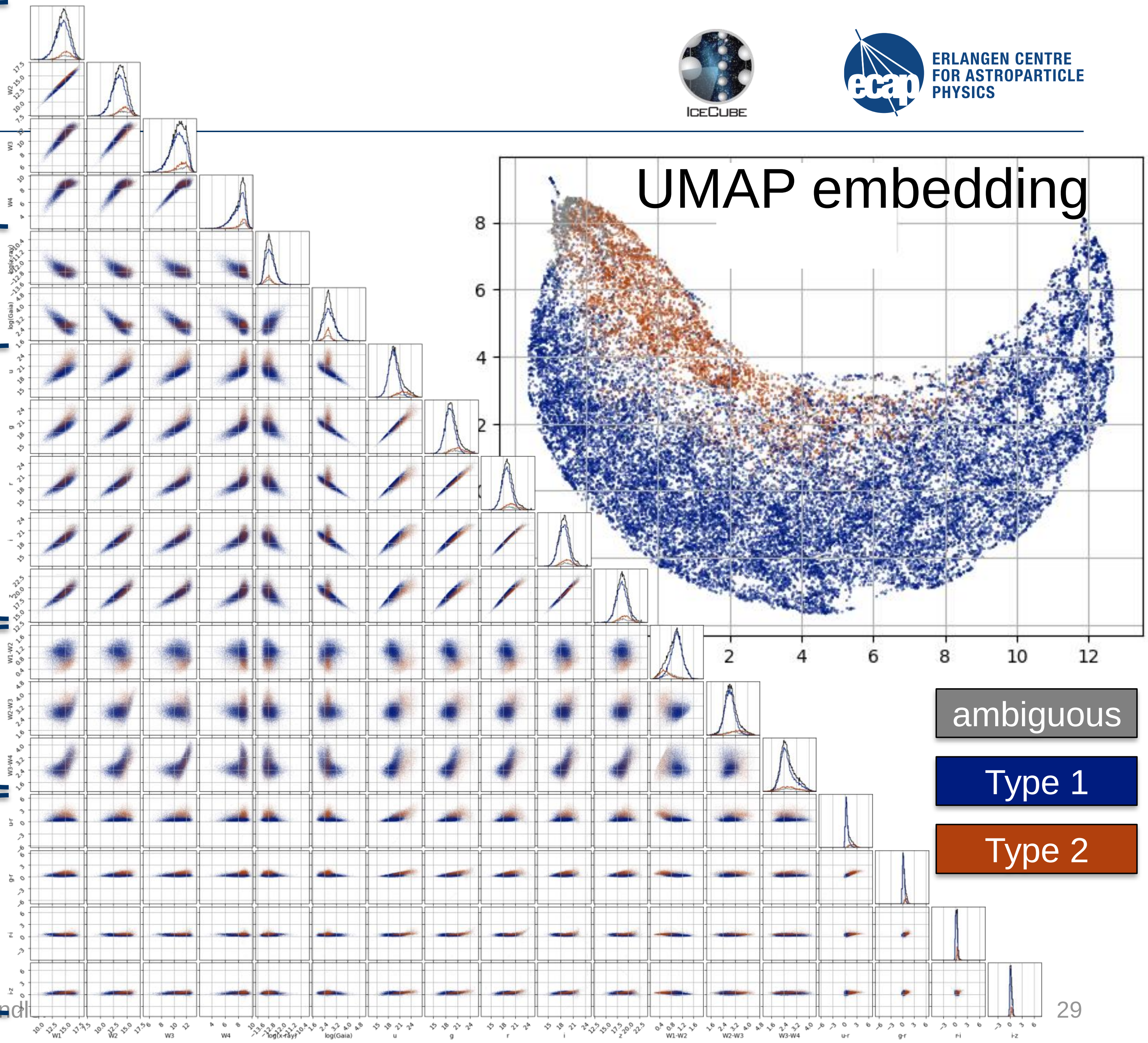
log(x-ray)

log(optical mean)

optical bands (u, g, r, i, z)

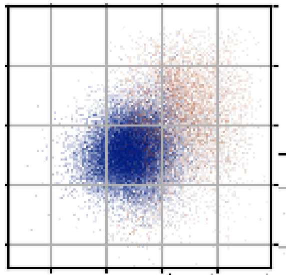
infrared differences

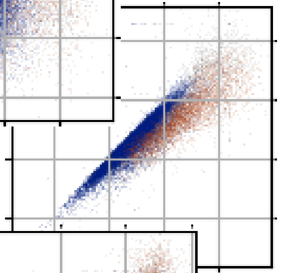
optical differences



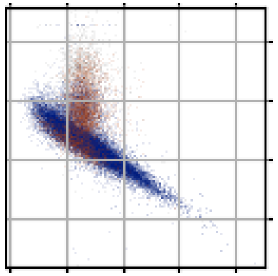
Classifier (Mechbal et al.)

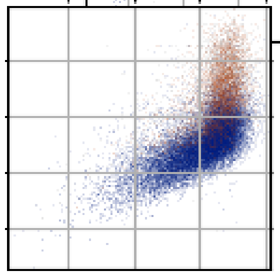
- separation between Type 1 / 2 somewhat similar to clusterin
  - Type 1 more in dense center area
  - Type 2 more at the fringes
  - also the case in UMAP





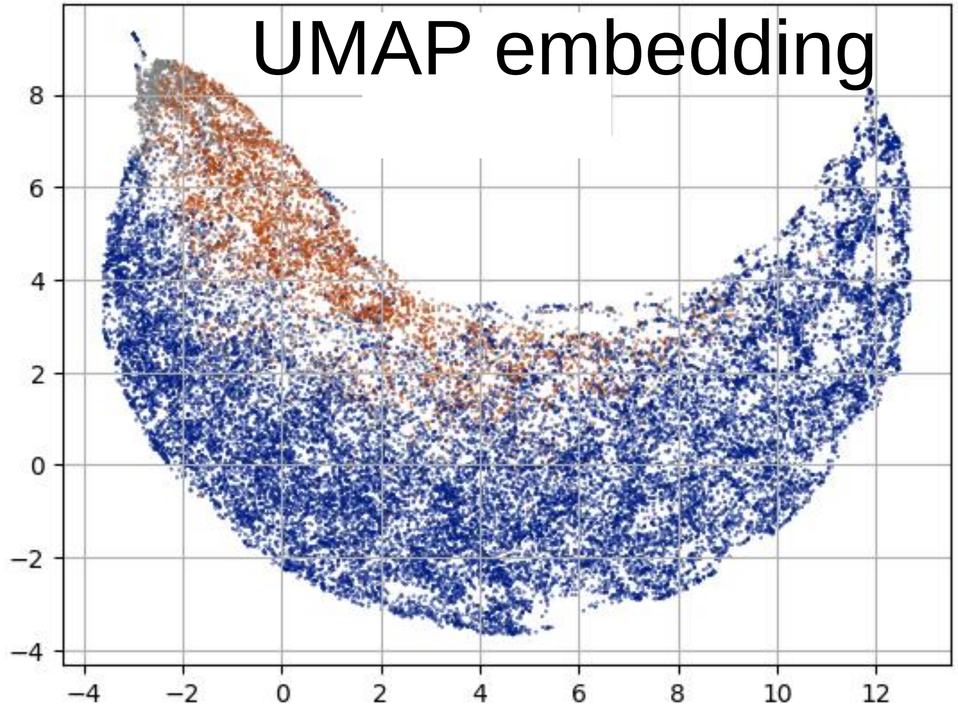
- some smaller features can be mapped to Type 1 / 2





→ for **comparison with clustering**, associate…

- unclustered with Type 2
- main cluster with Type 1



ambiguous

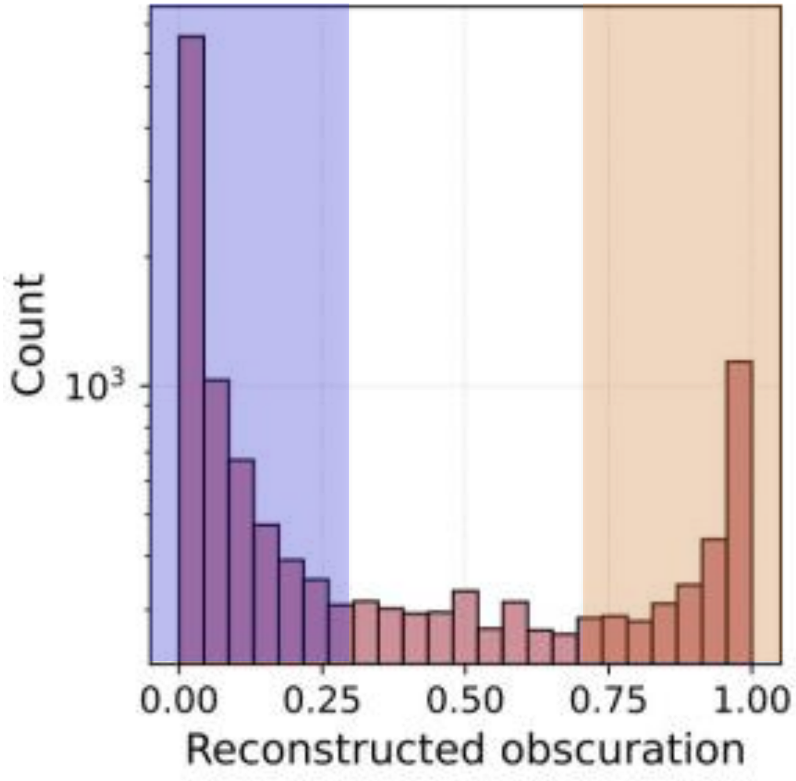
Type 1

Type 2

Classifier (Mechbal et al.)

obscuration score μ

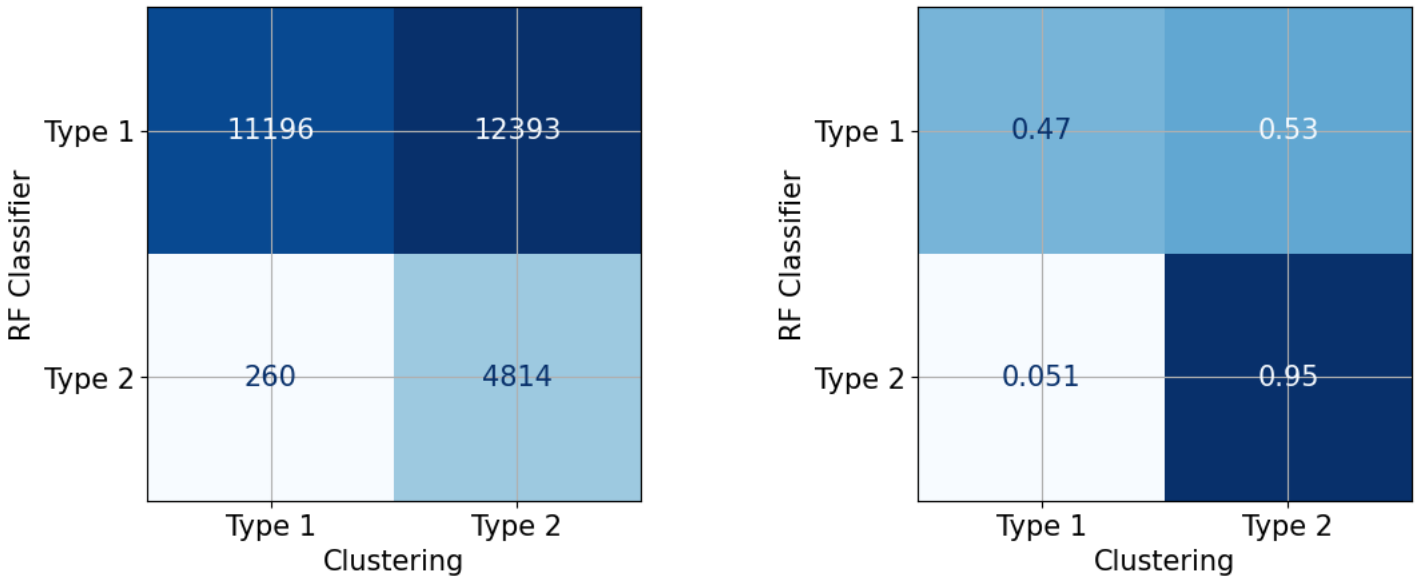
- $\mu < 0.3$ → Type 1
- $\mu > 0.7$ → Type 2
- $0.3 < \mu < 0.7$ → ambiguous type



Performance of clustering

current work in progress:

compare cluster result (prediction) to classifier (as truth):



→ clusterer…

- finds most Type 2 (95 %)
- but is just guessing for Type 1 (47 %)

however:
Type 1 more numerous!