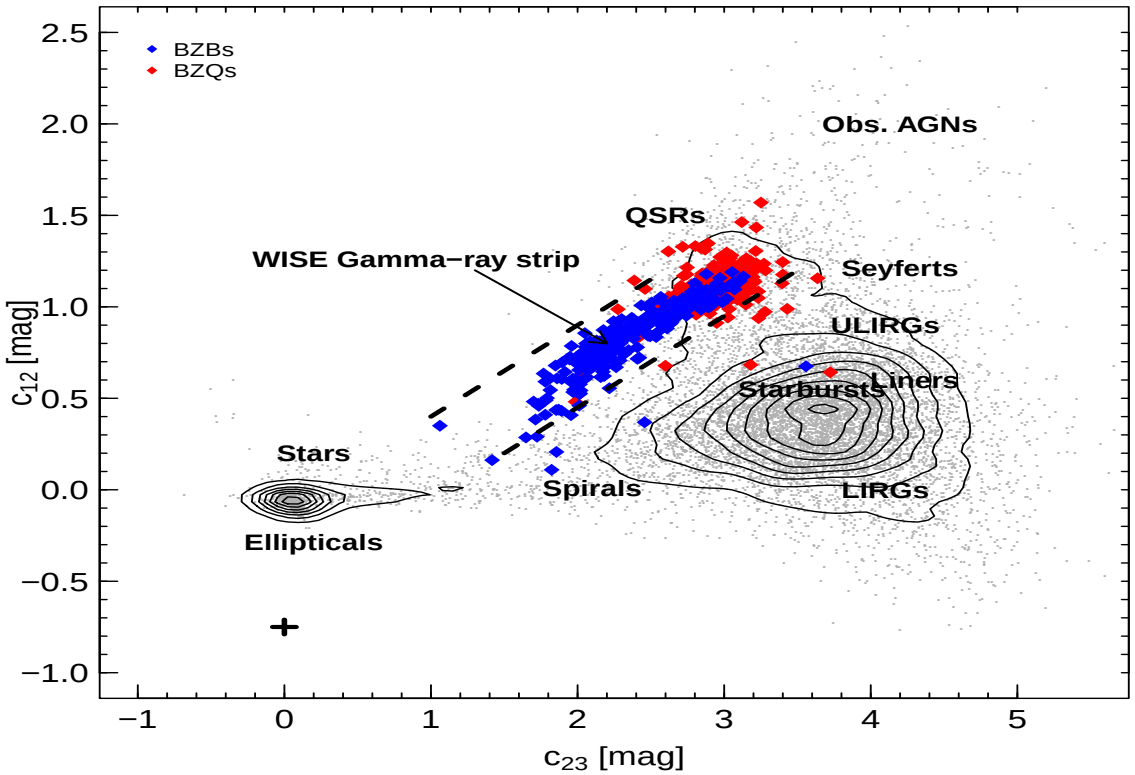
A new method for the extraction of mid-infrared γ-ray emitting candidate blazars

Raffaele D'Abrusco
(in collaboration with **F. Massaro, A. Paggi** and
G. Tosti, H.A. Smith, J. Grindlay and P. Cowperthwaite)

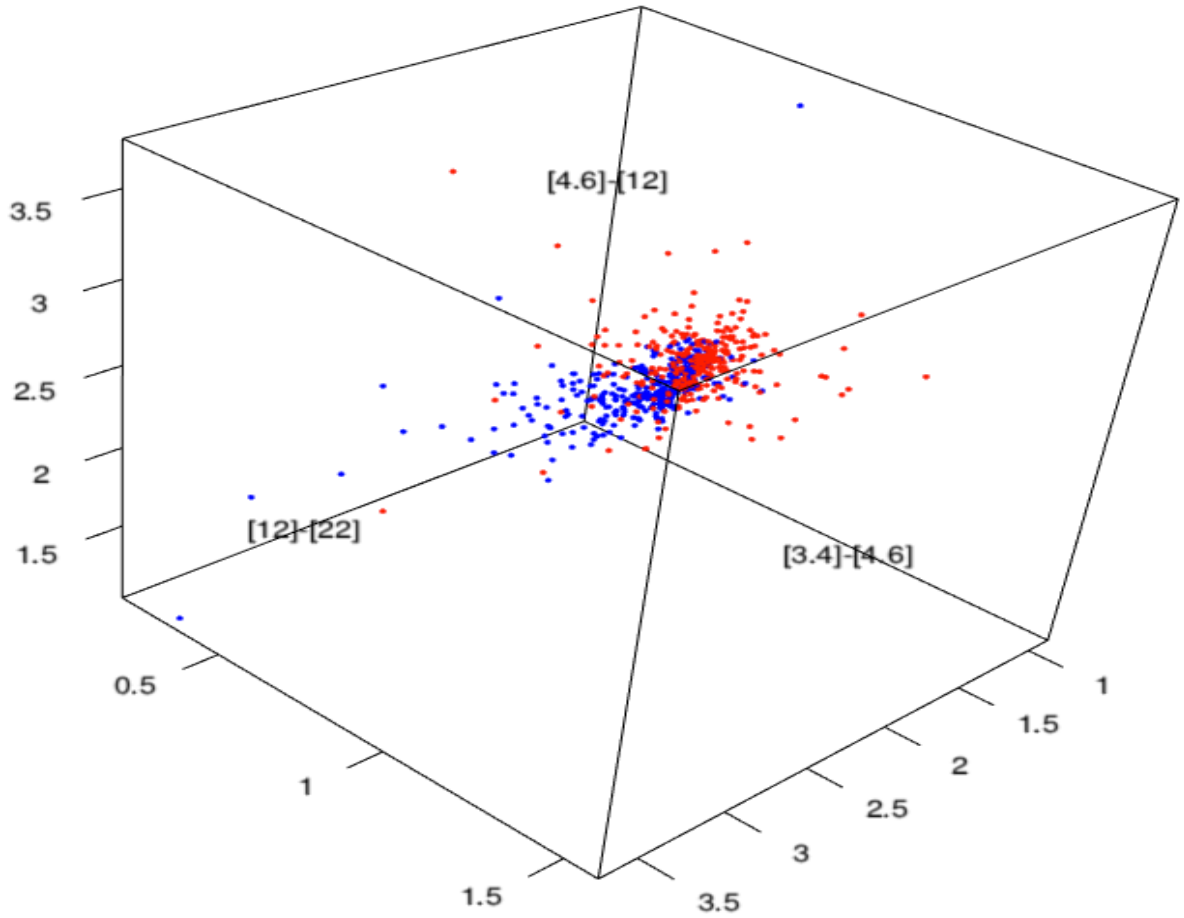
Harvard-Smithsonian Center for Astrophysics

The 3D *locus* of the *WISE* γ-ray detected blazars

As previously shown by Francesco, the *WISE gamma*-ray detected blazars occupy a well-defined and narrow region of the 3-dimensional *WISE* colors space ([3.4]-[4.6], [4.6]-[12], [12]-[22] μm), namely the 3D *WISE* **WGS *locus*** (D'Abrusco et al. 2012).



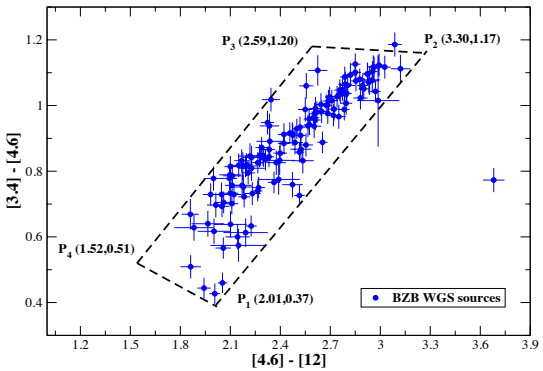
The 3D *locus* of the *WISE* γ-ray detected blazars



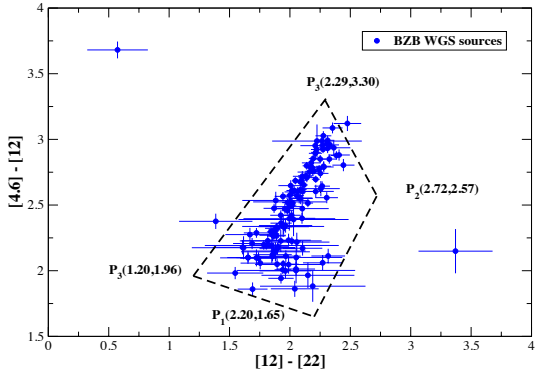
The strip as a selection method

A *model* (parametric description) of the *WISE* blazars strip can be used select IR candidate blazars. In other words, **we assume that the behavior of all γ-ray emitting blazars in *WISE* color is embodied by the behavior of the WGS sample**. We proposed (Massaro et al. 2012) a first model of the *locus* based on three 2D projections of the *locus* in various *WISE* color-color planes, and applied it to unidentified Fermi sources (Massaro et al. 2012b).

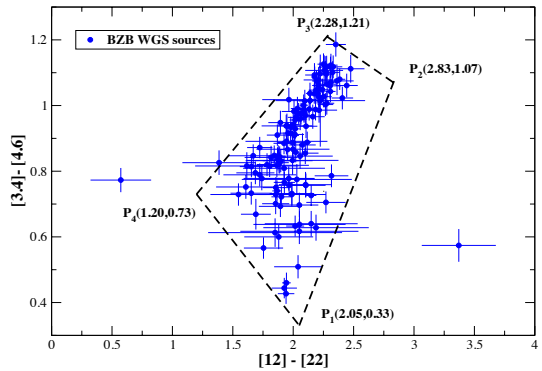
BZB population: [3.4]-[4.6]-[12] projection



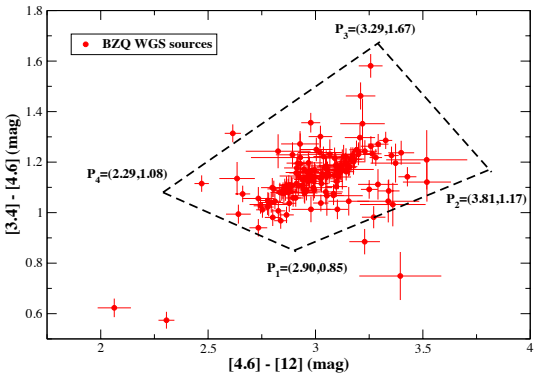
BZB population: [4.6]-[12]-[22] projection



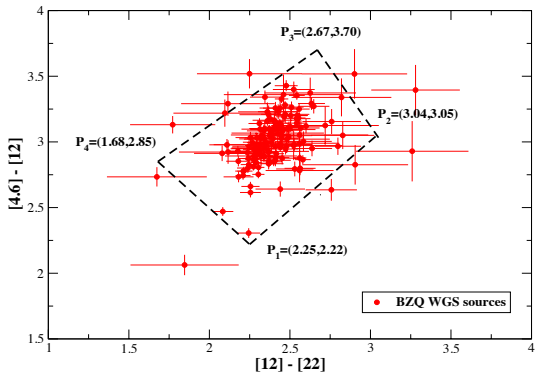
BZB population: [3.4]-[4.6]-[12]-[22] projection



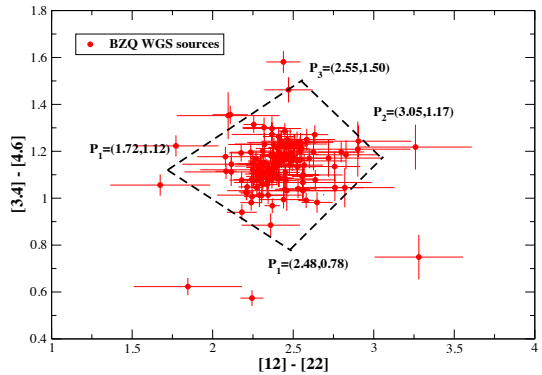
BZQ population: [3.4]-[4.6]-[12] projection



BZQ population: [4.6]-[12]-[22] projection

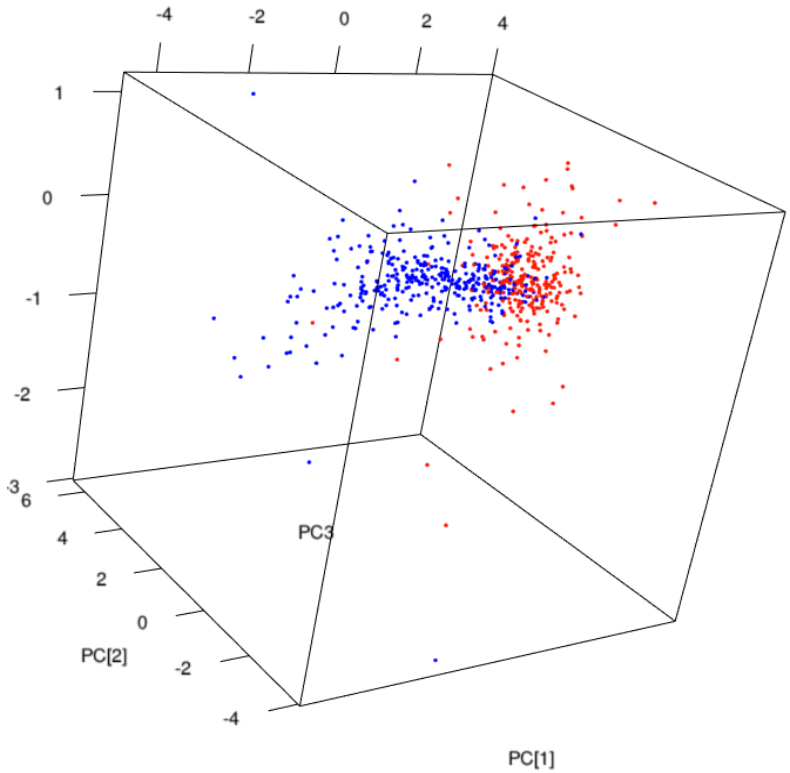


BZQ population: [3.4]-[4.6]-[12]-[22] projection



A new *WISE* Blazars *locus* model

The almost perfectly axisymmetric distribution of WGS sources suggests a simpler way to model the WGS *locus* in a different space, i.e. the space associated to an orthogonal transformation of the of WGS color space distribution that transform into linearly uncorrelated coordinates where the variance along one axis is maximized: the **the Principal Components Analysis.**

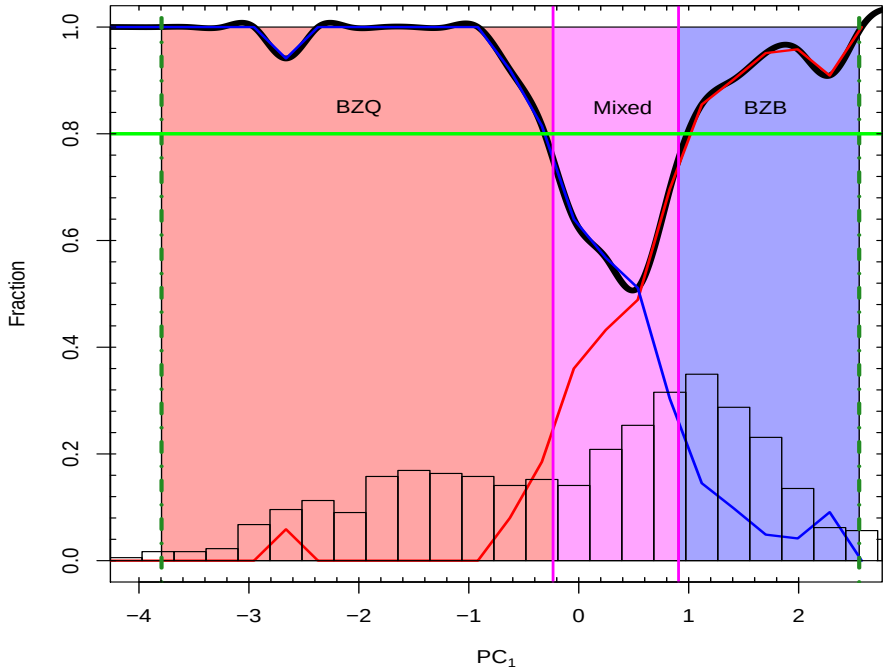


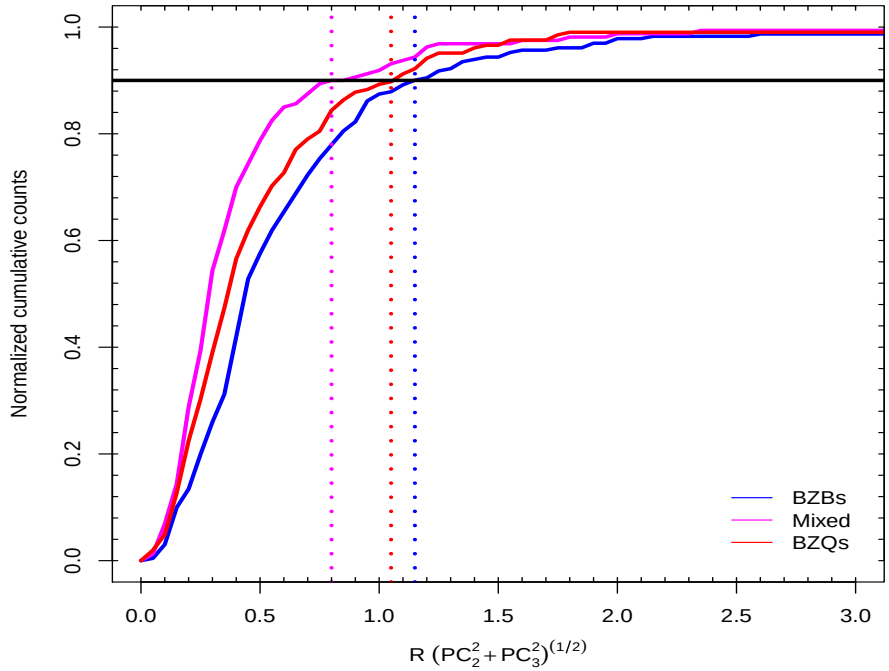
Why a *locus* model in the PC space is better than the old model?

- This approach provides a **much simpler model of the WGS** that takes advantage of the symmetry of the *locus*;
- This model can be **fine-tuned** by tweaking its parameters in order to improve particular aspects of the selection;
- The slight deviation from axis-symmetry in the BL Lac dominated region of the *locus* are not significant;

The model and its parameters

Regions of the WGS *locus* dominated by BL-Lac or FSRQ are modeled independently. Each section is modeled as a **cylinder with axis along the first PC_1 axis and radius** $R = \sqrt{PC_2^2 + PC_3^2}$. The free parameters of the model are the heights of the cylinders (1x3) and the radii of the cylinders (1x3).





The values of the 6 parameters of the new WGS *locus* model are determined based on the WGS sample. **A "mixed" region** located between the BL Lac and FSRQs dominates cylinders, **is introduced to highlight the region of the locus where the two families overlap**.

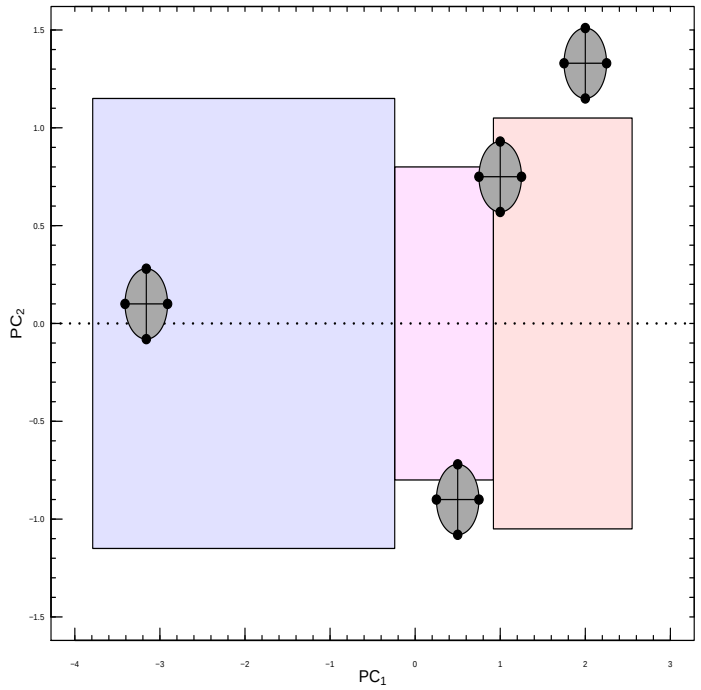
The basic rules

The uncertainties affecting the *WISE* colors are accounted for as **each source is represented by the uncertainty ellipsoid** centered on the point of the PCs space associated to the observed colors and with axes equal to the transformed error bars of the *WISE* color.

A few cases

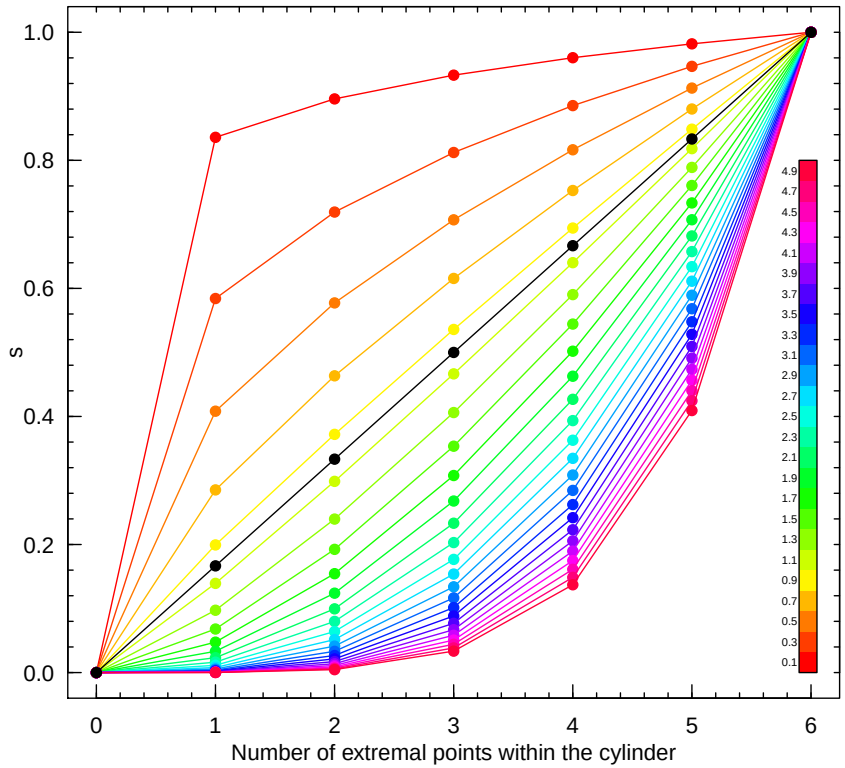
- All (six) extremal points within the cylinder;
- Five extremal points within the cylinder;
- Three extremal points within the cylinder;
- One extremal point within the cylinder;
- ~~source with no extremal points within the cylinder~~.

Sources with uncertainty ellipsoid overlapping two different regions will be classified independently for the two sub-regions.



The score

The position of the uncertainty ellipsoid of each source relatively to each of the cylinders can be translated into a quantitative estimate of its consistency with the *locus* model, **the "score"**. The score is a monotonic function of the number of extremal points of the ellipsoid contained in each cylinder of the model.



Score

$$s = \frac{1}{6^{\phi}} \cdot n^{\phi}$$

where n is the number of extremal points within a given cylinder and **ϕ is the *index* of the score assignment law**.

Weighted score

$$s_w = s \cdot w_{WGS}.$$

where:

$$w_{WGS} = \frac{1}{\max(4\pi\sigma_{PC_1}\sigma_{PC_2}\sigma_{PC_3})}$$

for the WGS (the largest volume of all uncertainty ellipsoids of the WGS sources).

The strip as a *classifier*

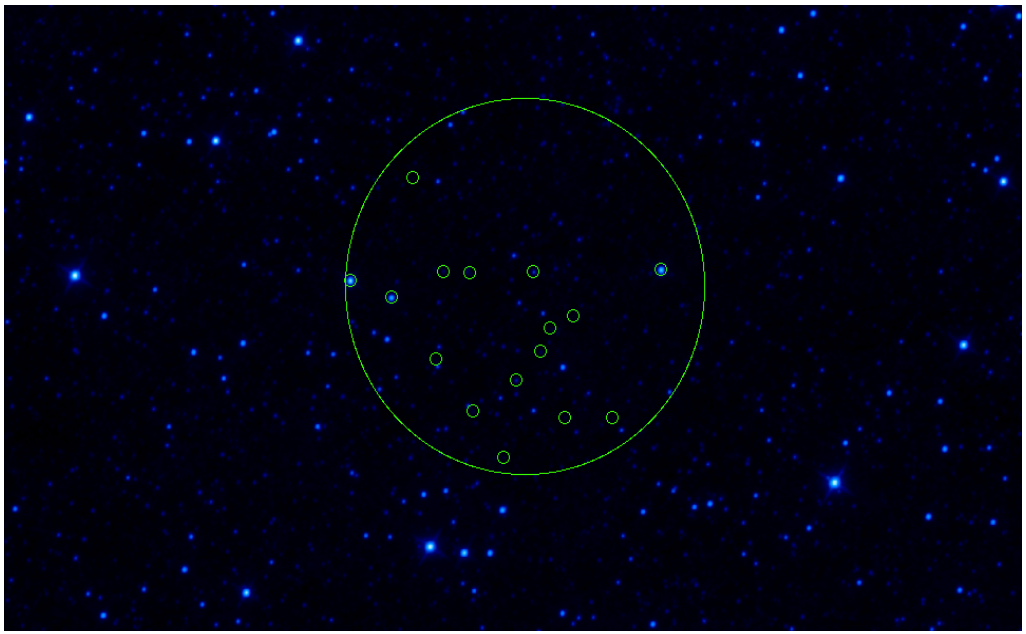
The *WISE* blazars *locus* can be considered a *classifier* whose parameters can be optimized by supervised learning (like a **neural network**) and whose performances (accuracy of the reconstruction) can be assessed consistently.

This systematic approach provides:

- A quantitative criterion to pick the "optimal" values of the parameters of the WGS *locus* model:
  - Accuracy of the reconstruction of the WGS sample;
  - **Completeness** vs **efficiency** of the classification;
- An easy way to adapt the model to different goals;
- Extensibility (adding new constraints from other wavelengths/*features*);
- Easily updatability (new version of blazars catalog, *WISE* photometric dataset, release, etc.);

Association of high-energy sources

Our method for the extraction of candidate blazars can be used to associate unidentified γ-ray sources. **All *WISE* sources in the searching regions (SRs)** (99% uncertainty region) from 2FGL was used to define the SRs) **with non zero are considered and ranked according to multiple criteria**.



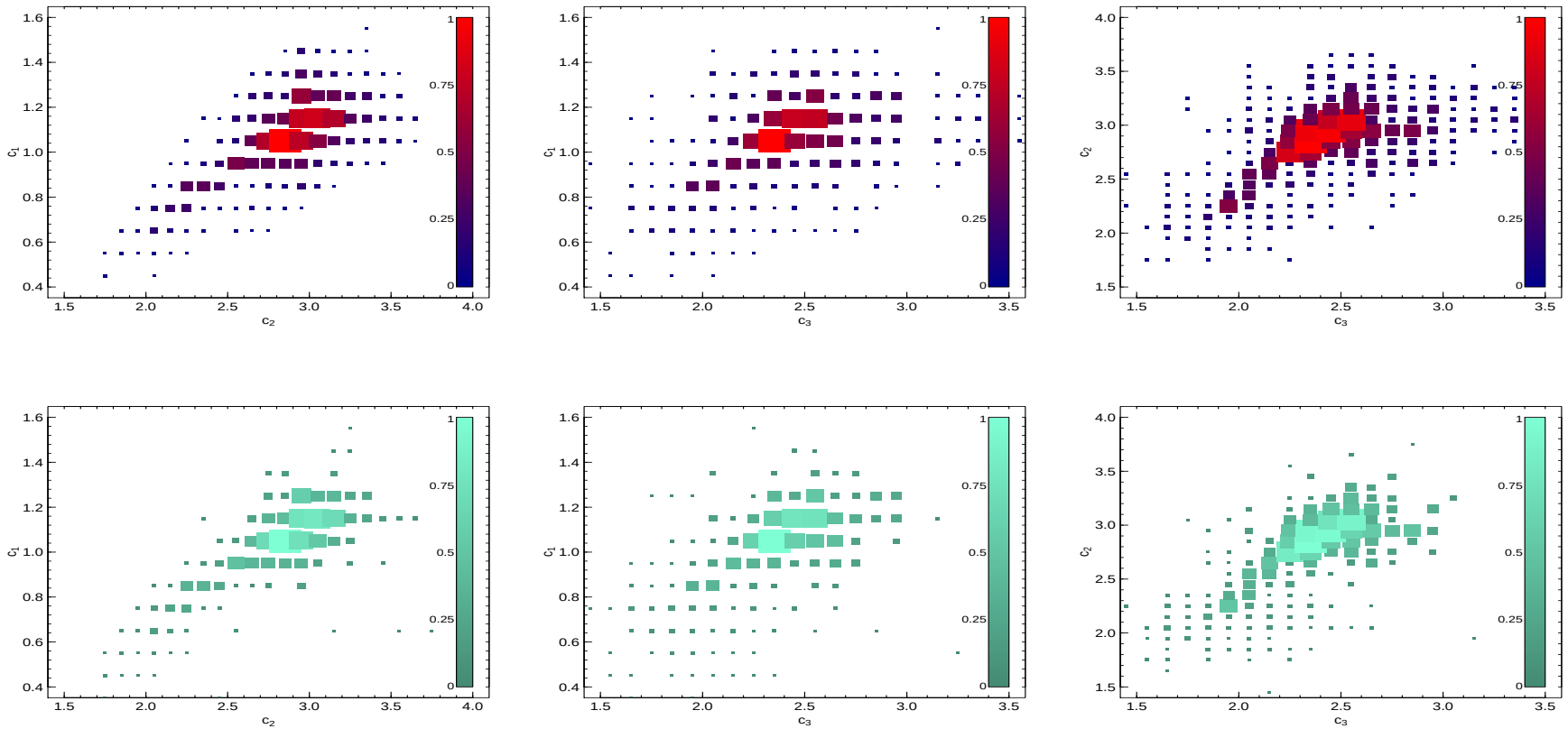
- **Sources are classified in BZB, BZQ or "mixed"**, and they divided in **three classes** by comparing their scores with the scores distribution of the WGS;
- With multiple candidates, the source with the largest class/score is considered the best candidates;
- For candidates with same class/scores (very unlikely), the tie-breaker is the distance from the center of the SR.

Efficiency and completeness

Two numbers convey all the information about the performance of a classification method:

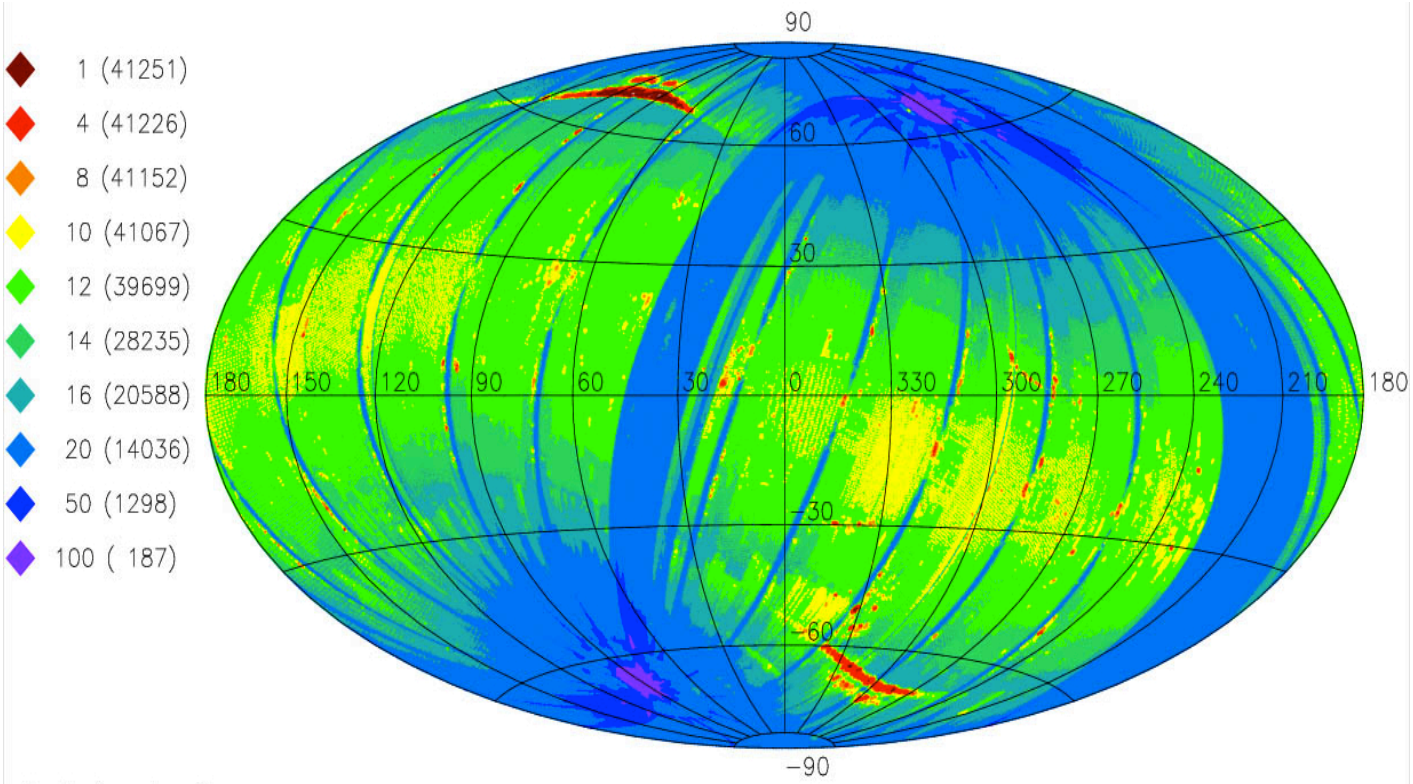
- **Efficiency**: "how may sources selected as candidates blazars turn out to be real blazars?"
- **Completeness**: "which fraction of the known blazars does this method select?"

The **integrated *e* and *c* are not as useful as local *e* and *c***, though (5% cross-validation on the WGS).



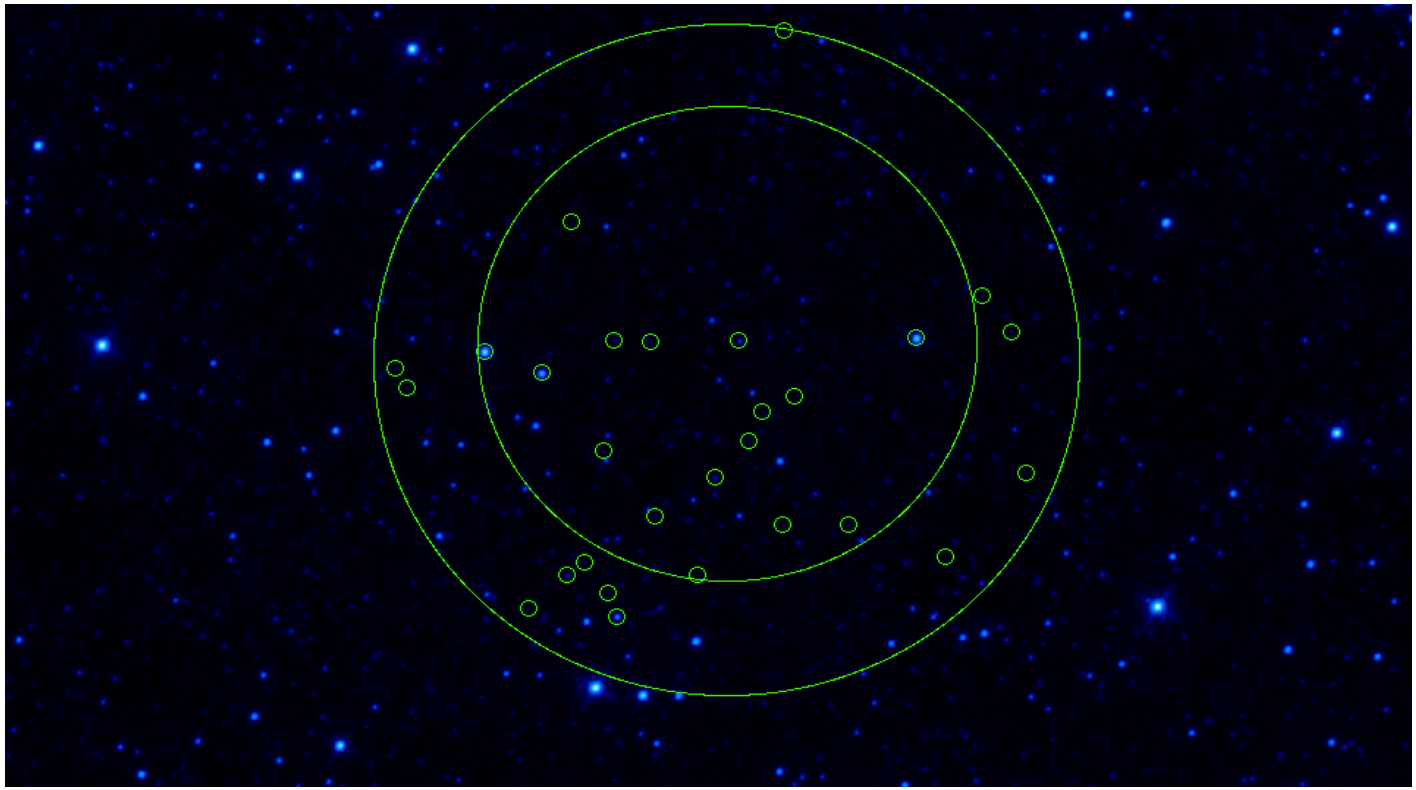
The importance of being *glocal*

The depth of the *WISE* All Sky observations (and, in turn, the density of sources in the catalog) rapidly varies in the sky as function of the galactic latitude and the number of exposure. This suggests that contamination of the association procedure based on the **global** model of WGS has to be evaluated **locally**.

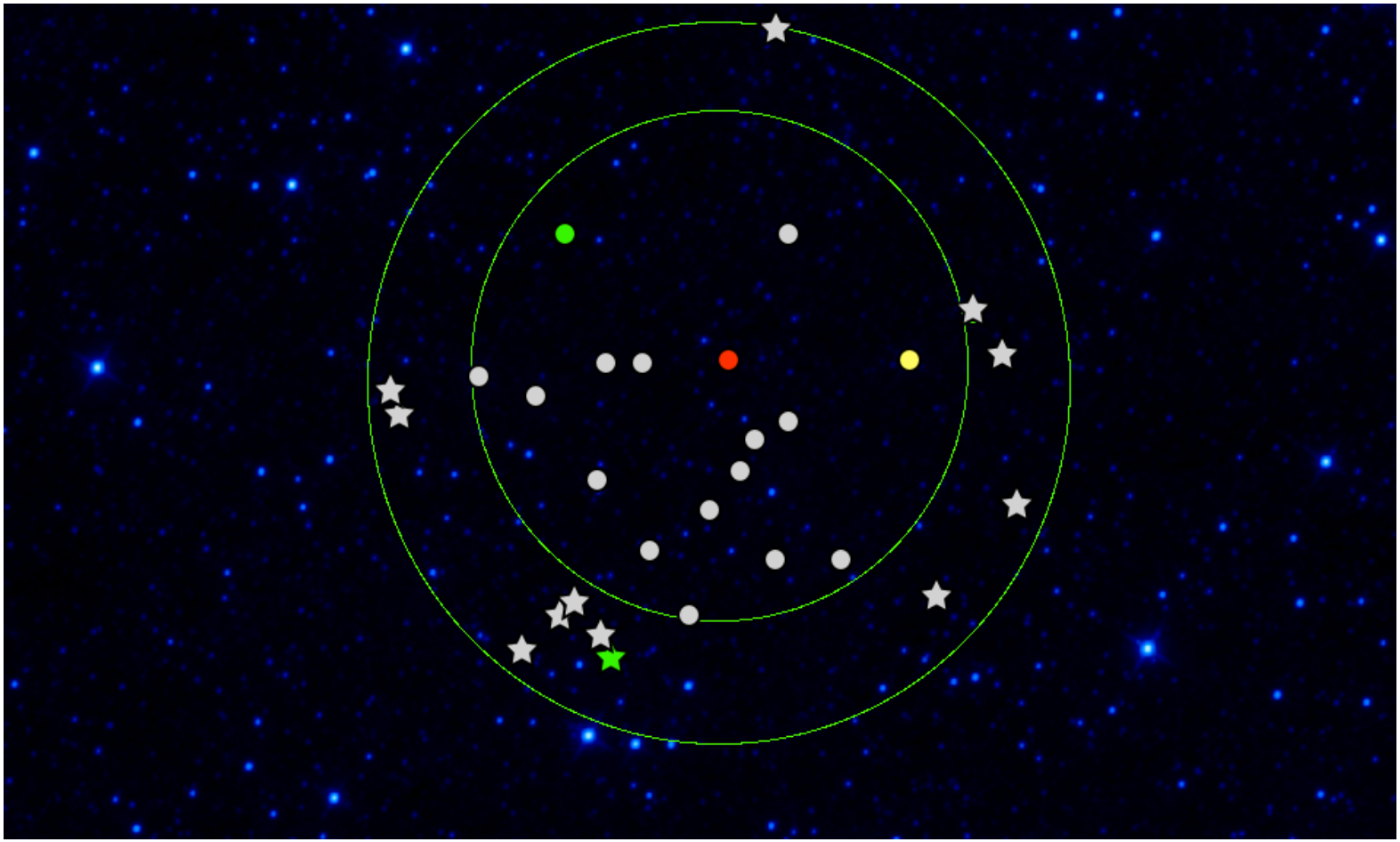


Locally estimated contamination

We evaluate the scores of a **local "background" region surrounding the SR**, and keep track of the sources with not zero score. While these might be interesting *per se* (and allow , the number and score distribution of such sources is an estimate of the local contamination.



The final picture



Conclusions

The peculiar distribution of γ-ray emitting blazars in the *WISE* color space can be used to search blazars candidates in *WISE* photometric catalog

The new **parametric model of the 3D color-space WGS *locus*** and the new **association procedure** improve over the previous methods as the new one:

- based on the 2FGL and *WISE* All Sky Catalog, yielding a more statistically robust determination of the *locus* position in a first place;
- classification of blazar candidates in BL Lac or FSRQs;
- can be easily tweaked to improve one aspect of the association procedure (the **purity** vs **completeness** trade-off);
- estimates the efficiency, completeness and contamination locally (spatially) and as a function of the *WISE* colors of the sources.

Thanks!