Class 12: machine learning & the blob pt 1

Miguel F. Morales
Bryna Hazelton

Final Presentation

- 20 min (5 min intro; 10 min analysis you are doing; 5 min questions)
- Quiz

No additional homework

- Work on final project
- Lots of office hours

Road ahead

Discussion of more advanced ideas

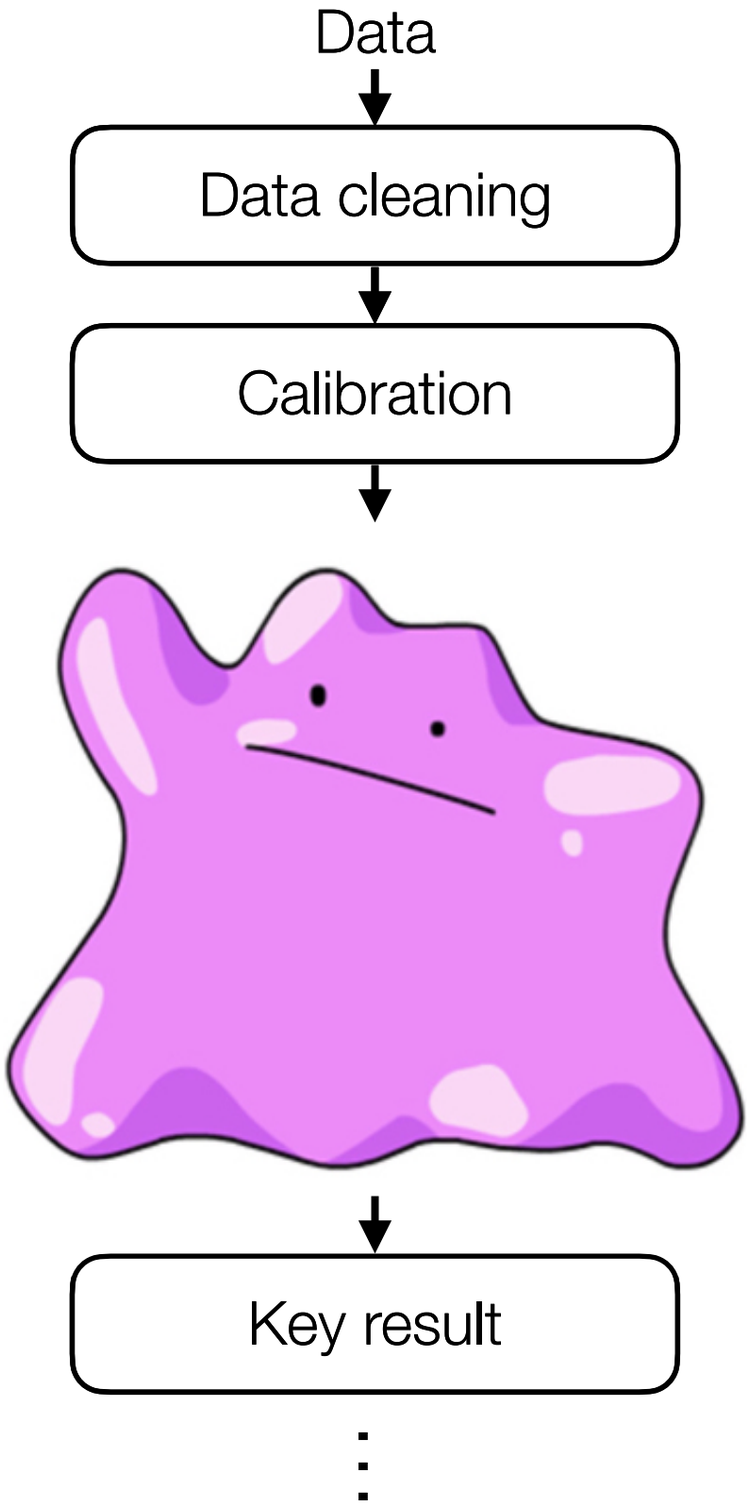
- Machine learning
- Parameters & inherited analyses
- Blind & semi-blind analyses
- Plots as a language
- Data rampages (killing trees for science)

Student presentations

Analysis blobs

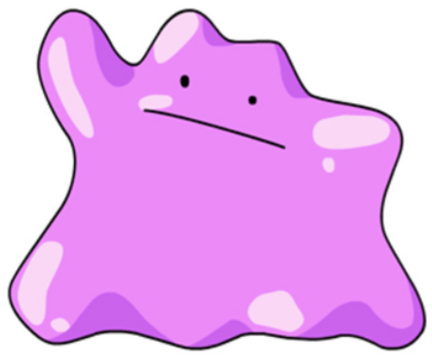
The analysis blob

Machine Learning



Inherited Analysis

Machine Learning (blob pt 1)



Kinds of machine learning

A) Analysis without thinking (saves time!)

B) Teaching a machine to do repetitive work

C) ML assisted research

Repetitive work

Freeze dried undergrads!

- Take about 6 months to train
- Can then unthawed to do repetitive task
- Work in non-OSHA approved situations
- Don’t graduate, Easily replicated

Problem characteristics

- Physics well understood
- You know the right answer
- With care you can figure out how to teach ML to recognize pattern (art of teaching)
- Throwback
- Very powerful for sifting vast amounts of information
- Enables you to concentrate on more advanced questions

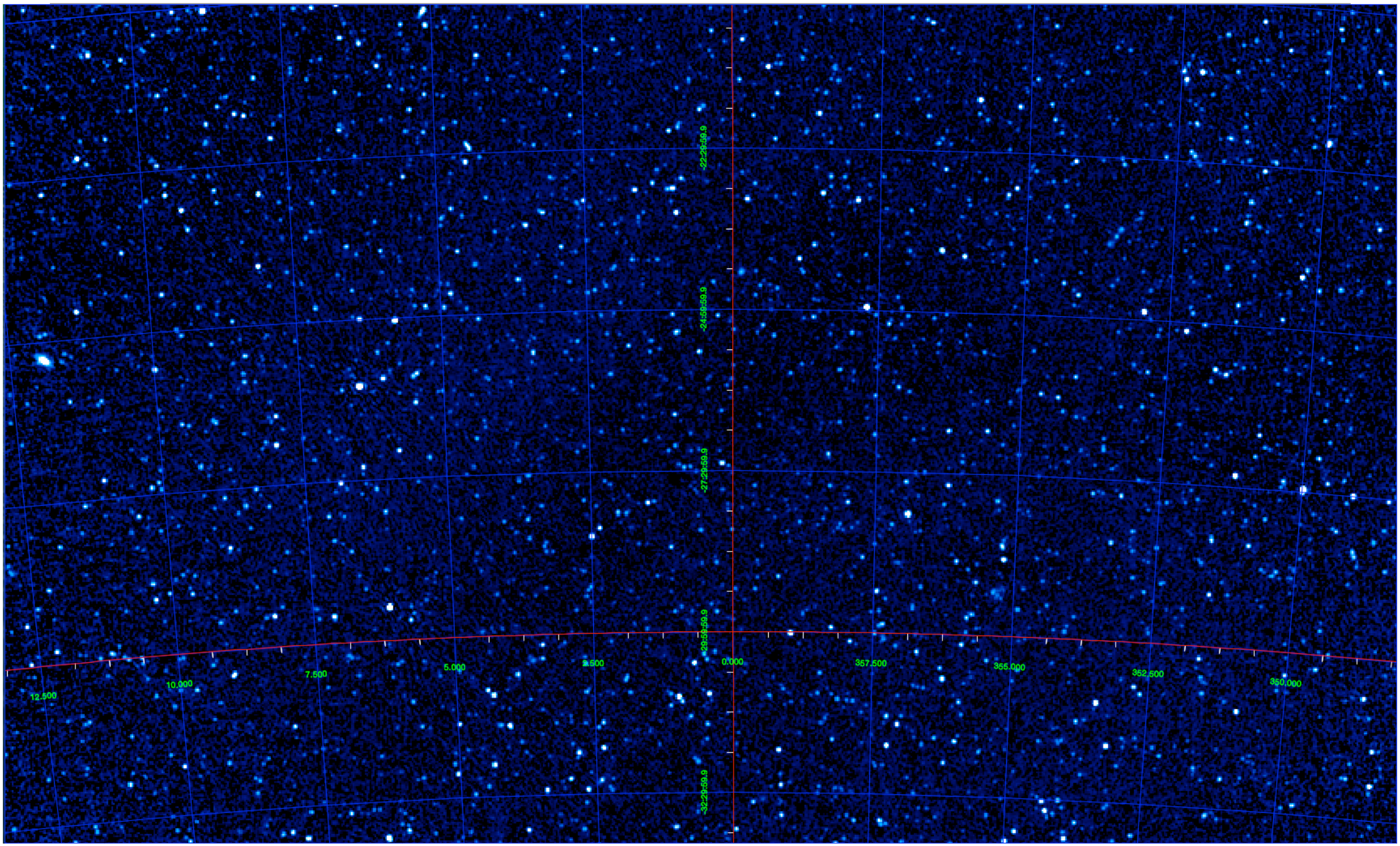
ML assisted research

Iterative ML training as part of research loop

- Searching for increasingly rare events
- Untangling the physics

Inputs and ML

Catalog example



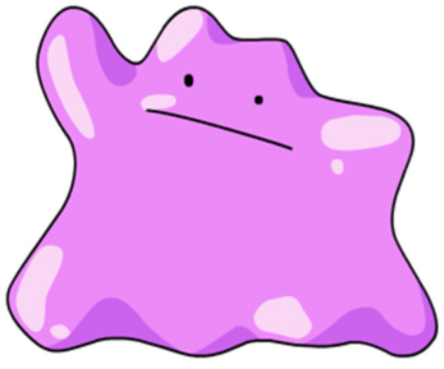
Iterative development

Think about inputs

Data

Deconvolution

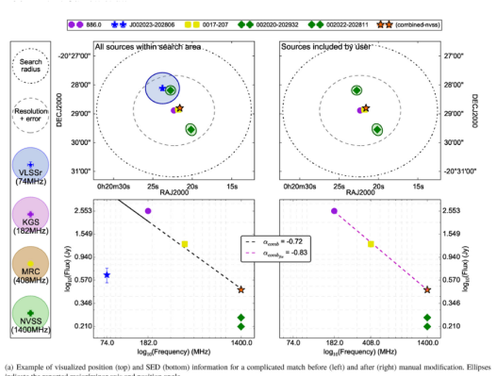
Candidate sources



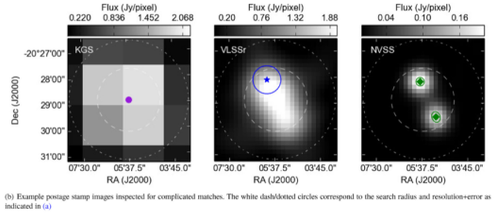
catalog

Train ML

Examine errors



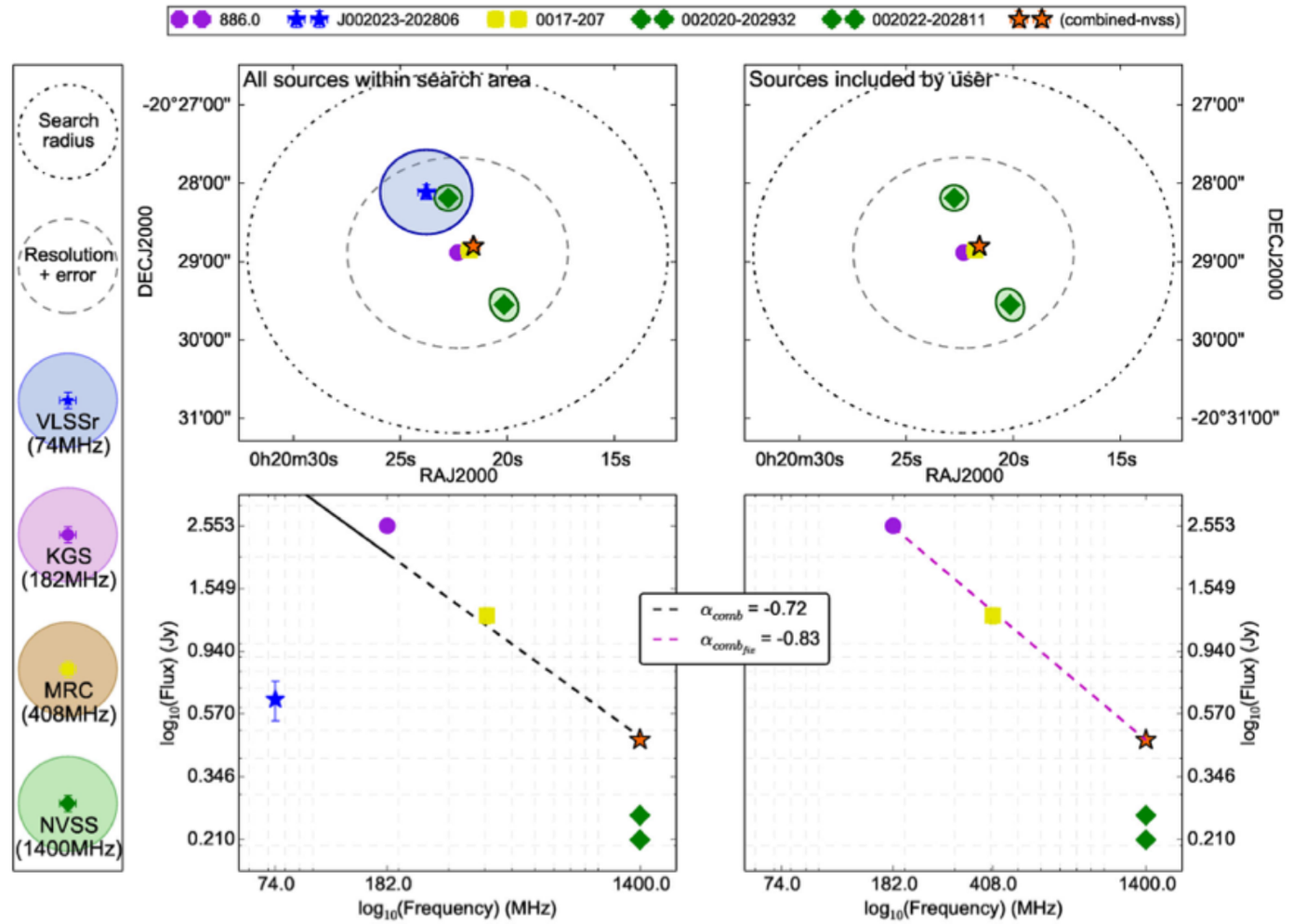
(a) Example of visualized position (top) and SED (bottom) information for a complicated match before (left) and after (right) manual modification. Ellipses indicate the reported major/minor axis and position angle.



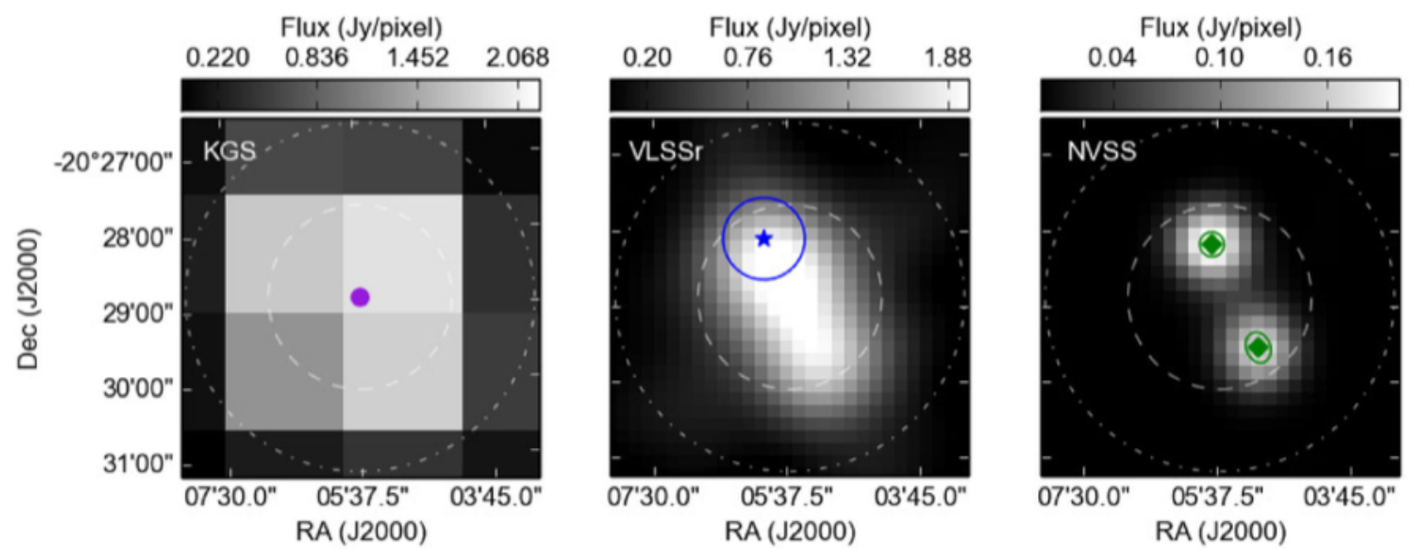
(b) Example postage stamp images inspected for complicated matches. The white dash/dotted circles correspond to the search radius and resolution+error as indicated in (a)

Errors

- Bayesian matching analysis
- Outliers examined by eye using diagnostic plots
- Patterns in errors identified



(a) Example of visualized position (top) and SED (bottom) information for a complicated match before (left) and after (right) manual modification. Ellipses indicate the reported major/minor axis and position angle.



(b) Example postage stamp images inspected for complicated matches. The white dash/dotted circles correspond to the search radius and resolution+error as indicated in (a)

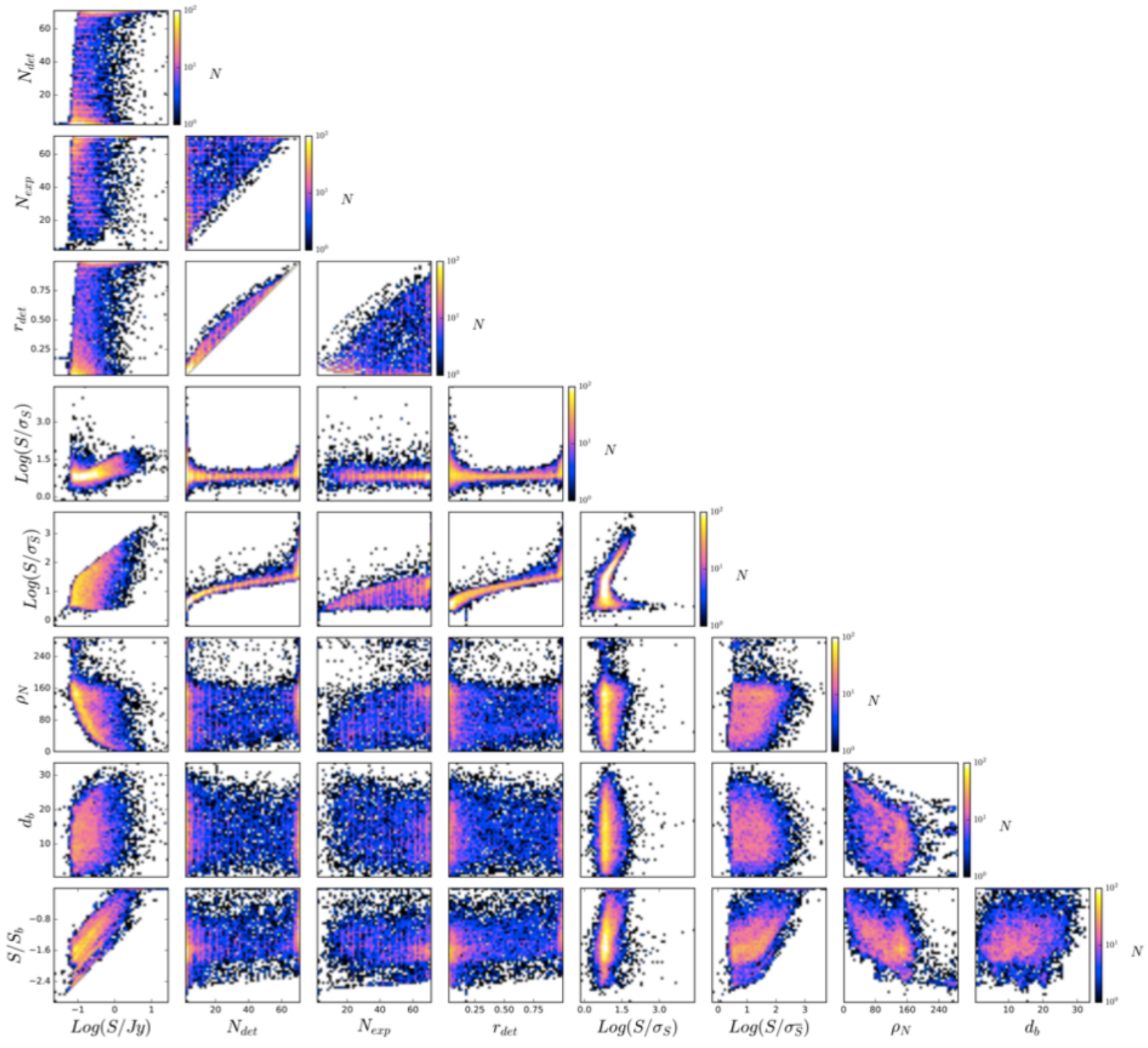
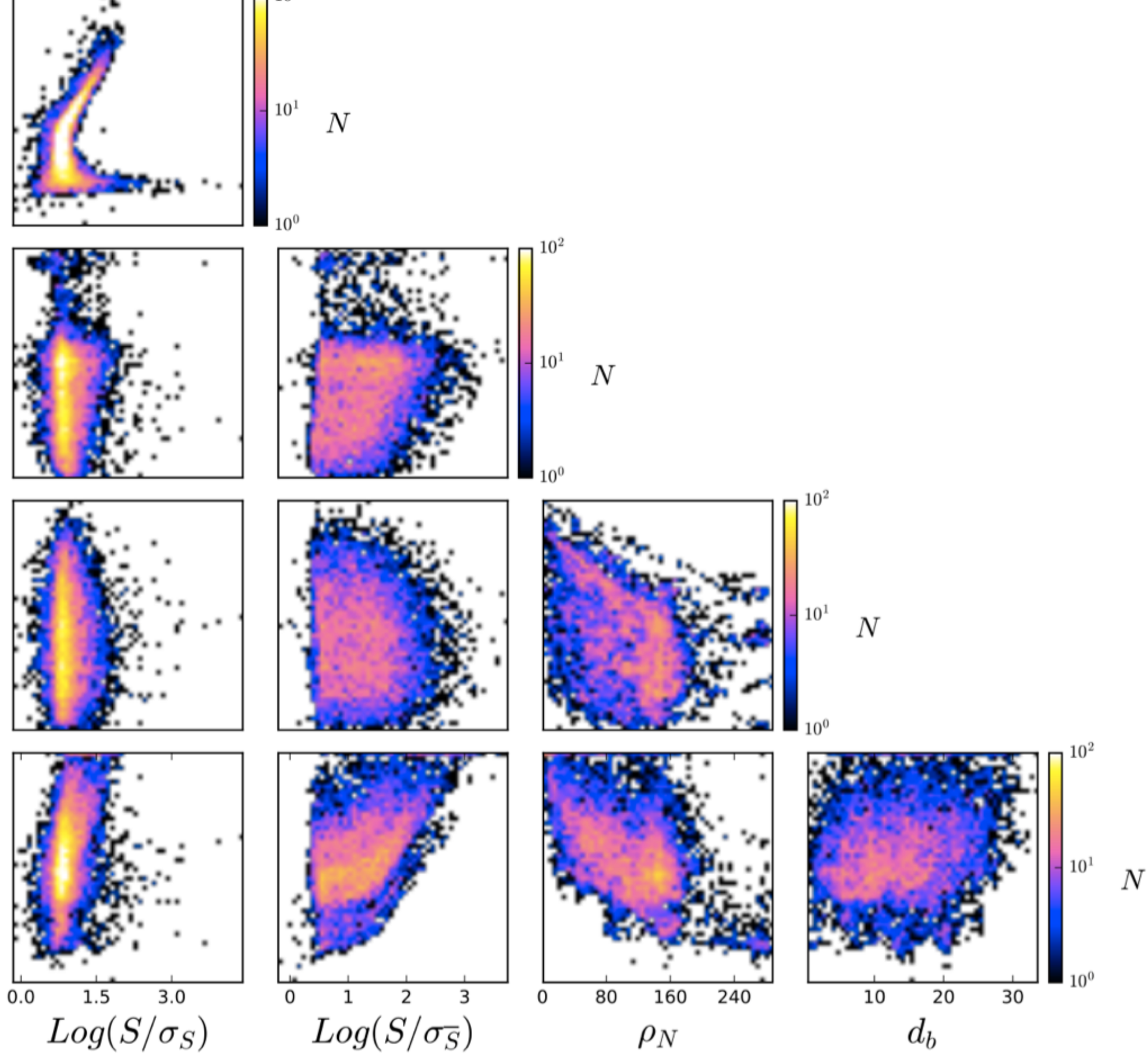


Figure 4.2: The 2D distributions of all 9 input features. The color is log-scaled to highlight structure.


10^0
10^1
N
10^2
0.0
1.5
3.0
$Log(S/\sigma_S)$
0
1
2
3
$Log(S/\sigma_{\overline{S}})$
0
80
160
240
ρ_N
10
20
30
d_b

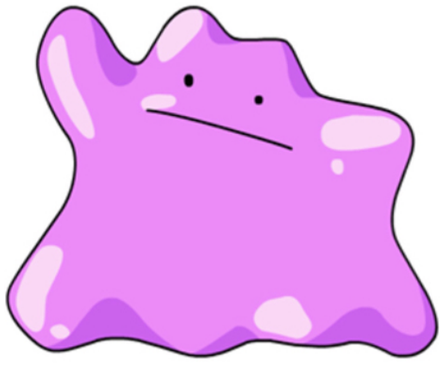
Iterative development

Think about inputs

Data

Deconvolution

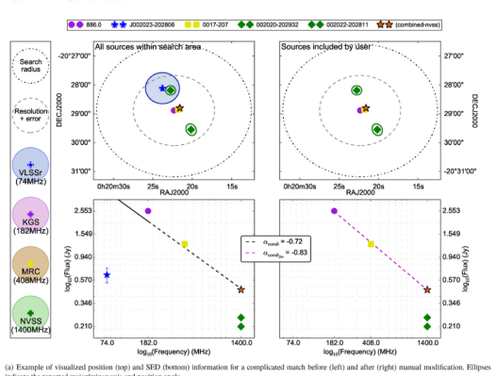
Candidate sources



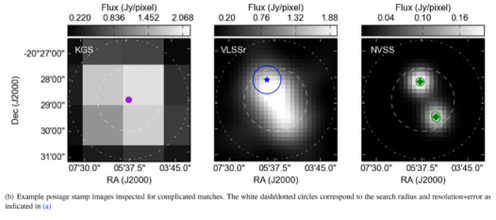
catalog

Train ML

Examine errors



(a) Example of visualized position (top) and SED (bottom) information for a complicated match before (left) and after (right) manual modification. Ellipses indicate the reported major/minor axis and position angle.



(b) Example postage stamp images inspected for complicated matches. The white dash/dotted circles correspond to the search radius and resolution+error as indicated in (a)

ML assisted analysis

- Increasingly rare events
- Develop sophisticated non-linear weighting
- Untangle the key physics (parameters encode science)

Significance

- Throw many signal-free events (data or MC), plot the output.