Class 14: more dragons & machine learning

Miguel F. Morales
Bryna Hazelton

Road ahead

Discussion of more advanced ideas

- Machine learning
- Deconvolution, forward modeling, CLEAN
- Blind & semi-blind analyses
- Plots as a language
- Data rampages (killing trees for science)

Student presentations

Deconvolution, forward modeling, CLEAN & chaos

Information loss

$$\text{data} = \text{Measuremnt} \times \text{world}$$

$$d = Mw$$

$$\hat{w} = M^T d$$

Instrument response function embodies information lost in the measurement process

- PSF in telescopes
- Energy resolution
- etc.

Deconvolution, forward modeling, CLEAN

Typically used when both:

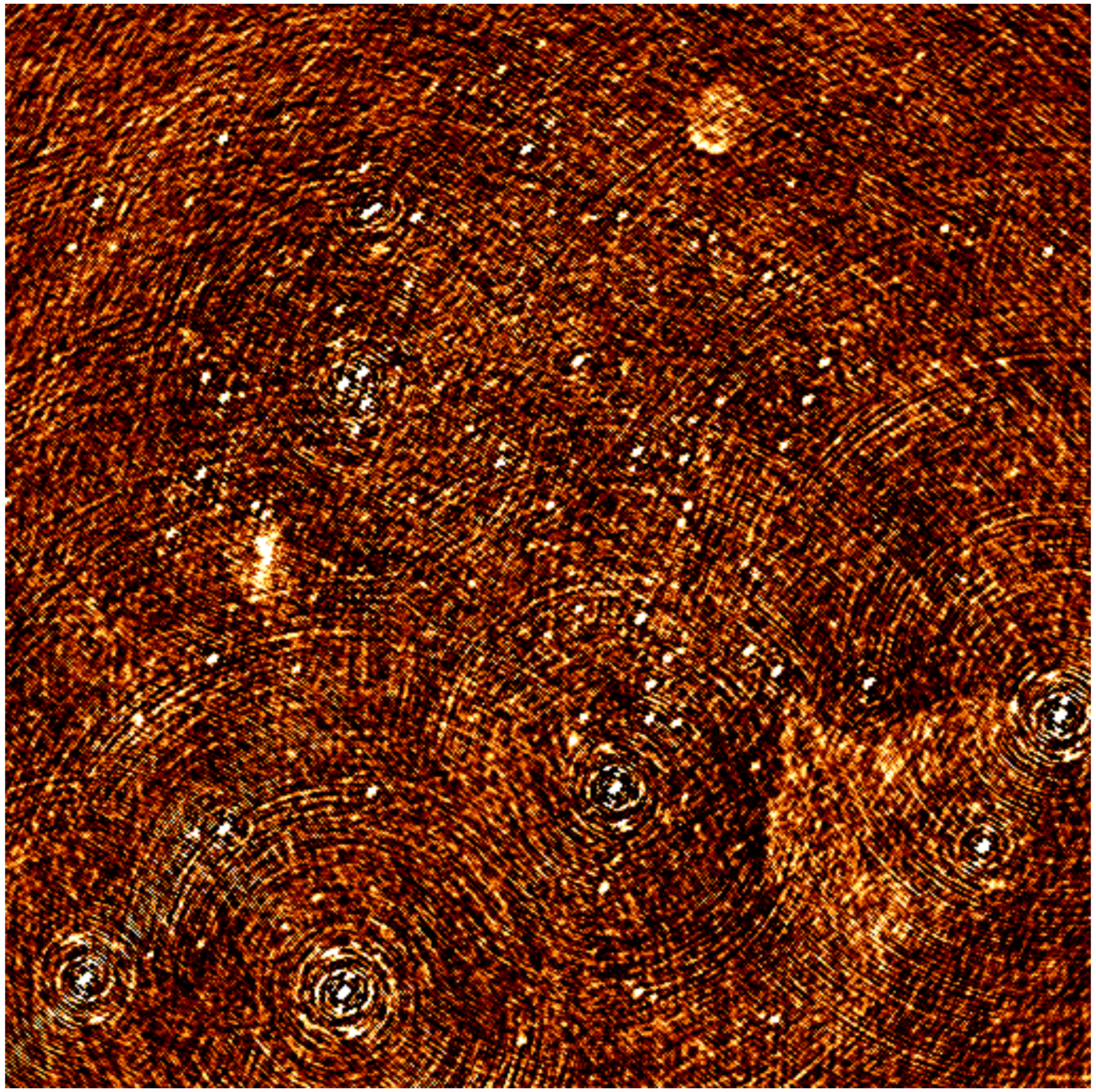
- Instrument response is ‘ugly’ (multiple peaks, wildly non-linear, etc.)
- and the world is sparse (isolated stars, a few emission lines, most of $\hat{w}$ vector are zeros
- Look for ‘peaks’ and forward model data

Typical deconvolution loop

- Look for feature (e.g. star, emission line)
- Guess feature is there, partially update model of the world but at only fraction of observed strength (loop gain)
- Forward model data from updated model
- Repeat until data-model data no longer improves (almost never converge)

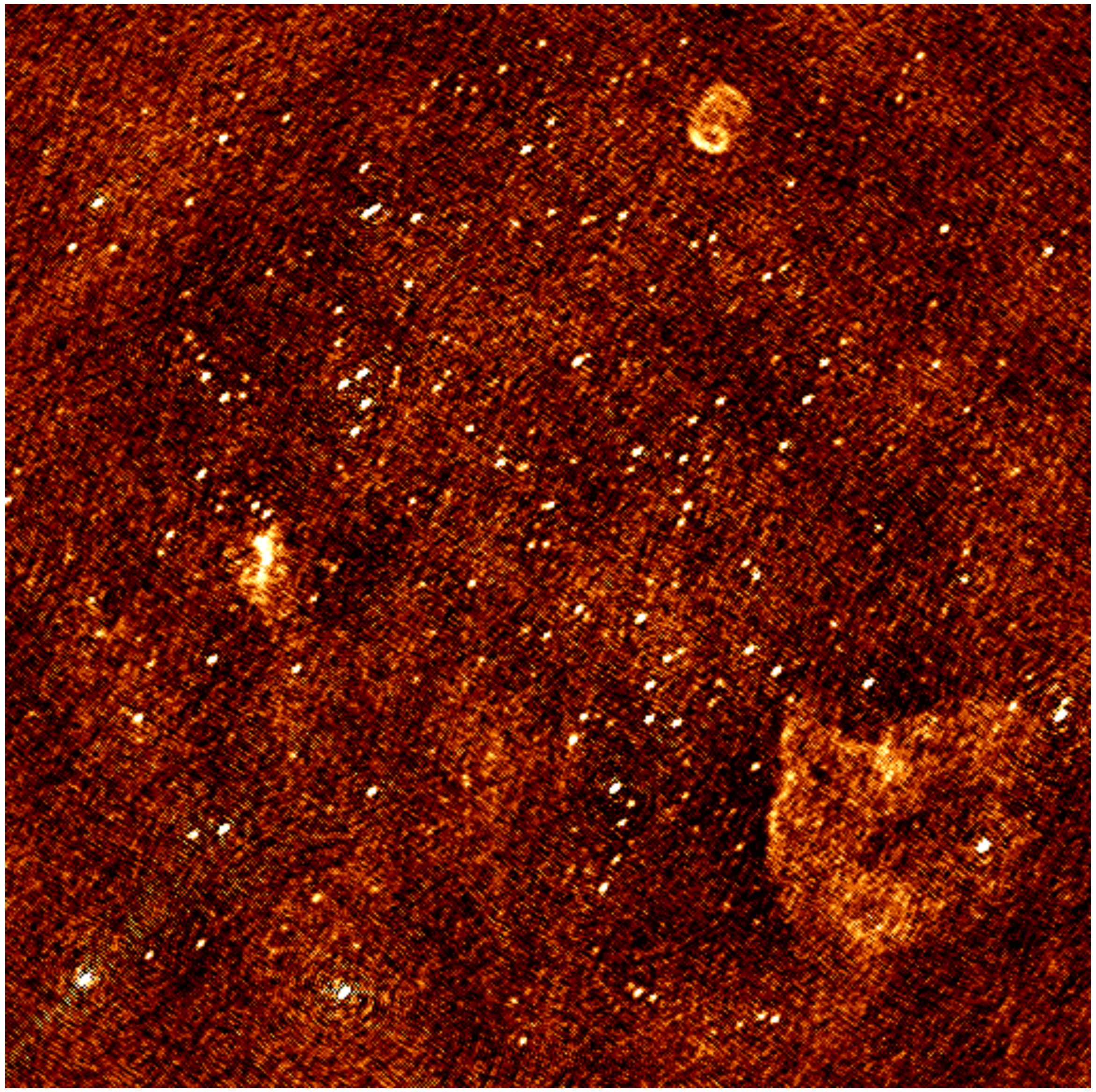
CLEAN

clean-dirty.fits



$$\hat{w} = M^T d$$

clean-image.fits

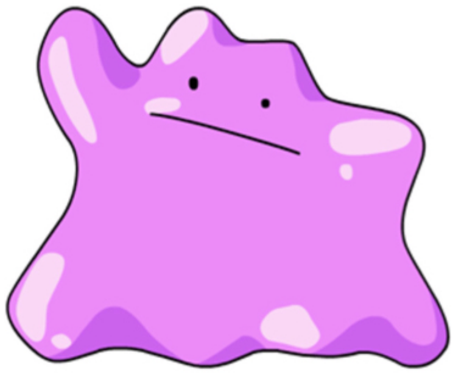


$$\hat{w} = Dd$$

Deconvolution

- A form of pseudo inverse
- Can work great (nasty response; simple world)
- Why they are unpleasant
  - ‘Decision points’ in algorithm make them formally *chaotic!*
  - Loop gain tuned to help convergence
  - Tend to have algorithm settings, which depend on implicit assumptions
  - Errors extremely hard to predict
  - False or catastrophic errors a common concern

Machine Learning



Kinds of machine learning

A) Analysis without thinking (saves time!)

B) Teaching a machine to do repetitive work

C) ML assisted research (data sieve)

D) ML assisted research (surfacing physics)

Do it with ML!

- 80% of ML projects
- Implicit goal is to not have to think about data
- Never works

Rule: ML takes more time to implement
than a traditional analysis

Repetitive work

Freeze dried undergrads!

- Take about 6 months to train
- Can be 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 to early days of physics & astronomy
- Very powerful for sifting vast amounts of information
- Can enable you to concentrate on more advanced questions

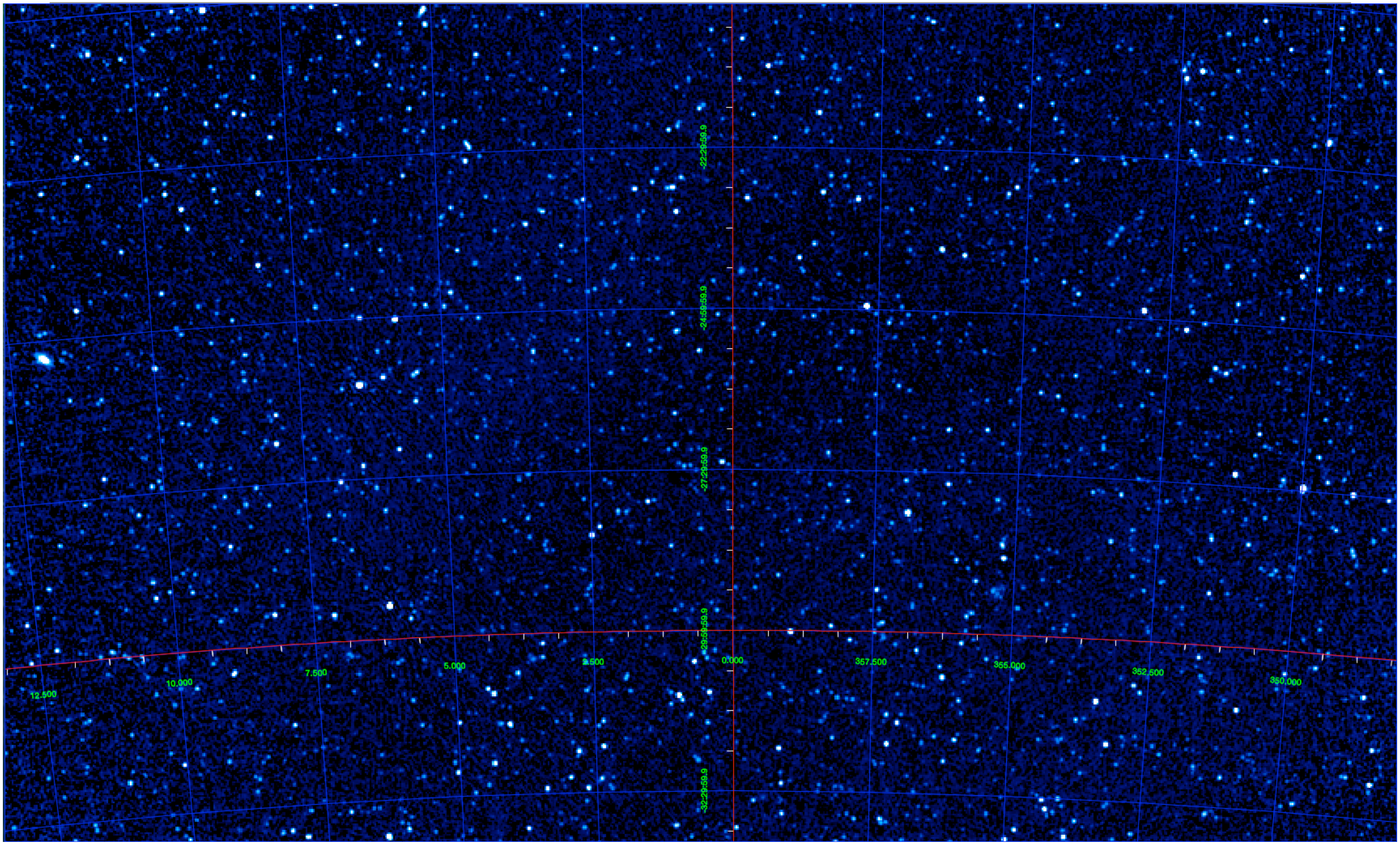
ML assisted research

Iterative ML training as part of research loop

- Searching for increasingly rare events
- Untangling the physics

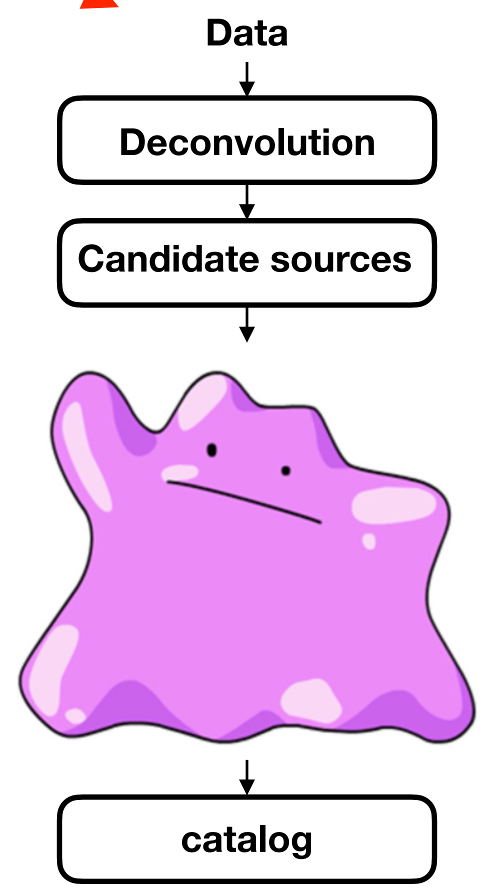
Inputs to ML & repetitive work example

Catalog example



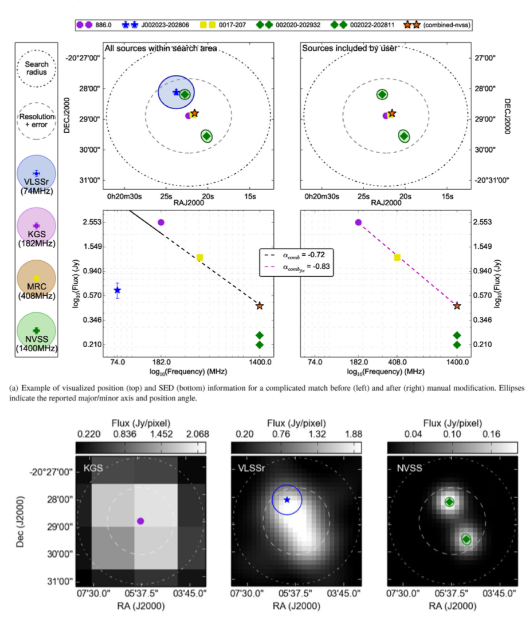
Iterative development

Think about inputs



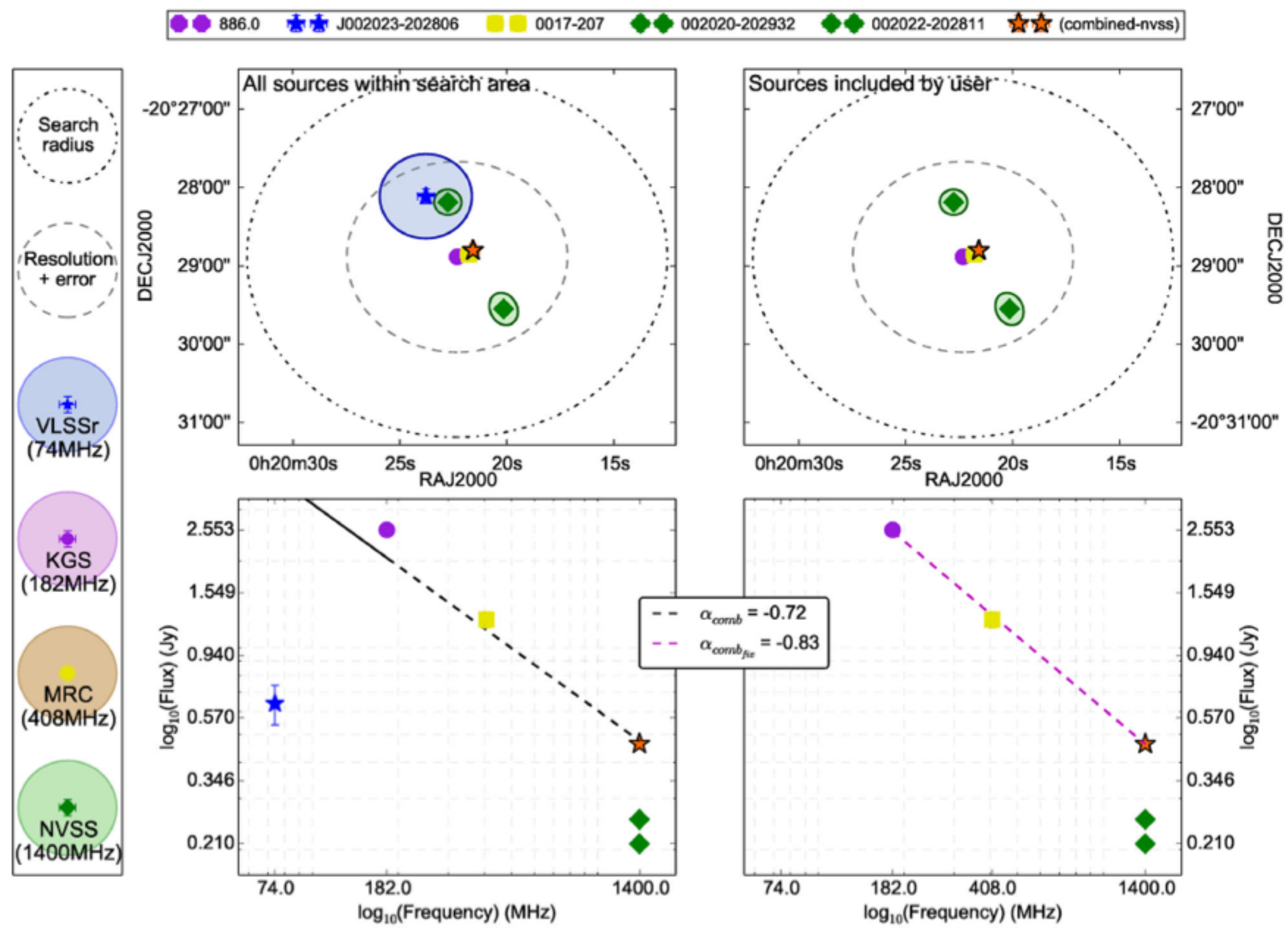
Train ML

Examine errors

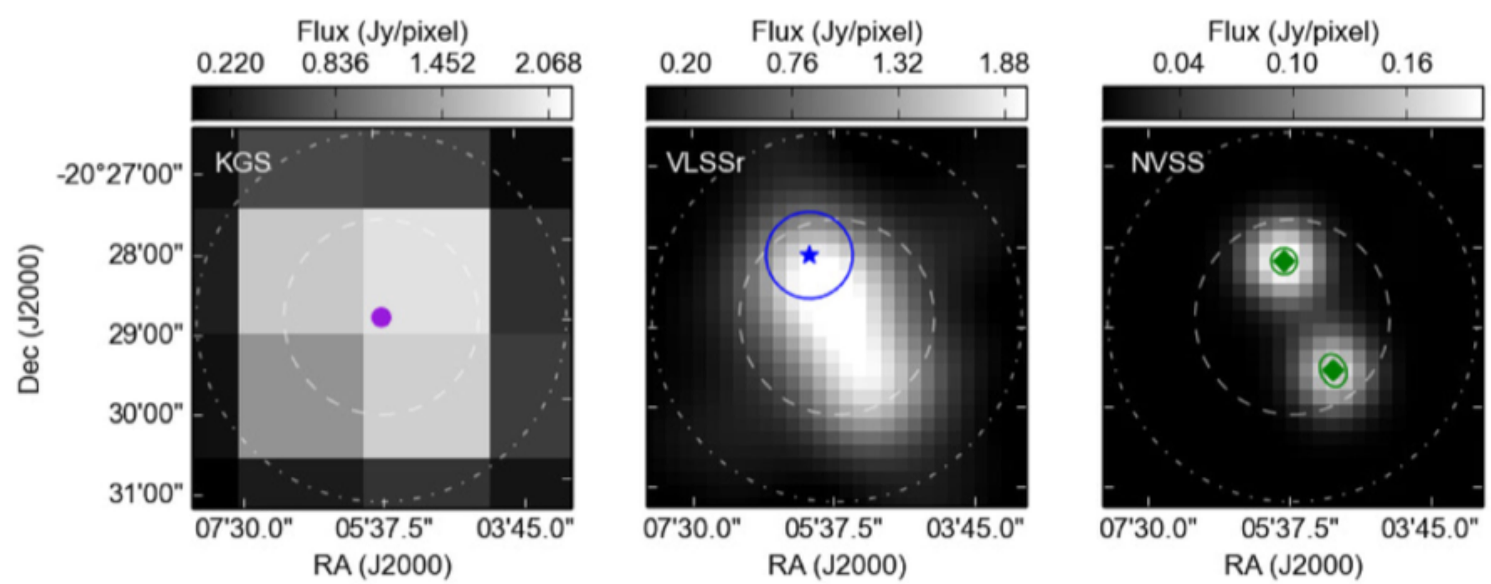


Errors

- Bayesian matching analysis (other catalogs help with training)
- 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)

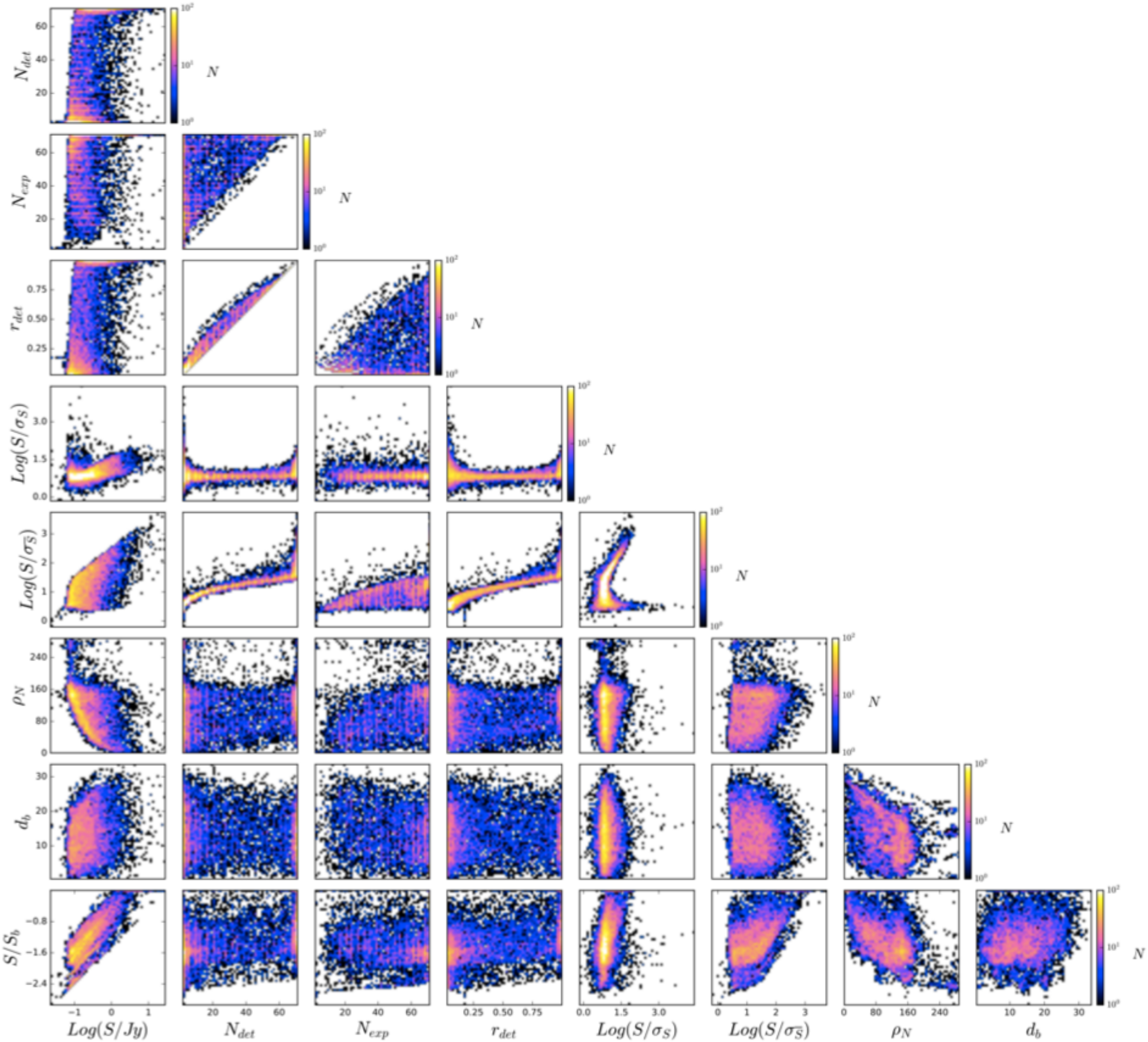
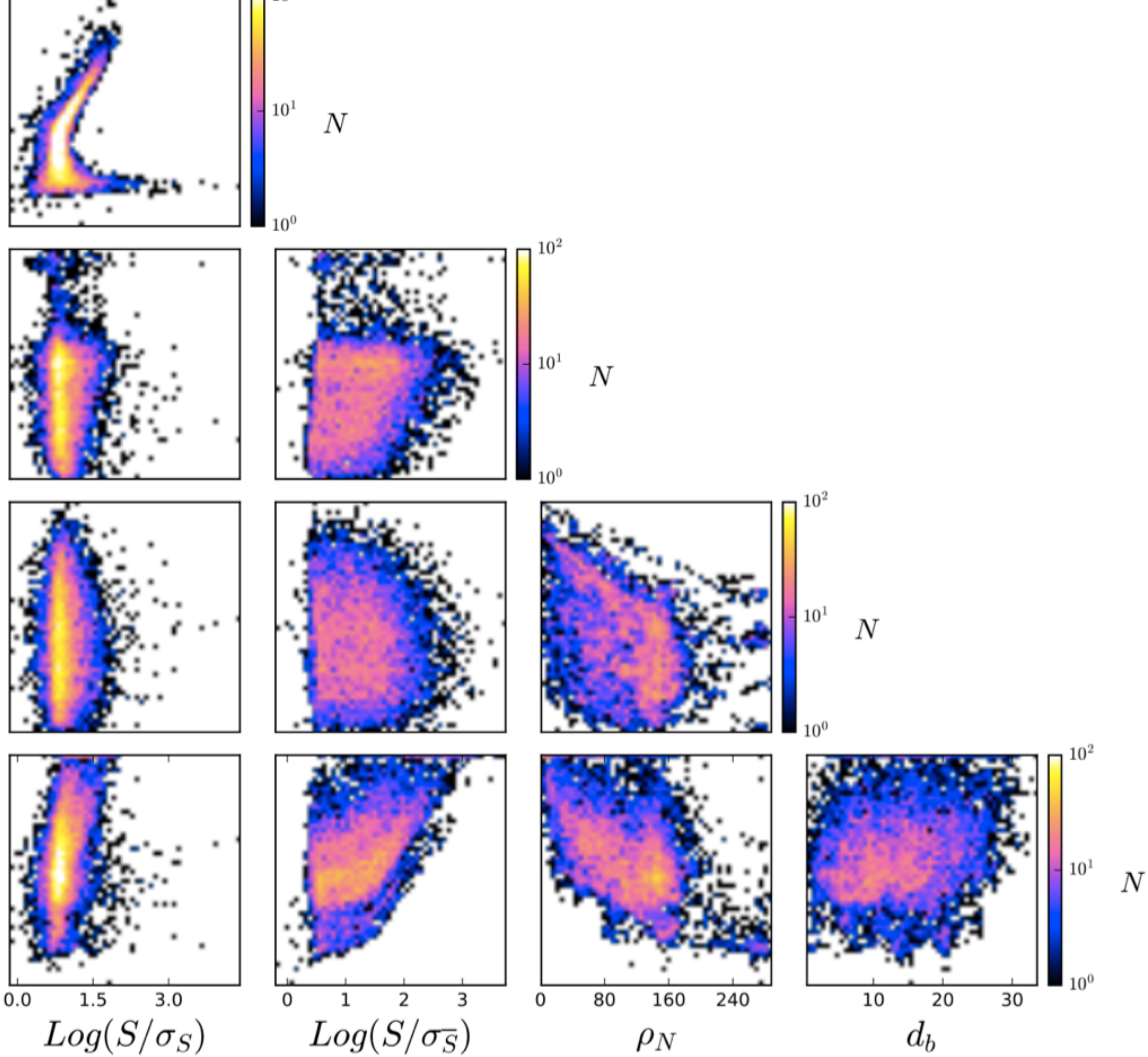


Figure 4.2: The 2D distributions of all 9 input features. The color is log-scaled to highlight structure.



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

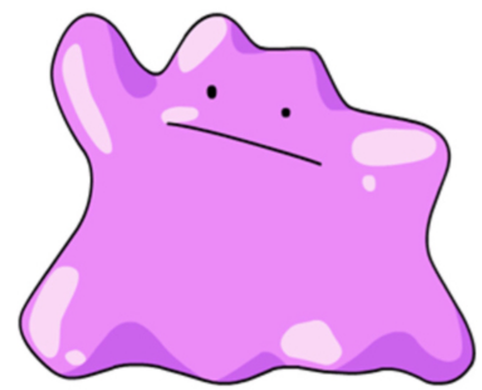
Iterative development

Think about inputs

Data

Deconvolution

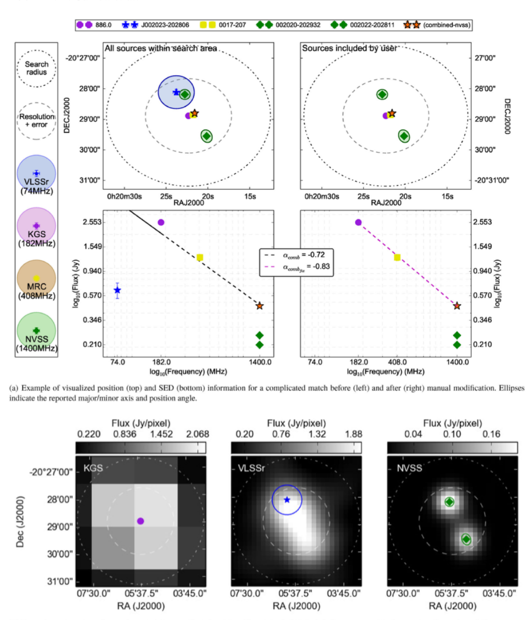
Candidate sources



catalog

Train ML

Examine errors



Kinds of machine learning

A) Analysis without thinking (saves time!)

B) Teaching a machine to do repetitive work

C) ML assisted research (data sieve)

D) ML assisted research (surfacing physics)

Sam Tetef Thursday

ML assisted analysis

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