Probabilistic Cataloguing

Stephen K N Portillo
with Benjamin C G Lee, Tansu Daylan and
Douglas P Finkbeiner
29 September 2017
INPA Seminar

Outline

- Why use probabilistic cataloging?
- What is probabilistic cataloging?
- Application to faint source populations
- Application to crowded stellar fields
- Extending probabilistic cataloging to star + galaxy fields

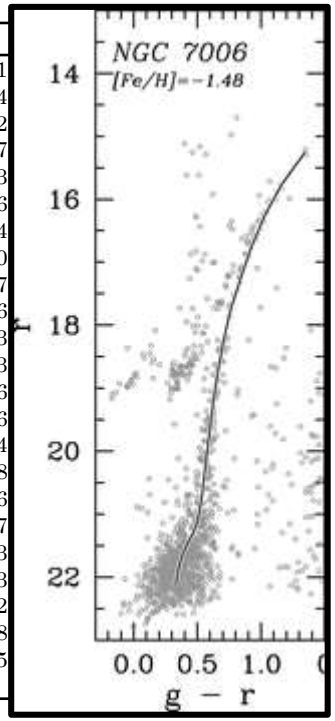
Why use probabilistic cataloging?

Telescopes don't make catalogues!

Slide title stolen from Hogg and Lang, EAS Publication Series 45, 351 (2011)
Image: SDSS DR 12

People make catalogues

RA (J2000)	DEC (J2000)	g	r
229.4351	2.010923	19.344	19.1
229.4353	1.990166	23.070	21.4
229.4358	2.033374	21.809	21.2
229.4361	2.070269	20.107	19.7
229.4362	1.997957	22.894	21.3
229.4364	2.048578	22.386	21.6
229.4366	2.053515	20.853	20.4
229.4369	2.103516	21.827	21.0
229.4369	2.043476	23.067	21.7
229.4370	2.051732	19.960	19.6
229.4371	2.102266	20.813	20.3
229.4373	2.052342	20.785	20.3
229.4374	1.996688	21.161	20.6
229.4376	2.133210	22.476	21.6
229.4378	2.039289	20.883	20.4
229.4380	2.077996	22.682	21.8
229.4380	2.043483	22.884	21.6
229.4381	2.045585	20.111	19.7
229.4382	2.011463	22.069	21.3
229.4382	2.029807	19.625	19.3
229.4382	2.030182	17.835	17.2
229.4385	2.157053	22.193	21.8
229.4385	2.147021	22.492	21.5
...	...	...	...



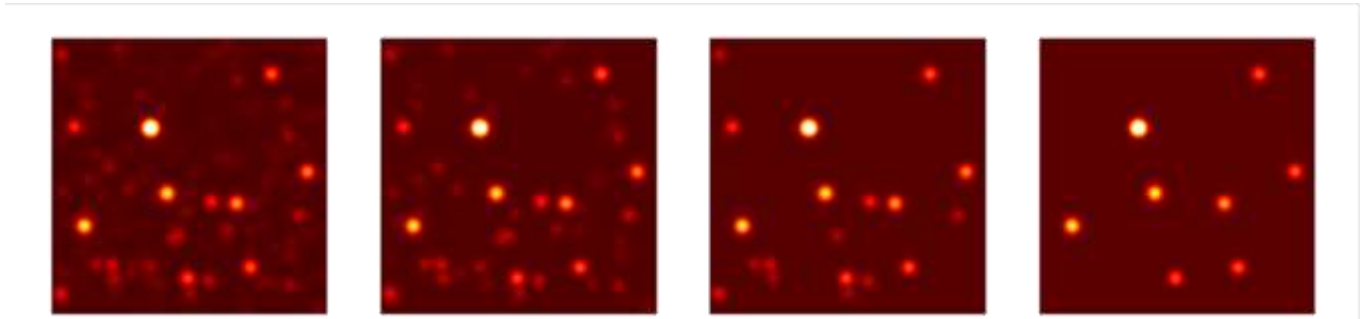
images → catalogues → science

Catalogue and color-magnitude diagram: An et al. ApJS 179, 328 (2008)

Faint Source Populations

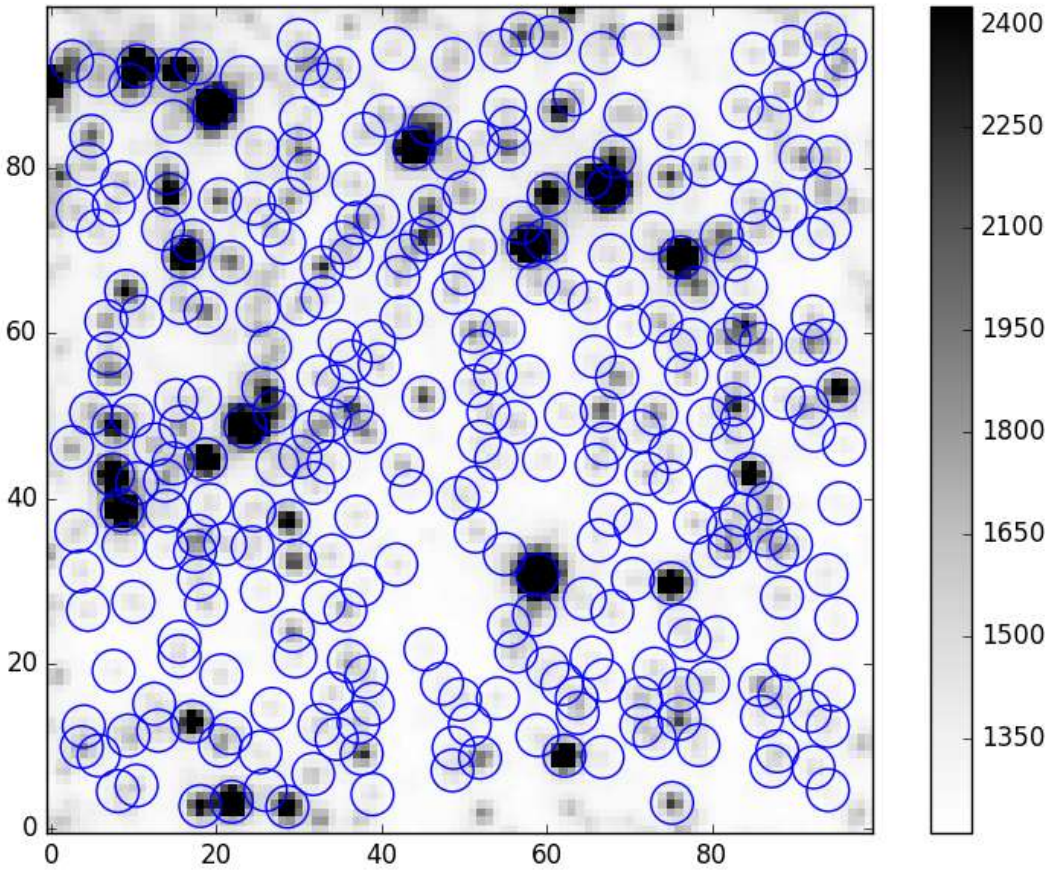
increasing F_{min}, I_{iso}
constant $\int I \, d\Omega$

finite counts infinite counts



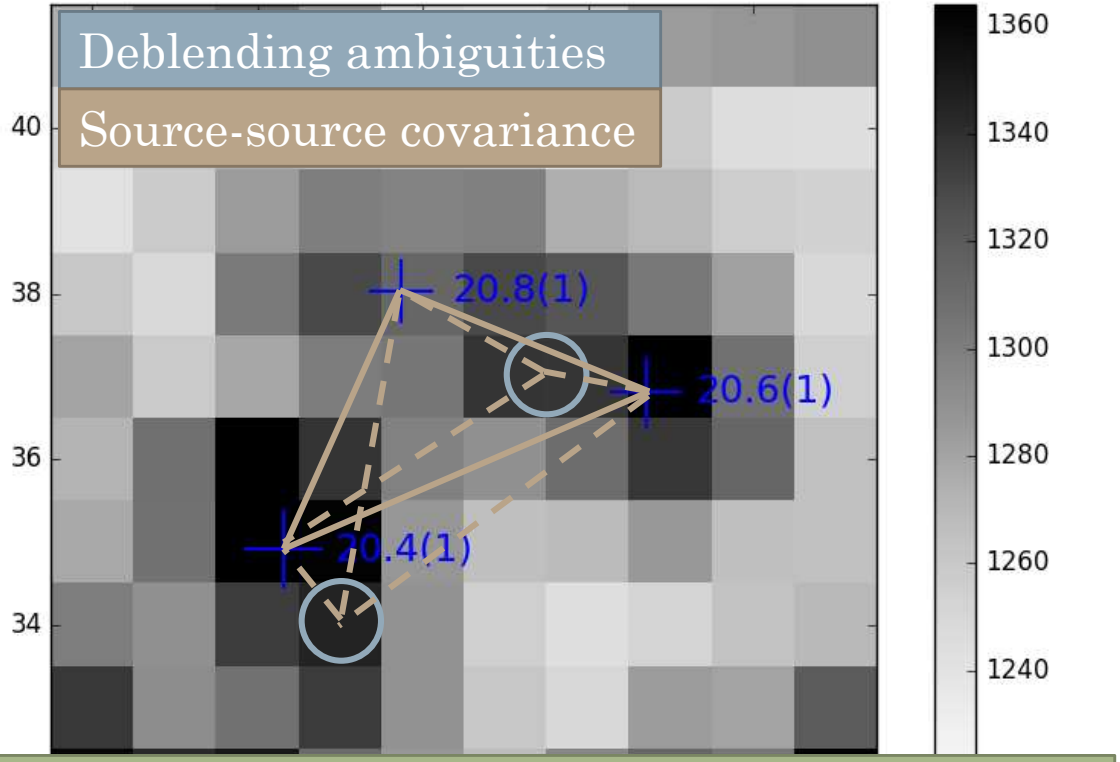
Counting statistics introduce a point source detection threshold
But even sources fainter than this threshold can affect image

Crowded Field Cataloguing



Data: SDSS DR12
Catalog: An et al. ApJS 179, 328 (2008)

Traditional Catalogue Issues

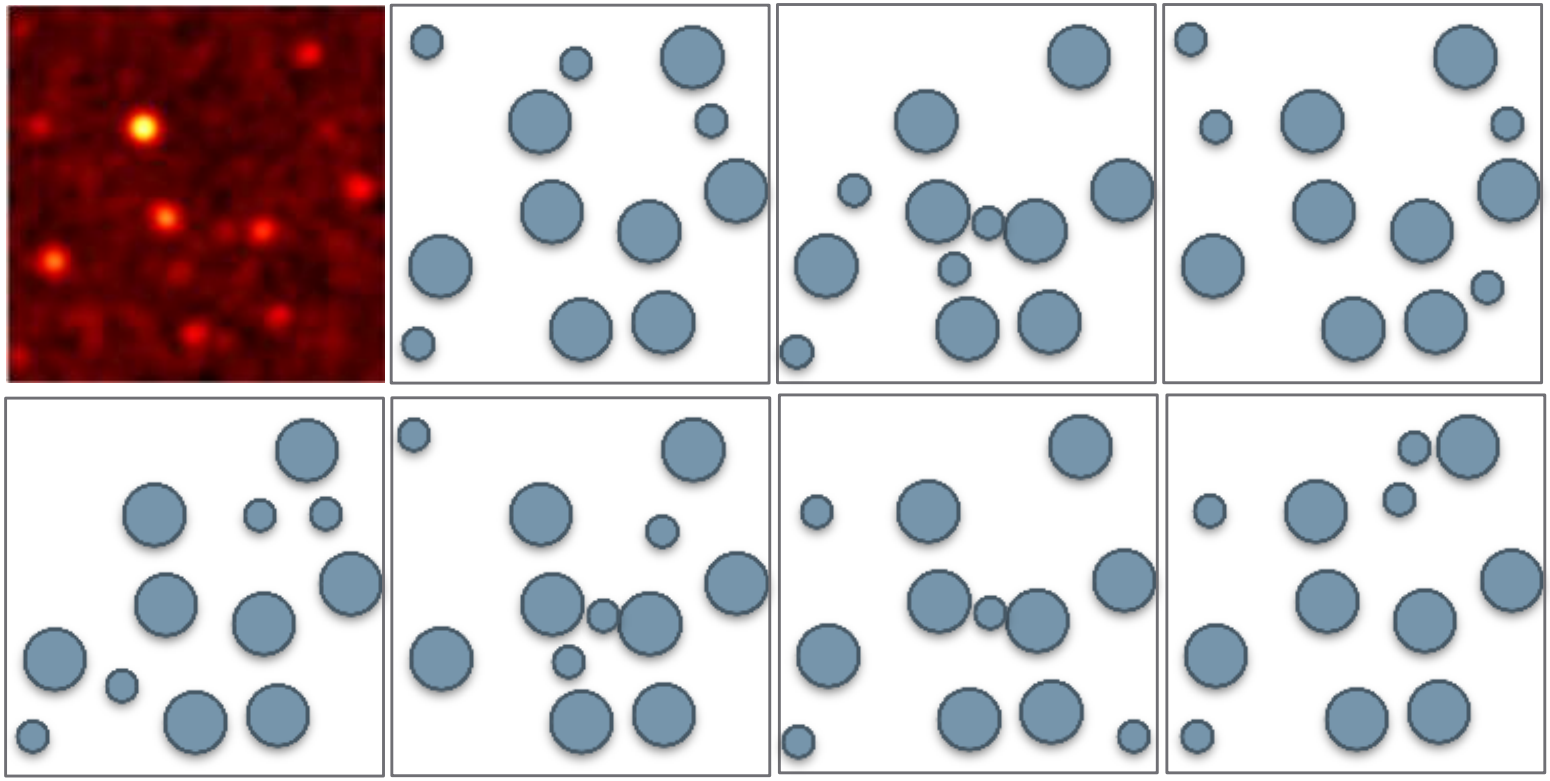


These issues will become more relevant as we build more sensitive ground-based telescopes

What is probabilistic cataloging?

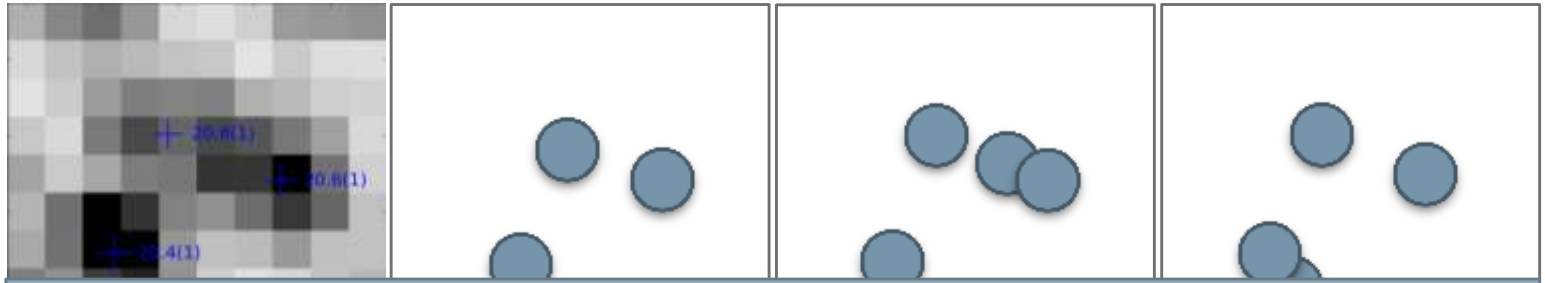
Probabilistic Cataloguing

- Infer an ***ensemble of catalogues***
- Reflects populations of faint sources that are compatible with the data, without overfitting

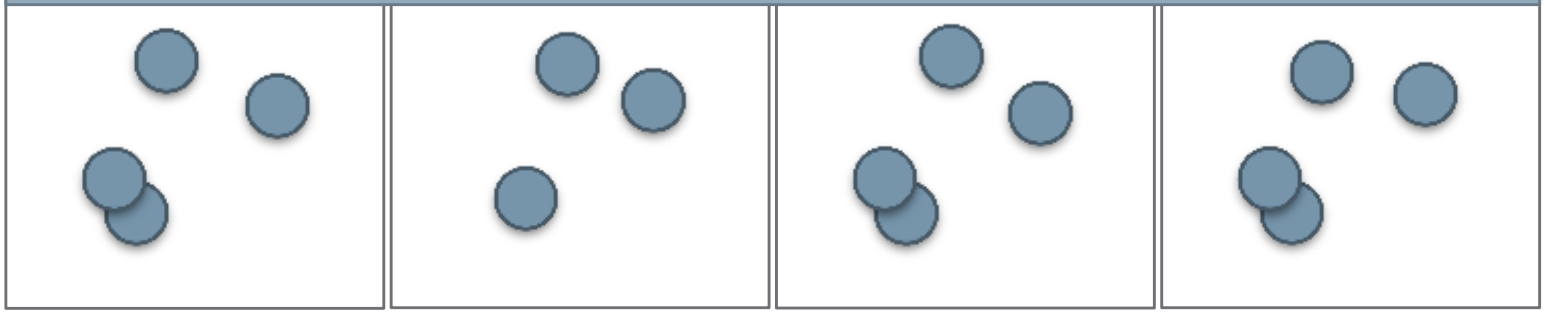


Probabilistic Cataloguing

- Infer an ***ensemble of catalogues***
- Naturally handles deblending ambiguities and source-source covariance in crowded fields



- Ensemble samples Bayesian posterior distribution of catalogues
  $P(\text{Catalogue}|\text{Image}) \propto \pi(\text{Catalogue})\mathcal{L}(\text{Image}|\text{Catalogue})$
- Explore catalogue space with Markov Chain Monte Carlo

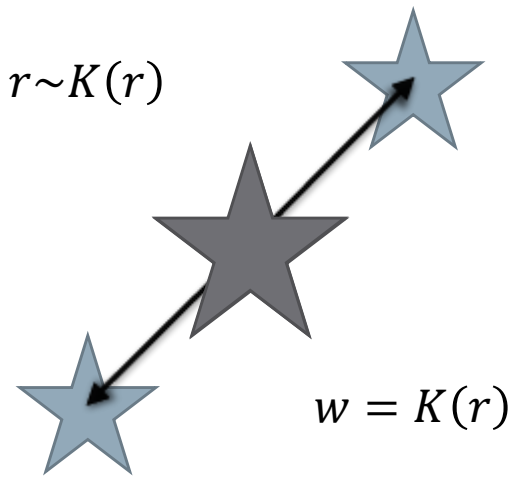


Bayesian Framework

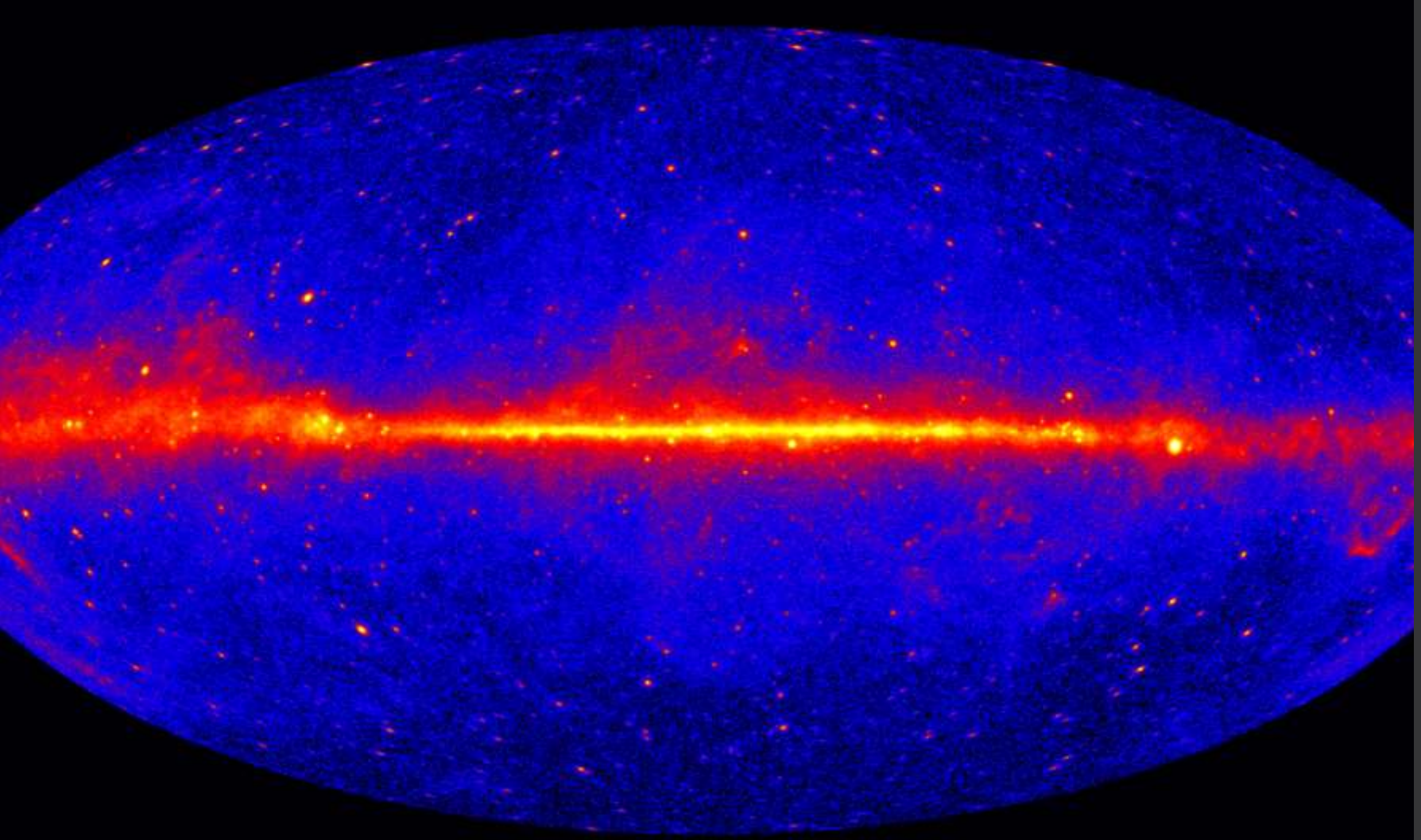
- Likelihood
  - Calculate the model image based on the catalogue
  - Compare the model image to the data and calculate likelihood
- Prior
  - Set priors for nuisance parameters like sky level
  - Set population distributions for source parameters
  - Set prior on the ***number*** of sources
    - Flat prior, no minimum flux: image will be overfit
    - Physical prior above some minimum flux
    - Parsimony prior to penalize unimportant sources

Transdimensional Proposals

- Birth
  - Propose new source, drawing from prior population distributions
- Death
  - Choose a source at random to remove



- Split
  - Choose a source at random to split
  - Split it into two, conserving flux and centre of flux and separate them according to some kick distribution
- Merge
  - For all pairs, use the value of the kick distribution for their separation as weights
  - Choose a pair using these weights and merge them, conserving flux and centre of mass

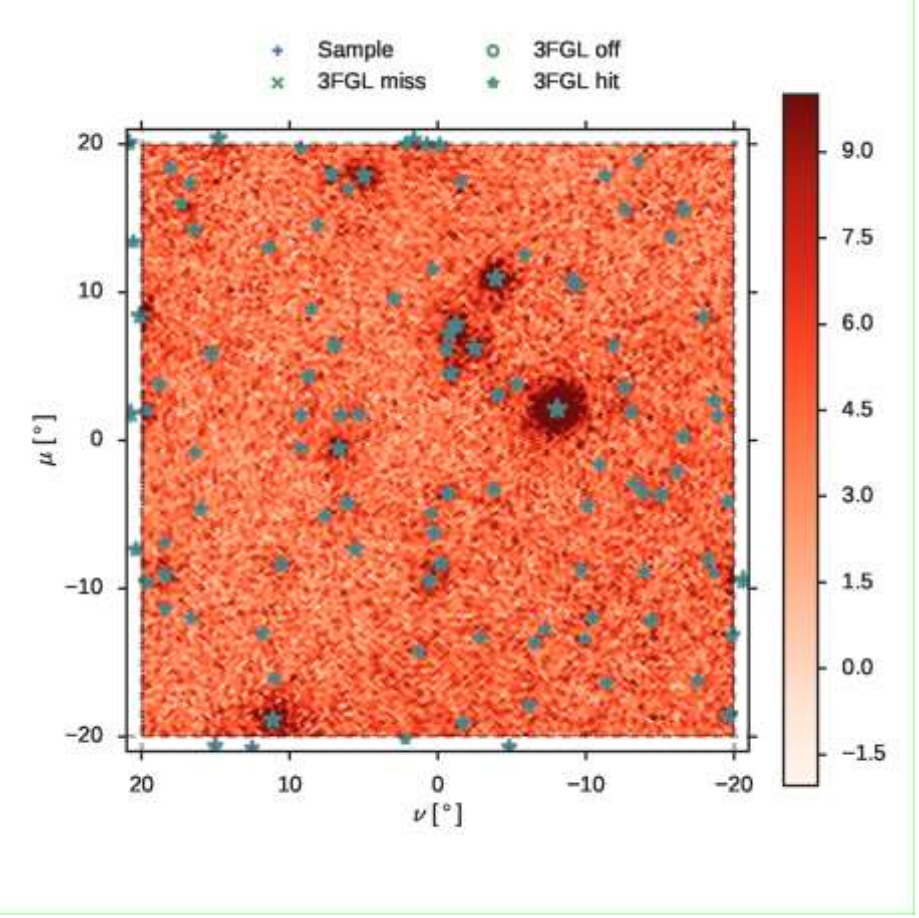


Application to faint source populations

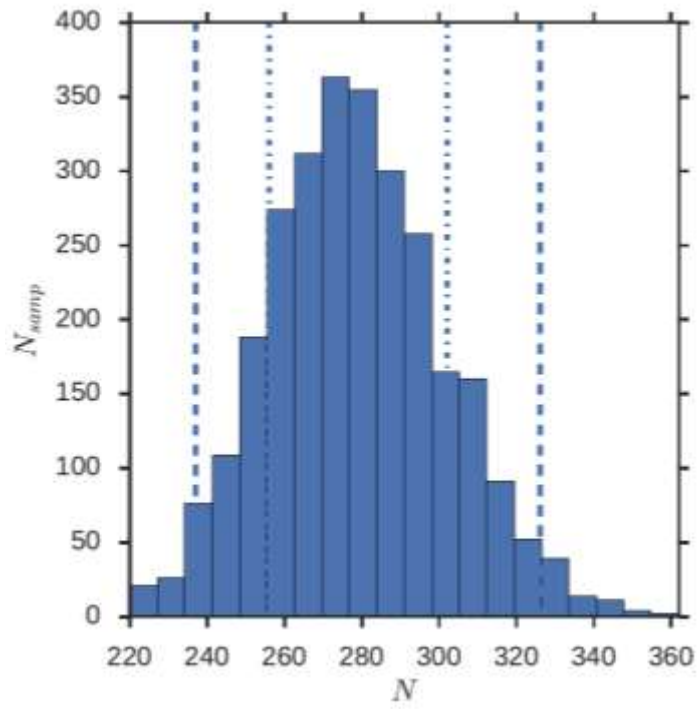
Daylan, **Portillo**, & Finkbeiner, 2017, ApJ, 839, 4

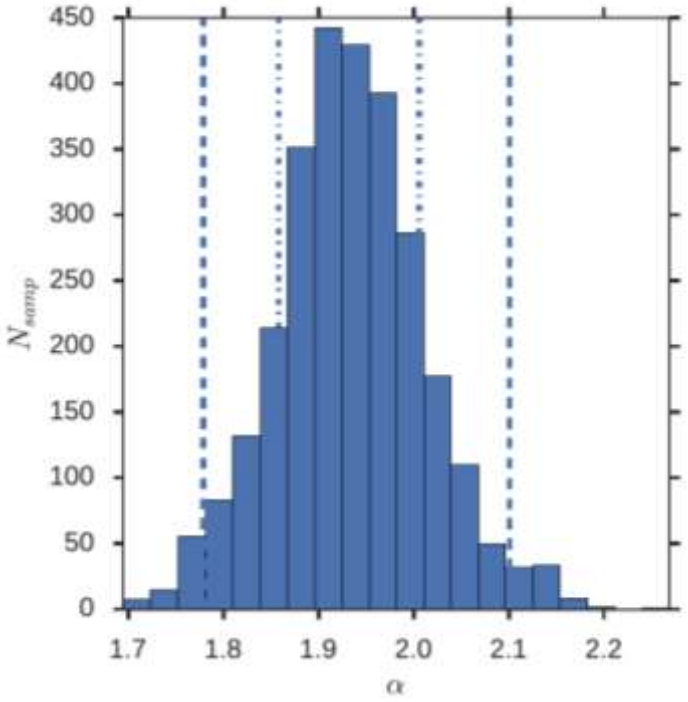
Image: Fermi

Probabilistic Catalogue

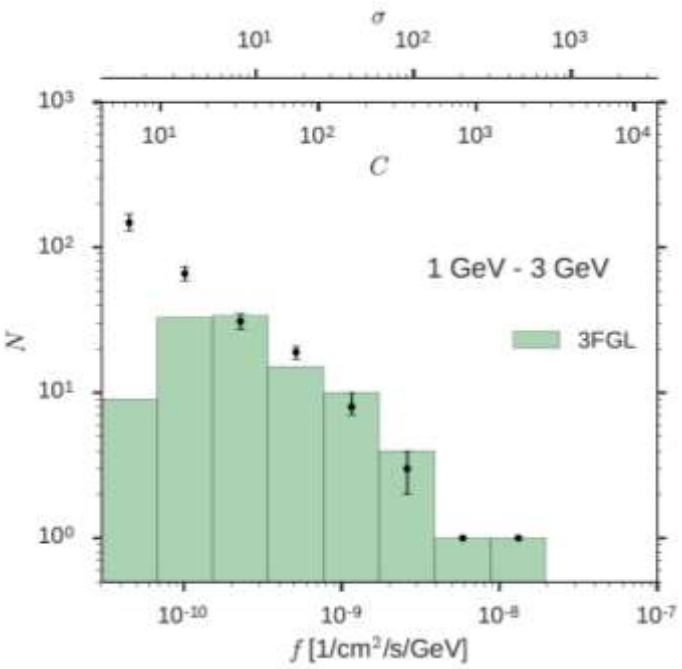


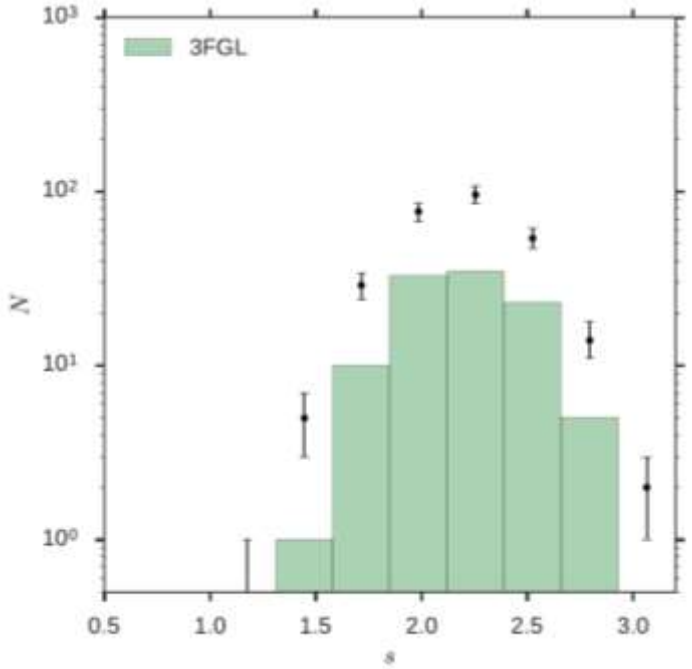
Population Constraints



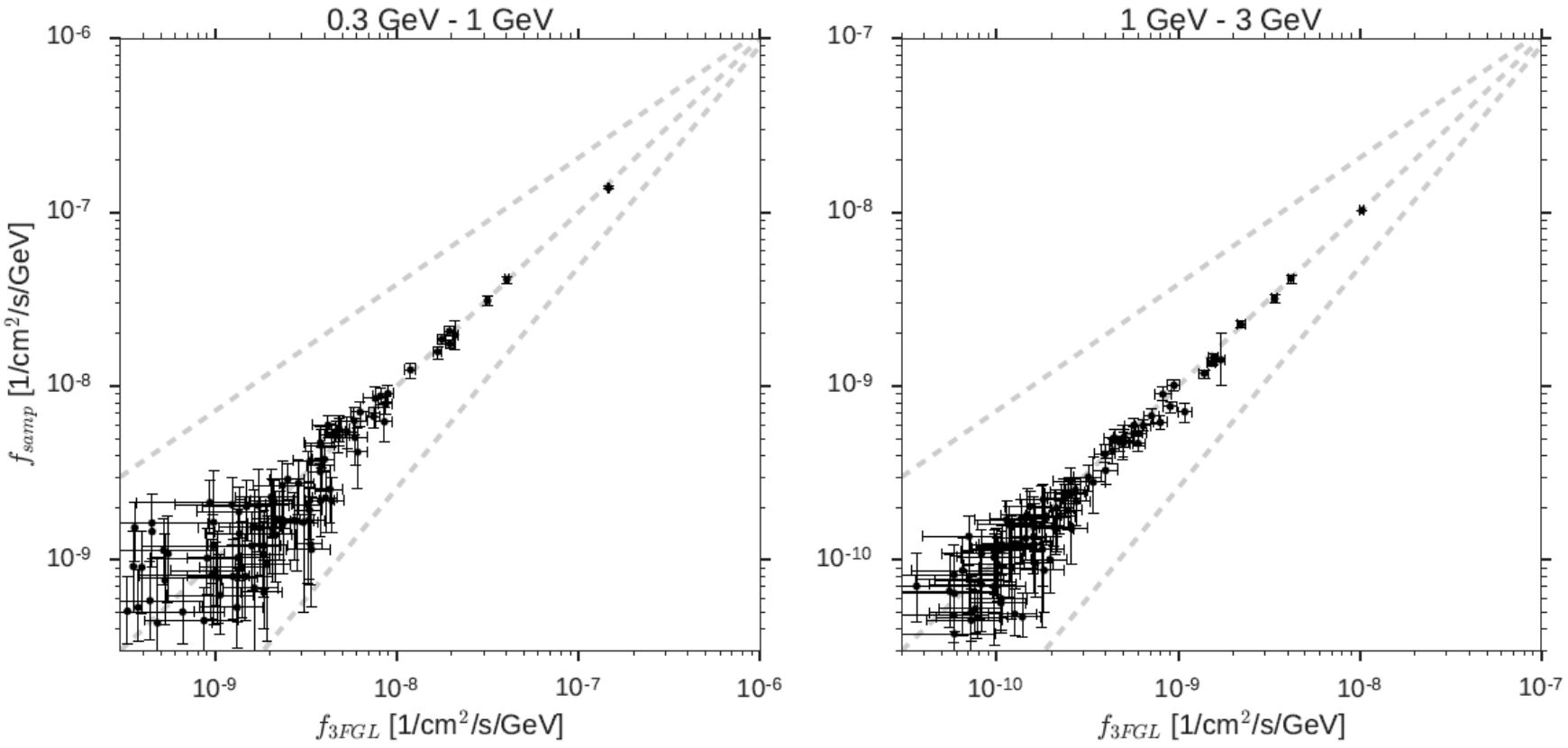


Flux and Colour Distributions





Comparison with Traditional Catalogue

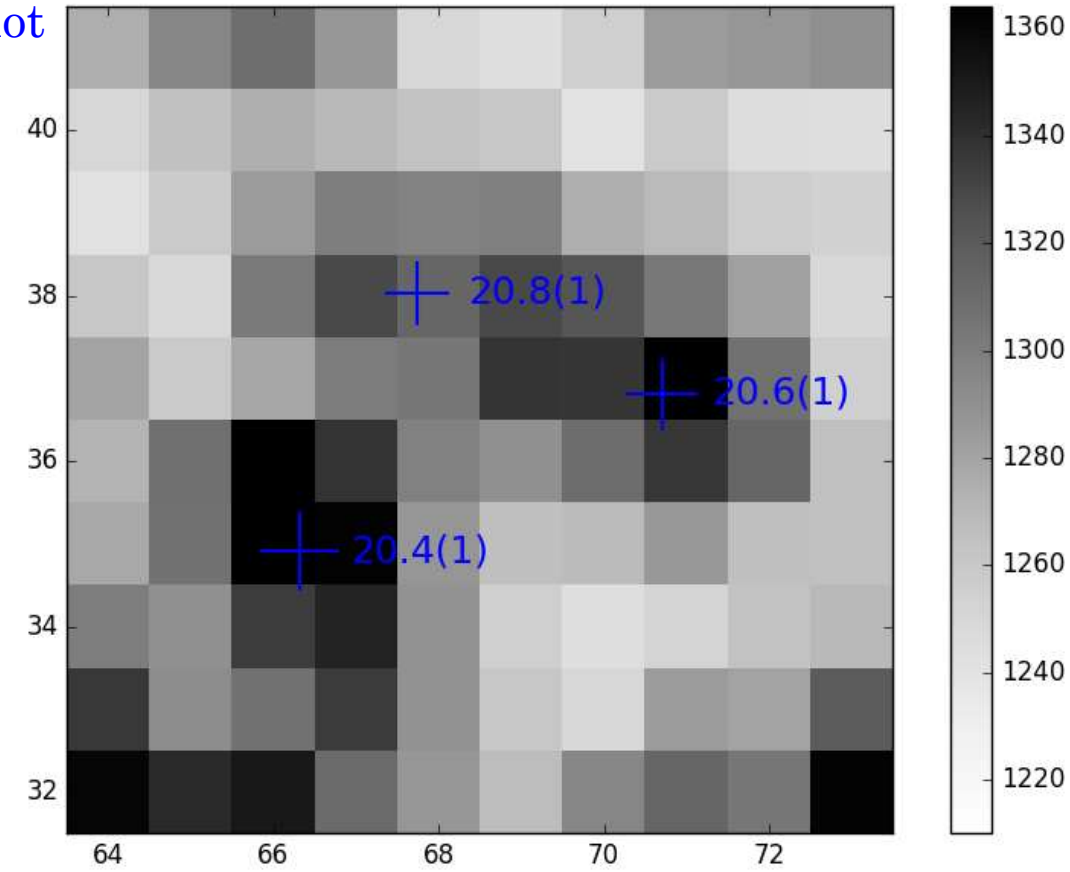


Application to crowded stellar fields

Portillo, Lee, Daylan, & Finkbeiner, 2017, AJ, 154, 4

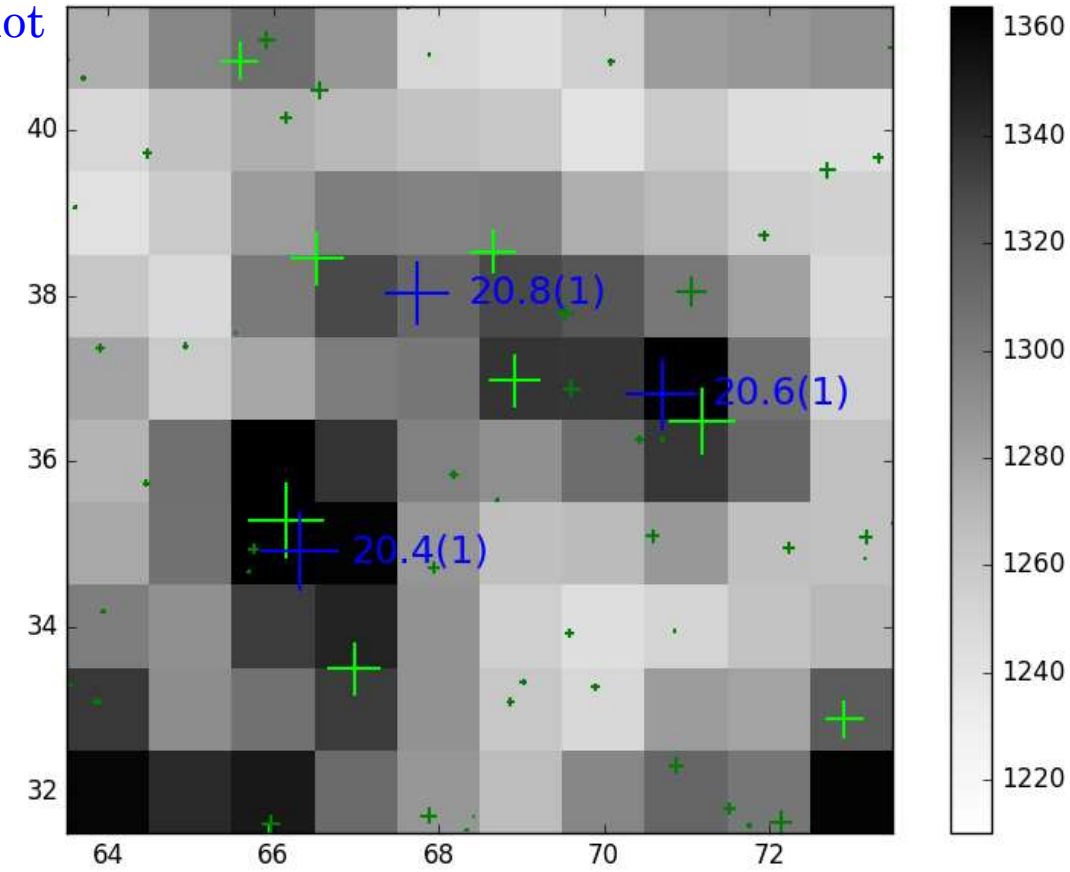
Traditional Catalogue

DAOPhot



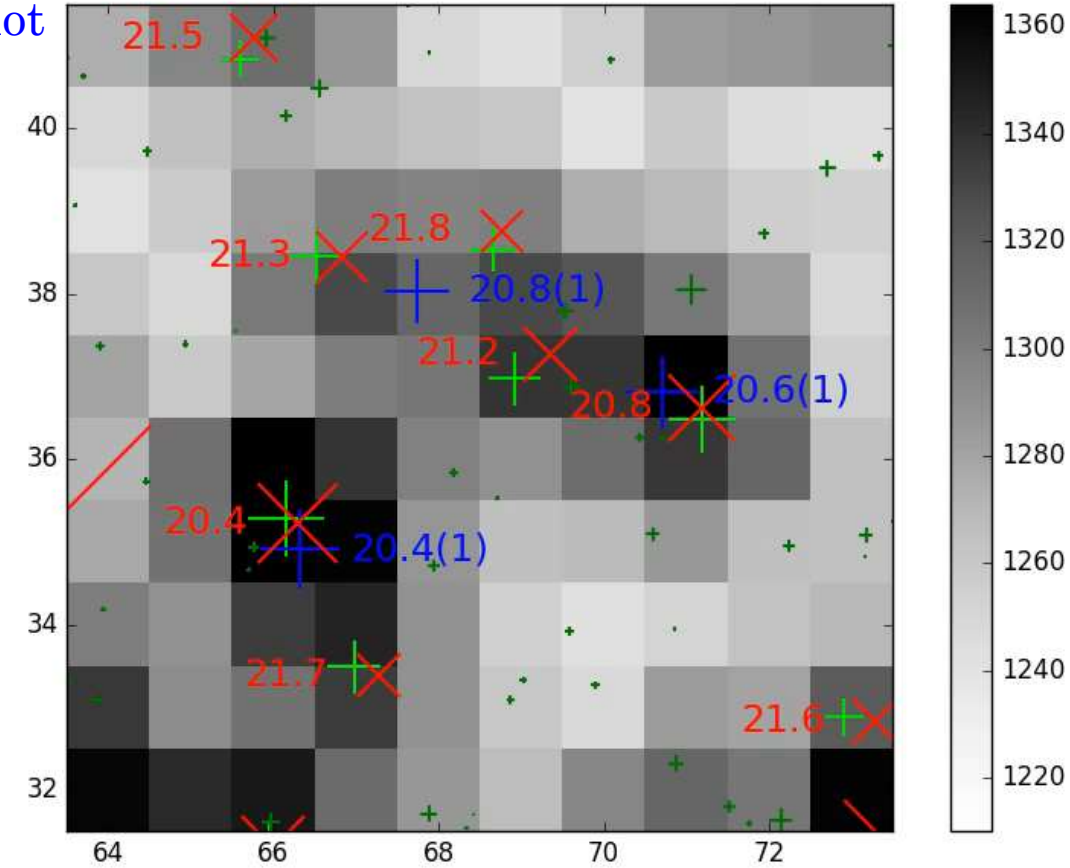
Compared to Hubble

DAOPhot
HST

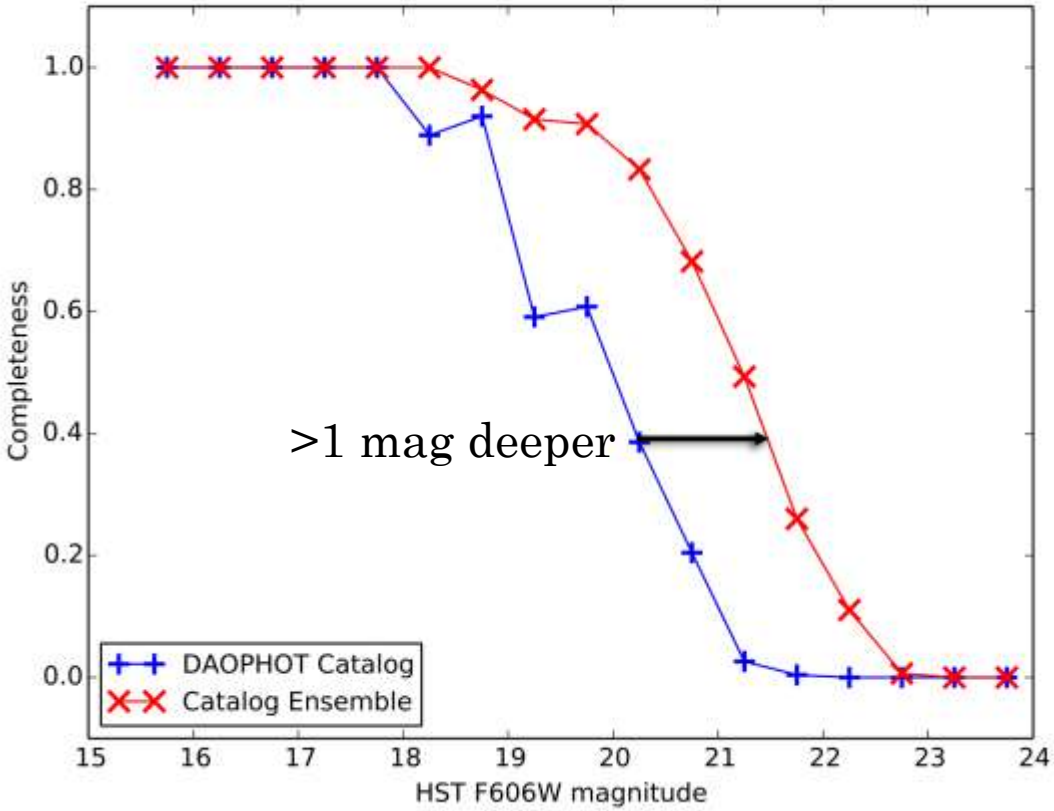


Probabilistic Catalogue

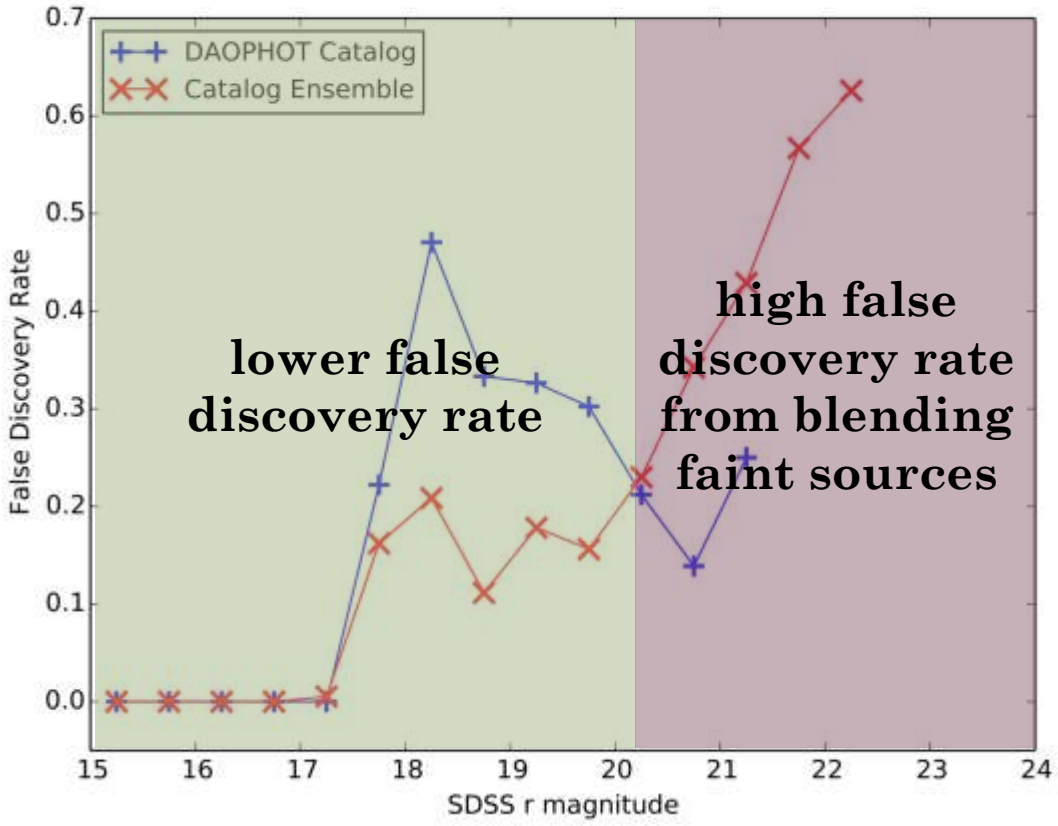
DAOPhot
HST
PCat



Completeness

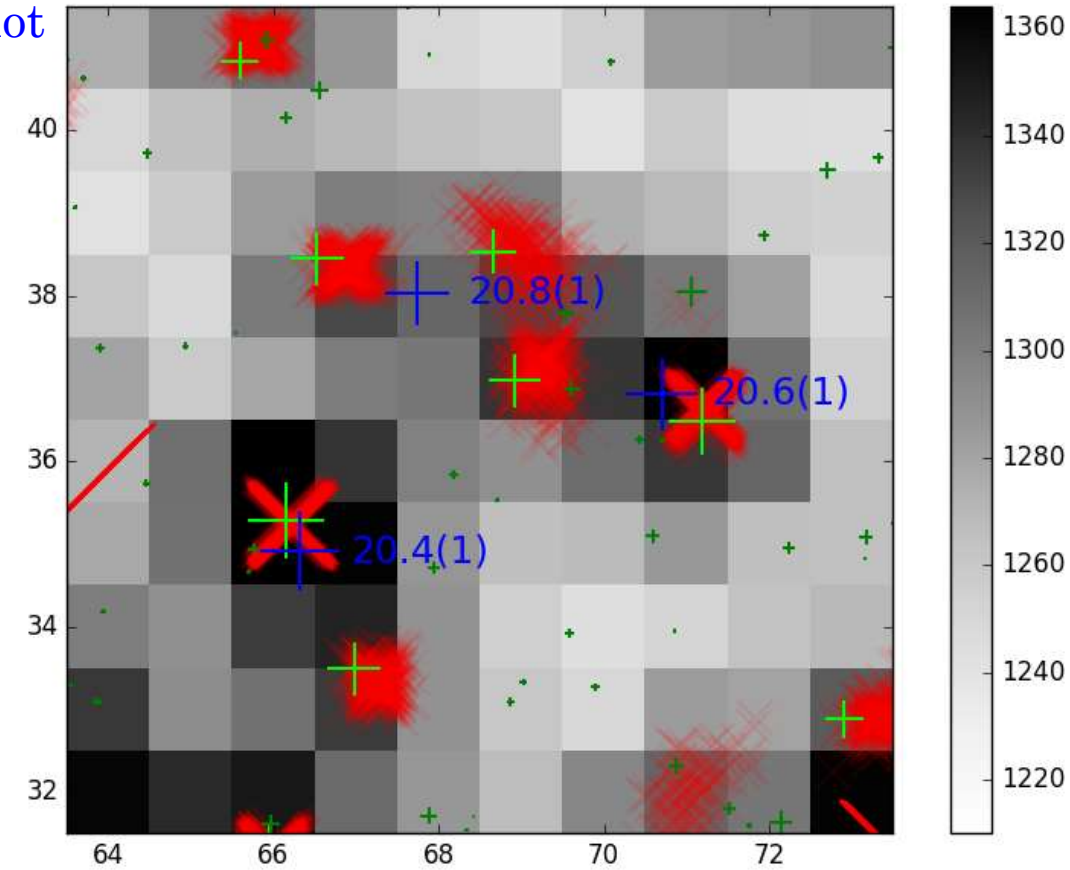


False Discovery Rate



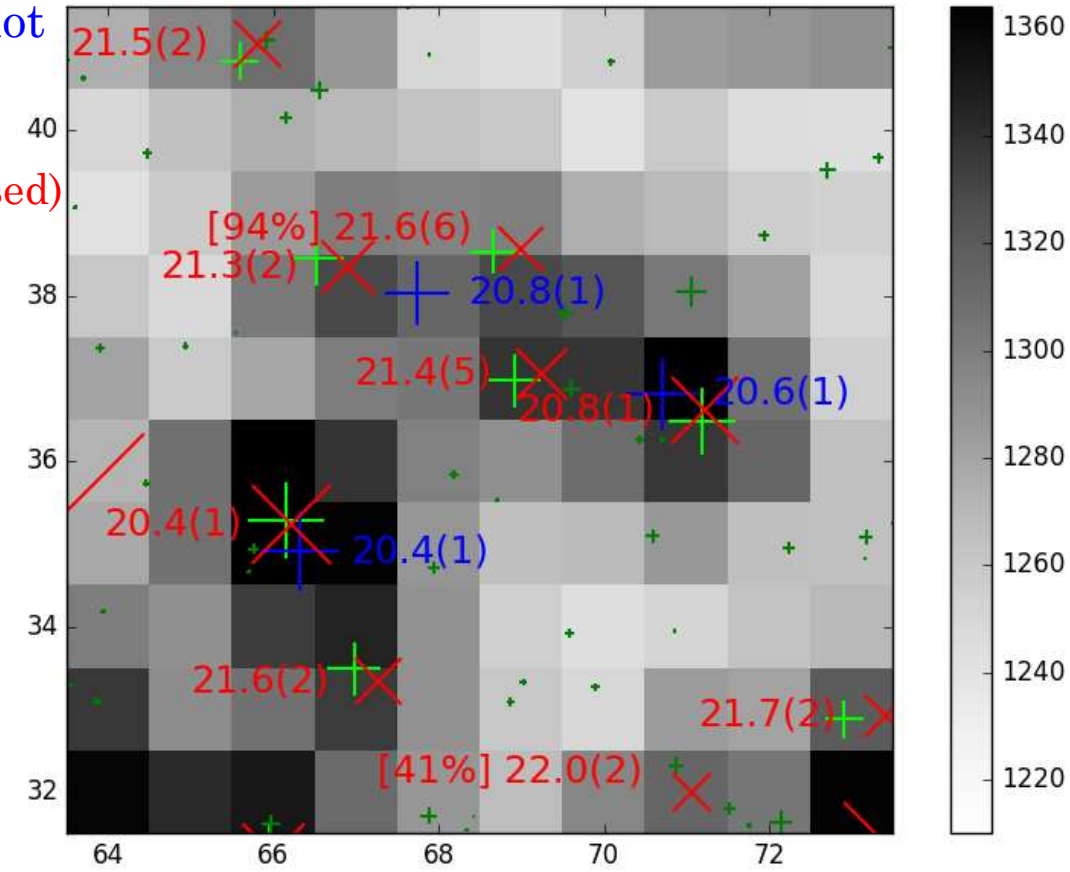
Stacked Catalogue Ensemble

DAOPhot
HST
PCat

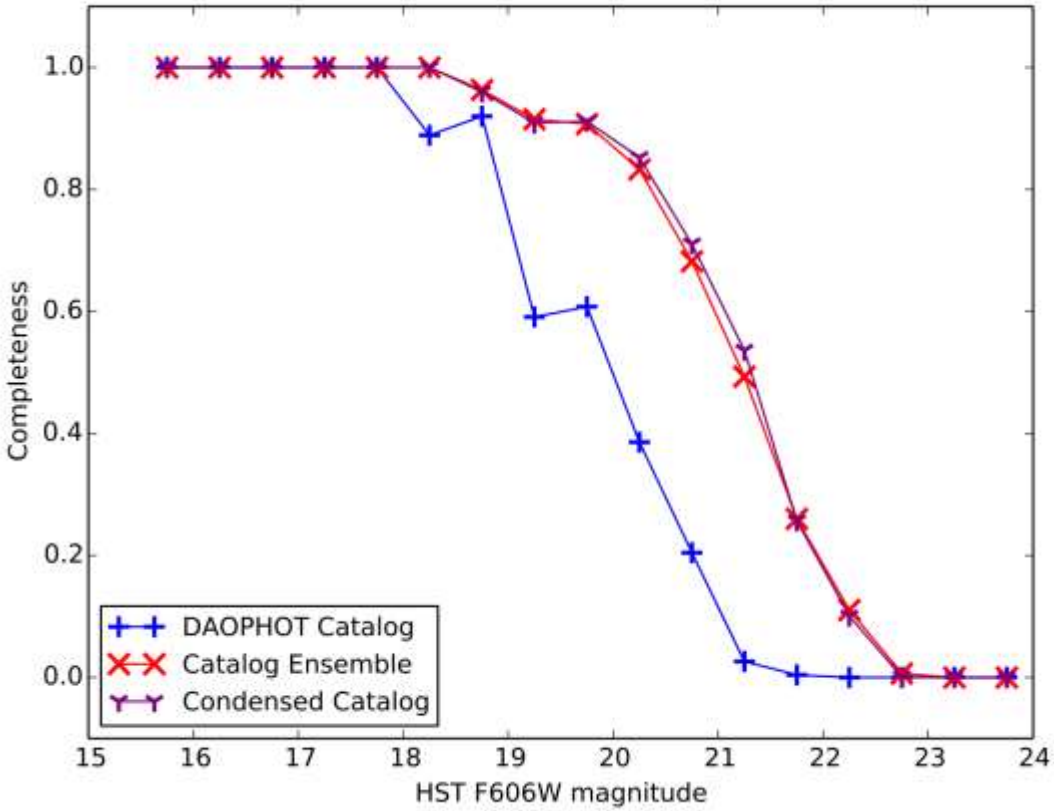


Condensed Catalogue

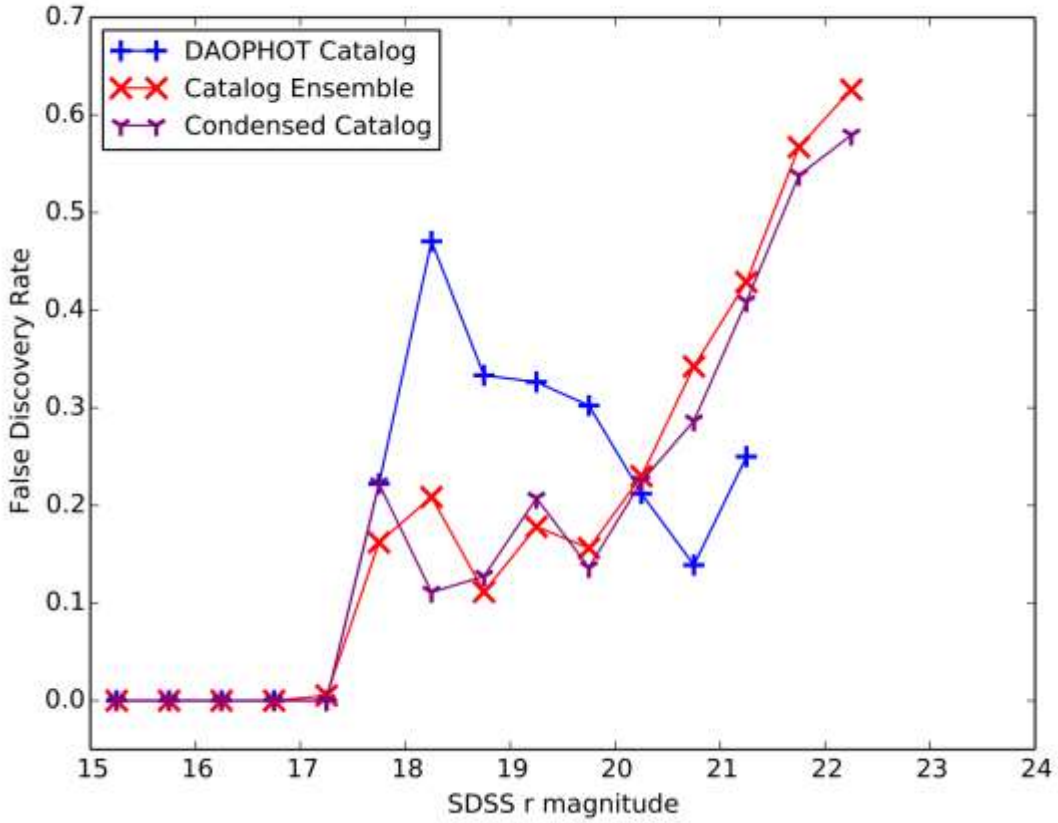
DAOPhot
HST
PCat
(condensed)

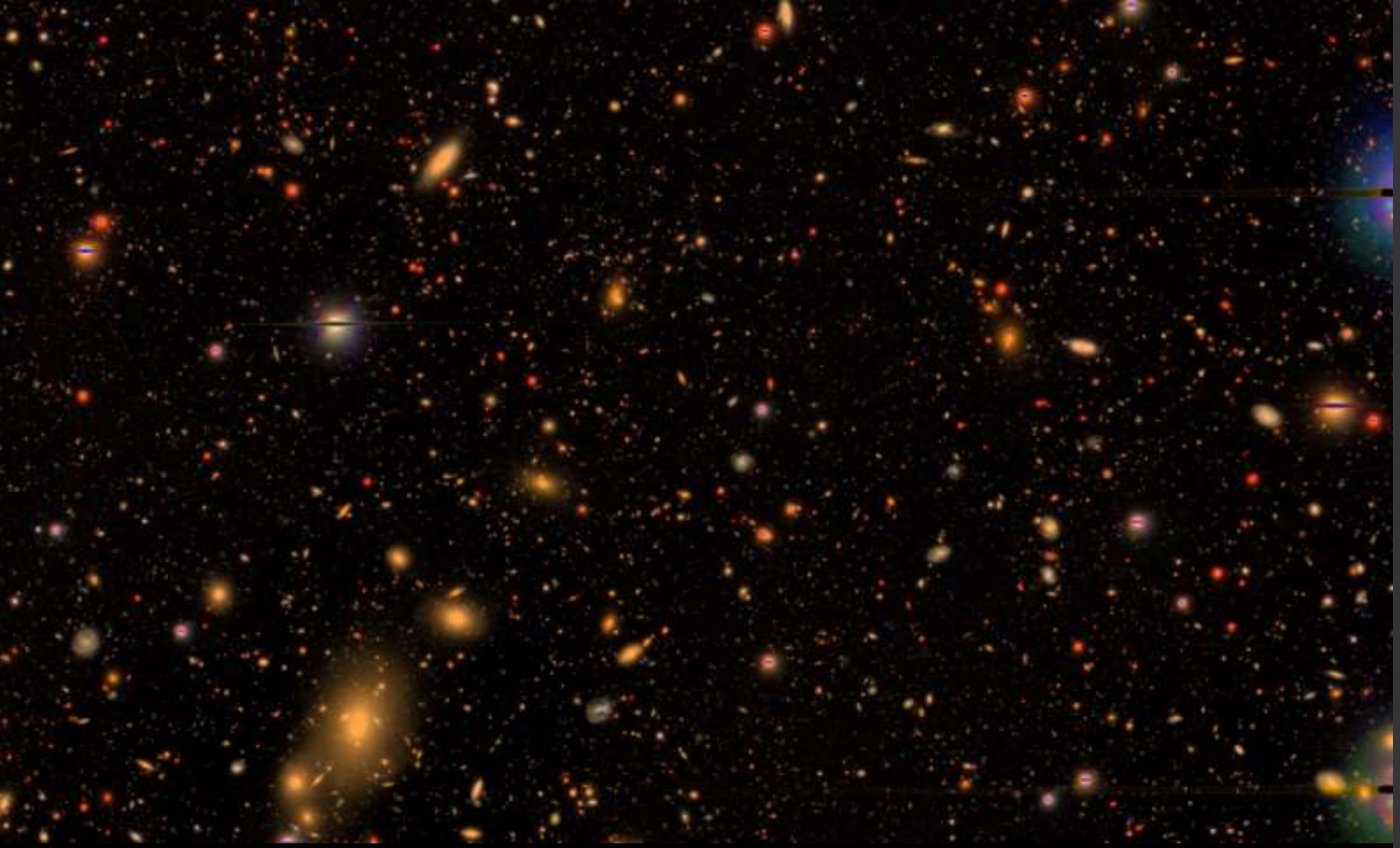


Completeness (Condensed Catalogue)



False Discovery Rate (Condensed Catalogue)





Extending probabilistic cataloging to star + galaxy fields

Image: Princeton/Hyper Suprime Cam Project

Bayesian Framework

- Likelihood
  - Calculate the model image based on the catalogue
  - Compare the model image to the data and calculate likelihood
- Prior
  - Set priors for nuisance parameters
  - Set population distributions for source parameters
  - Set prior on the number of sources
- Need to parameterize sources
- Galaxies are often fit with Sérsic profiles

$$I(r; I_e, n, r_e) = I_e \exp\left(-b_n\left[\left(\frac{r}{r_e}\right)^{1/n} - 1\right]\right)$$

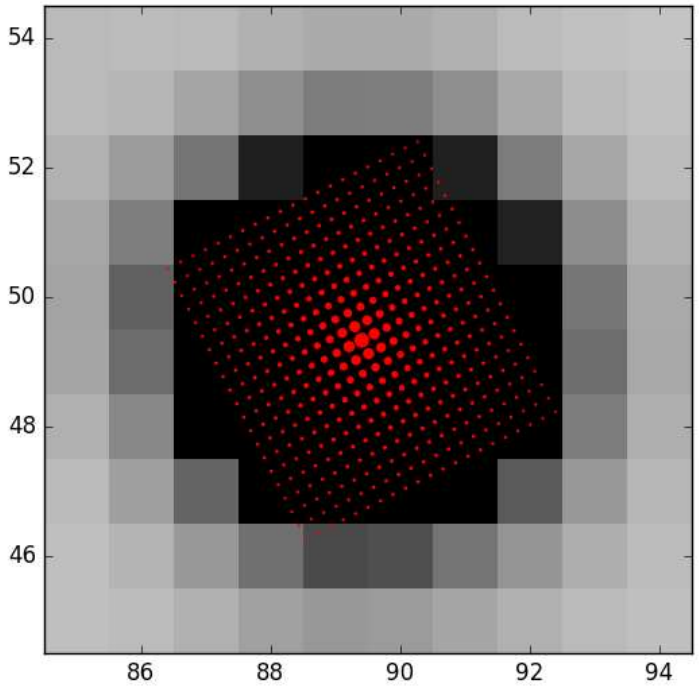
- Profile is approximate – most useful when detailed structure is not discernable

Transdimensional Proposals

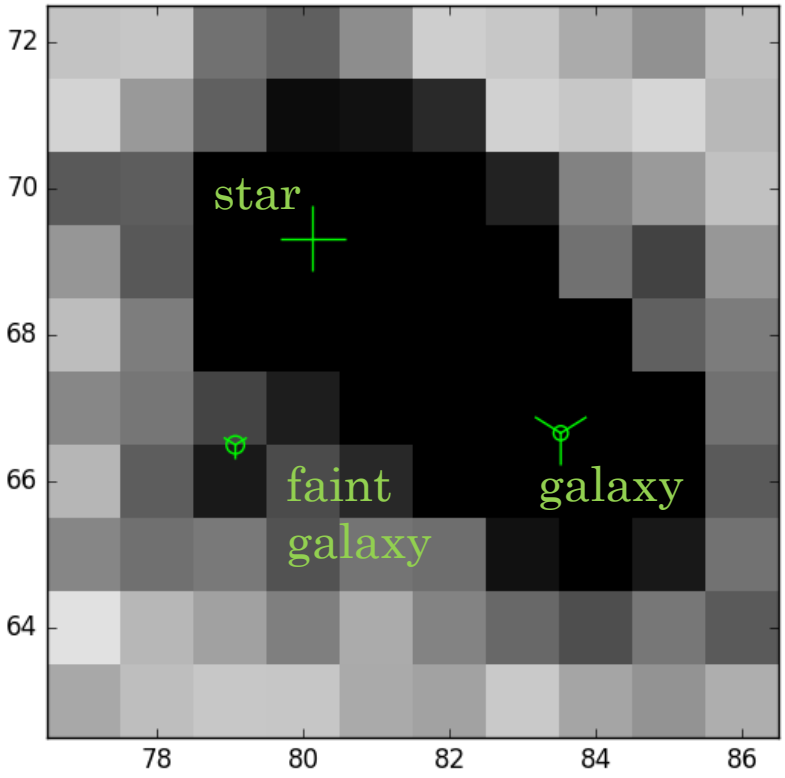
- Birth
  - Star
  - Galaxy
- Death
  - Star
  - Galaxy
- Star → Galaxy
- Galaxy → Star
- Split
  - Star → 2 Stars
  - Galaxy → 2 Galaxies
  - Galaxy → Galaxy + Star
- Merge
  - 2 Stars → Star
  - 2 Galaxies → Galaxy
  - Galaxy + Star → Galaxy

Galaxy Model Images

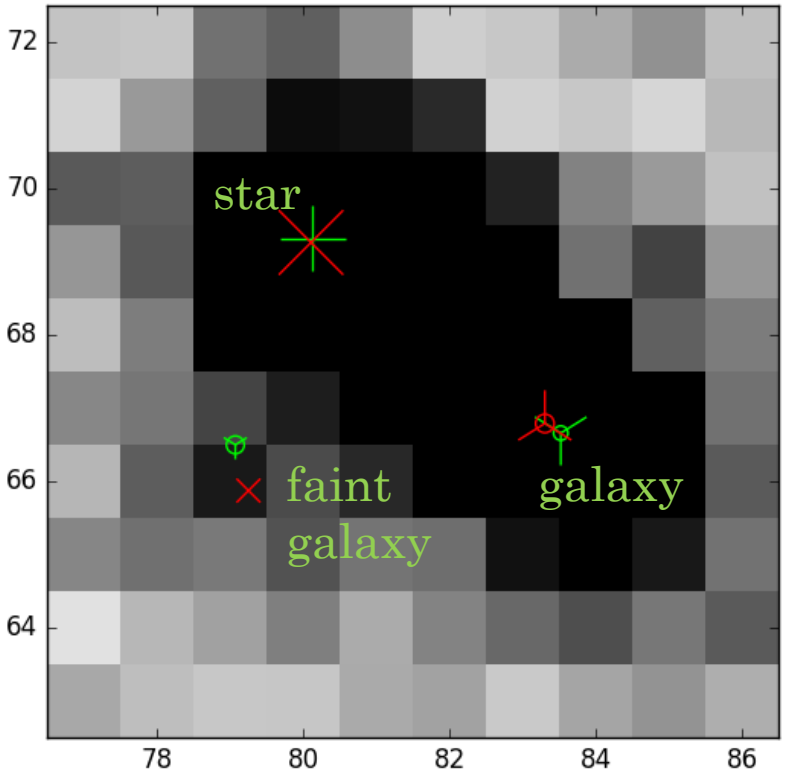
- Calculating model image is the slowest part for stars – even more so for galaxies
- We sped up our point source model image calculation
- So approximate galaxy profiles with a collection of point sources?



Deblending Stars + Galaxies



Deblending Stars + Galaxies



Conclusion

- Traditional cataloguing loses information on faint source populations and does not capture deblending ambiguities and source-source covariance
- Probabilistic cataloguing infers an ensemble of catalogues and captures the information lost by traditional catalogues
- Probabilistic cataloguing performs better in crowded stellar fields than traditional cataloguing
- The problem of crowded field photometry will be very relevant in the LSST era
- We are working on speed improvements and extending the method to star + galaxy fields