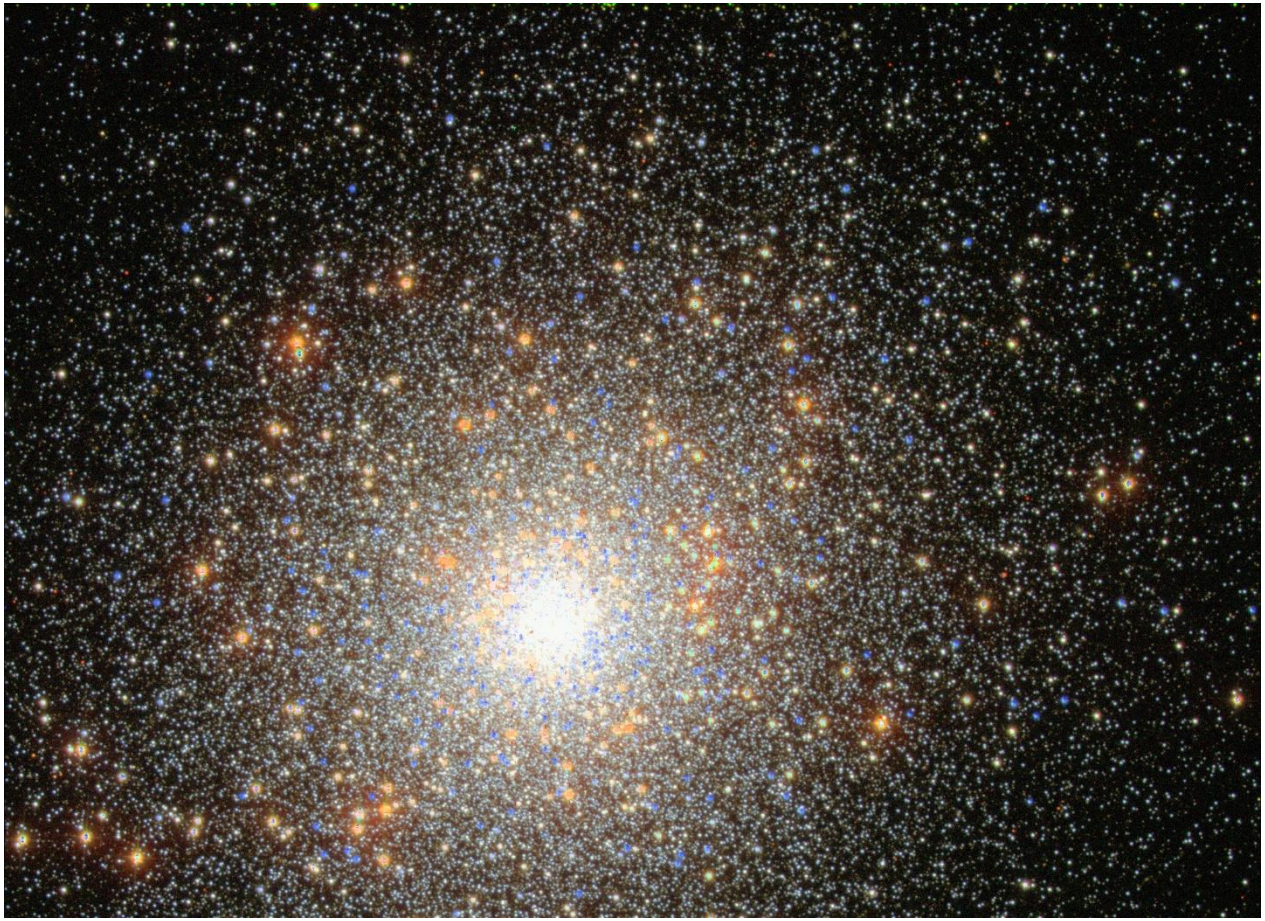
Probabilistic Cataloguing in Crowded Fields

Stephen K N Portillo
with Benjamin C G Lee, Tansu Daylan and
Douglas P Finkbeiner (arXiv:1703.01303)
13 June 2017
Getting ready to do science with LSST data, Lyon

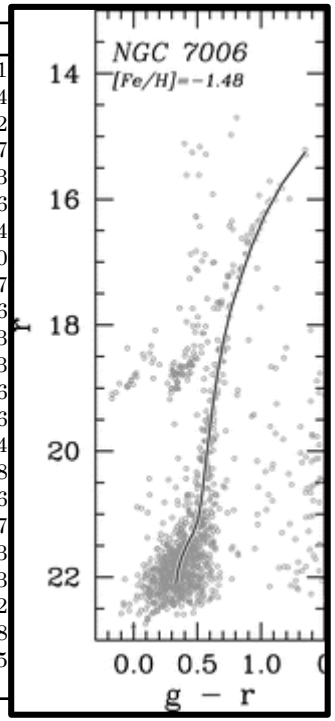
Telescopes don’t make catalogues!



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

People make catalogues

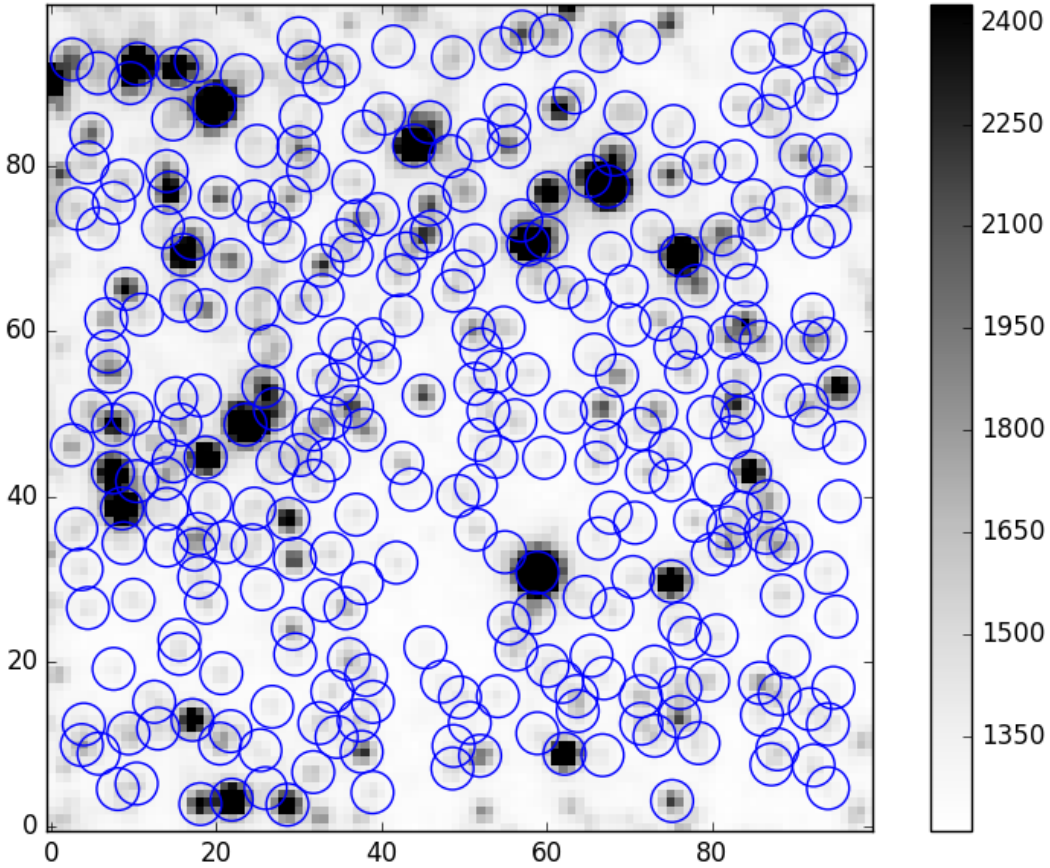
RA (J2000)	DEC (J2000)	g	r
229.4351	2.010923	19.344	19.1
229.4353	1.990166	23.070	21.4
229.4358	2.033374	21.809	21.2
229.4361	2.070269	20.107	19.7
229.4362	1.997957	22.894	21.3
229.4364	2.048578	22.386	21.6
229.4366	2.053515	20.853	20.4
229.4369	2.103516	21.827	21.0
229.4369	2.043476	23.067	21.7
229.4370	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

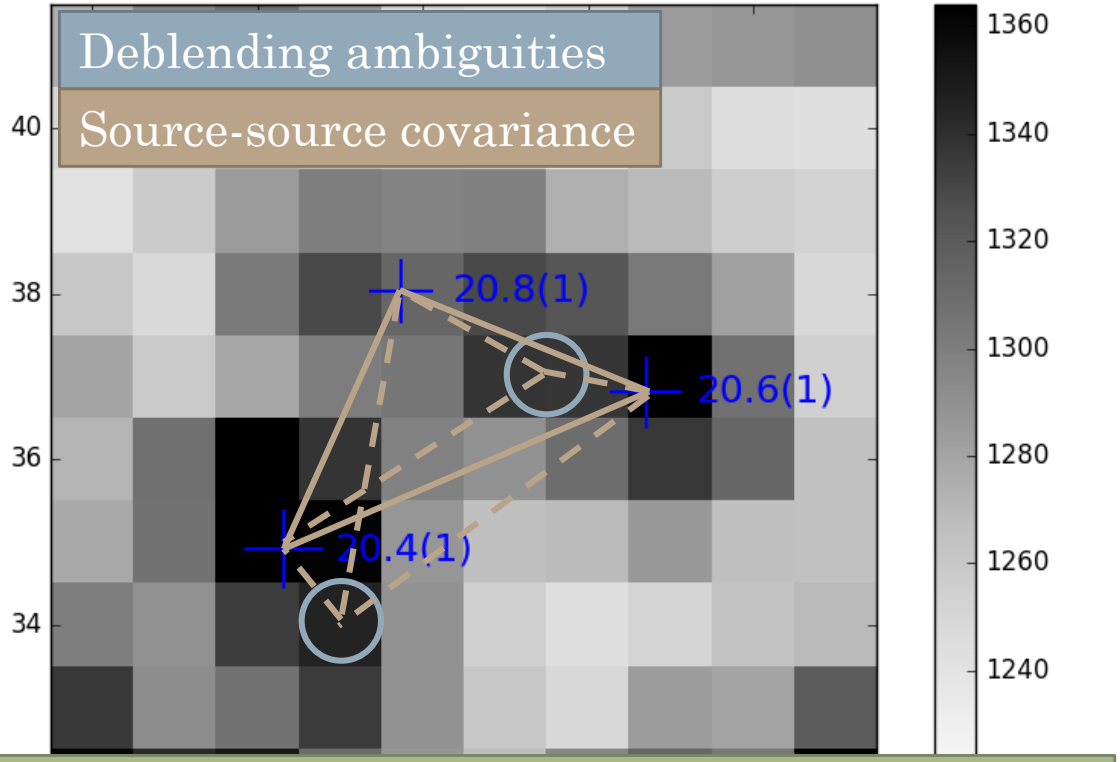
Catalogue and color-magnitude diagram: An et al. ApJS 179, 328 (2008)

Crowded Field Cataloguing



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

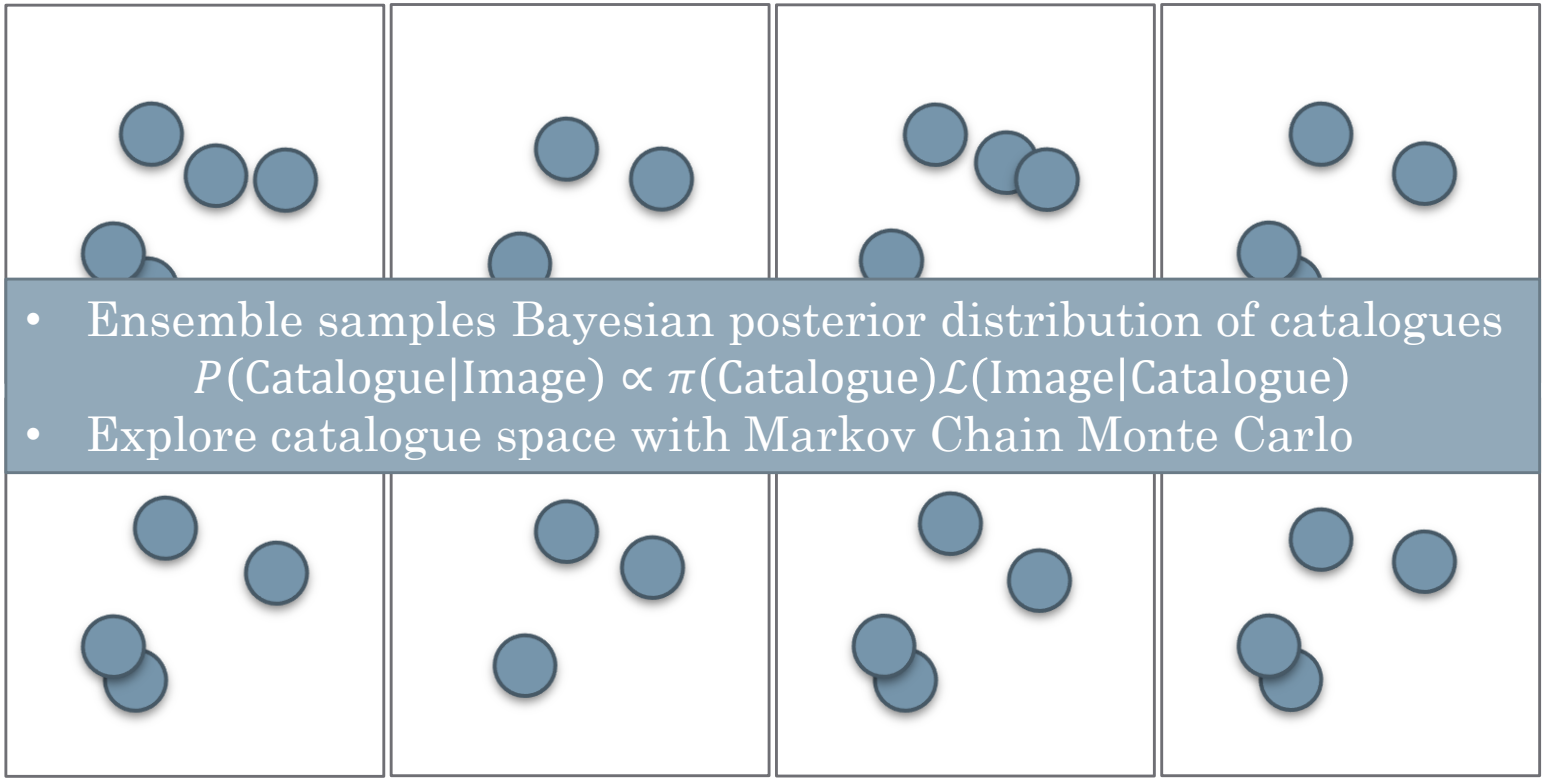
Traditional Catalogue Issues



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

Probabilistic Cataloguing

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



- 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

Catalogue Priors

- Prior that sources are independent and described by population parameters β:

$$\pi\left(\{\ell_i, b_i, F_i\}_{i=1}^{N}, \beta\right) = \pi(\beta)\pi(N|\beta)\prod_{i=1}^{N}\pi(\ell_i, b_i, F_i|\beta)$$

- β can describe both spatial and flux distributions
- What should the prior on the number of sources look like? What do we mean by "the number of sources"?

How many sources are there with a flux above F_{min}?

- A prior that penalizes the $(N+1)^{\text{th}}$ source based on the expected improvement in χ^2 under the null hypothesis that there are N sources:

$$\log\frac{\pi(N+1)}{\pi(N)} = -\frac{3}{2}$$

Reversible Jump MCMC

- Allows proposals to change dimensionality of model
  - Move m takes x and generates auxillary u to propose x'
  - Move m' takes x' and generates auxillary u' to propose x
  - $\dim x + \dim u = \dim x' + \dim u'$ and $(x, u) \leftrightarrow (x', u')$ one-to-one

$$\alpha(x \to x') = \min\left(1, \frac{\pi(x')}{\pi(x)} \frac{\mathcal{L}(x'|D)}{\mathcal{L}(x|D)} \frac{j_{m'}(x')}{j_m(x)} \frac{g(u')}{g(u)} \left|\frac{\partial(x', u')}{\partial(x, u)}\right|\right)$$

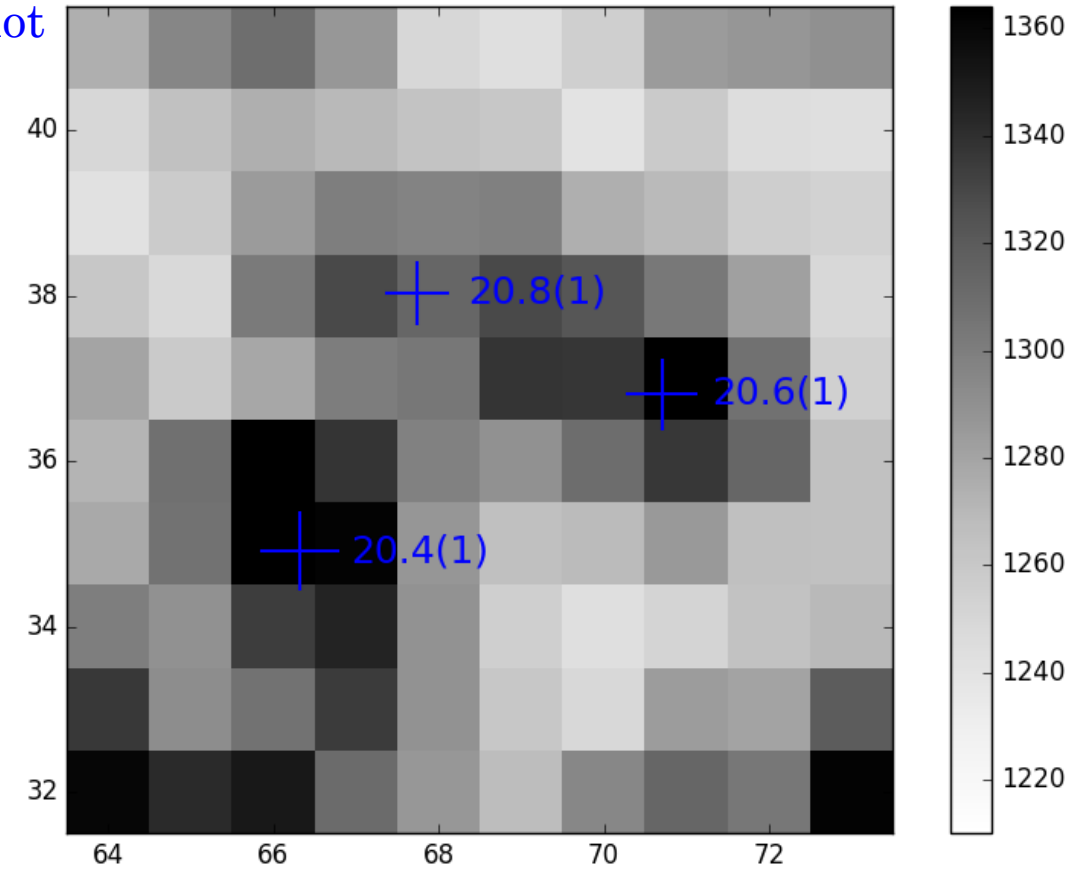
- For example, birth/death between $x = \{x_1, \ldots, x_N\}$ and $x' = \{x_1, \ldots, x_{N+1}\}$ has $u = x_{N+1}$ and $u' = \emptyset$
  - If birth and death equally likely, sources independent in prior and new source x_2 generated from prior

$$\alpha(x \to x') = \min\left(1, \frac{\pi(N+1)}{\pi(N)} \frac{\mathcal{L}(x'|D)}{\mathcal{L}(x|D)}\right)$$

P.J. Green, Biometrika 82, 711 (1995)

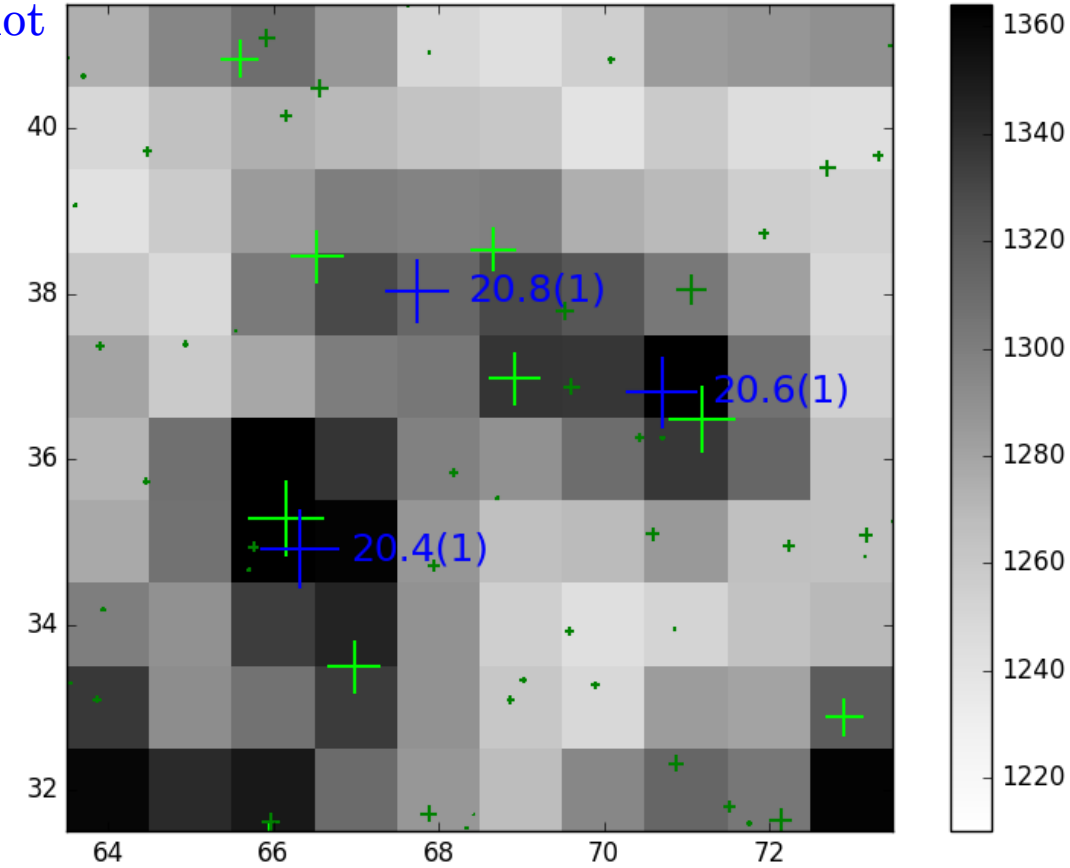
Traditional Catalogue

DAOPhot



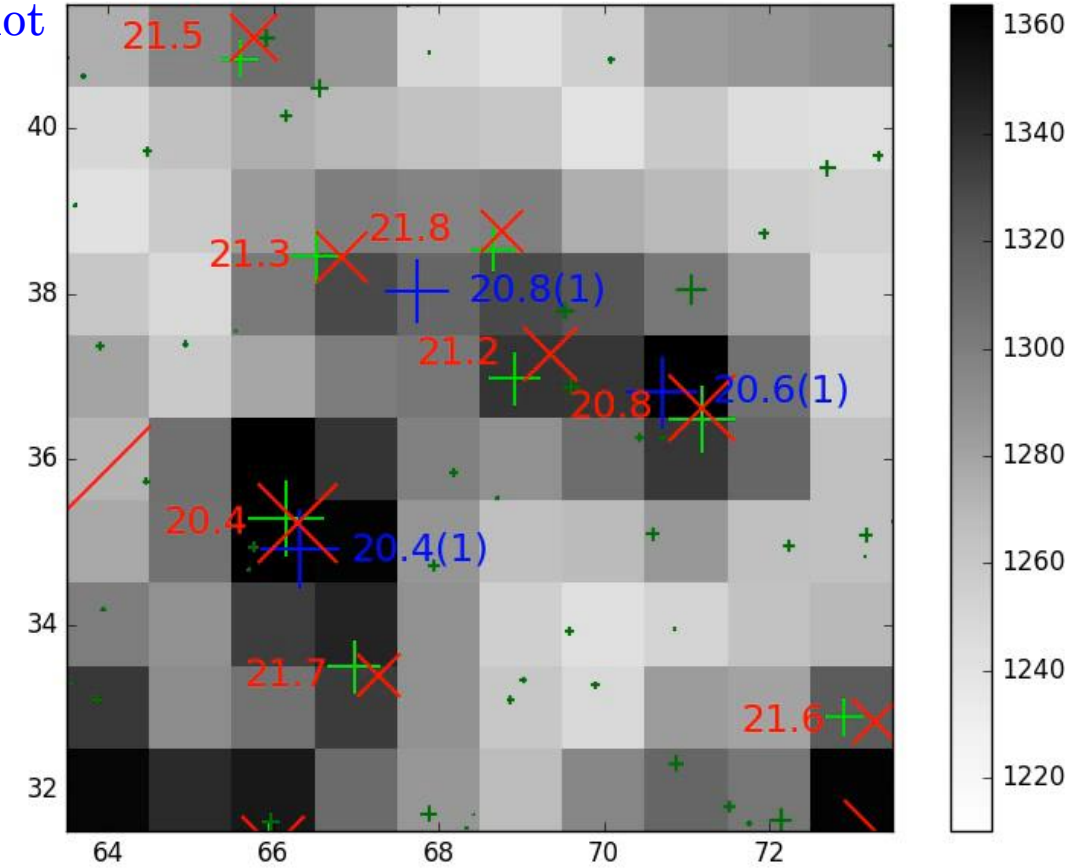
Compared to Hubble

DAOPhot
HST

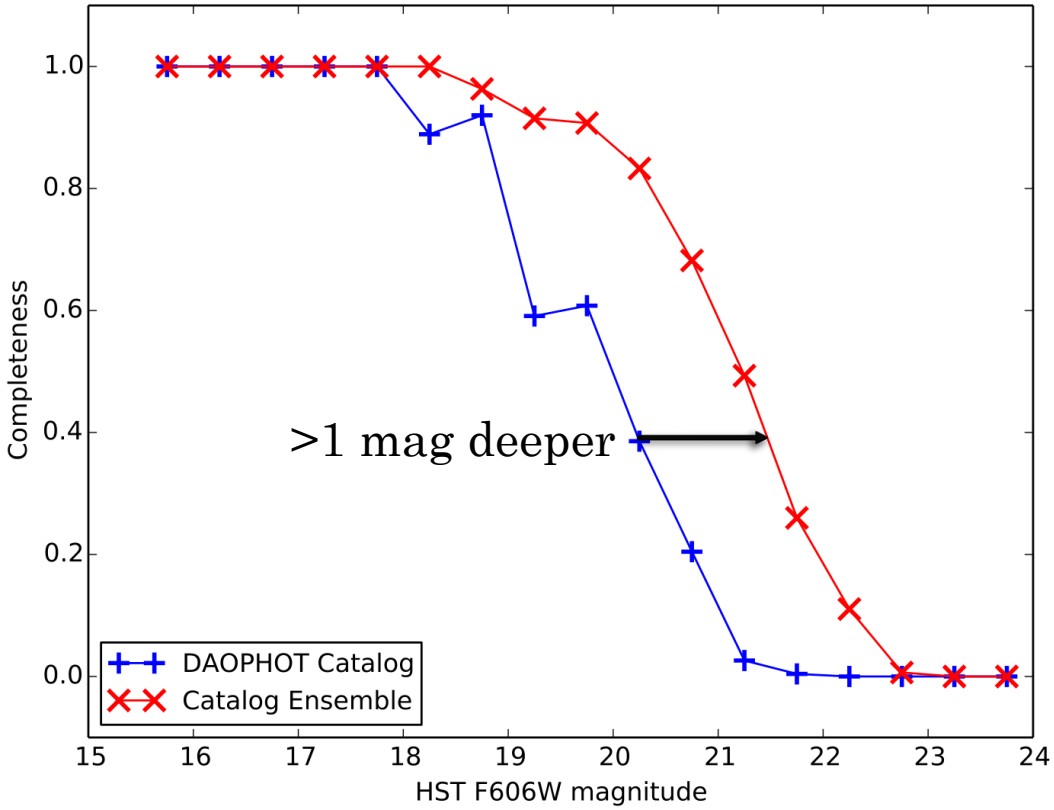


Probabilistic Catalogue

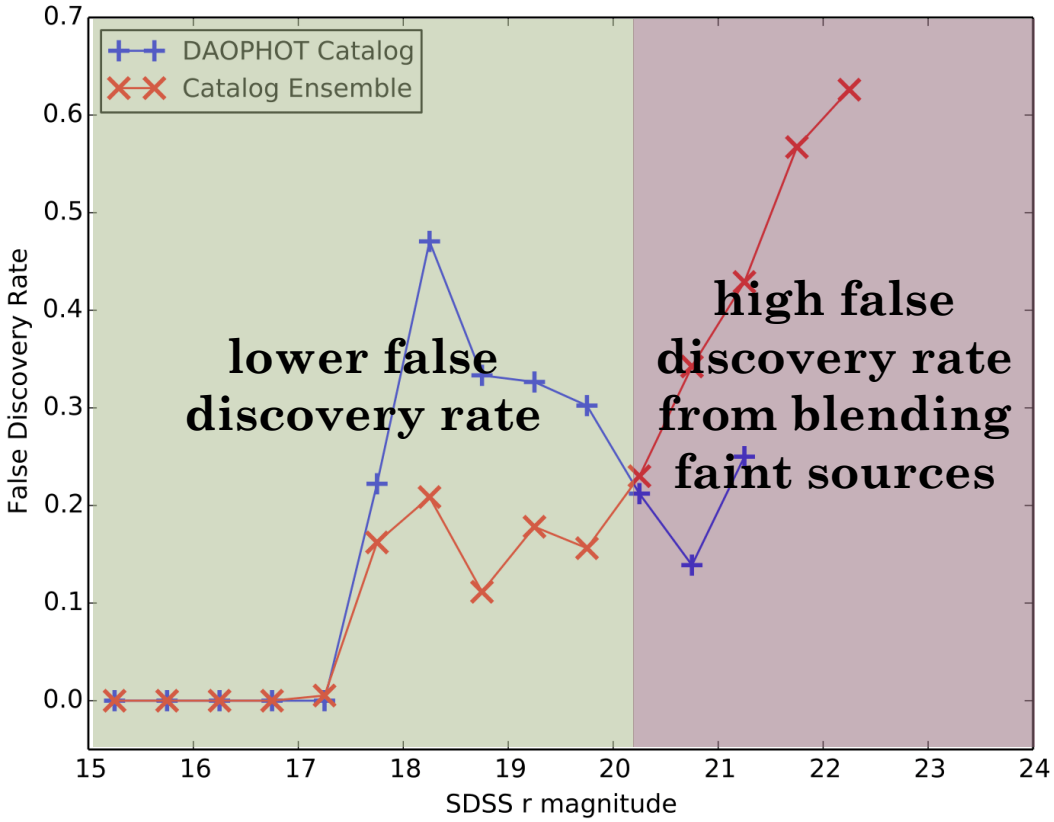
DAOPhot
HST
PCat



Completeness

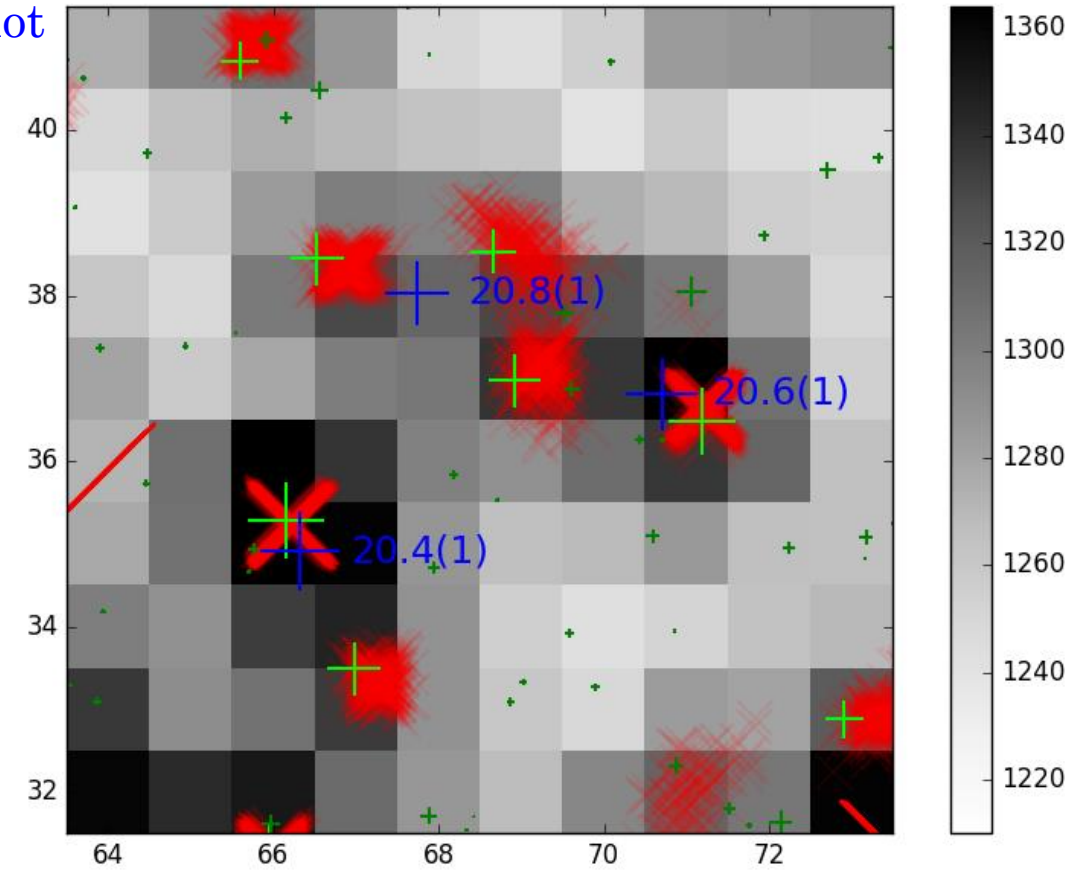


False Discovery Rate



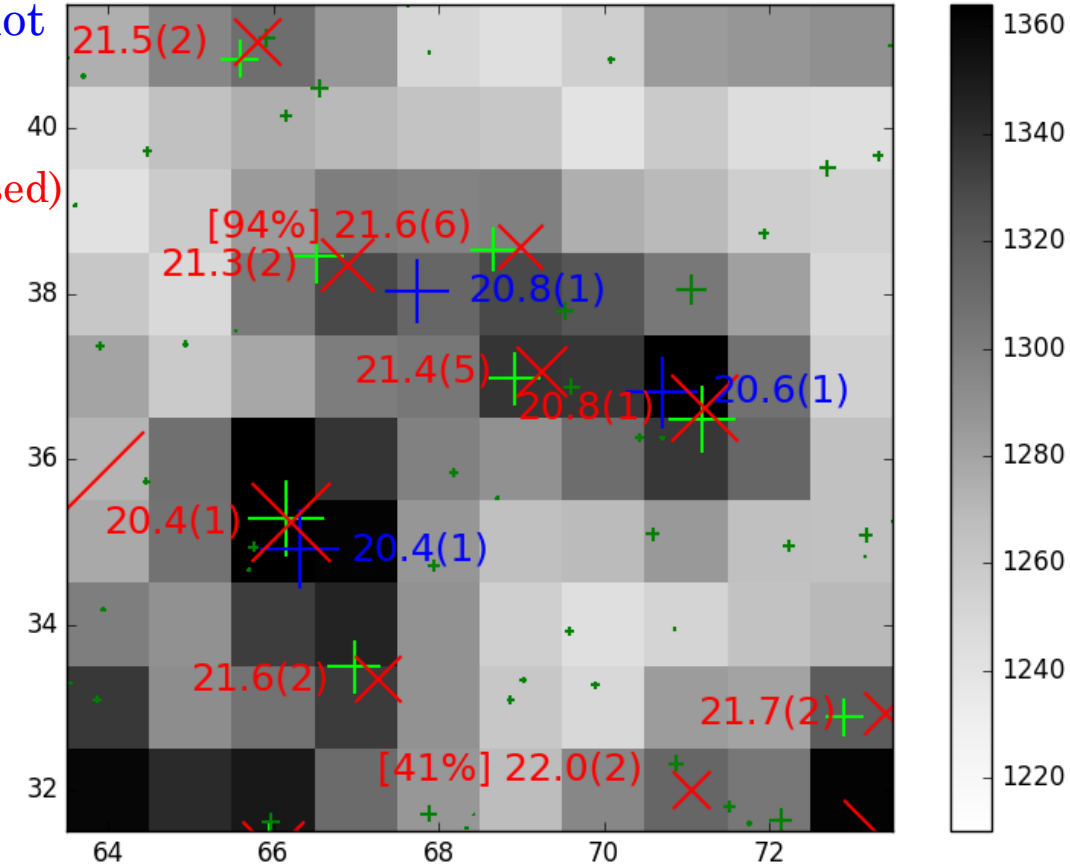
Stacked Catalogue Ensemble

DAOPhot
HST
PCat



Condensed Catalogue

DAOPhot
HST
PCat
(condensed)



Conclusion

- The problem of crowded field photometry will be very relevant in the LSST era
- Probabilistic cataloguing infers an ensemble of catalogues, capturing deblending ambiguities and source-source covariance
- The ensemble of catalogues can be collapsed into a condensed catalogue that is *marginalized over deblending ambiguities and source-source covariance*
- Our probabilistic cataloguing implementation finds more sources in crowded fields than DAOPHOT
- While the current implementation is very slow, we expect to be able to speed it up considerably

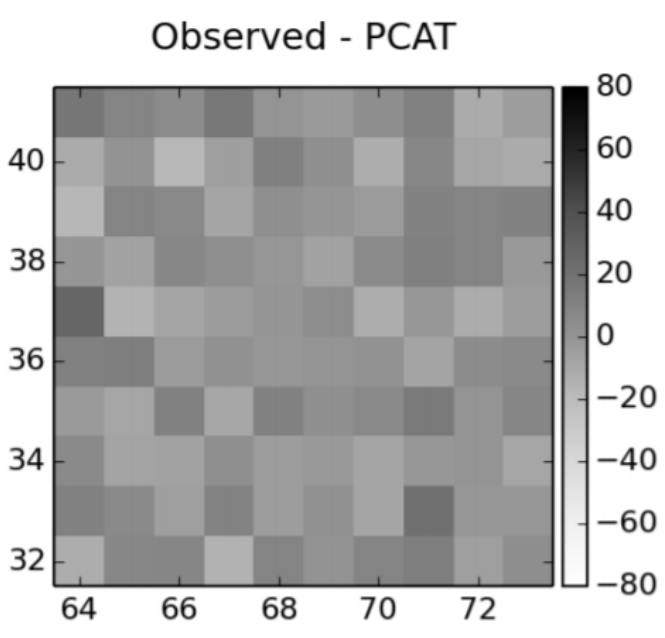
For more details see arXiv:1703.01303

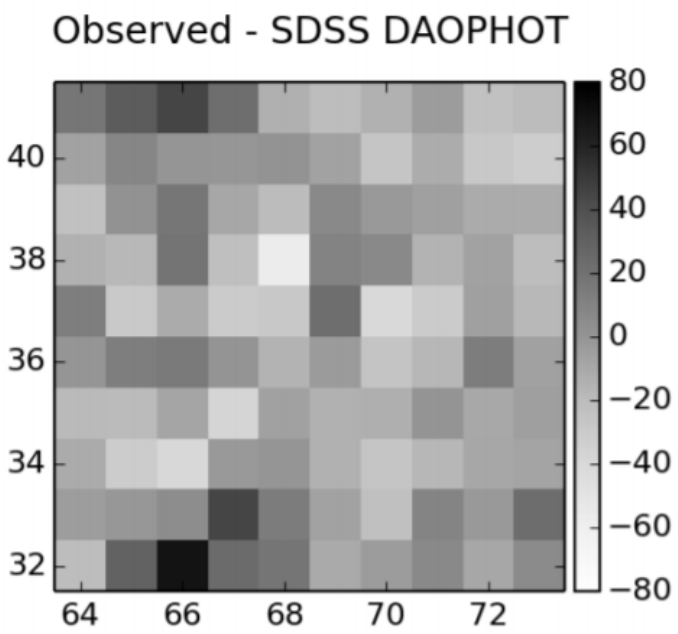
Backup Slides

Use Case Specifications

- $40'' \times 40''$ from Messier 2 ($N_{pix} = 10\,000$), 2’ from centre
- Core radius 0.34’, half-light radius 1.08’
- DAOPhot catalogue identifies 337 DAOPhot sources
- HST catalogue identifies 1 000 sources
- Run with about 250 CPU-hours (10^9 model evaluations @ 1 CPU-ms each)

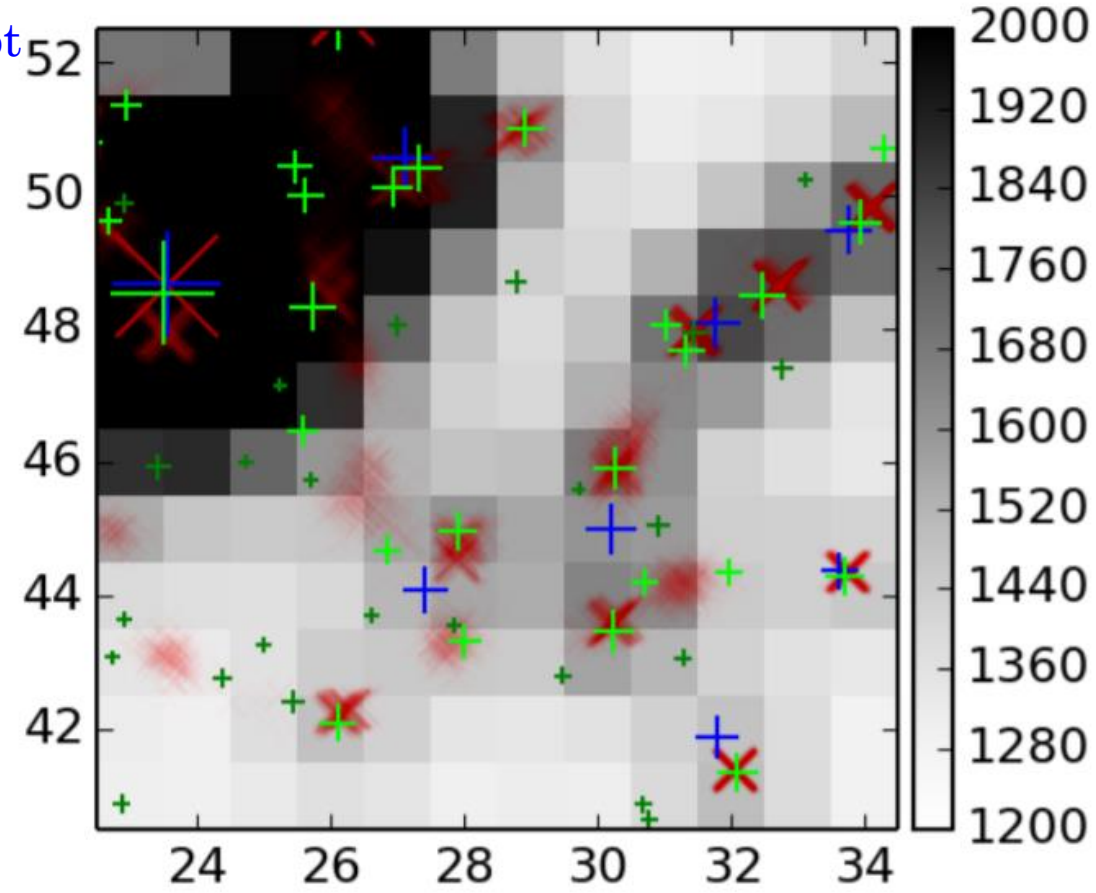
Residuals


Observed - PCAT
40
38
36
34
32
64
66
68
70
72
80
60
40
20
0
−20
−40
−60
−80

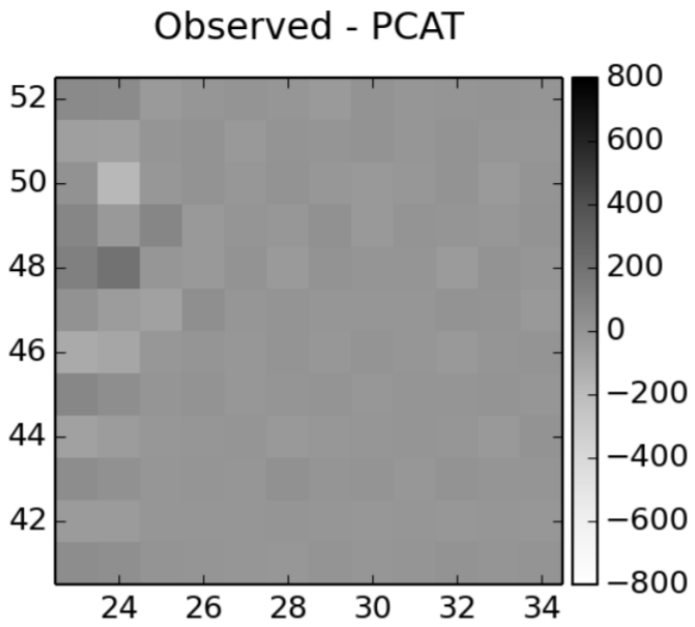

Observed - SDSS DAOPHOT
40
38
36
34
32
64
66
68
70
72
80
60
40
20
0
−20
−40
−60
−80

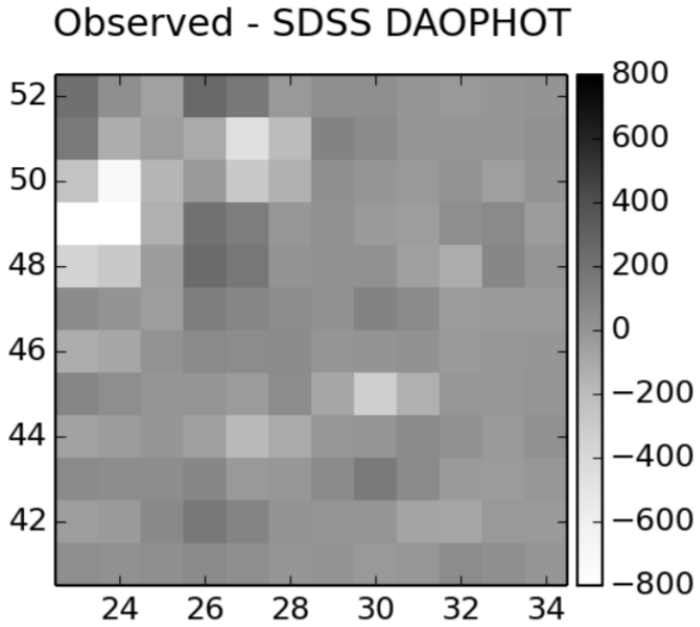
Worst-Case Scenario

DAOPhot
HST
PCat

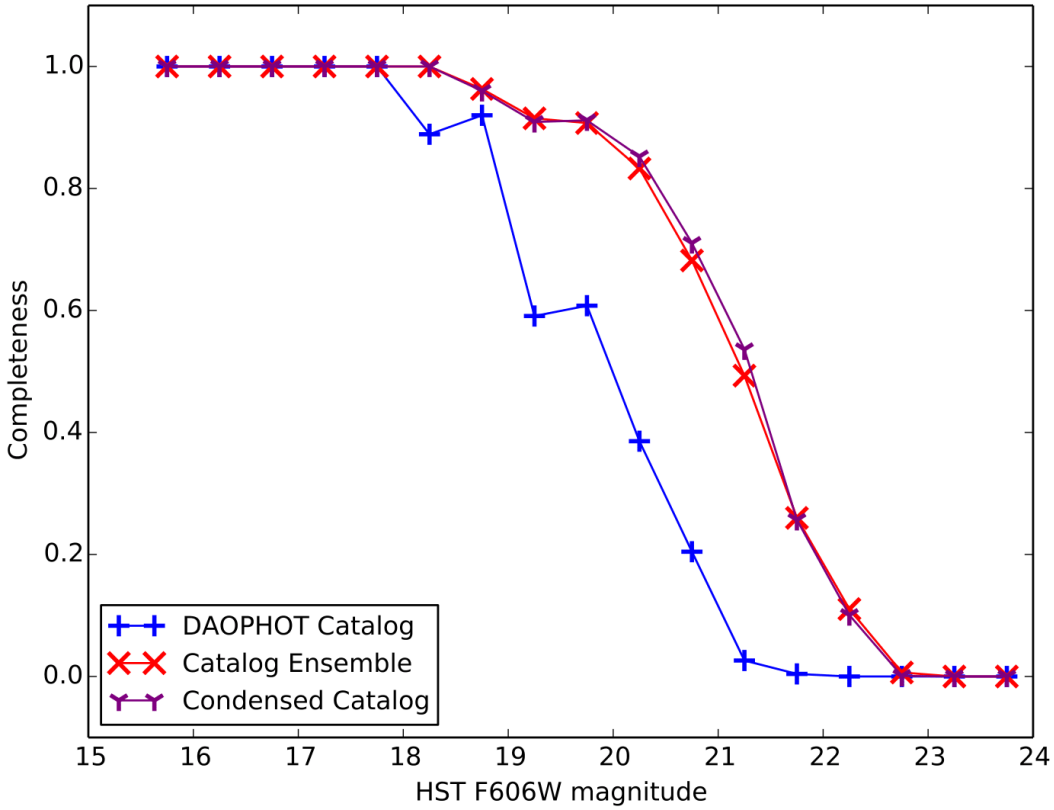


Worst-Case Residuals

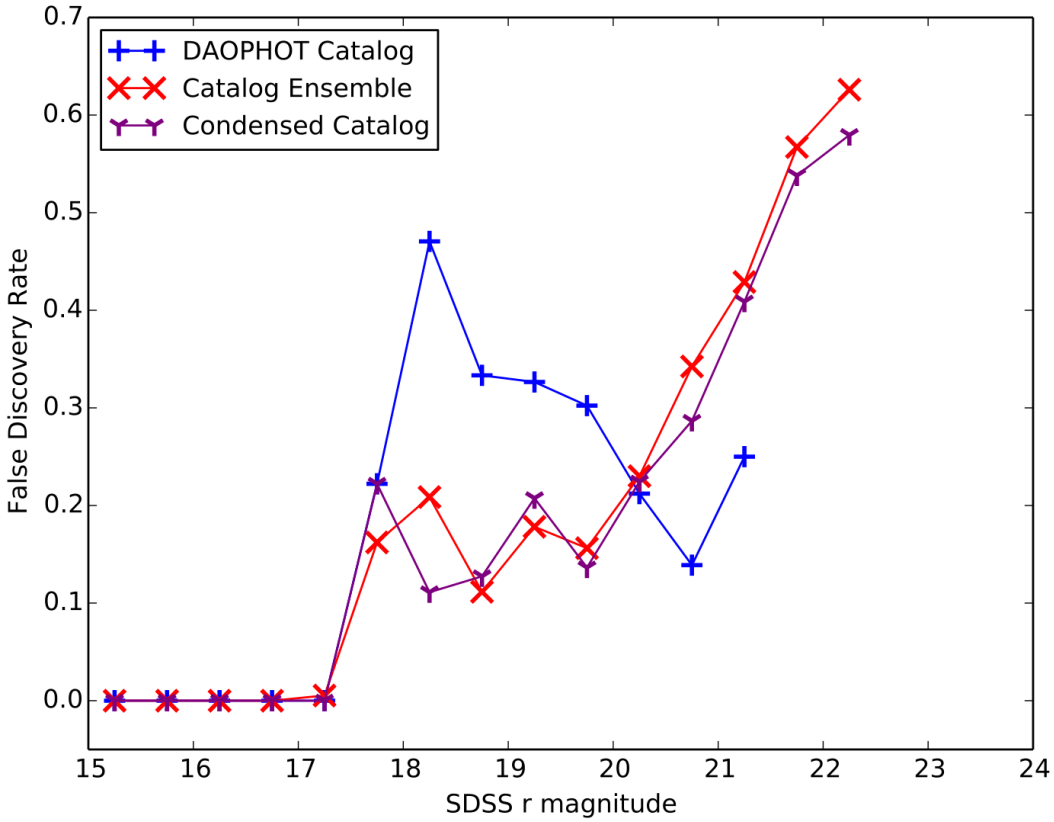




Completeness (Condensed Catalogue)



False Discovery Rate (Condensed Catalogue)



Receiver Operating Curve