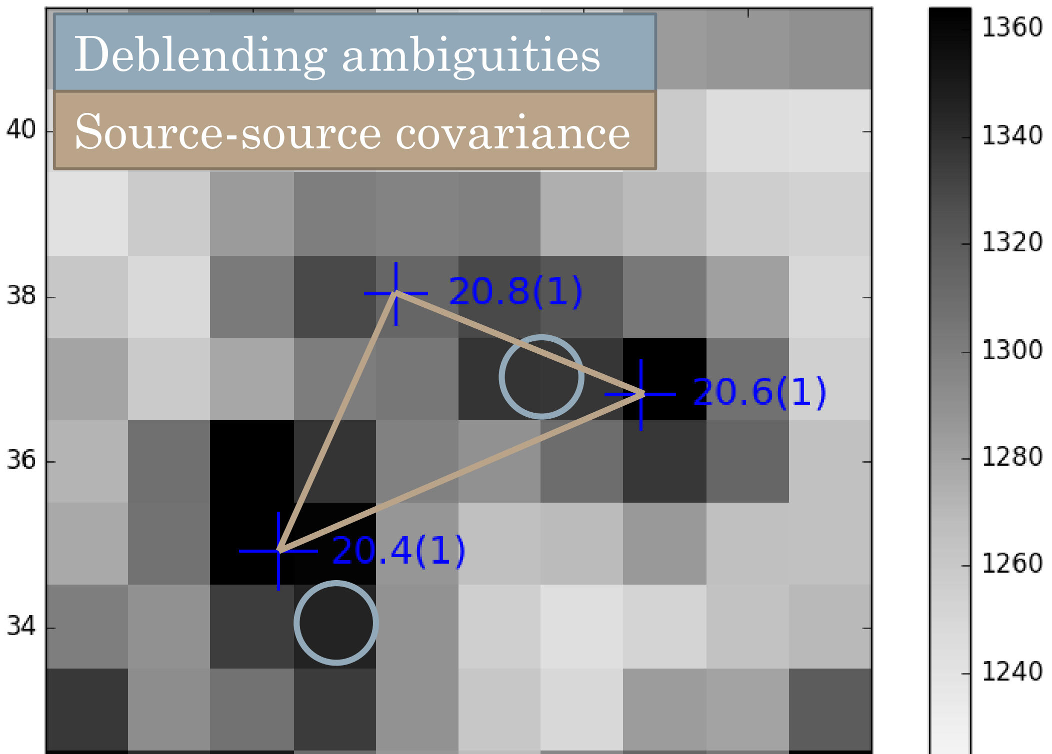
Probabilistic Cataloguing in Crowded Fields

Stephen K N Portillo
with Benjamin C G Lee, Tansu Daylan and Douglas P Finkbeiner
19 September 2017
AstroStat Day

Cataloguing Crowded Stellar Fields

Portillo et al. 2017, AJ, 154, 4

Traditional Catalogue Issues



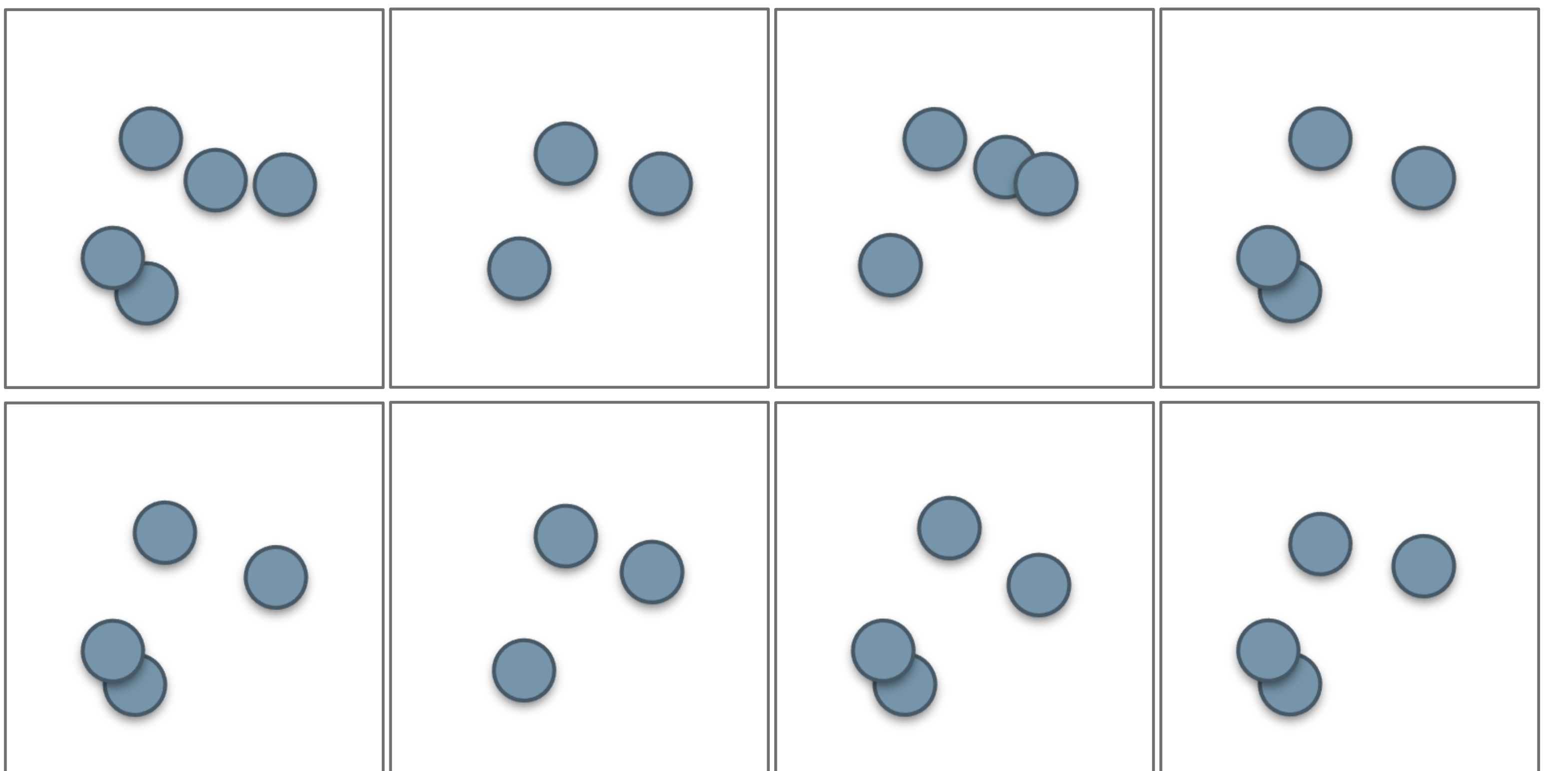
Deblending ambiguities

Source-source covariance

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

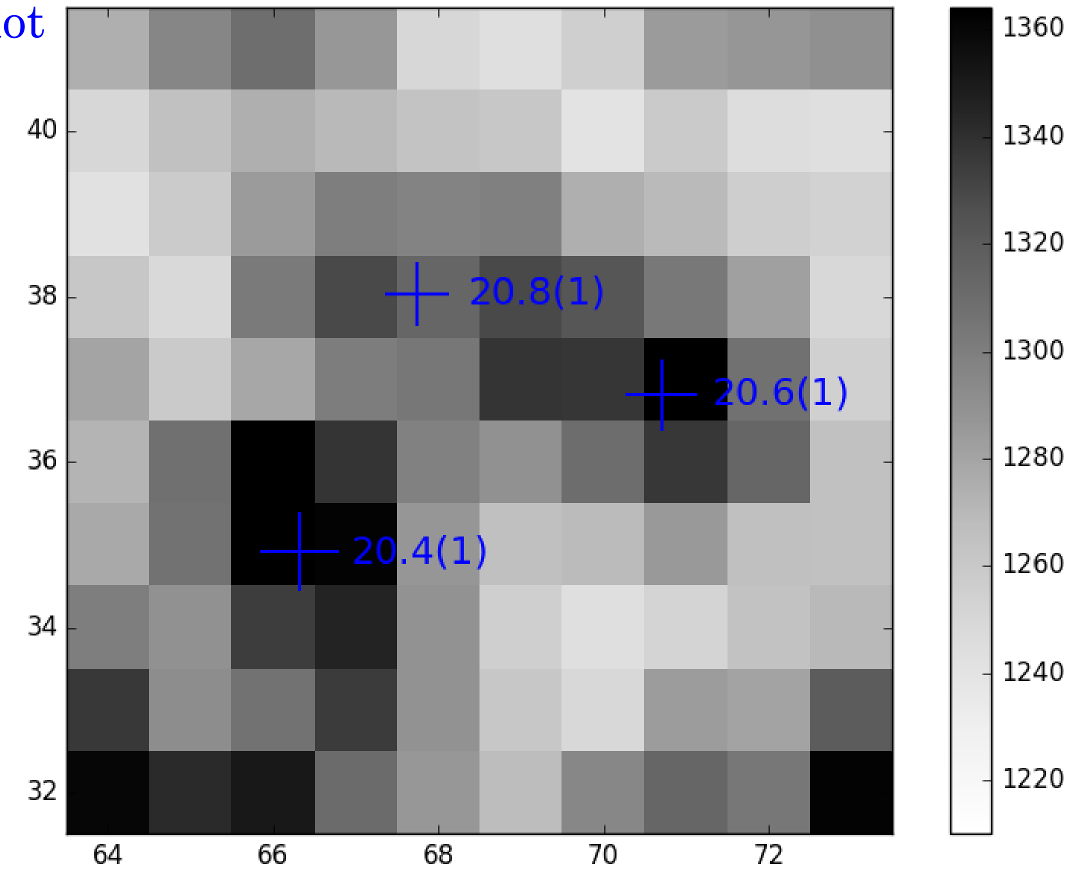
Probabilistic Cataloguing

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



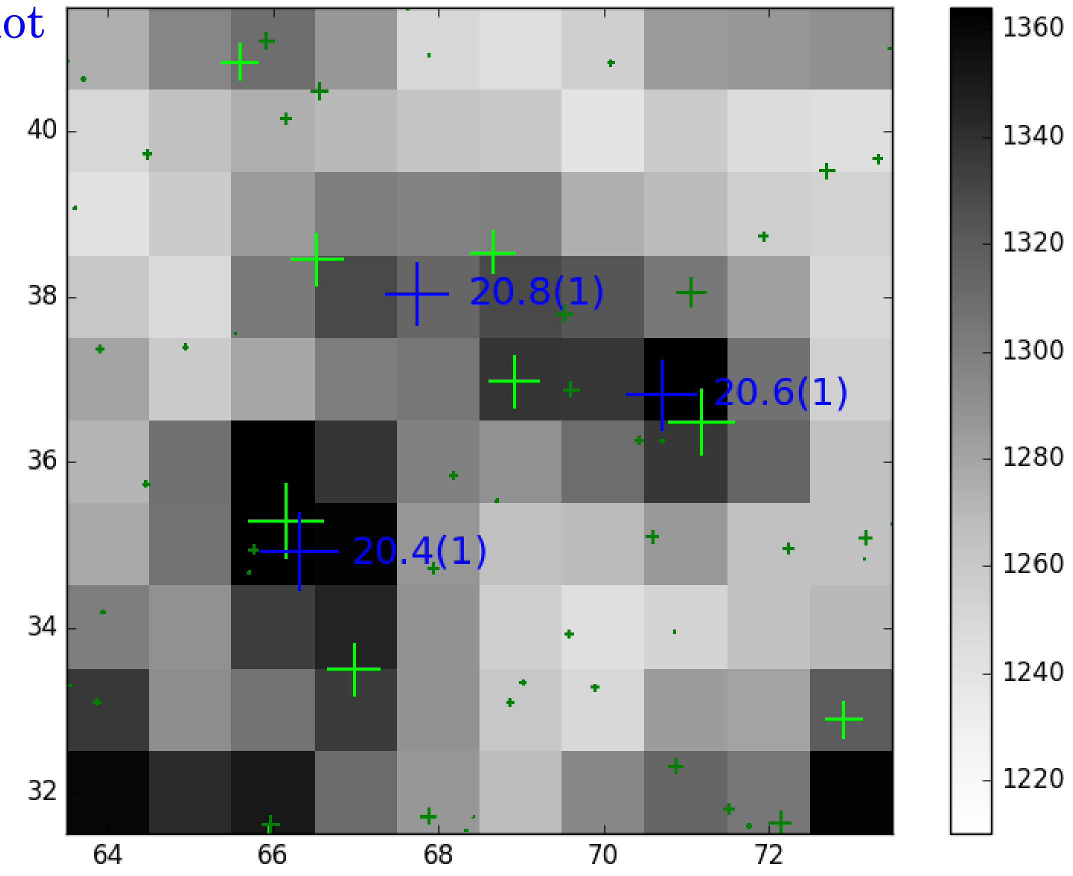
Traditional Catalogue

DAOPhot



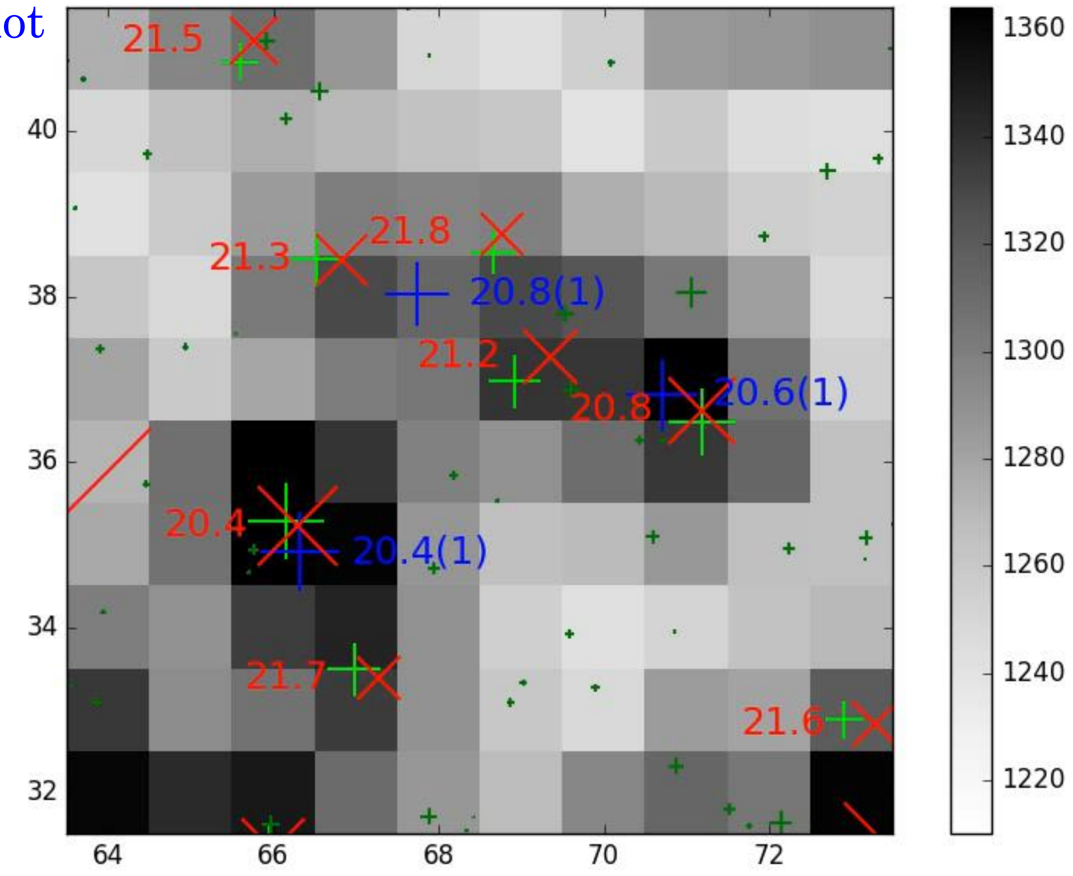
Compared to Hubble

DAOPhot
HST

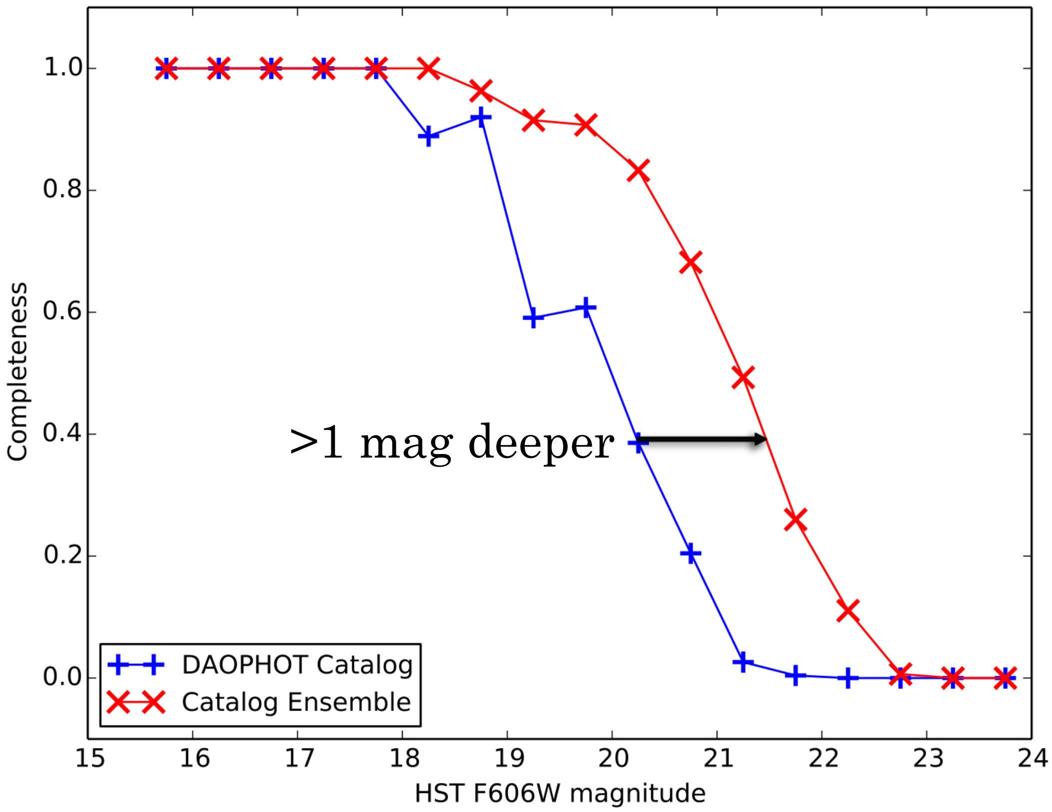


Probabilistic Catalogue

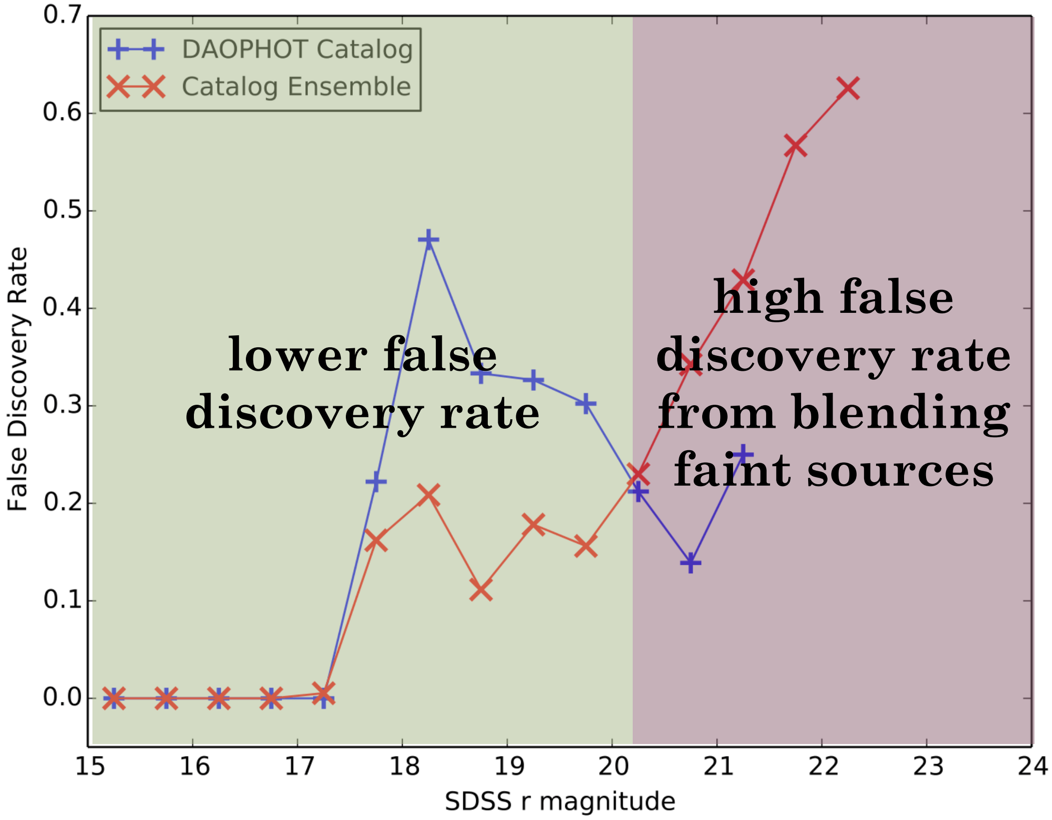
DAOPhot
HST
PCat



Completeness

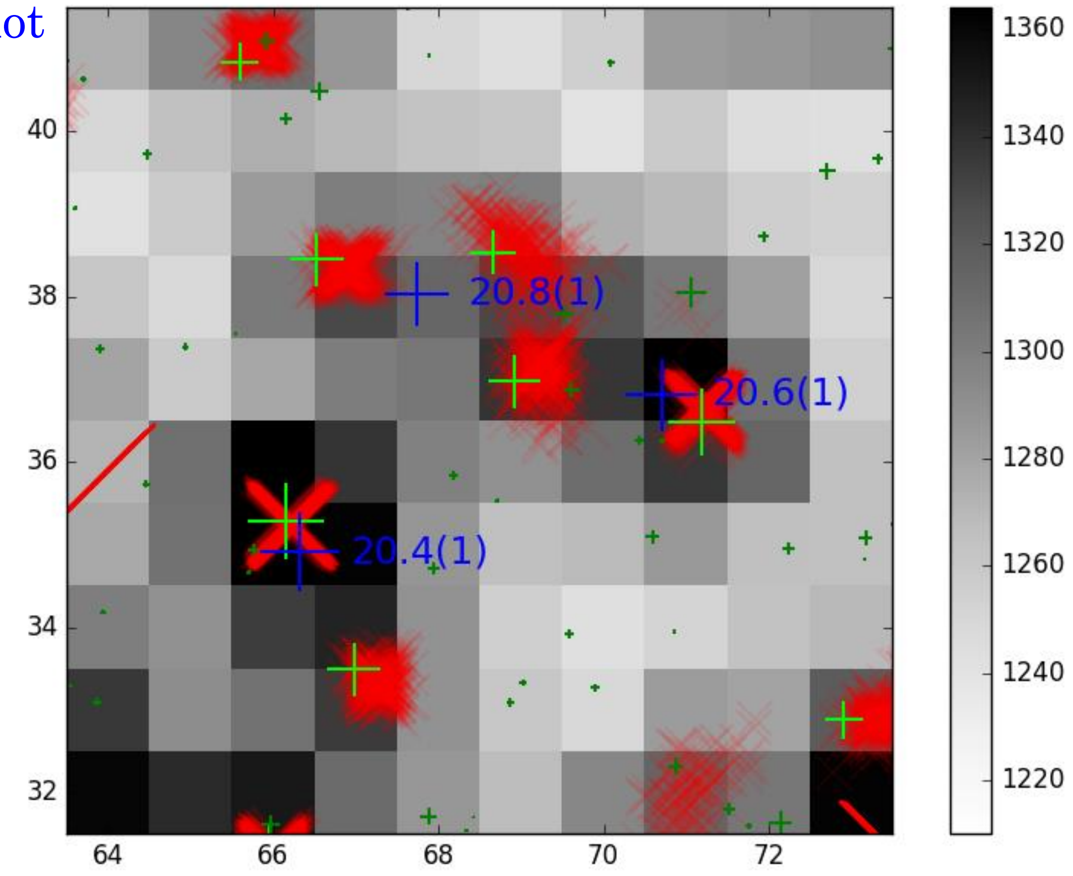


False Discovery Rate



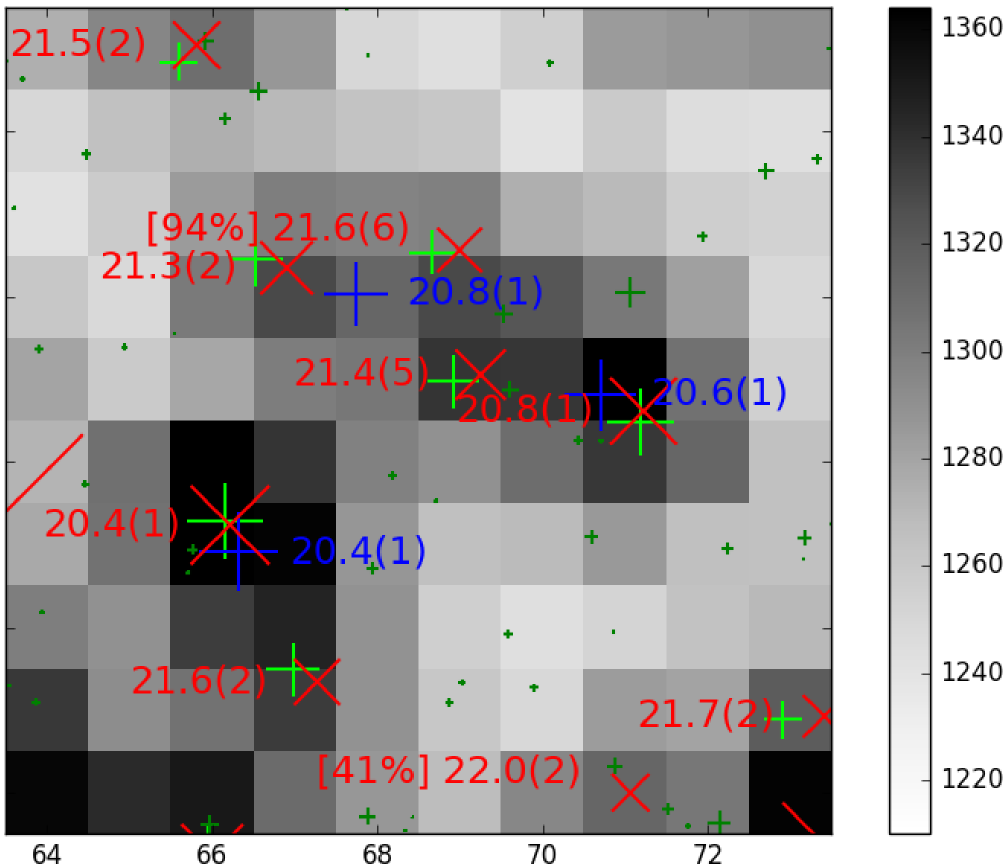
Stacked Catalogue Ensemble

DAOPhot
HST
PCat



Condensed Catalogue

DAOPhot
HST
PCat
(condensed)



Cataloguing Crowded Stellar Fields

Cataloguing Crowded ~~Stellar~~ Fields

Extending Probabilistic Cataloguing

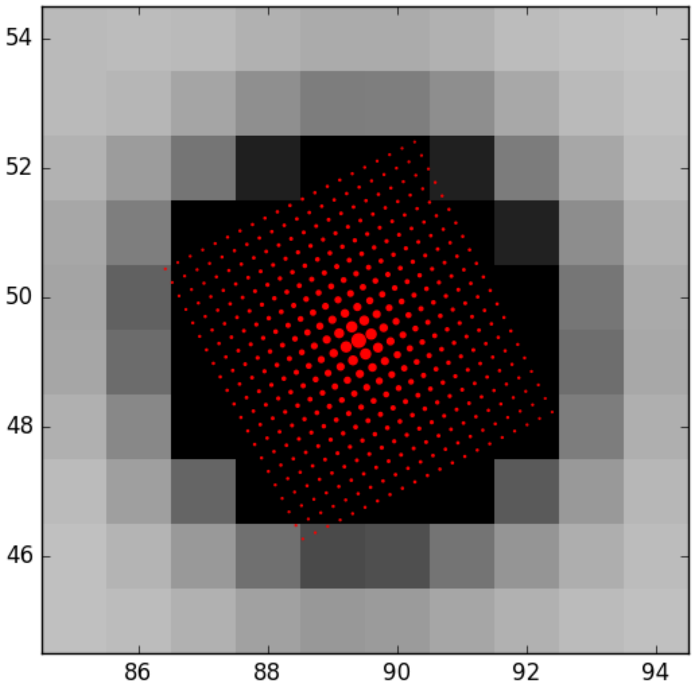
- Probabilistic cataloguing is a flexible framework
- Can incorporate any class of sources that can be parameterized so that:
  - The data can be forward modelled from a catalogue, in order to calculate the **likelihood**
  - **Priors** can be placed on these parameters
- Galaxies are often fit with Sérsic profiles

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

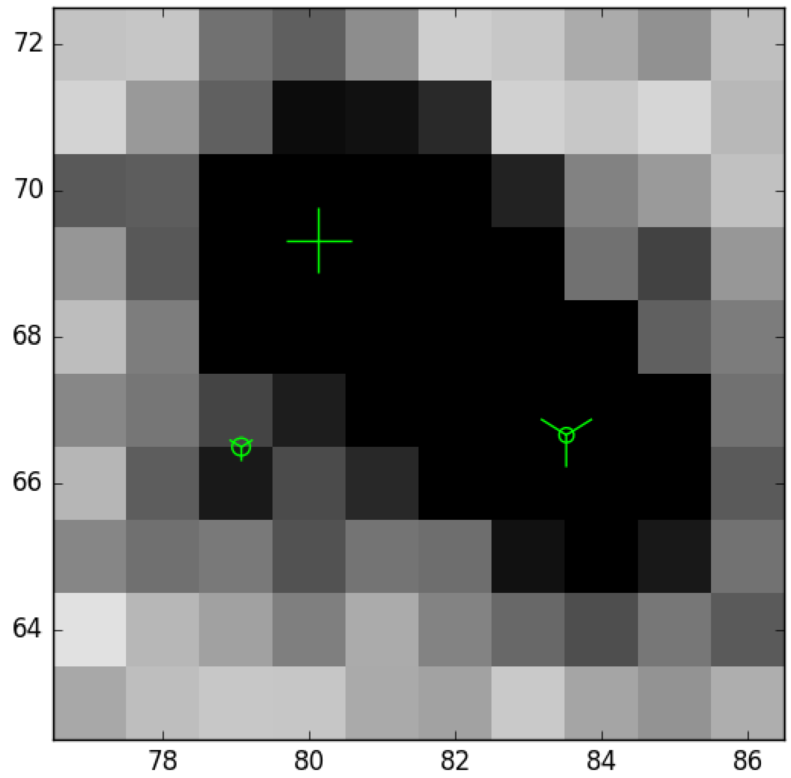
- These profiles are approximate – most useful for galaxies where detailed structure is not discernable

Galaxy Model Images

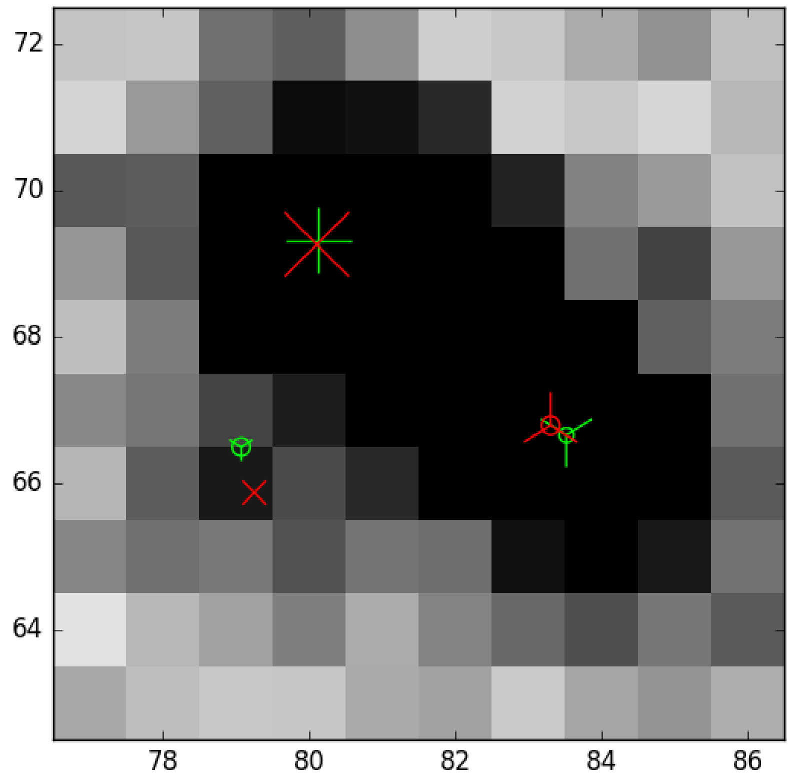
- Calculating model image is already the slowest part for stars – even more so for galaxies
- We developed a faster way to calculate model images of point sources
- So approximate galaxy profiles with a collection of point sources?



Deblending Stars + Galaxies



Deblending Stars + Galaxies



Conclusion

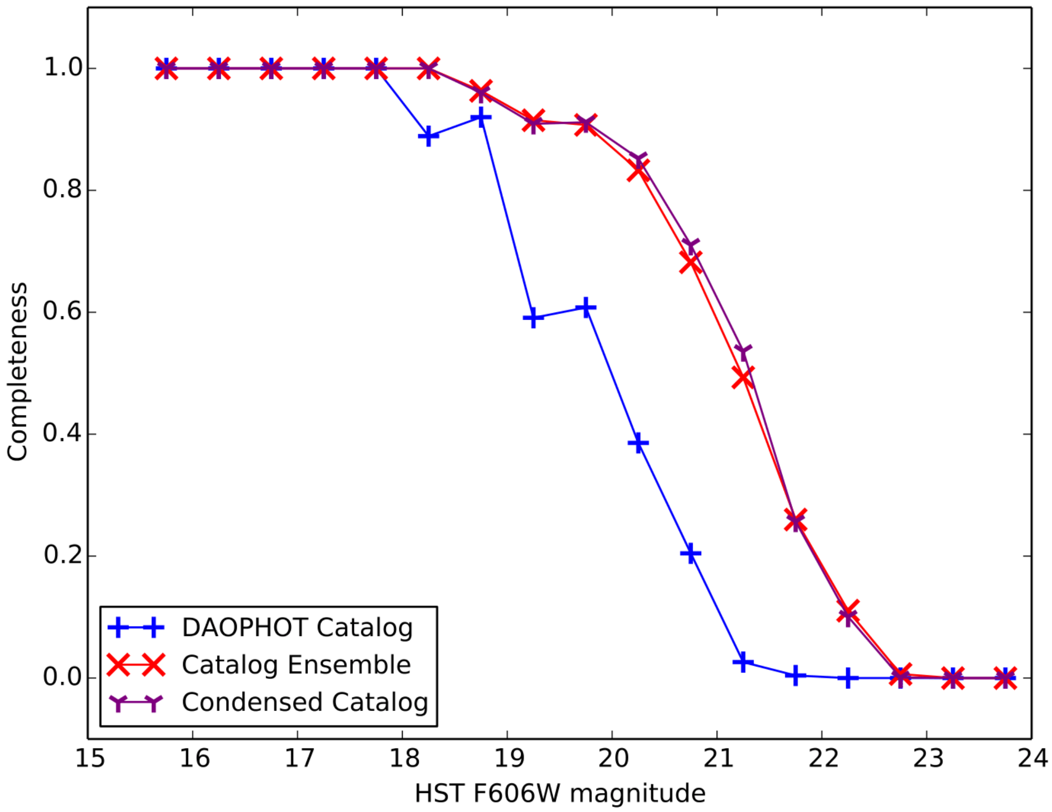
- The problem of crowded field photometry will be very relevant in the LSST era
- Probabilistic cataloguing is well-suited for crowded fields, capturing deblending ambiguities and source-source covariance
- We have demonstrated that probabilistic cataloguing outperforms traditional cataloguing in crowded stellar fields
- We have been able to speed up our probabilistic cataloguing implementation and are extending it to galaxies

Backup Slides

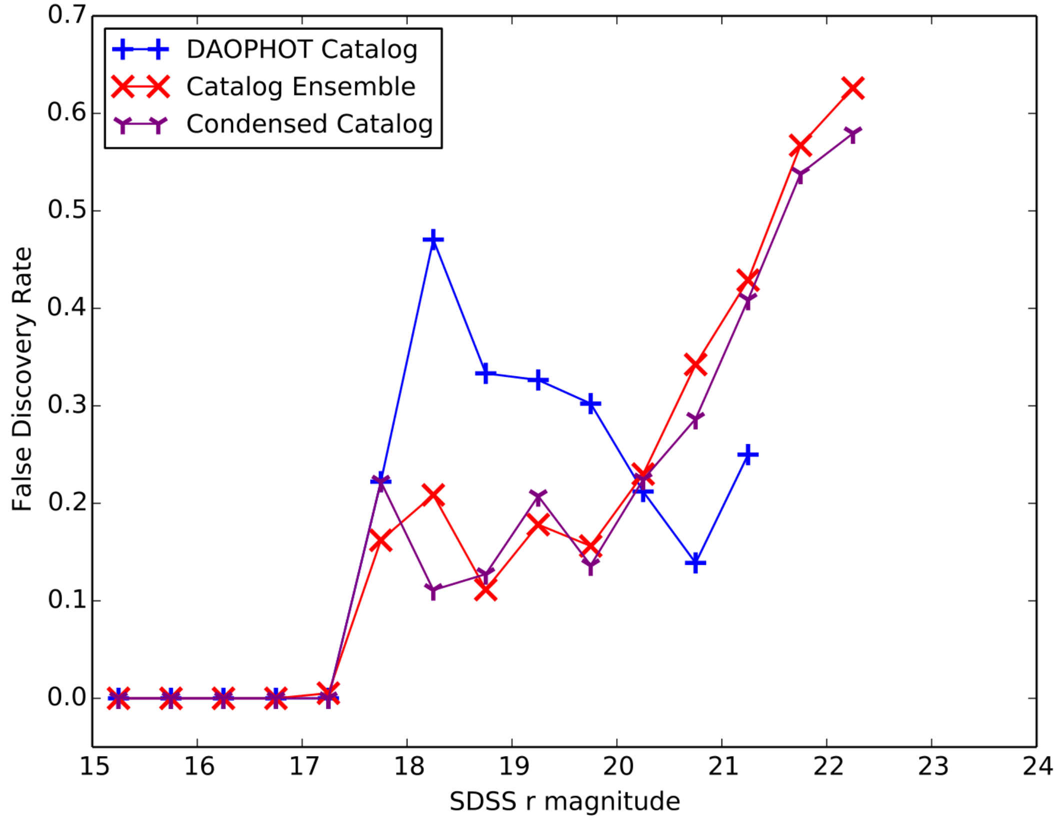
Use Case Specifications

- $40'' \times 40''$ 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)

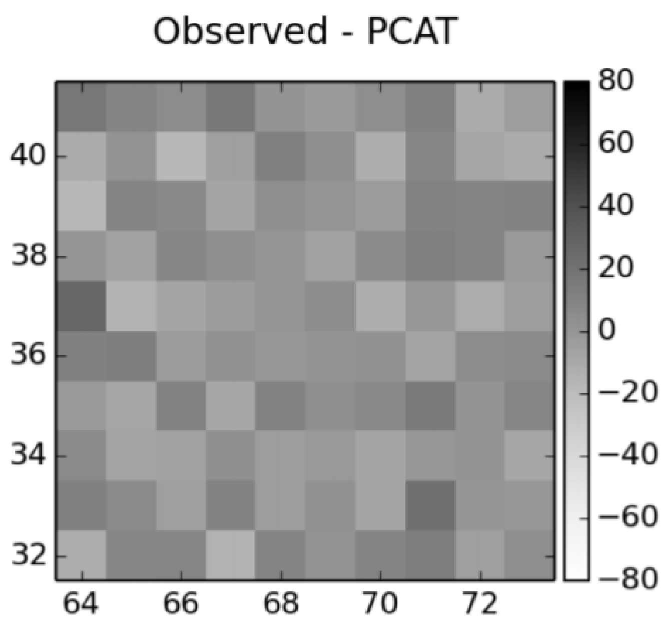
Completeness (Condensed Catalogue)



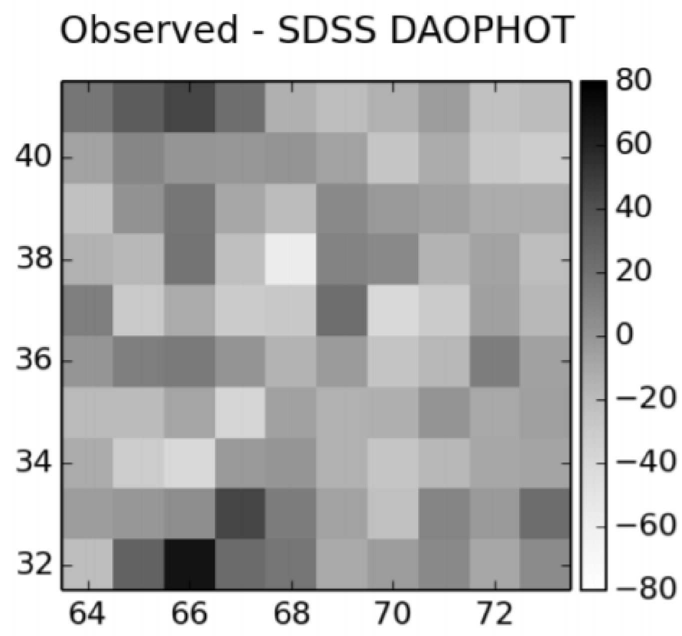
False Discovery Rate (Condensed Catalogue)



Residuals

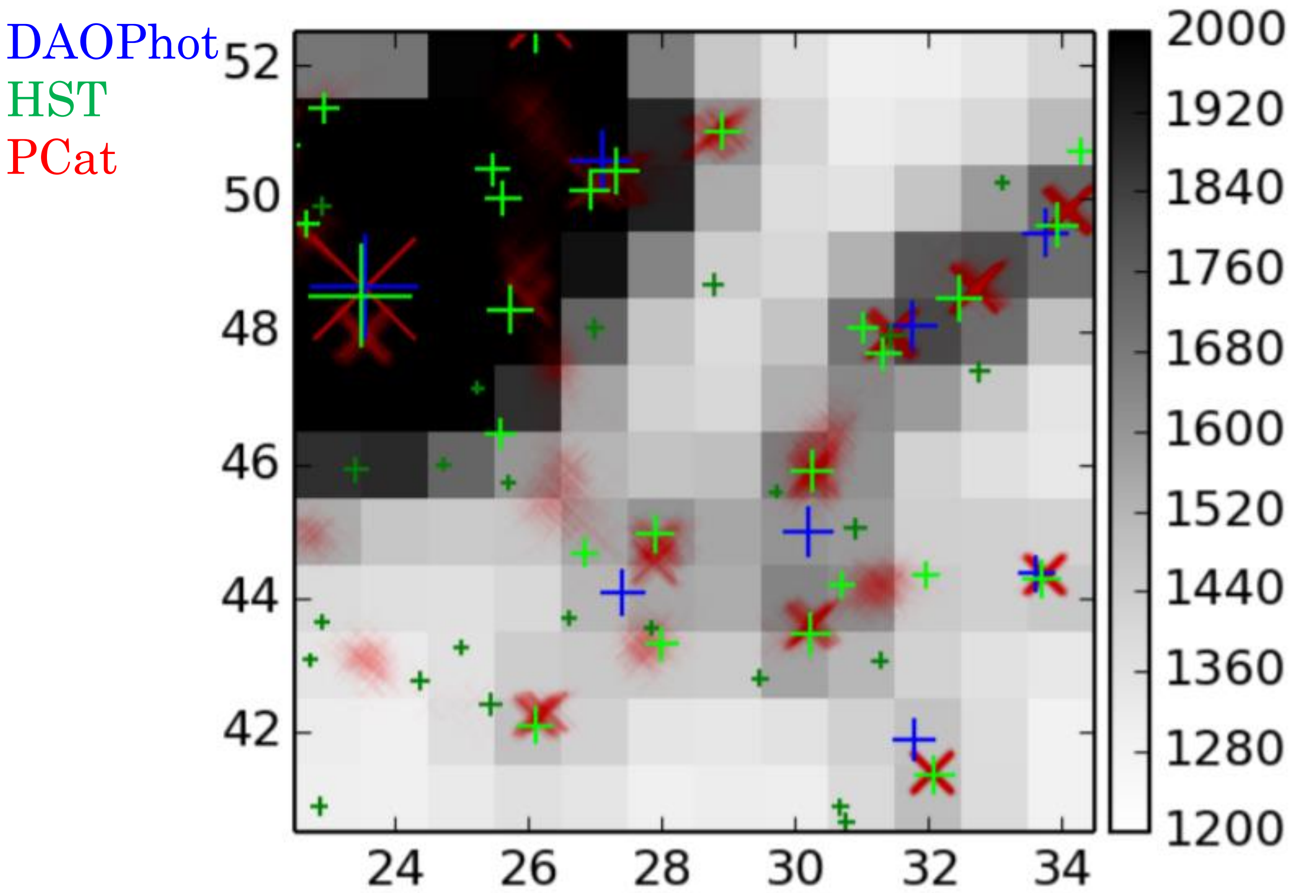


Observed - PCAT
40
38
36
34
32
64
66
68
70
72
80
60
40
20
0
−20
−40
−60
−80

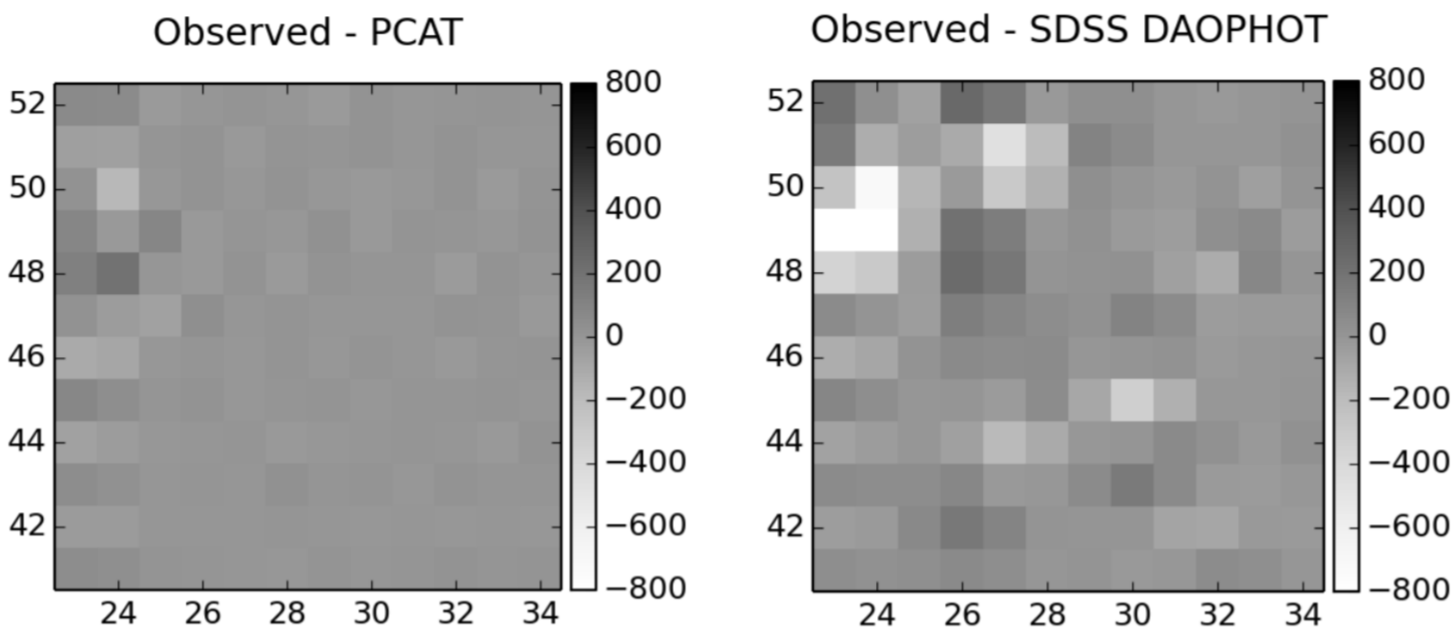


Observed - SDSS DAOPHOT
40
38
36
34
32
64
66
68
70
72
80
60
40
20
0
−20
−40
−60
−80

Worst-Case Scenario



Worst-Case Residuals



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 $x = \{x_1, \dots, x_N\}$ and $x' = \{x_1, \dots, x_{N+1}\}$ 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)

Catalogue Star Priors

- Prior that sources are independent:

$$\pi\left(\{x_i, y_i, F_i\}_{i=1}^{N}\right) = \pi(N) \prod_{i=1}^{N} \pi(x_i, y_i)\pi(f_i)$$

- Prior on location is flat
- Prior on flux is a power law of index 2, above f_{min}
- Prior on number penalizes the $(N+1)^{\text{th}}$ star based on the expected improvement in χ^2 under the null hypothesis that there are N stars:

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

Catalogue Galaxy Priors

- Galaxies have more parameters: modelled as circular disks with scale radius r and 3D orientation (θ, φ)
- Galaxies are given Sérsic profiles with fixed index
- Similar priors as stars for location and flux
- Prior on scale radius is a power law with index 2 with a minimum of 1 pixel
- Prior on orientation is spherically symmetric
- Prior on number penalizes the $(N+1)^{\text{th}}$ galaxy based on the expected improvement in χ^2 under the null hypothesis that there are N galaxies:

$$\log \frac{\pi(N+1)}{\pi(N)} = -\frac{6}{2}$$

Receiver Operating Curve