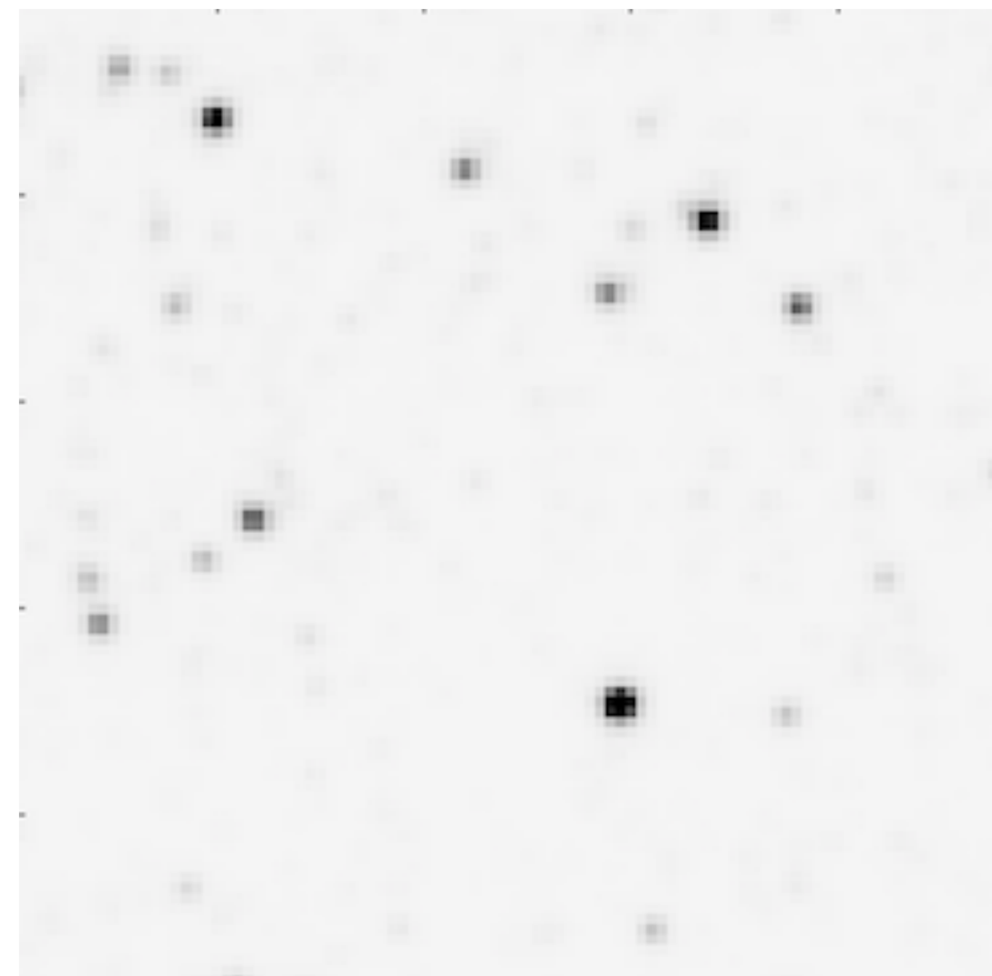
Probabilistic Catalogs for photometry in *Extremely* Crowded Fields

Douglas Finkbeiner
with Brendan Meade,
Stephen Portillo, Ben Lee, and Tansu Daylan

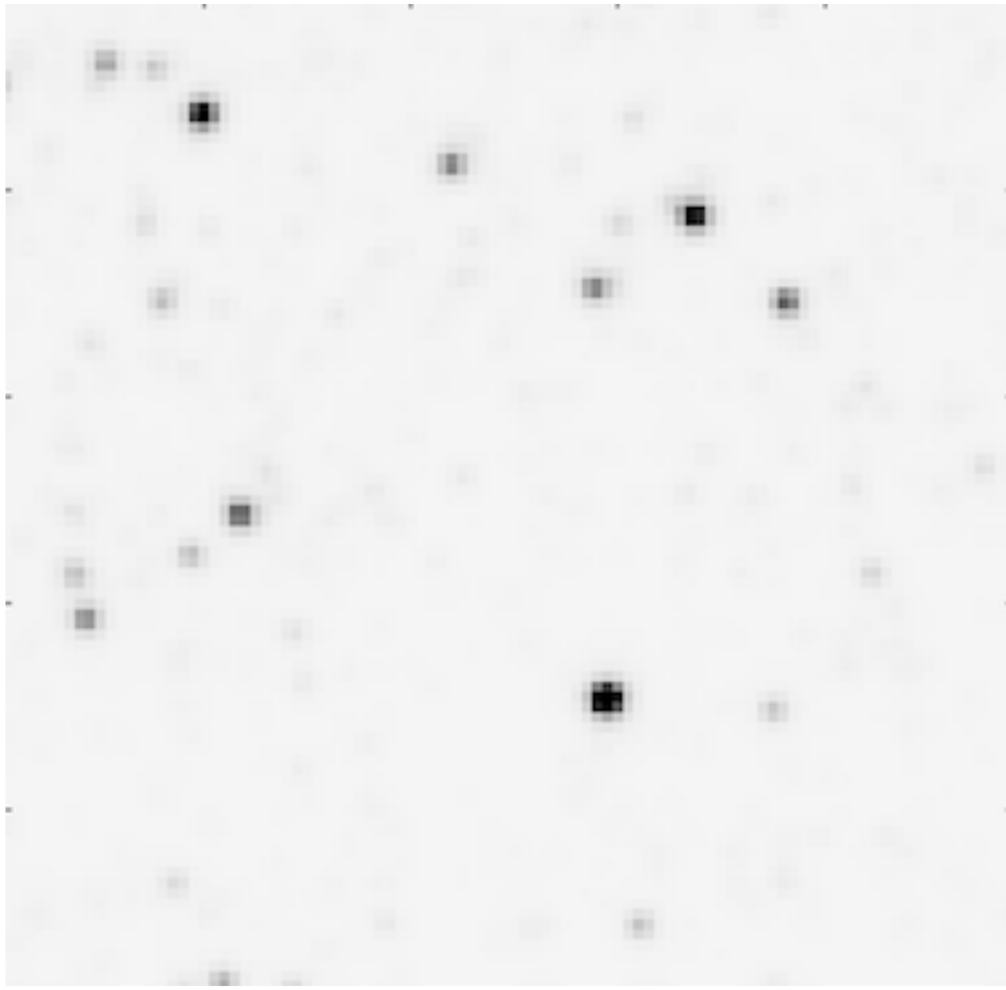
6 December, 2016

Stellar photometry:



Stellar photometry:

Identify objects

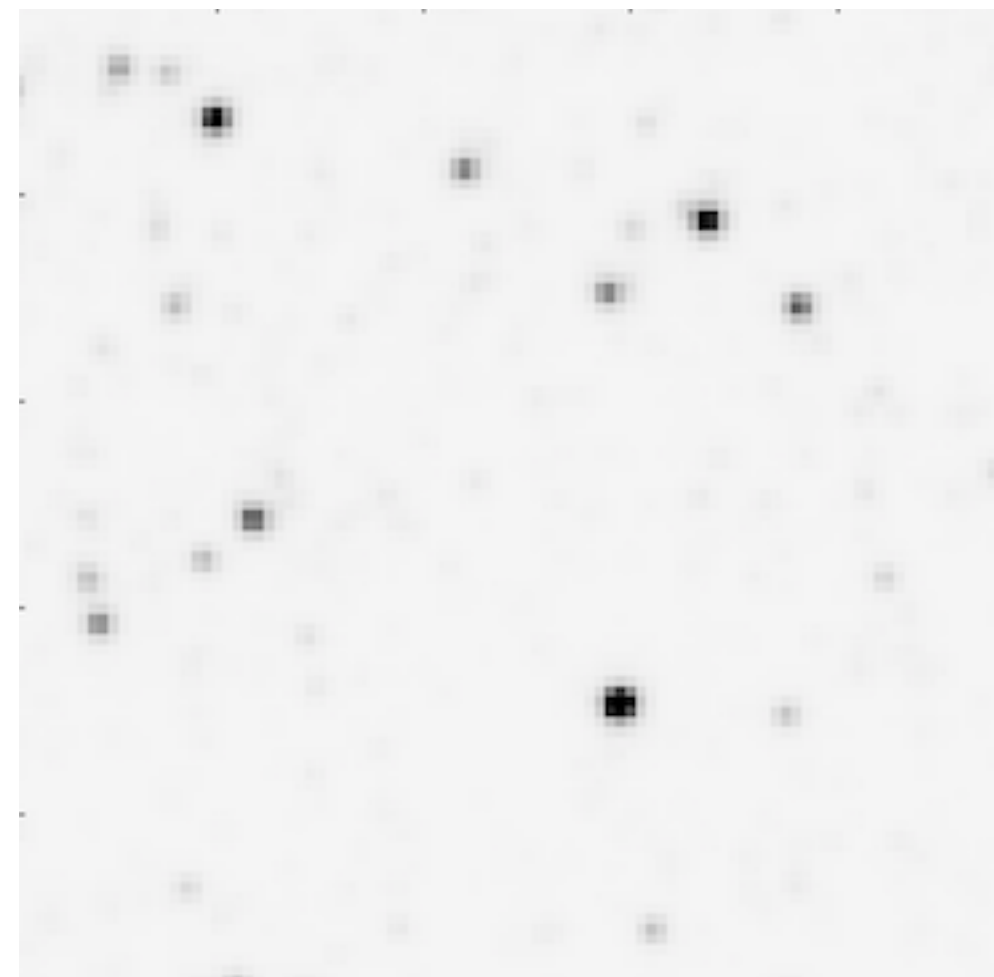


Stellar photometry:

Identify objects

Measure {x,y,flux} of each
(and sky level, psf shape…)

Optimize those to maximize likelihood of reconstructed image.
(Gaussian or Poisson)



Stellar photometry:

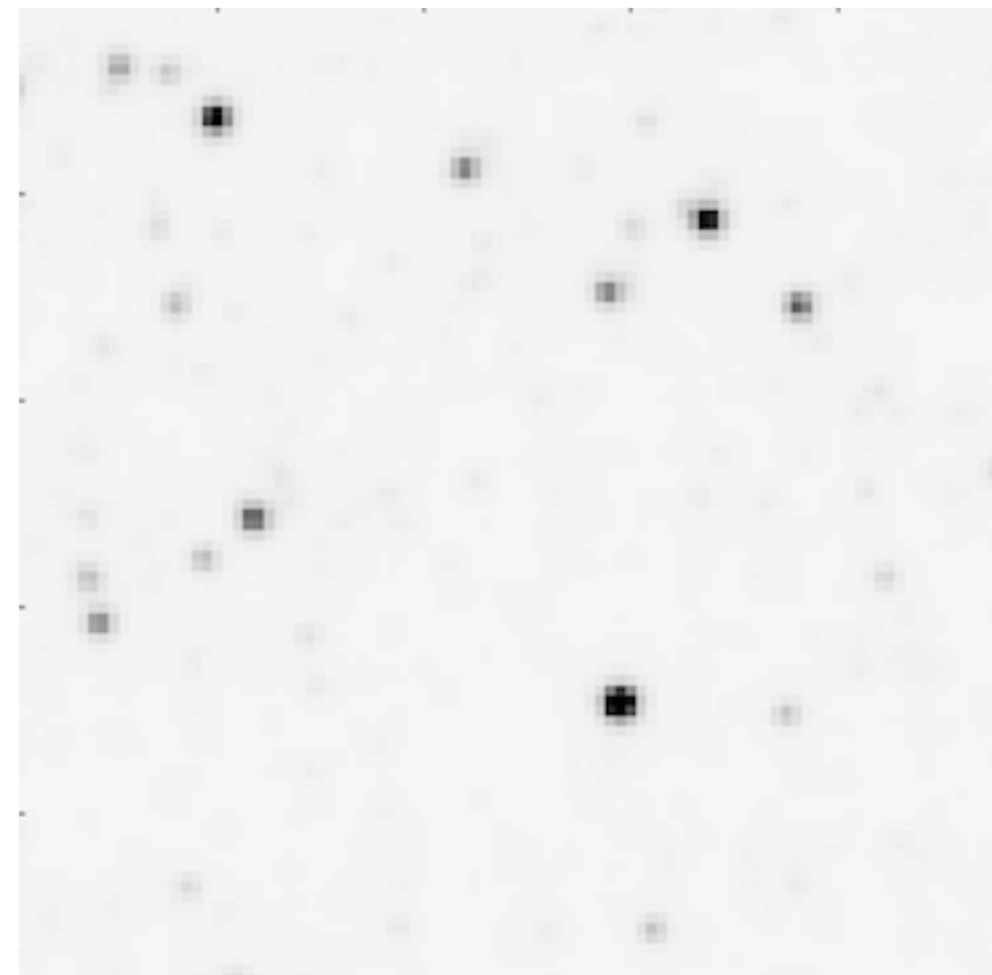
Identify objects

Measure {x,y,flux} of each
(and sky level, psf shape…)

Optimize those to maximize likelihood of reconstructed image.
(Gaussian or Poisson)

Inputs: astrometric / photometric cal, prior on flux distribution.

(Difference between optimal reconstruction of image and optimal reconstruction of catalog!)



There are many algorithms for this:

DAOPhot
DoPhot
SExtractor
SDSS pipeline
Pan-STARRS pipeline
DECam/DES pipeline
etc, etc…

All make different assumptions, try different approaches (representation of PSF) but all are attempting ~ the same thing.

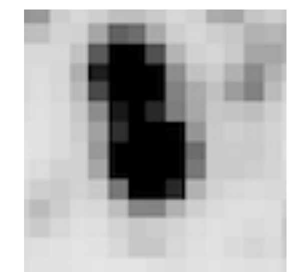
Crowded field photometry:

{x,y,flux} of each source are covariant with those of the neighbors. Can we track that?

Crowded field photometry:

{x,y,flux} of each source are covariant with those of the neighbors. Can we track that?

Could linearize the problem and make an x,y,flux covariance matrix, then marginalize over uncertainties in neighboring sources.

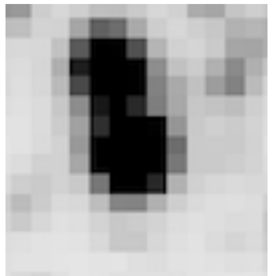


Need to ID all the neighbors first.

Crowded field photometry:

{x,y,flux} of each source are covariant with those of the neighbors. Can we track that?

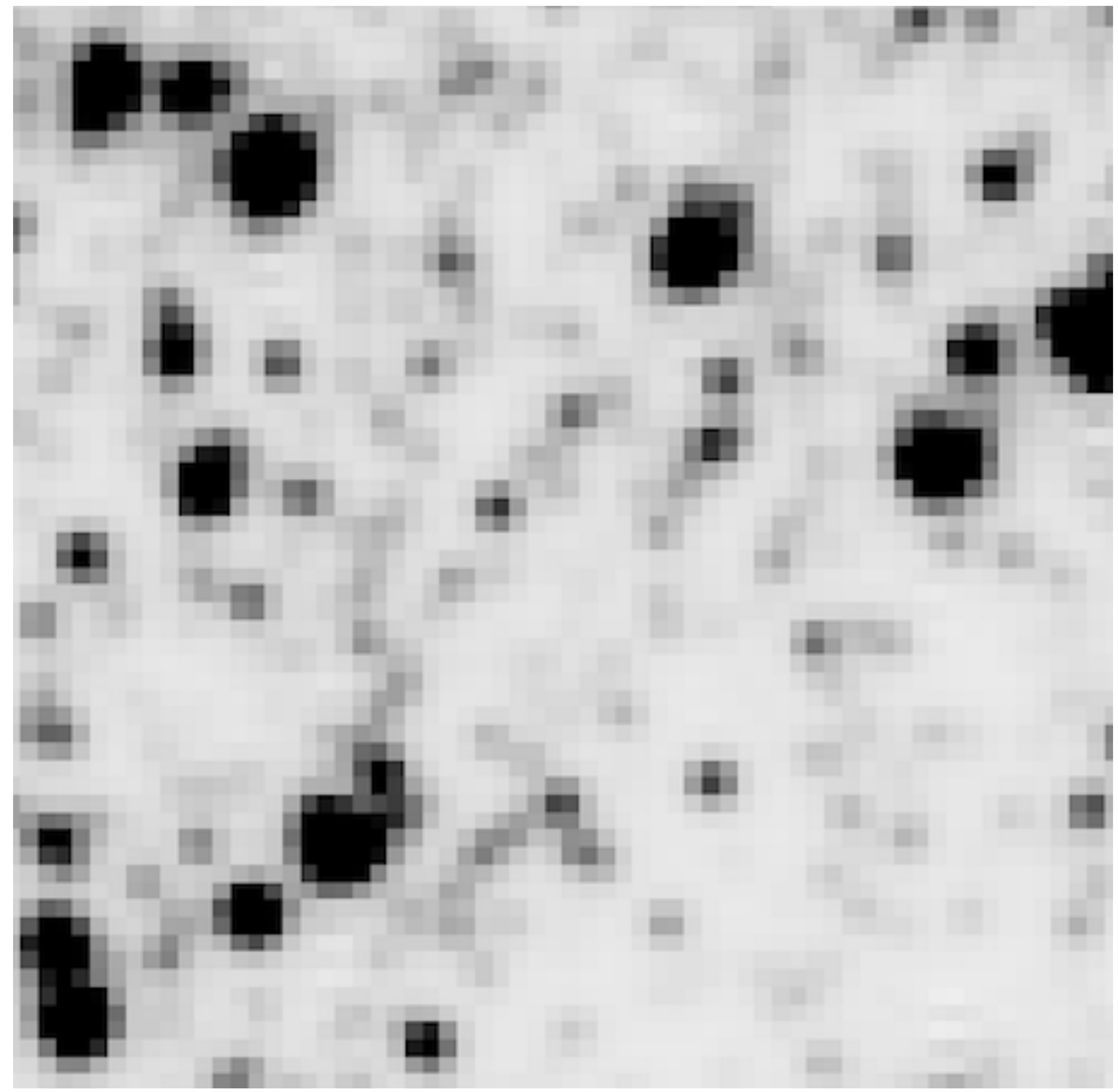
Could linearize the problem and make an x,y,flux covariance matrix, then marginalize over uncertainties in neighboring sources.



Need to ID all the neighbors first.

I.e. only max L (for a given source!) in the context of some parameterization.

But in general?



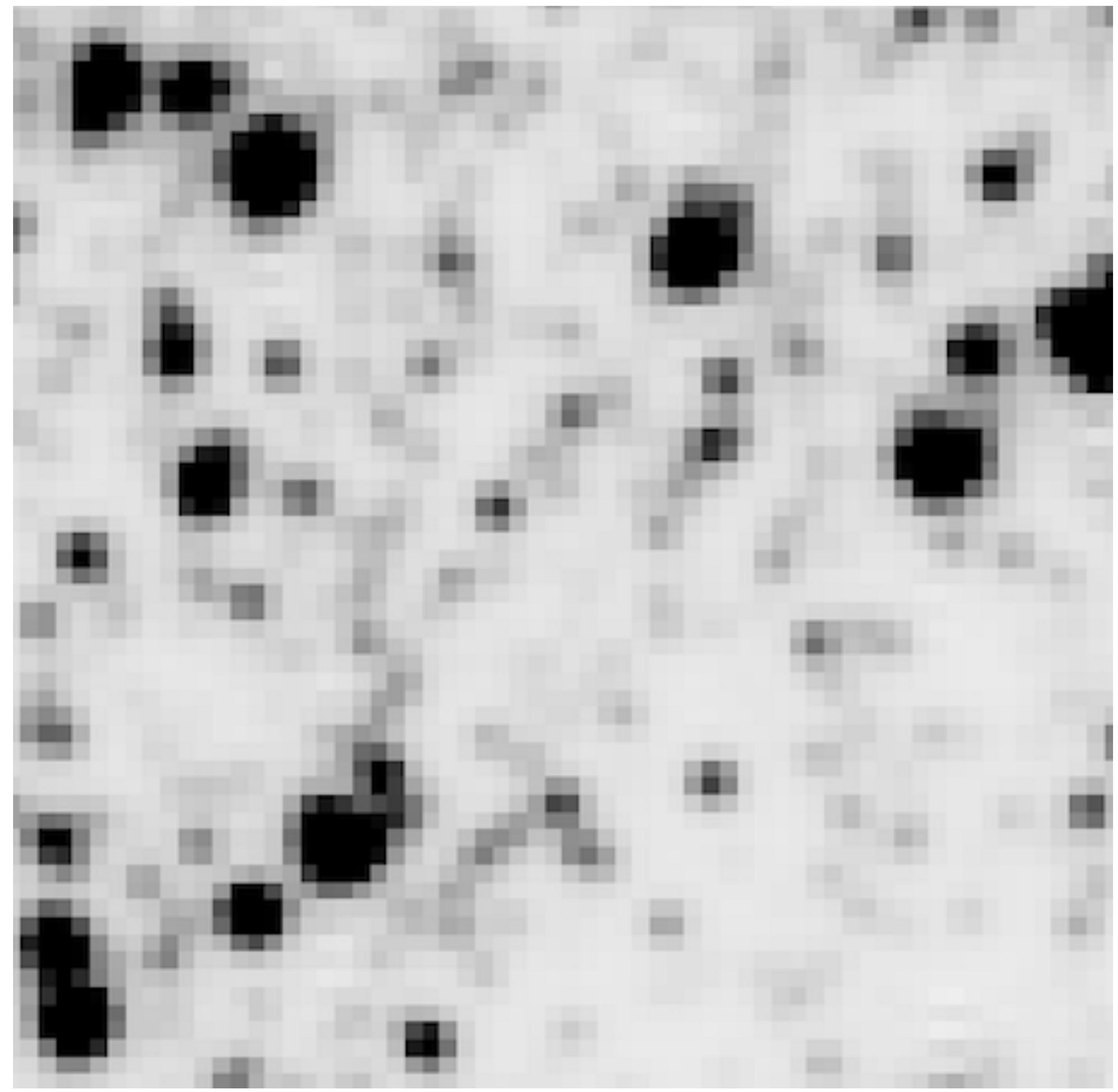
But in general?

Can keep "throwing sources at it" but when to stop?

How to propose births (and deaths)?

How to try all permutations of possible neighbors?

Correct uncertainty estimate must marginalize over all the options.



This sounds like MCMC

in a variable-dimension parameter space!

This sounds like MCMC

in a variable-dimension parameter space!

Trans-dimensional search

Probabilistic Catalogs for Crowded Stellar Fields

Brendon J. Brewer, Daniel Foreman-Mackey, David W. Hogg

(Submitted on 25 Nov 2012 (v1), last revised 20 Apr 2013 (this version, v2))

We present and implement a probabilistic (Bayesian) method for producing catalogs from images of stellar fields. The method is capable of inferring the number of sources N in the image and can also handle the challenges introduced by noise, overlapping sources, and an unknown point spread function (PSF). The luminosity function of the stars can also be inferred even when the precise luminosity of each star is uncertain, via the use of a hierarchical Bayesian model. The computational feasibility of the method is demonstrated on two simulated images with different numbers of stars. We find that our method successfully recovers the input parameter values along with principled uncertainties even when the field is crowded. We also compare our results with those obtained from the SExtractor software. While the two approaches largely agree about the fluxes of the bright stars, the Bayesian approach provides more accurate inferences about the faint stars and the number of stars, particularly in the crowded case.

Imagine the space of all possible (star) catalogs,
with

N = {0,1,2,3,…Nmax} sources. Define a likelihood function (or posterior) in that space.

Sample from it.

Proposals to perturb
x,y,flux

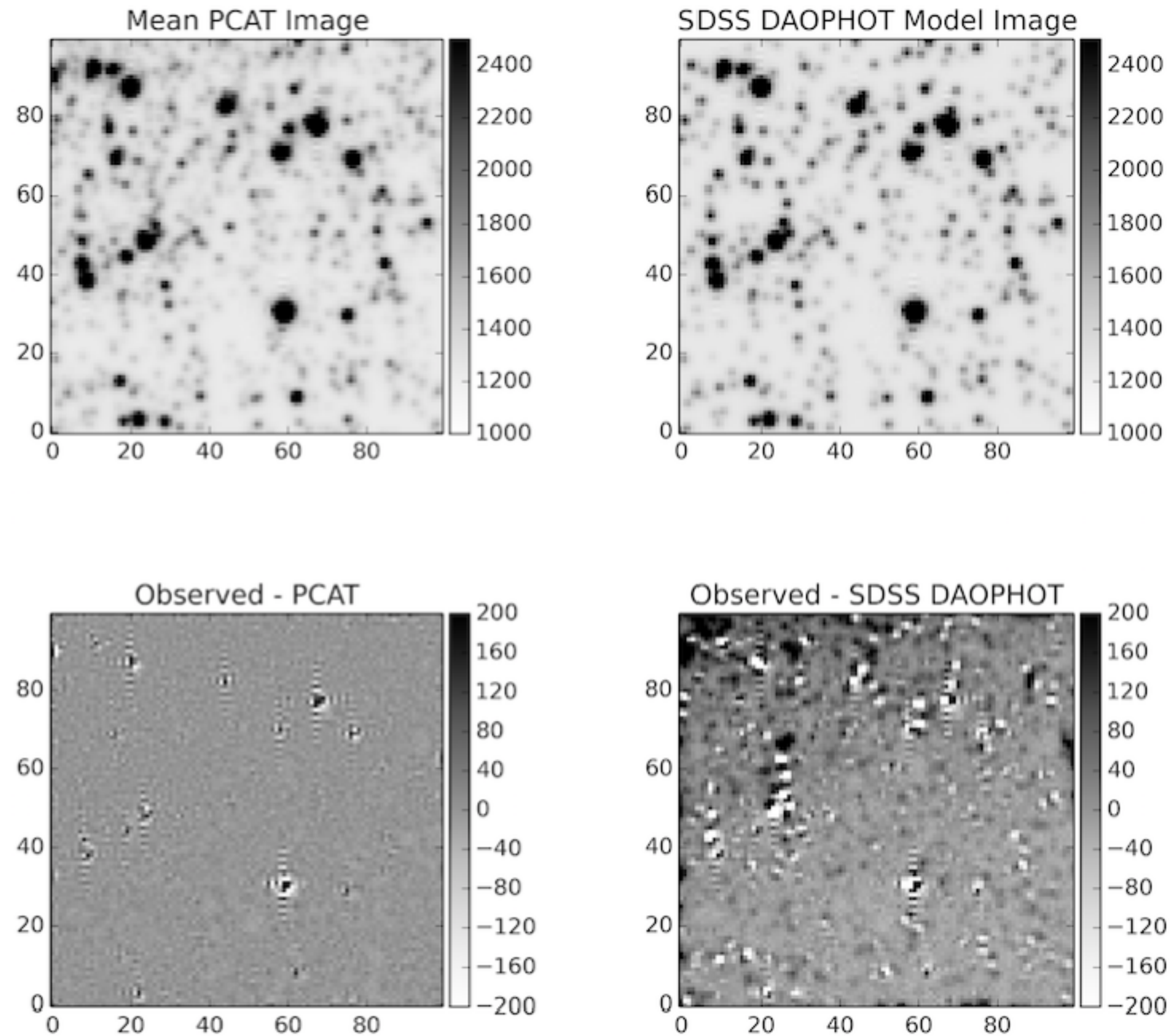
but also
add stars
remove stars
split stars
merge stars

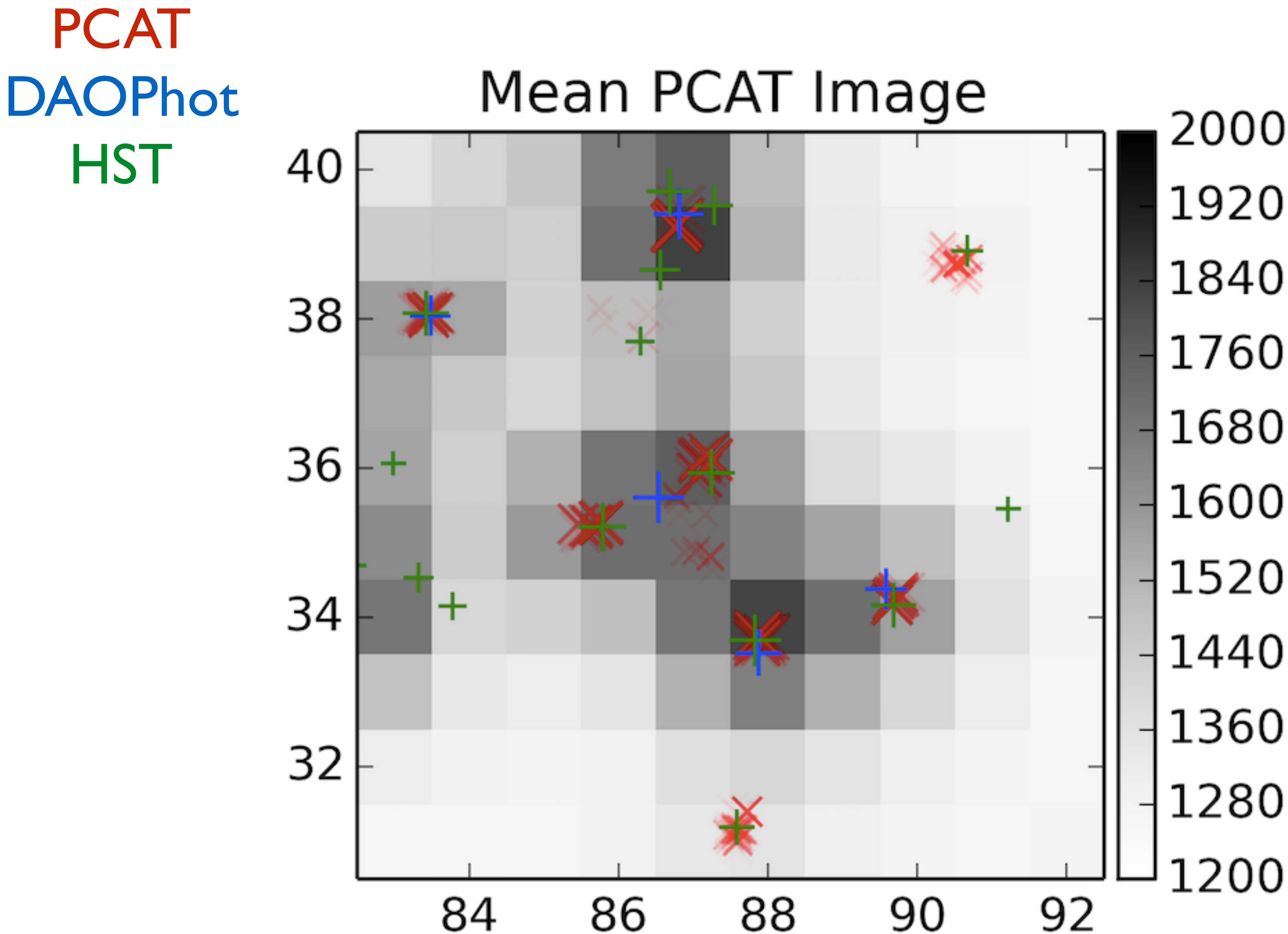
Every type of move must be reversible
-> detailed balance.

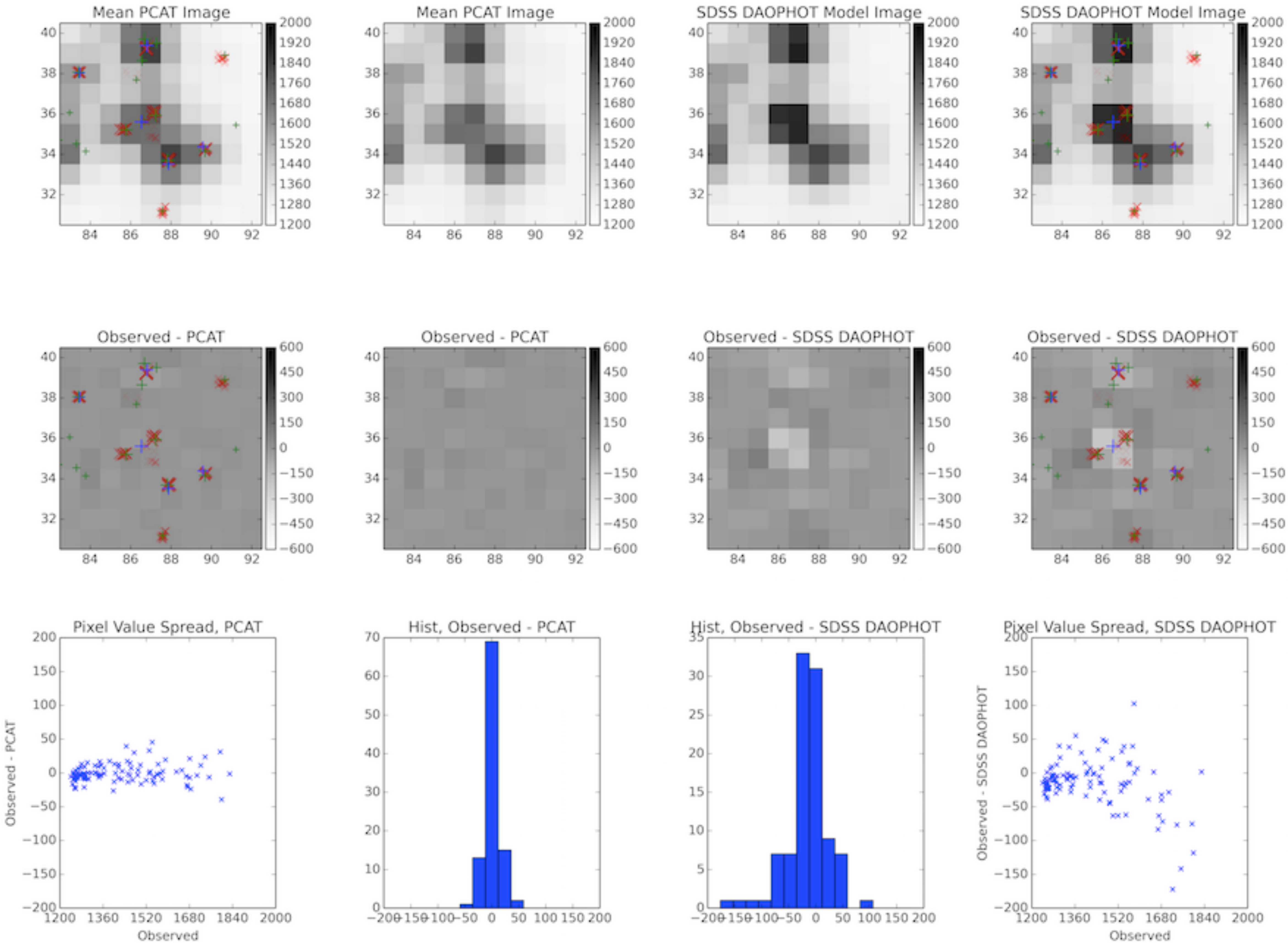
We do this with our code “PCAT” using
the DNEST3 sampler by Brendon Brewer

Test case:
Messier 2 (globular cluster)
on SDSS Stripe 82 (lots of data)
Also HST data (for reality check)

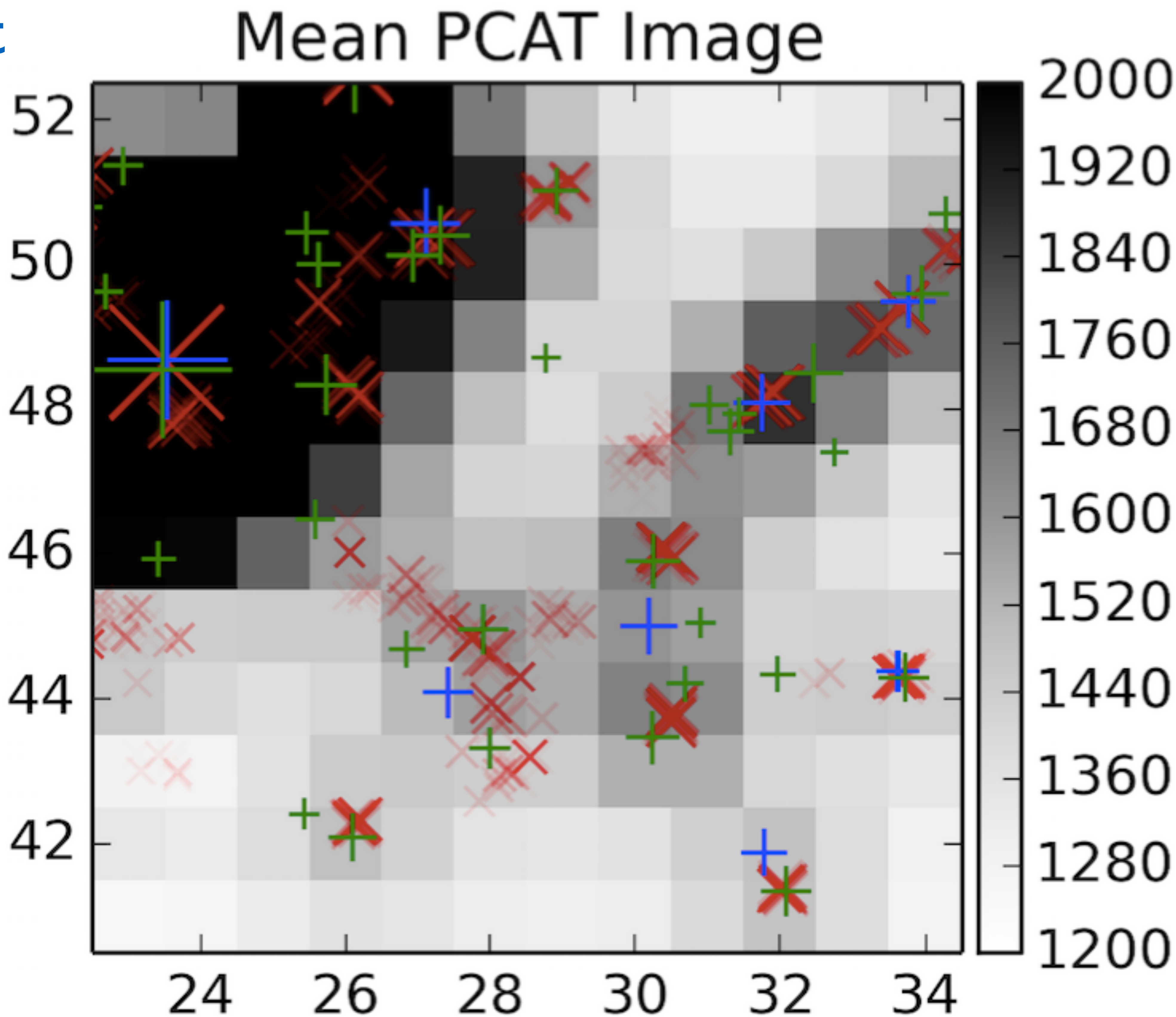
SDSS pipeline failed, but
An etal. (2007) provide DAOPhot
catalogs in this field.

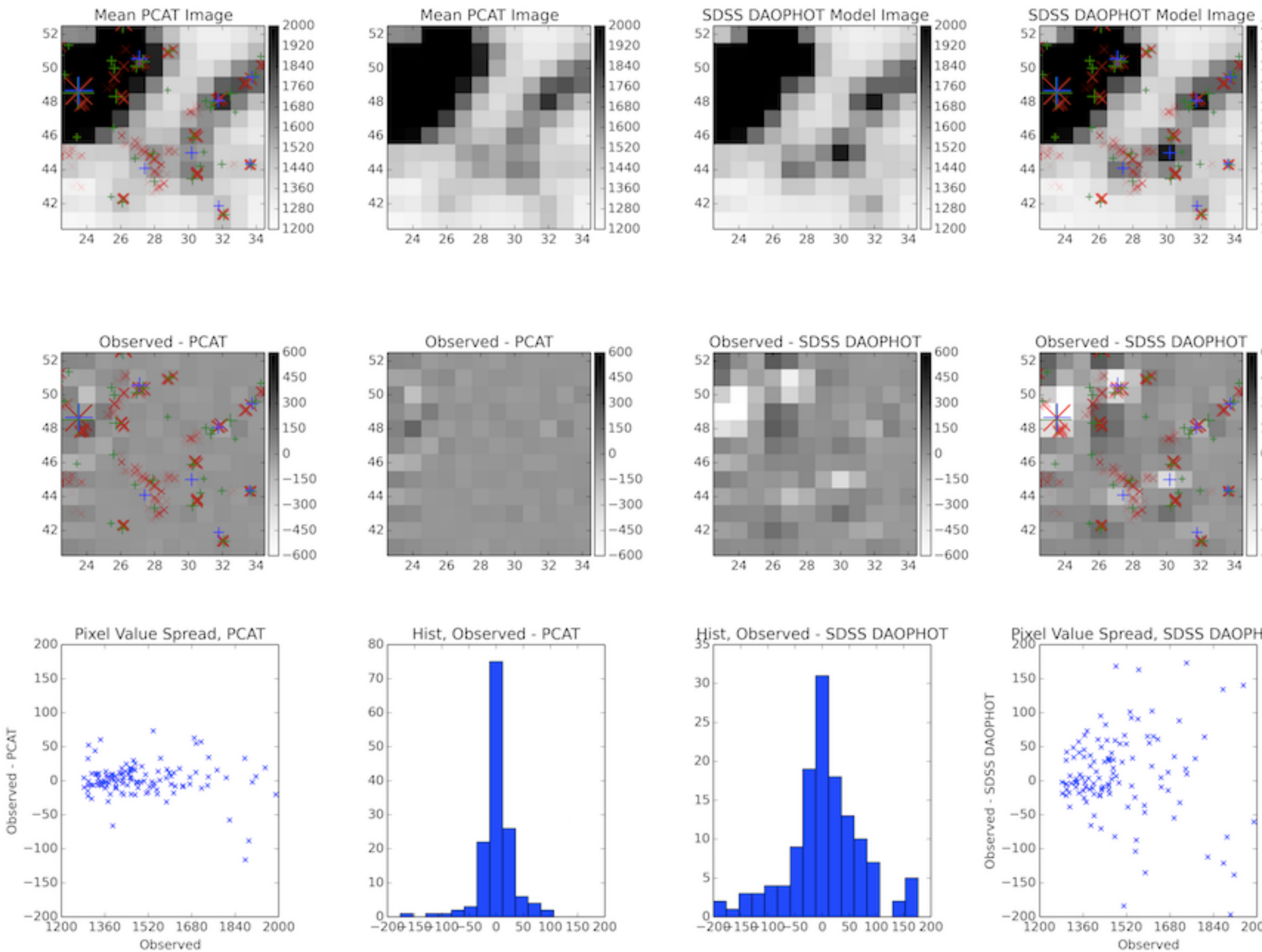

Mean PCAT Image
SDSS DAOPHOT Model Image
Observed - PCAT
Observed - SDSS DAOPHOT
0
20
40
60
80
1000
1200
1400
1600
1800
2000
2200
2400
−200
−160
−120
−80
−40
0
40
80
120
160
200


PCAT
DAOPhot
HST
Mean PCAT Image
40
38
36
34
32
84
86
88
90
92
2000
1920
1840
1760
1680
1600
1520
1440
1360
1280
1200

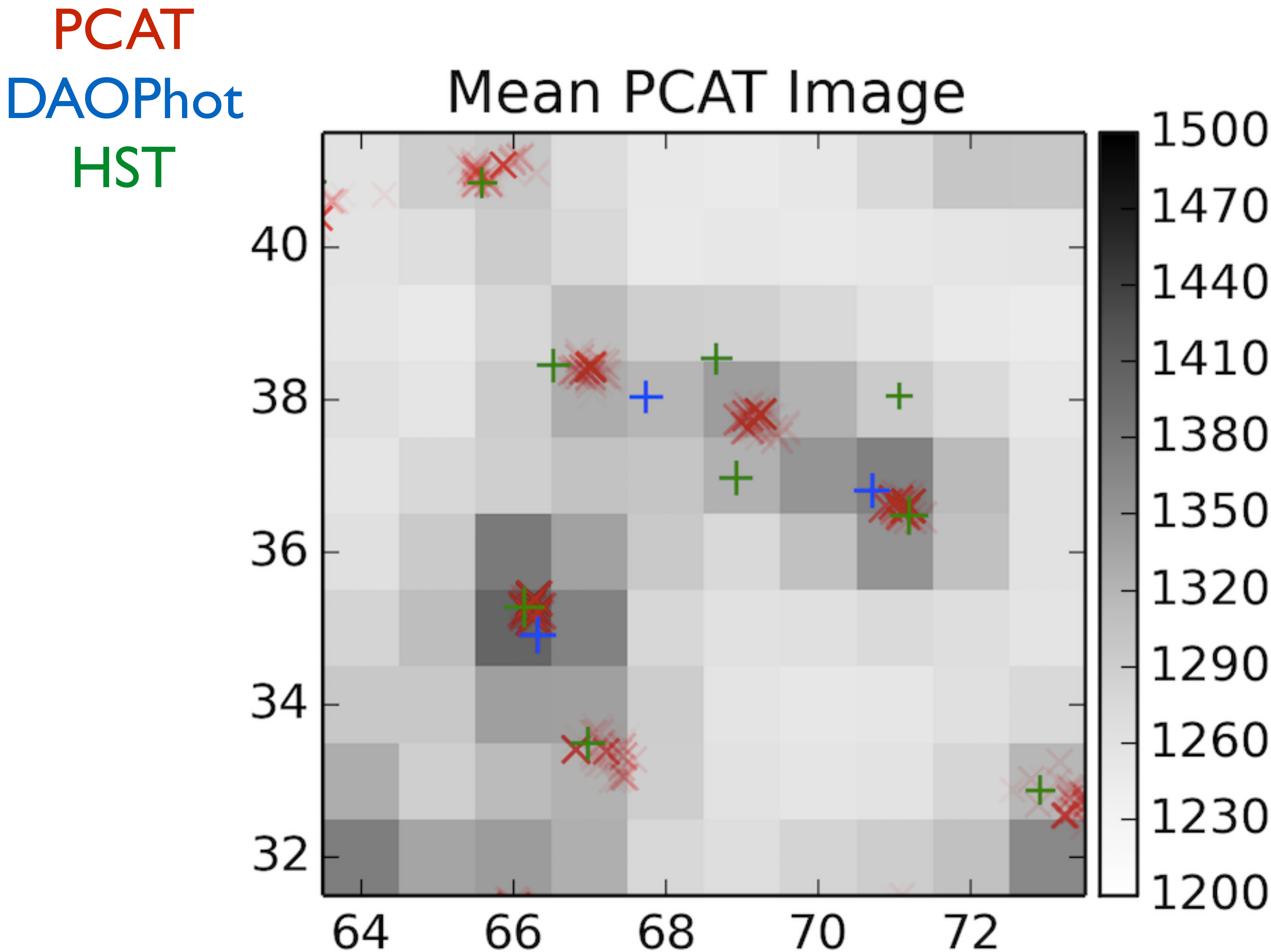

Mean PCAT Image
Mean PCAT Image
SDSS DAOPHOT Model Image
SDSS DAOPHOT Model Image
Observed - PCAT
Observed - PCAT
Observed - SDSS DAOPHOT
Observed - SDSS DAOPHOT
Pixel Value Spread, PCAT
Hist, Observed - PCAT
Hist, Observed - SDSS DAOPHOT
Pixel Value Spread, SDSS DAOPHOT
Observed - PCAT
Observed
Observed - SDSS DAOPHOT
Observed

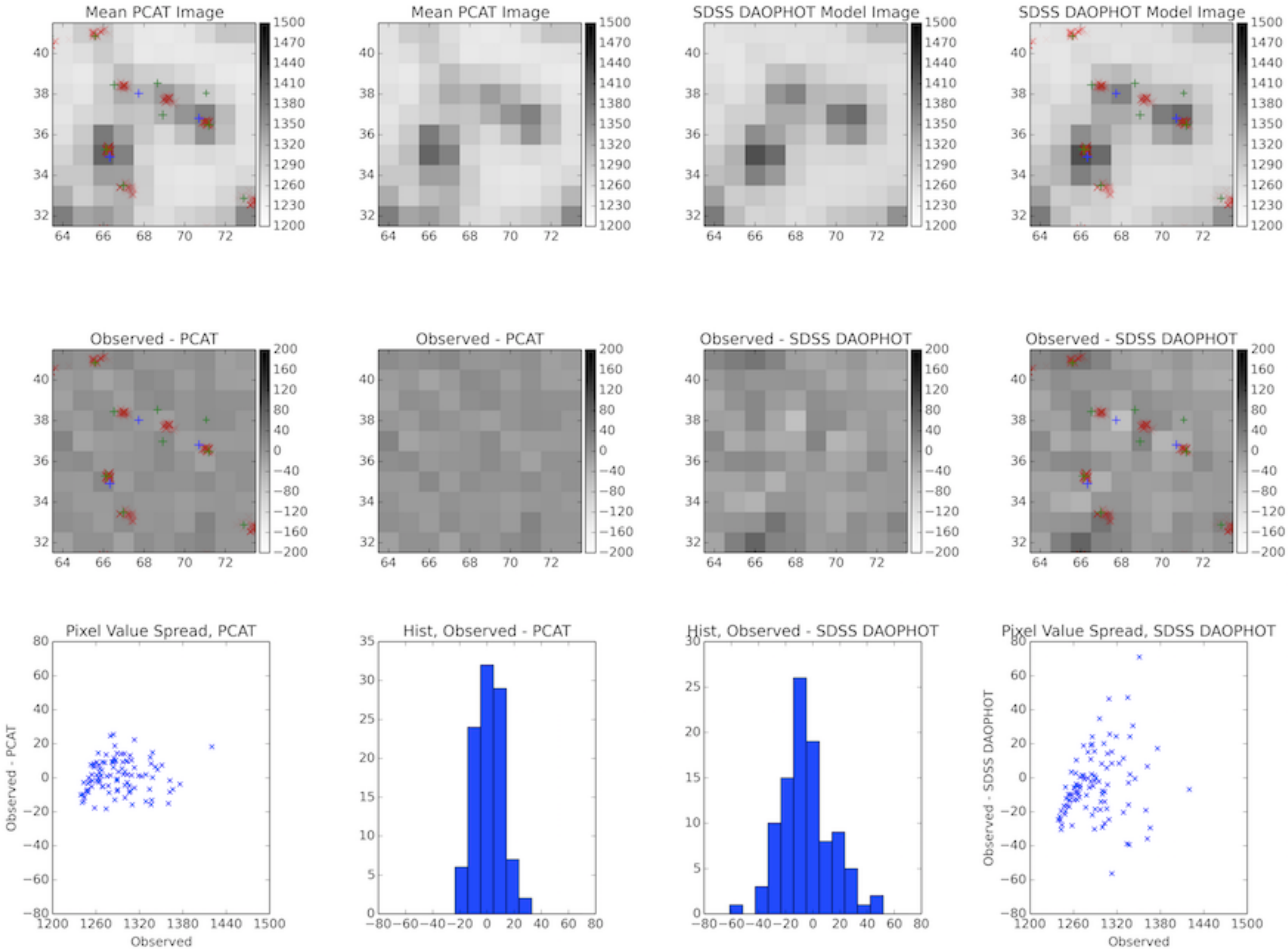
PCAT
DAOPhot
HST



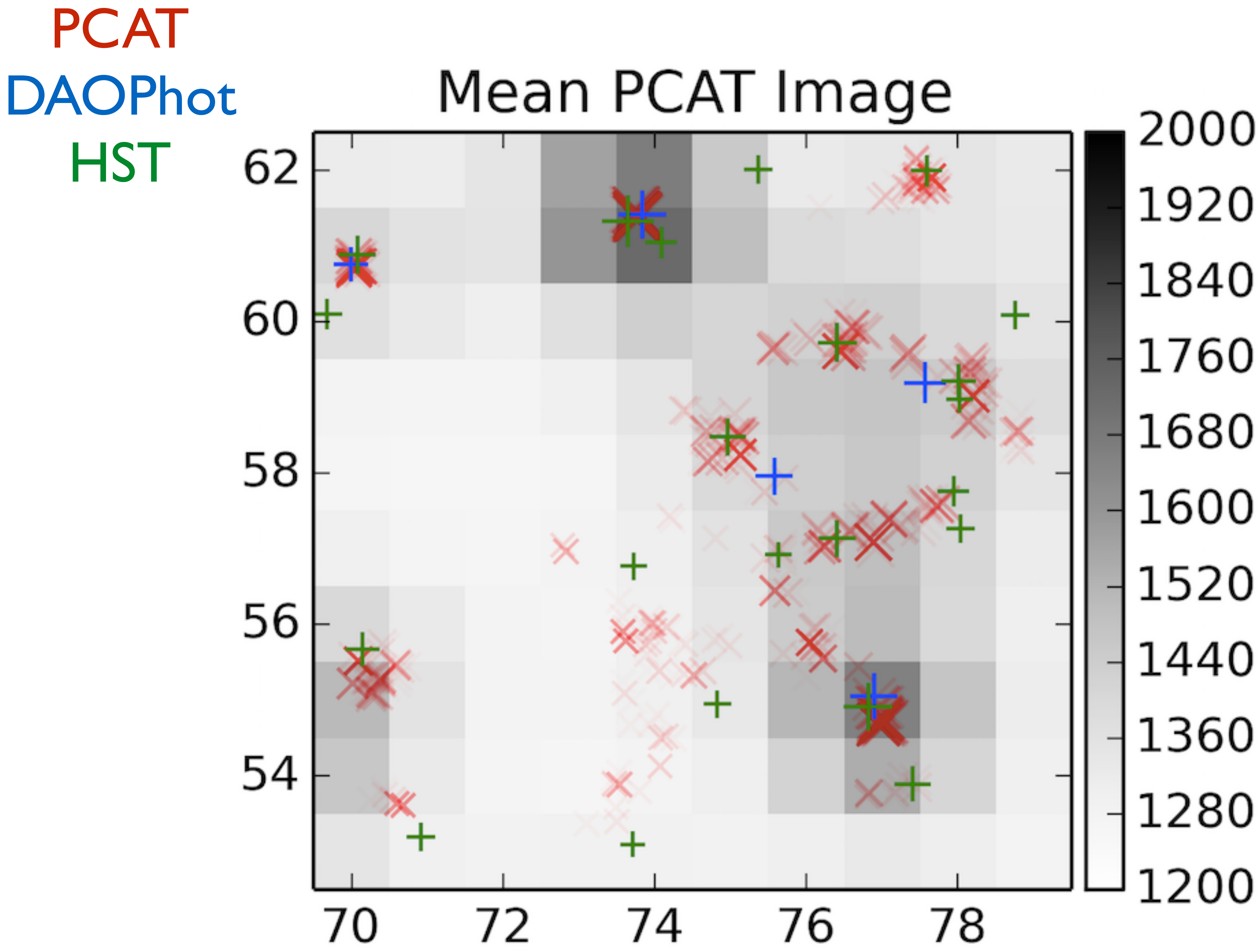


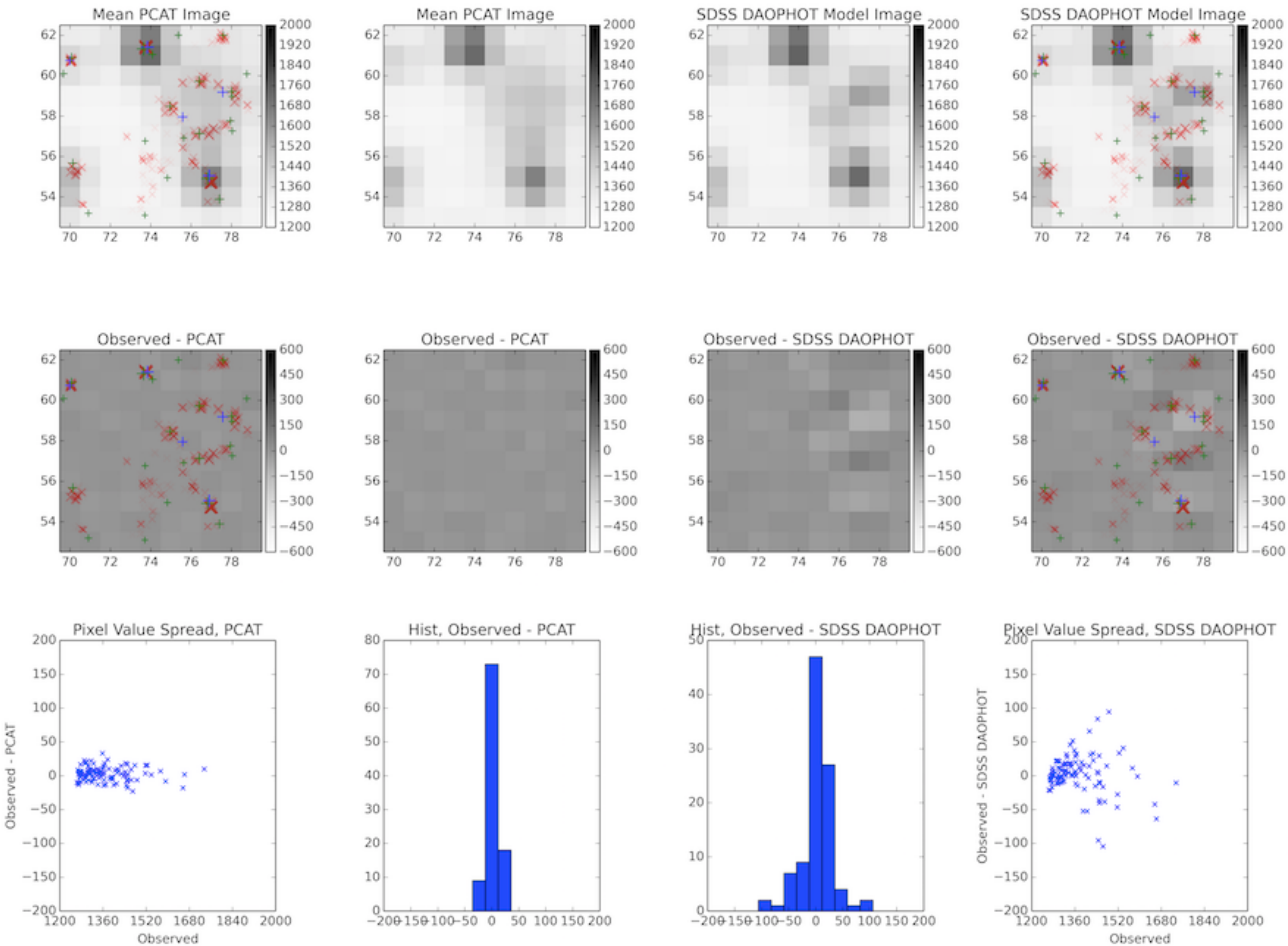
Mean PCAT Image
Mean PCAT Image
SDSS DAOPHOT Model Image
SDSS DAOPHOT Model Image
Observed - PCAT
Observed - PCAT
Observed - SDSS DAOPHOT
Observed - SDSS DAOPHOT
Pixel Value Spread, PCAT
Hist, Observed - PCAT
Hist, Observed - SDSS DAOPHOT
Pixel Value Spread, SDSS DAOPHOT
Observed - PCAT
Observed
Observed - SDSS DAOPHOT
Observed


PCAT
DAOPhot
HST
Mean PCAT Image
40
38
36
34
32
64
66
68
70
72
1500
1470
1440
1410
1380
1350
1320
1290
1260
1230
1200



Mean PCAT Image
Mean PCAT Image
SDSS DAOPHOT Model Image
SDSS DAOPHOT Model Image
Observed - PCAT
Observed - PCAT
Observed - SDSS DAOPHOT
Observed - SDSS DAOPHOT
Pixel Value Spread, PCAT
Hist, Observed - PCAT
Hist, Observed - SDSS DAOPHOT
Pixel Value Spread, SDSS DAOPHOT
Observed - PCAT
Observed
Observed - SDSS DAOPHOT
Observed


PCAT
DAOPhot
HST
Mean PCAT Image
62
60
58
56
54
70
72
74
76
78
2000
1920
1840
1760
1680
1600
1520
1440
1360
1280
1200

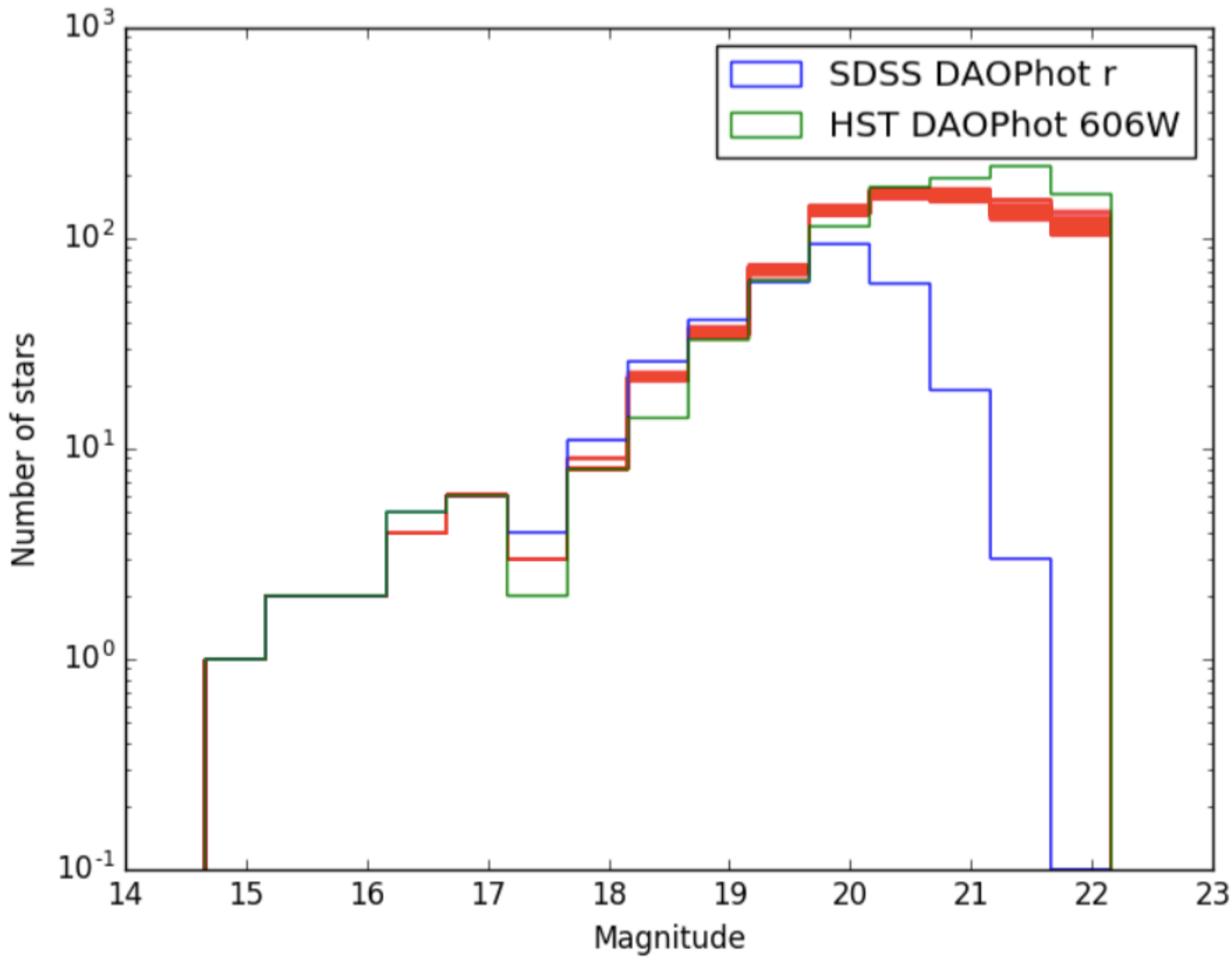


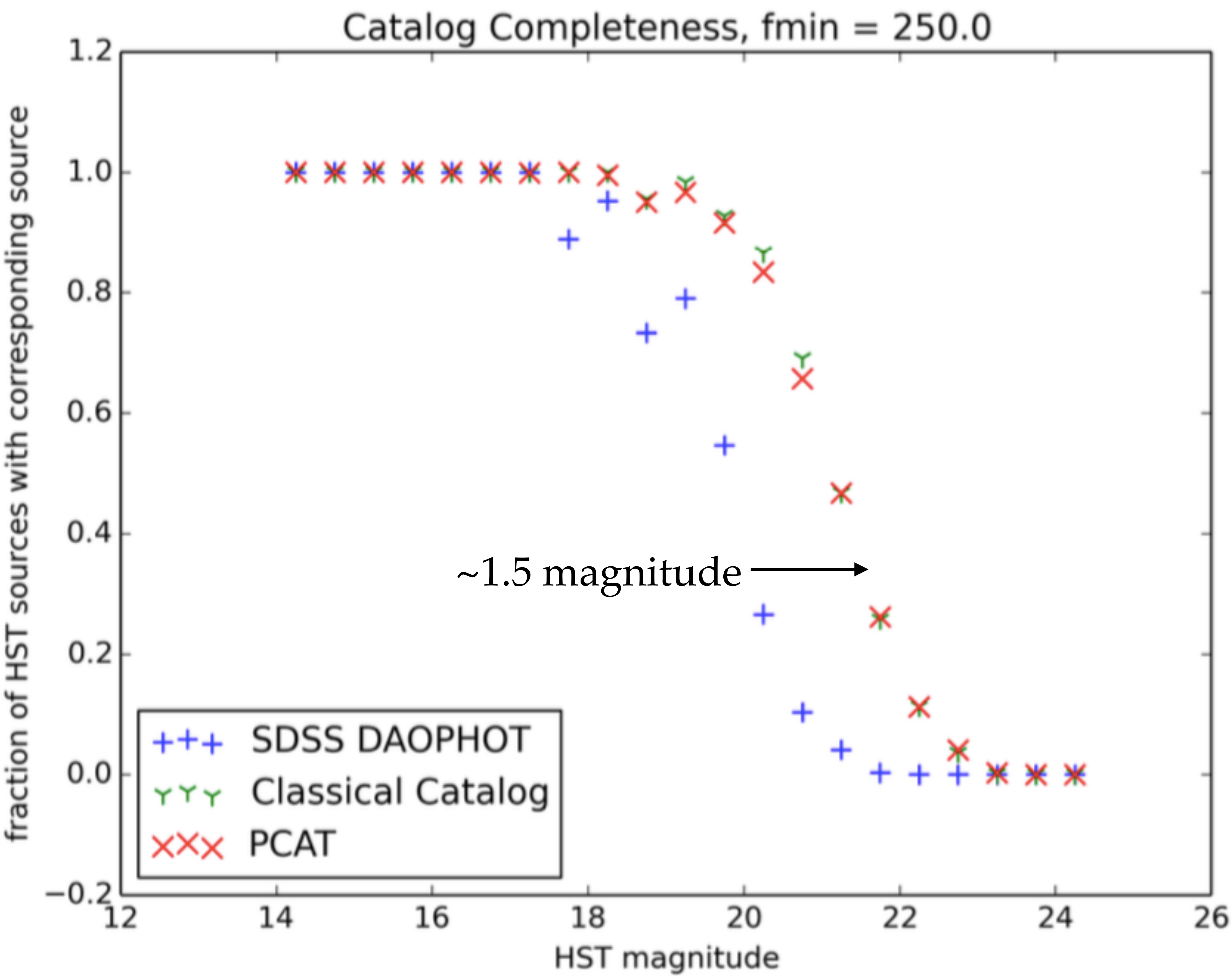
Mean PCAT Image
Mean PCAT Image
SDSS DAOPHOT Model Image
SDSS DAOPHOT Model Image
Observed - PCAT
Observed - PCAT
Observed - SDSS DAOPHOT
Observed - SDSS DAOPHOT
Pixel Value Spread, PCAT
Hist, Observed - PCAT
Hist, Observed - SDSS DAOPHOT
Pixel Value Spread, SDSS DAOPHOT
Observed - PCAT
Observed
Observed - SDSS DAOPHOT
Observed

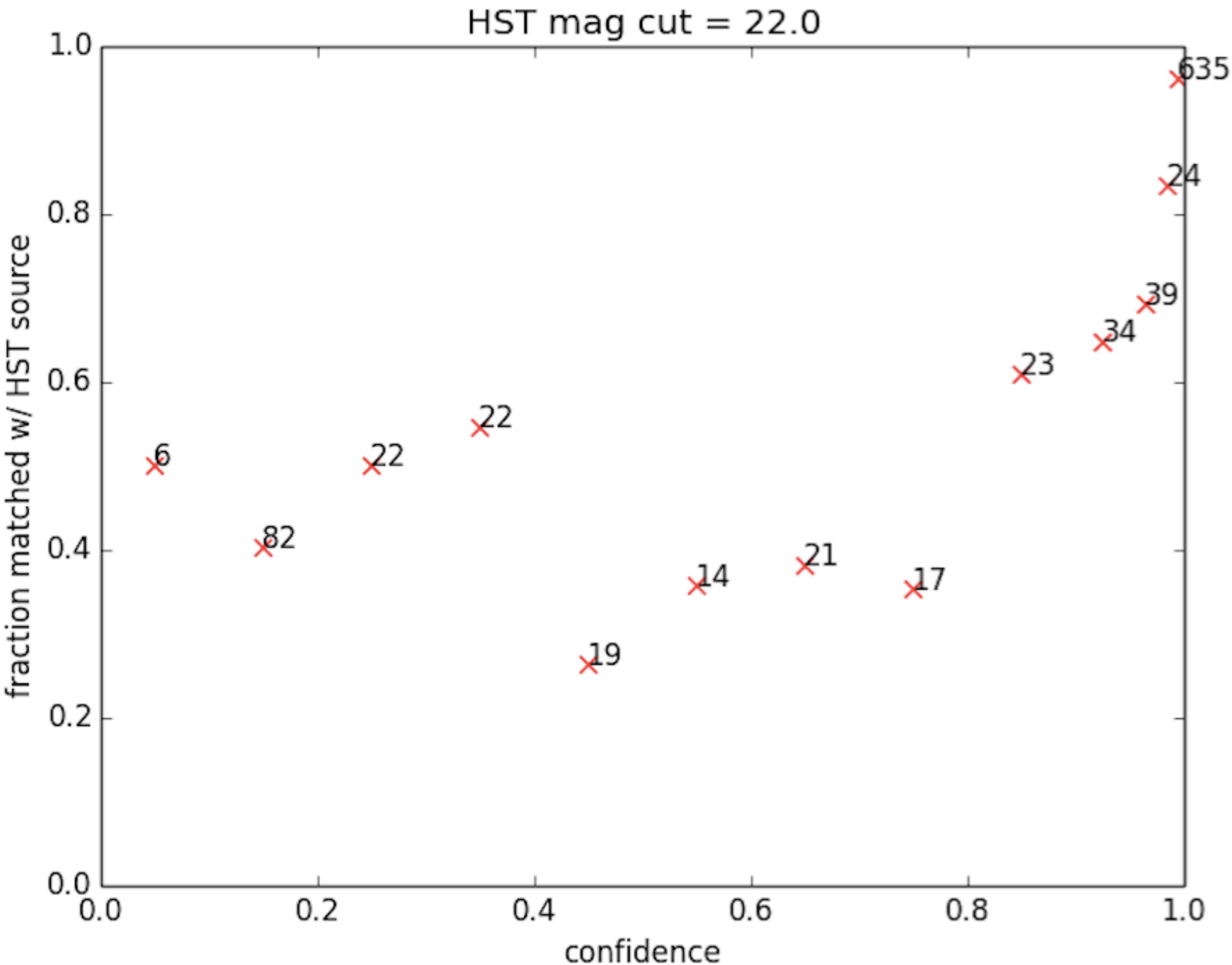
Can we recover a "classical" catalog from this?

- Confidence (% catalog samples with this object)
- {x, y, flux}
- {sigma_x, sigma_y, sigma_flux} (marginalized !)
- sigfac (by what factor is the flux error higher?)

Can do this to compare to other catalogs, e.g. HST.


SDSS DAOPhot r
HST DAOPhot 606W
Number of stars
Magnitude
10^3
10^2
10^1
10^0
10^{-1}
14
15
16
17
18
19
20
21
22
23


Catalog Completeness, fmin = 250.0
fraction of HST sources with corresponding source
HST magnitude
~1.5 magnitude
SDSS DAOPHOT
Classical Catalog
PCAT
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
12
14
16
18
20
22
24
26



HST mag cut = 22.0
fraction matched w/ HST source
confidence
0.0
0.2
0.4
0.6
0.8
1.0
635
24
39
34
23
22
22
6
82
21
14
17
19

Is this too slow to ever use?

We aspire to have it be 1000x as much CPU (in core-seconds per pixel) as the SDSS pipeline. In 2025 or 2030, ~ as much a computational challenge as SDSS was in 2000. (in $$$)

Advantages of a probabilistic (or ensemble) catalog:

- They are explicit about priors and hyper-priors.
- Covariances are embedded in the ensemble.
- Marginalizing over nuisance parameters is trivial.
- Propagation of errors to summary statistic is trivial.

Cons:
- More compute time / storage
- Difficult to interpret as an actual list of sources.

This might be how we will do things in 10 years.

The end