Multiband Probabilistic Cataloging: A Joint Fitting Approach to Point Source Detection and Deblending

Richard M. Feder, Stephen K. N. Portillo, Tansu Daylan, and Douglas P. Finkbeiner

Introduction

While simple in principle, cataloging becomes challenging in the limit of high covariance between neighboring sources and low signal to noise. Traditional cataloging methods perform poorly in crowded fields, and the enhanced depth of future telescope surveys ensures that issues from crowding will become more exacerbated. An alternative is to exploit a transdimensional, hierarchical, and Bayesian framework, *probabilistic cataloging* (PCAT), that samples from the posterior probability distribution of a *metamodel* (union set of emission models with different dimensionality) given observed data. PCAT has been shown to outperform traditional cataloging methods in application to single band optical data (Portillo et al. 2017). This paper demonstrates the advantages and limitations that arise by extending the framework of PCAT to accommodate multiple band photometric datasets.

Generative Model and Likelihood

We express the expected counts at each pixel grid coordinate $(x, y) = (l, m)$ for band b as λ^b_{lm}. This is calculated as a sum of the background in band b, I^b_{sky}, and the contributions of nearby sources:

$$\lambda^b_{lm} = I^b_{sky} + \sum_{i=1}^{N} f_{bi}\mathcal{P}_b(l - x_{bi}, m - y_{bi})$$

The full likelihood is

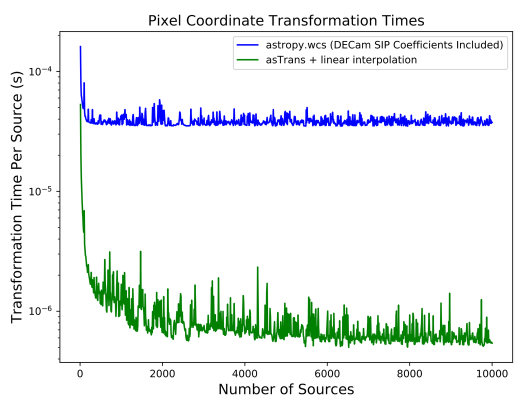
$$\mathcal{L} = \prod_{b=1}^{n_b}\prod_{l=1}^{W}\prod_{m=1}^{H} \frac{1}{\sqrt{2\pi\lambda^b_{lm}}}\exp\left(-\frac{(k^b_{lm} - \lambda^b_{lm})^2}{2\lambda^b_{lm}}\right).$$

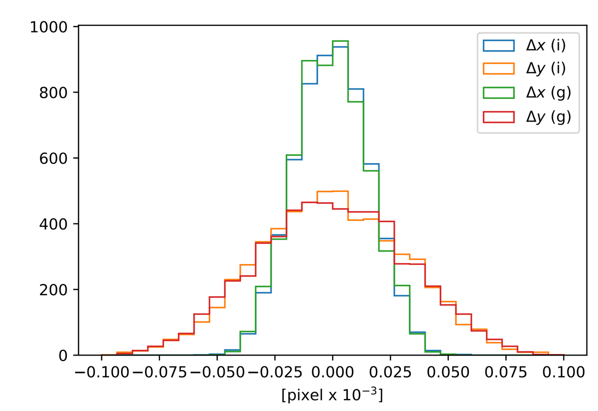
When sampling with RJMCMC, it is much easier to evaluate the log-likelihood, such that our products turn into sums over pixels and bands:

$$\log\mathcal{L} \approx \sum_{b=1}^{n_{bands}}\sum_{l=1}^{w}\sum_{m=1}^{H} -\frac{(k^b_{lm} - \lambda^b_{lm})^2}{2\lambda^b_{lm}}$$

Multi-band Astrometric Calibration

We use linearized astrometric transformations to compute cross-band source positions. We verify that this approximation is accurate to 10^{-4} pixels (right), while nearly two orders of magnitude faster than *astropy.wcs* (left). The linearized transformations use $\sim 5\%$ of PCAT's computation.





While faster approximations can be made, we found in mock tests that astrometric miscalibration at the level of $\sim 10^{-2}$ pixel resulted in oversplitting effects for the brightest sources.

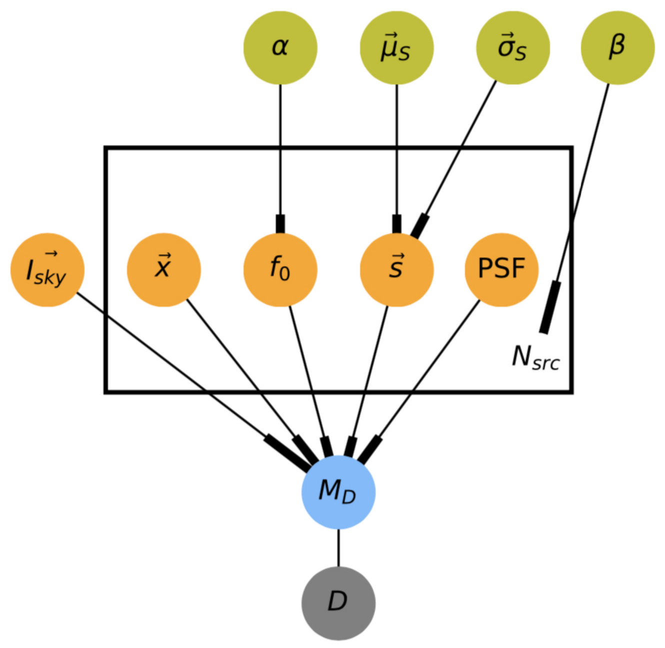
PCAT is a...

Bayesian Hierarchical Model

We infer our catalog by sampling from the posterior of the model given the data, which is obtained through Bayes' rule:

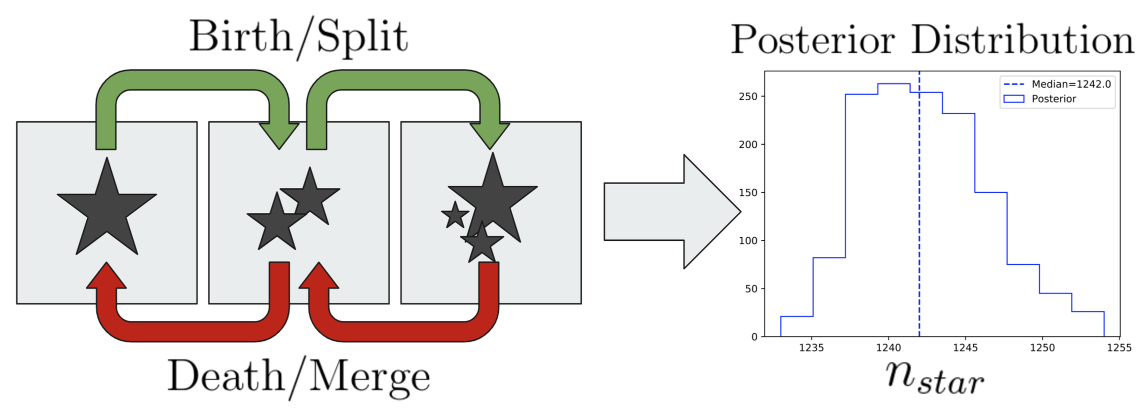
$$P(\vec{\theta}|D) \propto P(\vec{\theta})P(D|\vec{\theta}).$$

We can then construct a hierarchical model, in which individual source parameters are conditioned and marginalized over by higher level parameters, as shown below:



Transdimensional Sampler

By sampling within models of fixed dimension and *across* models of varying dimension with Reversible-Jump Markov Chain Monte Carlo (RJMCMC), we naturally marginalize over source-source covariances and can statistically infer the number of sources in a given image:

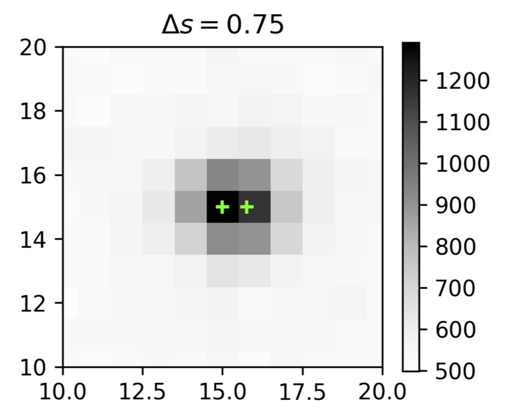


To maintain detailed balance, we impose a parsimony prior on $N_\star$, the number of sources in the image:

$$\pi(N_\star) \propto \exp\left(-\frac{N_\star(\text{d.o.f. per star})}{2}\right) = \exp\left(-\frac{N_\star(2+n_{bands})}{2}\right)$$

Directly Disentangling Point Sources

How well can we deblend highly covariant sources, with and without color information? We test on several blended mock configurations to answer this. **Right:** A mock image of two blended sources with positions given by green crosses.



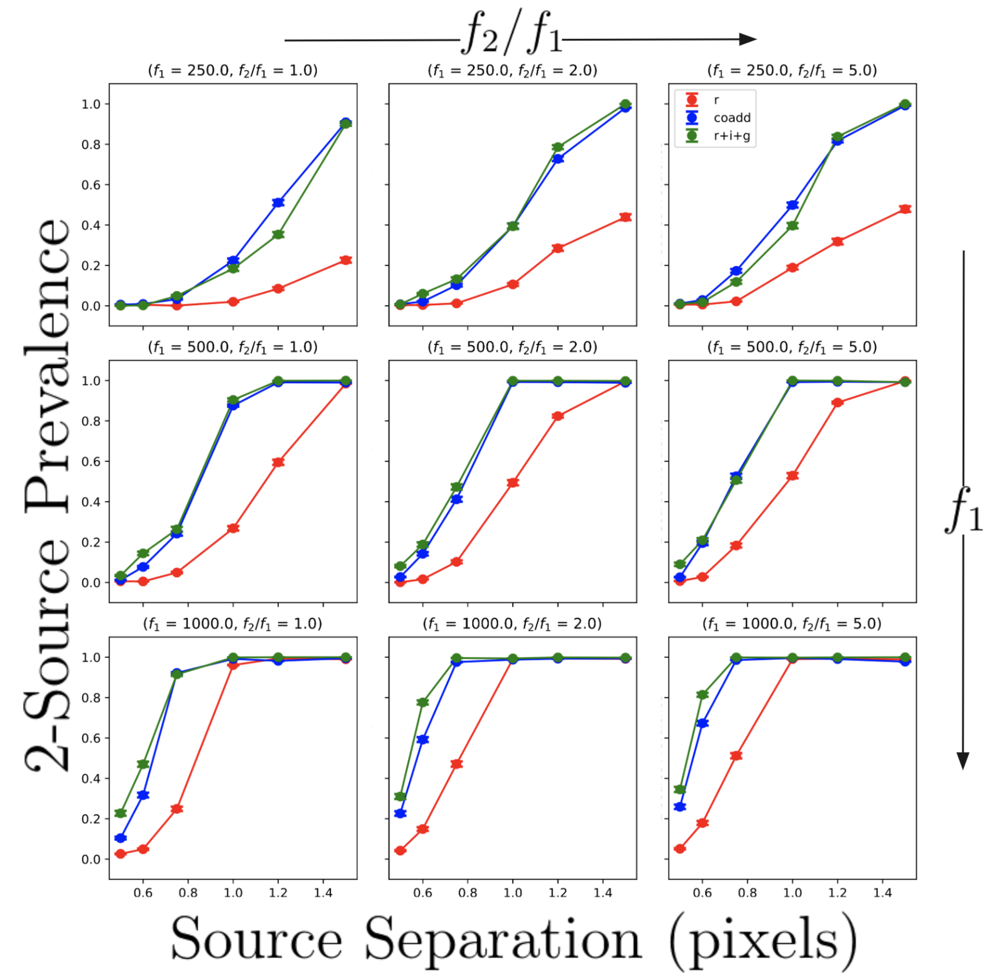


Figure 1: Plots of two-source prevalence as a function of source separation. Prevalence p_n is defined here as the fraction of catalog samples from the Markov chain with n sources. Error bars are standard errors from 10 noise realizations.

Mock Data Tests

Testing on a mock SDSS 100×100 pixel image with 2000 sources, we can assess the impact of additional bands and a weak color prior:

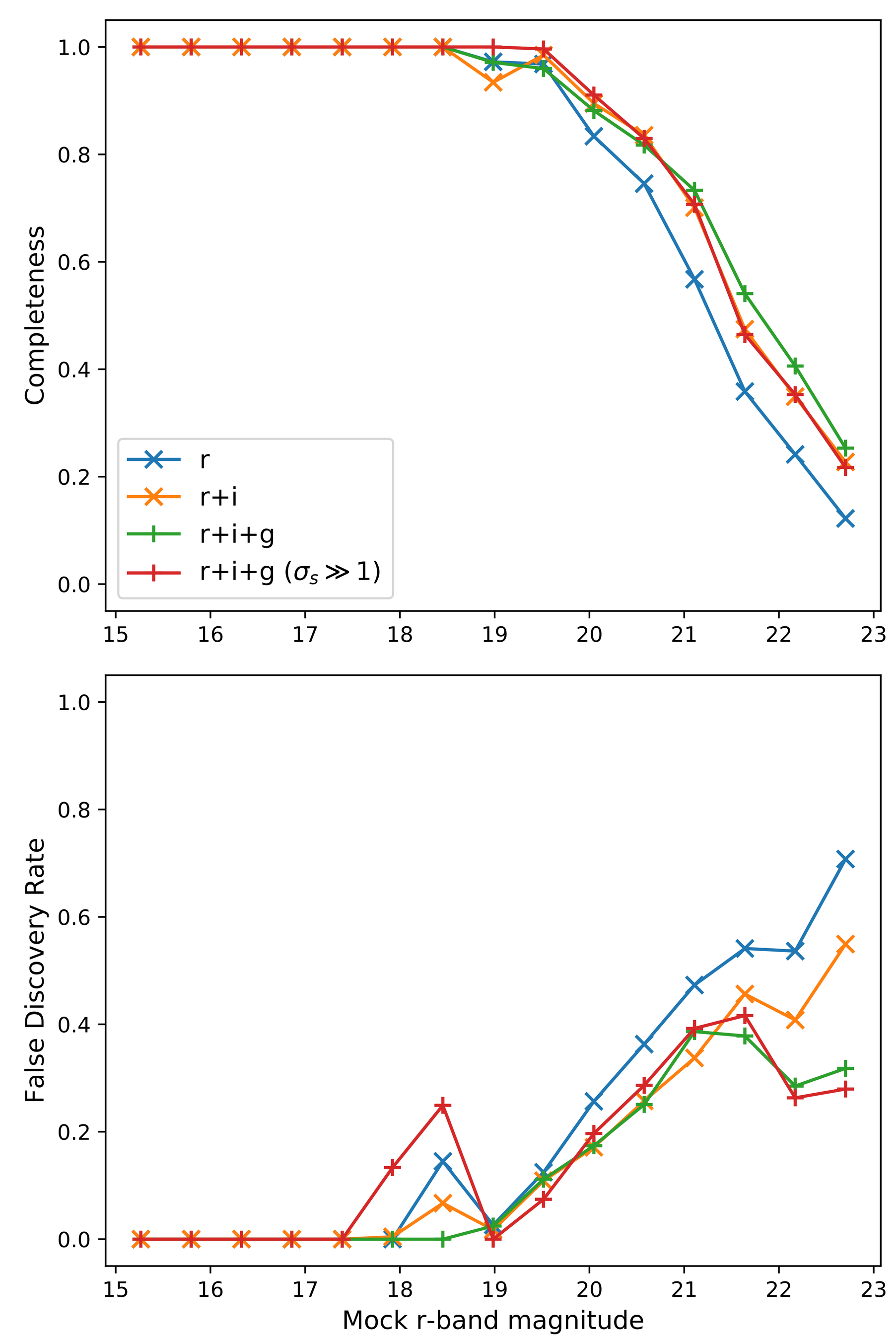


Figure 2: Completeness (top) and false discovery rates (bottom) for one (blue), two (orange), and three band (green, red) probabilistic catalogs.

Color-Magnitude Posteriors on Messier 2

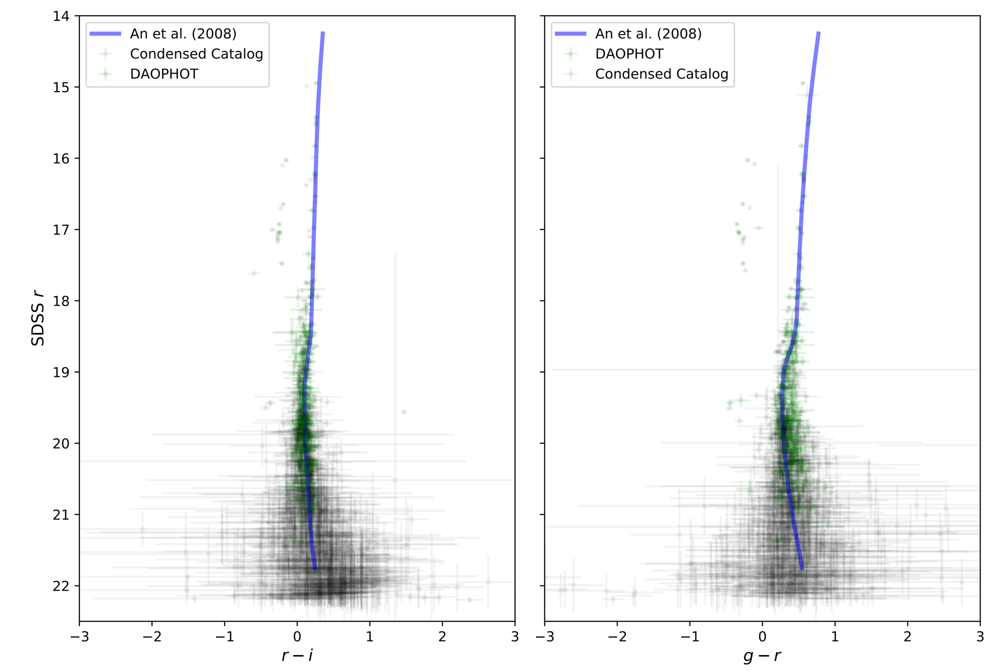
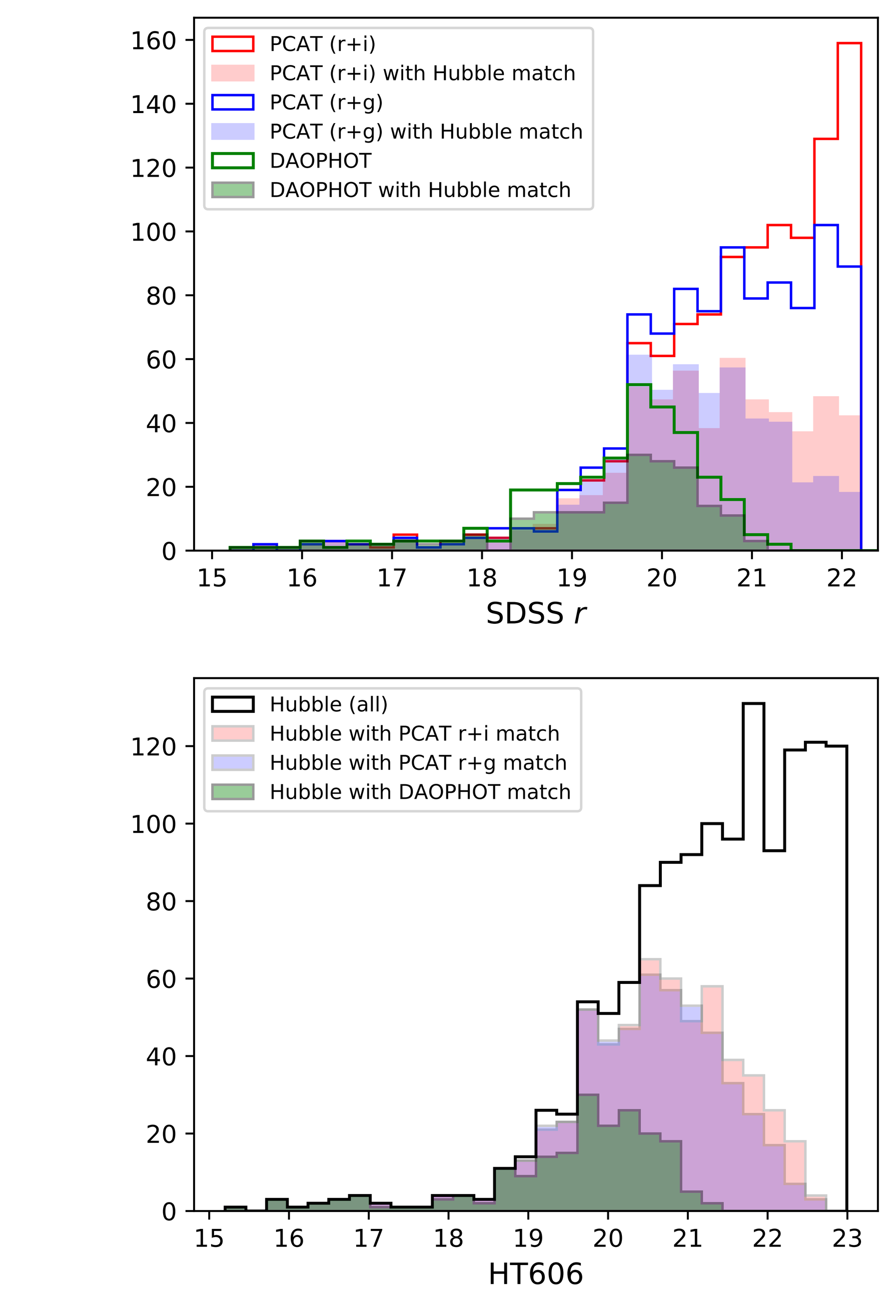


Figure 3: Color magnitude diagrams for $r + i$ band (left) and $r + g$ band (right) condensed catalogs. Error bars on the condensed catalog come from marginalizing over catalog ensemble samples. Green points mark DAOPHOT catalog sources. Plotted in blue is the fiducial sequence for the full cluster M2 obtained by An et al. (2008).

Results on SDSS Messier 2 Globular Cluster

We test our method on a 100×100 SDSS image of the globular cluster Messier 2. Using Hubble as "ground truth", we find that PCAT goes 1.5 magnitudes deeper than DAOPHOT *using the same data*, and has a lower false discovery rate for $r \leq 20.5$.



References

Hogg & Lang (2010) (1008.0738)
Brewer et al. (2014) (1411.3921)
Portillo et al. (2017) (1703.01303)
Daylan et al. (2017a) (1607.04637)
Daylan et al. (2017b) (1706.06111)
An et al. (2008) (ApJ Supp 179)
Green, P. (1995) (Biometrika, 82-4)
Sarajedini et al. (2007) (AJ 133-1658)