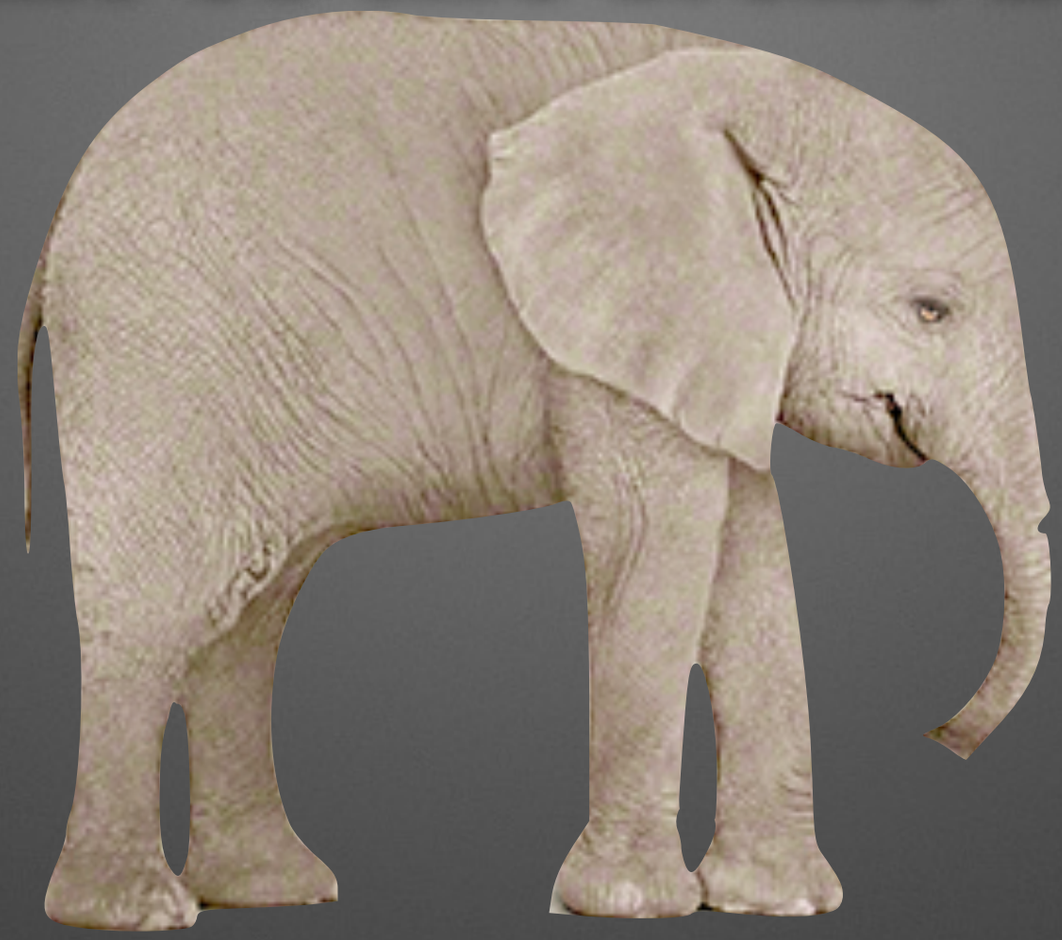
In Pursuit of the Next Big Thing



Adam A Miller
JPL/Caltech

2016 Hubble Symposium
15 March 2016

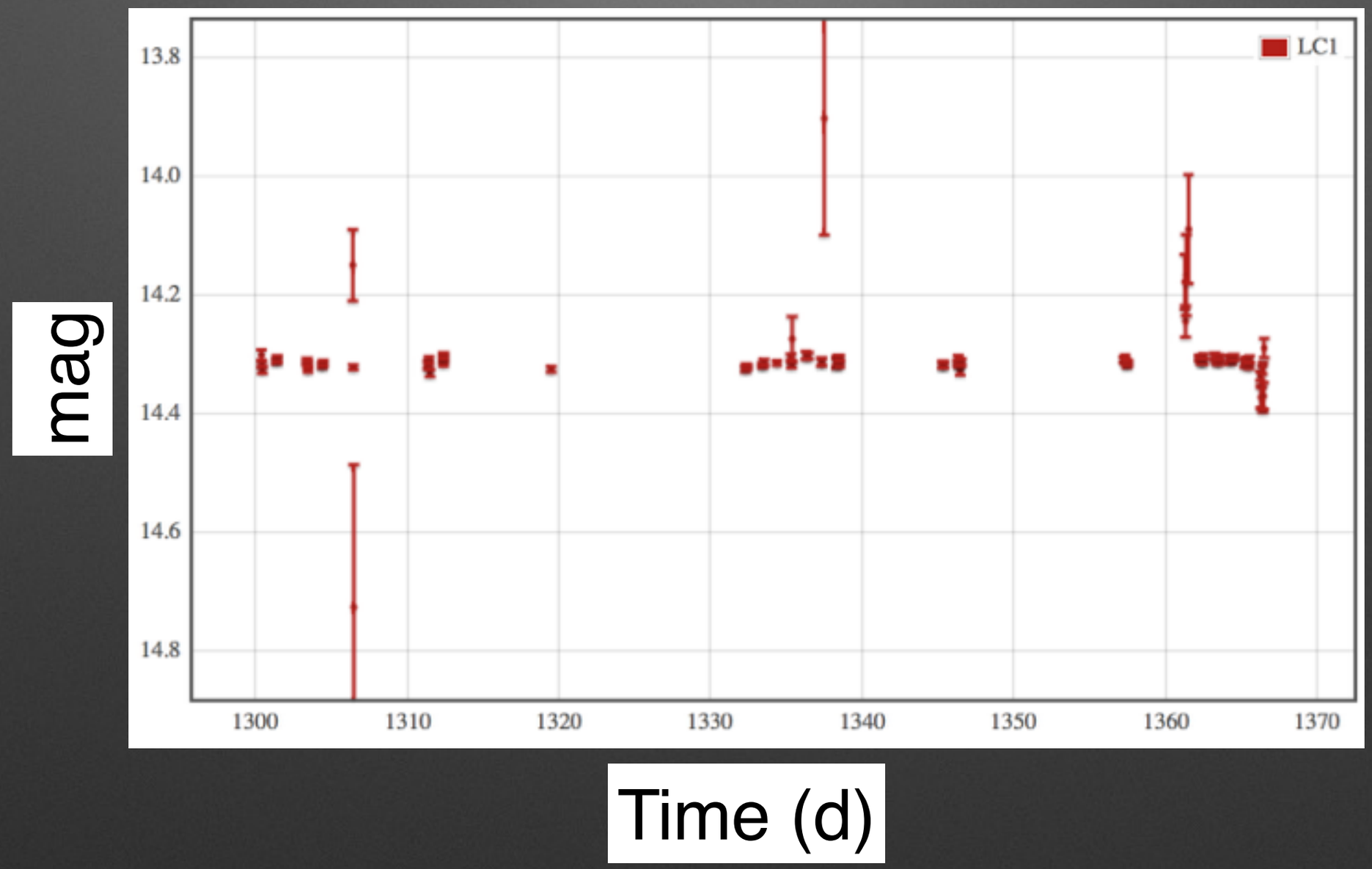
Jet Propulsion Laboratory
California Institute of Technology

Copyright 2016 California Institute of Technology.
U.S. Government sponsorship acknowledged.

It’s much easier to teach an astronomer
about machine learning,
than it is to teach astronomy
to a data-miner.

— Sanjay Chawla, Prof. Statistics

Heteroskedastic Errors



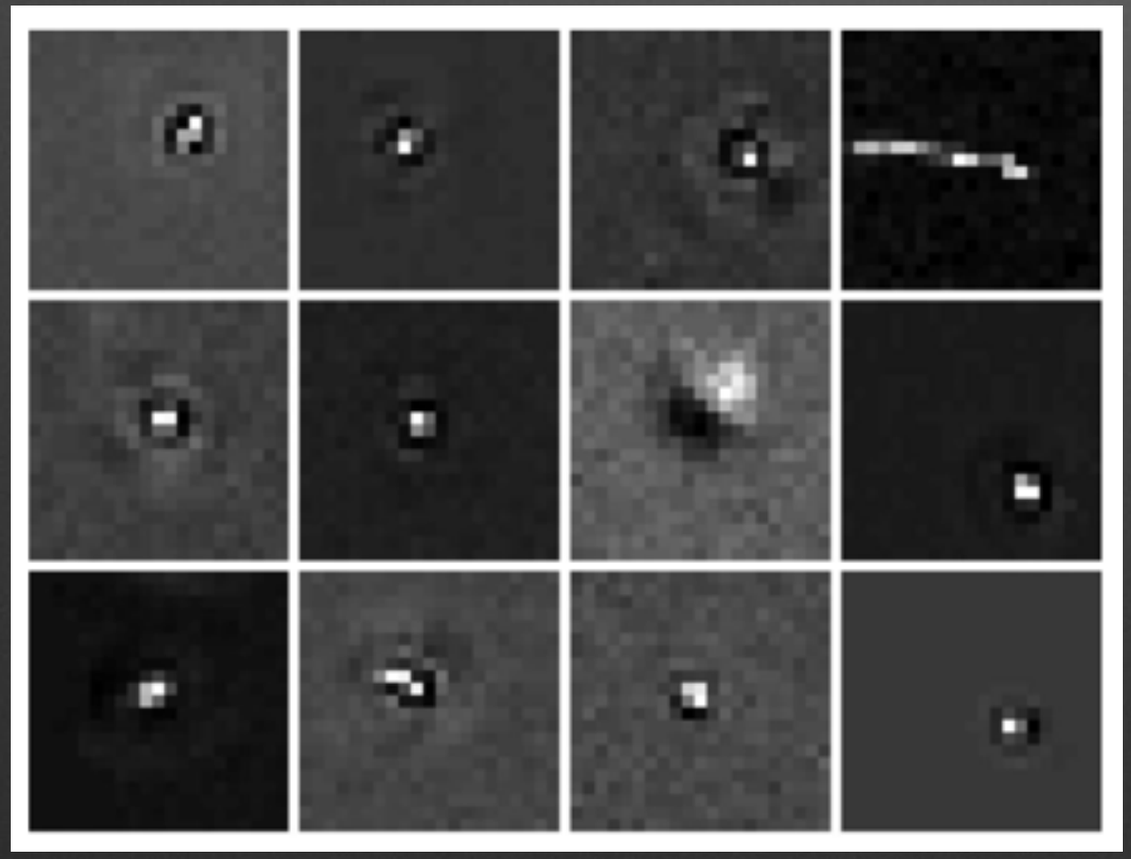
Faint Objects

Faint Objects

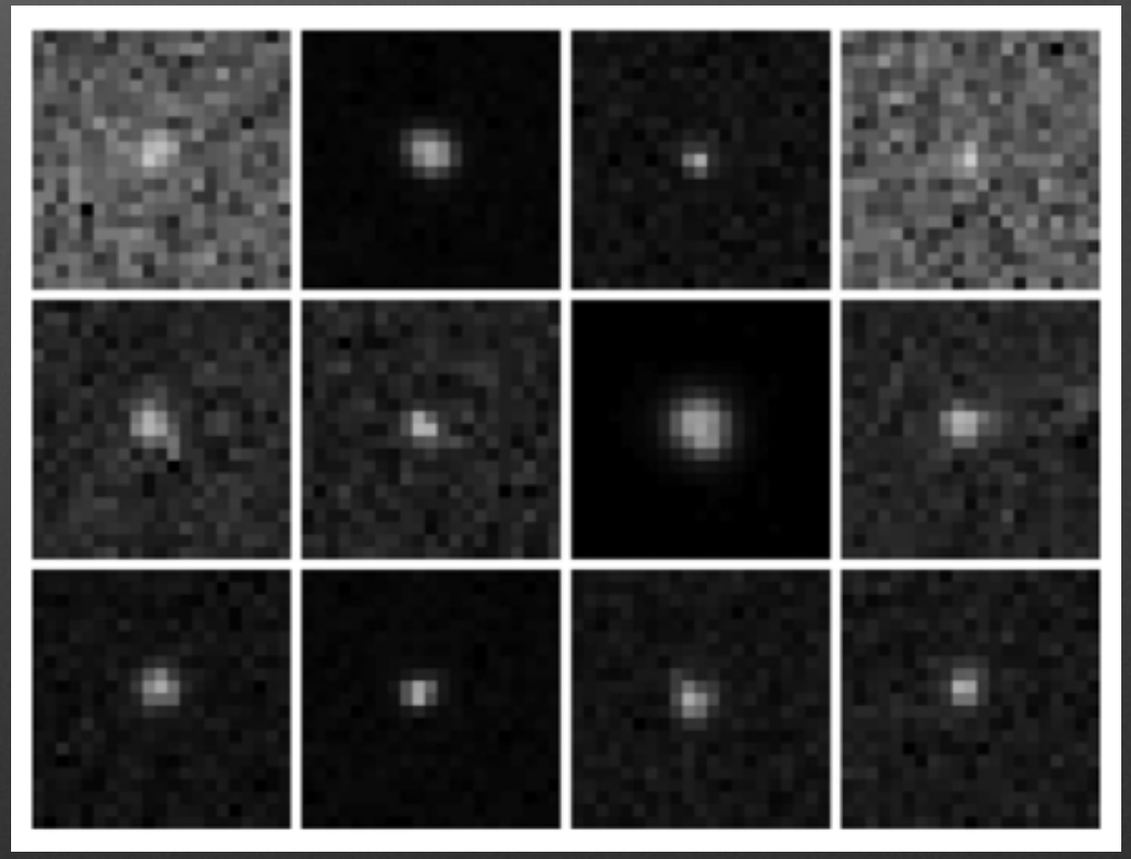
ML success: Real-Bogus

- PTF - 1M transient candidates/night
- ~1k real variable sources/night
- Lots of human hours needed (wasted?) filtering candidates

1M of these



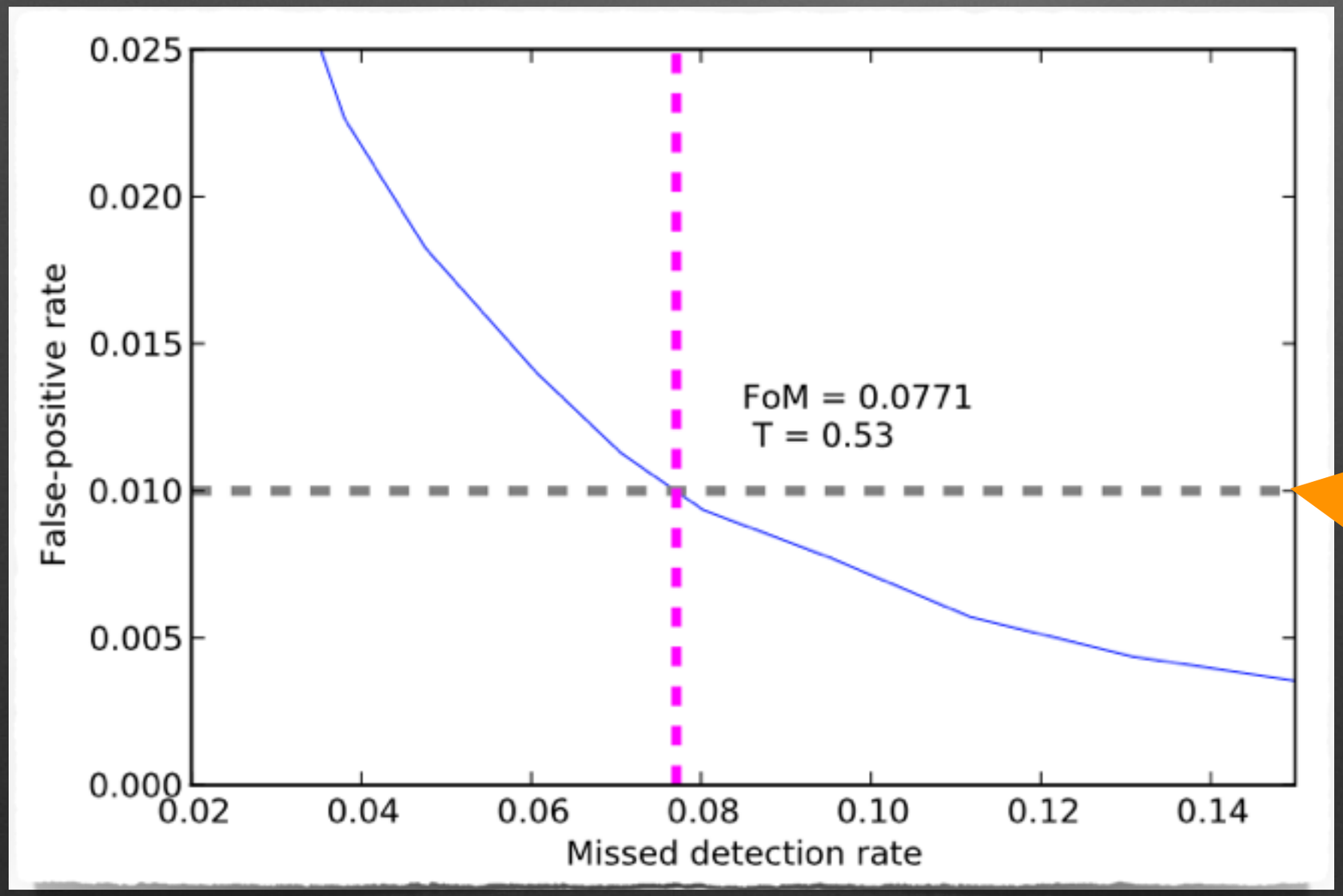
only ~1k of these



Brink+13

ML success: Real-Bogus

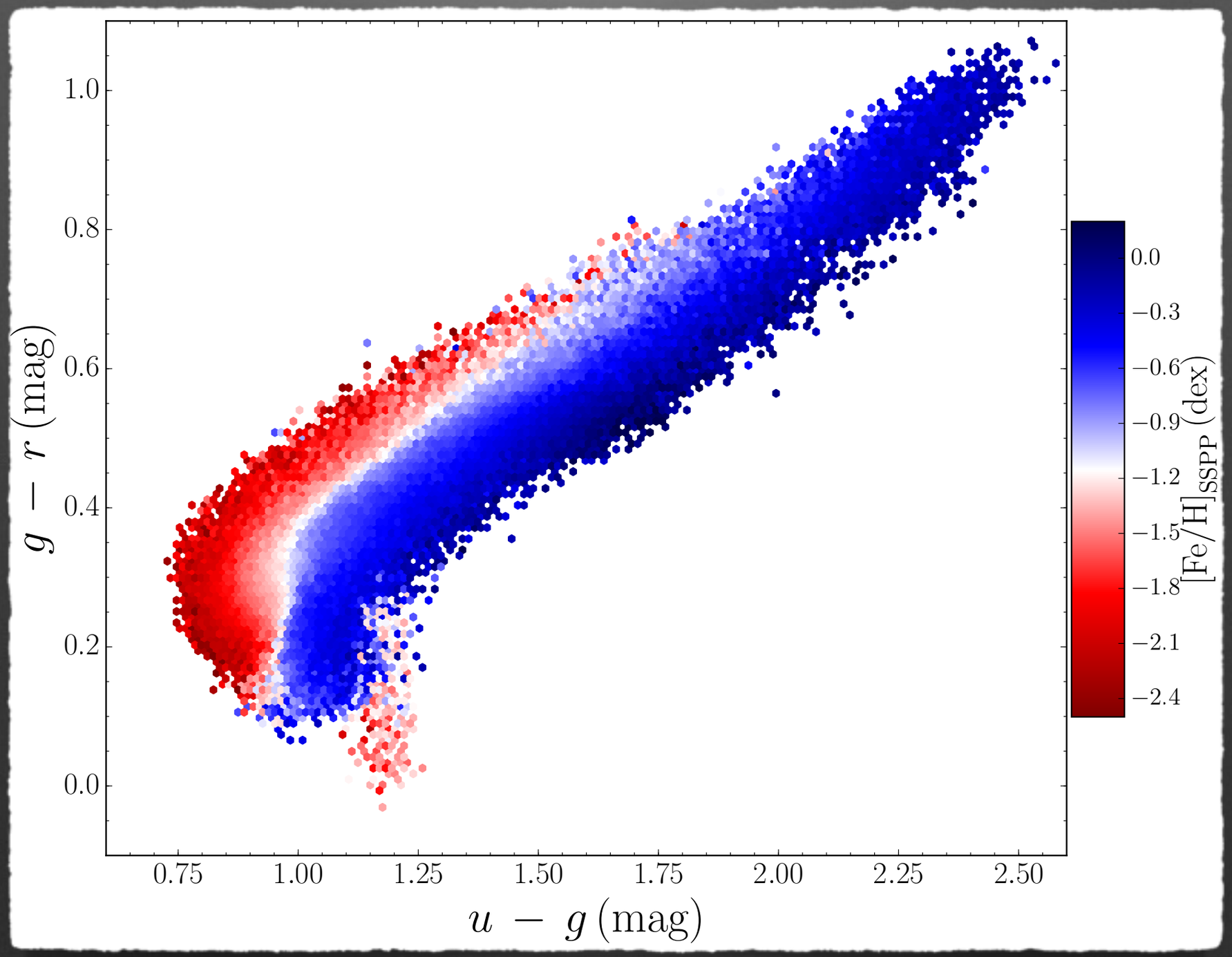
- Determine acceptable N_{cand} for human vetting
- Optimize model to detect max(real) within N_{cand}



Model tuned for ~1% FPs

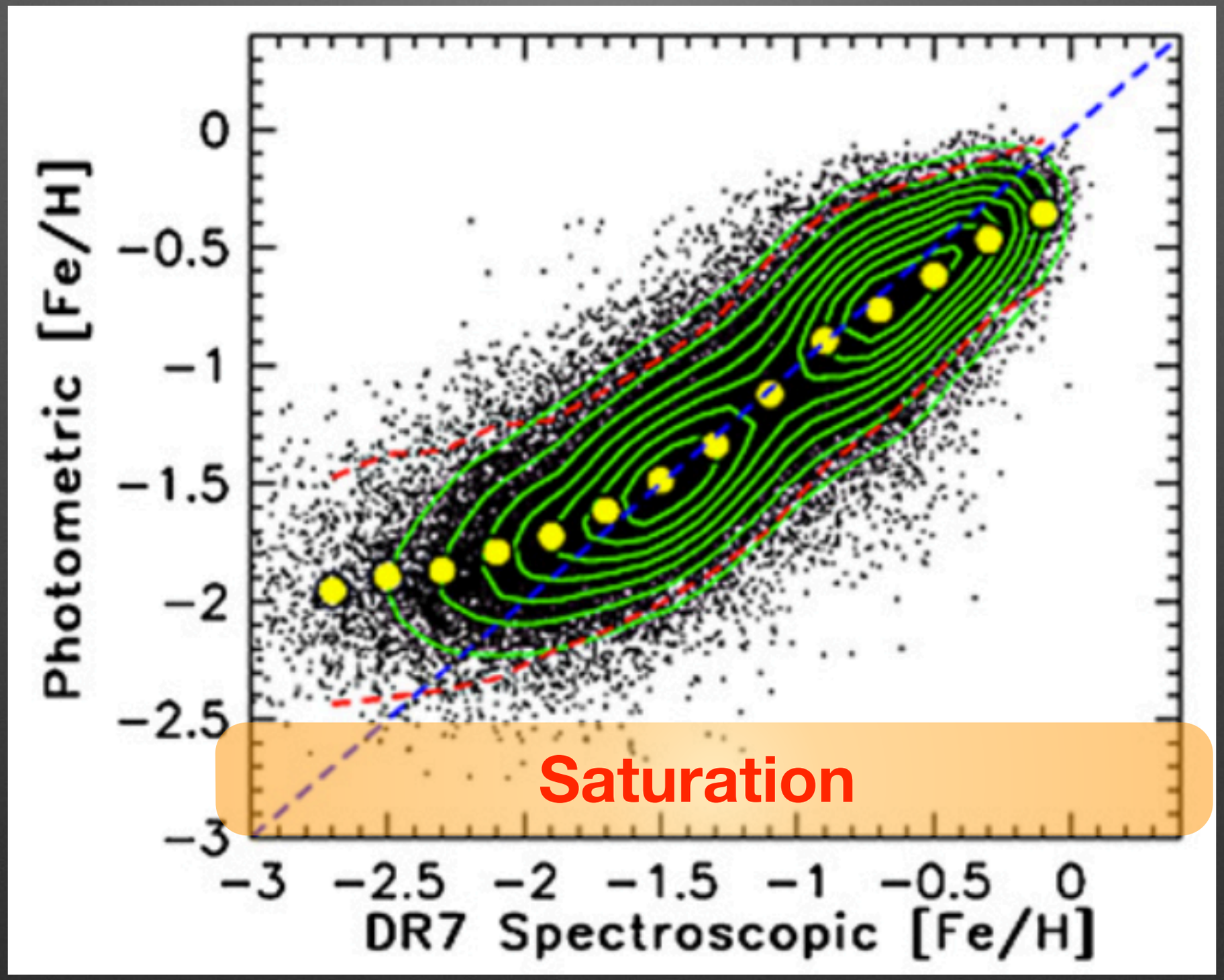
Brink+13

ML success(?): EMP stars



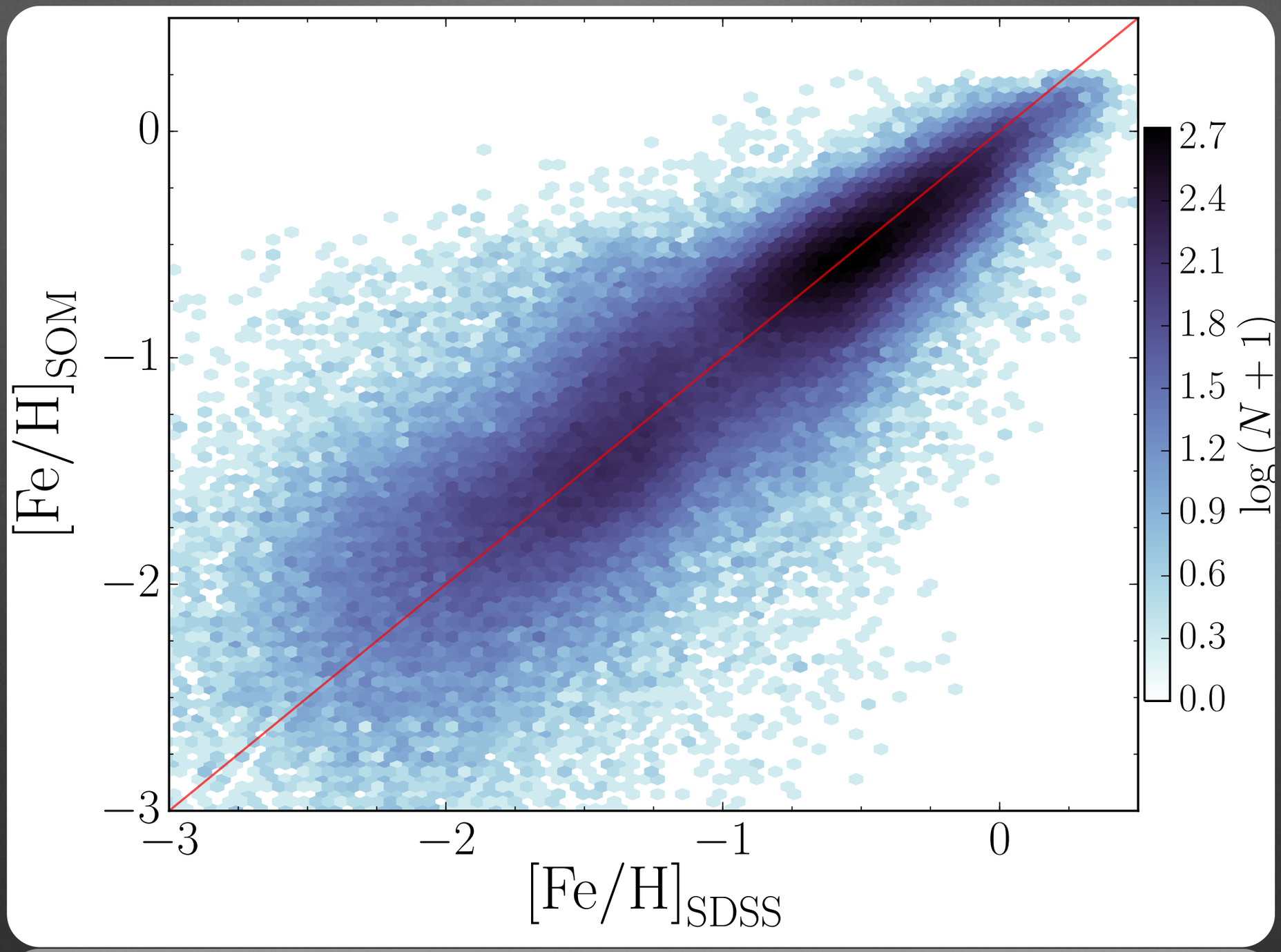
Miller 15

ML success(?): EMP stars



Bond+10

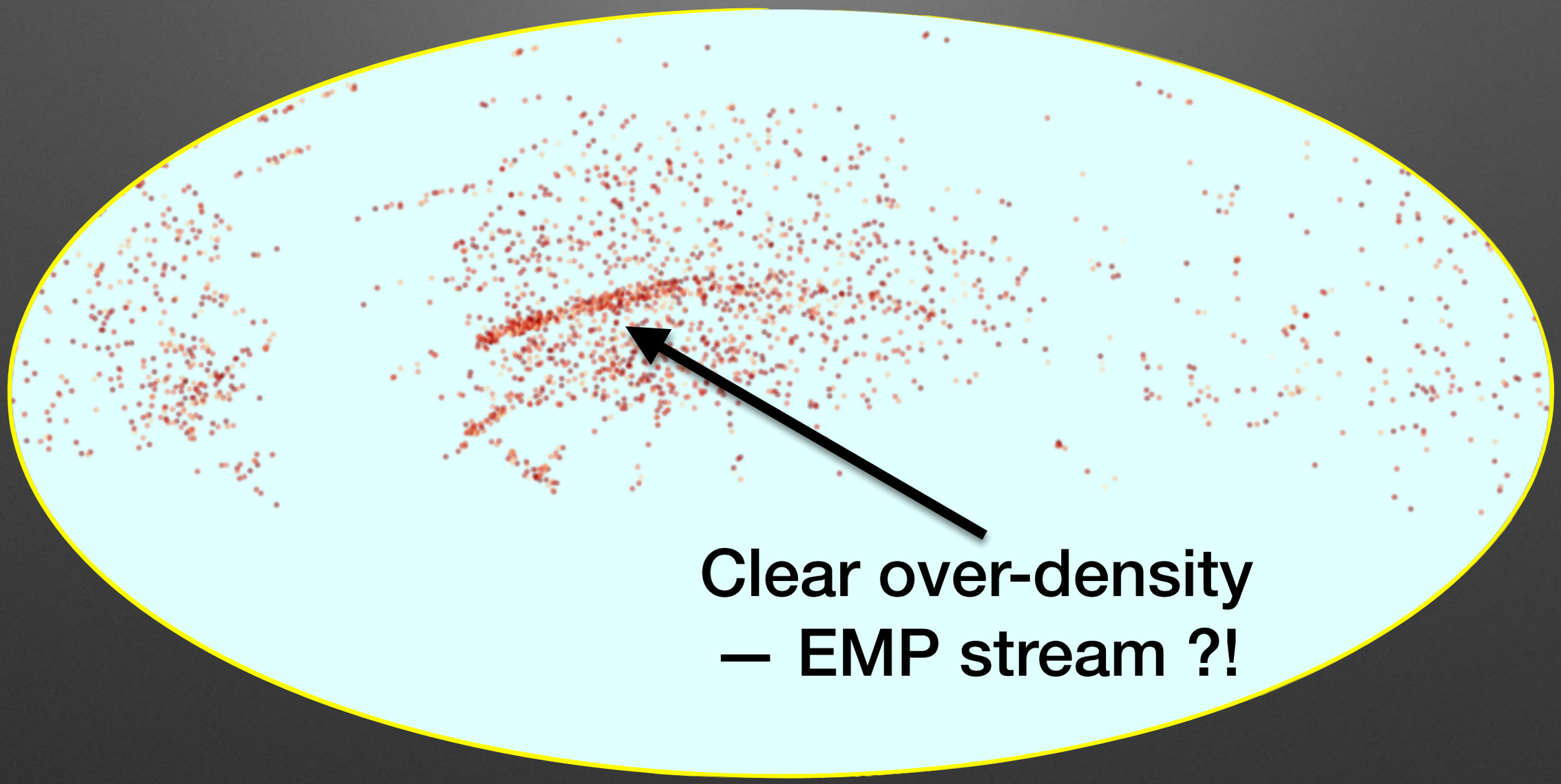
ML success(?): EMP stars



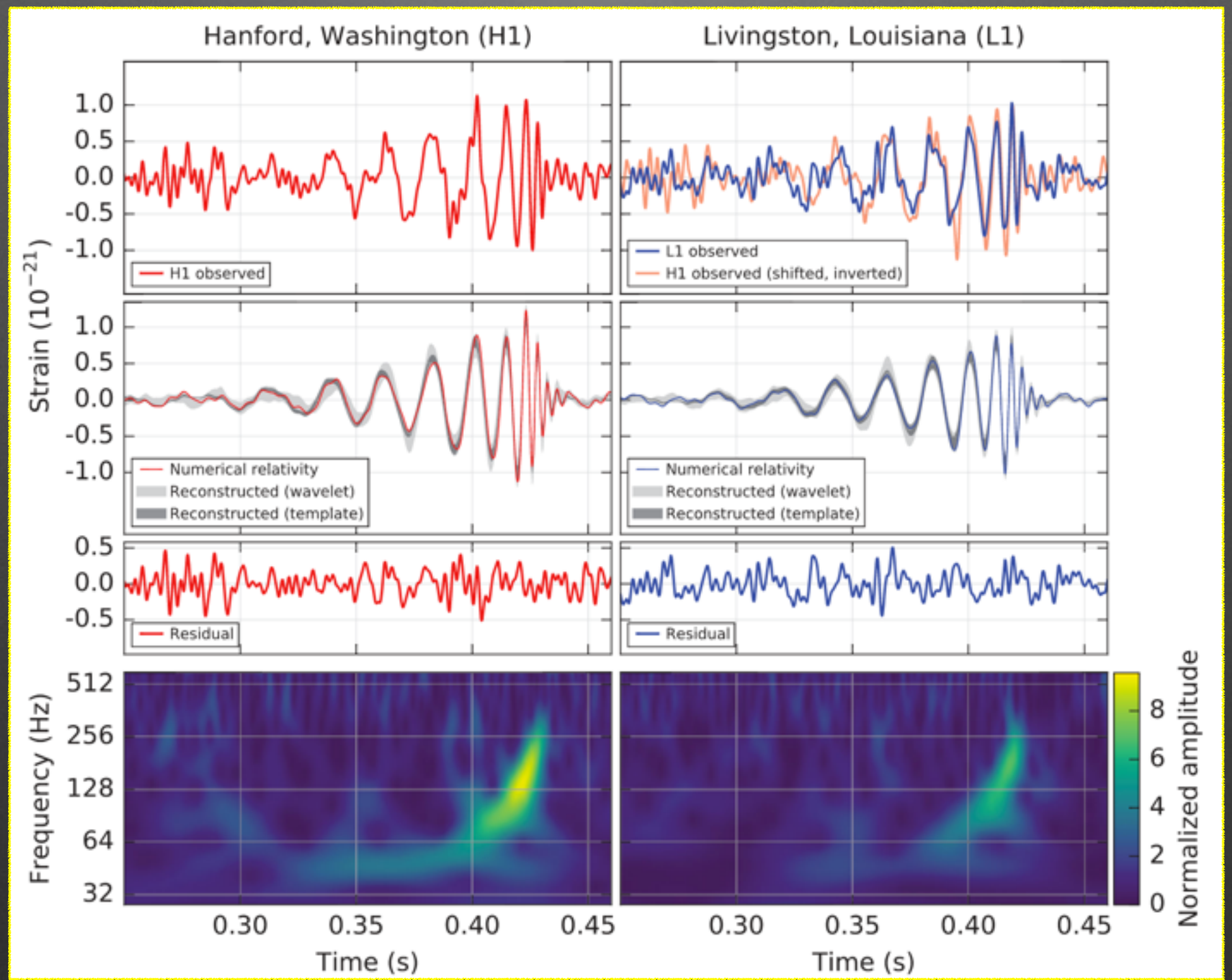
Miller 15

ML success(?): EMP stars

Map location of highest confidence EMP candidates

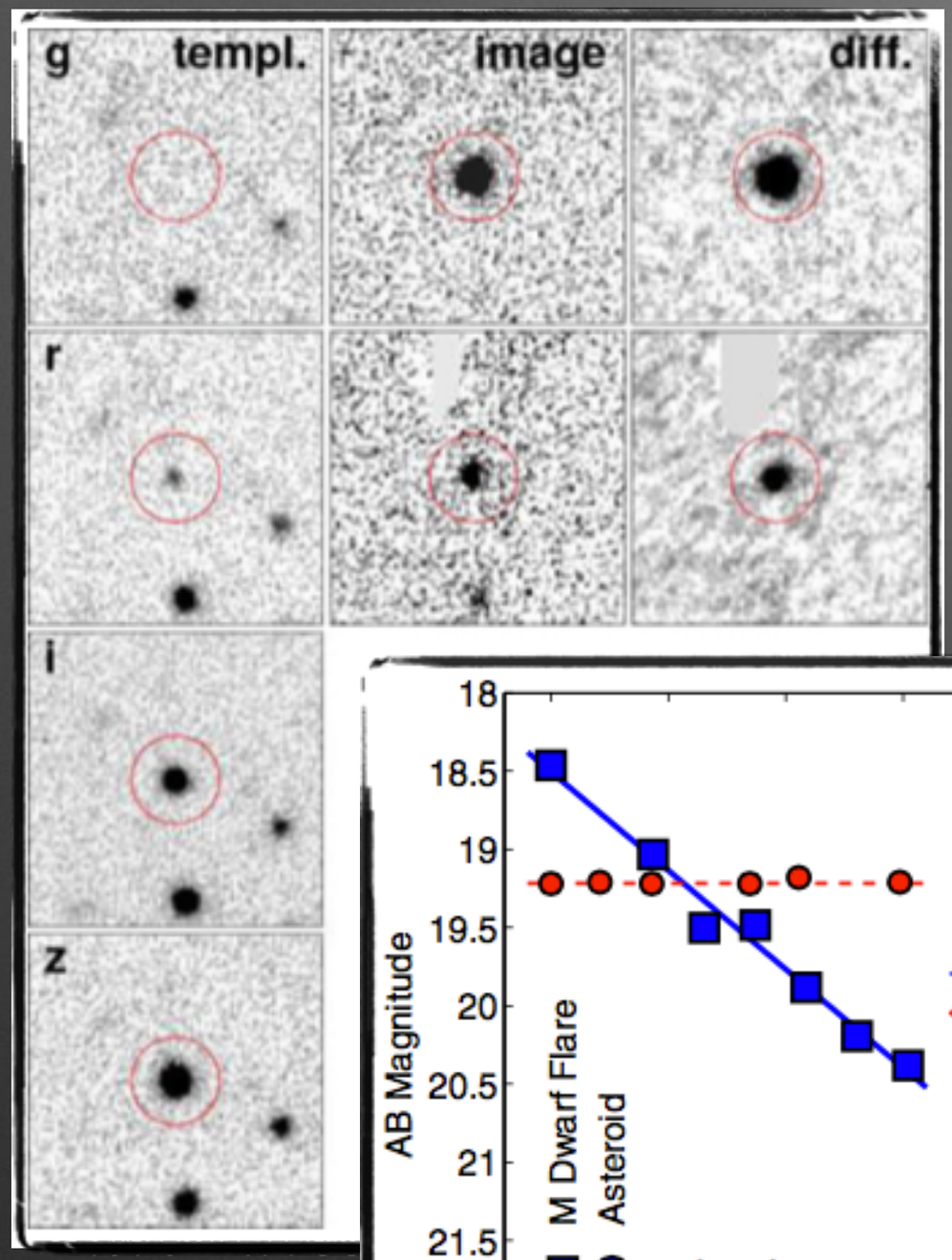


ML success(?): star/galaxy separation

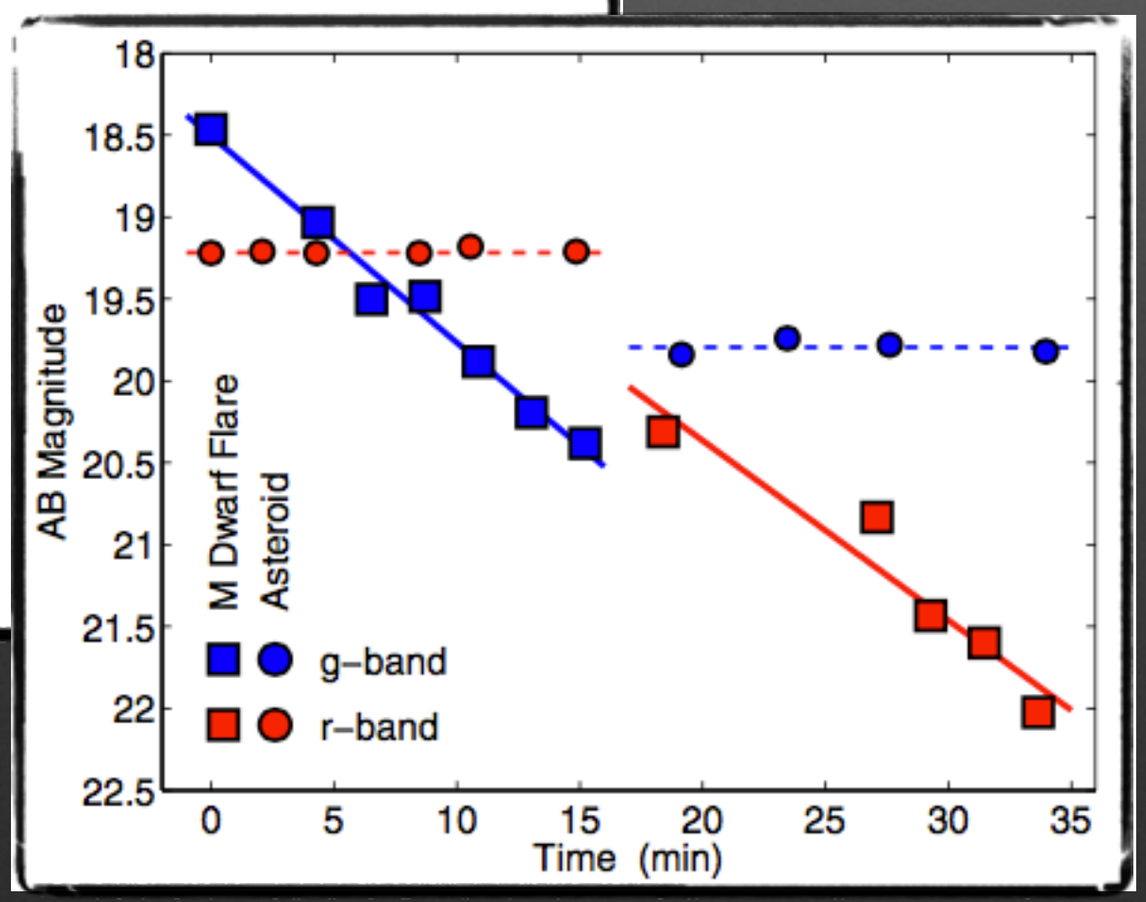


LVC+16

ML success(?): star/galaxy separation



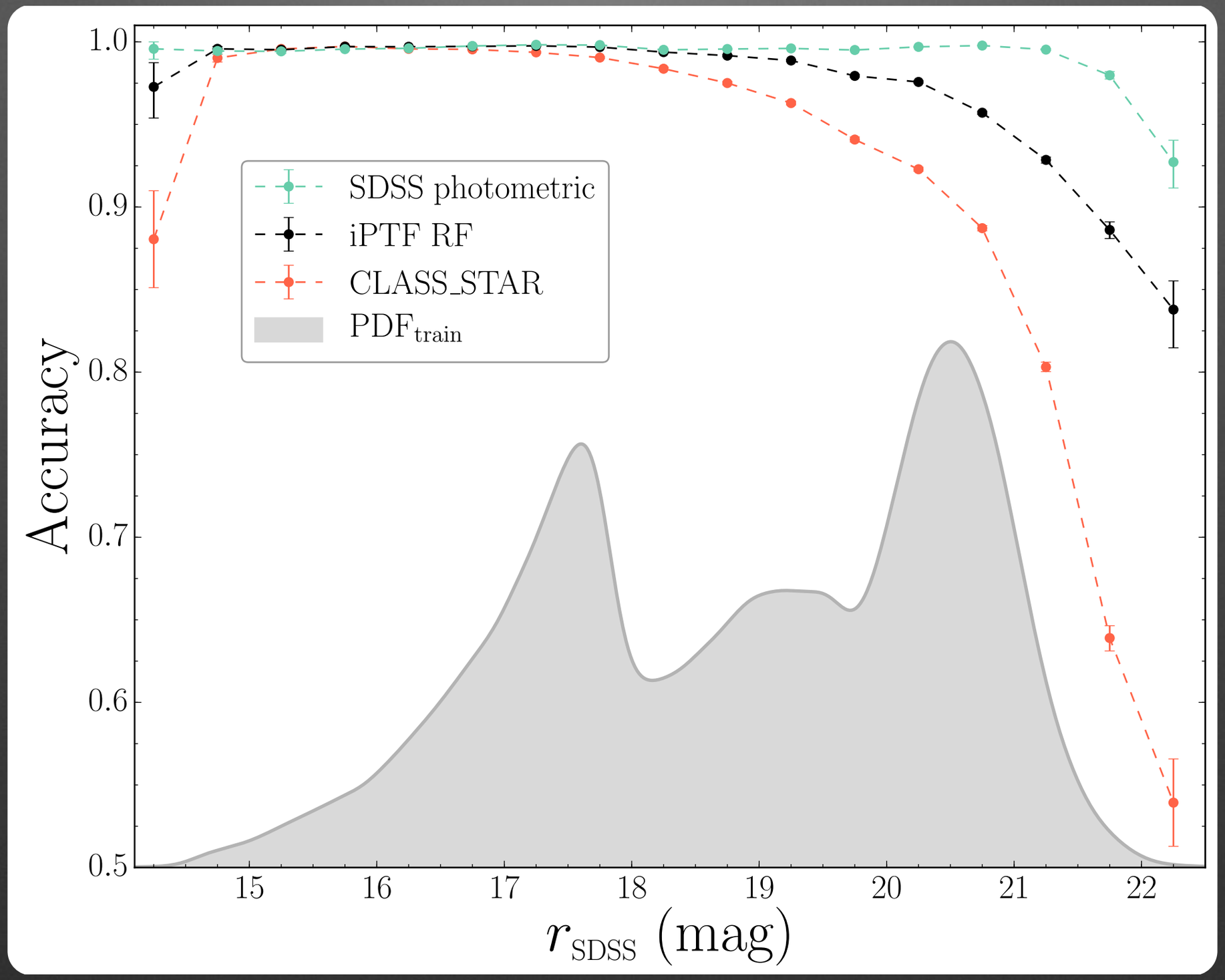
- Fast transients dominated by M dwarf flares
- Most likely false positive when searching for GW-EM



s/g model eliminates FPs

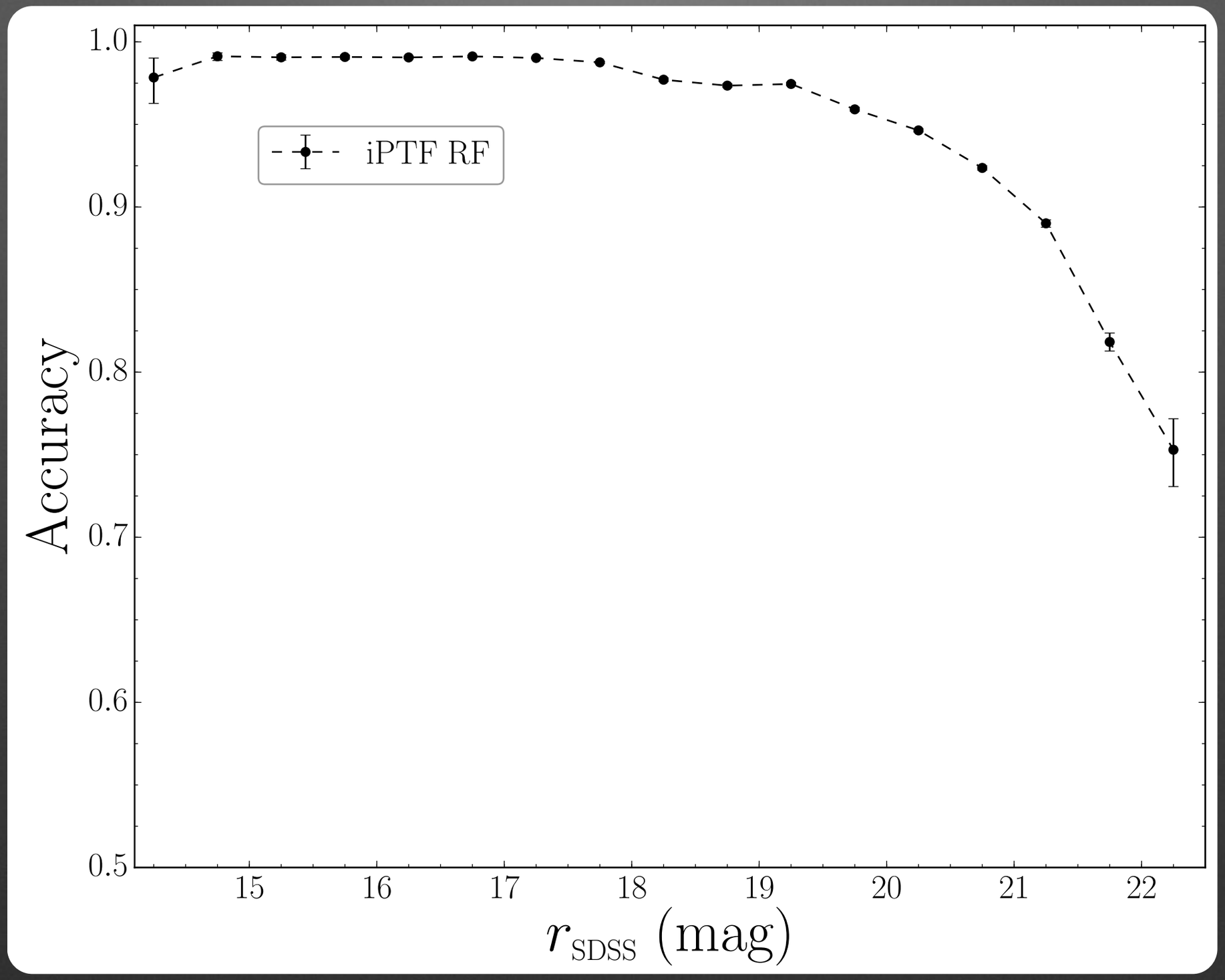
Berger+13

ML success(?): star/galaxy separation



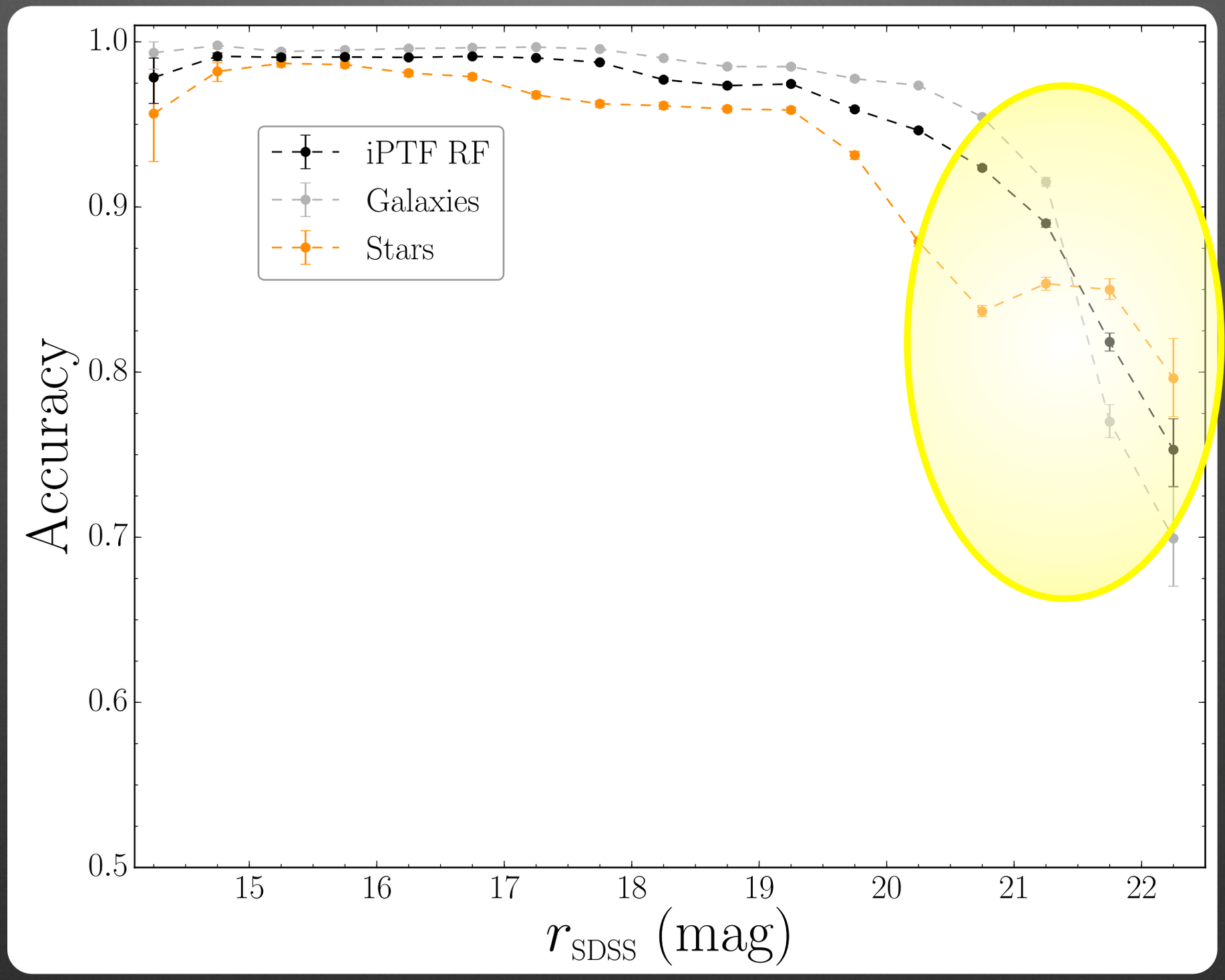
Kulkarni, AAM 16

ML success(?): star/galaxy separation



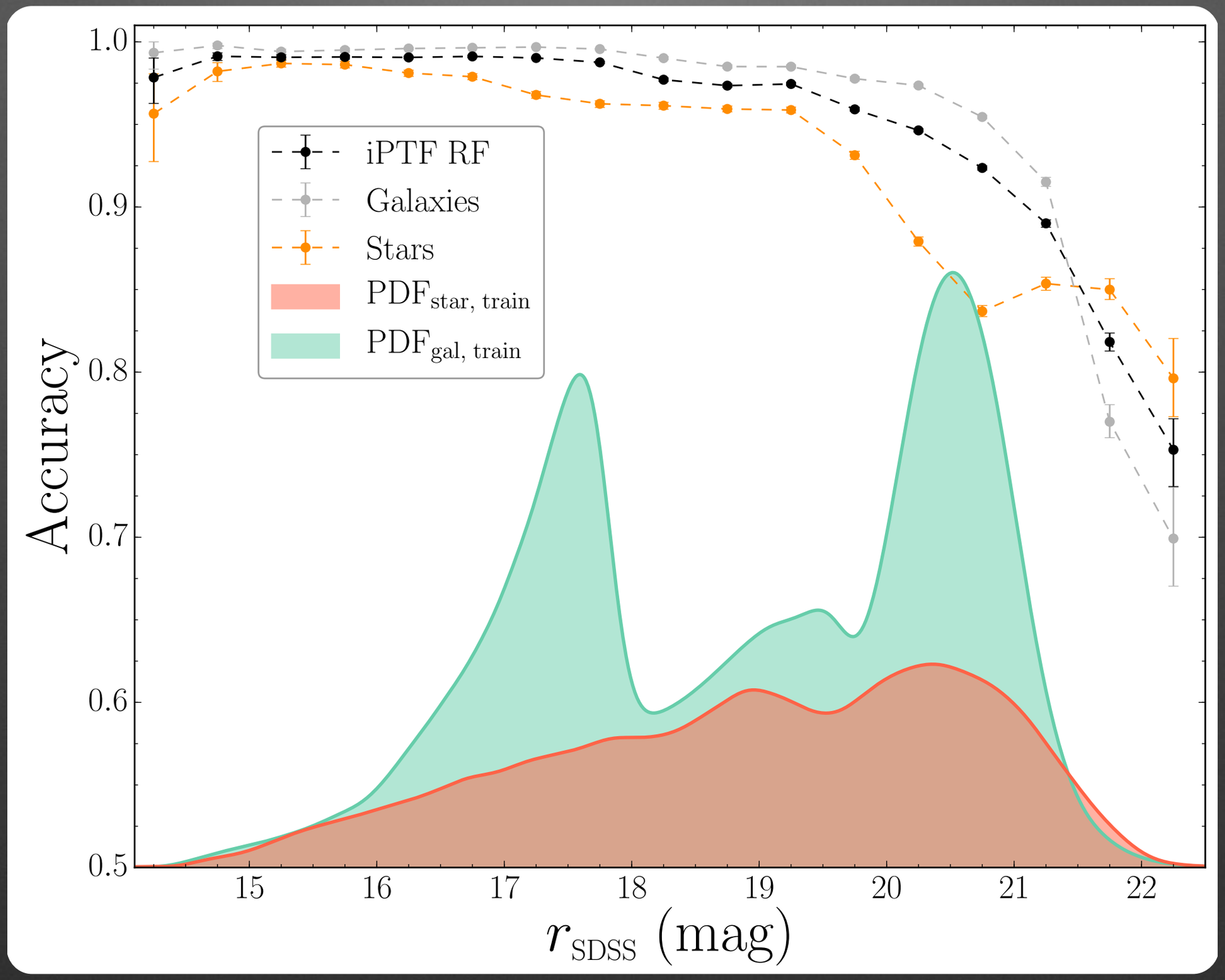
Kulkarni, AAM 16

ML success(?): star/galaxy separation



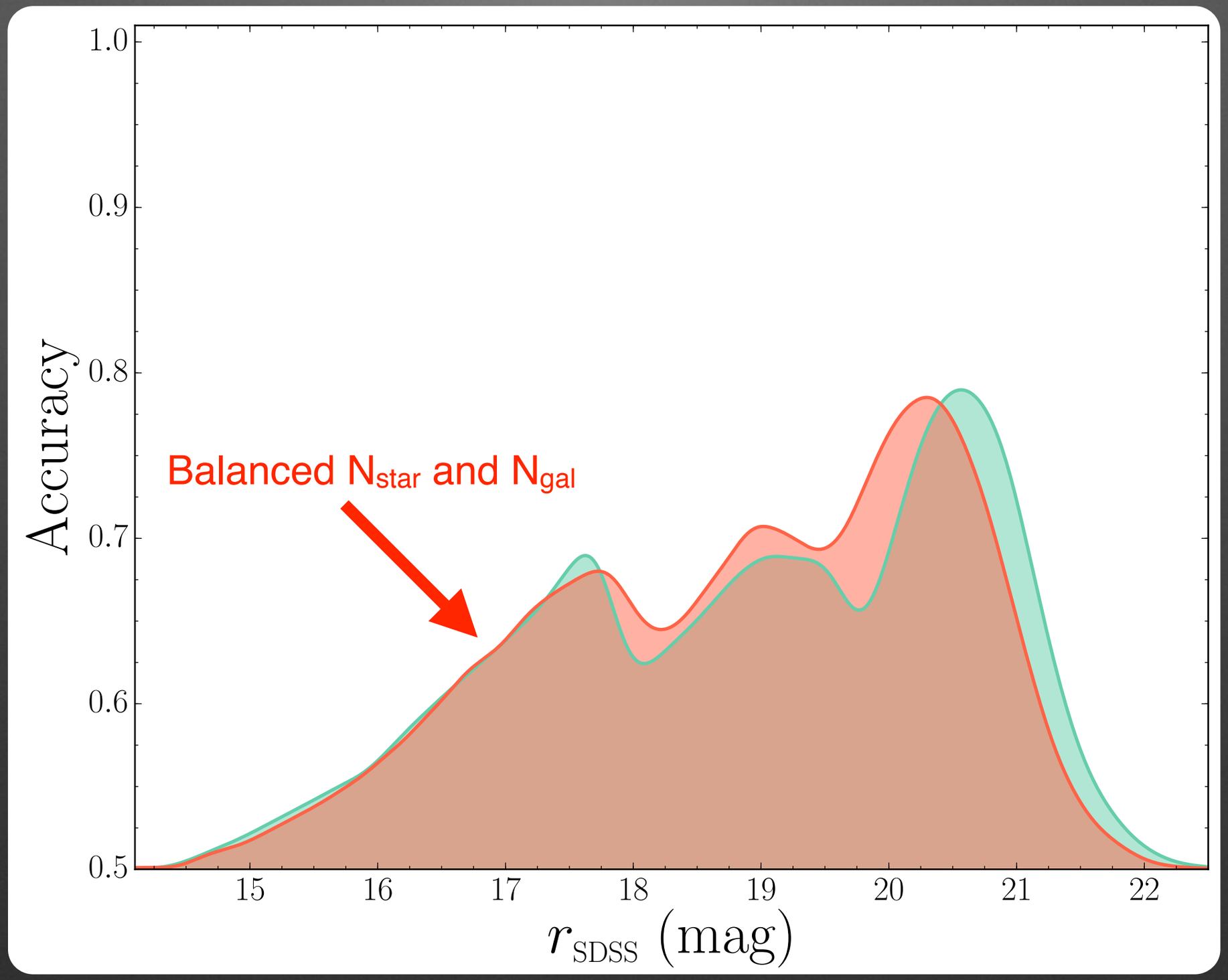
Kulkarni, AAM 16

ML success(?): star/galaxy separation



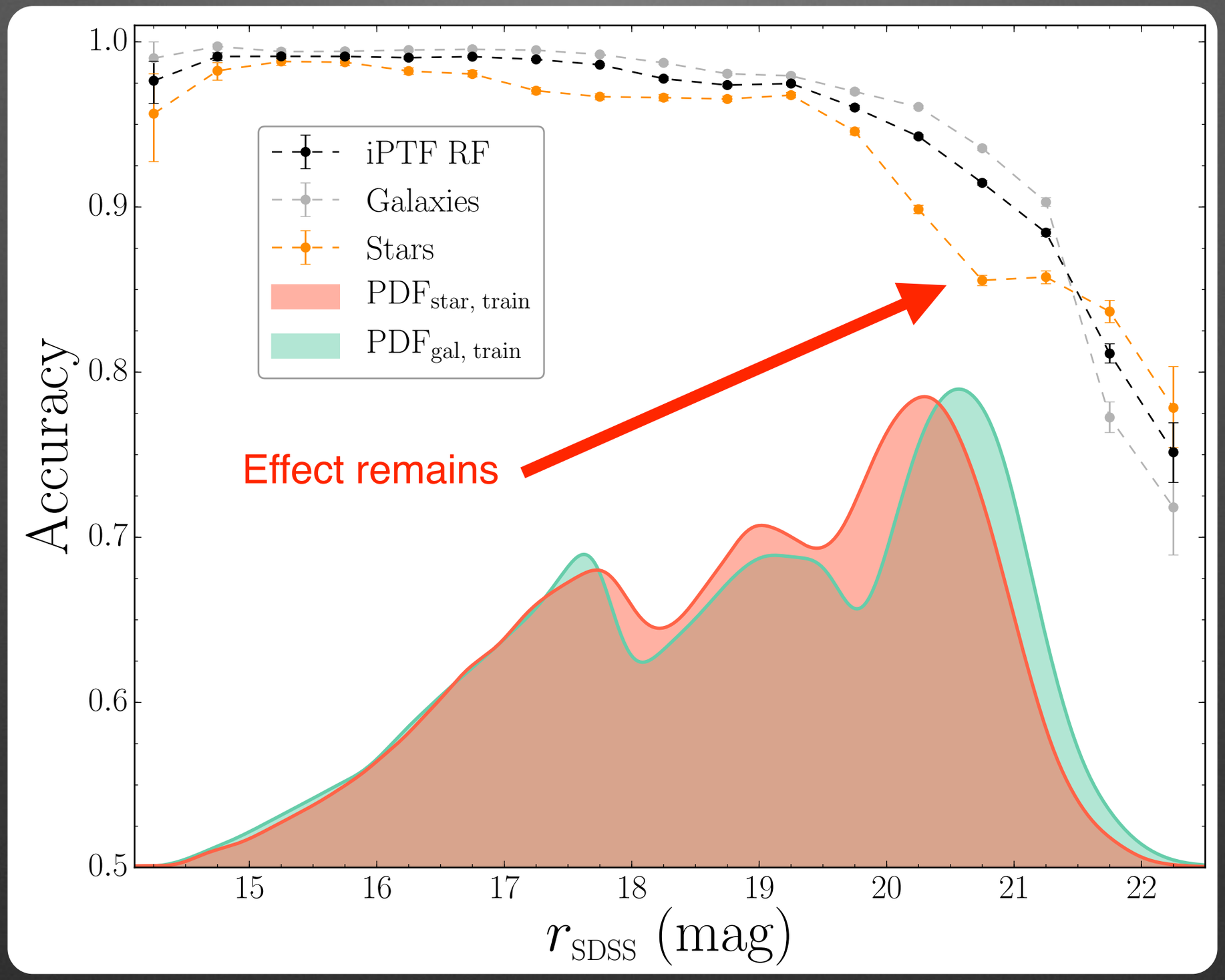
Kulkarni, AAM 16

ML success(?): star/galaxy separation



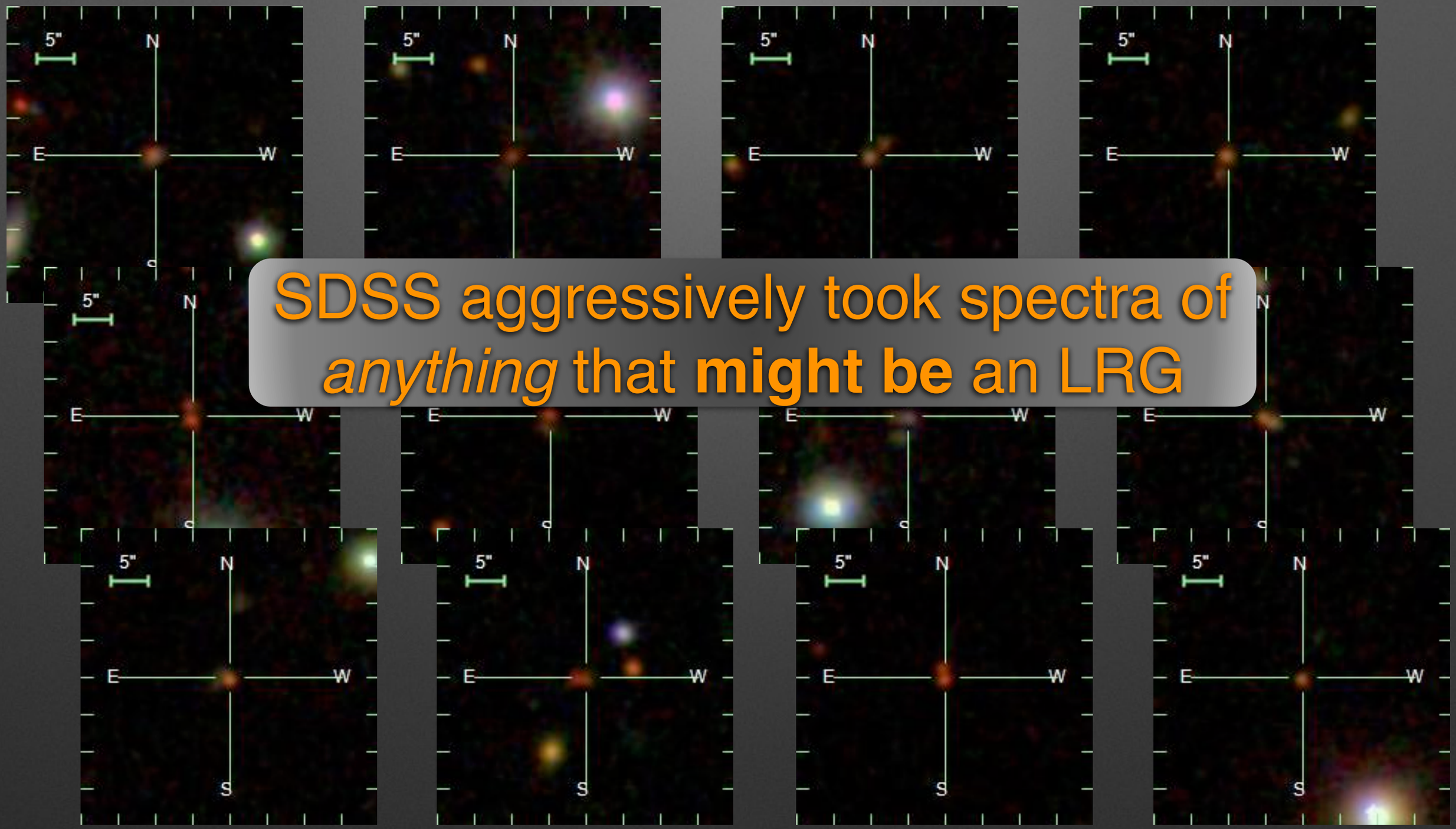
Kulkarni, AAM 16

ML success(?): star/galaxy separation



Kulkarni, AAM 16

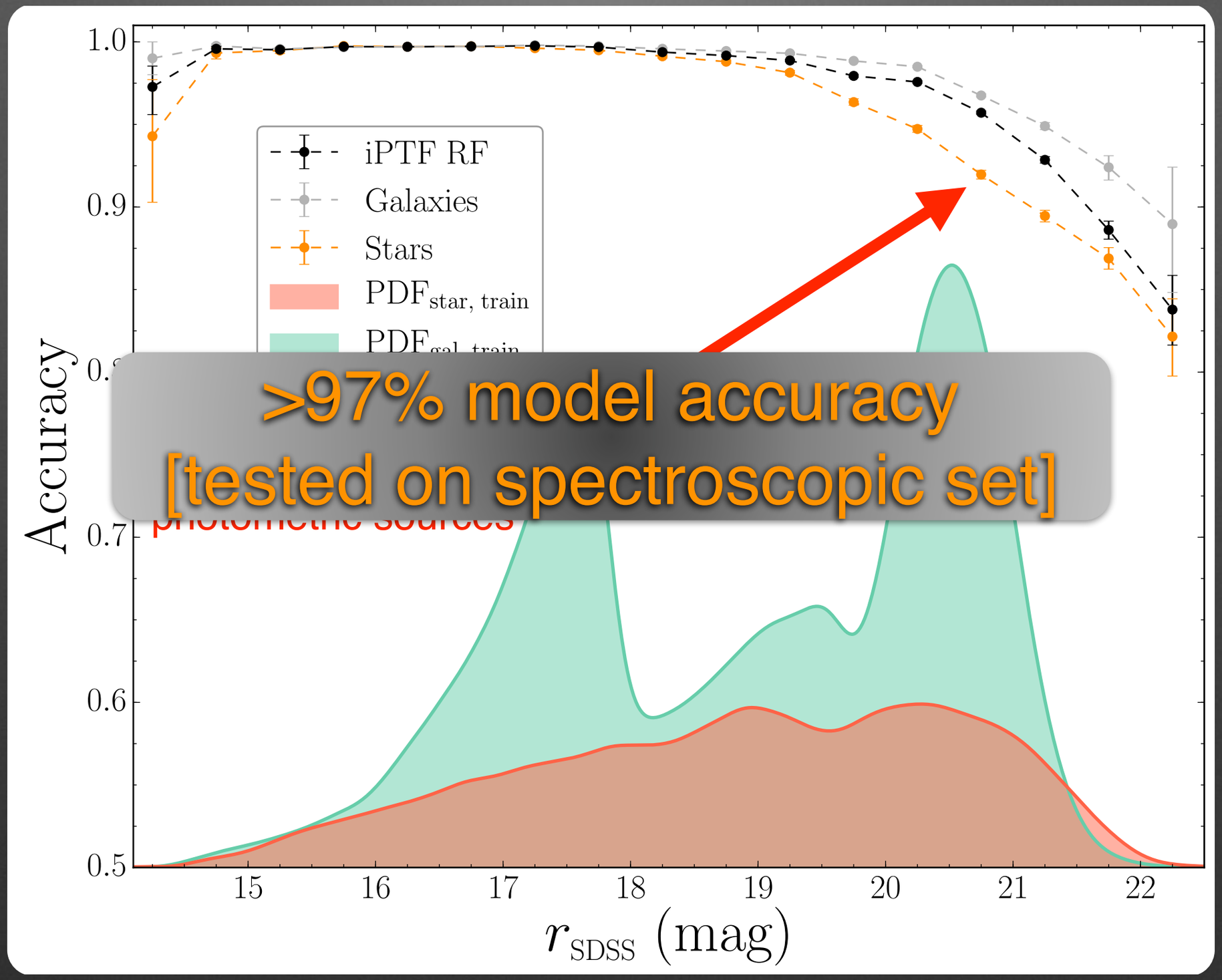
ML success(?): star/galaxy separation



SDSS aggressively took spectra of *anything* that **might be** an LRG

(C) SDSS

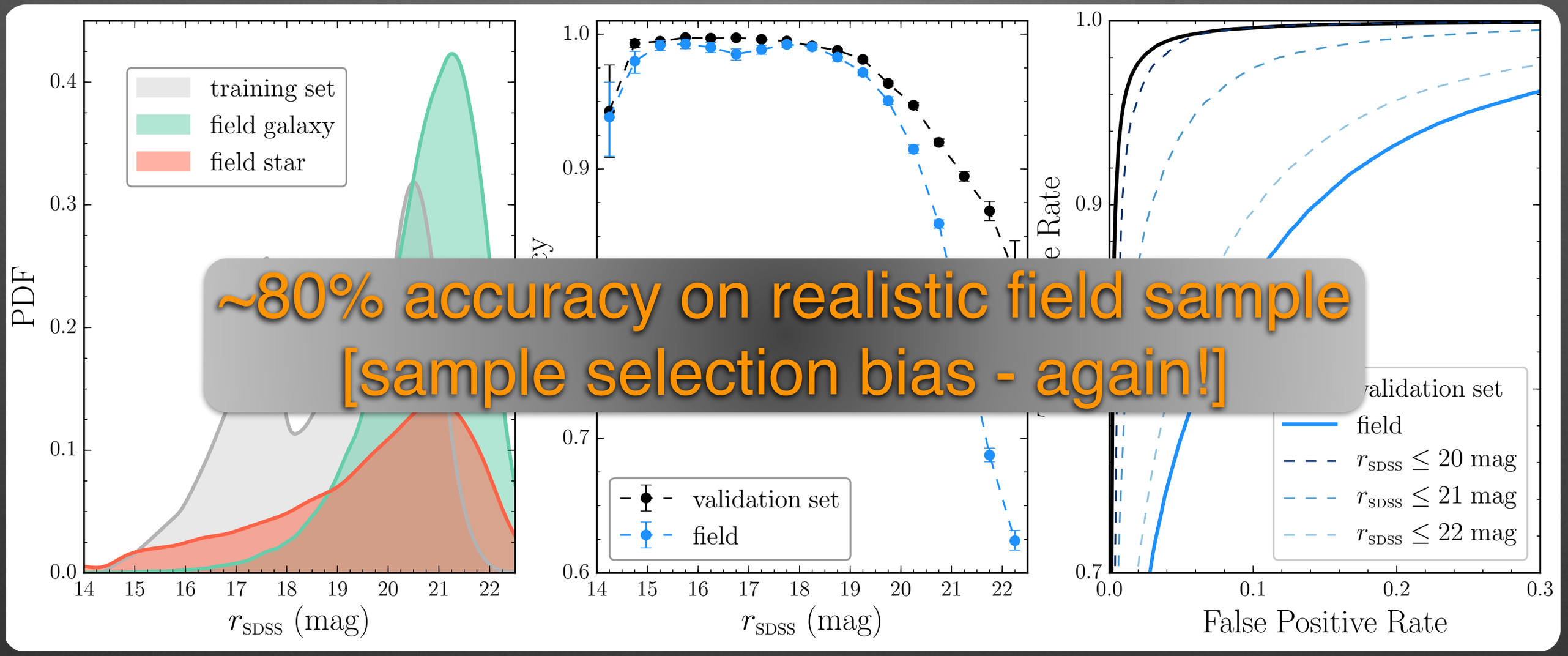
ML success(?): star/galaxy separation



>97% model accuracy
[tested on spectroscopic set]

Kulkarni, AAM 16

ML success(?): star/galaxy separation

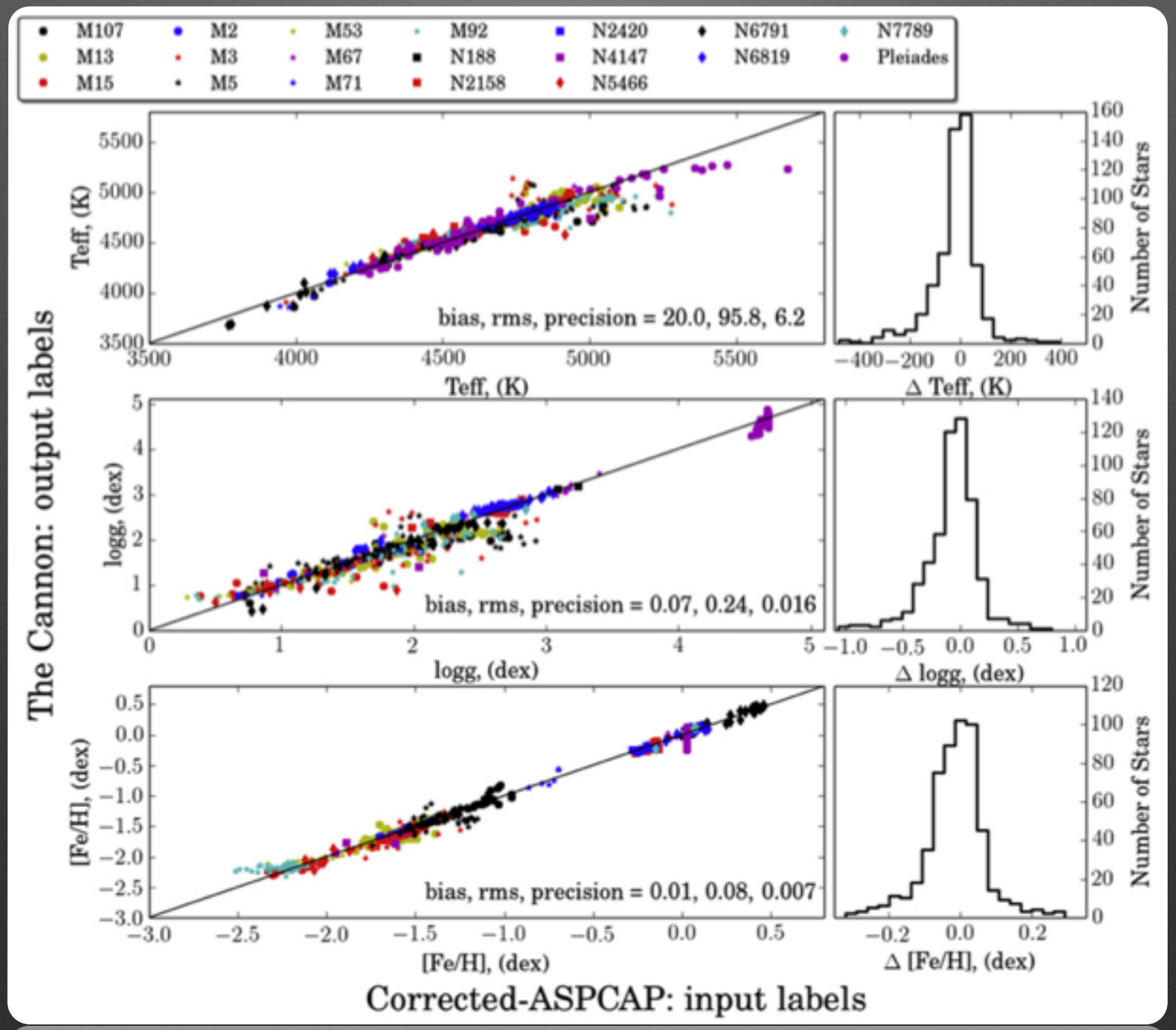


~80% accuracy on realistic field sample
[sample selection bias - again!]

Kulkarni, AAM 16

ML-like success: The Cannon

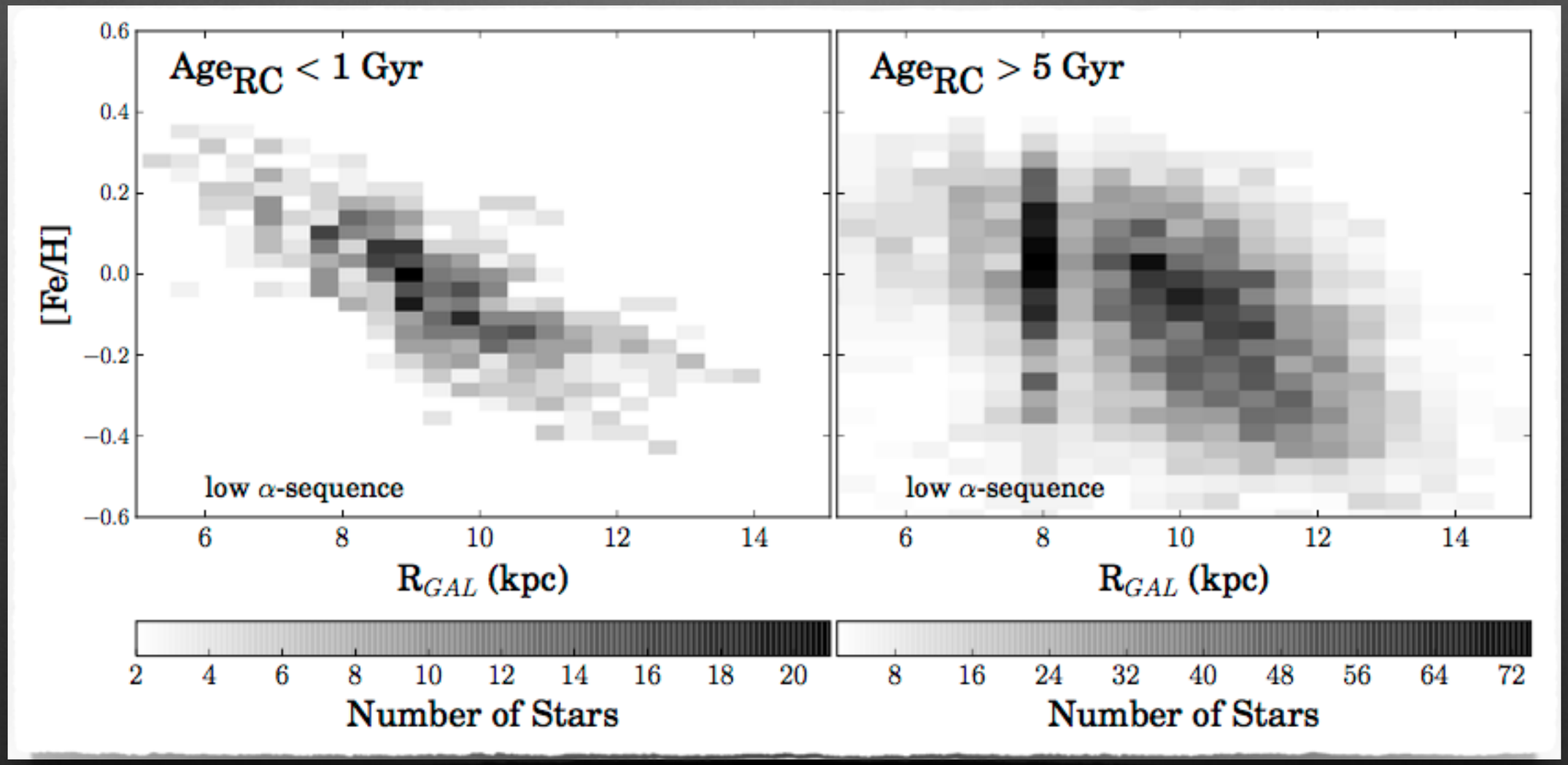
- ML-inspired data-driven method to analyze APOGEE spectra (Ness+15)
  - ➡ T_{eff}, log g, [Fe/H], abundances, also M_*, age (! - for RGs) (Ness+16)



Ness+15

ML-like success: The Cannon

- Direct evidence for stellar migration in the Milky Way (Ness+16)



Ness+16

Conclusions

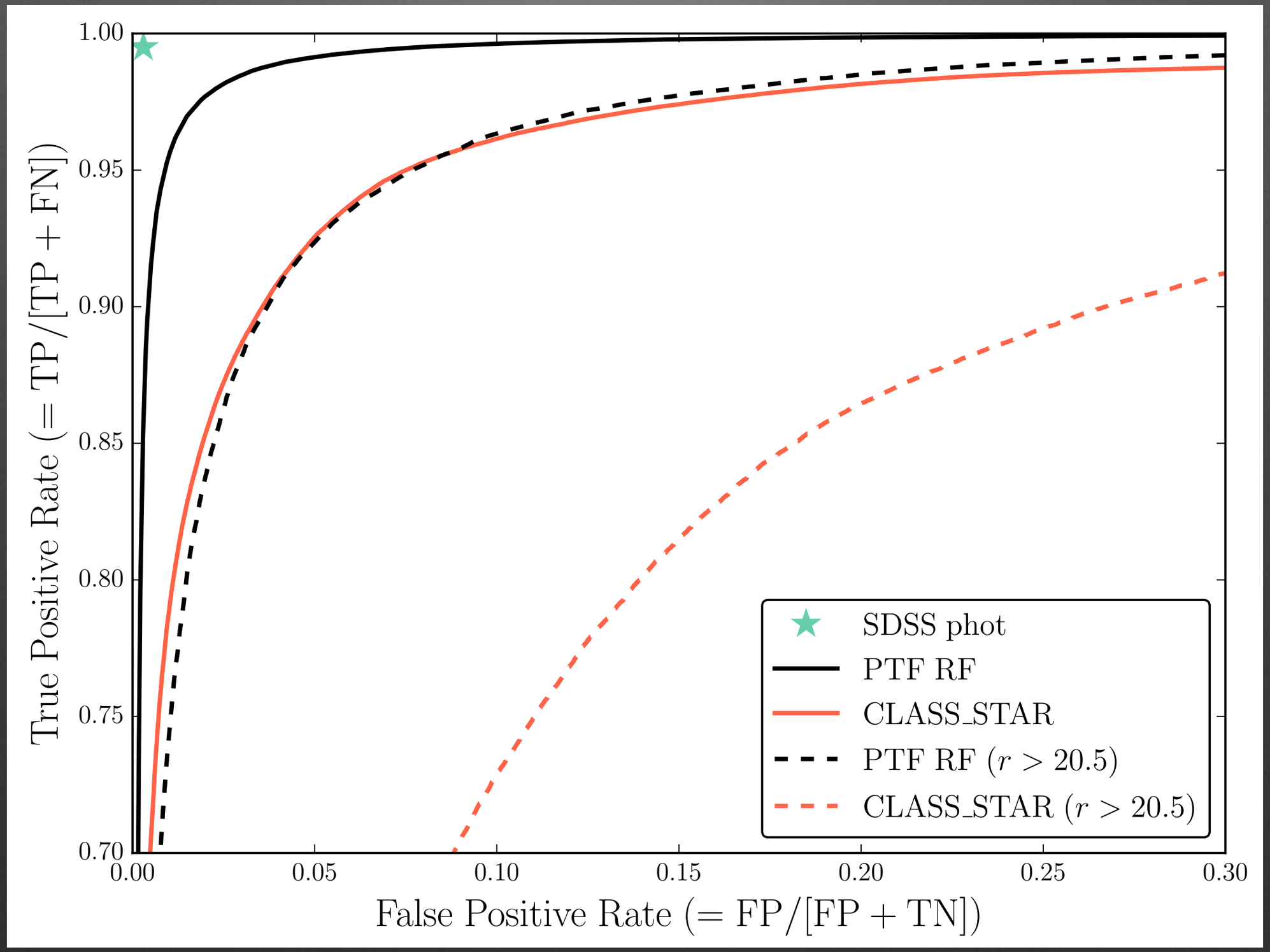
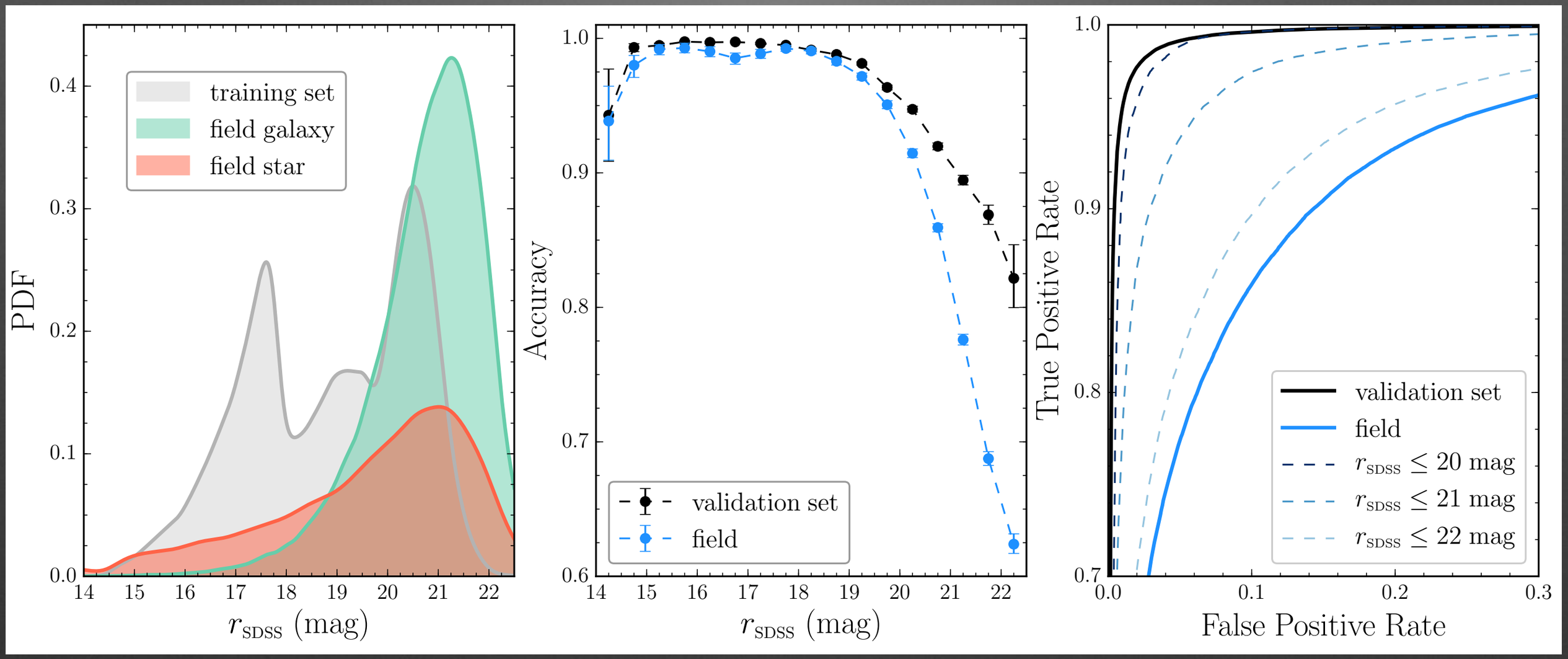
- Data-driven solutions are a necessity for ever-growing wide-field surveys (ZTF, LSST, etc)
  - ➡ e.g. real-bogus for transients
- Off-the-shelf ML algorithms are rarely plug+play for astro
  - ➡ nasty systematics (heteroskedastic errors & targeting bias)
  - ➡ e.g., small calibration errors in SDSS for EMP discovery
  - ➡ e.g., SDSS LRG bias
- Specific targets for one-off projects most likely to succeed
  - ➡ properly account for and propagate (non-ML like) uncertainties
  - ➡ e.g., The Cannon - measuring ages for >10k giants

Awwwwwwwww

SNAP!

You just asked a dope question!

Dope Question Time



Dope Question Time