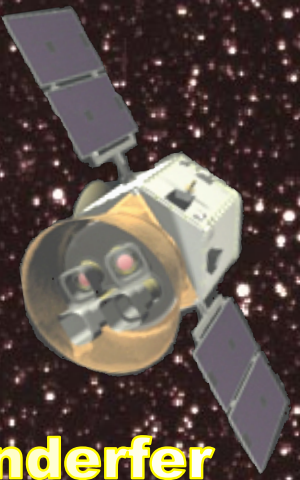
Machine Learning Approaches to Vetting Transiting Planet Signatures in Photometric Light Curves

Jon M. Jenkins

NASA Ames Research Center

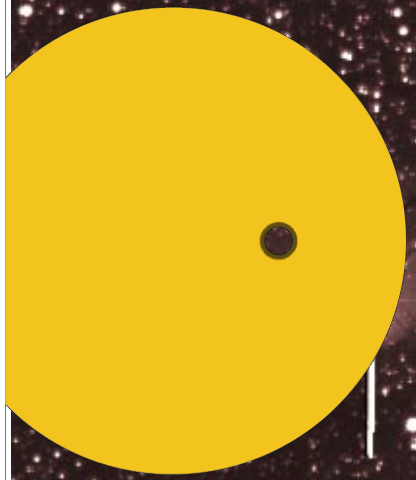
Sean McCauliff, Joseph Catanzarite,
Sean Seader, Joseph Twicken & Dwight Sanderfer



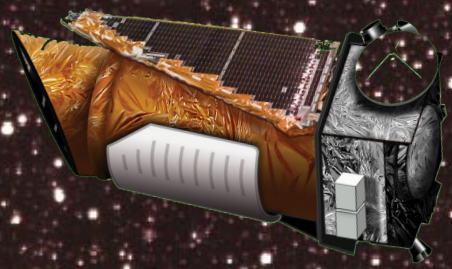
Monday September 15, 2014

Toward Other Earths II

Porto, Portugal



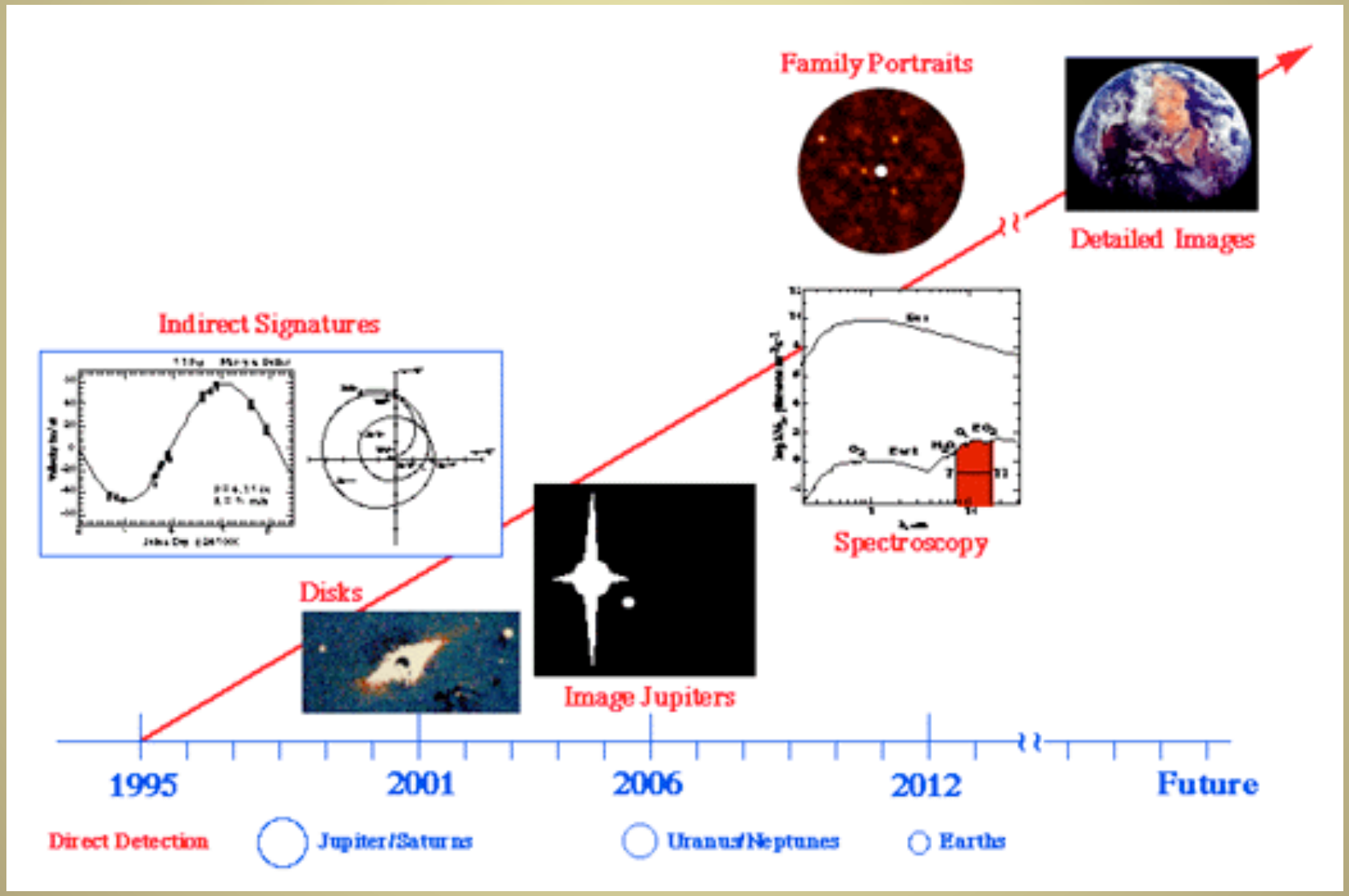
TOWARDS OTHER EARTHS II
The Star-Planet Connection



NASA's 1995 ExNPS Report

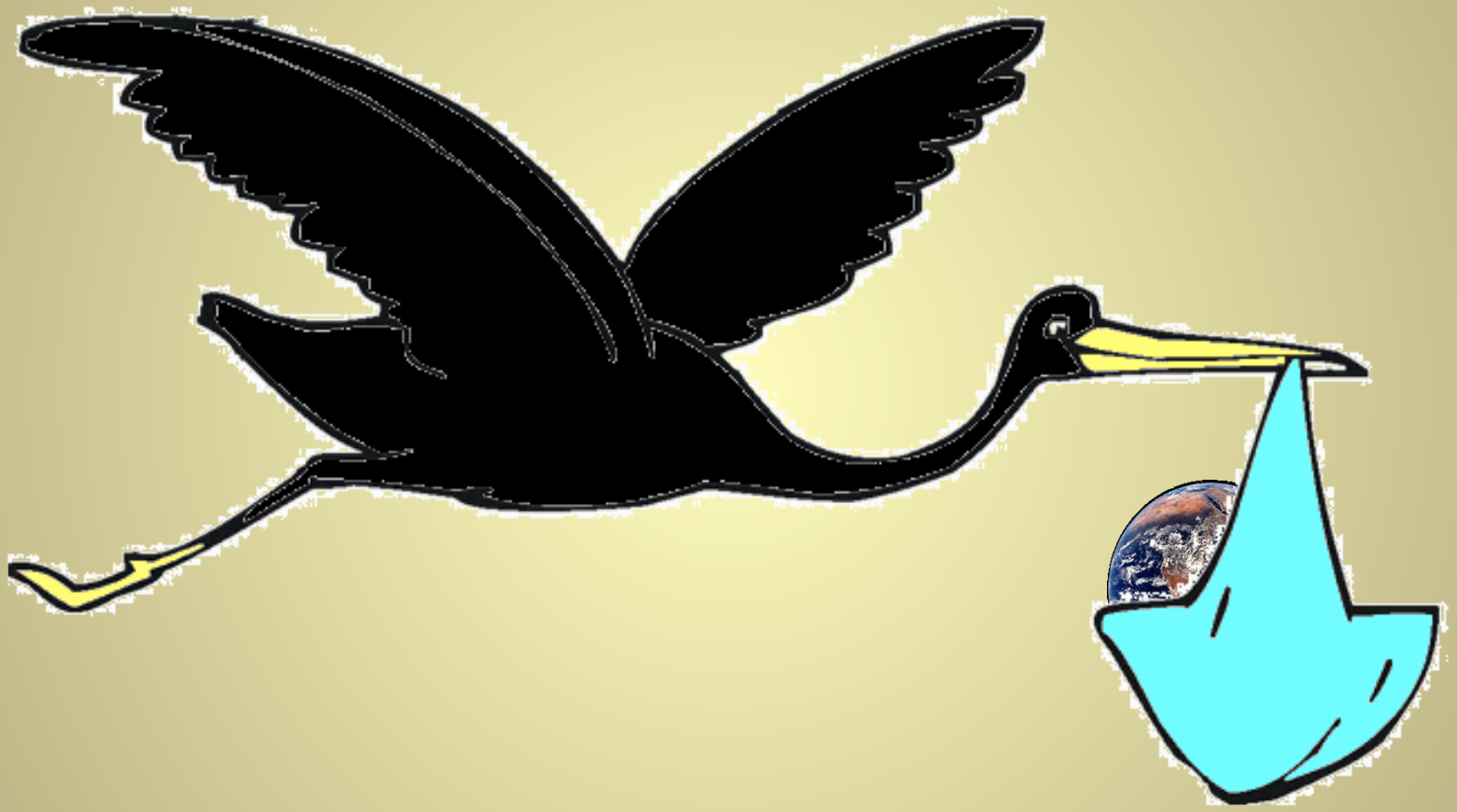
Kepler
A Search for Earth-size Planets

Transit Photometry not Recommended



Where Do Planets Come From?

Kepler
A Search for Earth-size Planets

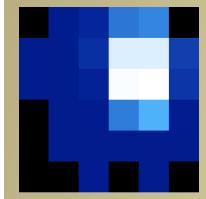


The *Kepler* Pipeline*

Kepler
A Search for Earth-size Planets

Artificial Transit & BEB Injection Machine

Artificial Transits and Eclipses



Raw Data

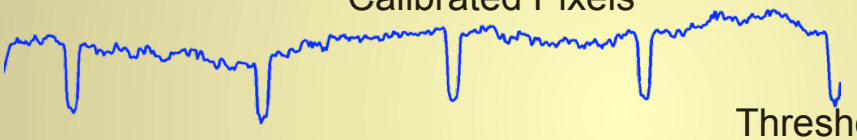
CAL
Pixel level Calibration

PA
Photometric Analysis

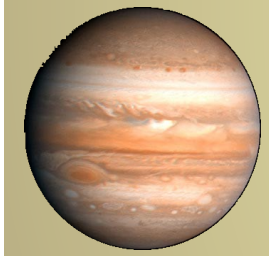
Raw Light Curves & Centroids

PDC
Pre-search Data Conditioning

Calibrated Pixels



Corrected Light Curves



Threshold Crossing Events (TCEs)

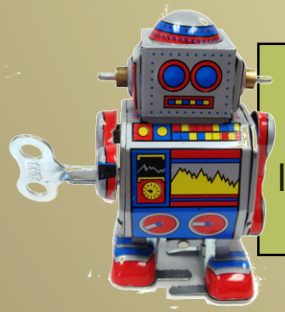
TPS
Transiting Planet Search

Diagnostic Metrics & Reports

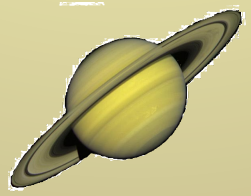
DV
Data Validation

Planet, or dud?

TCERT
Threshold Crossing Event Review Team



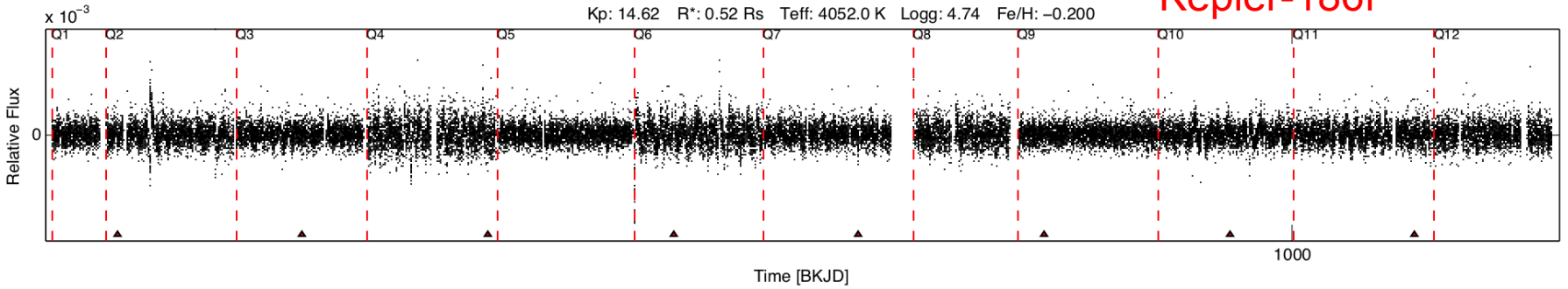
Auto-Vetting
Applying machine learning to candidate evaluation

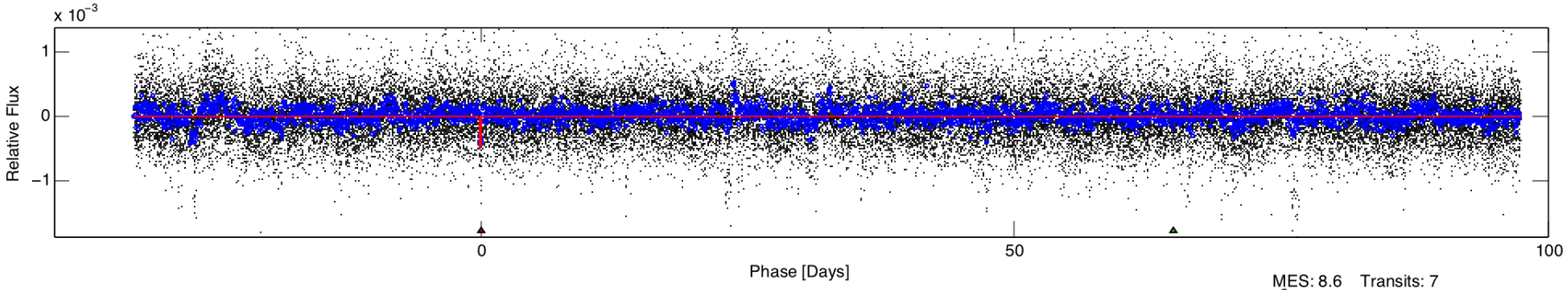


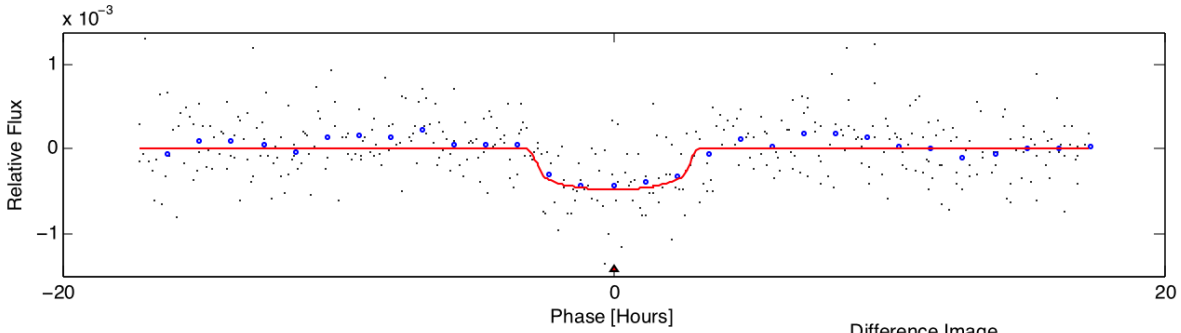
*+ a lot of follow up, characterization, and validation

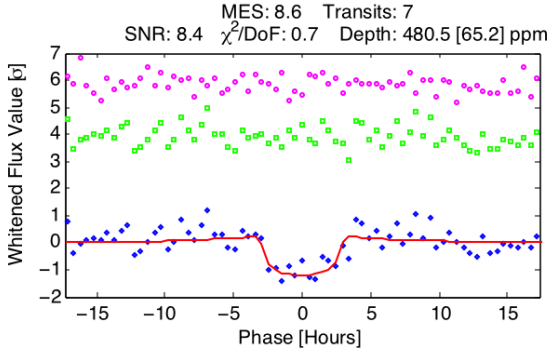
KIC: 8120608 Candidate: 5 of 5 Period: 129.945 d

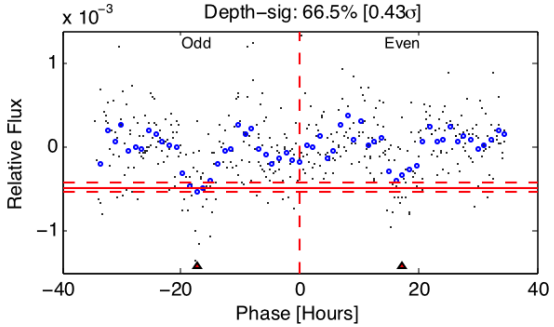
Kepler-186f

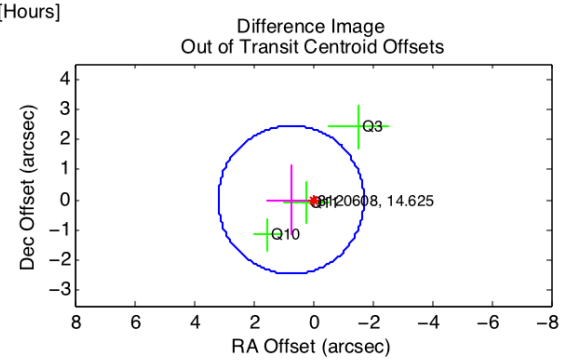












DV Fit Results:

Period = 129.94545 [0.00238] d
Epoch = 176.8272 [0.0115] BKJD
Rp/R* = 0.0200 [0.0271]
a/R* = 165.05 [882.88]
b = 0.34 [13.80]

Teq = 202 K
Rp = 1.14 Re
a = 0.4104 AU

DV Diagnostic Results:

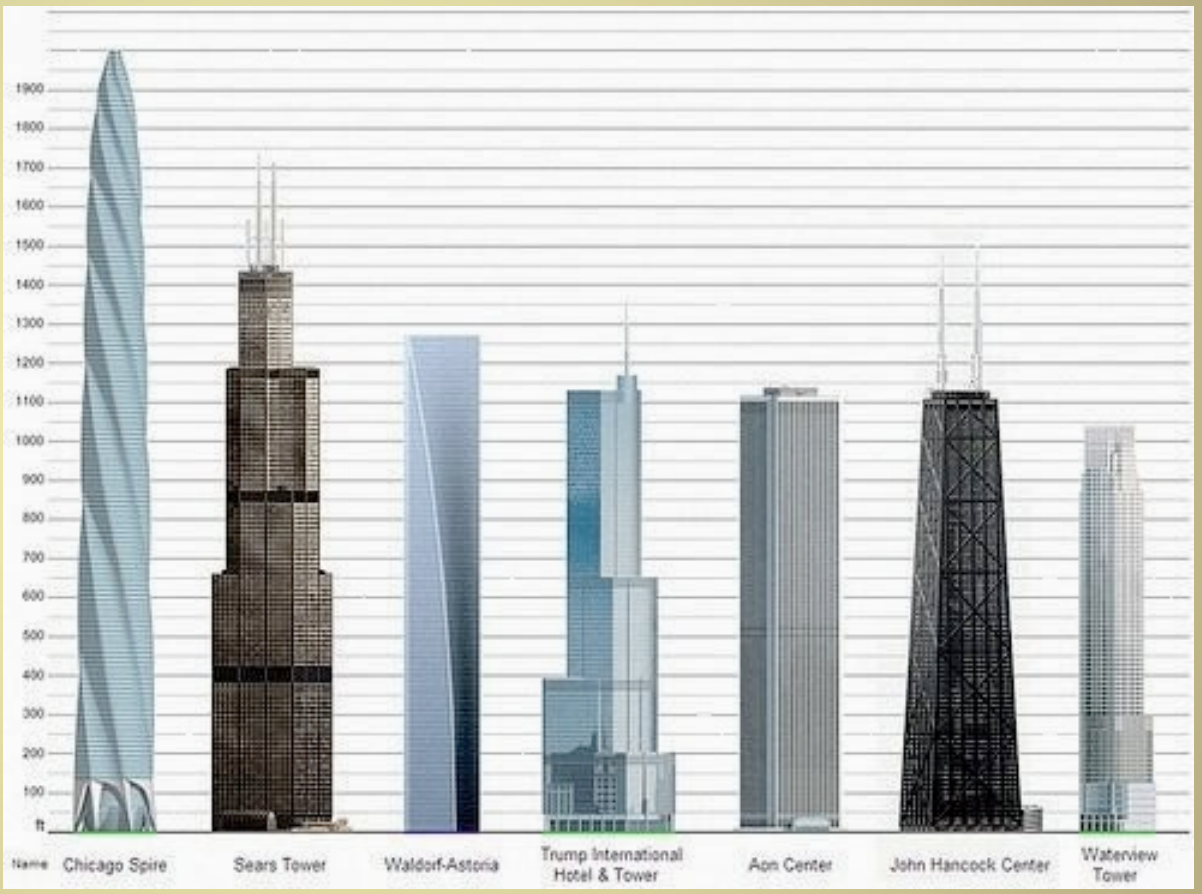
Epoch−sig: 78.0% [0.28σ]
ShortPeriod−sig: 100.0% [0.00σ]
LongPeriod−sig: N/A
Centroid−sig: 4.6% [1.99σ]
Bootstrap−pfa: N/A

OotOffset−rm: 0.735 arcsec [0.90σ]
KicOffset−rm: 0.869 arcsec [1.17σ]
OotOffset−bf: N/A
KicOffset−bf: N/A

30-Oct-2012

A Stack of DV Reports

Kepler
A Search for Earth-size Planets



…would reach ~140 stories high

A Plague of Planet-like Signatures

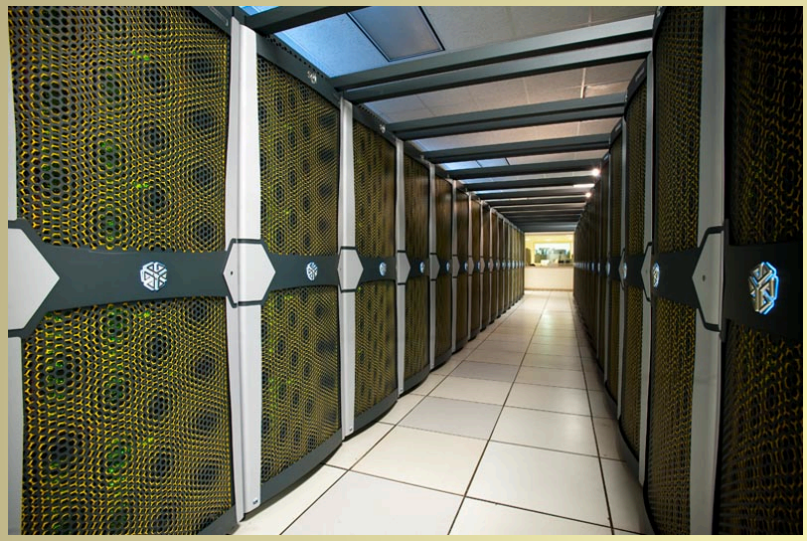
Hardware Architecture

Original computing clusters had ~500 computer cores

It would take ~2 years to re-process all Kepler data here

Hardware Architecture

Kepler
A Search for Earth-size Planets



NAS Pleiades Supercomputer has >126,000 computer cores

It takes ~2 months to re-process all Kepler data now

Distribution of TCEs

Kepler
A Search for Earth-size Planets

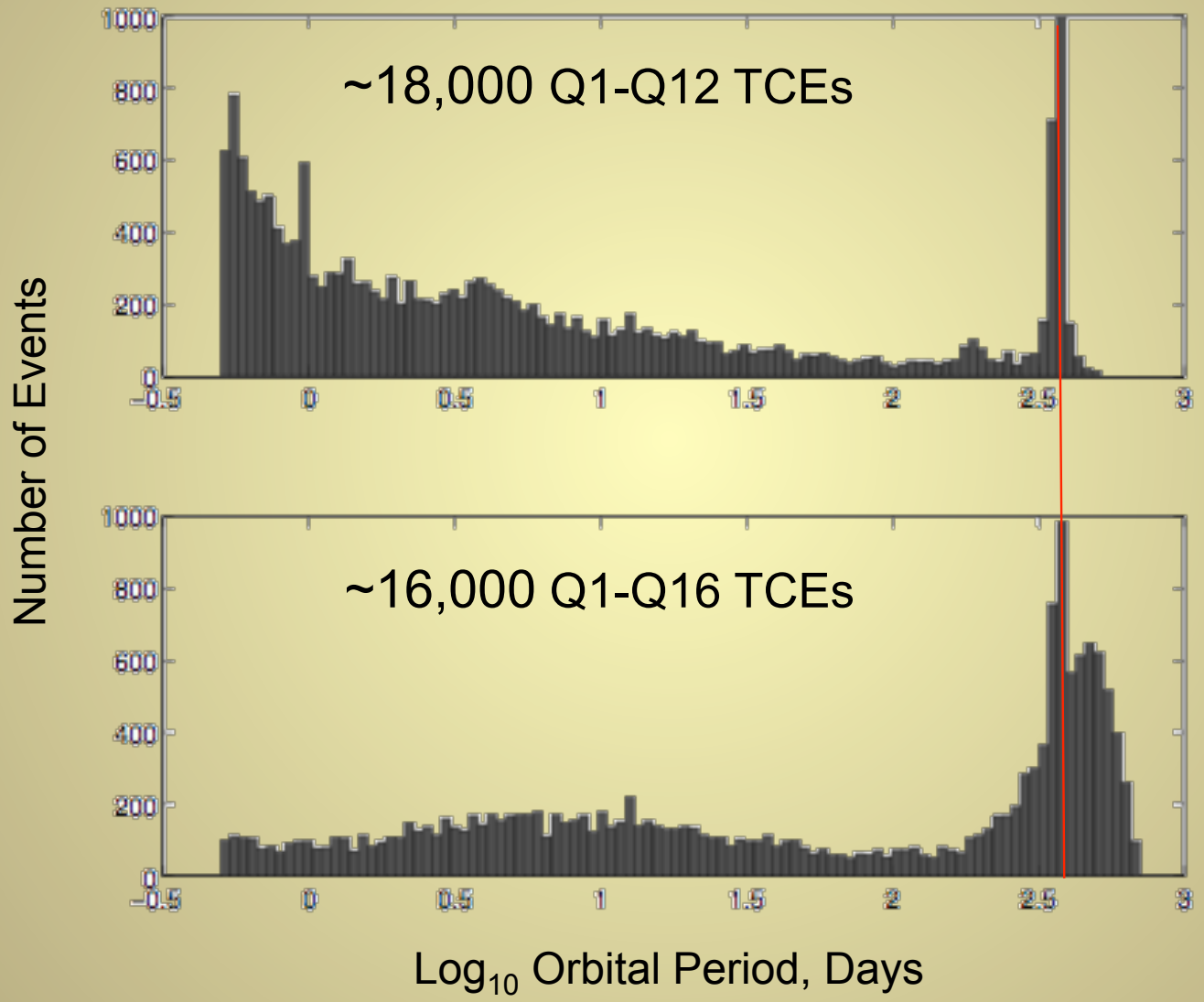


Image artifacts cause a large number of TCEs at 1 year periods!

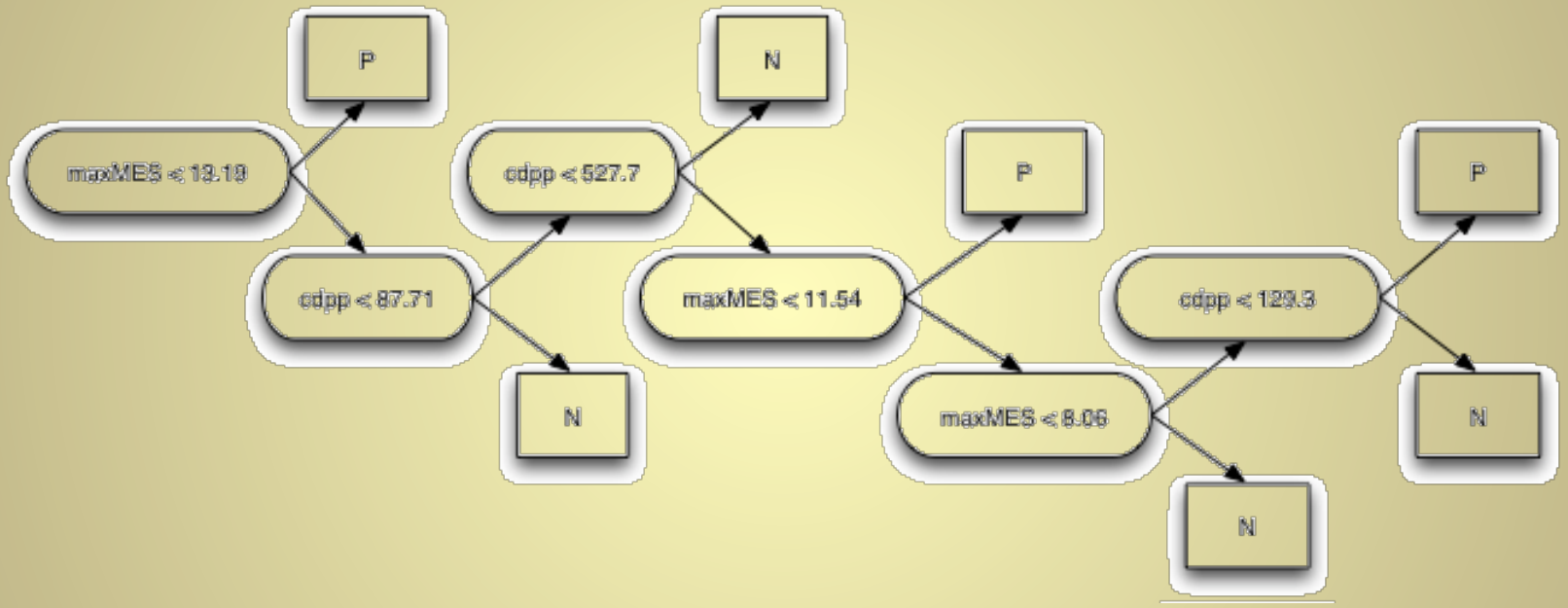
Random Forests for Auto-vetting

Random Forests for Auto-vetting

A Falling Tree in the Random Forest[1]

Kepler
A Search for Earth-size Planets

- The importance of each attribute can be characterized
- The voting record can be used to “score” the objects classified as “planet candidates” and “non-planetary candidates”

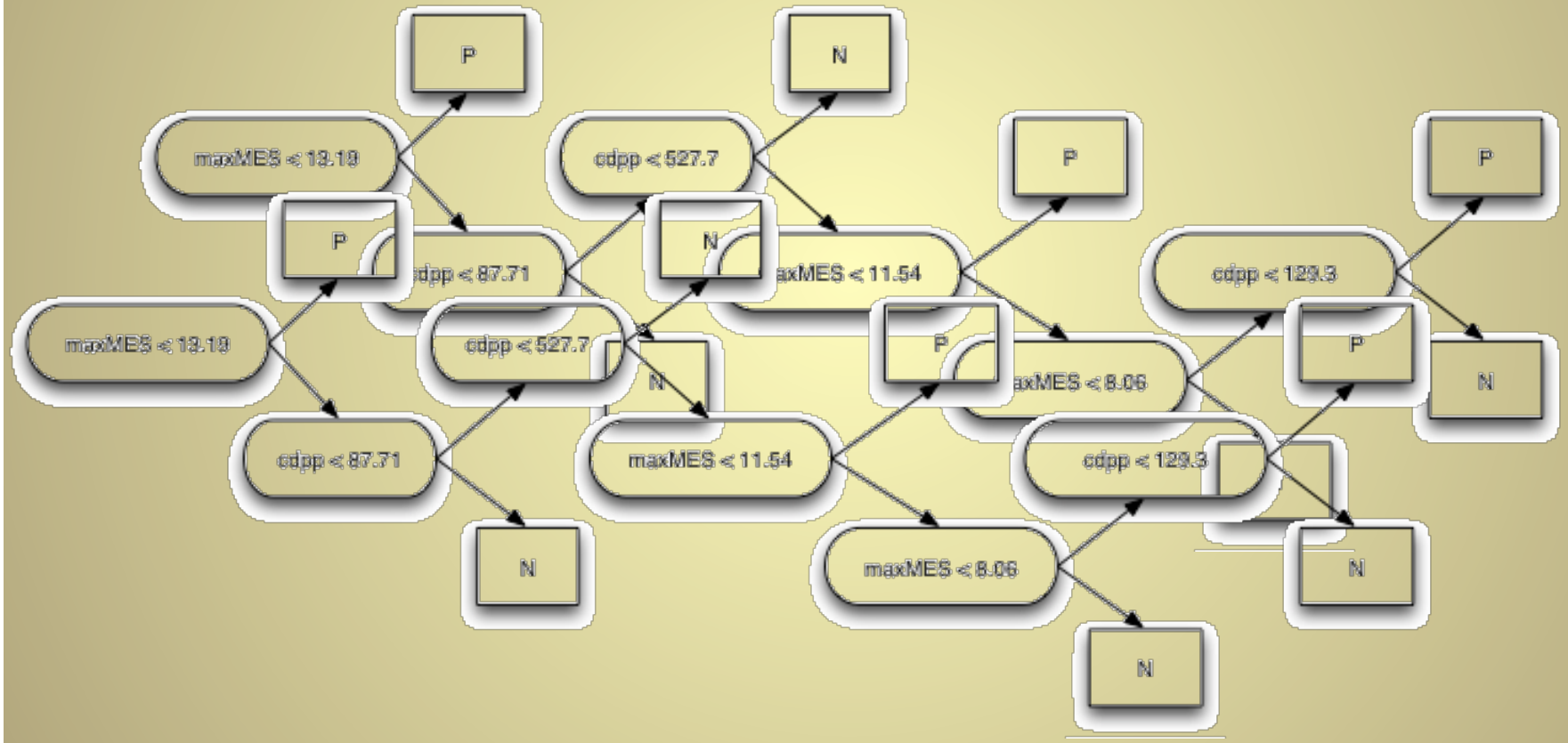


[1]Breiman, Leo (2001), "Random Forests”, Machine Learning **45** (1): 5–32

A Falling Tree in the Random Forest[1]

Kepler
A Search for Earth-size Planets

- The importance of each attribute can be characterized
- The voting record can be used to "score" the objects classified as "planet candidates" and "non-planetary candidates"

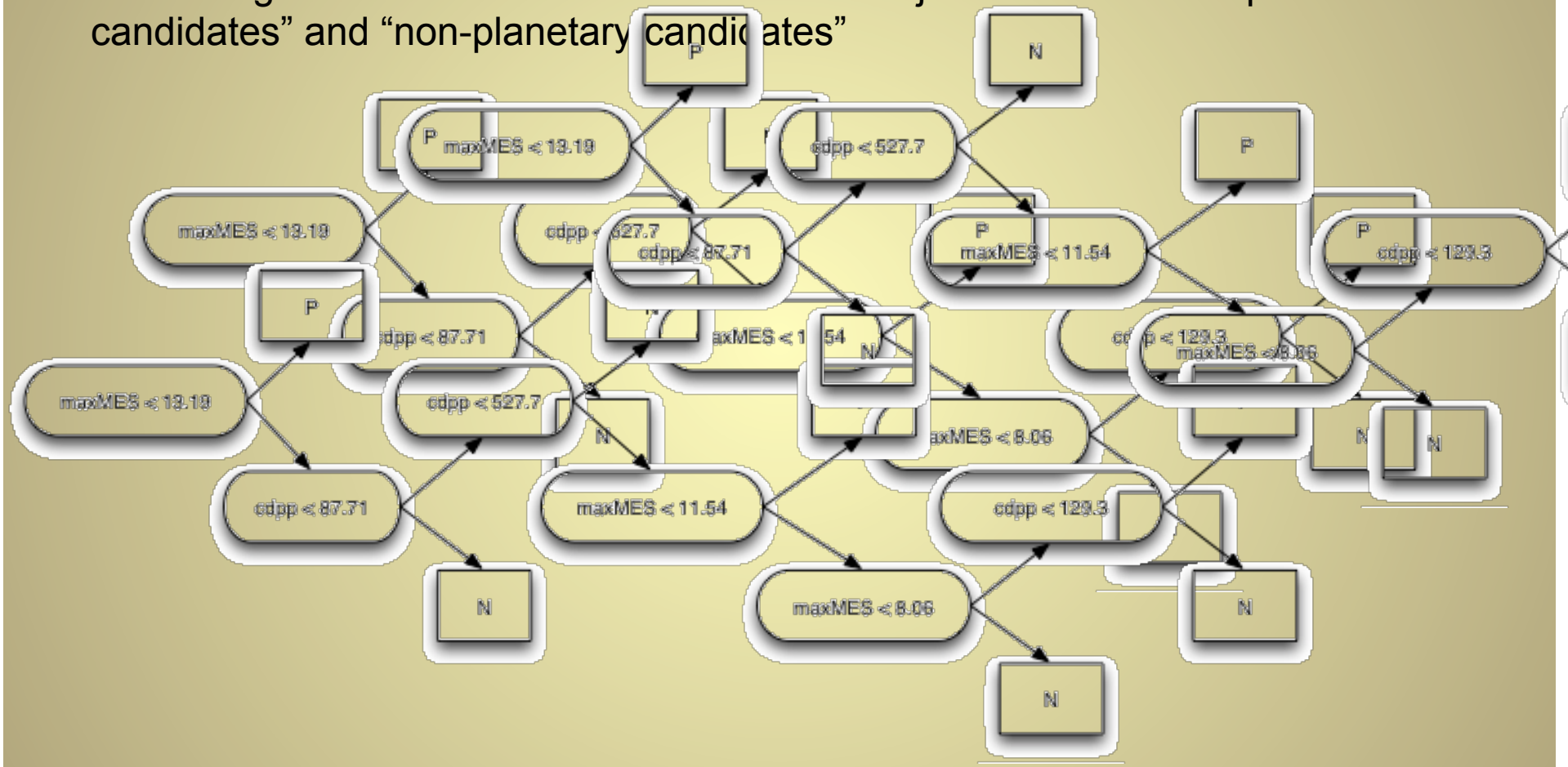


[1]Breiman, Leo (2001), "Random Forests", Machine Learning **45** (1): 5–32

A Falling Tree in the Random Forest[1]

Kepler
A Search for Earth-size Planets

- The importance of each attribute can be characterized
- The voting record can be used to "score" the objects classified as "planet candidates" and "non-planetary candidates"

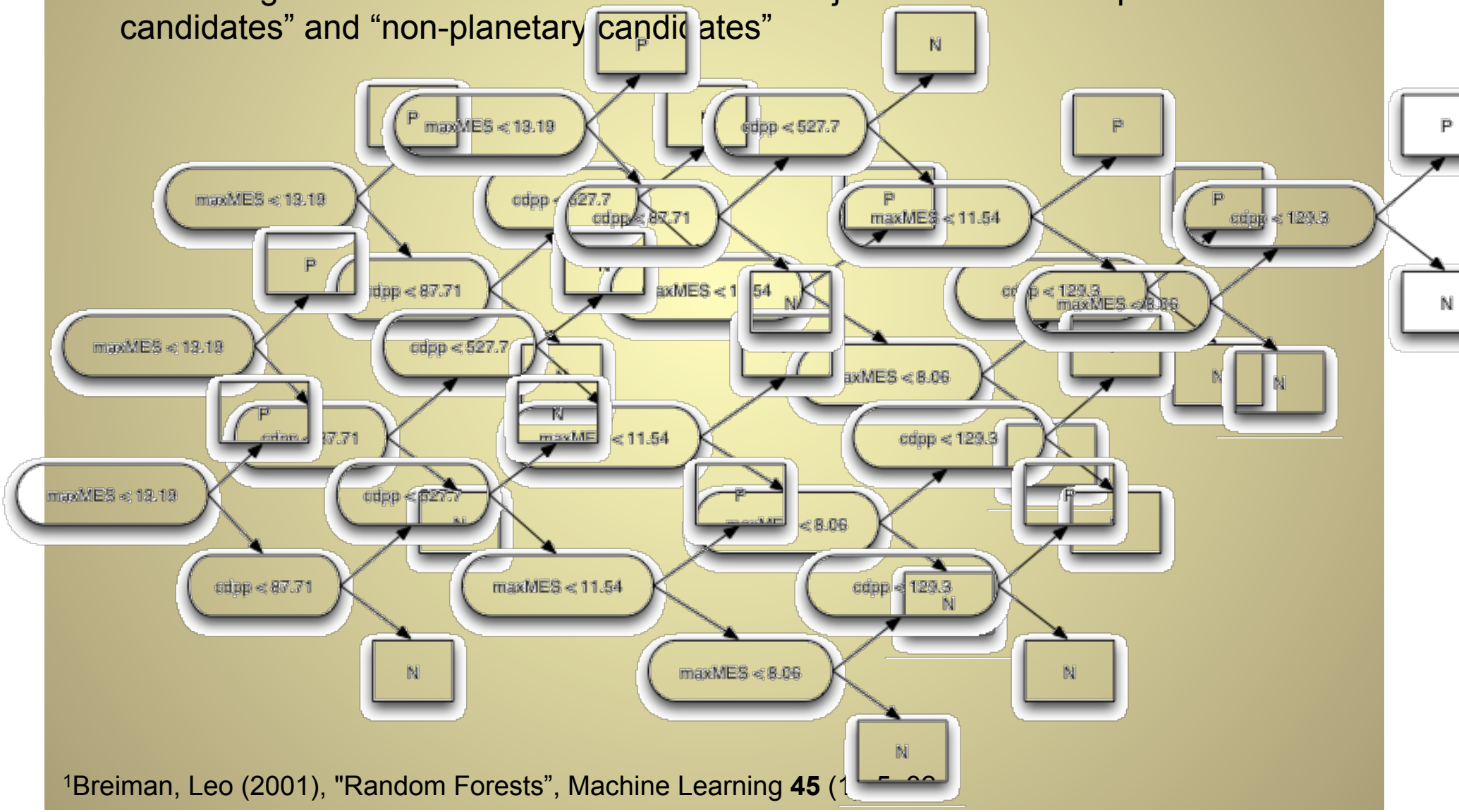


[1]Breiman, Leo (2001), "Random Forests", Machine Learning **45** (1): 5–32

A Falling Tree in the Random Forest[1]

Kepler
A Search for Earth-size Planets

- The importance of each attribute can be characterized
- The voting record can be used to "score" the objects classified as "planet candidates" and "non-planetary candidates"

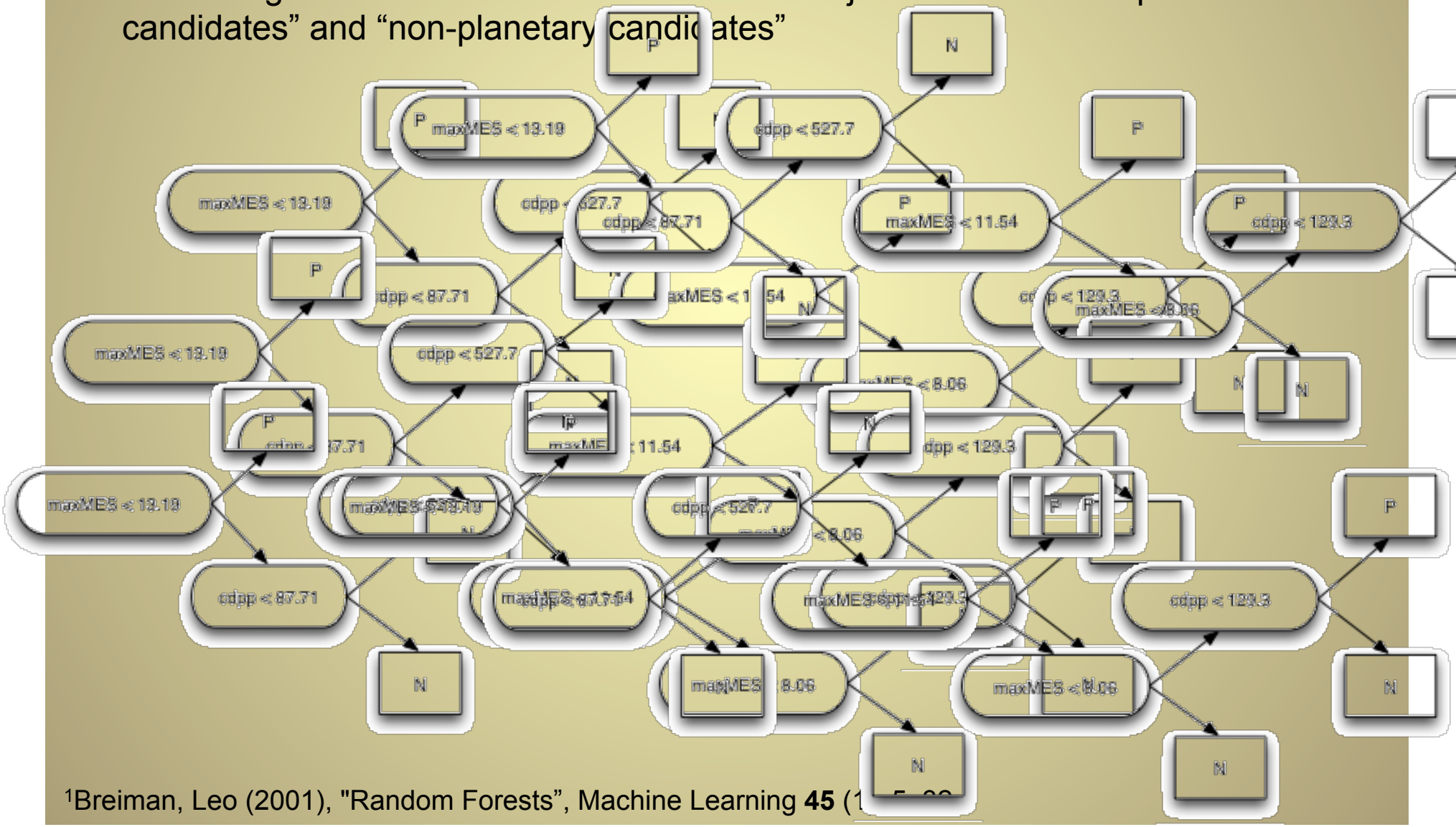


[1]Breiman, Leo (2001), "Random Forests", Machine Learning **45** (1

A Falling Tree in the Random Forest[1]

Kepler
A Search for Earth-size Planets

- The importance of each attribute can be characterized
- The voting record can be used to “score” the objects classified as “planet candidates” and “non-planetary candidates”

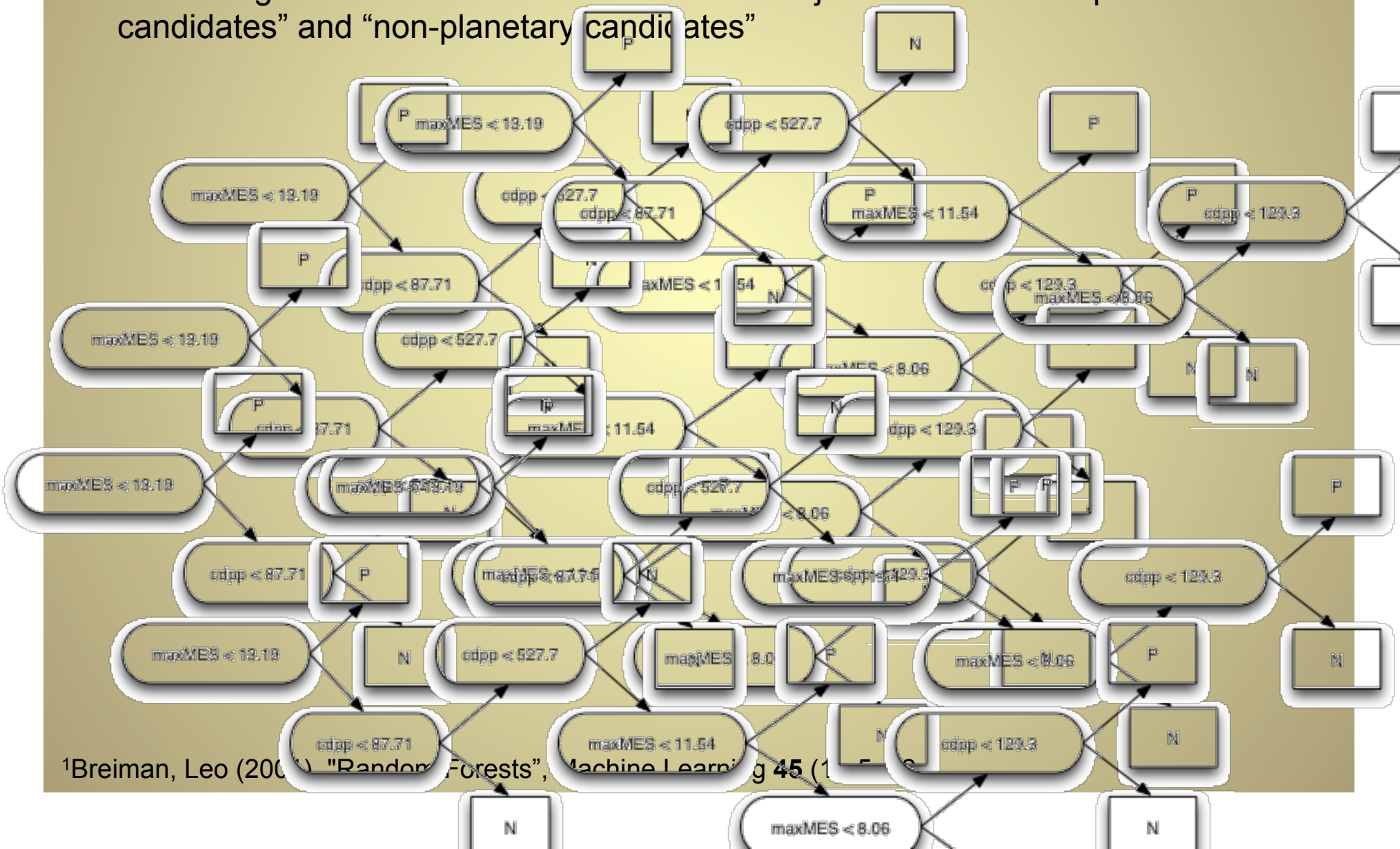


[1]Breiman, Leo (2001), "Random Forests”, Machine Learning **45** (1

A Falling Tree in the Random Forest[1]

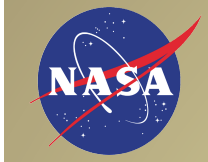
Kepler
A Search for Earth-size Planets

- The importance of each attribute can be characterized
- The voting record can be used to "score" the objects classified as "planet candidates" and "non-planetary candidates"



[1]Breiman, Leo (200). "Random Forests", Machine Learn g **45** (1 5 3

Initial Training Experiments



A Search for Earth-size Planets

"Planet Candidate" class established
1. Using confirmed and validated Kepler planets
2. Using planet candidates identified by TCERT

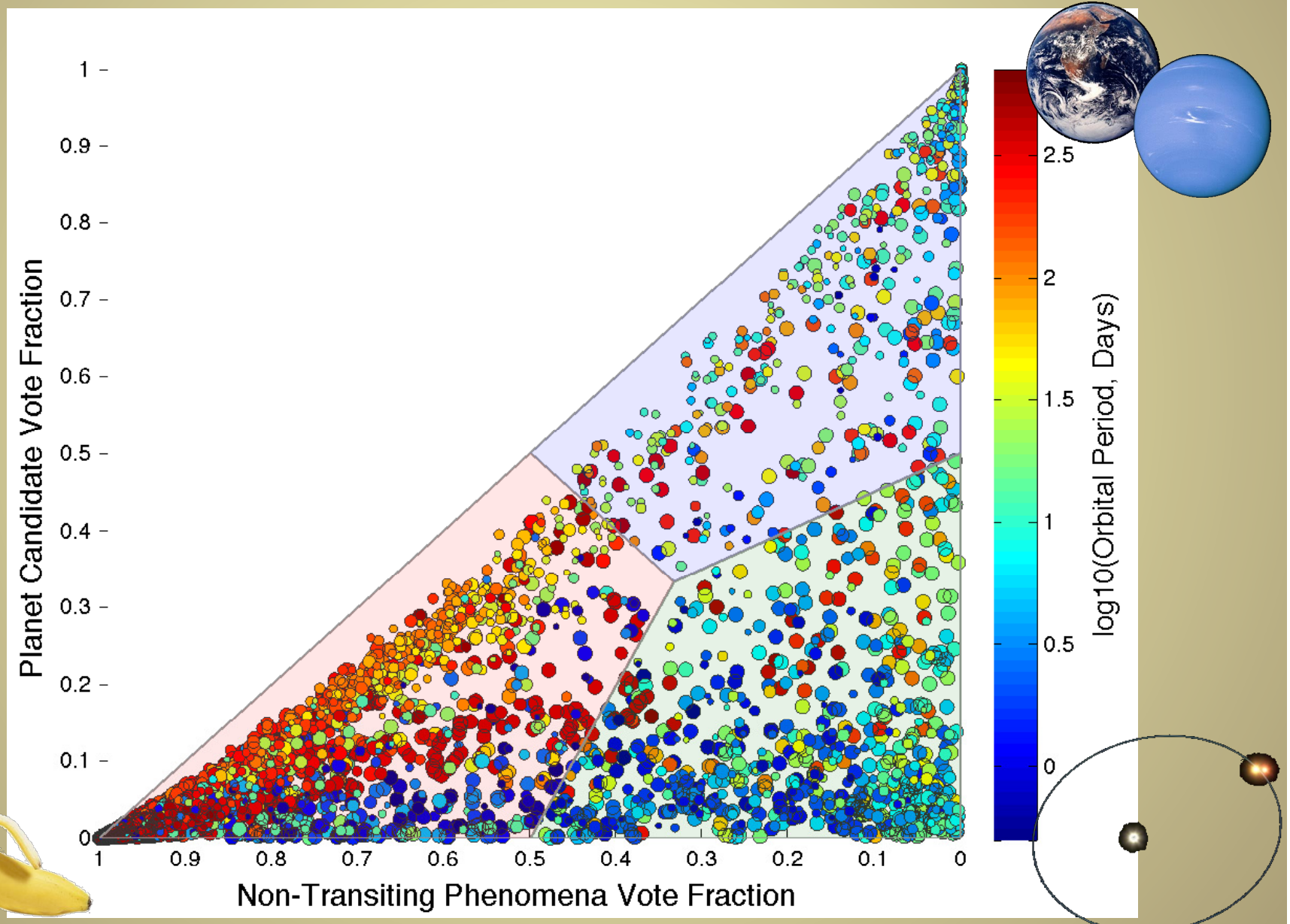
"Astrophysical False Positive" class established
1. Using list of eclipsing binaries from EBWG (less PCs)
2. Using period/epoch collisions (as per Coughlin et al. 2014 AJ 147)

"Non-Transiting Phenomena" class established
1. Using TCERT No-Nos re-identified in Q1-Q16 search
2. Augmented with random 200 "unknown" TCEs subjected to mini-triage

Results:
1. High recovery rates for PCs can be achieved (~96%)
2. AFPs not well defined by training set, given large dispersion
3. NTPs not fully defined by training set

Initial Training Results

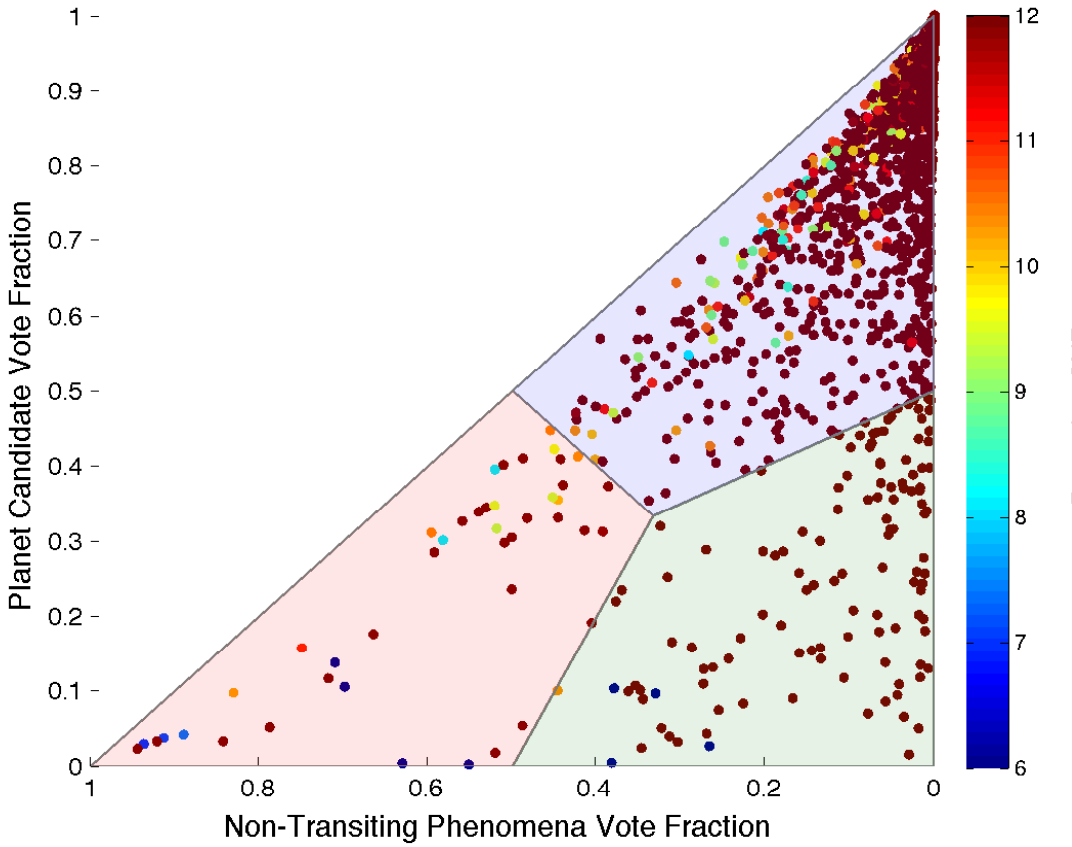
Kepler
A Search for Earth-size Planets

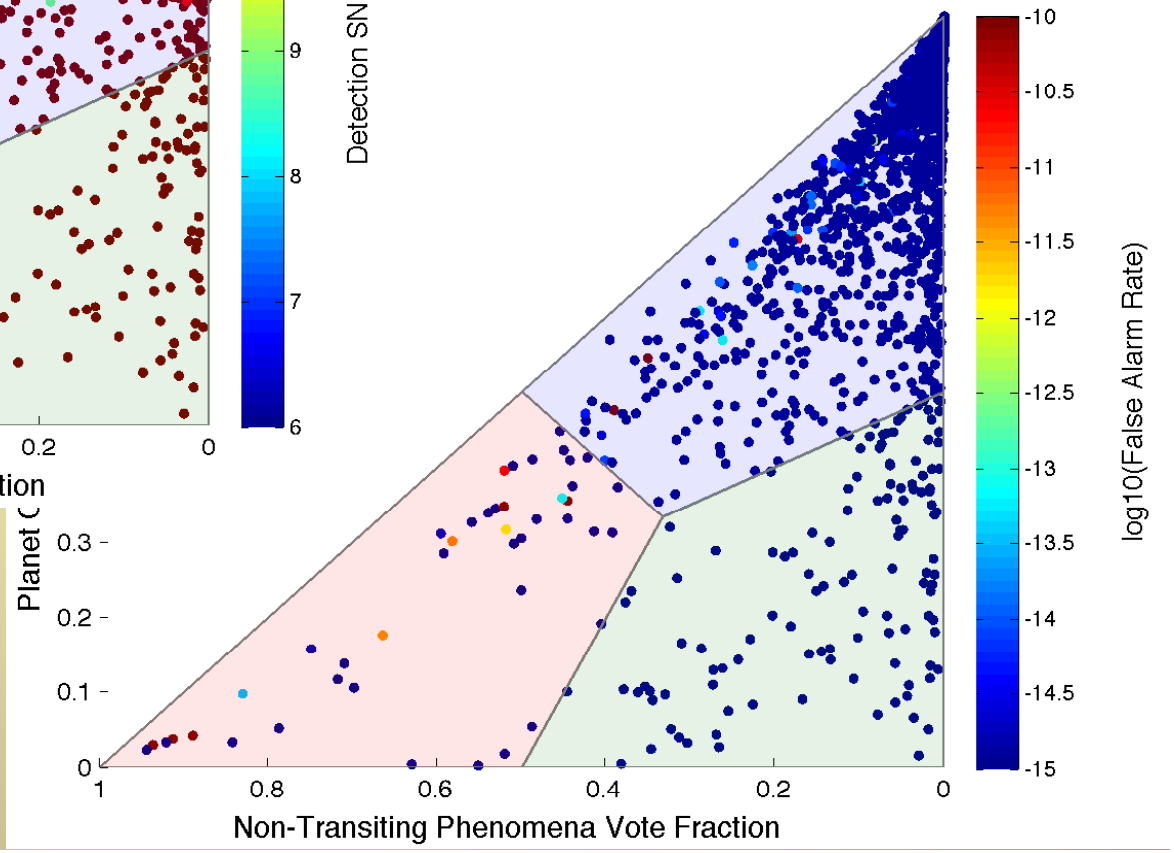


Initial Training Experiments: Planets

Kepler
A Search for Earth-size Planets

KOI Confirmed and Validated Planets and Pre-Q1-Q16 PCs

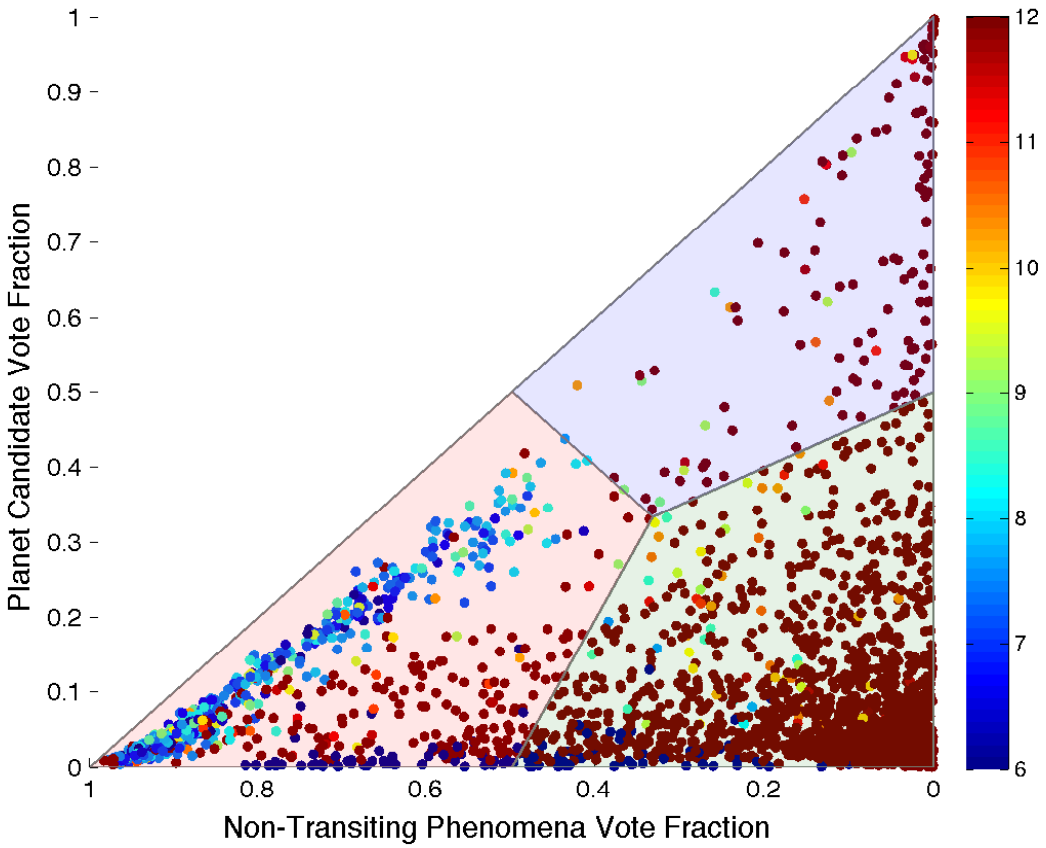


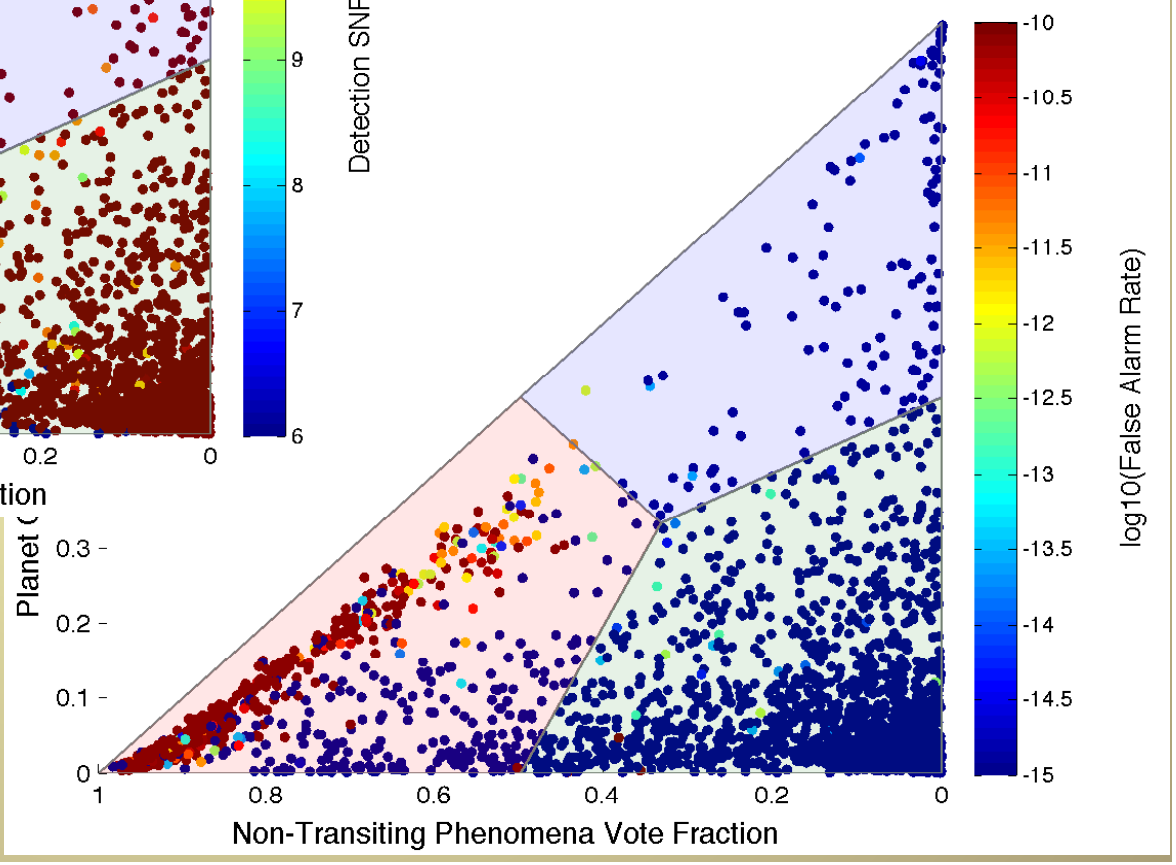


Initial Training Experiments: FPs

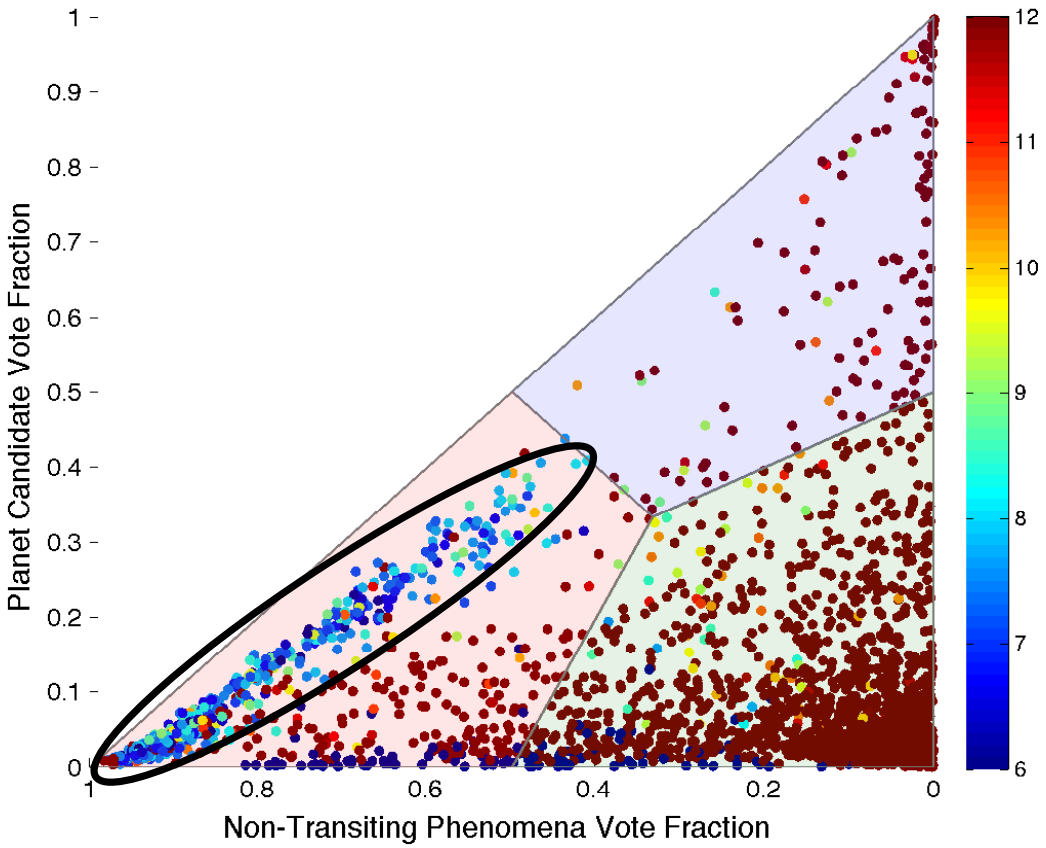
Kepler
A Search for Earth-size Planets

KOI False Positives

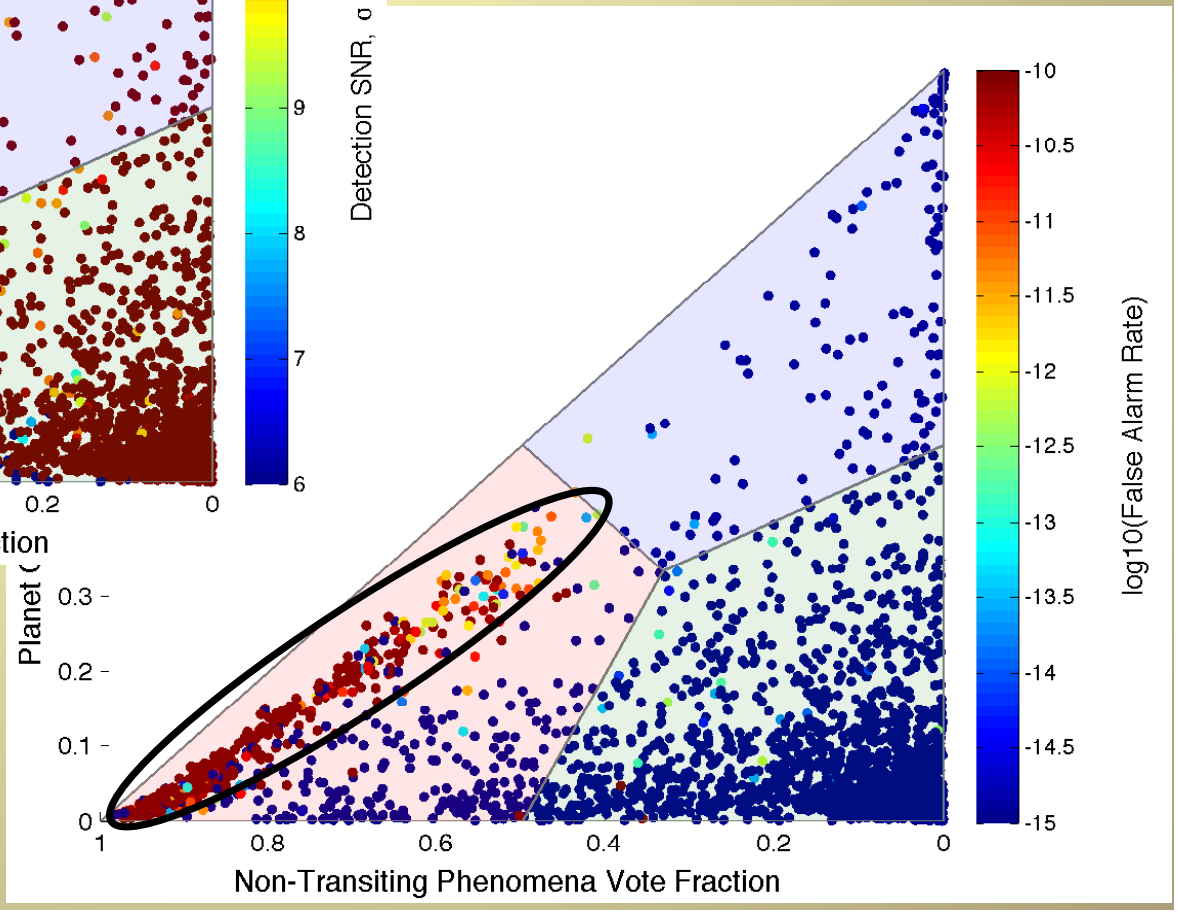




Initial Training Experiments: FPs



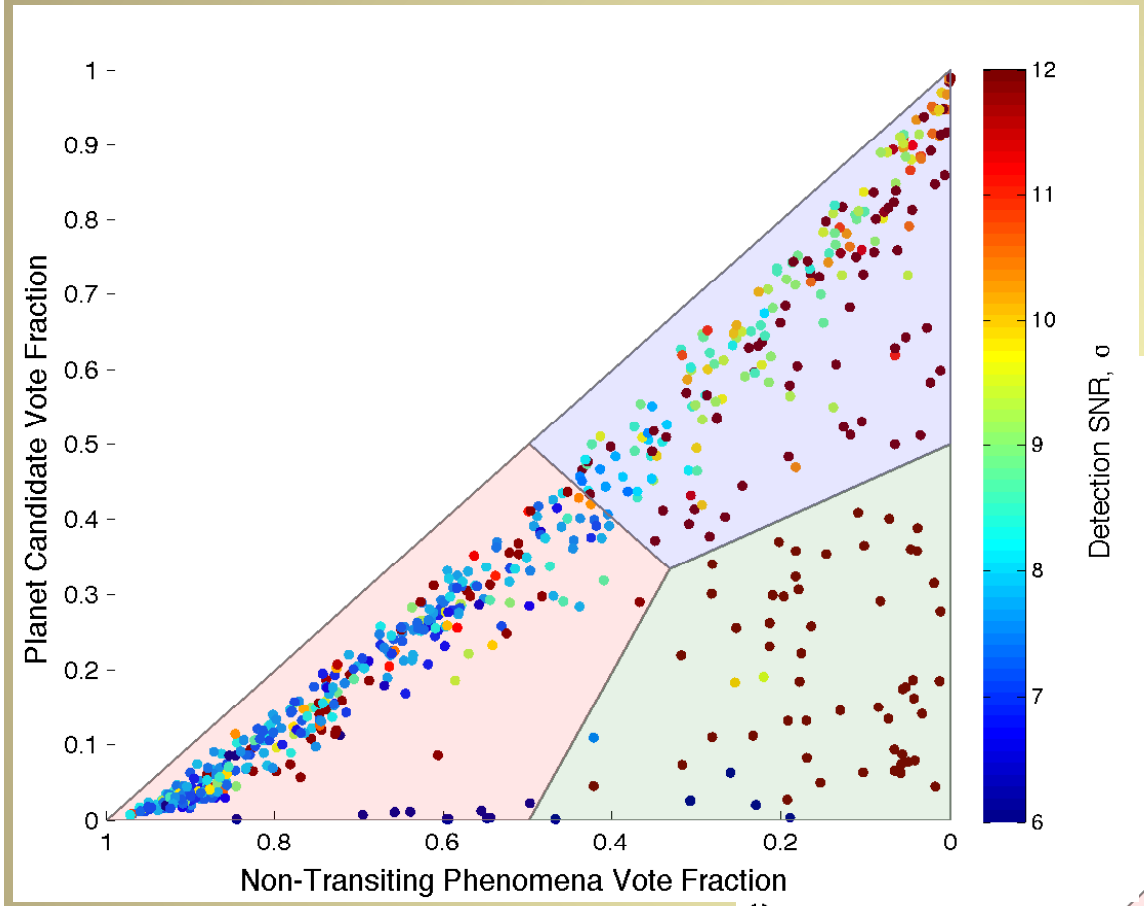
KOI False Positives



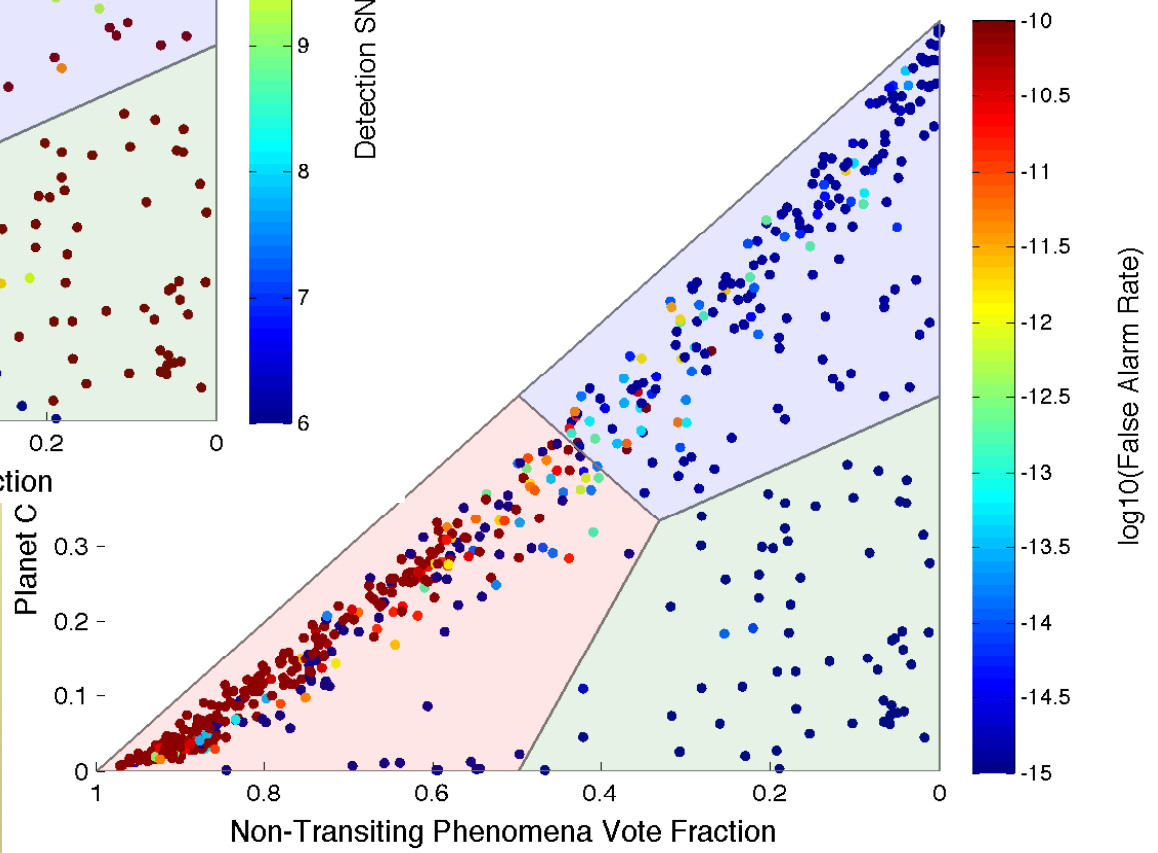
Bootstrap:
Jenkins and Seader 2014,
In preparation

Initial Training Experiments: PCs

Kepler
A Search for Earth-size Planets



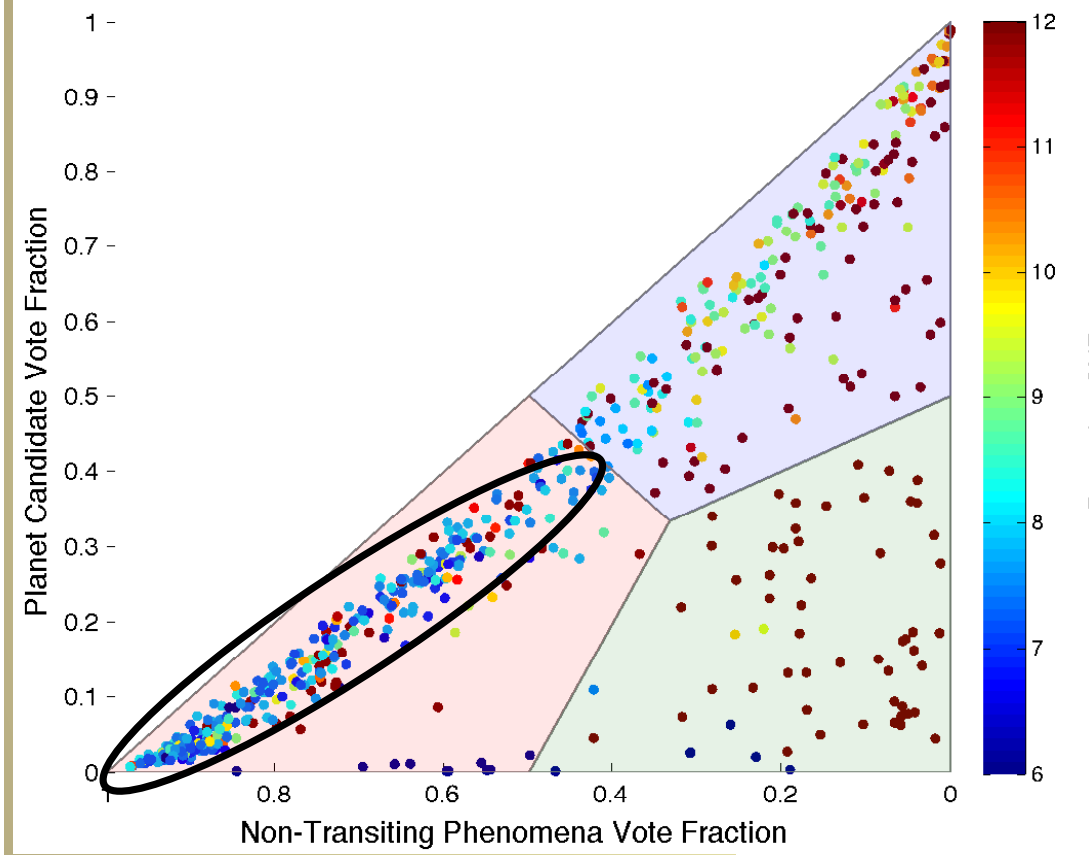
New KOI Planet Candidates*



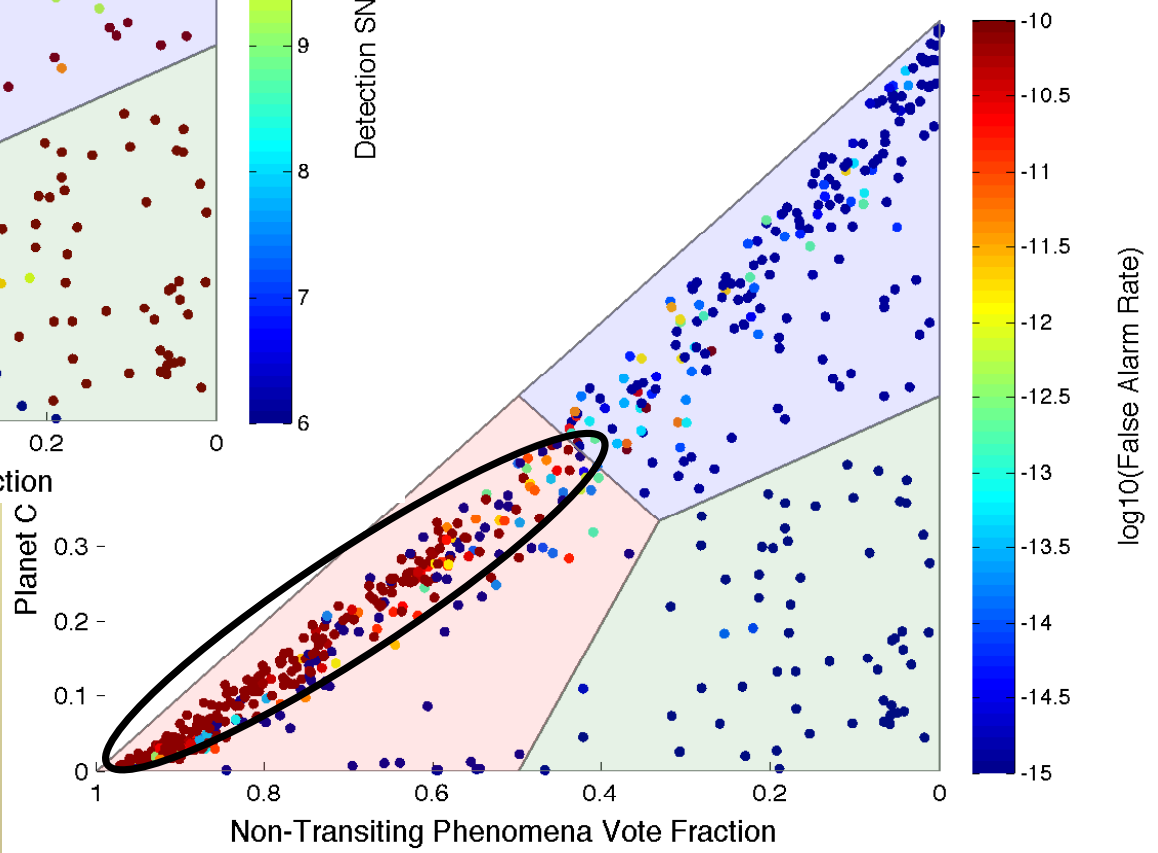
*Q1-Q16 Dispositions are not final and subject to change

Initial Training Experiments: PCs

Kepler
A Search for Earth-size Planets

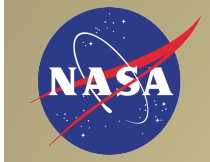


New KOI Planet Candidates*



*Q1-Q16 Dispositions are not final and subject to change

New Training Regimen



A Search for Earth-size Planets

"Planet Candidate" Class Retained

"Astrophysical False Positive" Class Strengthened
1. Shorter/Longer Period Test cut at 0.5
2. TCEs with huge radii (>25 Re) set as AFPs

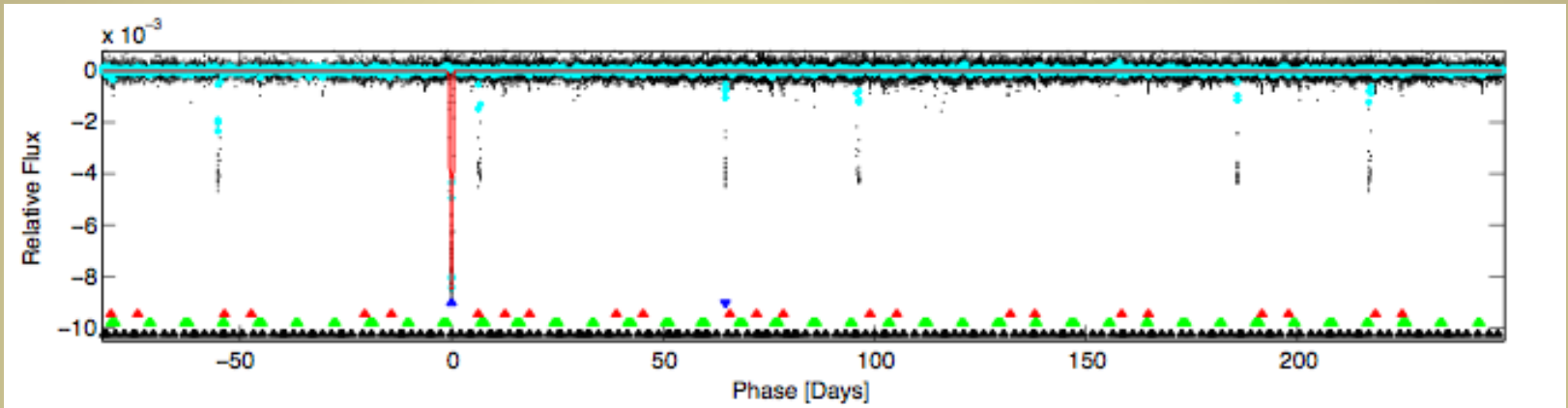
"Non-Transiting Phenomena Class Strengthened
1. Bootstrap FAR cut at $1x10^{-11}$ (conservative)
2. Ratio of robust detection statistic/detection SNR to maxMES cut at 0.5

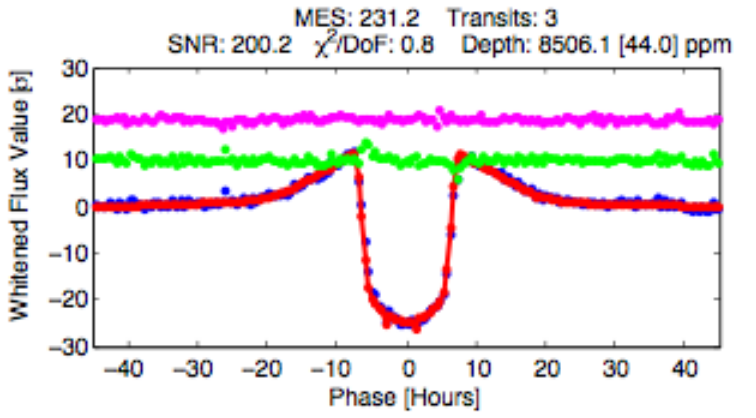
New Diagnostics added:
1. Combination of shorter/longer period tests
2. Ratios of robust detection stat/detection SNR to maxMES
3. Maximum correlation to any other TCE

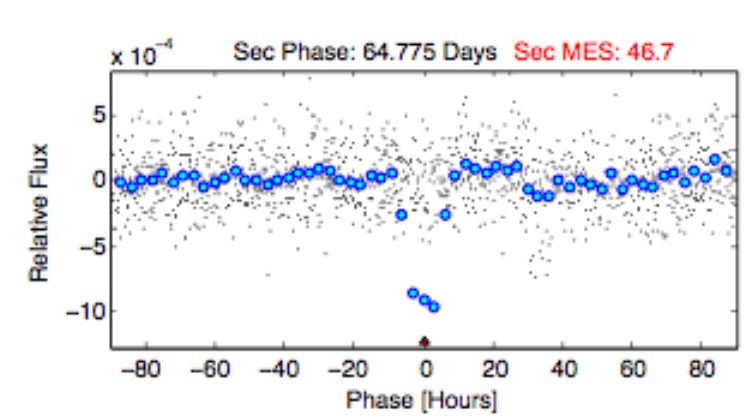
Some Caveats: Kepler-90

Kepler
A Search for Earth-size Planets

Only confirmed planet system failing the bootstrap



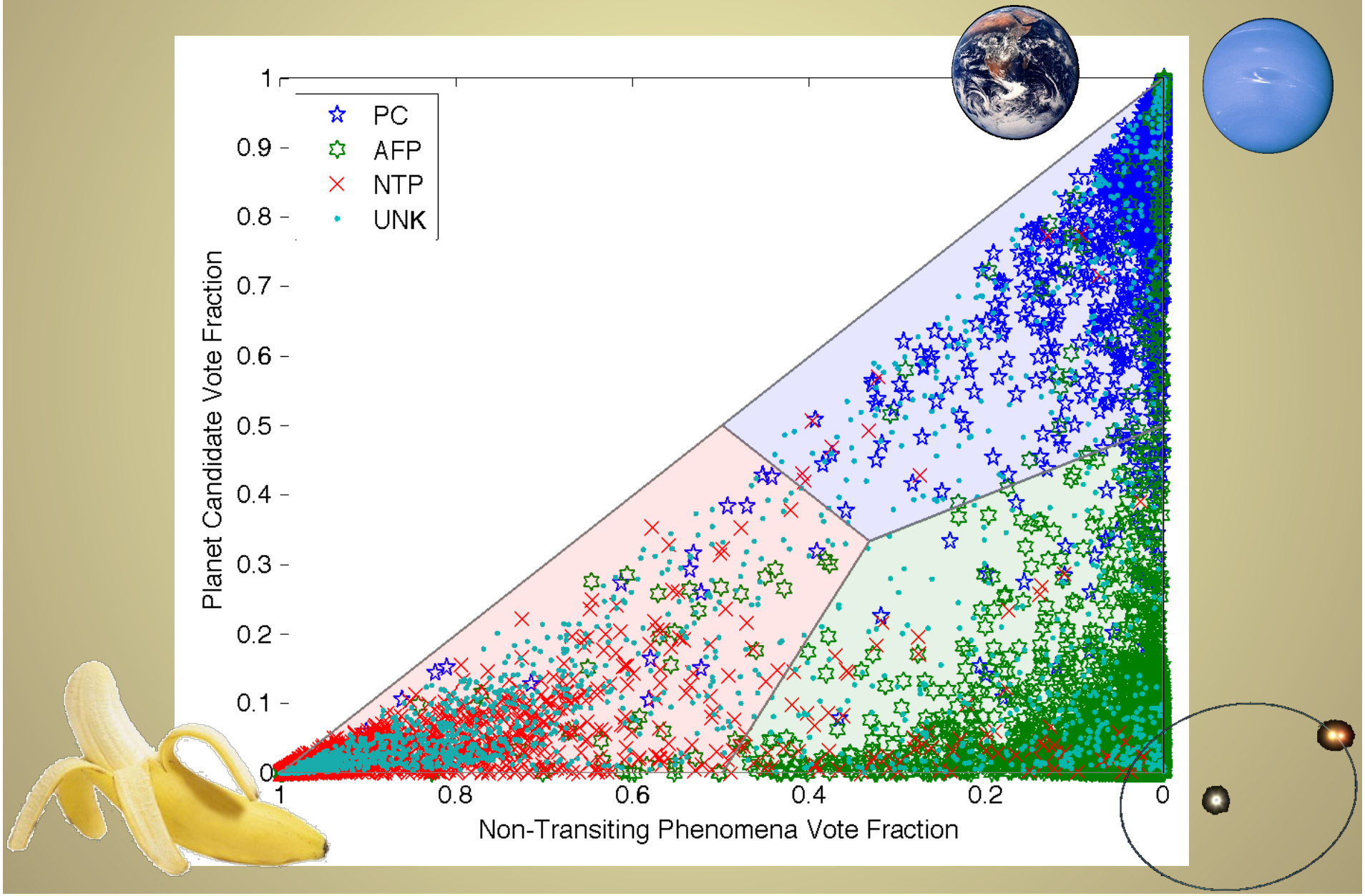




Strong TTVs may fool the bootstrap (and weak secondary test)

Three Class Vetting: Latest Training Results

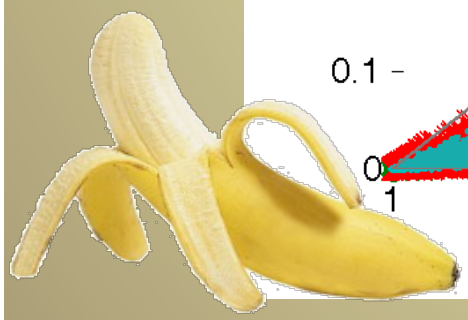
Kepler
A Search for Earth-size Planets

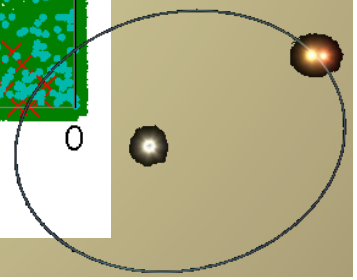


Three Class Vetting: Latest Training Results

Kepler
A Search for Earth-size Planets

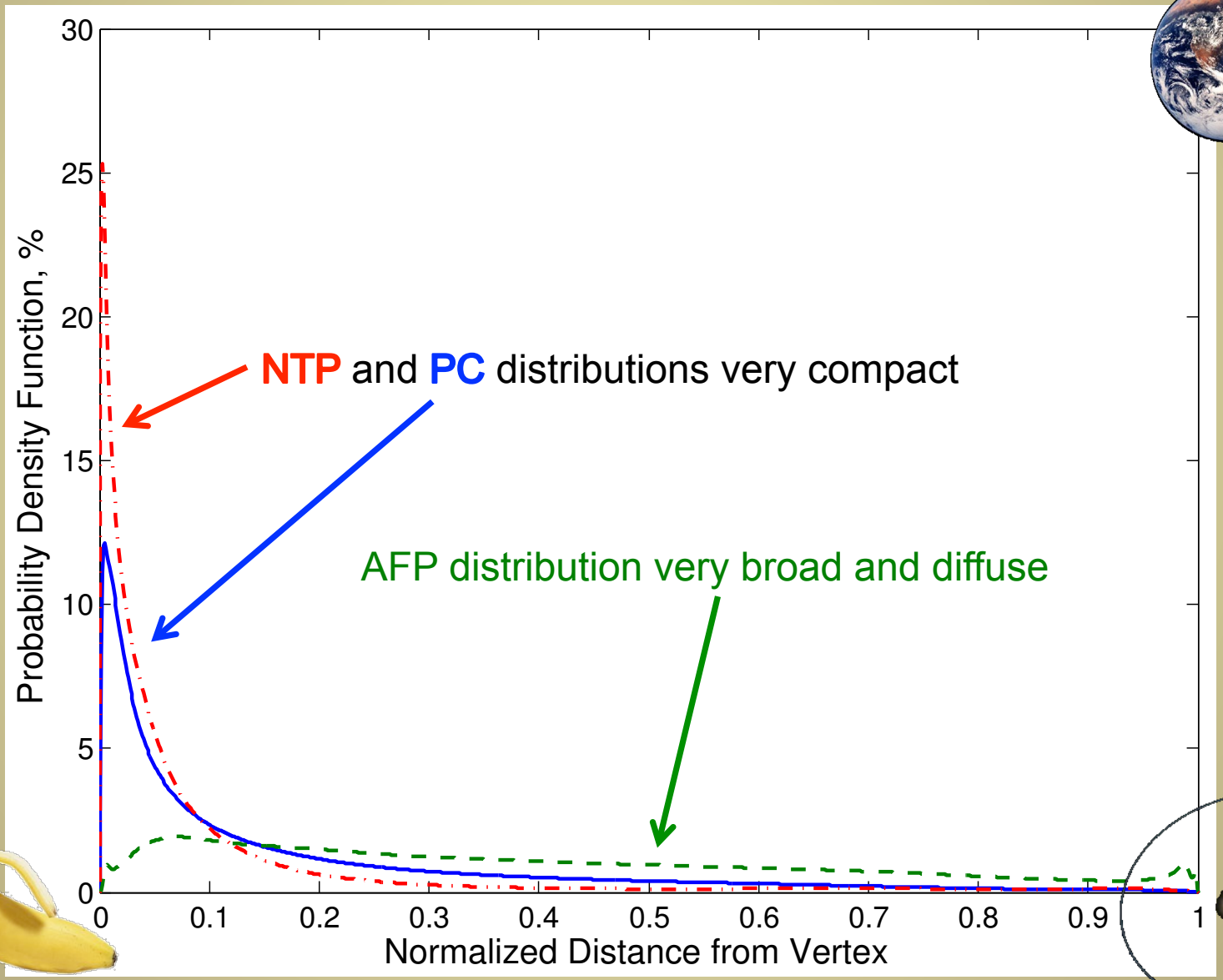
%	Class PC	Class AFP	Class NK	False Discovery Rate
Label PC	97.3	2.1	0.6	3.9
Label AFP	3.2	95.1	1.6	4.3
Label NTP	0.1	1.0	98.9	1.0





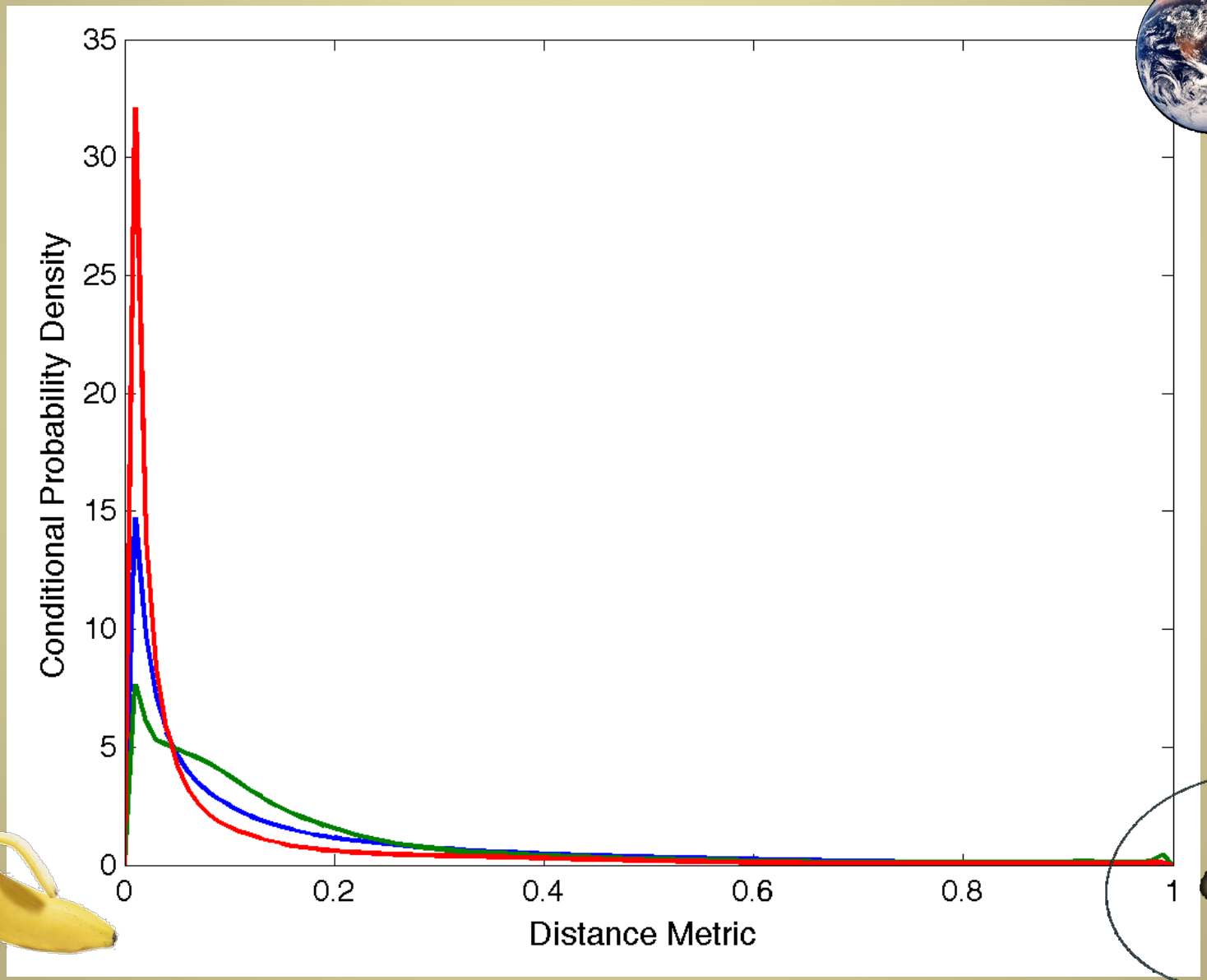
Original Class Conditional Densities

Kepler
A Search for Earth-size Planets



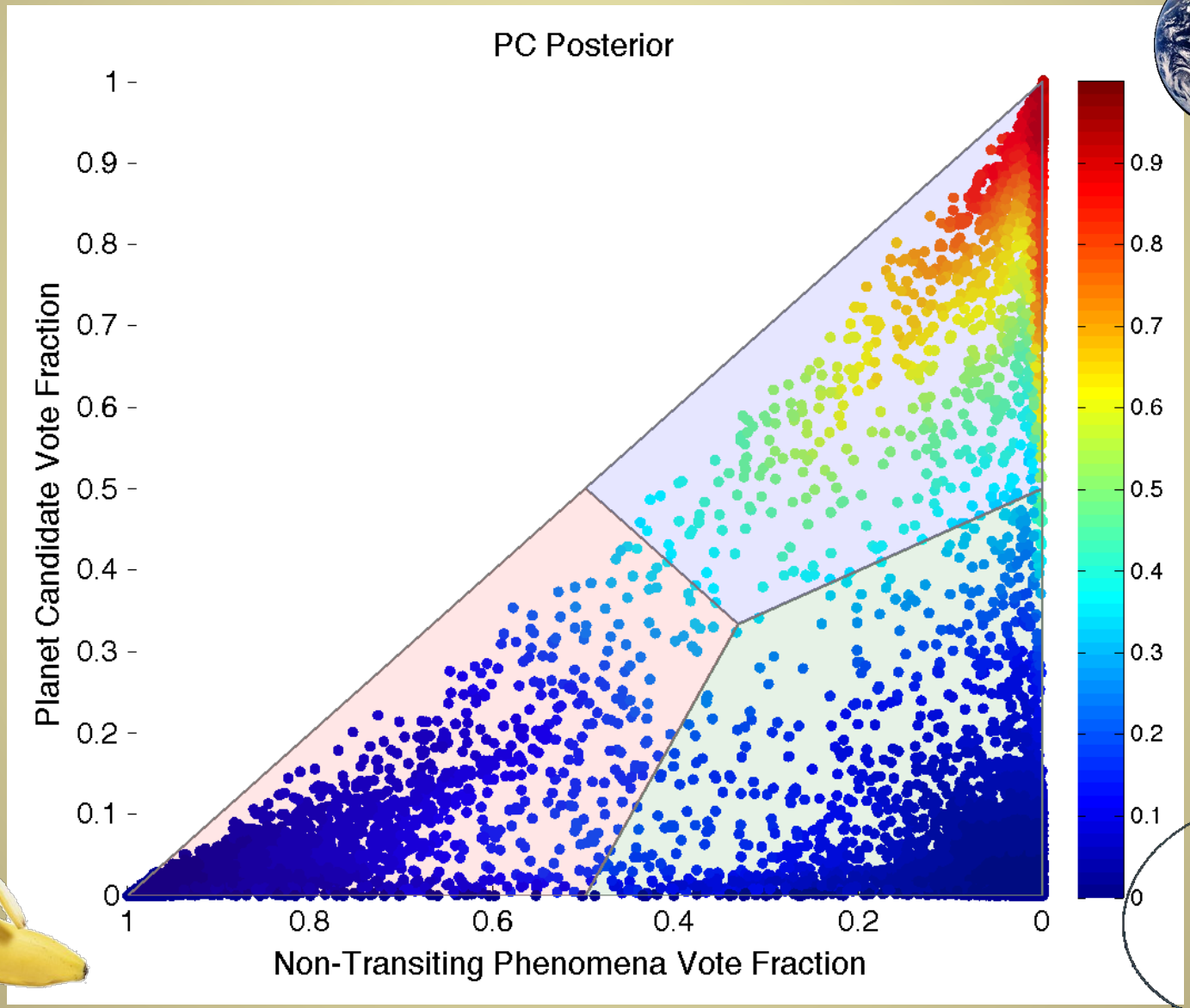
New Class Conditional Densities

Kepler
A Search for Earth-size Planets



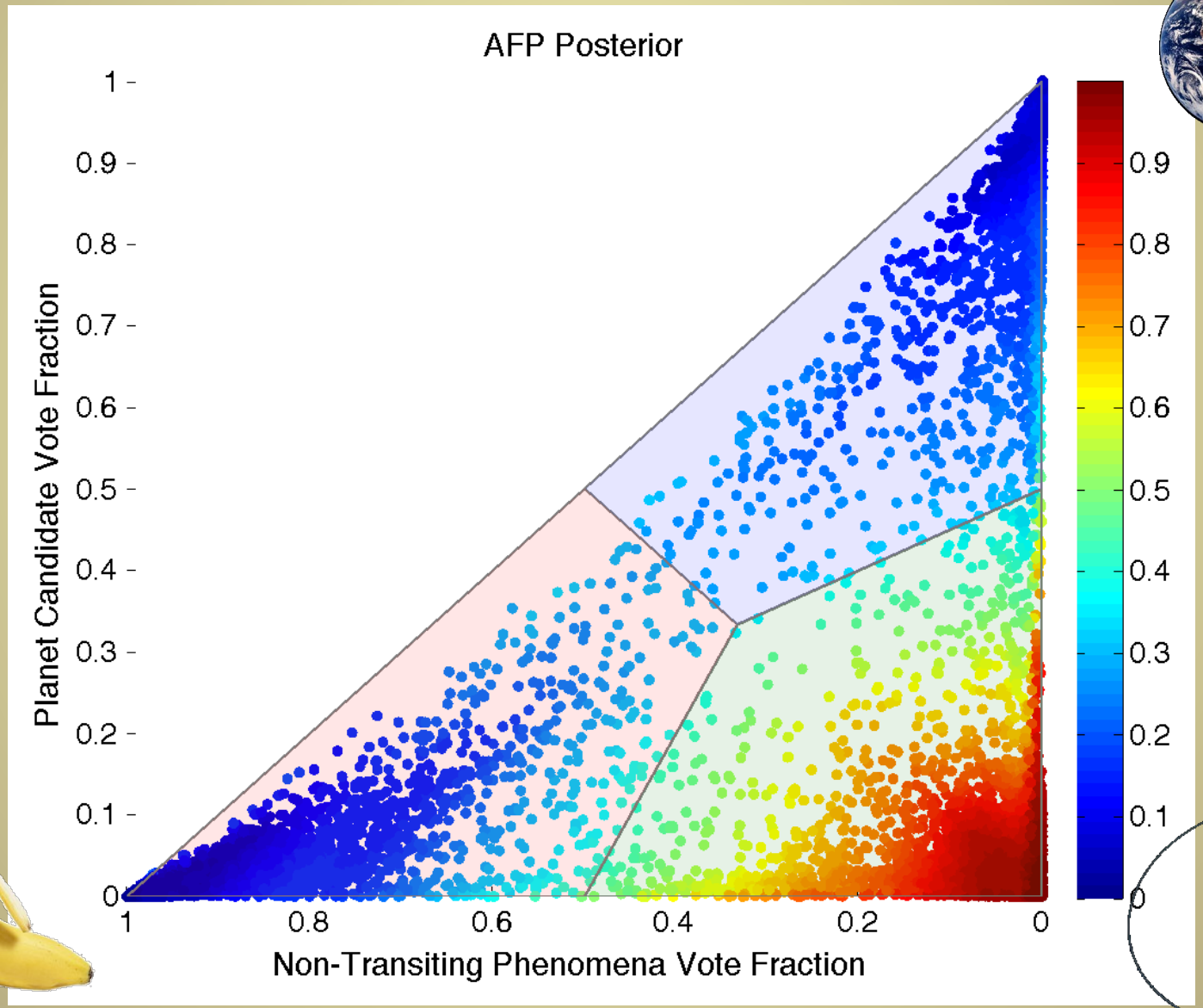
PC Posterior Probabilities

Kepler
A Search for Earth-size Planets



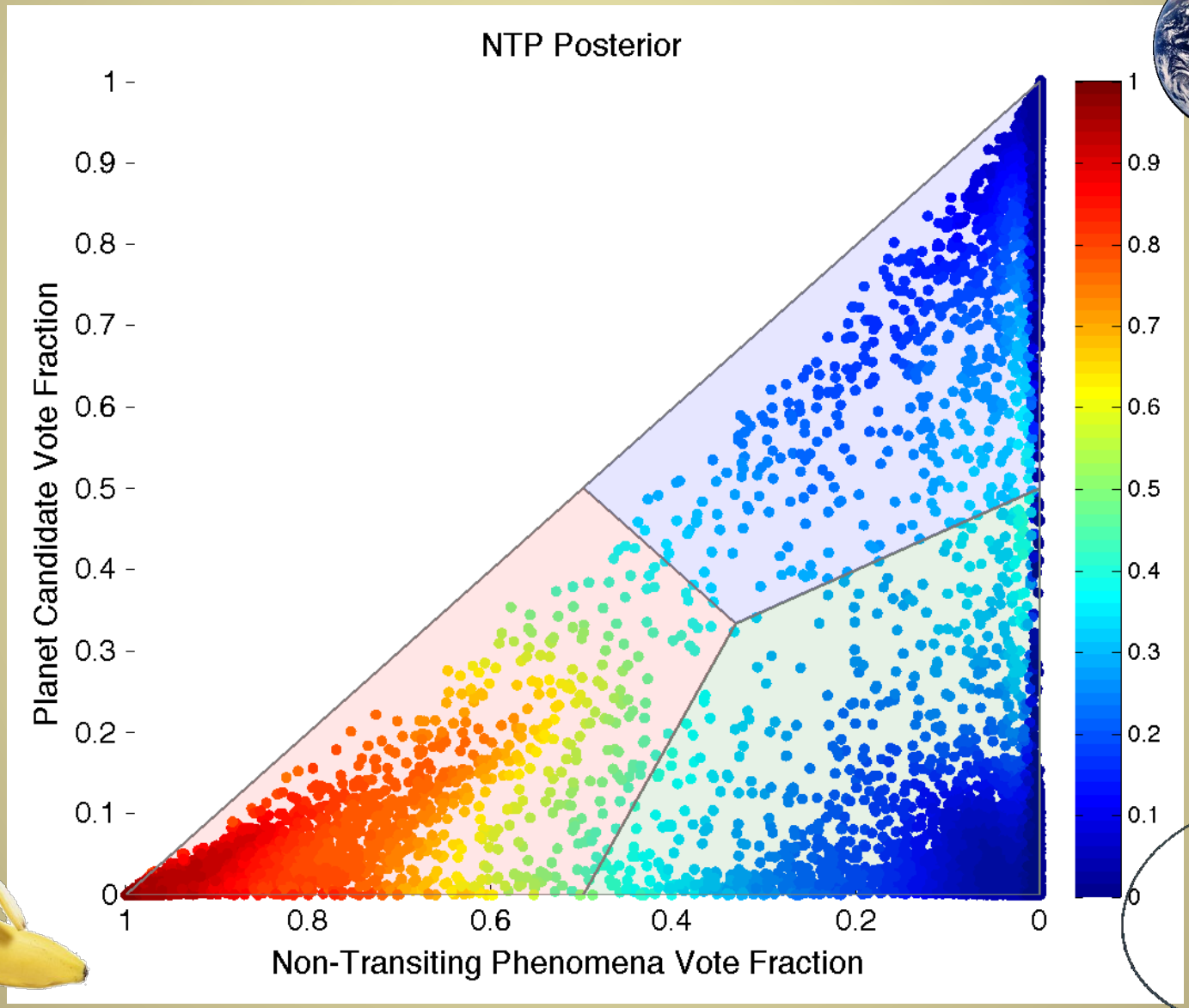
AFP Posterior Probabilities

Kepler
A Search for Earth-size Planets



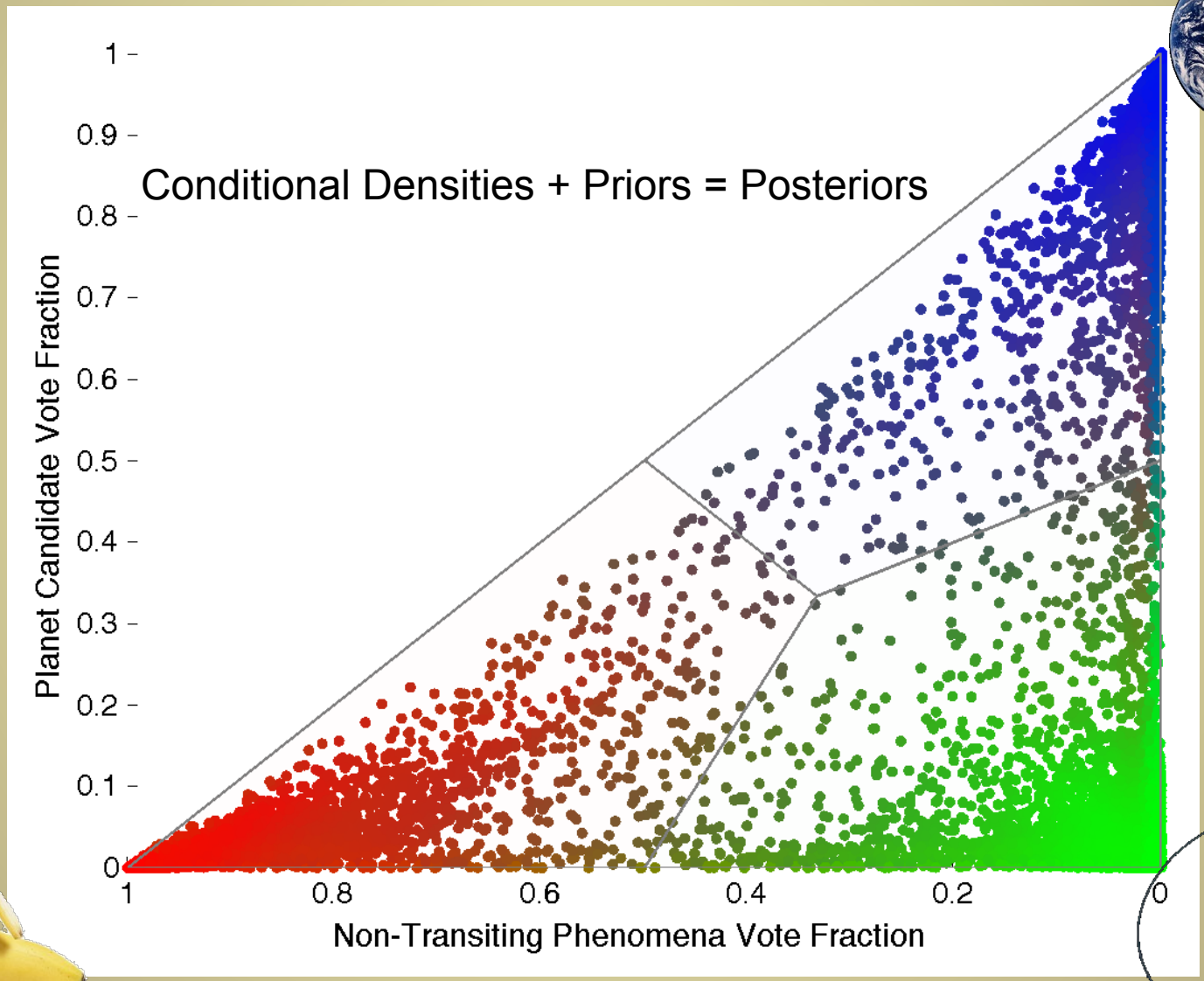
NTP Posterior Probabilities

Kepler
A Search for Earth-size Planets



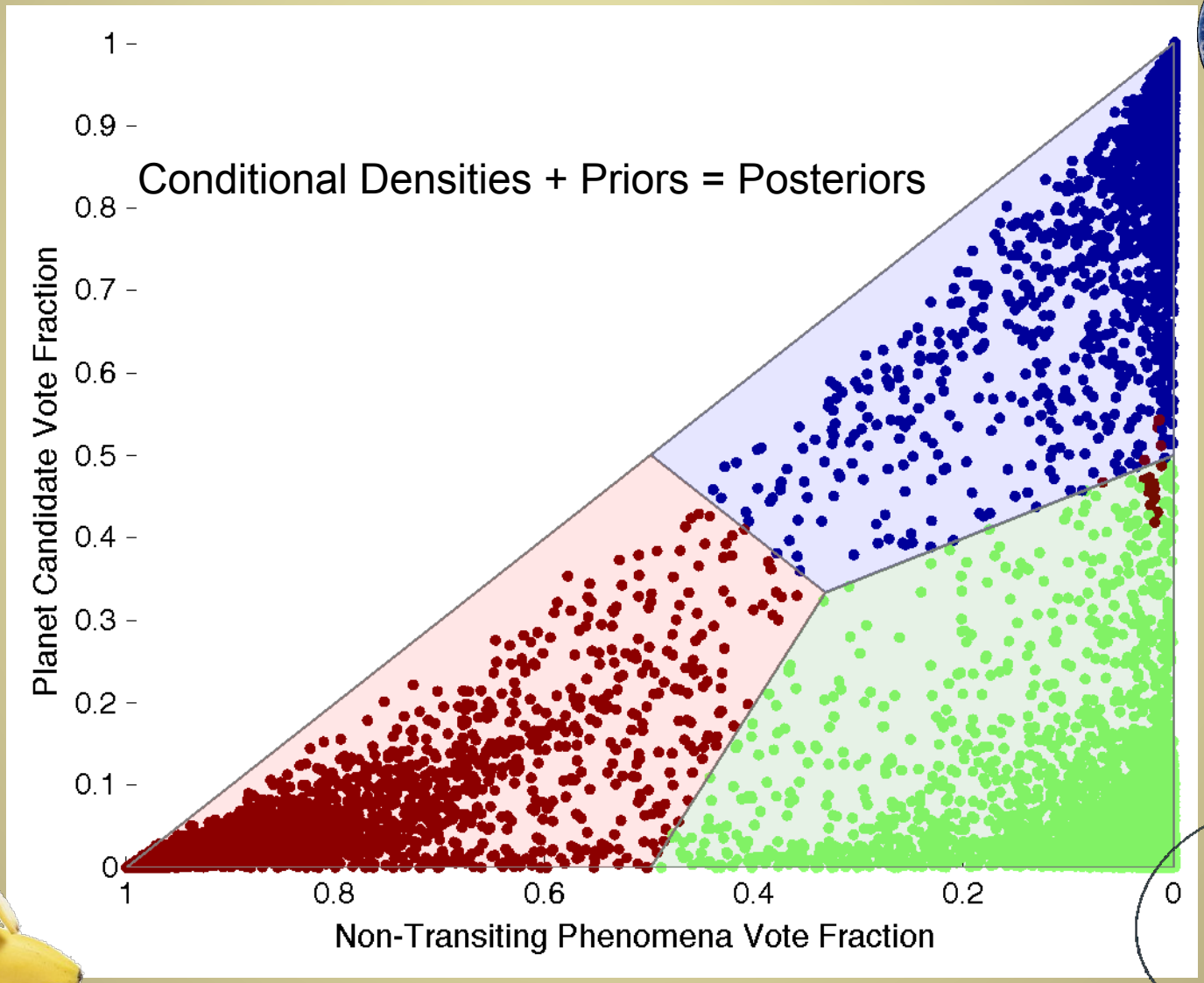
Random Forest Probabilities

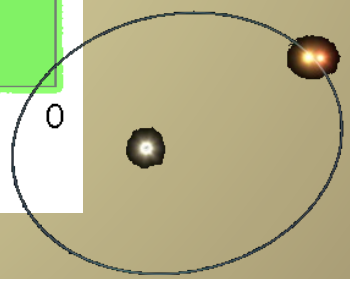
Kepler
A Search for Earth-size Planets



Bayes MAP Classifications

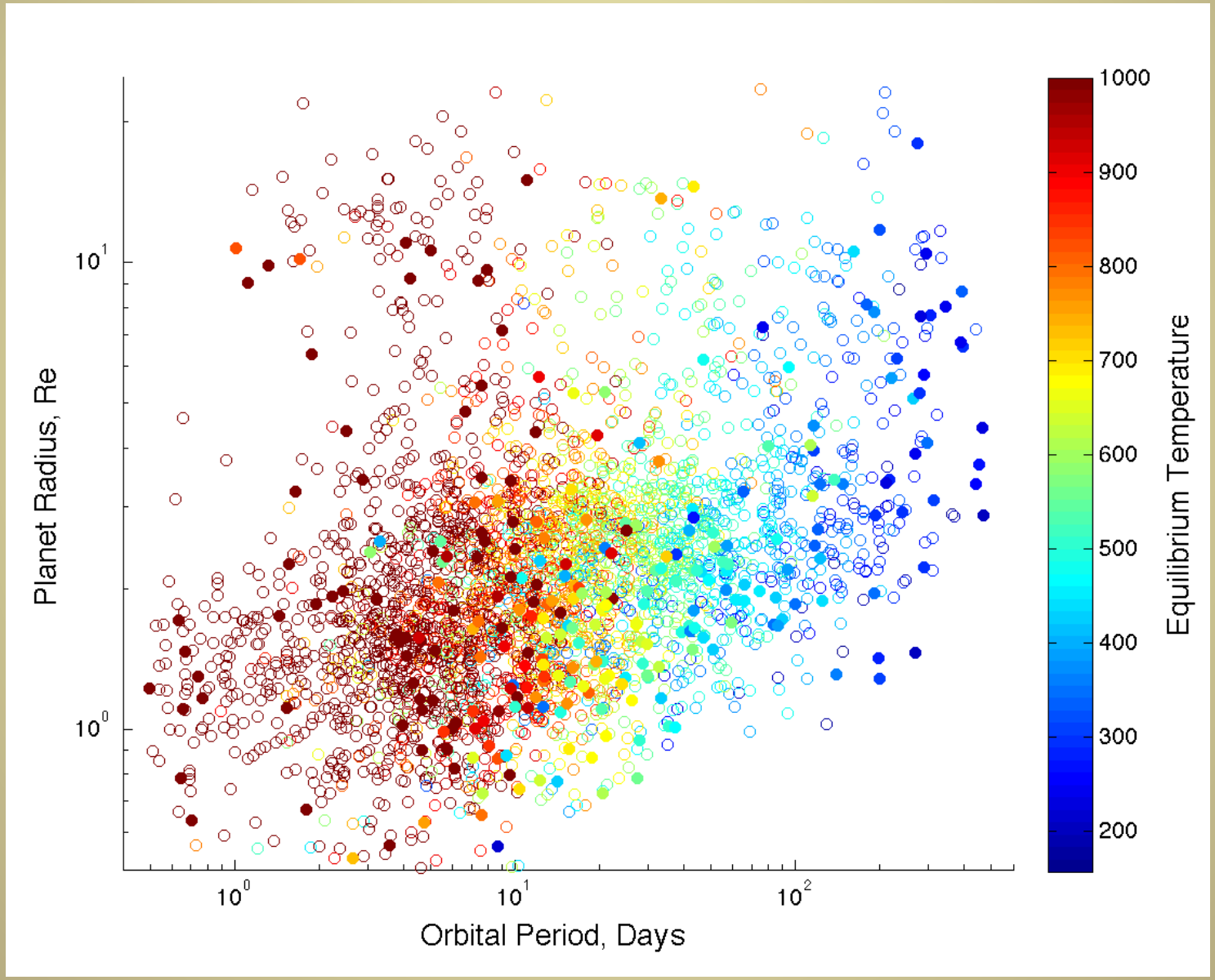
Kepler
A Search for Earth-size Planets

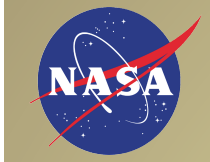




New Likely Planet Candidates

Kepler
A Search for Earth-size Planets





Summary

A Search for Earth-size Planets

- Vetting planet candidates is tough for Kepler and probably will tougher for TESS and PLATO
- Machine learning promises to significantly reduce the amount of time humans need to spend on “duds
- Strong training sets and key diagnostics are important for supervised learning systems
- Random forests can ascribe posterior probabilities to planet candidates identified by transit searches – can be used in occurrence rate calculations
- Transit injection can be used to train the autovetter