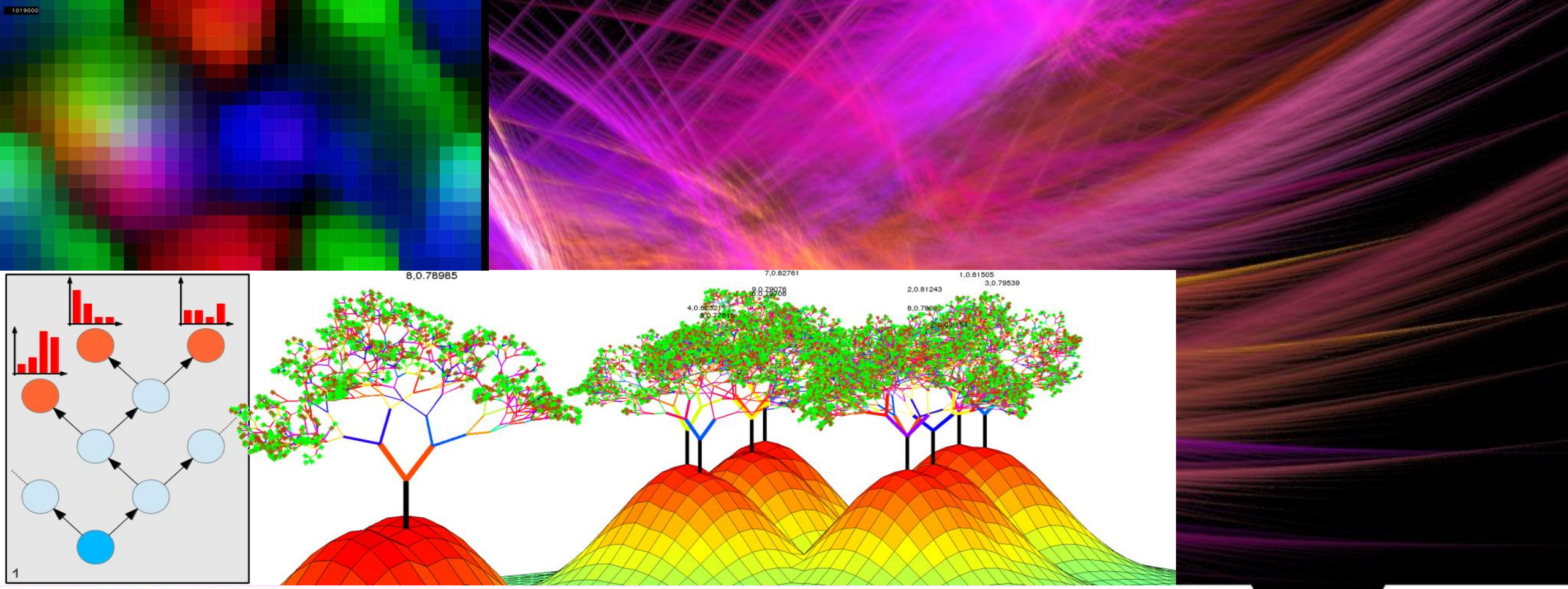


Automatic Candidate Validation in the PLATO Era

David Armstrong, Jevgenij Gamper, Theo Damoulas

WARWICK
THE UNIVERSITY OF WARWICK

05/09/2017

(an example) Planet Detection Process

Kepler DV pipeline (Wu et al 2010)

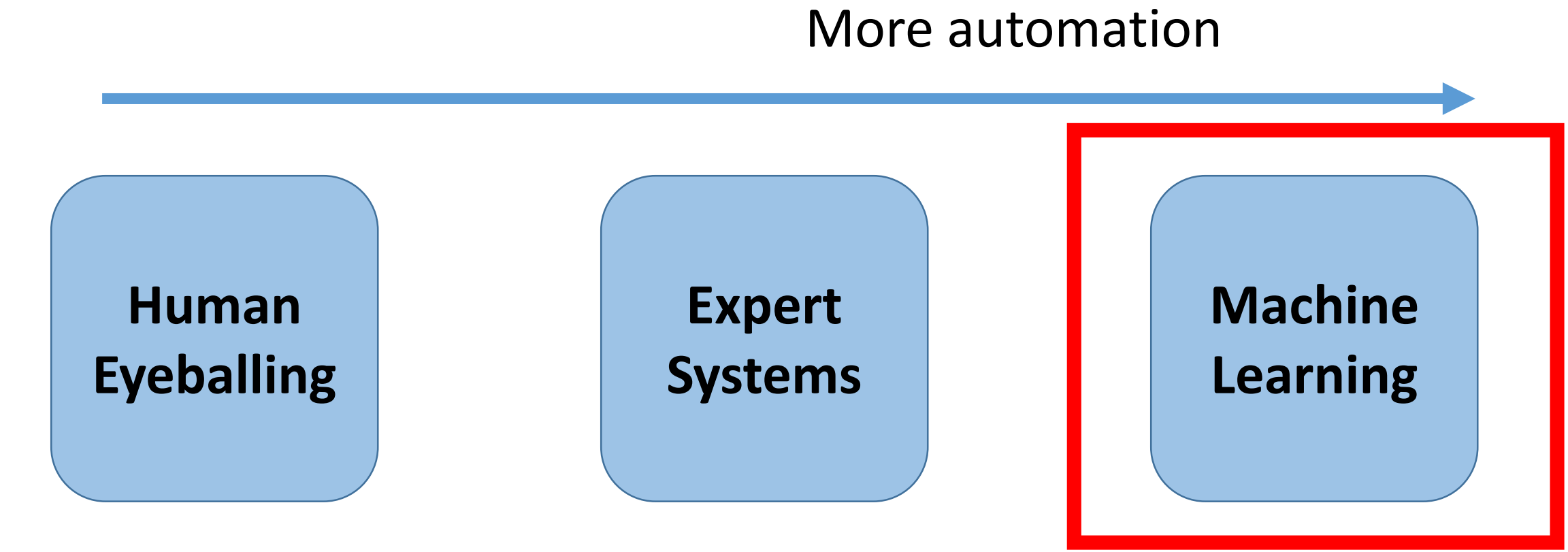
Kepler Autovetter (McCauliff et al 2015)

Eyeballing

Follow-up (RVs, AO, etc)

VESPA, BLENDER, PASTIS

Candidate Vetting



Time-consuming, statistically awkward

Hard to interpret

What are the best ML methods here?

Can we use ML for planet validation too?

Can we extract posterior probabilities?

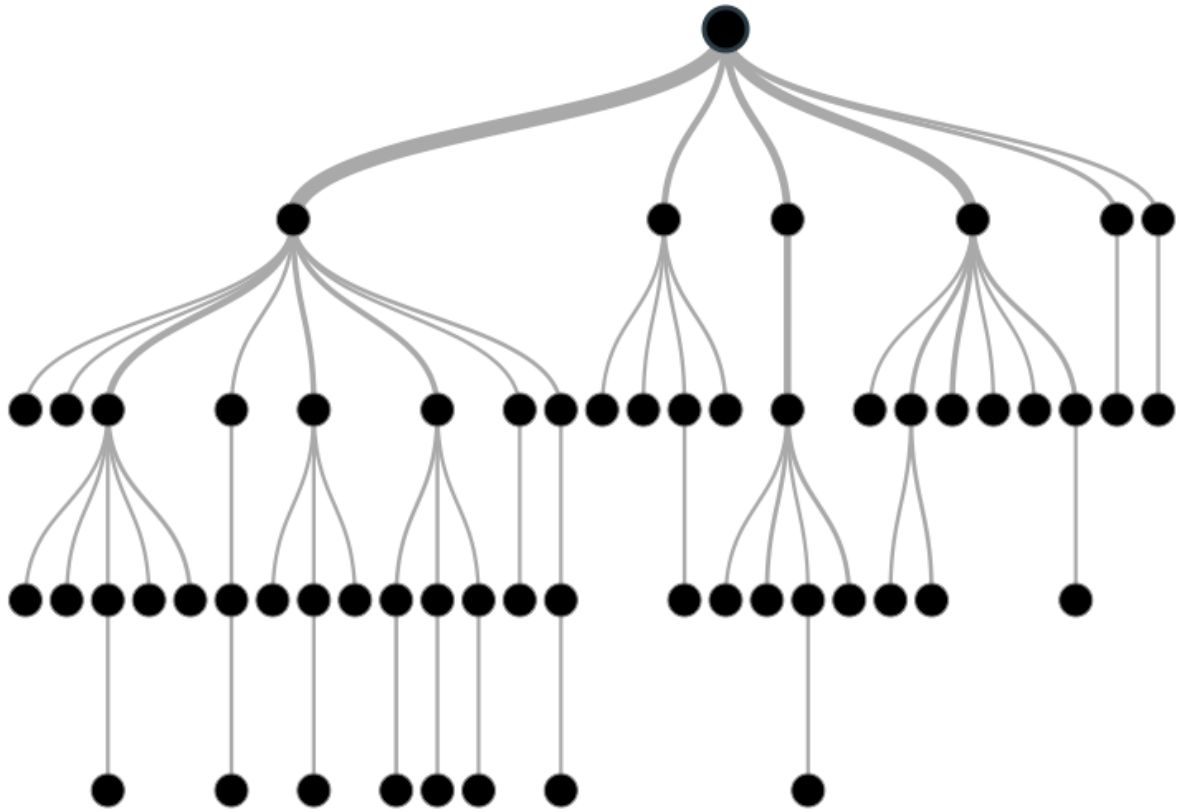
Kepler Testing

- Cumulative KOI catalogue
- 2238 confirmed planets, 1810 false positives
- Focus on separating confirmed and false positives – not identifying good candidates within the TCEs
- 149 features. 33 after importance and correlation cuts.

Methods

- 11 different methods, with a range of complexity

- Single decision tree, other simple methods (SVM, K-NN), neural net, random forest varieties



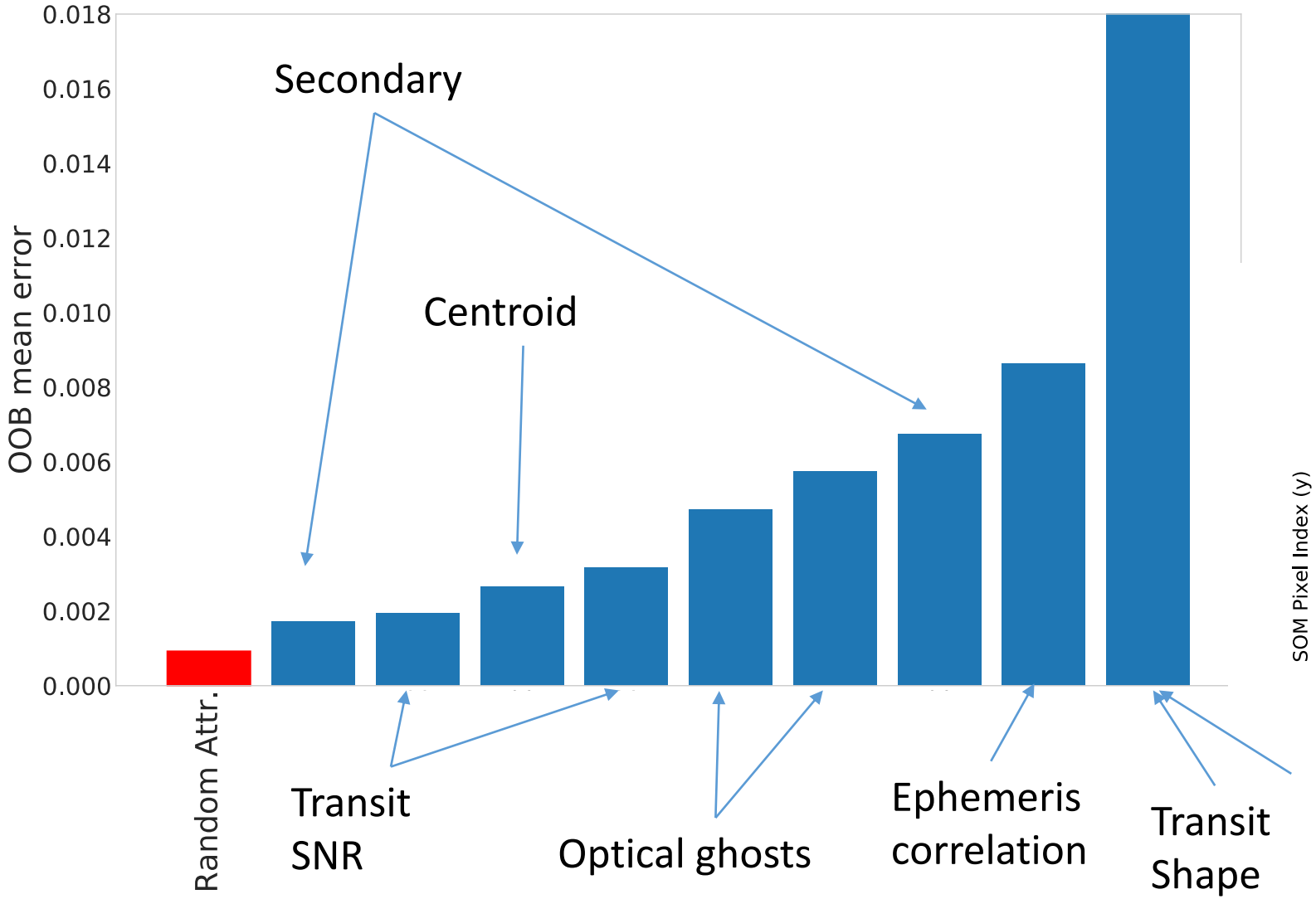
Decision Tree Ensembles

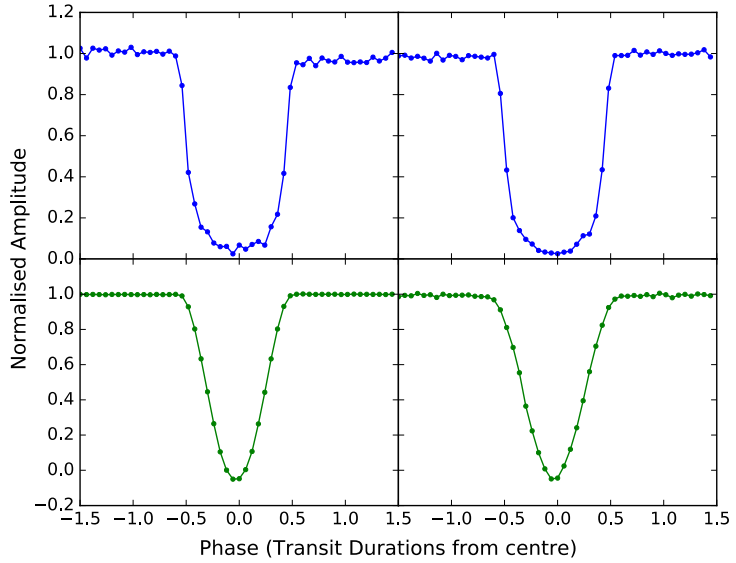
Simple linear classifiers

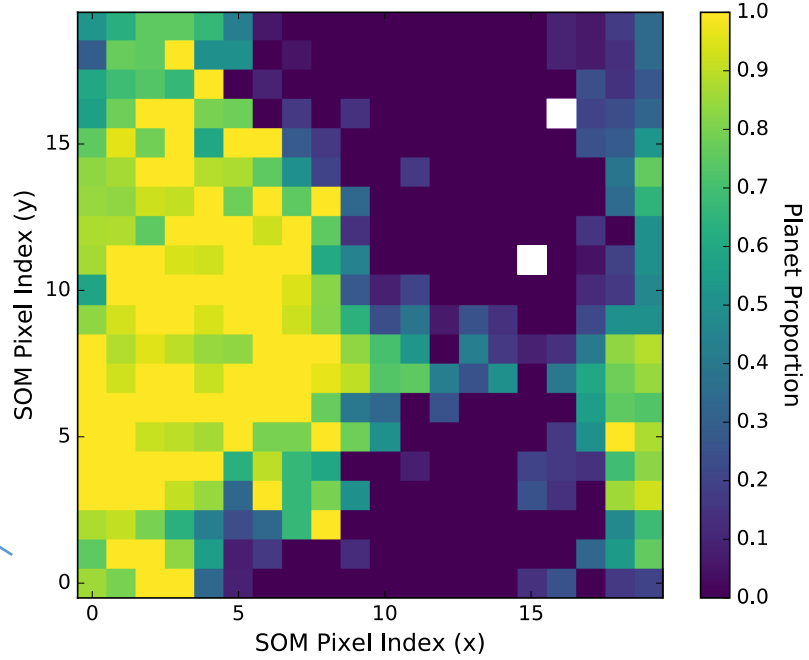
Attempted complexity

Classifier	AUC	Precision	Recall	Brier
Random Forest	0.99	0.96	0.94	0.03
Extra Trees	0.99	0.97	0.93	0.04
Logistic Regression	0.97	0.94	0.90	0.06
LDA	0.97	0.92	0.87	0.07
Ridge Classifier	0.97	0.92	0.87	0.10
SVM	0.97	0.92	0.90	0.17
K-NN	0.96	0.98	0.89	0.06
MLP	0.96	0.92	0.89	0.07
QDA	0.95	0.96	0.57	0.20
Ada-Boost (DT)	0.94	0.93	0.93	0.10
Decision Tree	0.92	0.95	0.93	0.06

Features



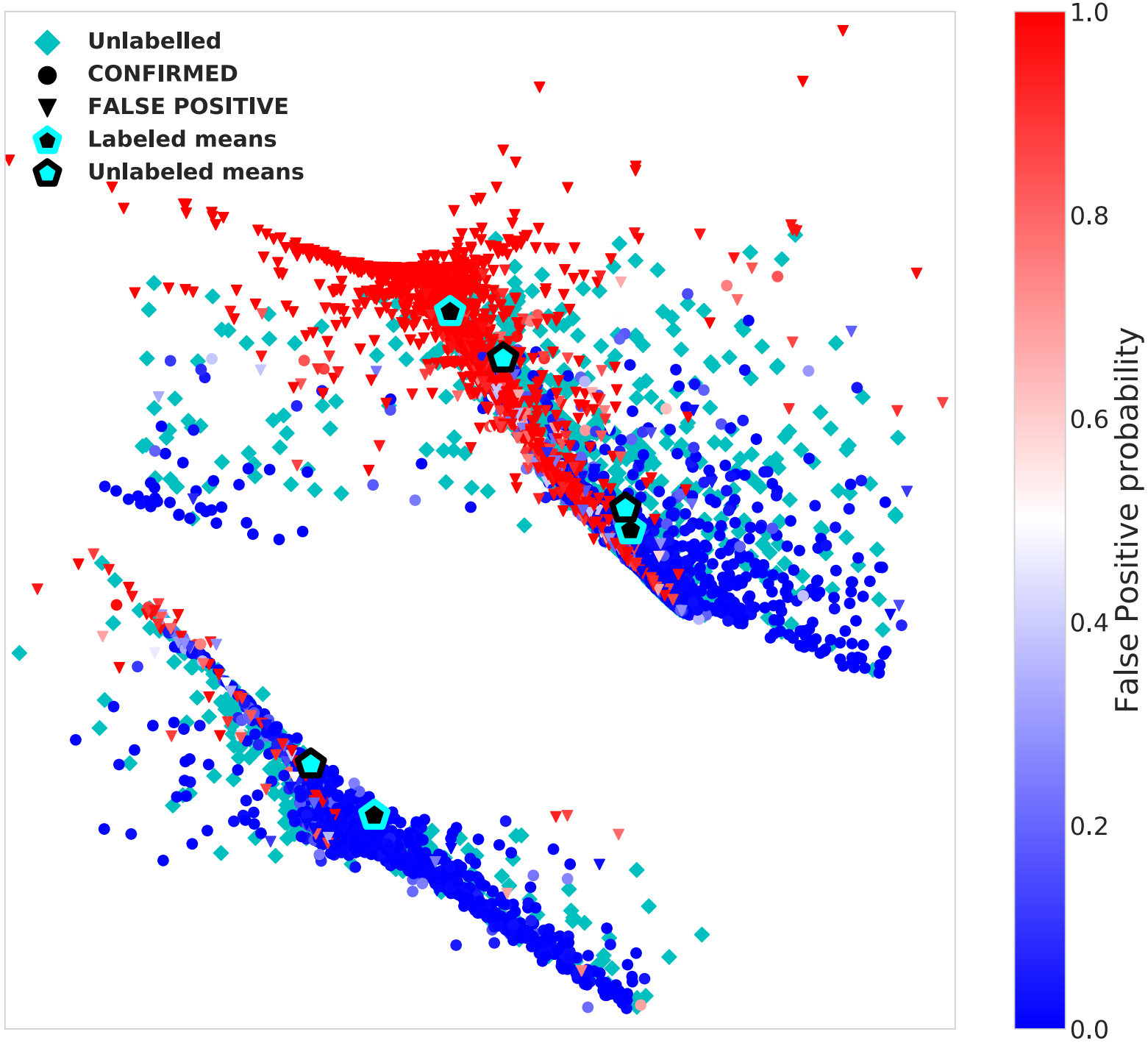




Armstrong et al 2017

Projections

- Visualisation of the feature space
- Labelled distributed like unlabelled
- Outliers – affect the random forest



Decision Tree Ensembles

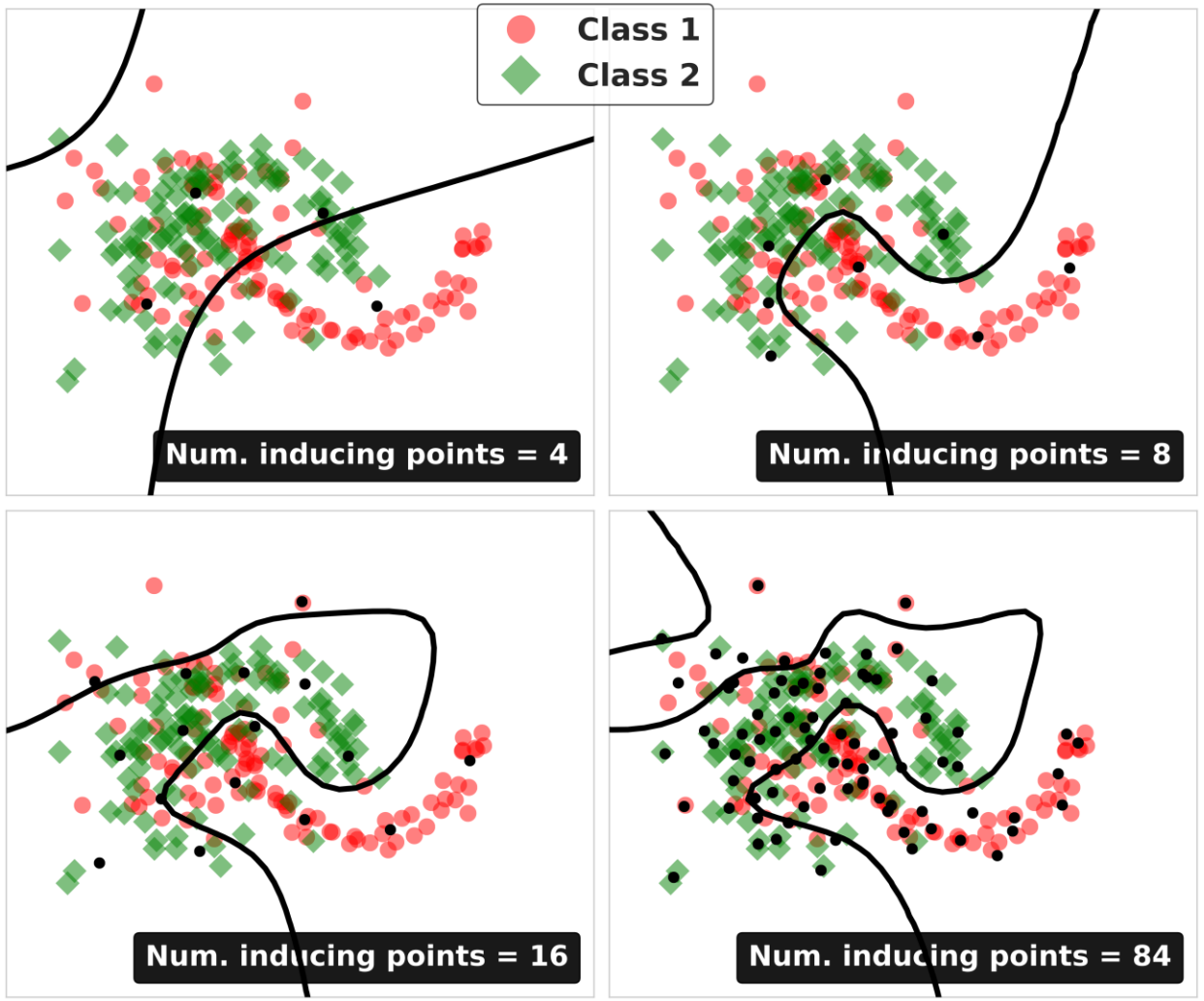
Simple linear classifiers

Attempted complexity

Classifier	AUC	Precision	Recall	Brier
Random Forest	0.99	0.96	0.94	0.03
Extra Trees	0.99	0.97	0.93	0.04
Logistic Regression	0.97	0.94	0.90	0.06
LDA	0.97	0.92	0.87	0.07
Ridge Classifier	0.97	0.92	0.87	0.10
SVM	0.97	0.92	0.90	0.17
K-NN	0.96	0.98	0.89	0.06
MLP	0.96	0.92	0.89	0.07
QDA	0.95	0.96	0.57	0.20
Ada-Boost (DT)	0.94	0.93	0.93	0.10
Decision Tree	0.92	0.95	0.93	0.06

GP Classifier

- Provide uncertainties and probabilities
- Forced simplicity - Linear kernel and inducing points
- **Still ~98% - but not overconfident where it shouldn’t be**

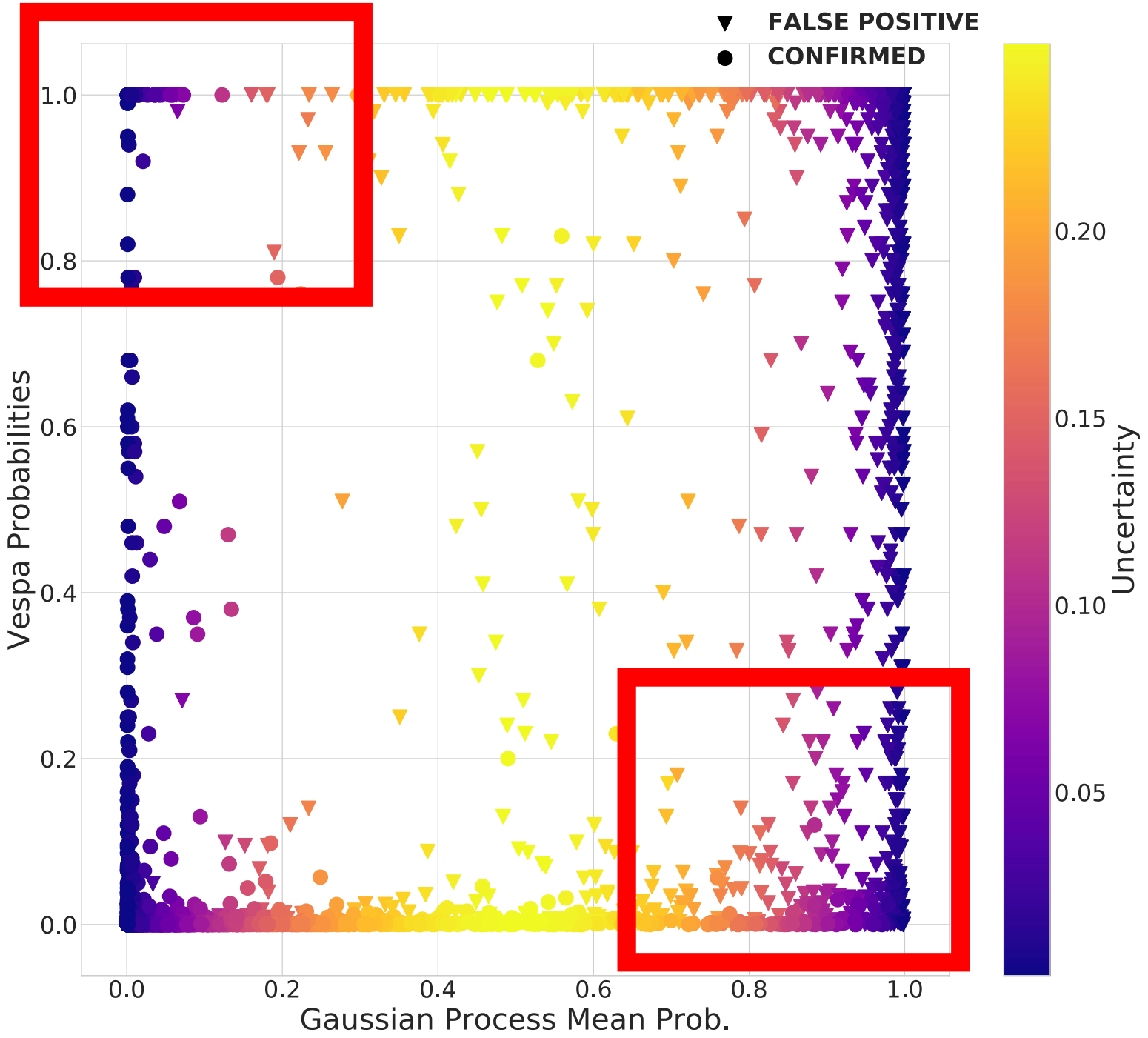


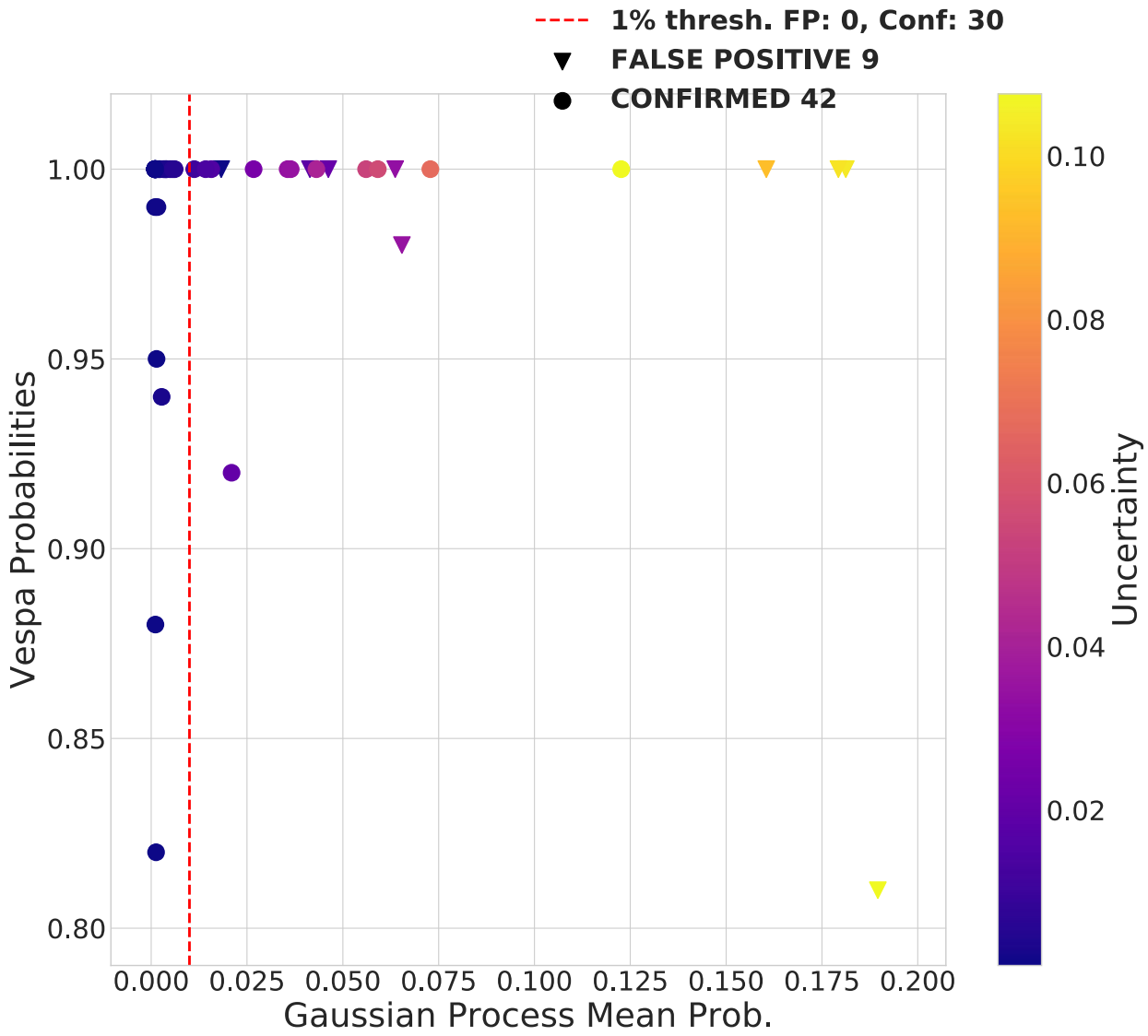
Are the Planets Planets?

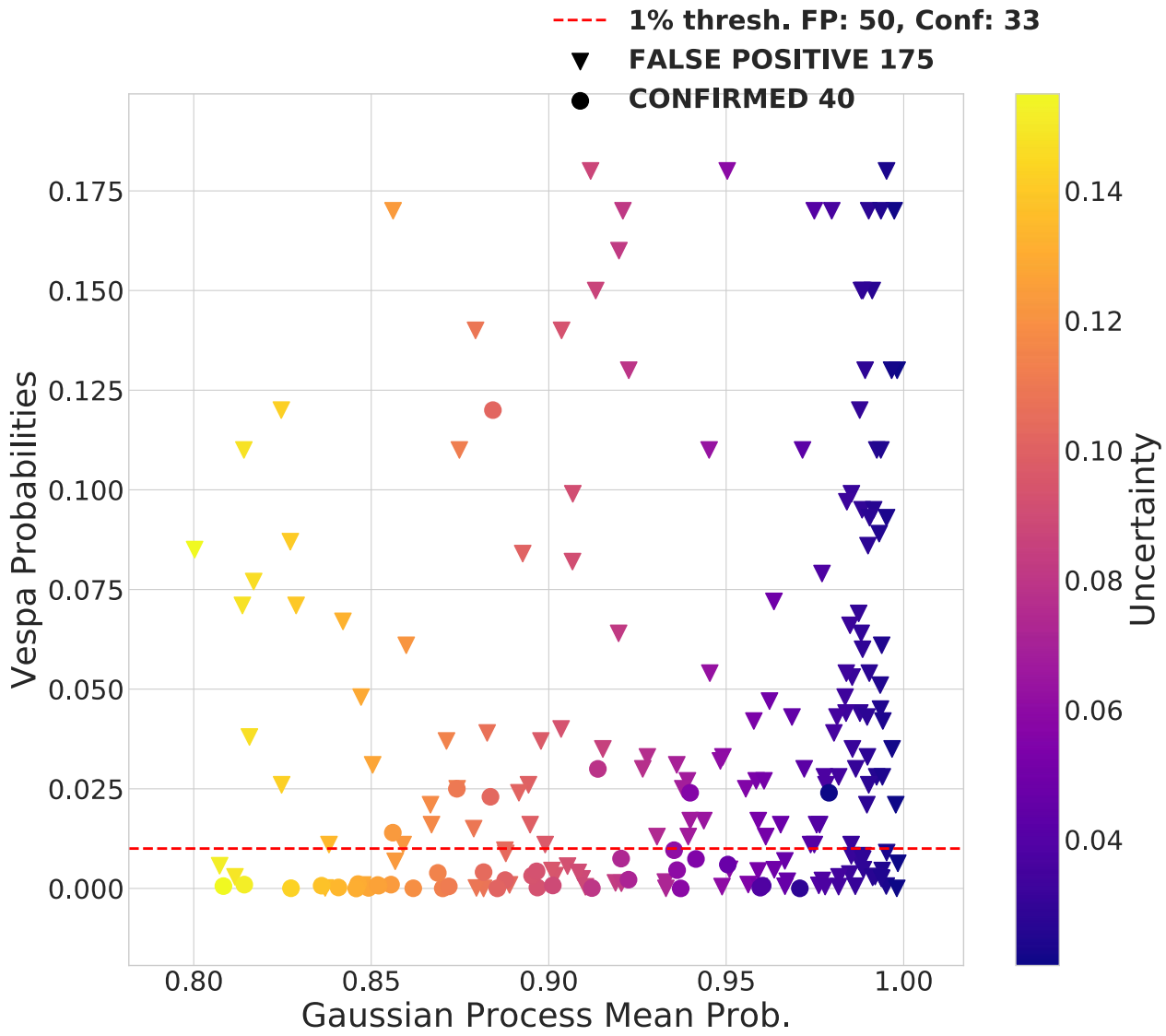
- Outputs based only on input data
- In this run: signal strength, signal shape, secondary eclipse, centroid offset, optical ghosts
- As such, FP designations are likely reliable, plus general ranking

Validation – Posterior Probabilities

- We have the lightcurve shape and vetting data.
- Lack of stellar observations or *a priori* blending likelihood compared to usual methods.
  - Could be added (GAIA?)
- At this point, extremely optimistic, but worth a try


FALSE POSITIVE
CONFIRMED
Vespa Probabilities
Gaussian Process Mean Prob.
Uncertainty
0.0
0.2
0.4
0.6
0.8
1.0
0.05
0.10
0.15
0.20


1% thresh. FP: 0, Conf: 30
FALSE POSITIVE 9
CONFIRMED 42
Vespa Probabilities
Gaussian Process Mean Prob.
Uncertainty
1.00
0.95
0.90
0.85
0.80
0.000
0.025
0.050
0.075
0.100
0.125
0.150
0.175
0.200
0.10
0.08
0.06
0.04
0.02


1% thresh. FP: 50, Conf: 33
FALSE POSITIVE 175
CONFIRMED 40
Vespa Probabilities
Gaussian Process Mean Prob.
Uncertainty
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
0.80
0.85
0.90
0.95
1.00
0.14
0.12
0.10
0.08
0.06
0.04

Summary

- ML methods extremely effective on the surface
- Significant risk of overfitting/overconfidence – can be mitigated
- Already good for ranking, FP exclusion
- Perhaps eventually for full planet validation?
  - As a complementary, independent and fast method

Development Possibilities

- Use simulated training sets
  - Large numbers possible
  - We control the inputs
- Feed in knowledge of target star
- Simulations need extreme care – to cover all FP scenarios with realistic input distributions
  - Time consuming to set up – very fast once done

Outliers

- Remove outliers from training set
- No large performance drop, but less overfitting