Implications of Completeness in Transit Surveys

Jessie Christiansen

NASA Exoplanet Science Institute/Caltech

jessie.christiansen@caltech.edu

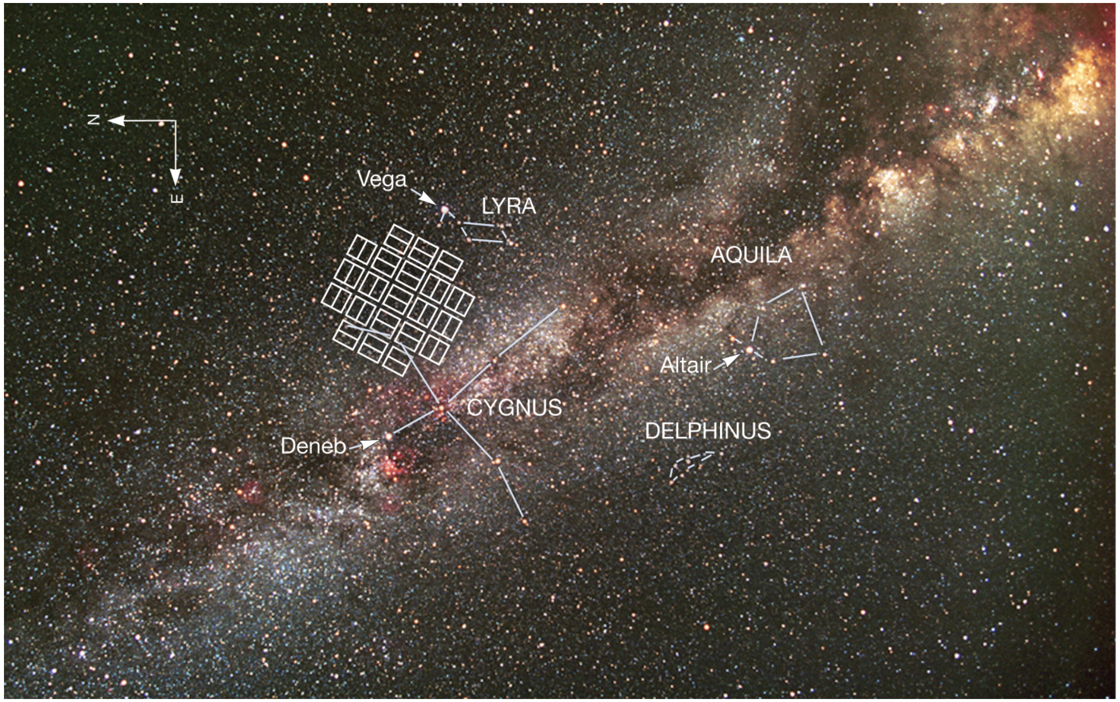
*available to chat over lunch today

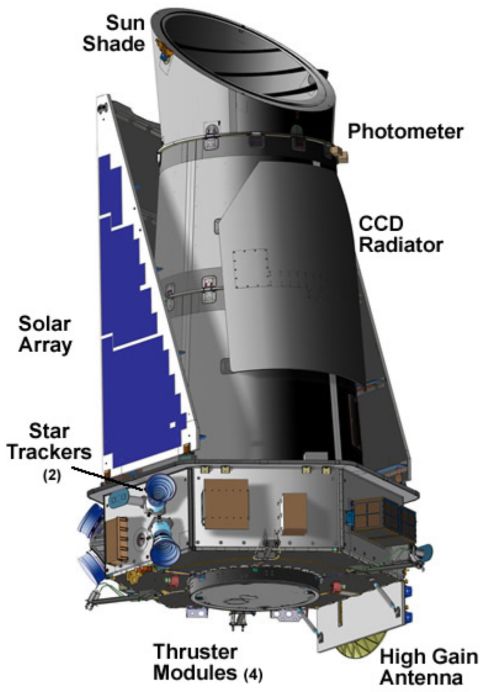
Overview

- Calculating planet occurrence rates (e.g. η_{Earth})
- An experiment to measure the *Kepler* pipeline completeness
- Implications for derived occurrence rates
- Other sources of systematic errors in occurrence rates

Designing a mission to measure η_{Earth}

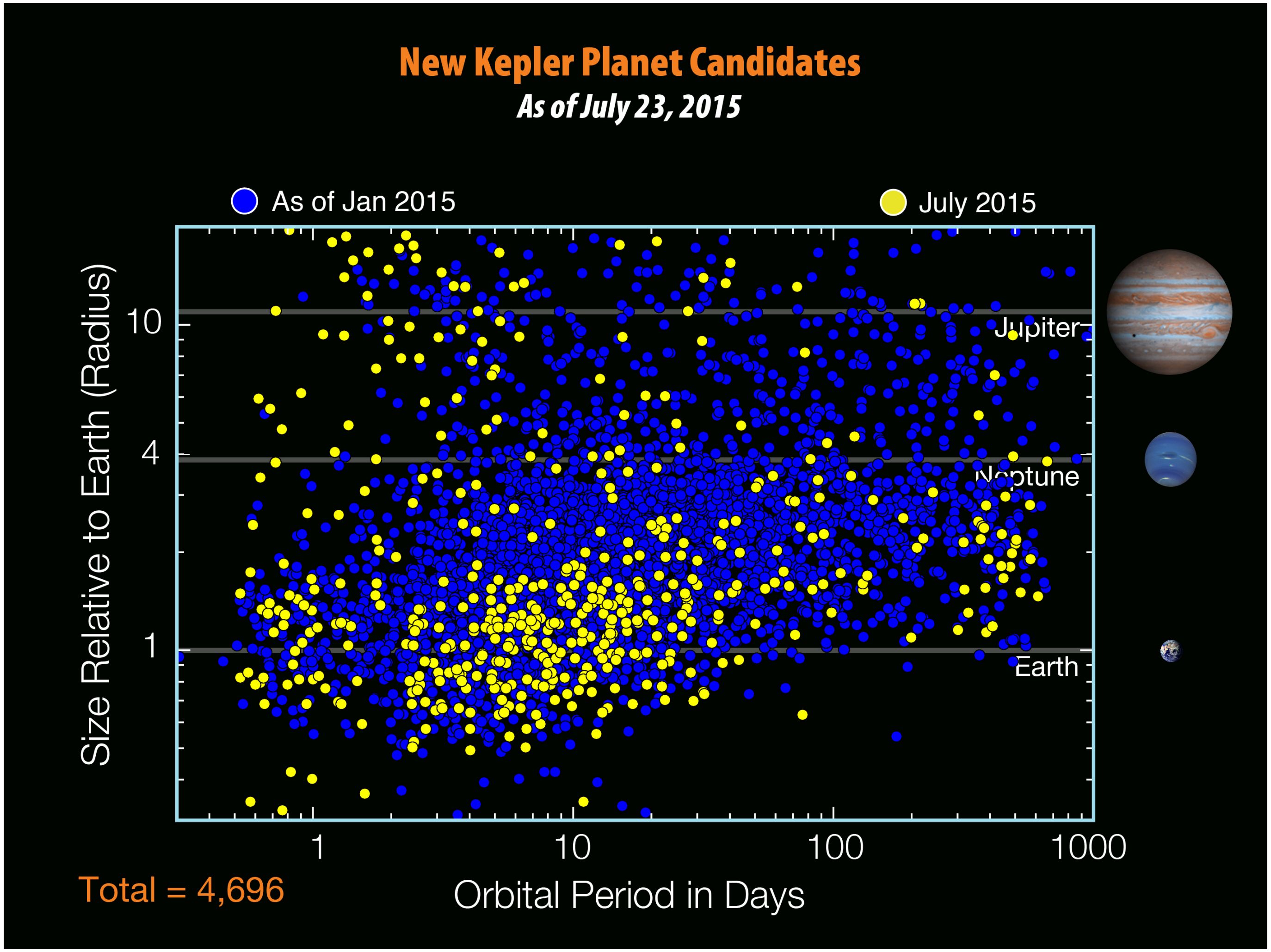
- Kepler was a transit survey optimized for finding terrestrial planets (0.5 to 10 Earth masses) in the habitable zone (out to 1 AU) of stars like the Sun
- Continuously, simultaneously monitored nearly 200,000 stars, 1m Schmidt telescope, 30min integrations, field-of-view of >100 sq deg with 42 CCDs
- Light curves, lists of candidates and pipeline products available at NASA Exoplanet Archive: everyone can play!





The story so far…

- Last week, hundreds of new candidates announced
- Four years of data, 4696 planet candidates,
- NB. Version ***9.2*** of the Kepler pipeline – important!!



Determining η_{Earth}

We need to calculate both:

$N_{measured}$: the number of real Earth-like planets in the Kepler sample (i.e. understanding the reliability, or false positive rate)

$N_{detectable}$: the number of stars around which the Kepler pipeline would have detected such planets (i.e. understanding the completeness)

The aim of my research has been to characterise $P_{i,SNR}$ for the Kepler pipeline, which we can then use to calculate the pipeline detection efficiency.

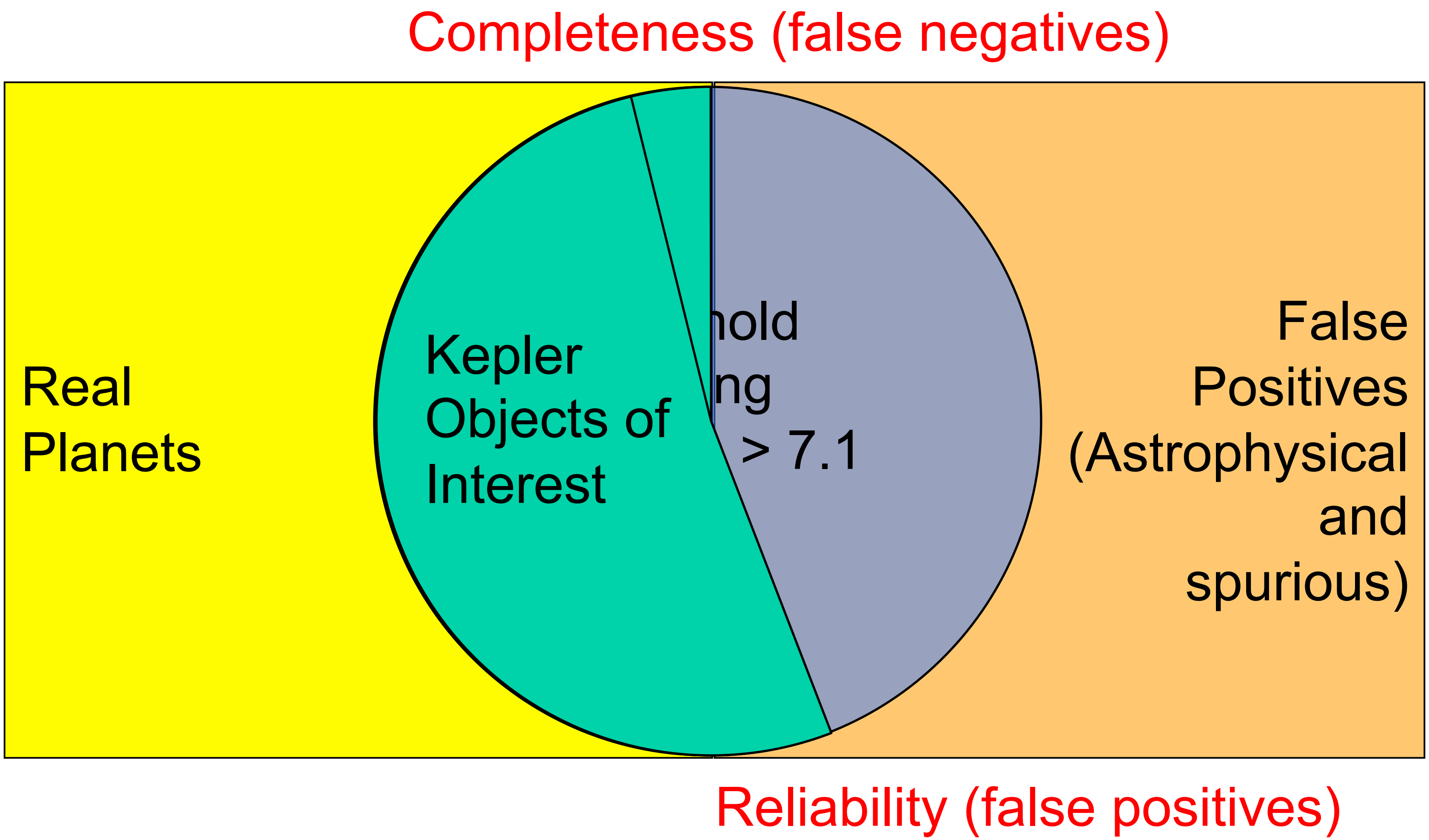
$$\eta_{EARTH} = \frac{N_{measured}}{N_{detectable}}$$, where

$$N_{detectable} = \sum_i P_{i,geo} P_{i,SNR}$$

Geometric probability of ith planet to transit

Probability of ith planet to having strength SNR being detected

Put another way…



Historical estimates of $P_{i,SNR}$

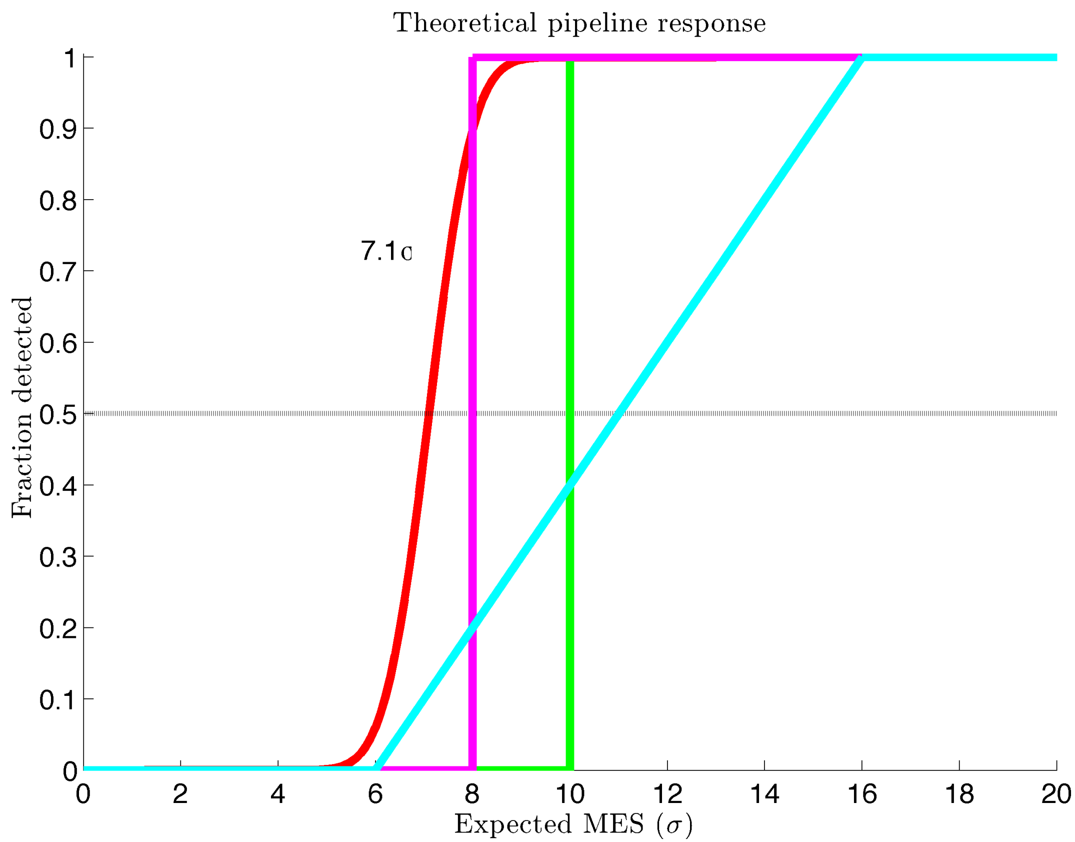
In a perfect world:

- Pipeline threshold = 7.1σ (Catanzarite & Shao 2011, Borucki et al 2011, Traub 2012, Dressing & Charbonneau 2013, Kopparapu 2013)

In a conservative world:

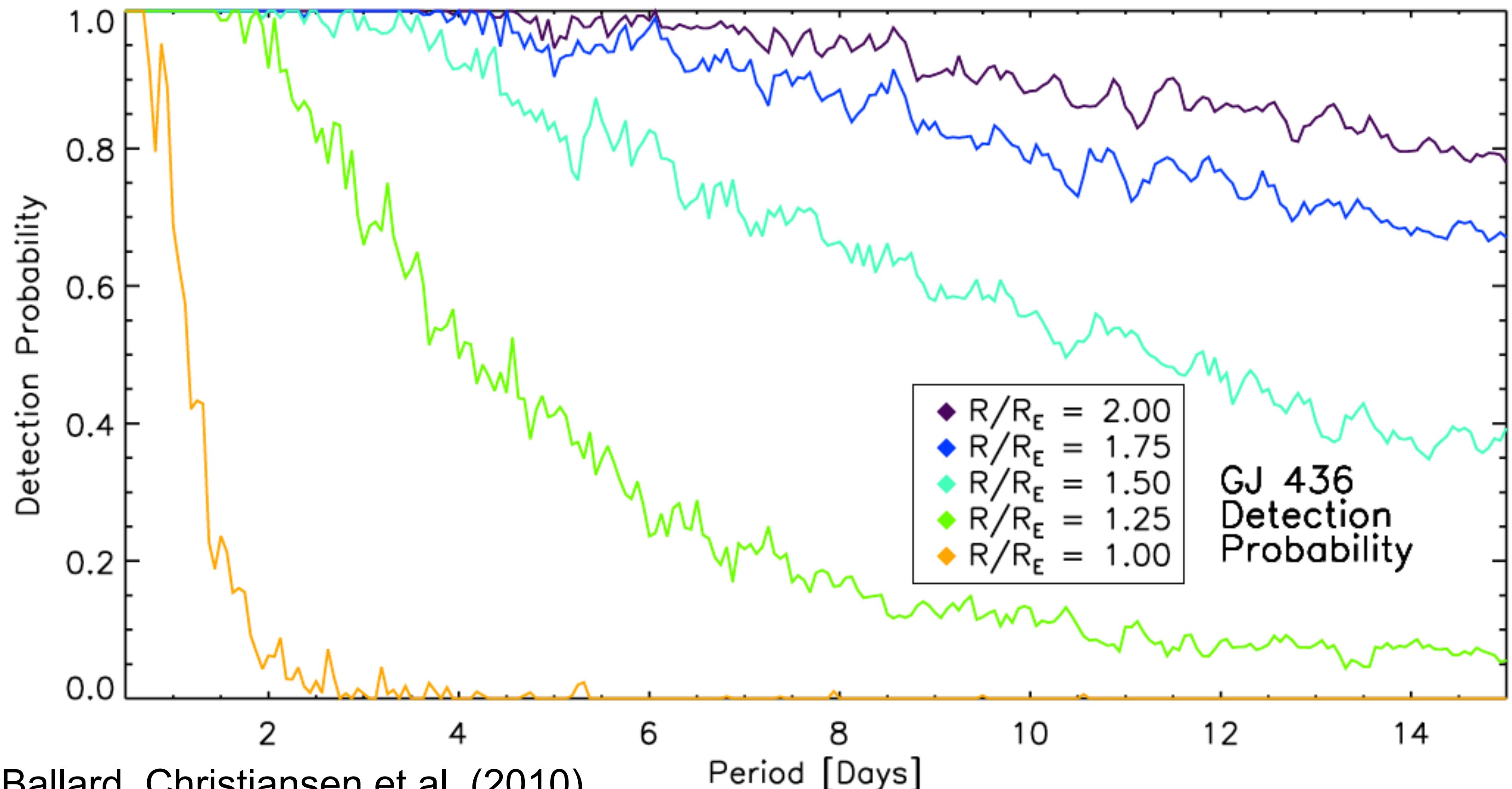
- Howard et al (2011, >2Re,<50d), Youdin (2011, >0.5Re,<50d)
- Dong & Zhu (2012, < 250d)
- Fressin et al (2013), Mulders et al (2014),

These teams report significant variation in derived occurrence rates, but examine different parameters spaces… hard to disentangle the impact of the decision re: detectability



Measuring $P_{i,SNR}$

- To characterize recoverability of signals, ideally we would perform a Monte Carlo analysis – for each star, inject a suite of fake transit signals into the pixels and find the limits of detectability
-
-



Ballard, Christiansen et al. (2010)

The 2nd transit injection experiment

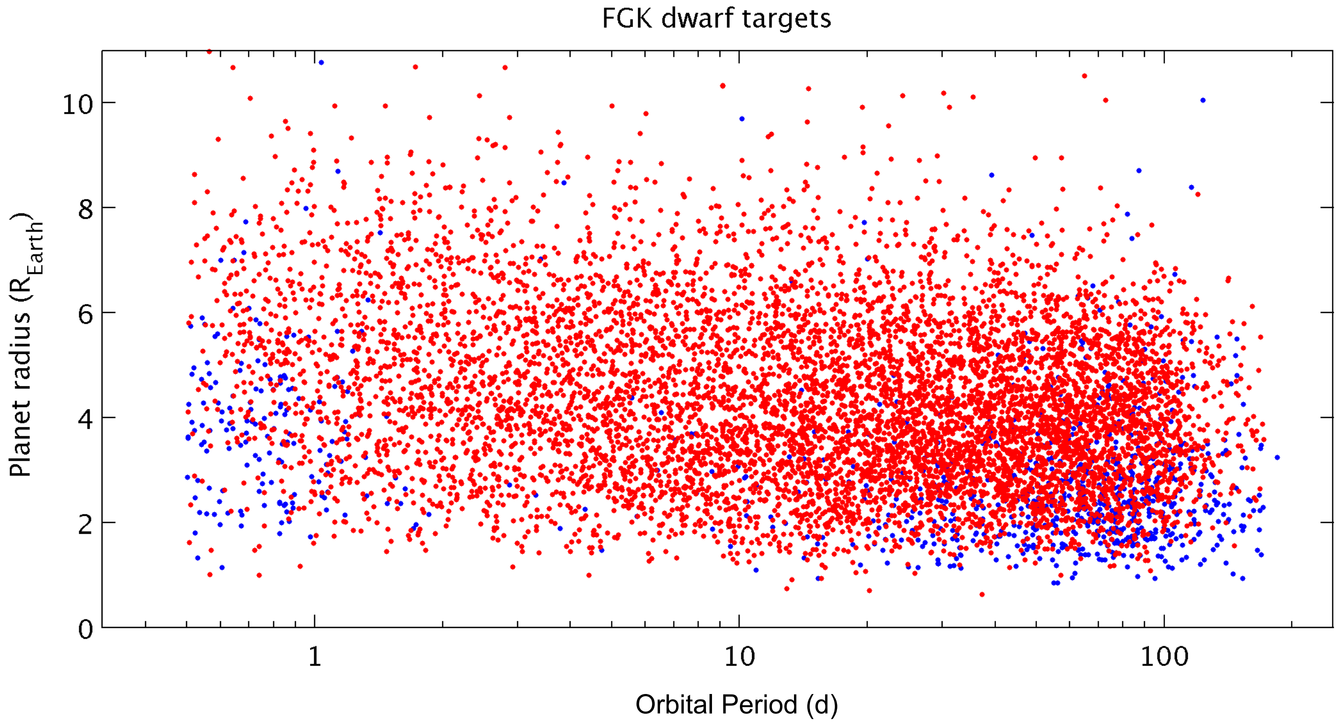
(Christiansen et al. 2015)

- Four quarters (one year) of data
- 15 CCDs (out of 84)
- Every target (10,341 targets)
- Into the pixels themselves (not the flux time series)
- Version ***9.1*** of the Kepler pipeline
- So, applicable to the Mullally et al (2015) catalogue, NOT to the Coughlin et al (2015) catalogue (***updates to the pipeline between 9.1 and 9.2, specifically in the way in which potential signals are evaluated, significantly change the detectability of long period, low SNR signals.)

Sky group	Channel	Description
32	4	Edge of field/worst focus
70	10	Variable black/bias correction
71	11	Edge of field/worst focus
9	13	FGS crosstalk
25	17	Nominal/best focus
48	24	Nominal/best focus
66	26	Rolling band artefacts/Moiré pattern drift
84	32	Edge of field/worst focus
62	46	Variable black/bias correction
78	50	FGS crosstalk
4	56	Edge of field/worst focus
18	58	Rolling band artefacts/Moiré pattern drift
19	59	Nominal/best focus
38	62	Rolling band artefacts/Moiré pattern drift
74	70	FGS crosstalk
53	81	Start-of-line ringing

Injected versus recoveries

0.5-11Re

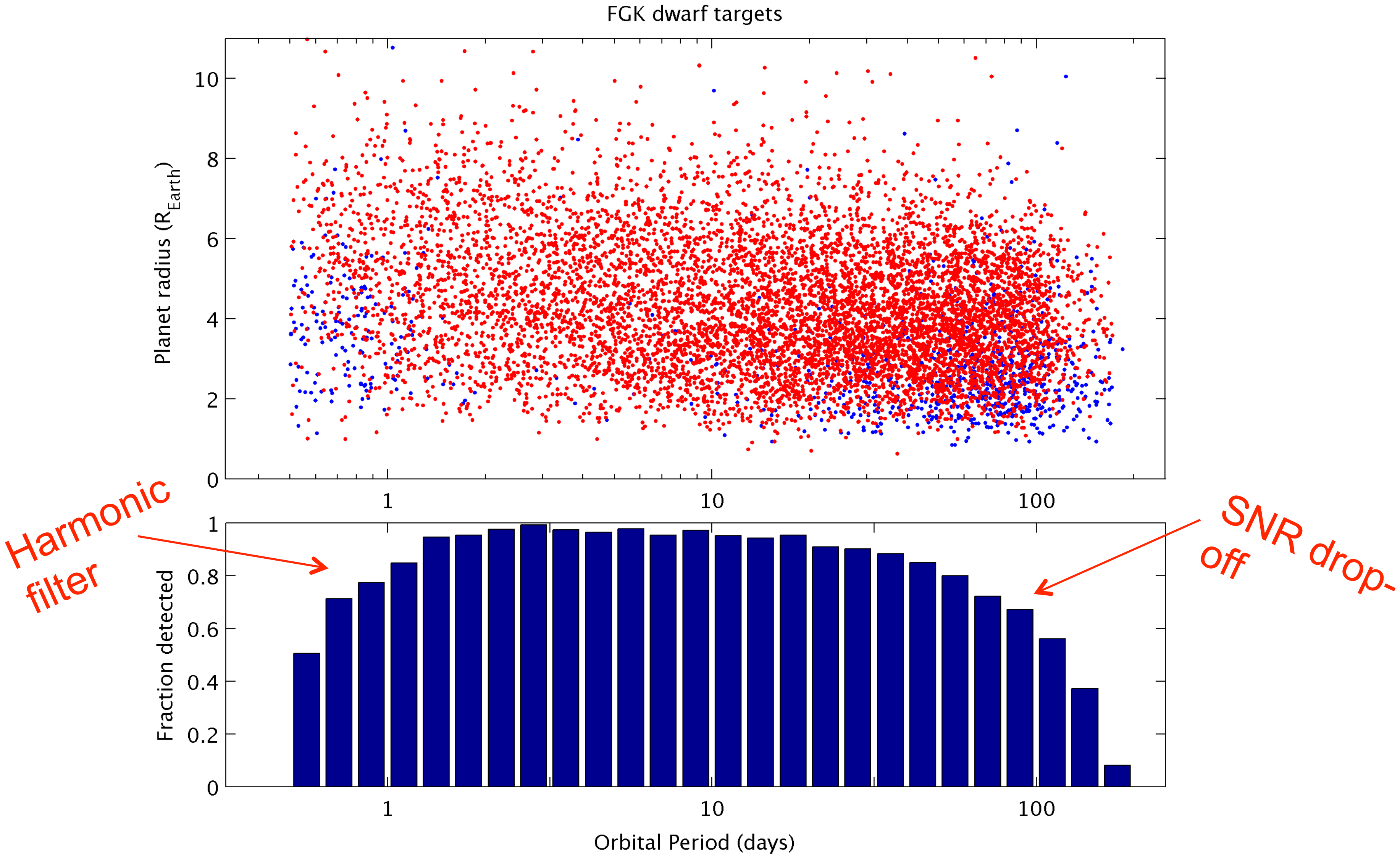


10,341 injections

9,123 recoveries

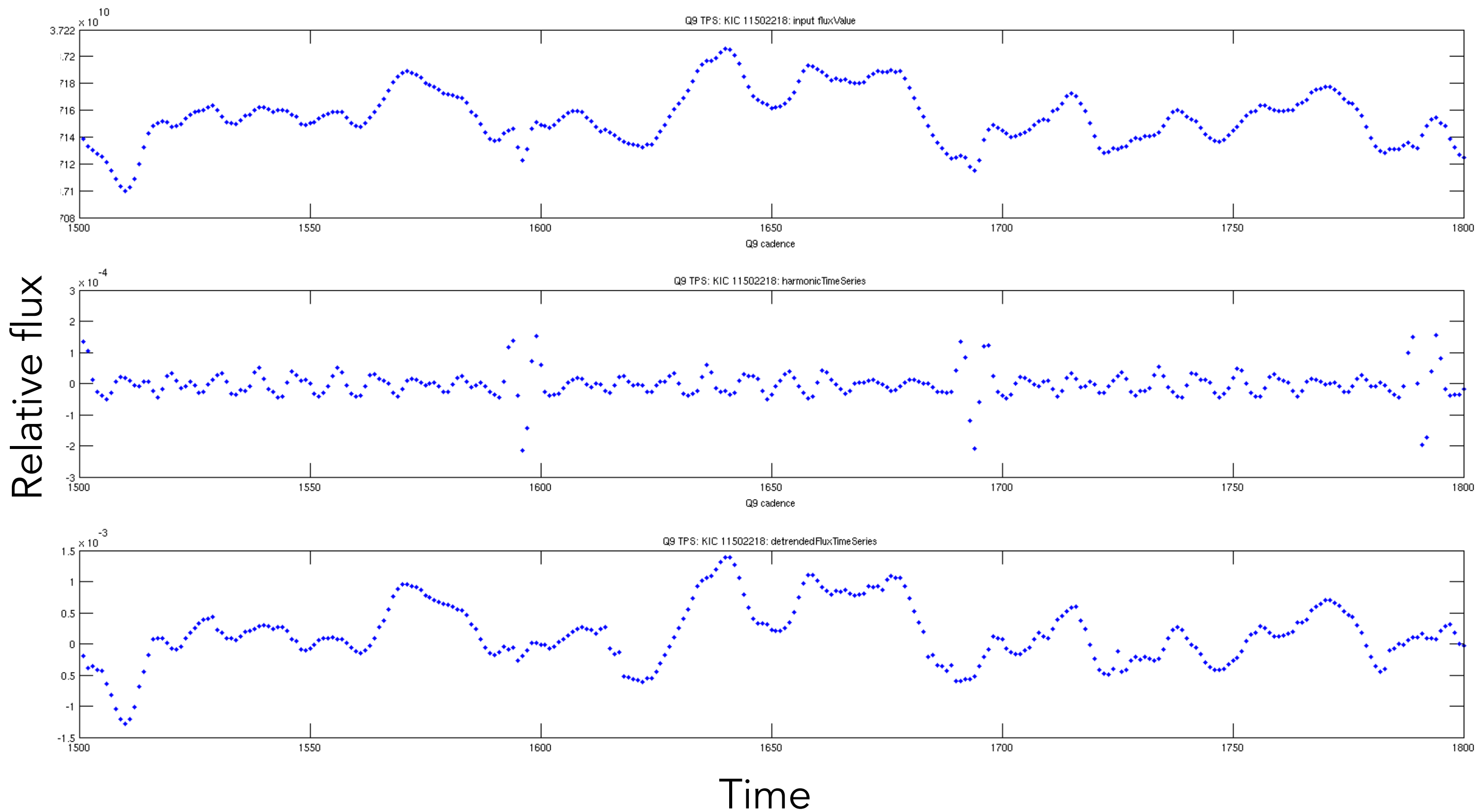
0.5-200d

Injected versus recoveries

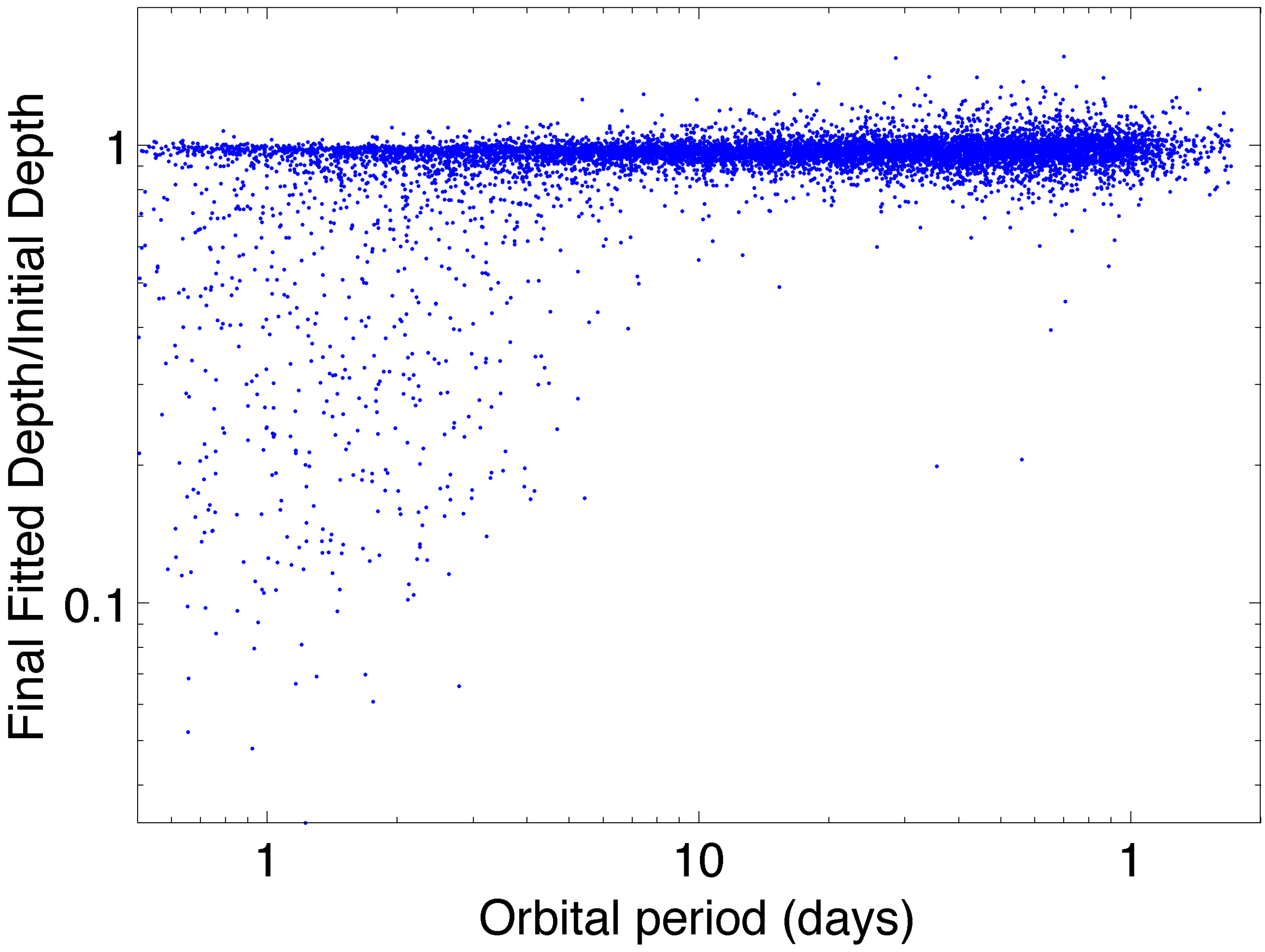


The harmonic filter removes short period signals

- Before whitening and folding, TPS fits out harmonics, to enable planet searches around active stars
- For transiting or eclipsing light curves with periods < 3 days, and especially < 1 day, the transits are modeled as a Fourier series and removed – important implications for completeness!

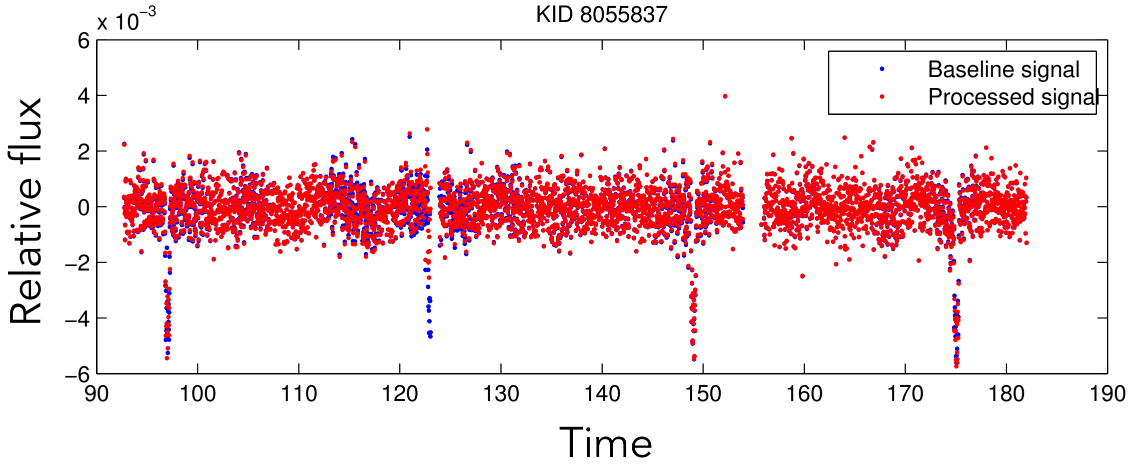


The harmonic filter also reduces the depths of recovered signals



Transits near data gaps are distorted

- We perform additional systematic correction near data gaps, where the most significant systematics (largely due to thermal changes) occur, which typically distorts transits within 2 days of data gaps



Calculated expected SNR

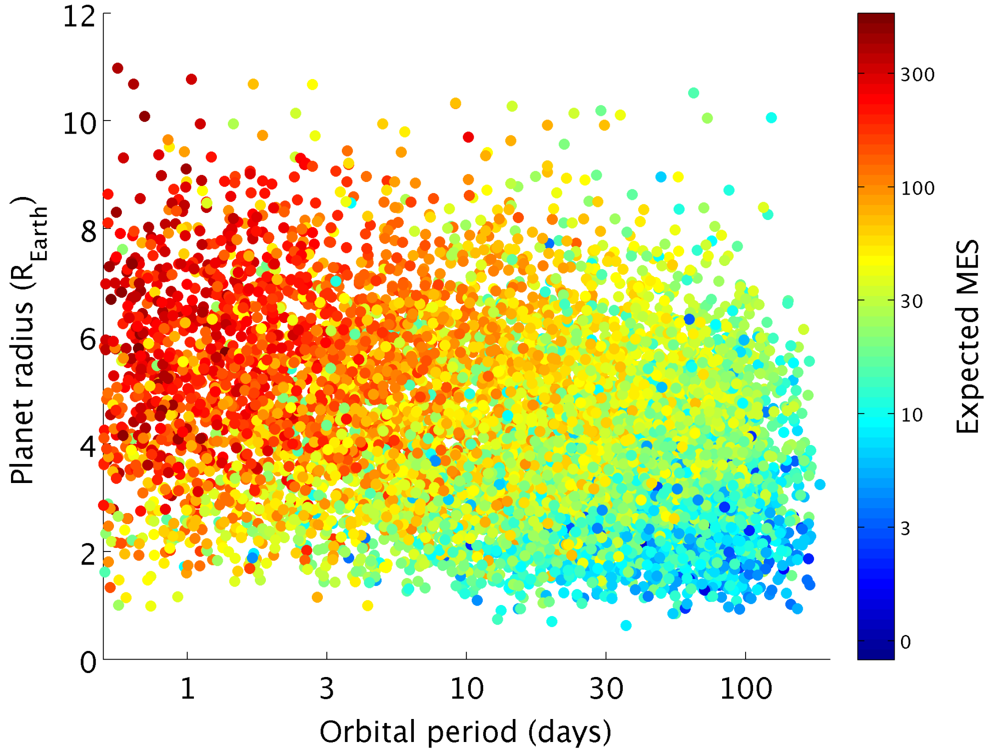
- MES = Multiple Event Statistic
  (signal to noise of the whitened transit model compared to the whitened light curve)
- 'Ground truth' MES:

  (Eq 4. Christiansen et al. 2012)

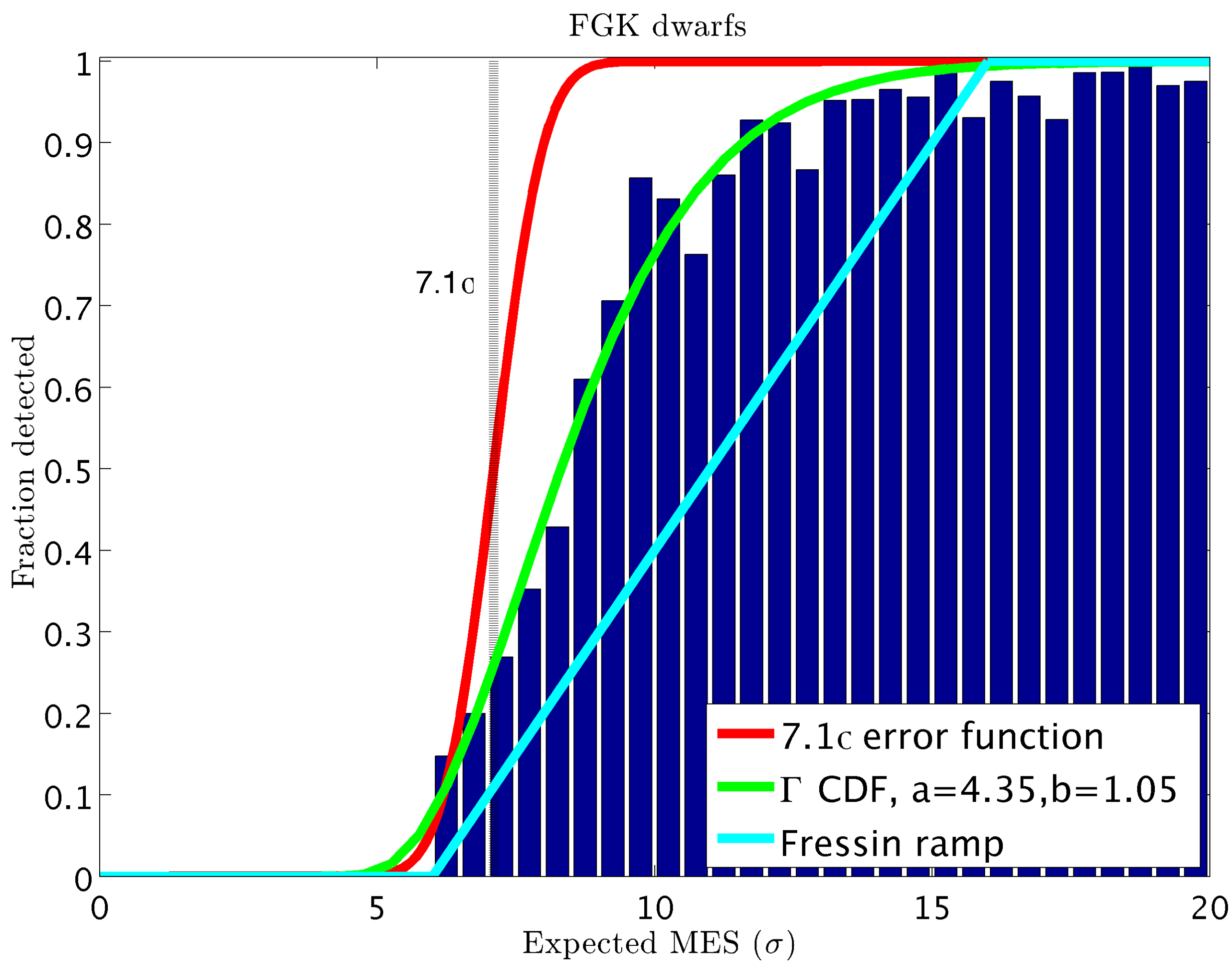
$$MES = \sqrt{\frac{t_{obs} * f_{obs}}{P}} * \frac{\delta}{CDPP_{eff}}$$

$$MES = \sqrt{N_t} * SNR_t$$

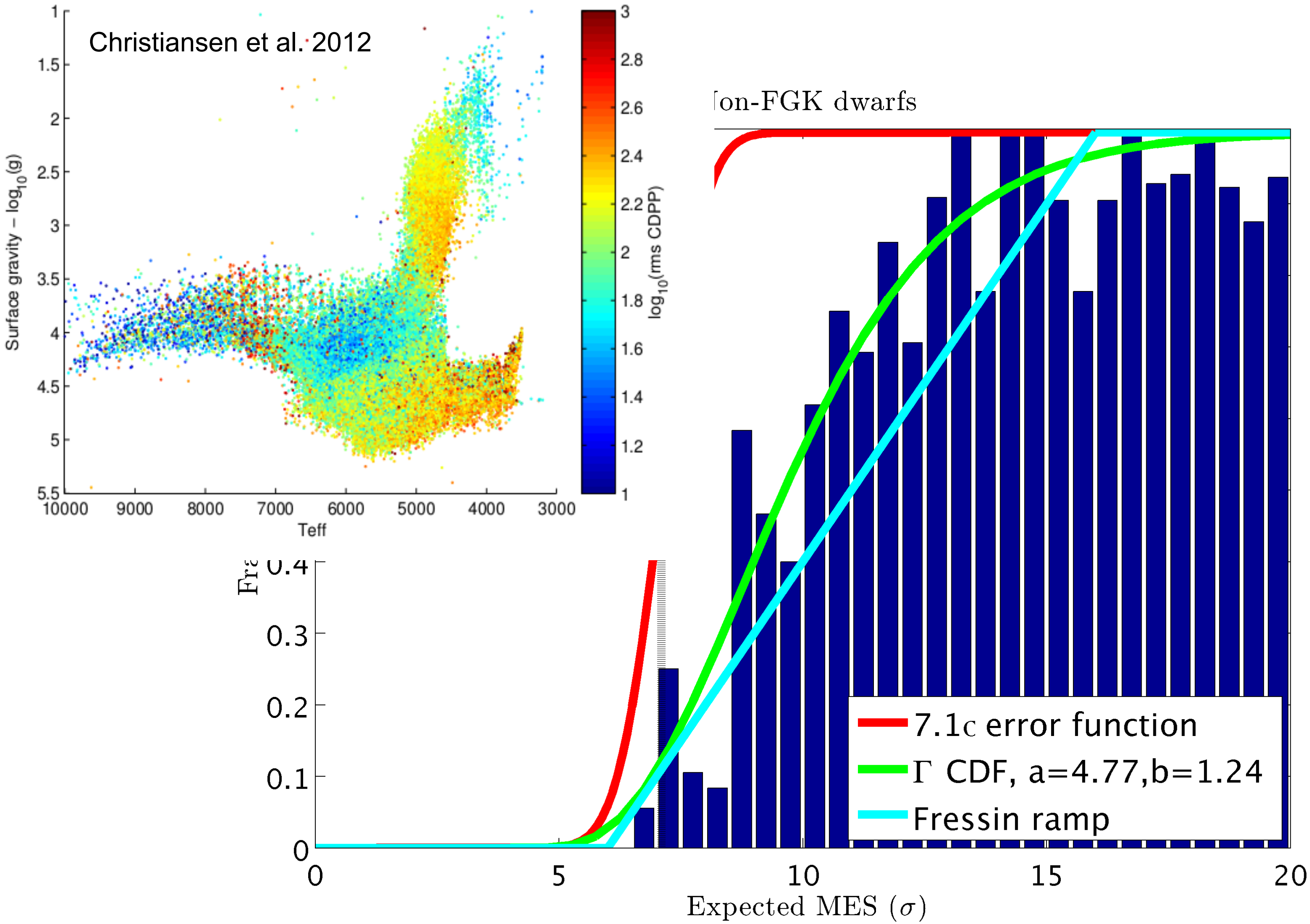
- For the expected MES, need to also include
  - dilution (crowding)
  - additional gapping/de-weighting of observations
  - shape of model template



Sensitivity function



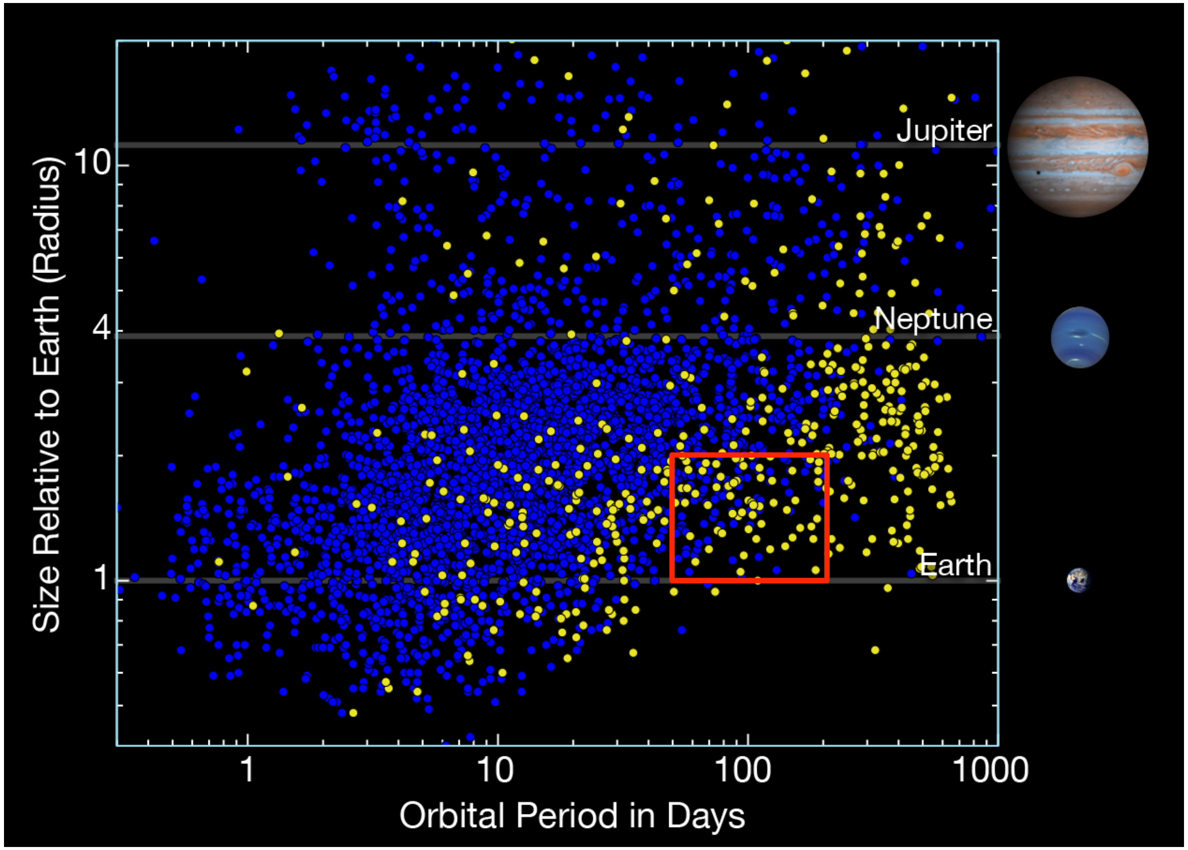
Sensitivity function



Giants and M-dwarfs – masking of real signals by correlated noise in the light curves

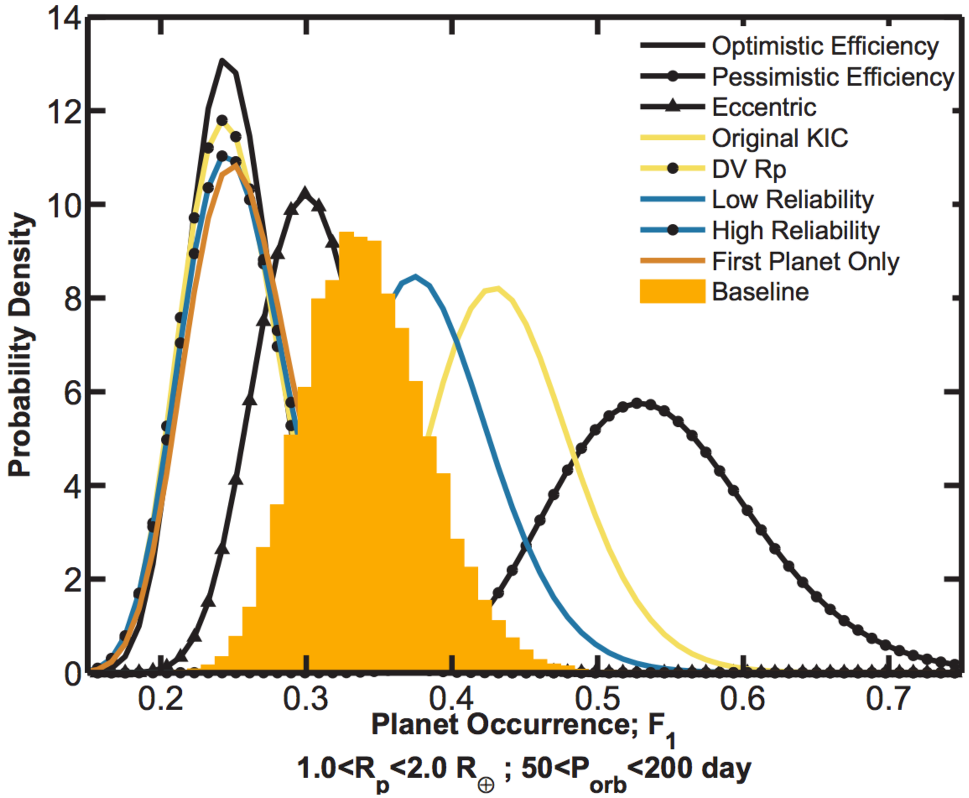
What does this mean for occurrence rate calculations?

- Using the method described by Youdin 2011; Burke, Christiansen et al. (2015) – parametric occurrence rate (best fit = broken power law in radius and power law in period)
- 50-200 days, 1-2Re planets, using Q1-Q16 catalogue (Mullally et al. 2015):

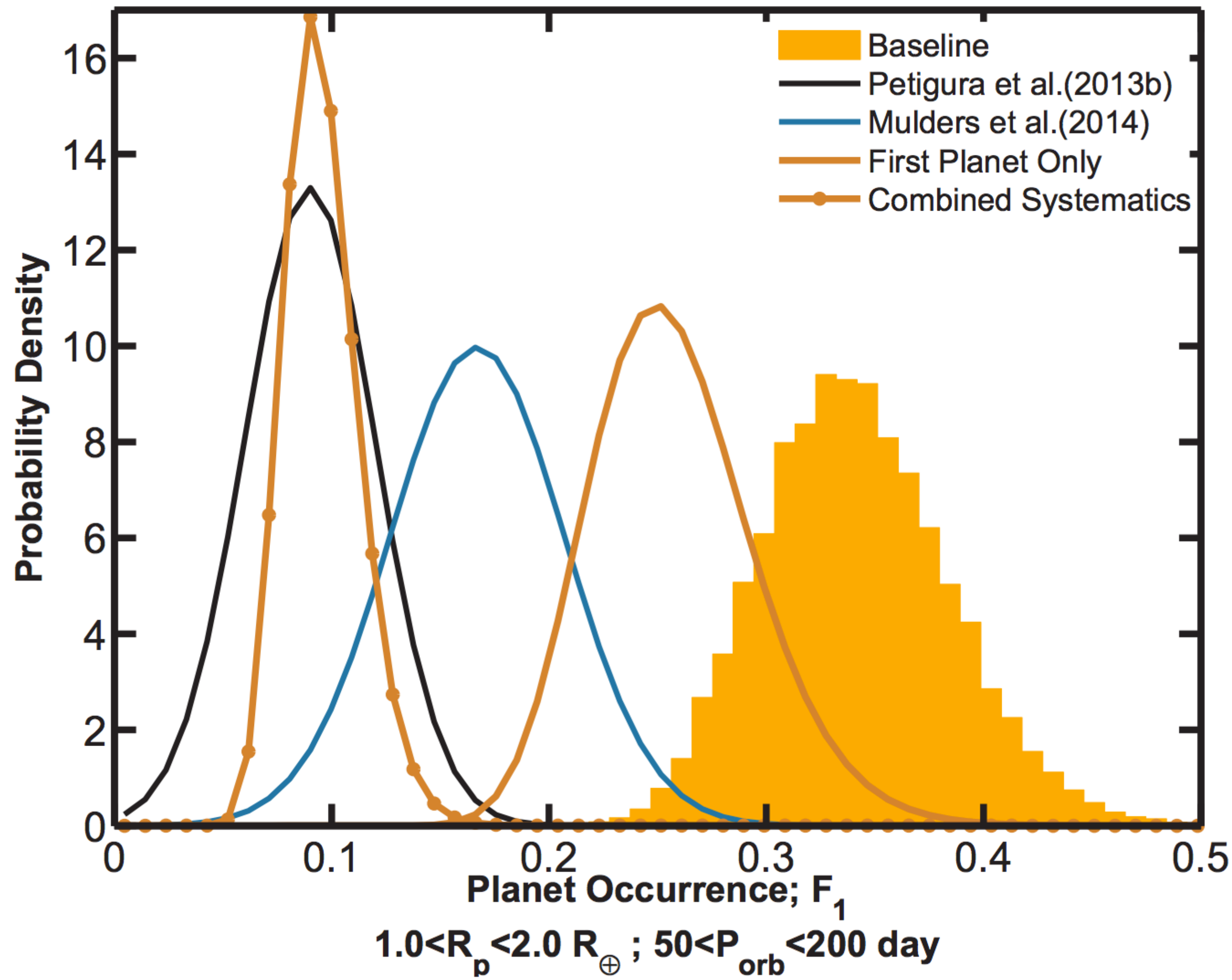


What does this mean for occurrence rate calculations?

- Using the method described by Youdin 2011; Burke, Christiansen et al. (sub) – parametric occurrence rate (best fit = broken power law in radius and power law in period)
- 50-200 days, 1-2Re planets, using Q1-Q16 catalogue (Mullally et al. 2015):



Comparison to other calculations



This is not the rug you are looking for

- ***Completeness doesn't end with the pipeline!***
  - Lists of Threshold Crossing Events (potential planet candidates – full lists available at the Exoplanet Archive for your perusal!) are passed on for further analysis
  - Up to and including the Mullally et al. 2015 catalogue, this vetting included human intervention
  - How much coffee have YOU had today??
  - From the Coughlin et al. 2015 catalogue onwards, there are two complementary autonomous vetting algorithms
    - A machine-learning algorithm thatlearns from a training set which light curves look like planets and which look like false positives
    - A(n) heuristic-based method which reproduces the decisions and cuts made by the human vetters
  - Both of these are being challenged by the latest, greatest transit injection run.

What to expect when you're expecting occurrence rate products

- ***Four-year, full-focal-plane transit injection run currently being analysed!***
- Pipeline version 9.2 – the same as the recent Q1-Q17 Coughlin et al. 2015 planet candidate catalogue that I showed earlier
- 159,000 targets across the focal plane
- 130,000 with transits injected at the target, 29,000 'false positives' with spatial offsets
- 0.5-500 days, 0.25-7Rearth
- Full table of injections and recovery status coming to an Exoplanet Archive near you!