Planet Detection: The *Kepler* Mission

Jon M. Jenkins
SETI Institute/NASA Ames Research Center

Friday February 25, 2011

GISS CESS 2011, NYC

SAO

STScI

Outline

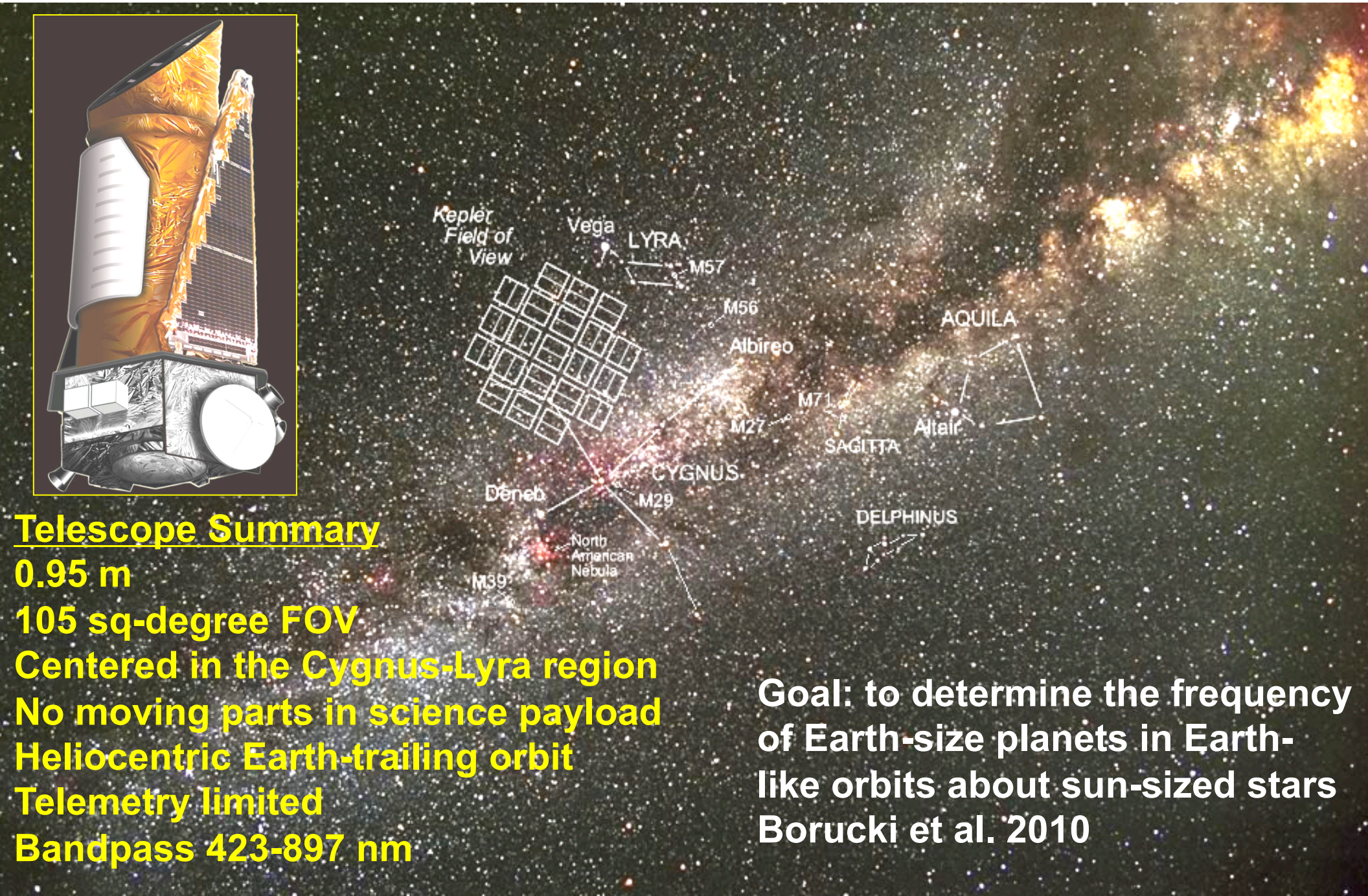
Kepler
A Search for Earth-size Planets

- Overview of the *Kepler* Mission and the Science Pipeline
- Transiting Planet Search – An Adaptive, Wavelet-Based Matched Filter
- Pre-Search Data Conditioning: What Went Wrong and How We Can Fix It
- Conclusions

NASA's first mission capable of finding Earth-size and smaller planets

Mission Overview

Kepler
A Search for Earth-size Planets



Telescope Summary

0.95 m
105 sq-degree FOV
Centered in the Cygnus-Lyra region
No moving parts in science payload
Heliocentric Earth-trailing orbit
Telemetry limited
Bandpass 423-897 nm

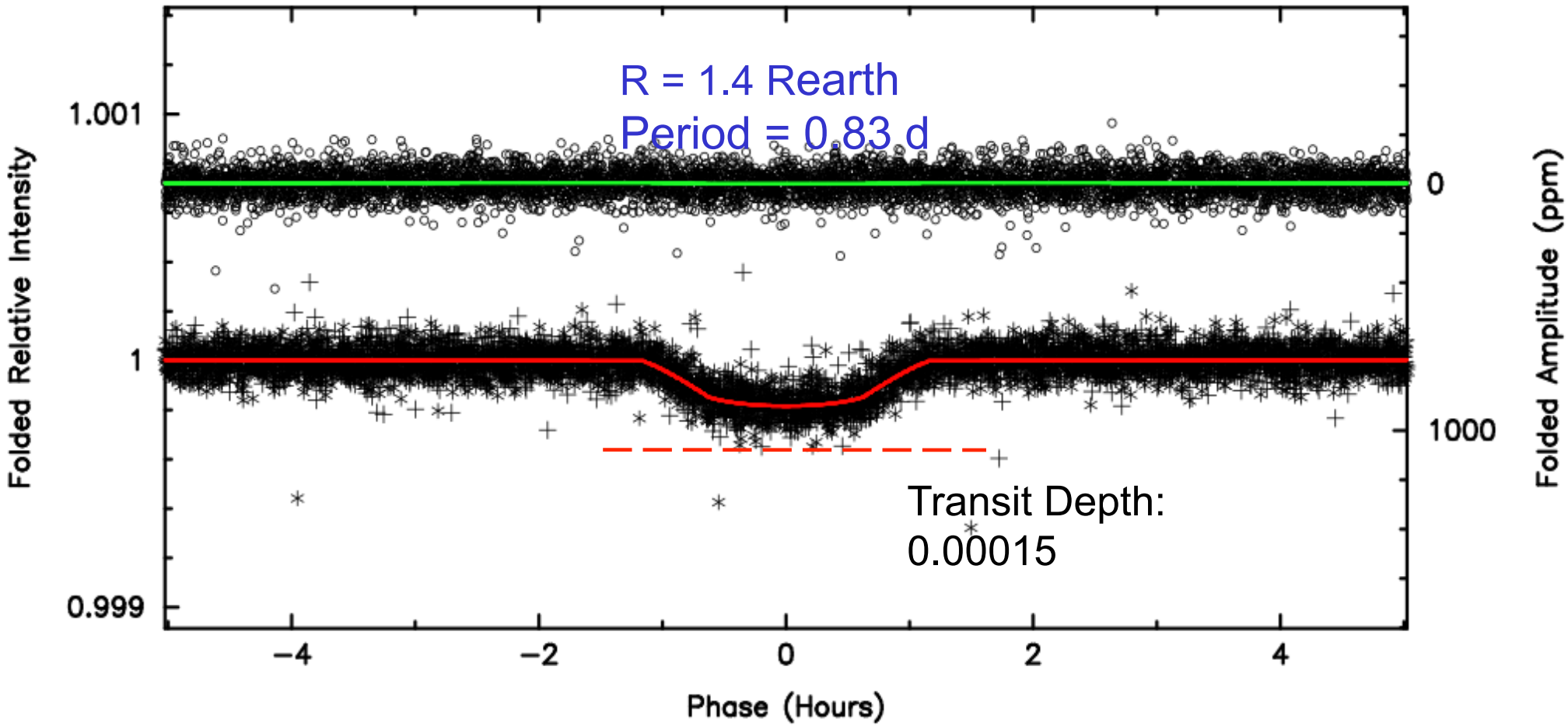
Goal: to determine the frequency of Earth-size planets in Earth-like orbits about sun-sized stars
Borucki et al. 2010

Recent *Kepler* News

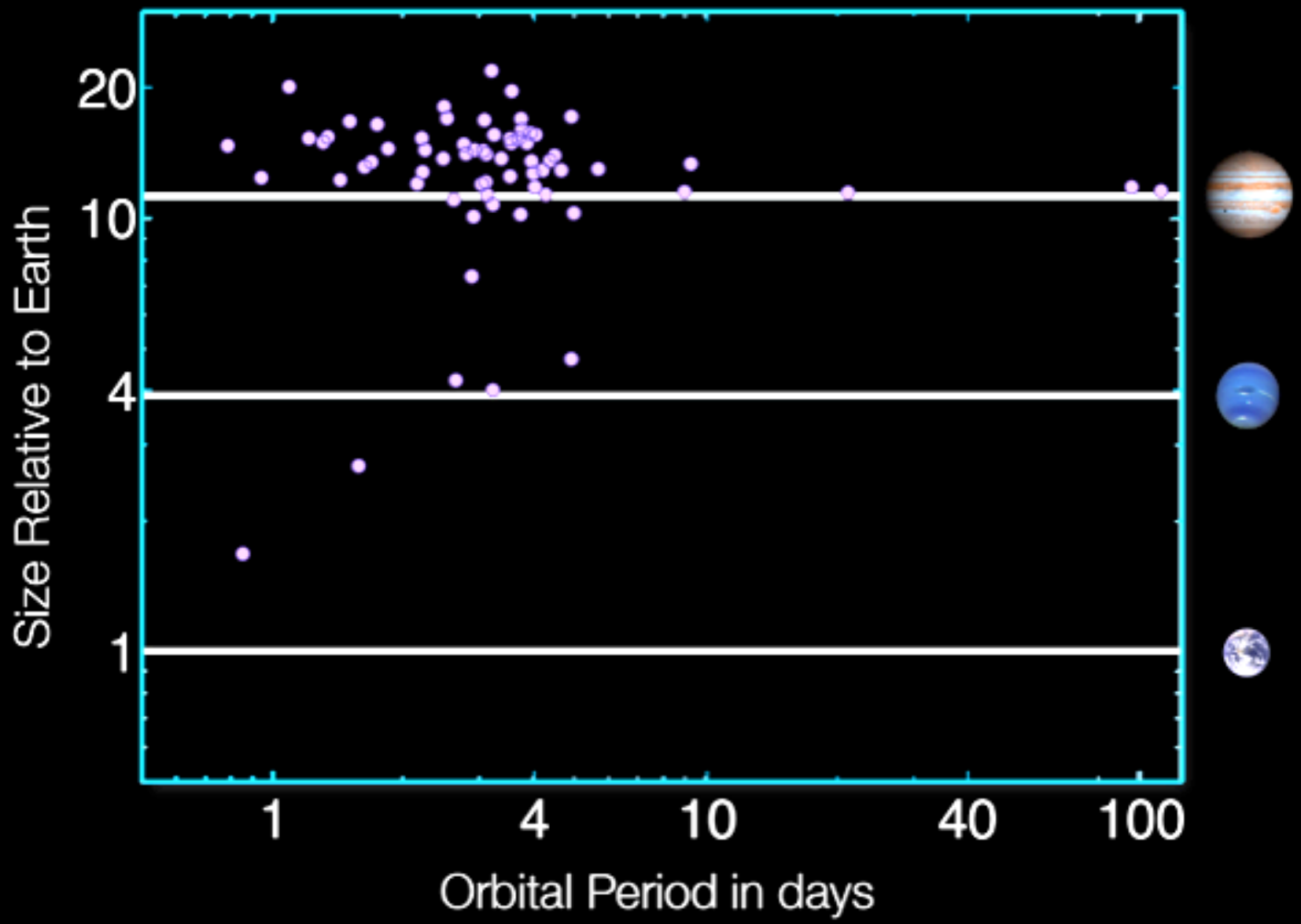
Kepler
A Search for Earth-size Planets

- First confirmed rocky planet announced: Kepler-10b
- 1235 New Planet Candidates Detected
- Most are nearly Earth-size: 2 - 5x Earth-Size
- One star with 6 transiting planets: Kepler-11

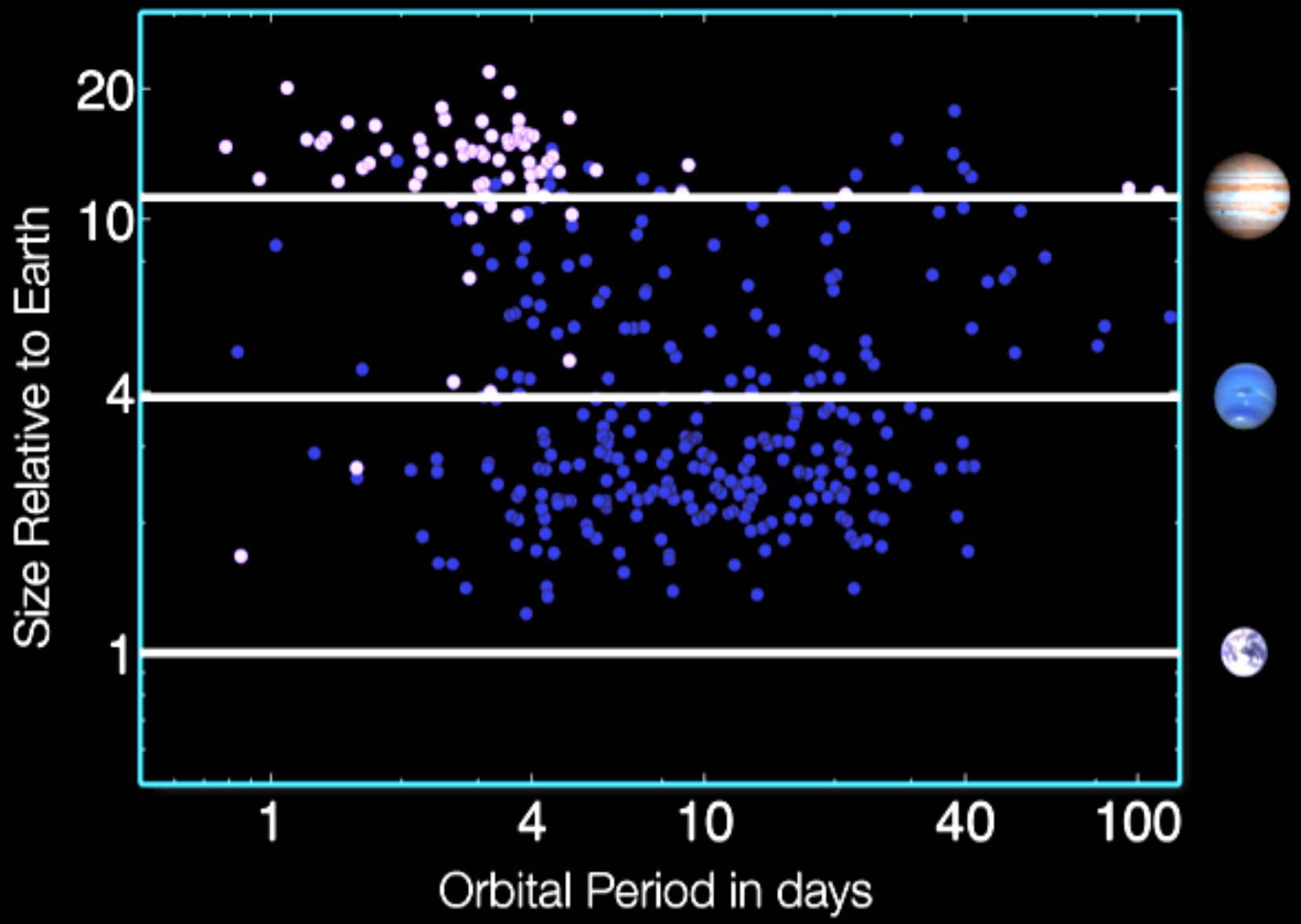
Kepler-10


R = 1.4 Rearth
Period = 0.83 d
Transit Depth:
0.00015
Folded Relative Intensity
1.001
1
0.999
Folded Amplitude (ppm)
0
1000
−4
−2
0
2
4
Phase (Hours)

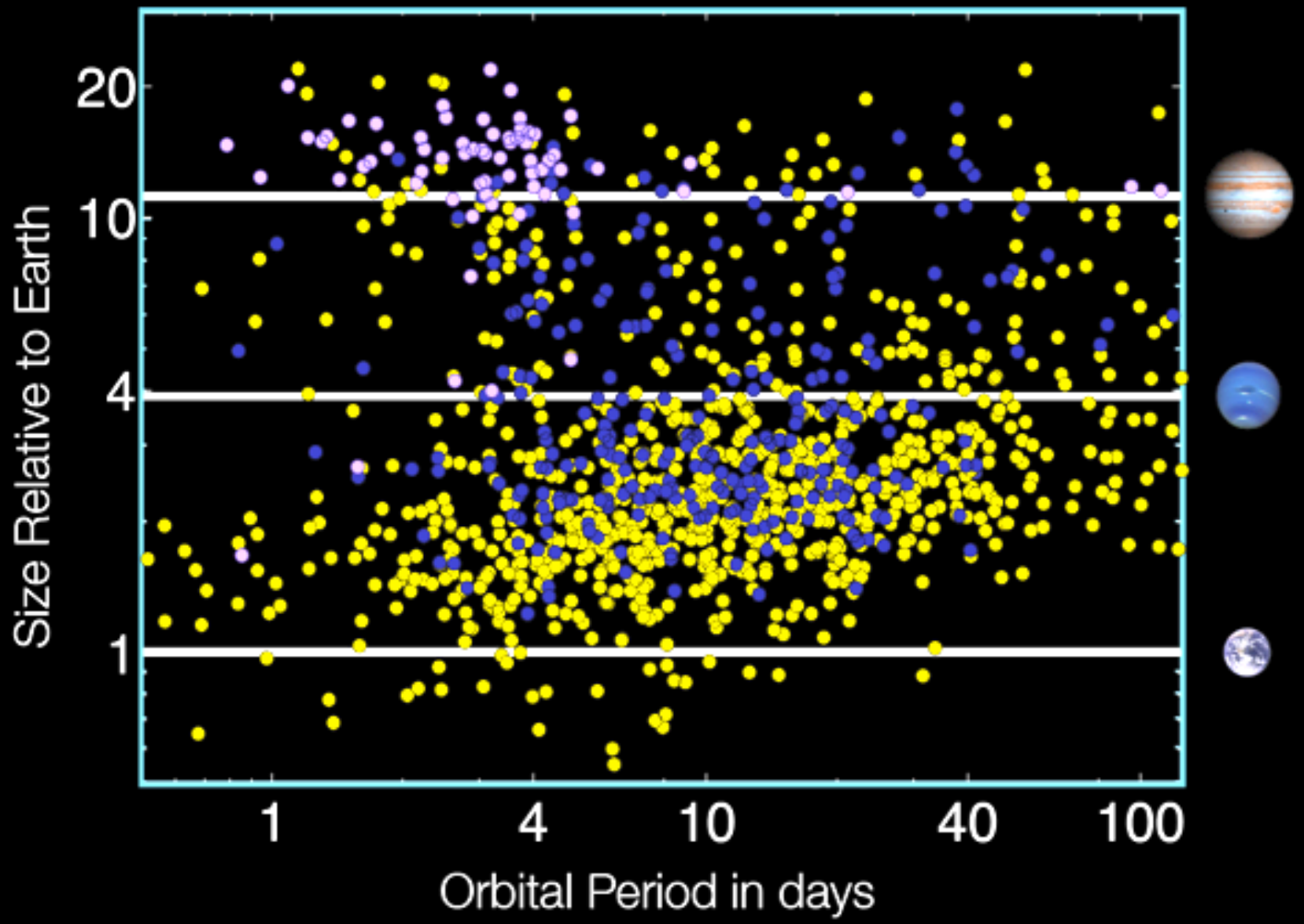
Pre-Kepler Transiting Planets - 2009



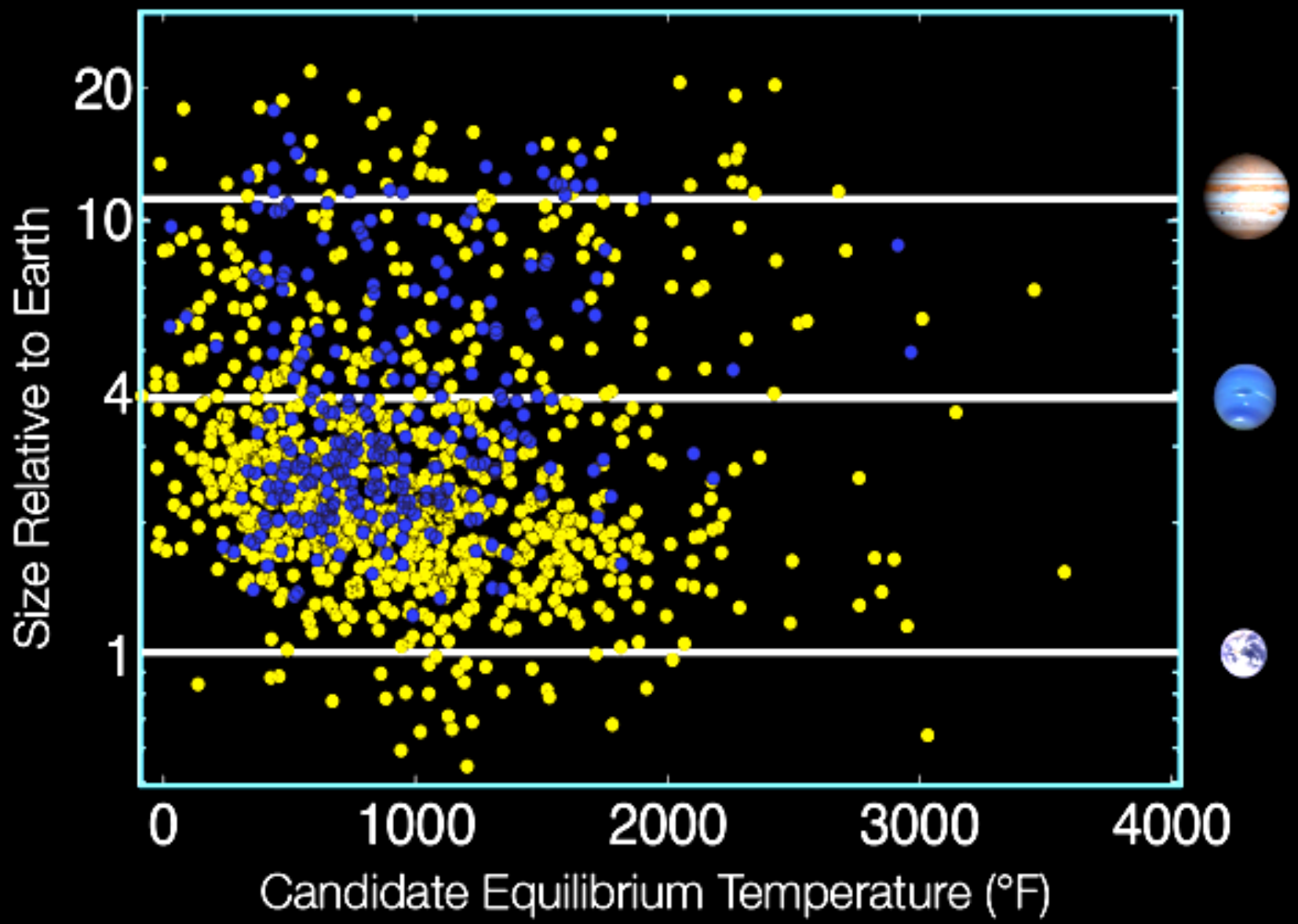
Kepler Candidates as of June 2010



Kepler Candidates as of February 1, 2011

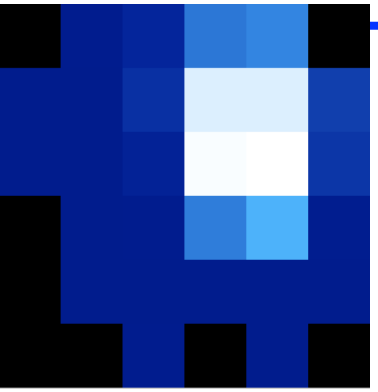


Kepler Candidates as of February 1, 2011



The Kepler Science Pipeline: From Pixels To Planets

Kepler
A Search for Earth-size Planets



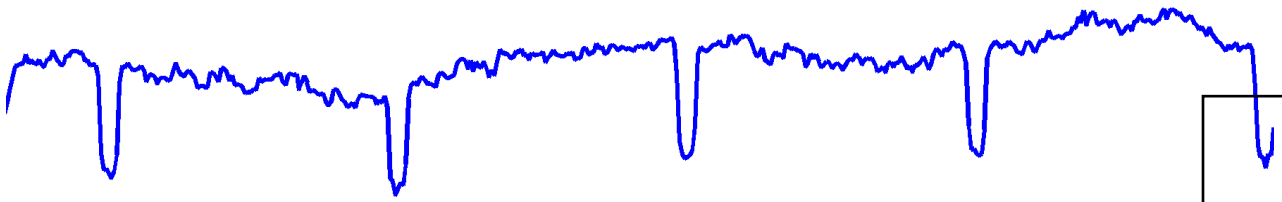
Raw Data

CAL
Pixel Level Calibrations

Calibrated Pixels

PA
Photometric Analysis

Sums Pixels Together/ Measures Star Locations



Raw Light Curves/ Centroids

PDC
Presearch Data Conditioning

Removes Systematic Errors

Corrected Light Curves

TPS
Transiting Planet Search

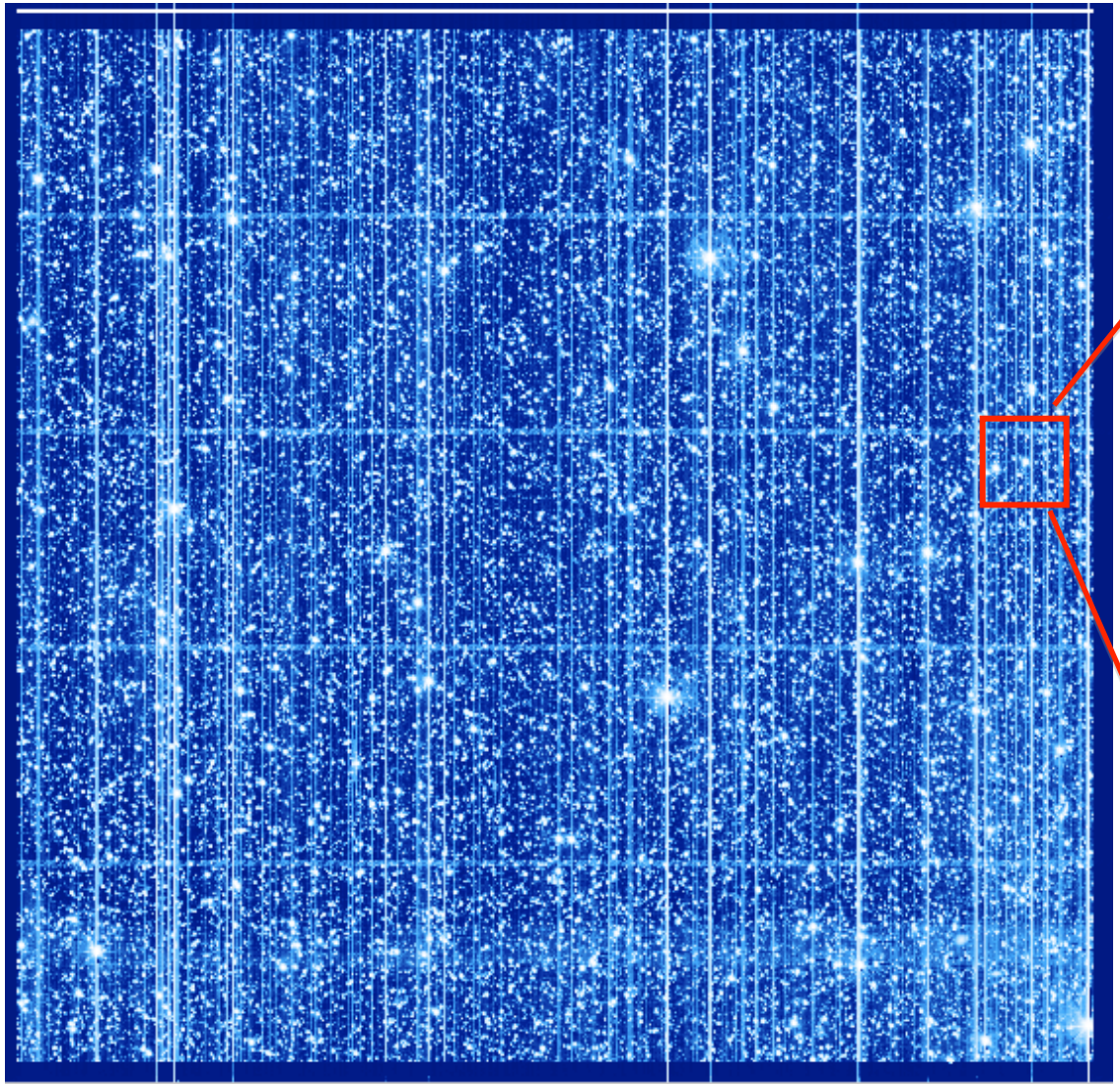
TCEs:
Threshold Crossing Events

DV
Data Validation

Diagnostic Metrics

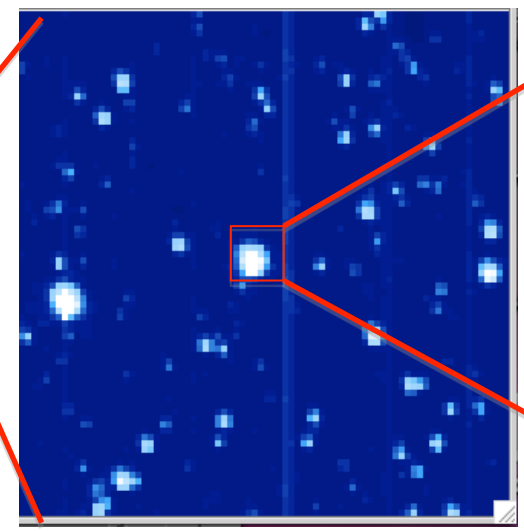
Pixel-Level Data From *Kepler*

Module 17 Output 2



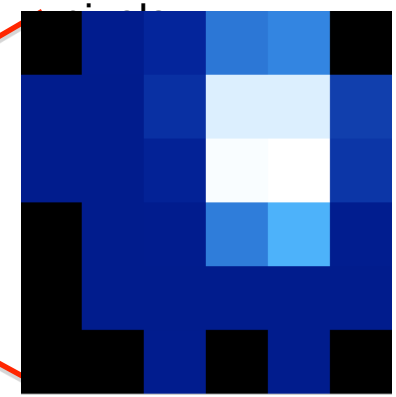
1.13 (h) x1.22 (w) degrees

Zoomed image near planet HAT-P-7b



0.09x0.09 degrees
80x80 pixels
6400 pixels total

HAT-P-7b

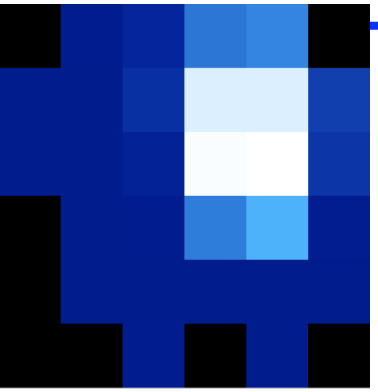


6.6x6.6 millidegrees
28 pixels collected
Black = no data

Scaled to show faint detail

The Kepler Science Pipeline: From Pixels To Planets

Kepler
A Search for Earth-size Planets



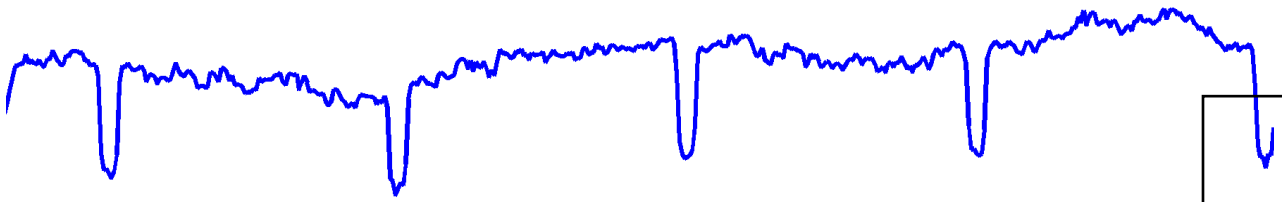
Raw Data

CAL
Pixel Level Calibrations

Calibrated Pixels

PA
Photometric Analysis

Sums Pixels Together/ Measures Star Locations



Raw Light Curves/ Centroids

PDC
Presearch Data Conditioning

Removes Systematic Errors

Corrected Light Curves

TPS
Transiting Planet Search

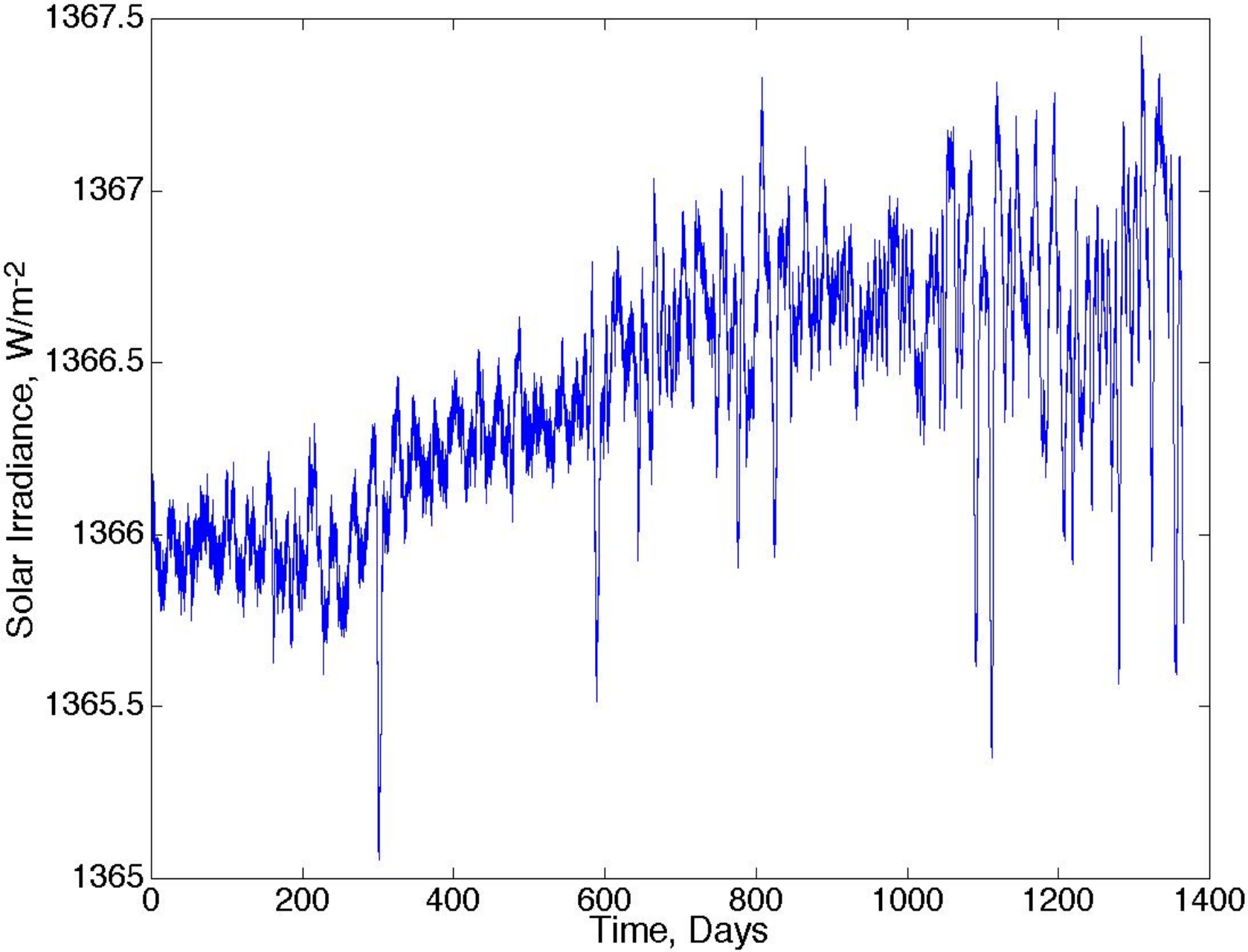
TCEs:
Threshold Crossing Events

DV
Data Validation

Diagnostic Metrics

Solar Variability

A Search for Earth-size Planets



Detection Statistics

Kepler
A Search for Earth-size Planets

w **is (colored) Gaussian noise with autocorrelation matrix** ***R***
x **is the data**
s **is the signal of interest**

Decide ***s*** **is present if**

$$T = \frac{x^T R^{-1} s}{\sqrt{s^T R^{-1} s}} > \gamma$$

How do we determine R?

If the noise is stationary, we can work in the frequency domain:

$$T = \int \frac{X(f) S^*(f)}{P(f)} df \Bigg/ \sqrt{\int \frac{S(f) S^*(f)}{P(f)} df}$$

PSDs for Solar-Like Variability

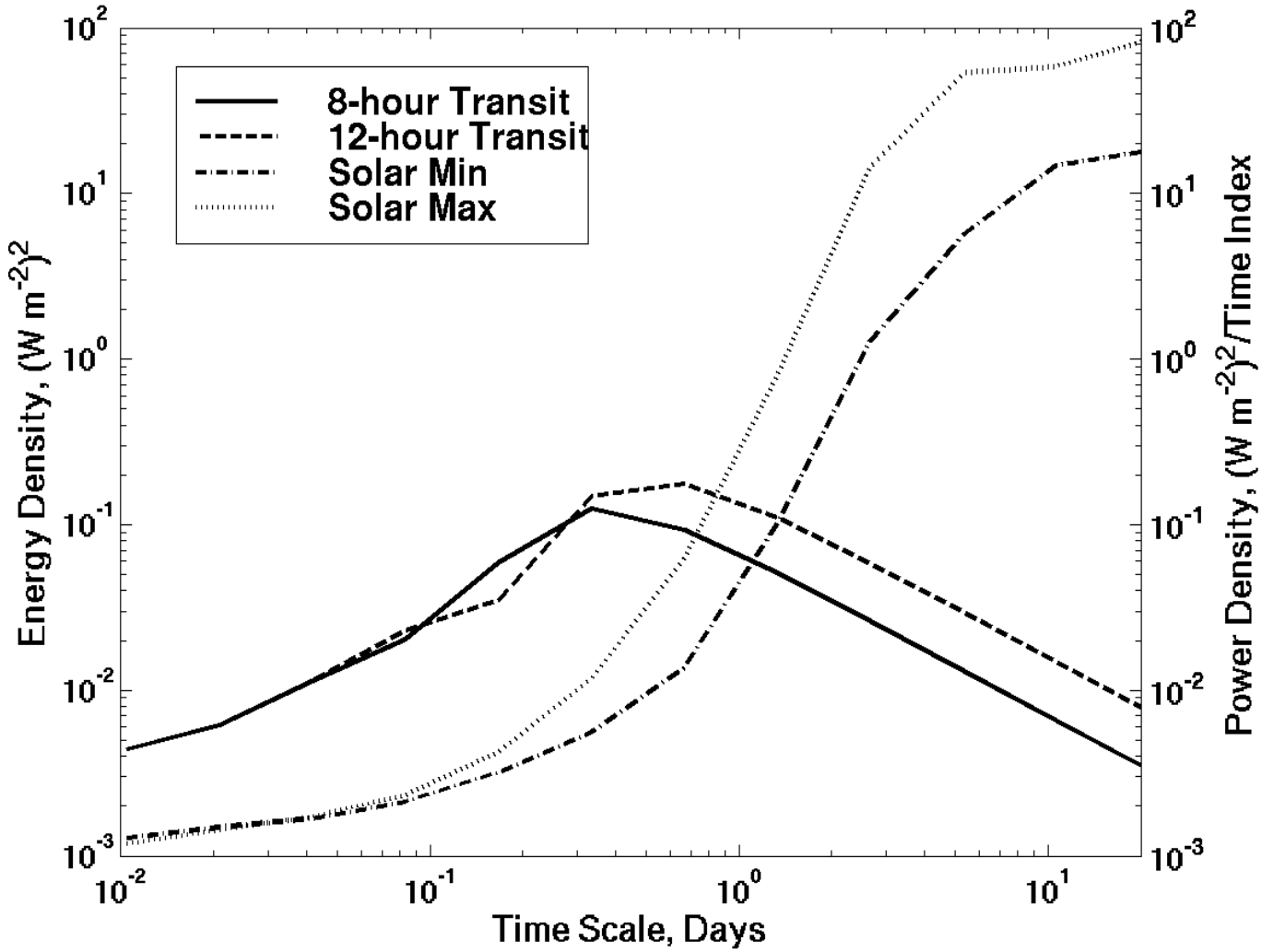
Kepler
A Search for Earth-size Planets

Is stellar variability stationary?

No!

We must work in a joint time-frequency domain

Wavelets are a natural choice



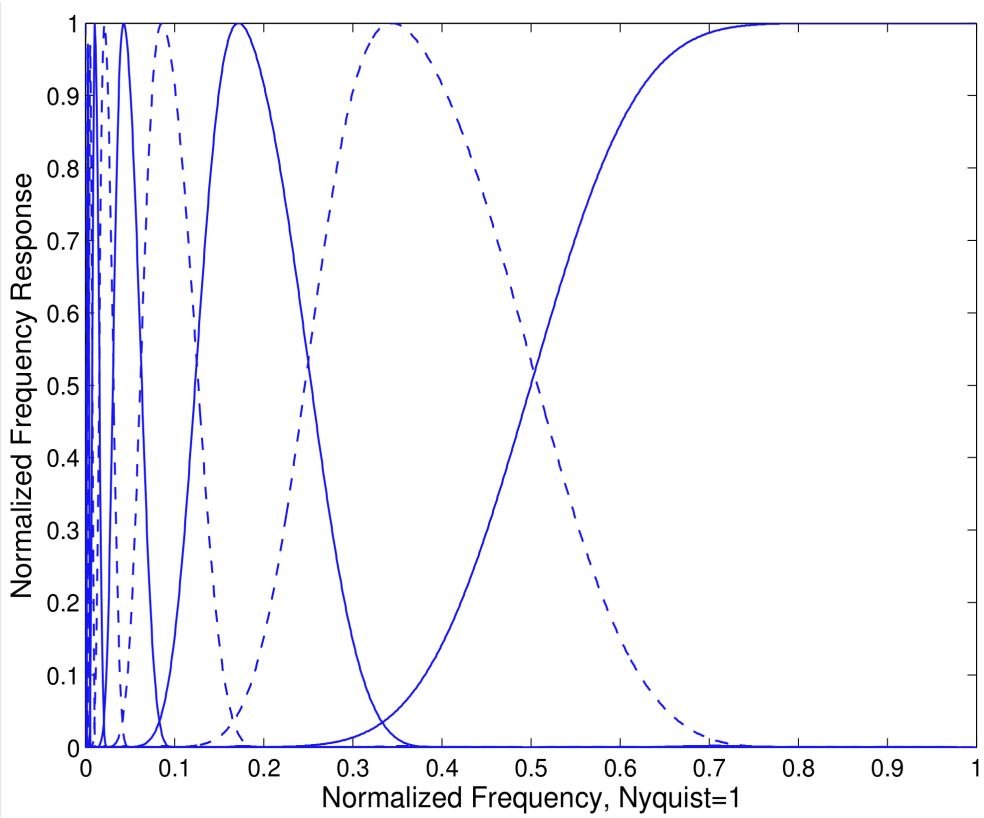
A Wavelet-Based Approach

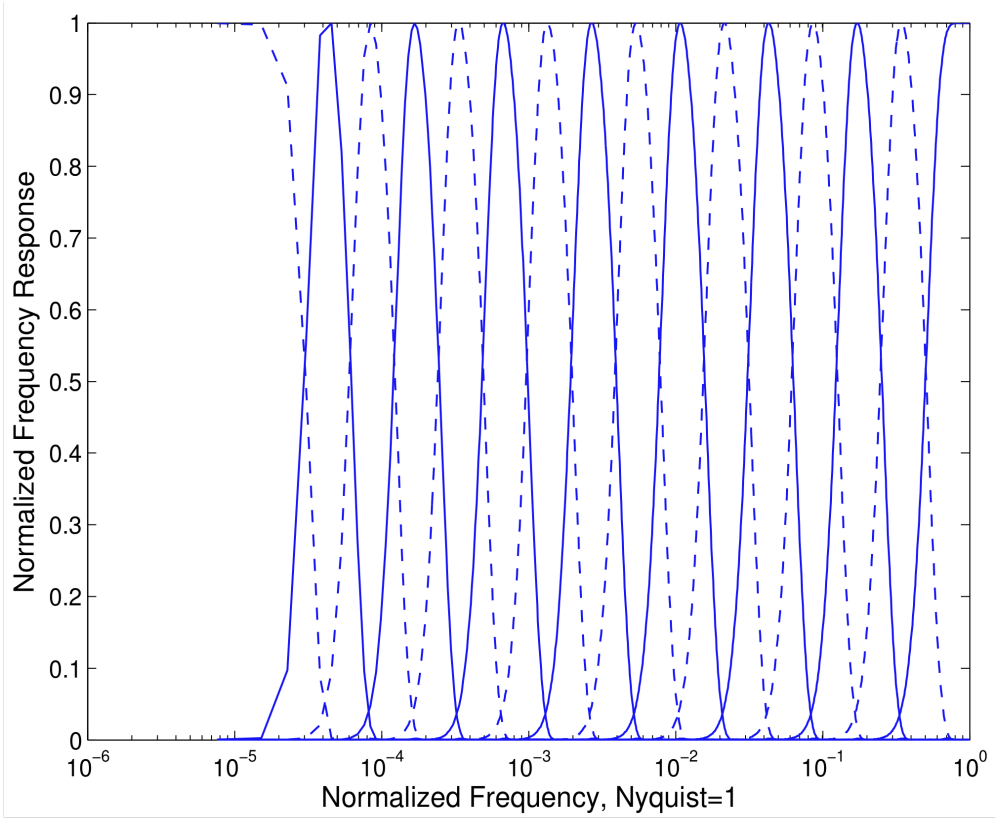
Kepler
A Search for Earth-size Planets

Filter-Bank Implementation of an Overcomplete Wavelet Transform

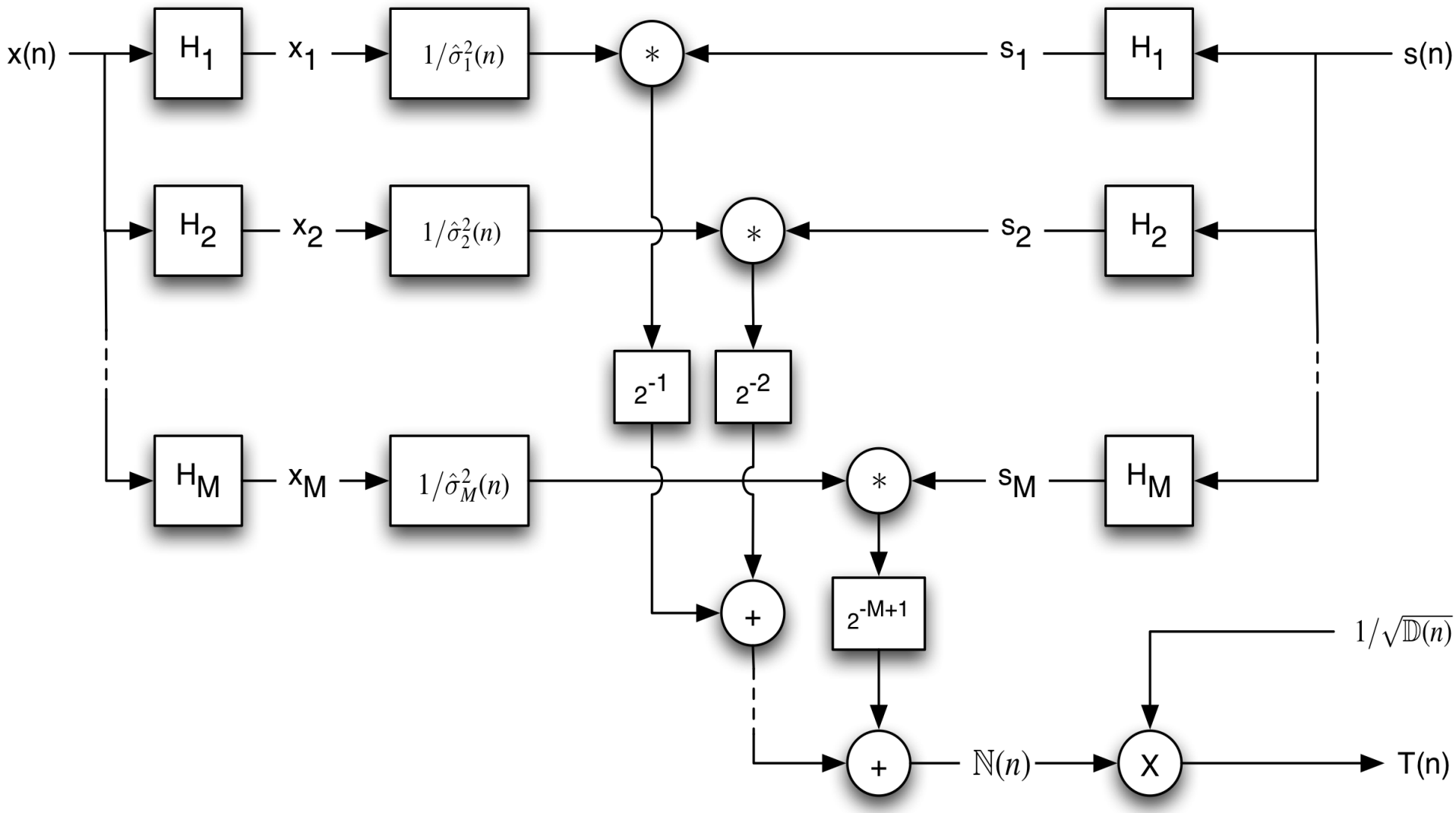
The time series *x(n)* is partitioned (filtered) into complementary channels

$$W_X(i,n) = \{h_1(n) * x(n), h_2(n) * x(n), \ldots, h_M(n) * x(n)\}$$
$$= \{x_1(n), x_2(n), \ldots, x_m(n)\}$$



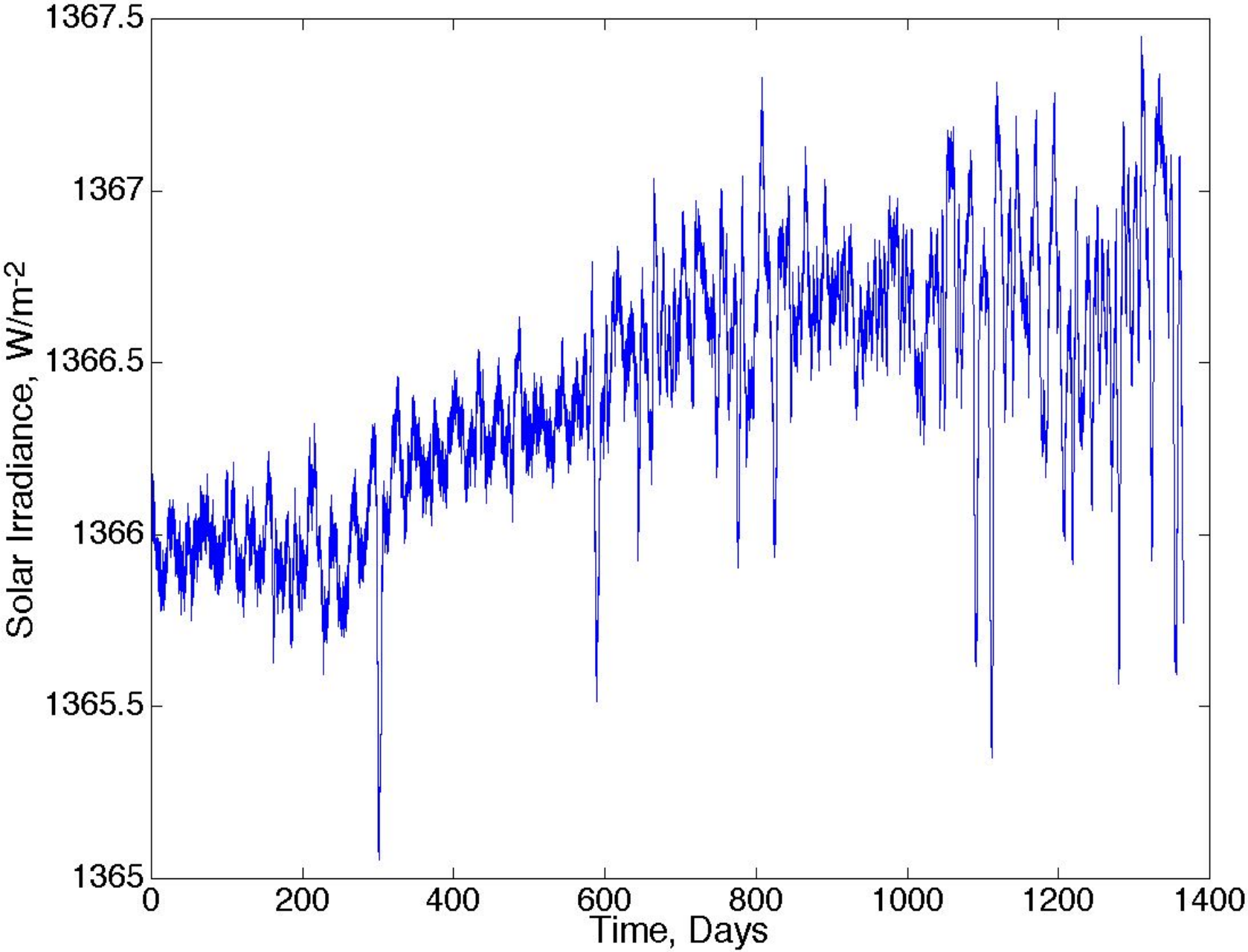


Signal Flow Diagram



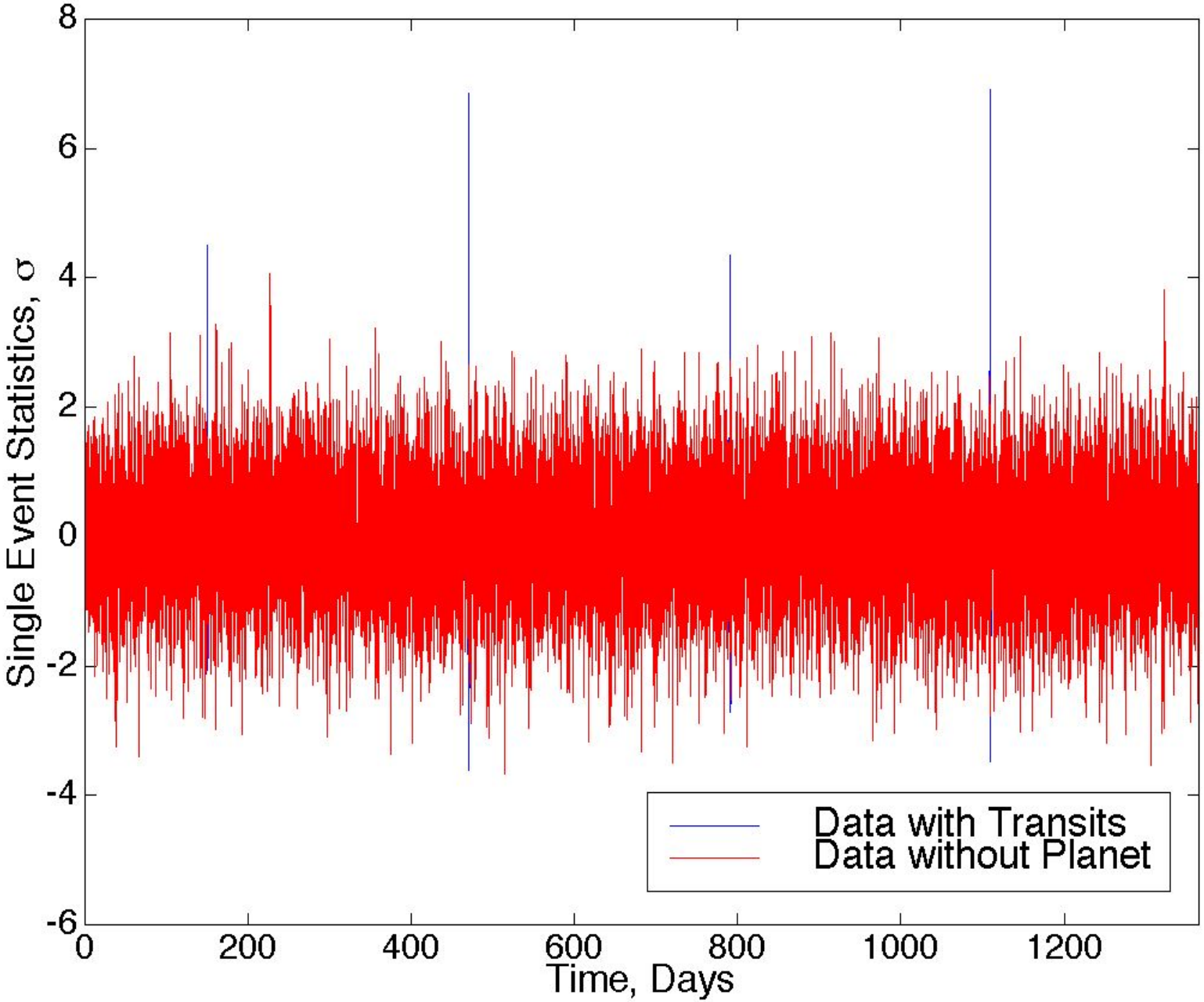
Kepler-like Noise + Transits

A Search for Earth-size Planets



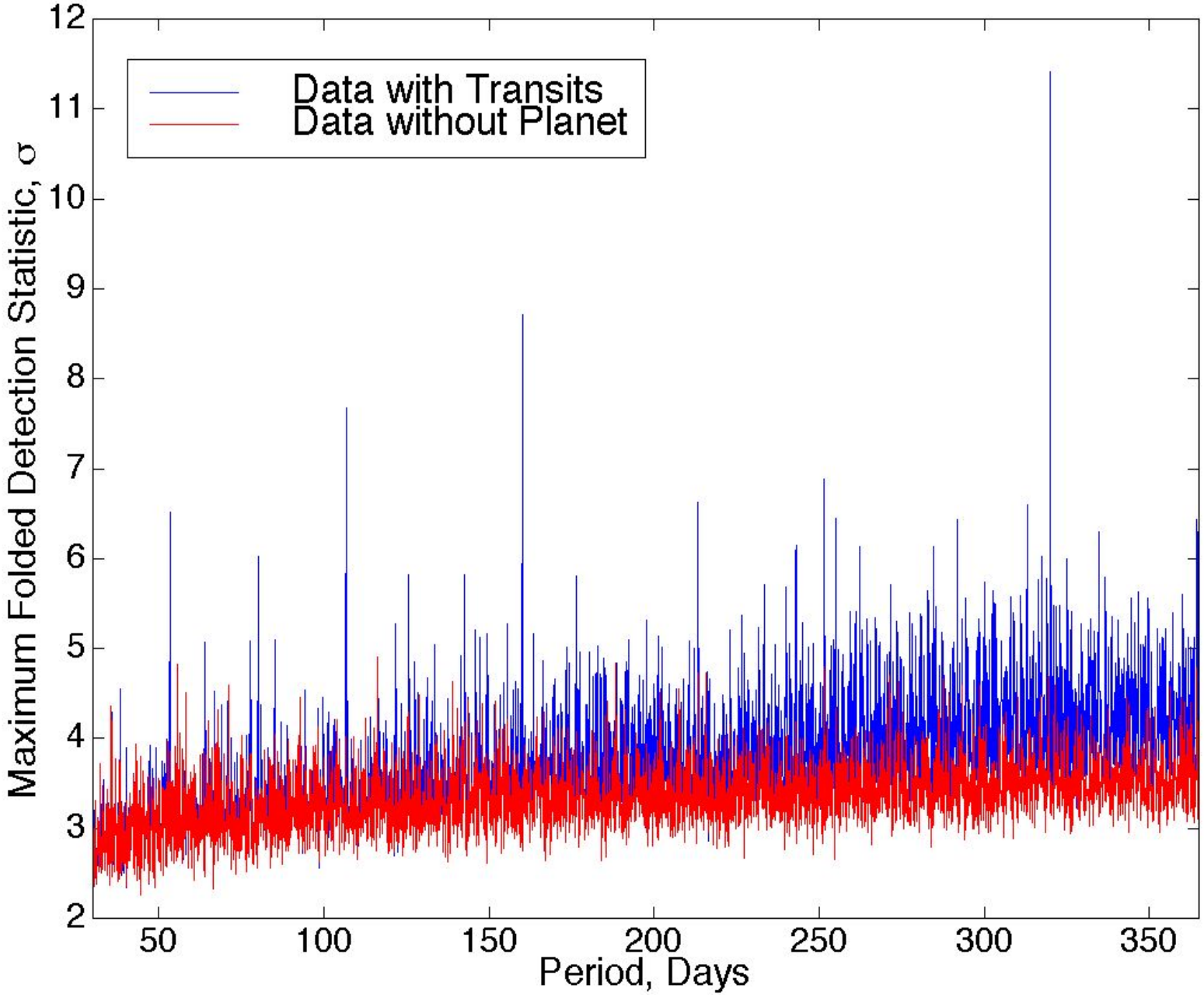
Single Transit Statistics

A Search for Earth-size Planets



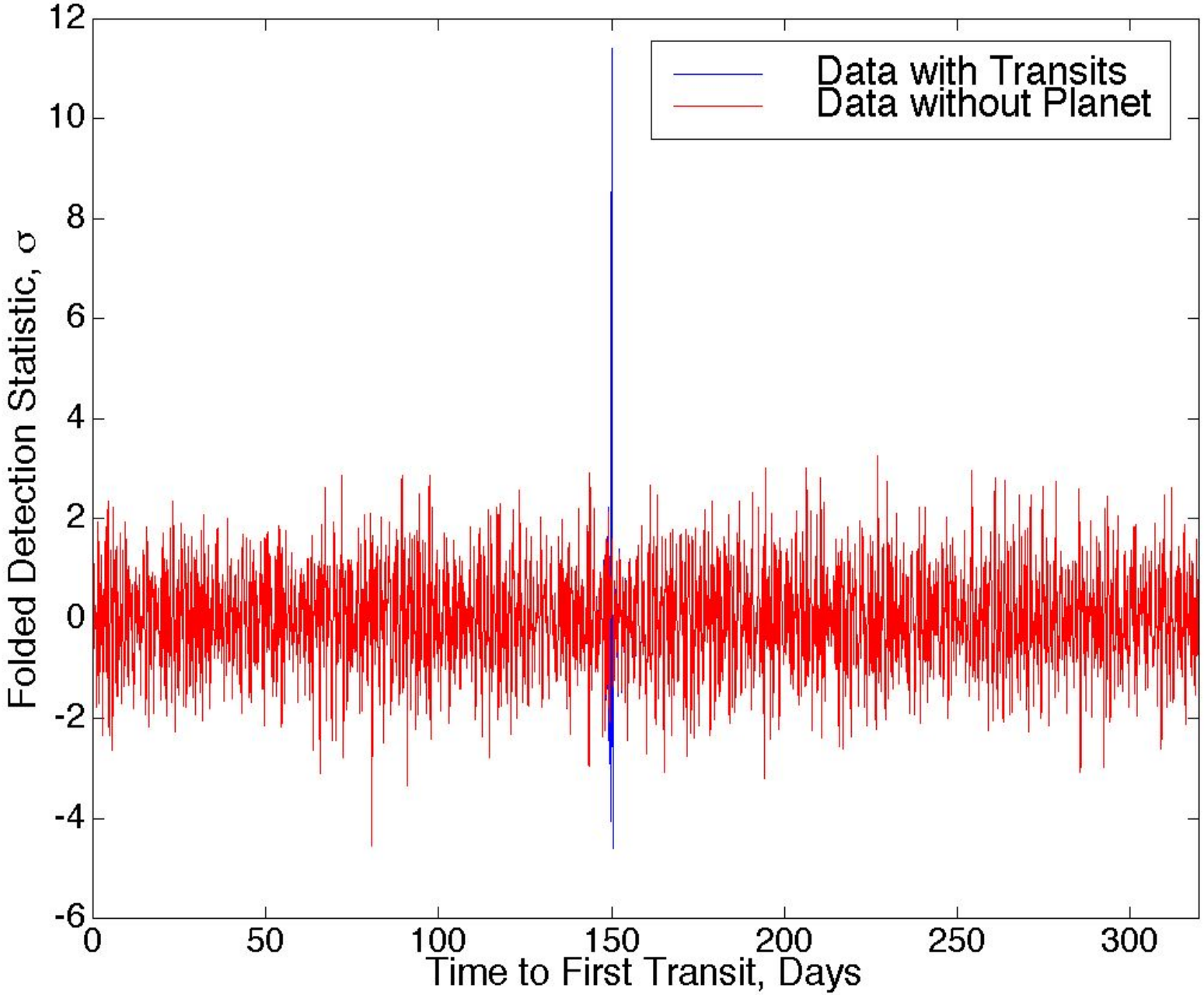
Folded Transit Statistics

Kepler
A Search for Earth-size Planets



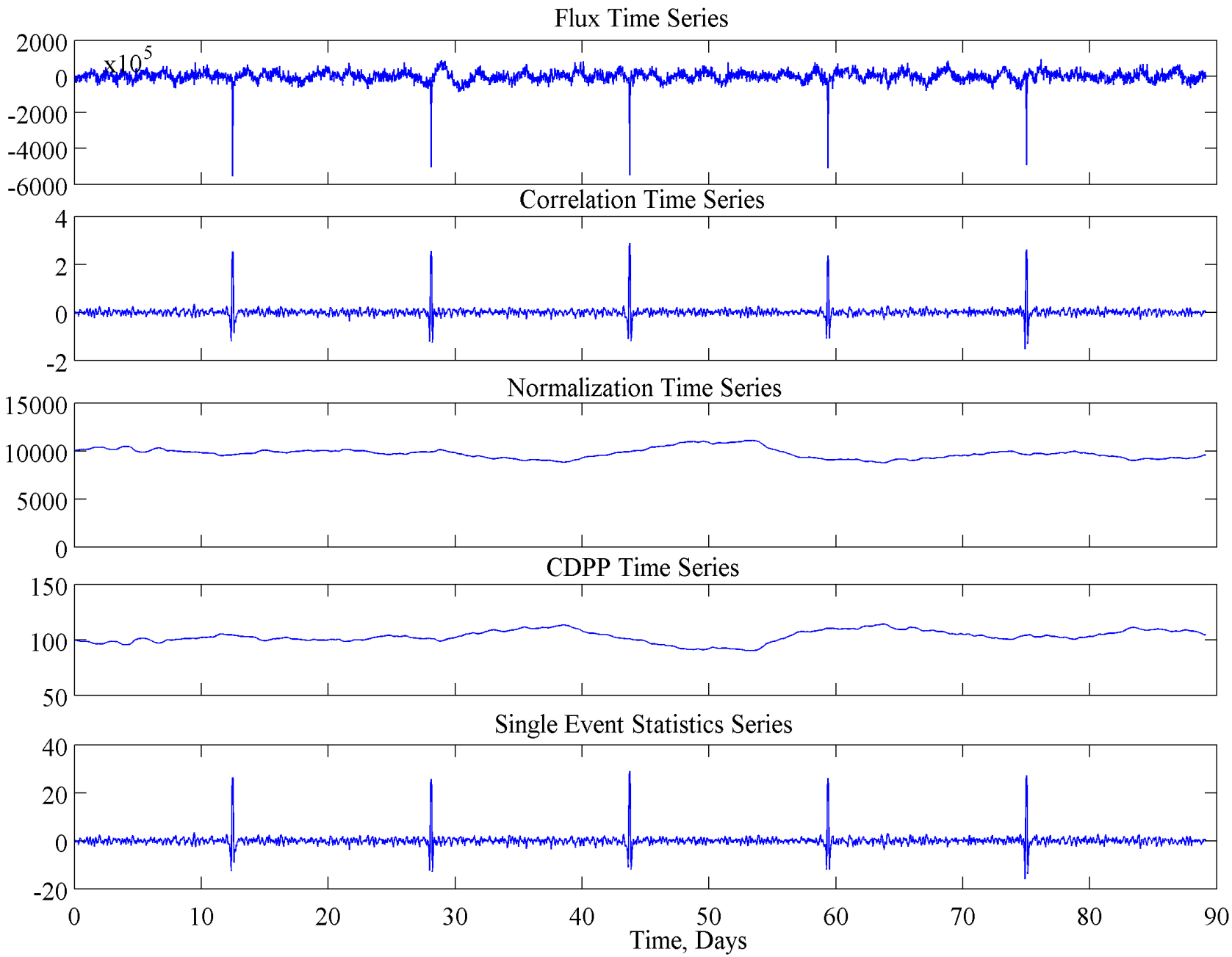
Folded Statistics at Best –Matched Period

A Search for Earth-size Planets

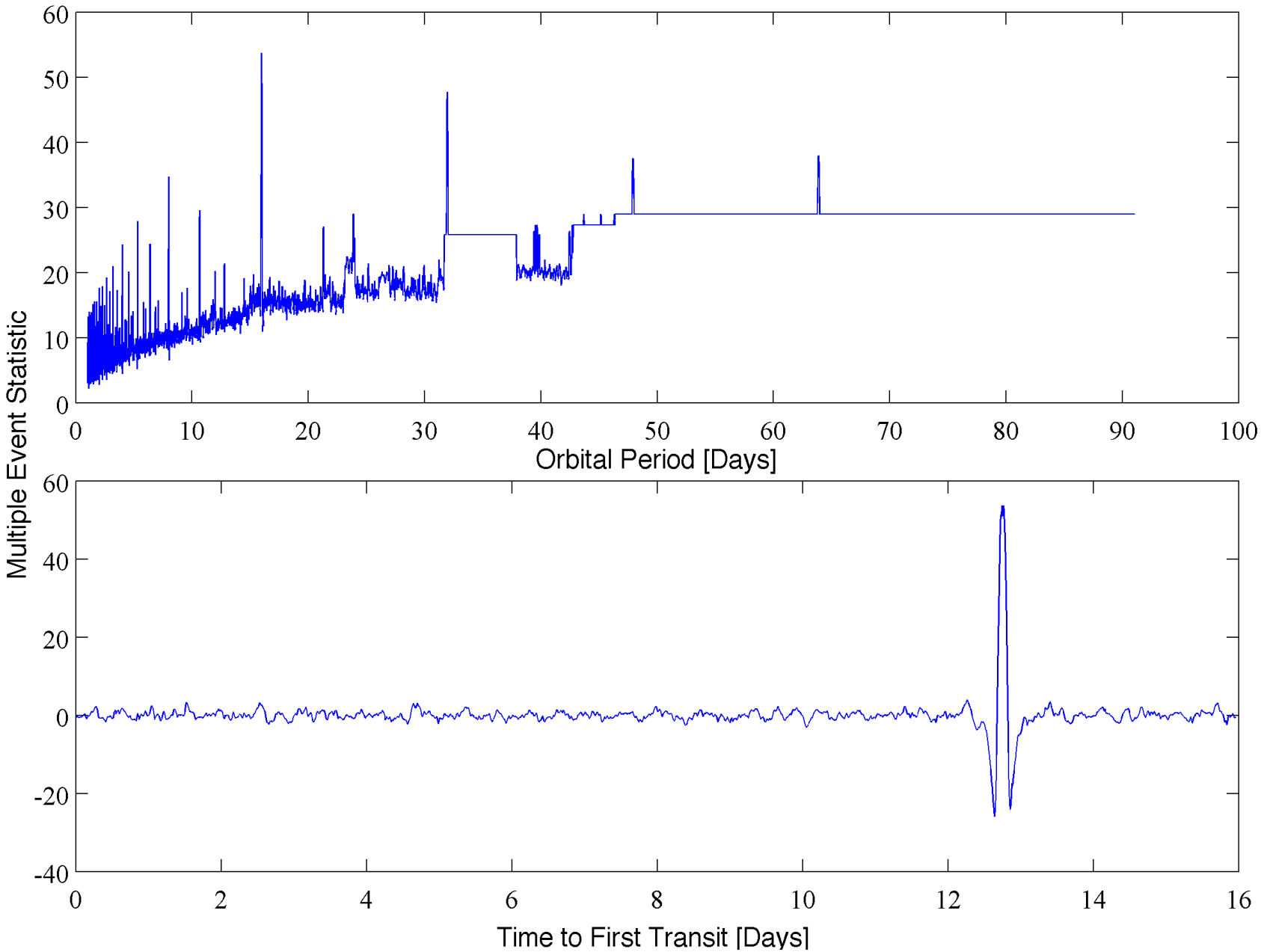


Analyzing a Light Curve

Kepler
A Search for Earth-size Planets

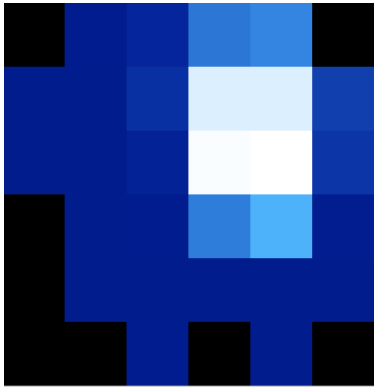


Folding Detection Statistics



The Kepler Science Pipeline: From Pixels To Planets

Kepler
A Search for Earth-size Planets



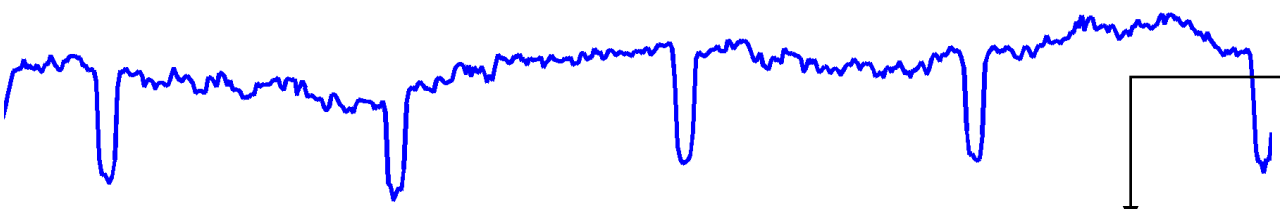
Raw Data

CAL
Pixel-level calibrations

Calibrated Pixels

PA
Photometric Analysis

Sums pixels/ measures star locations



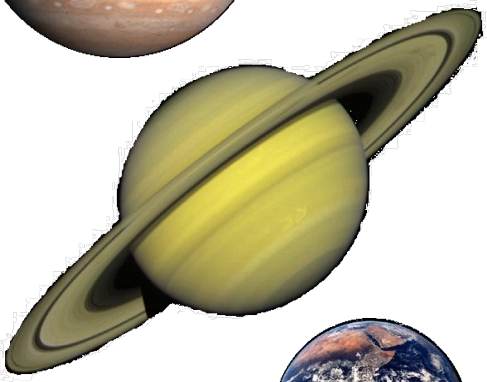
Raw Light Curves/ Centroids

PDC
Presearch Data Conditioning

Removes systematic errors

Corrected Light Curves

TPS
Transiting Planet Search

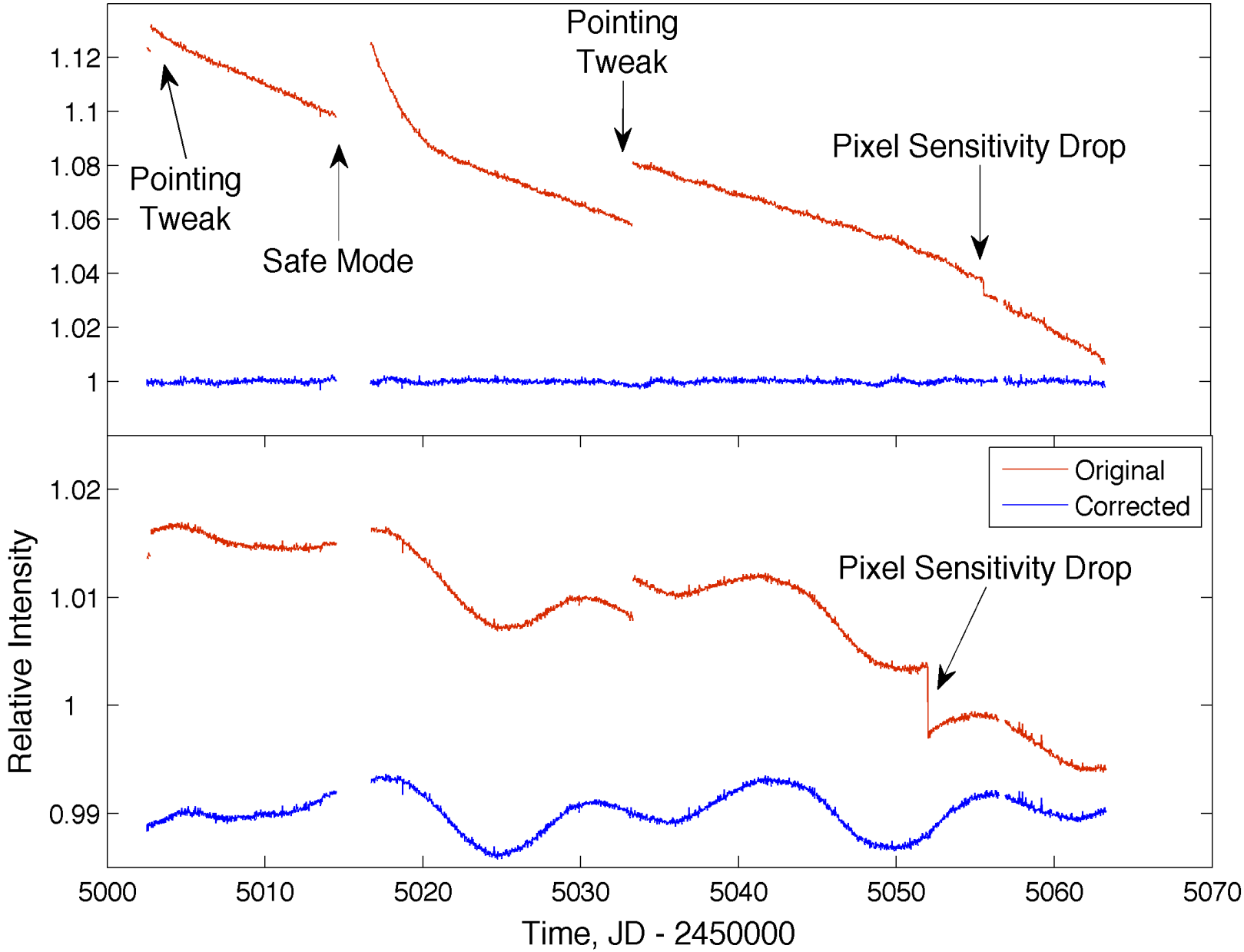


Diagnostic Metrics

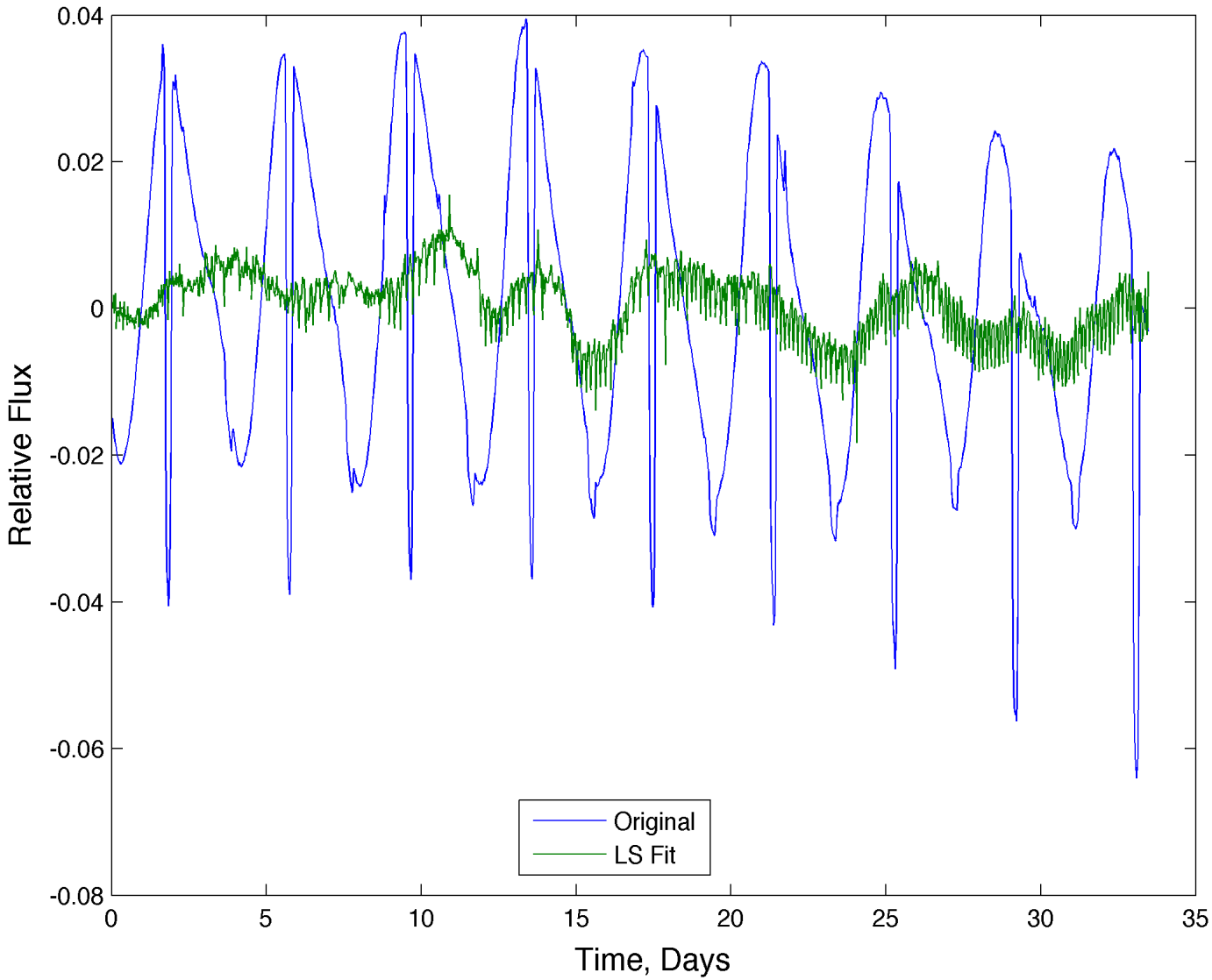
DV
Data Validation

TCEs:
Threshold Crossing Events

PDC Often Does a Good Job



PDC Often Over-Fits Variable Stars



A star fitted with systematics (green) extracted by SVD from 1400 “quiet” stars from Q1 on channel 2.1.

PDC Is Fundamentally Flawed

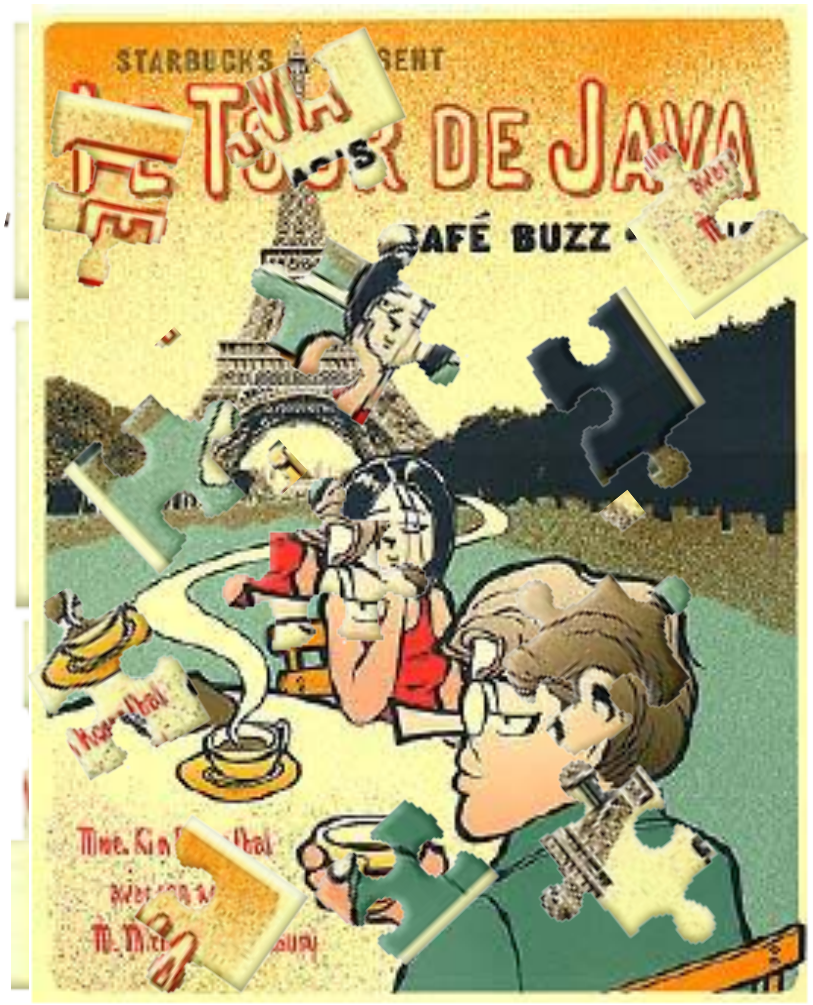
Kepler
A Search for Earth-size Planets

PDC co-trends against instrumental signatures using least squares (LS) approach

LS attempts to explain ***all*** of a given time series, not just the part the model can explain well

There is no way a simple LS fit can “put on the brakes”

PDC often trades bulk RMS for increased noise at short time scales



A Solution

The Problem: Find a model that best fits the data

$$\hat{y} = \mathrm{H}\theta$$

Bayesian approaches to fitting allow one to bring in side information to constrain the fit

We propose the Maximum A Posteriori approach wherein we formulate PDFs for the fit coefficients based on an analysis of the ensemble of stars behavior against a robust LS fit.

The MAP attempts to maximize the conditional expectation

$$\text{argmax } p(c|x)$$

Where c are the coefficients and x are the data.

Bayes' rule gives $p(c|x)*p(x) = p(x|c)*p(c)$ so we can instead maximize

$$\text{argmax } p(x|c)*p(c)$$

An Analytic Formulation

A Search for Earth-size Planets

For a Gaussian prior:

$$\hat{\theta}_{\mathrm{MAP}} = \left(\mathrm{H}^{\mathrm{T}}\mathrm{C}_{\mathrm{w}}^{-1}\mathrm{H} + \mathrm{C}_{\theta}^{-1}\right)^{-1}\left(\mathrm{H}^{\mathrm{T}}\mathrm{C}_{\mathrm{w}}^{-1}\mathrm{y} + \mathrm{C}_{\theta}^{-1}\mu_{\theta}\right)$$

For uniform white noise

$$\hat{\theta}_{\mathrm{MAP}} = \left(\mathrm{H}^{\mathrm{T}}\mathrm{H} + \sigma^{2}\mathrm{C}_{\theta}^{-1}\right)^{-1}\left(\mathrm{H}^{\mathrm{T}}\mathrm{y} + \sigma^{2}\mathrm{C}_{\theta}^{-1}\mu_{\theta}\right)$$

As $\sigma^{2} \rightarrow 0, \quad \hat{\theta}_{\mathrm{MAP}} \rightarrow \left(\mathrm{H}^{\mathrm{T}}\mathrm{H}\right)^{-1}\left(\mathrm{H}^{\mathrm{T}}\mathrm{y}\right) = \hat{\theta}_{\mathrm{MLE}}$

While as $\sigma^{2} \rightarrow \infty, \quad \hat{\theta}_{\mathrm{MAP}} \rightarrow \left(\mathrm{C}_{\theta}^{-1}\right)^{-1}\left(\mathrm{C}_{\theta}^{-1}\mu_{\theta}\right) = \mu_{\theta}$

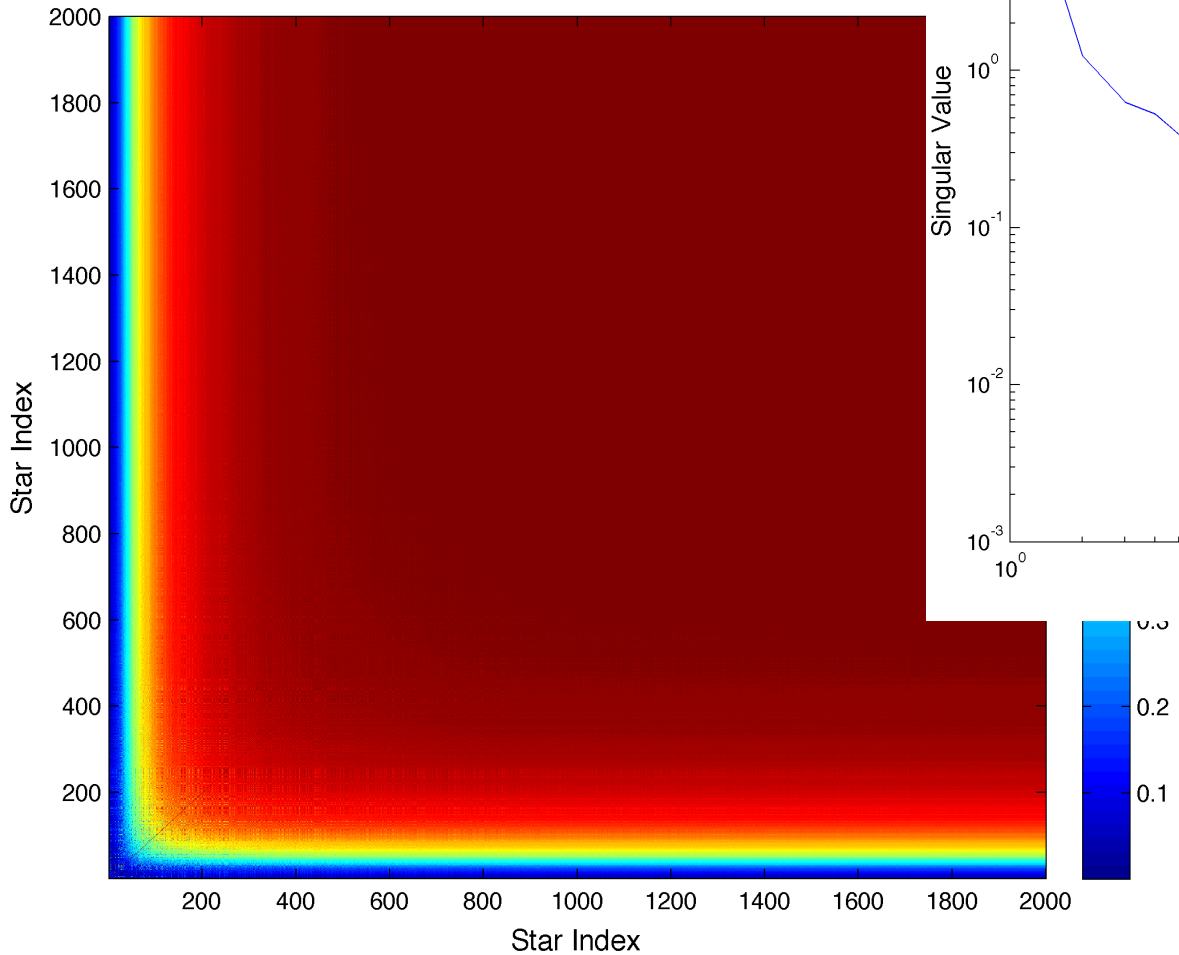
SVD decomposition of Quiet Stars

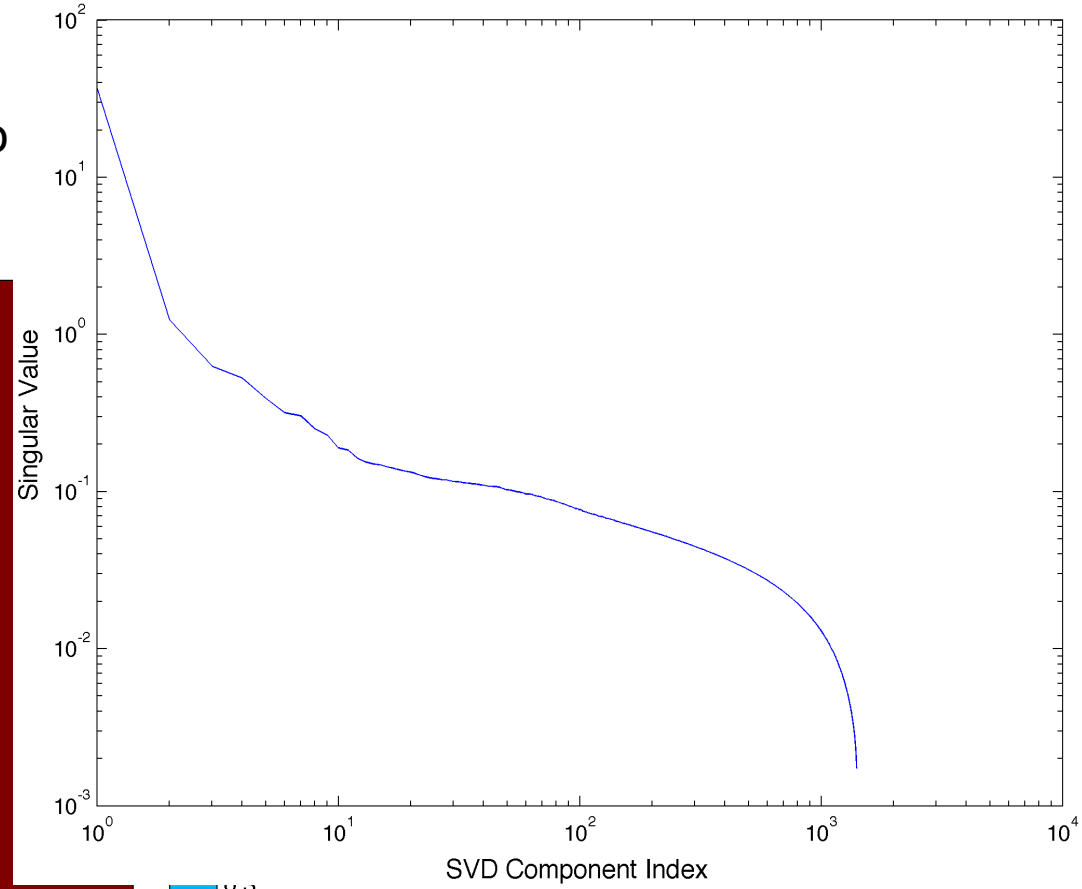
A Search for Earth-size Planets

600 stars unique and poorly correlated with other stars: these are the variable stars

SVD of 1400 most-correlated stars

The “knee” occurs at component index 10 or so





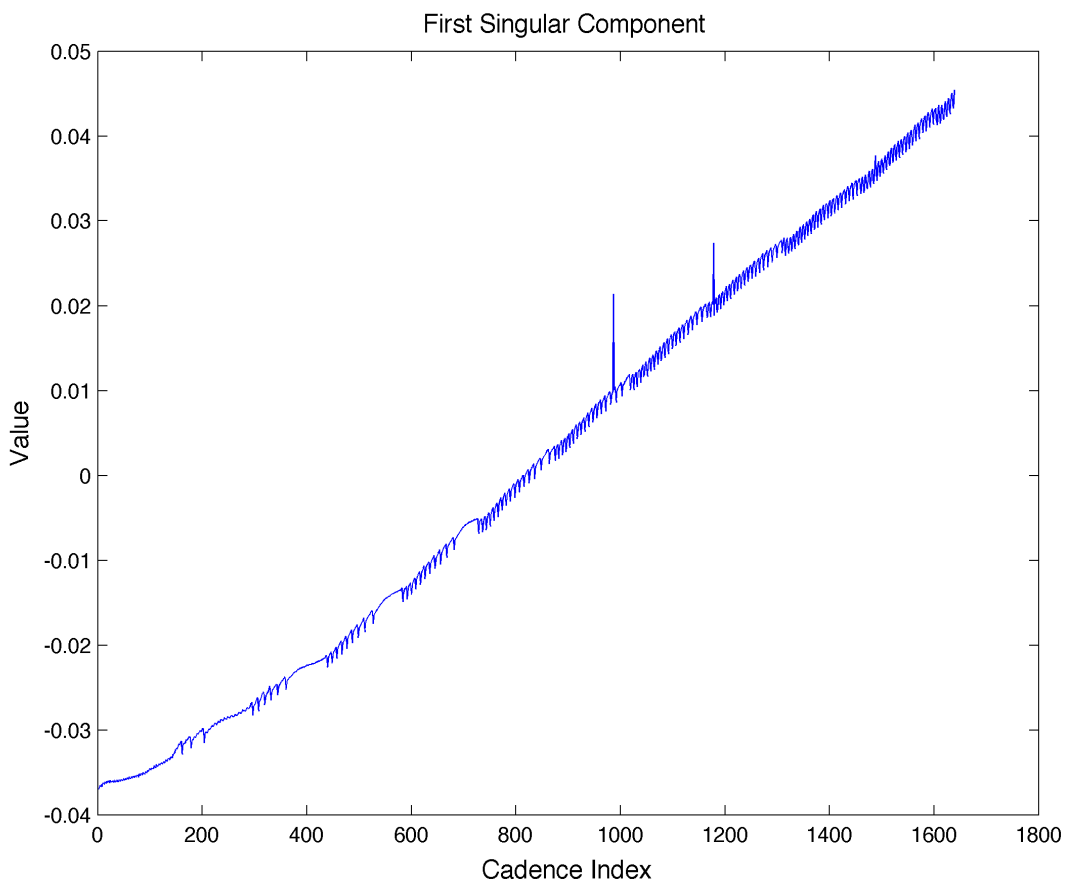
Determining the PDFs of the Fit Coeffs

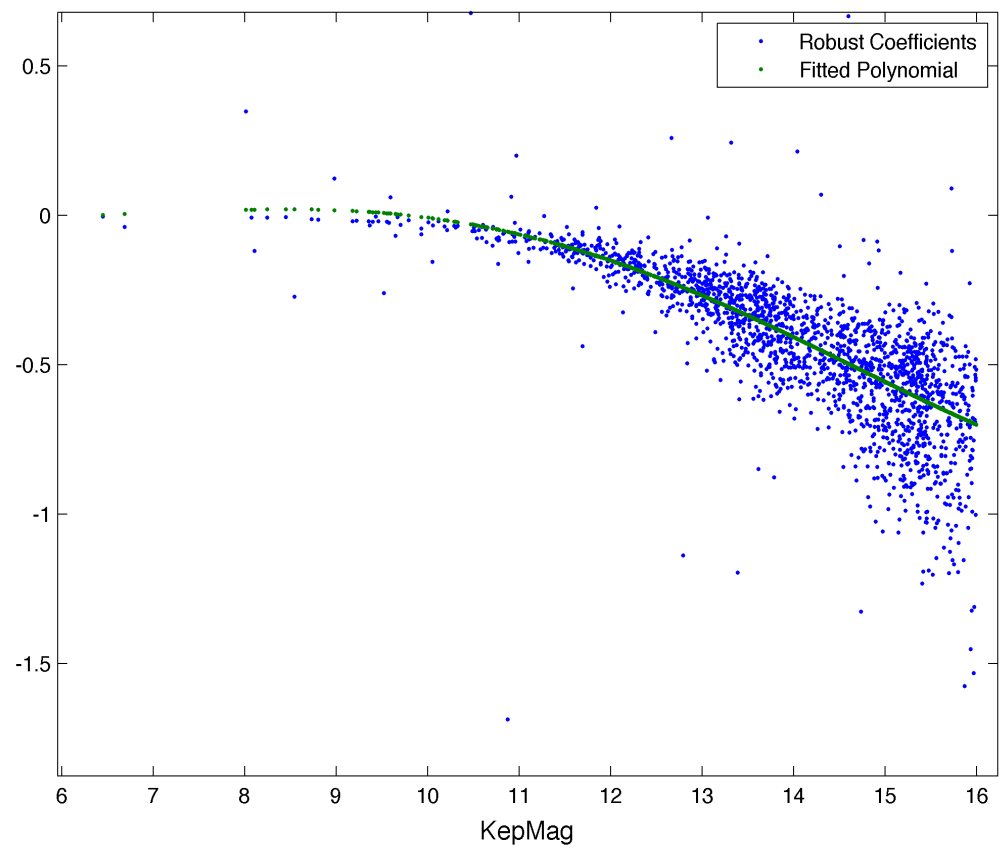
Kepler
A Search for Earth-size Planets

The first term looks like mostly focus variations, with Argabrightenings.

We observe dependencies in the systematic terms to stellar brightness and position on the sky

For this analysis we modeled the PDFs as Gaussian but with means dependent on magnitude

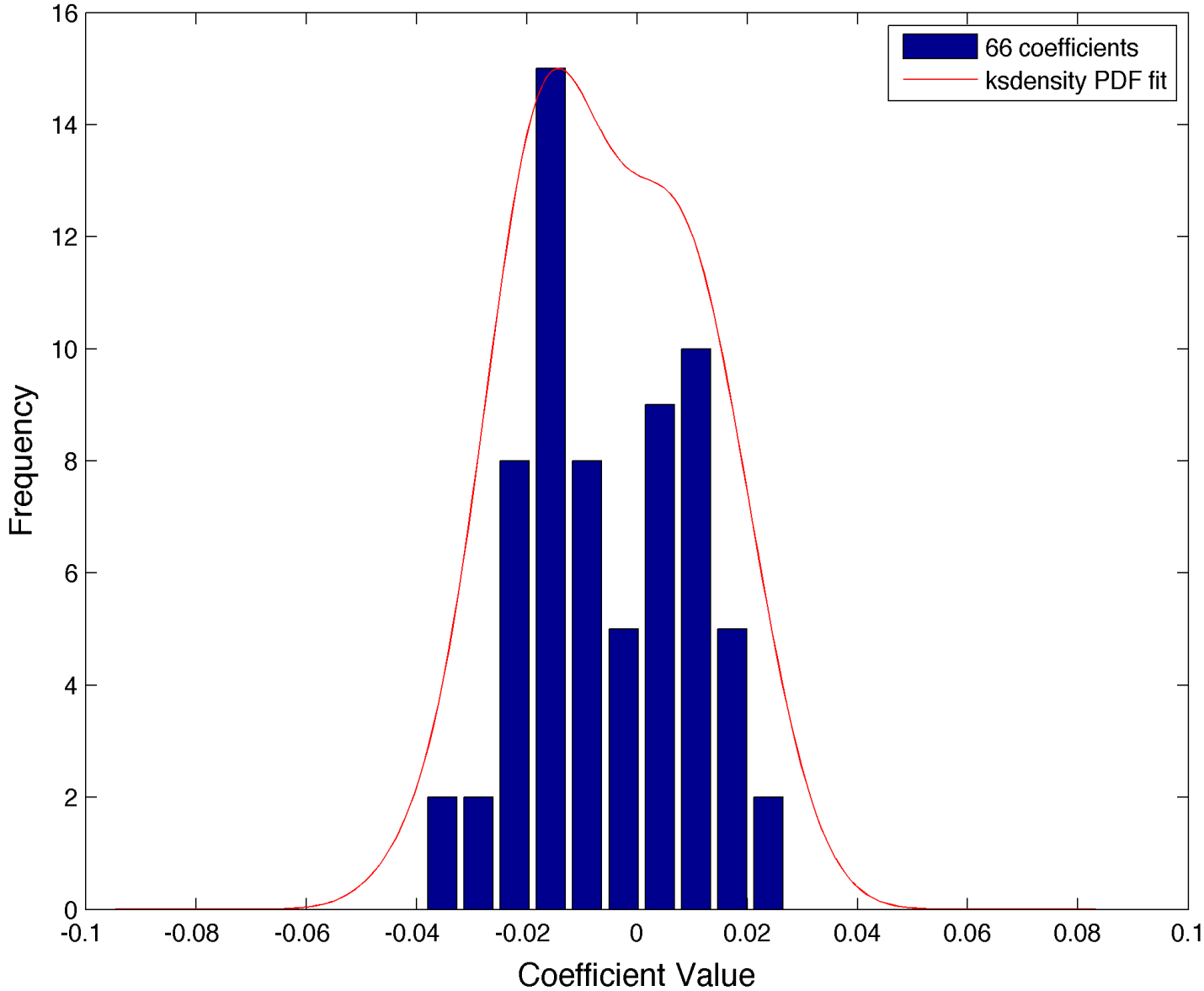




Actual PDFs Are Non-Gaussian

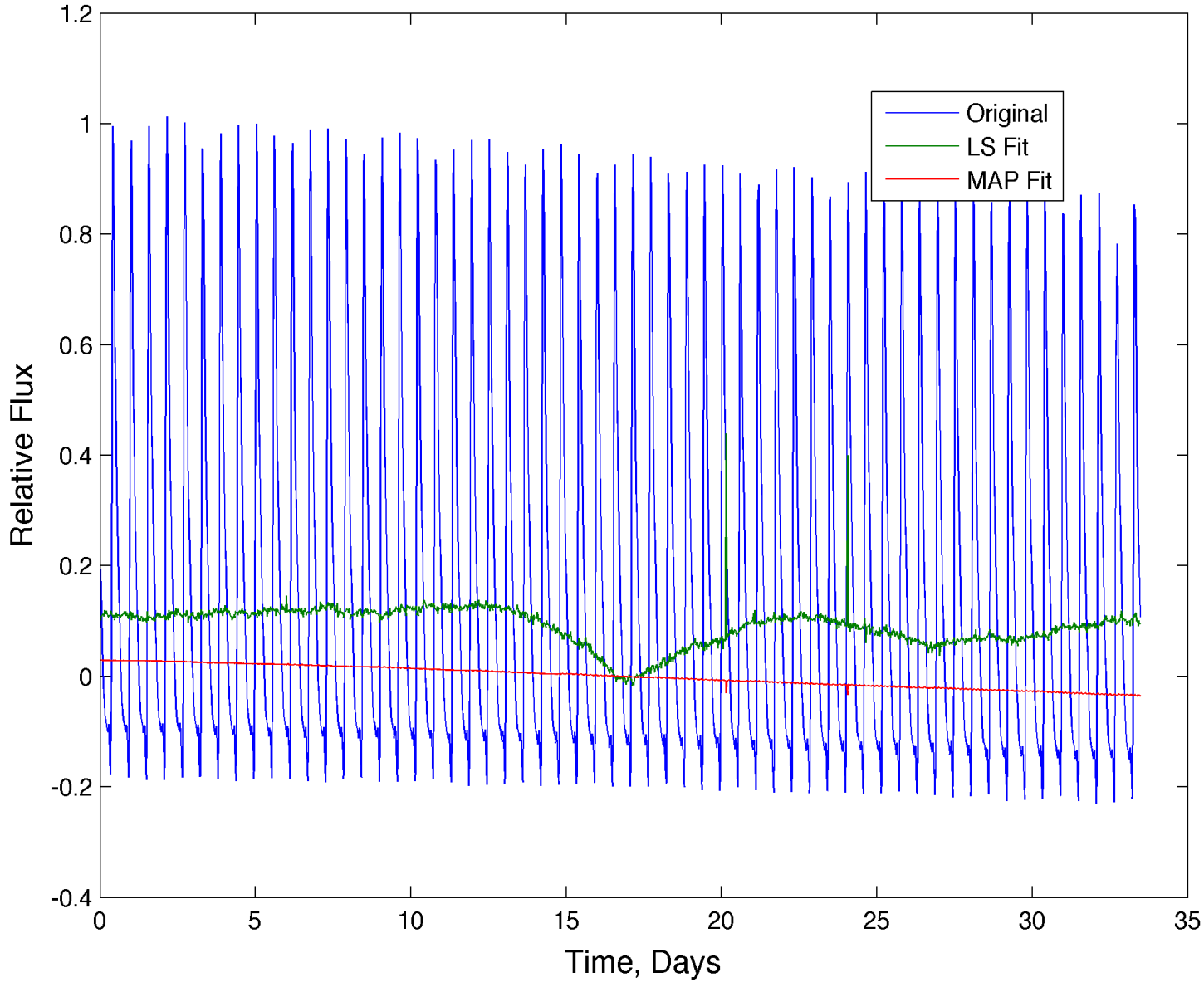
A Search for Earth-size Planets

An Example of an empirical PDF from the development underway at the SOC



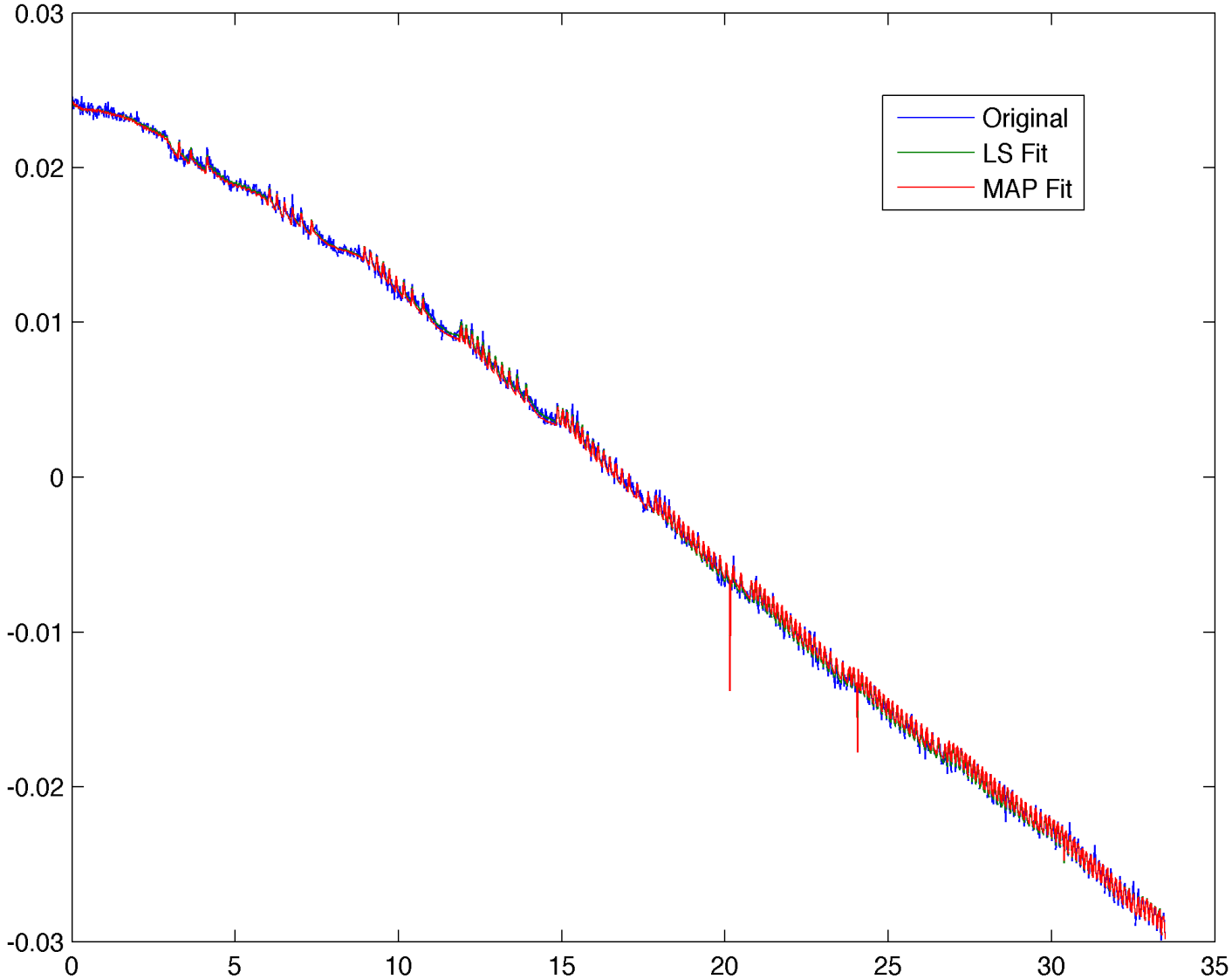
Example (1)

An RR Lyrae star that is treated poorly by robust LS, but is well-served by MAP



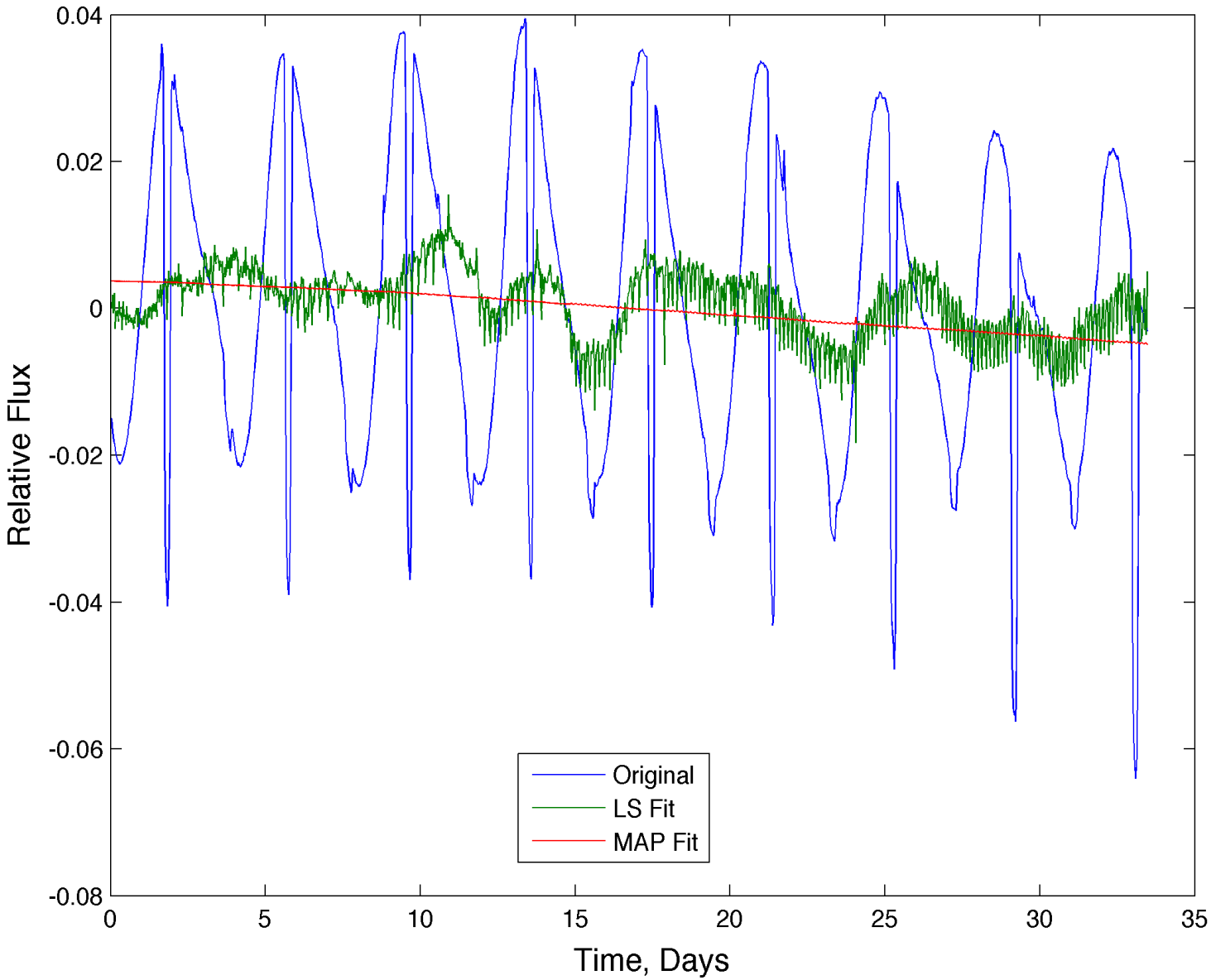
Example (2)

A "quiet" star dominated by systematics. The LS and the MAP fits are comparable.



Example (3)

A not-so-quiet star.



Conclusions

The *Kepler* Mission is returning a wealth of exoplanet science (and asteroseismology)

A Wavelet-based approach allows us to detect small planets in the presence of time-varying, non-white observation noise and stellar variability

The MAP approach can be applied to the current co-trending approach in PDC

MAP allows us to reconstruct the physical constraints underlying the propagation of instrumental signatures to stellar flux time series implicitly by observing how a LS fit behaves for most stars.

MAP promises to improve our ability to identify and remove instrumental signatures from the *Kepler* light curves while preserving intrinsic stellar variability