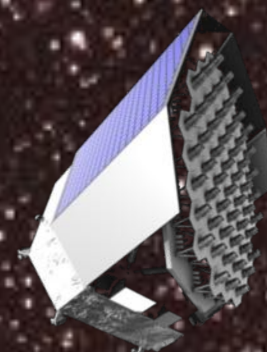
The Impact of Stellar Variability on the Detection of Transiting Earth-like Planets



Jon M. Jenkins

NASA Ames Research Center

Wednesday October 24 2018

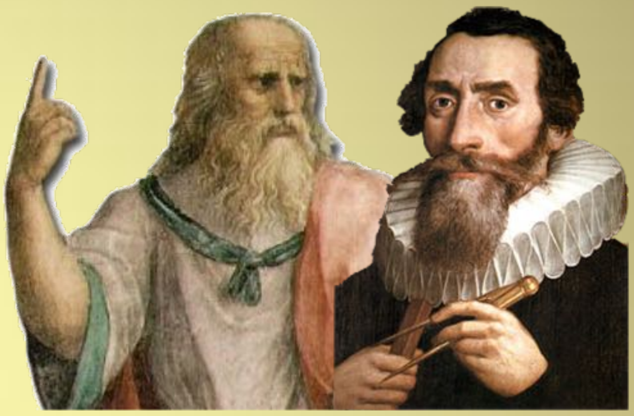
Institute for Planetary Science

Laboratoire d'Astrophysique de Marseill

Marseille, France

Overview

- What did it take to build the *Kepler* science pipeline?
- The SOC Pipeline
- Solar Variability
- Detection Theory
- A Wavelet-based Adaptive Matched Filter
- Observations of Stellar Noise on Transit Timescales
- Excess Stellar Variabilty
- Developing the TESS Pipeline
- Summary



A Search for Earth-size Planets


KEPLER
SCIENCE DATA PROCESSING PIPELINE
DYN
CAL
PA
PDC
DV
TPS

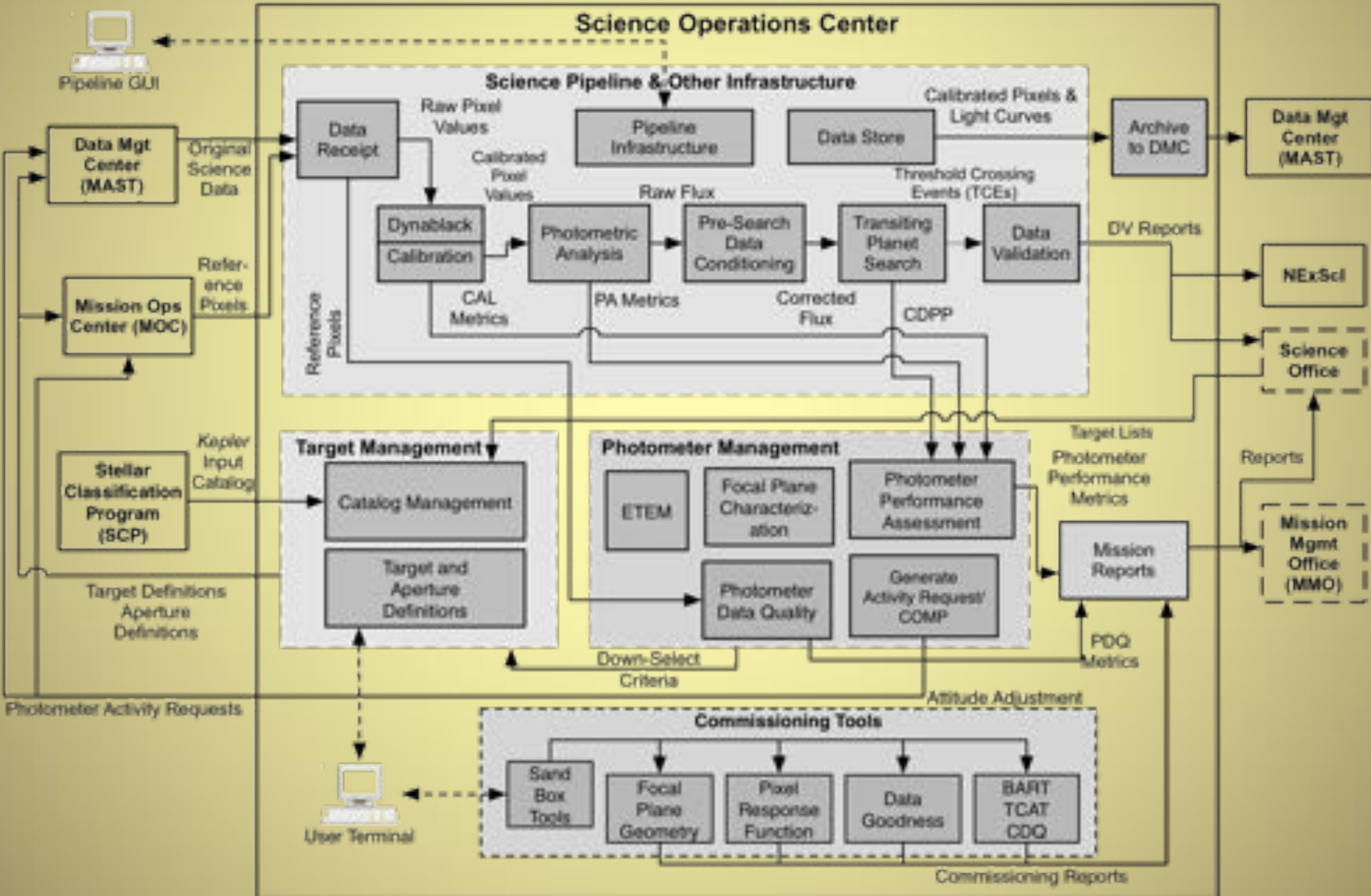
The Science Operations Center: What did it take?

Kepler
A Search for Earth-size Planets

- Design started in earnest in 2004 with launch in March 2009 and operations through May 2013 and reprocessing through 2017
- A total of ~100 person years of effort went into the first complete version of the pipeline (from pixels to planets)
- The staffing was at ~20 individuals per year through 2016, tapering off thereafter (~280 FTEs over project lifetime)
- Build 5.0 was the launch-ready software release
- There were 4 major builds thereafter, with substantive point releases to mitigate issues subsequently identified in flight or full volume re-processing
- Build 9.0, 9.1, 9.2, 9.3 really represented at least two full builds of effort (issues identified in full re-processing and in completeness and reliability processing)
- Unexpected instrumental effects/stellar variability/hardware failures motivated significant software modifications on orbit

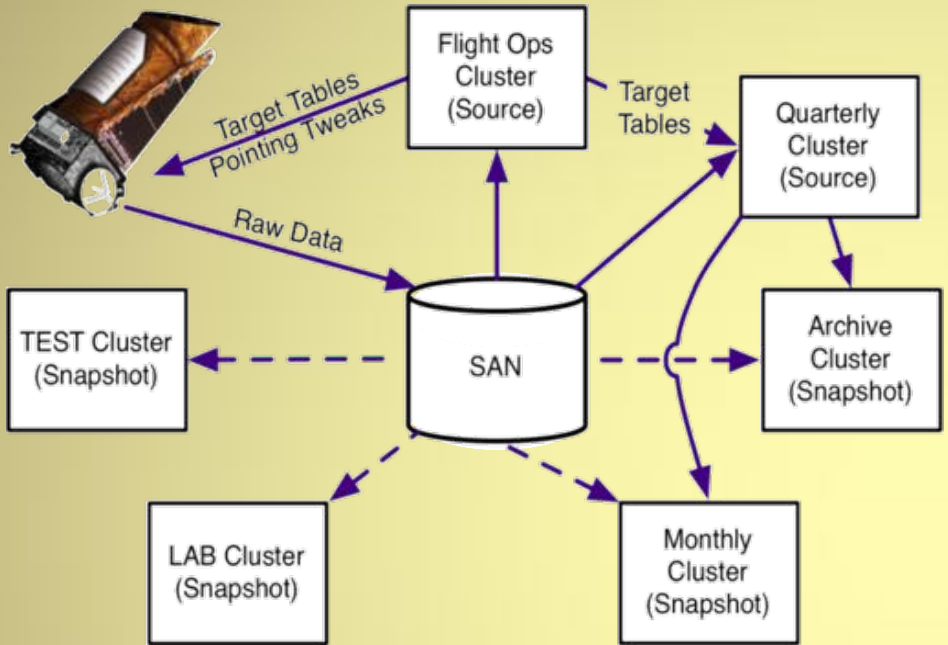
Science Operations Center Architecture

Kepler
A Search for Earth-size Planets



SOC Cluster Architecture

A Search for Earth-size Planets



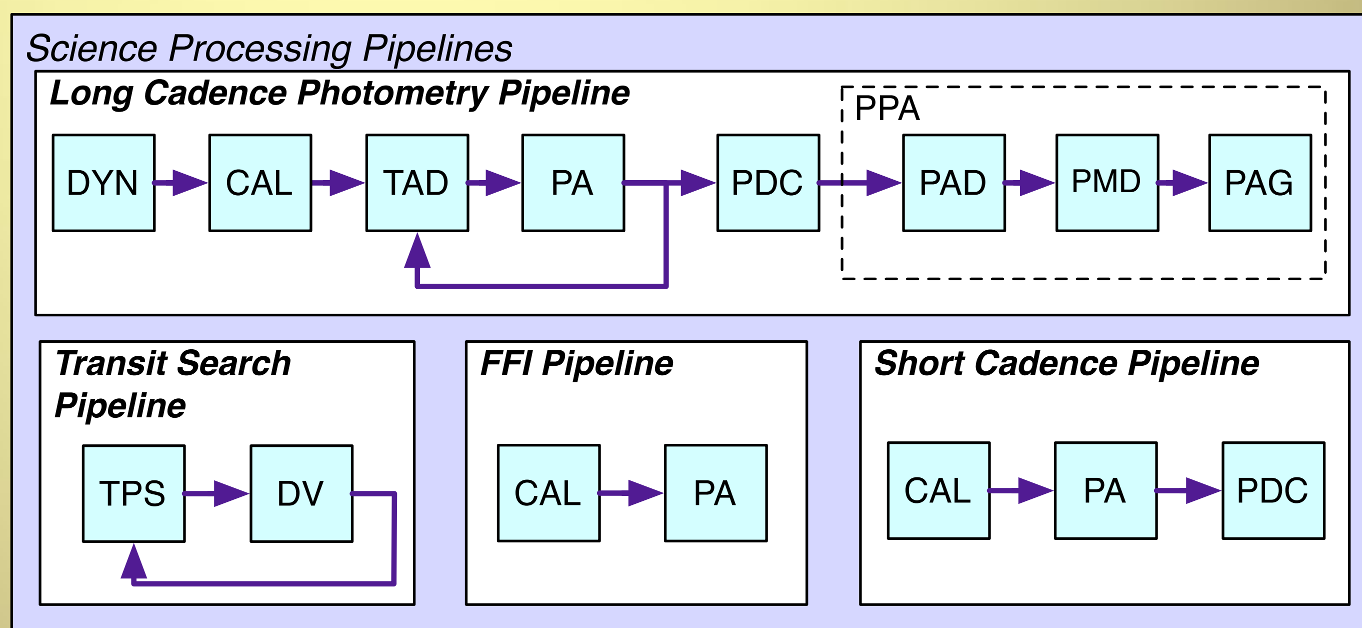
6 Clusters:

4 Operations Clusters:

Flight Ops, Quarterly, Monthly & Archive)

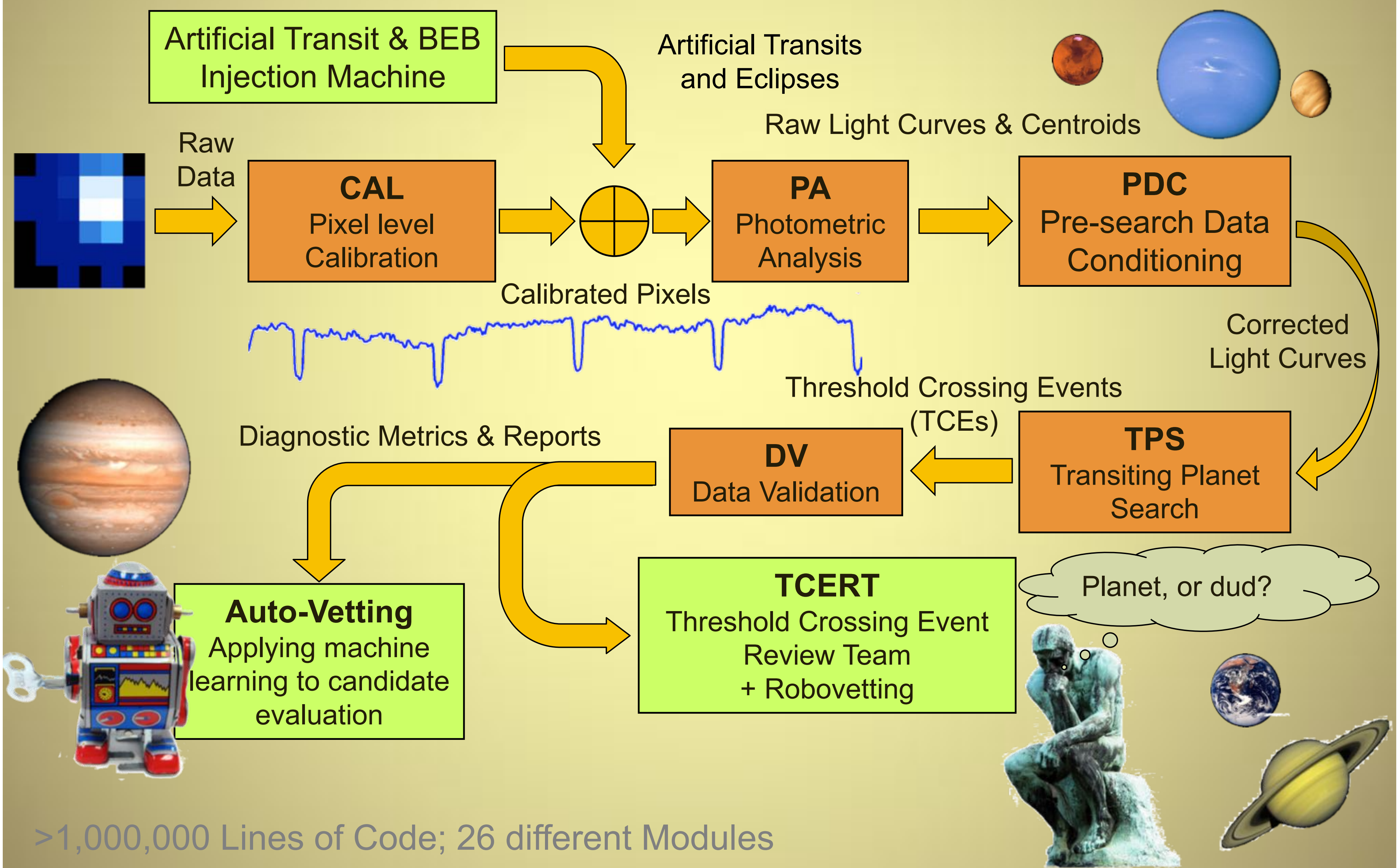
2 Test Clusters:

LAB & TEST



Kepler's Science Pipeline

Kepler
A Search for Earth-size Planets

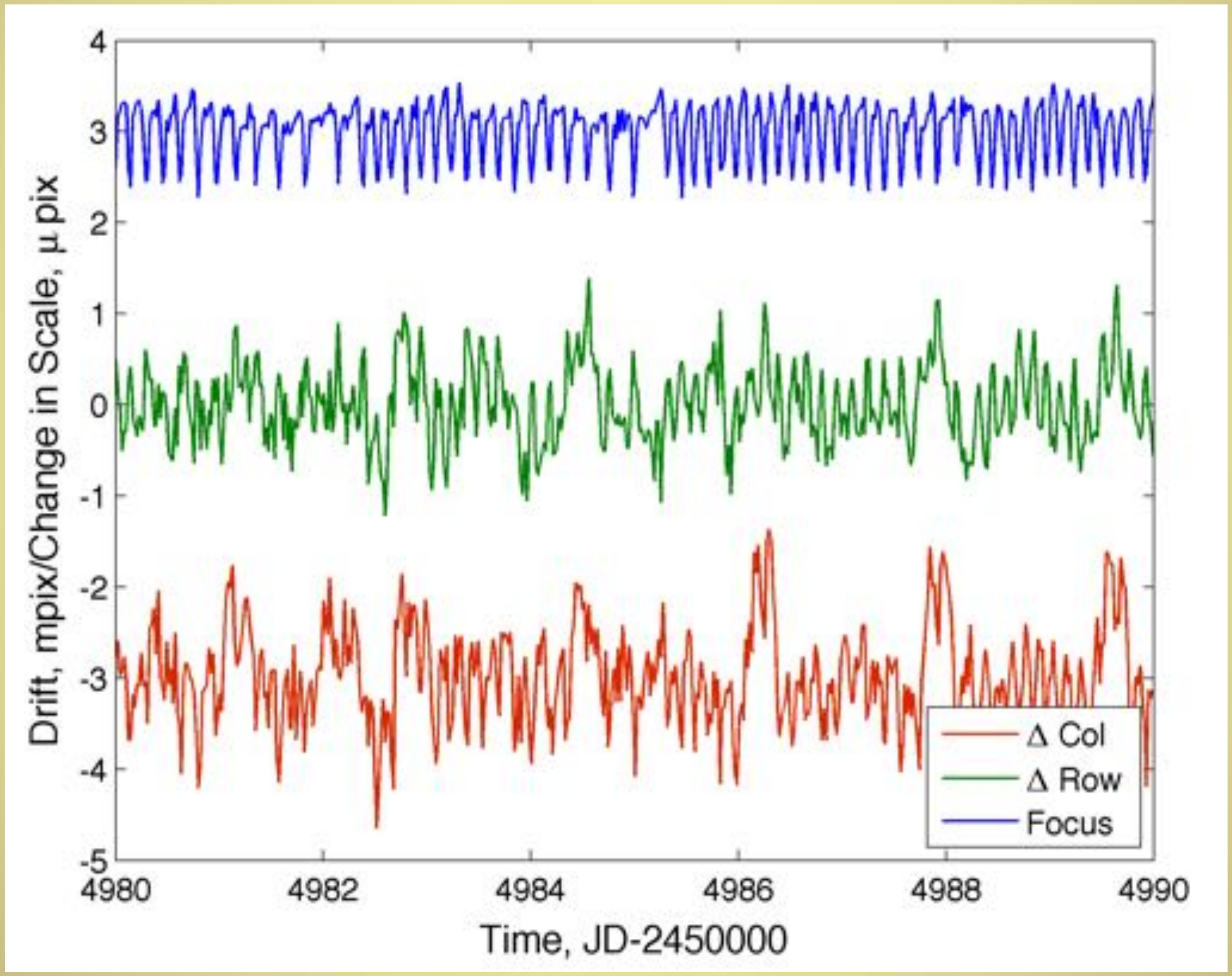


>1,000,000 Lines of Code; 26 different Modules

Short Timescale Instrumental Errors

A Search for Earth-size Planets

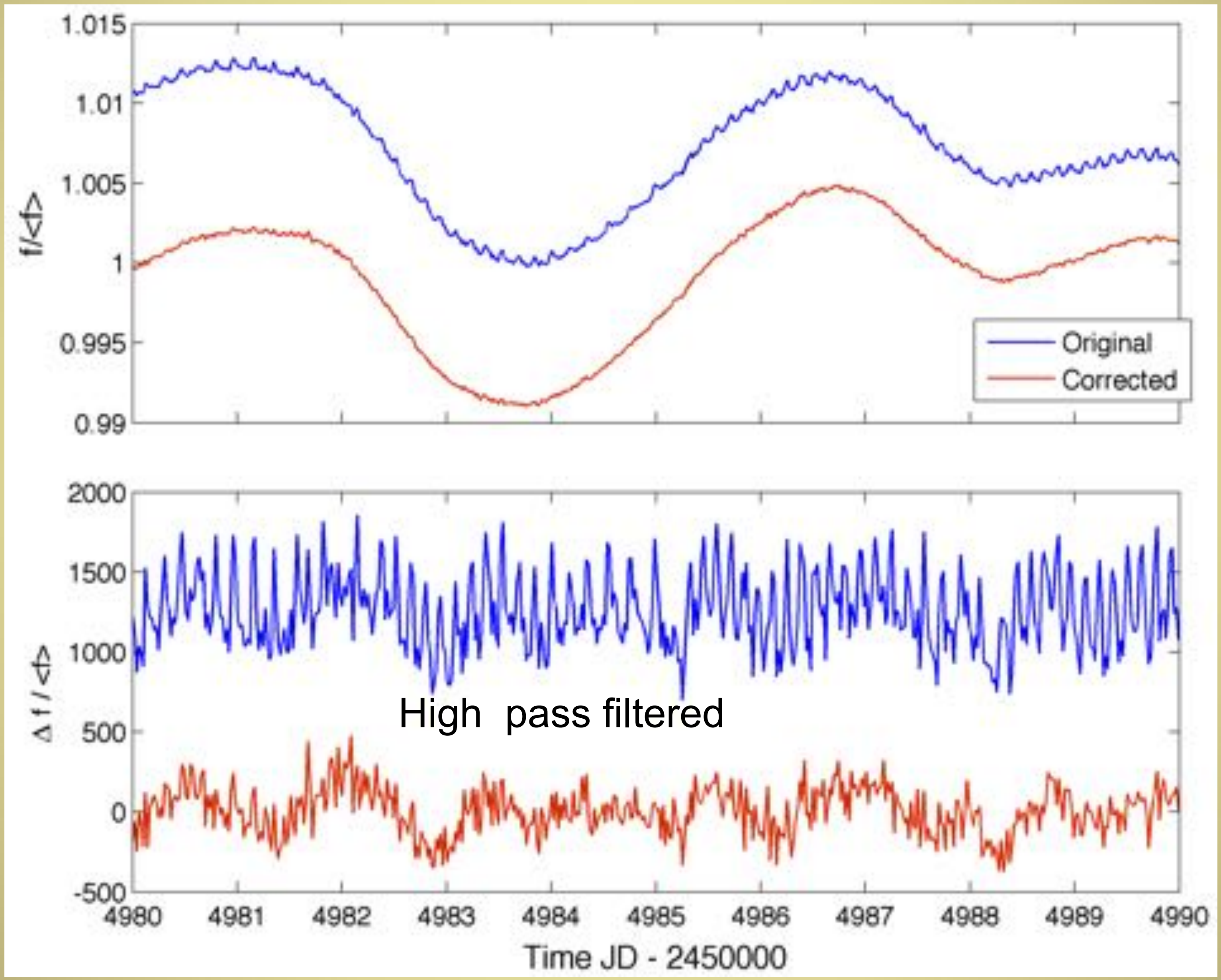
Signature of a heater cycling on the reaction wheels 3/4



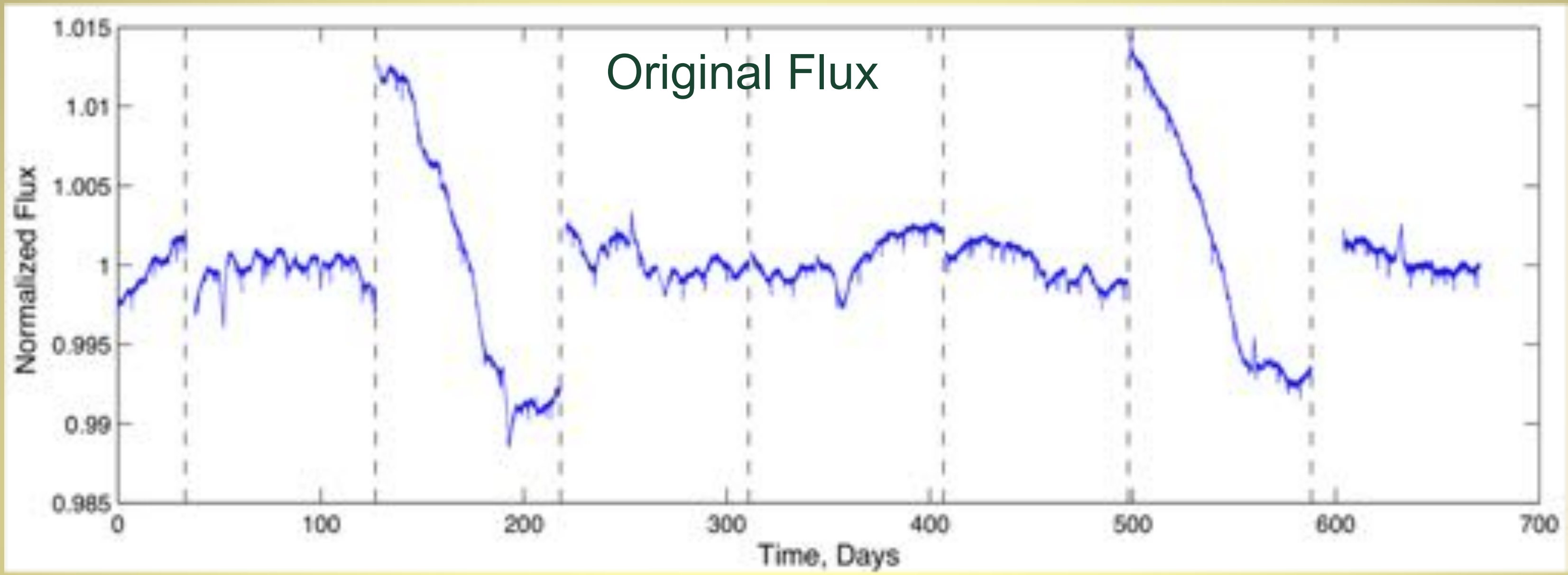
Kepler is sensitive to its thermal environment

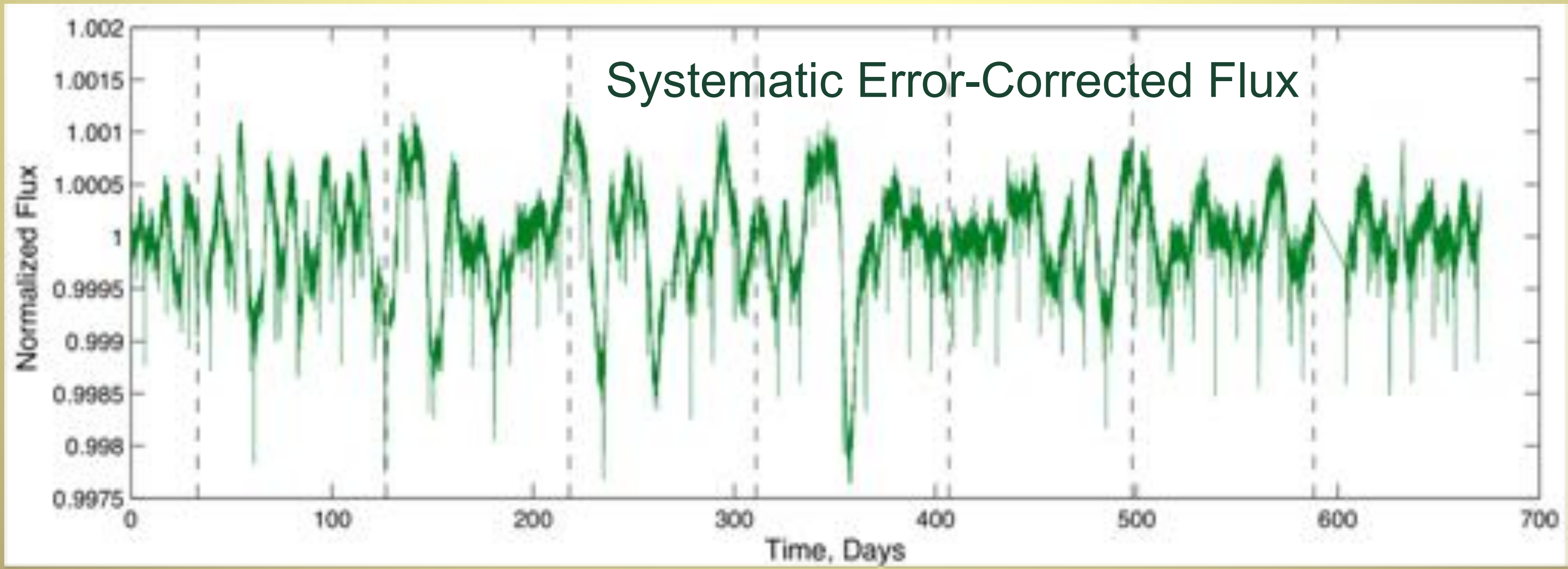
Instrumental Effects in Photometry

Kepler
A Search for Earth-size Planets



Correcting Systematic Errors

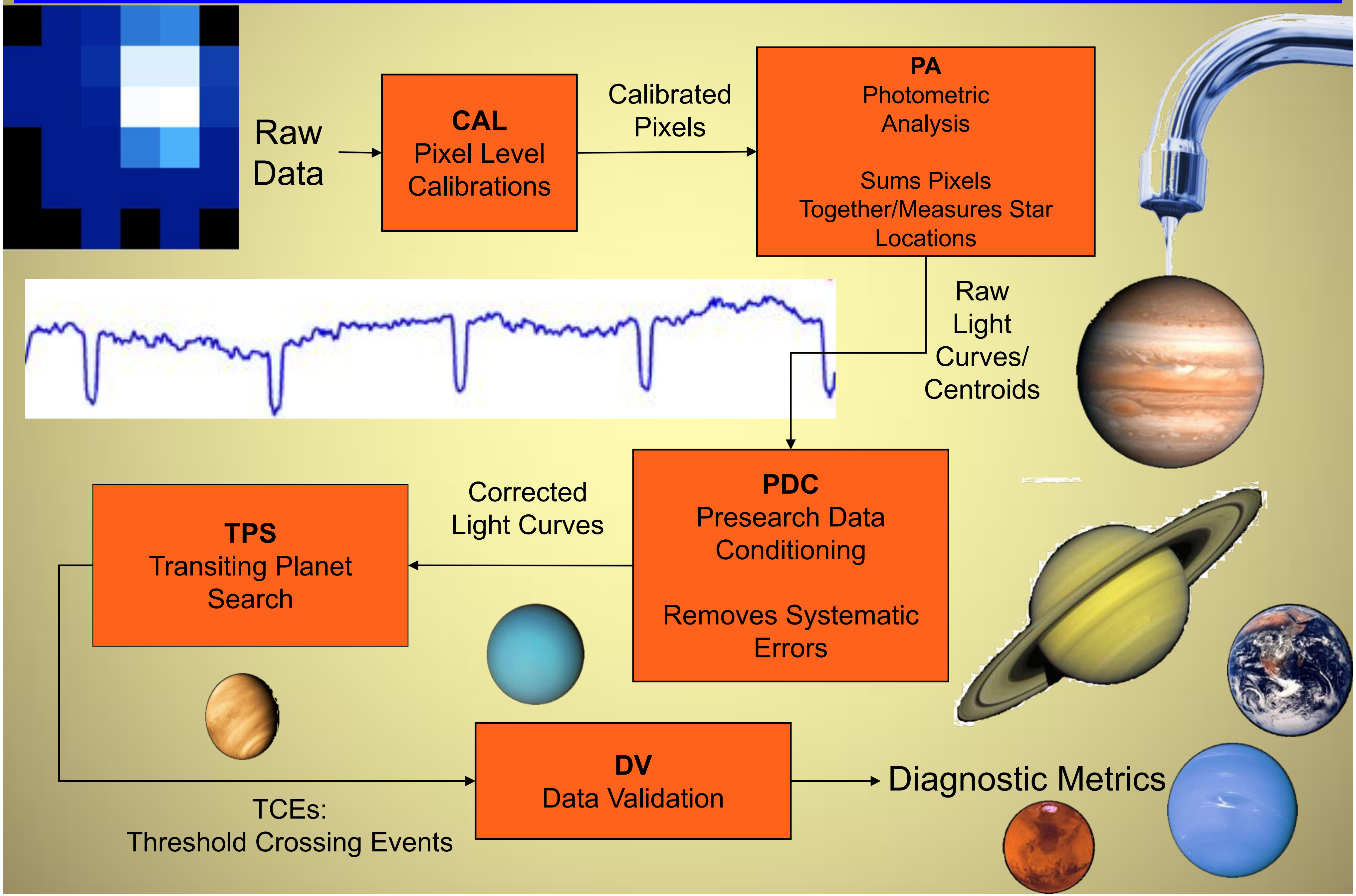




We apply a Maximum A Posteriori approach as per Stumpe et al. 2014

The Kepler Science Pipeline: From Pixels To Planets

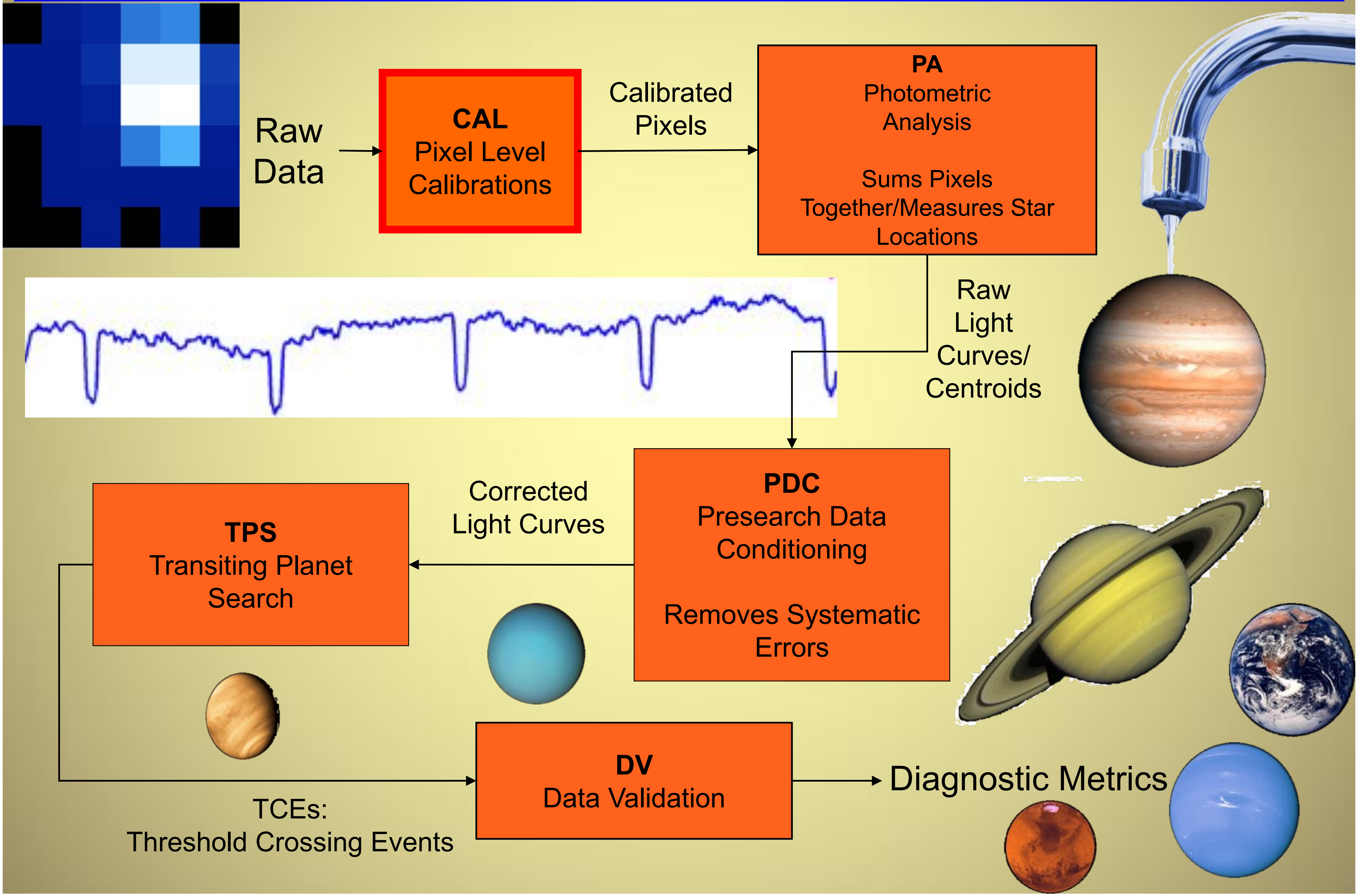
A Search for Earth-size Planets



The Kepler Science Pipeline: From Pixels To Planets

Kepler

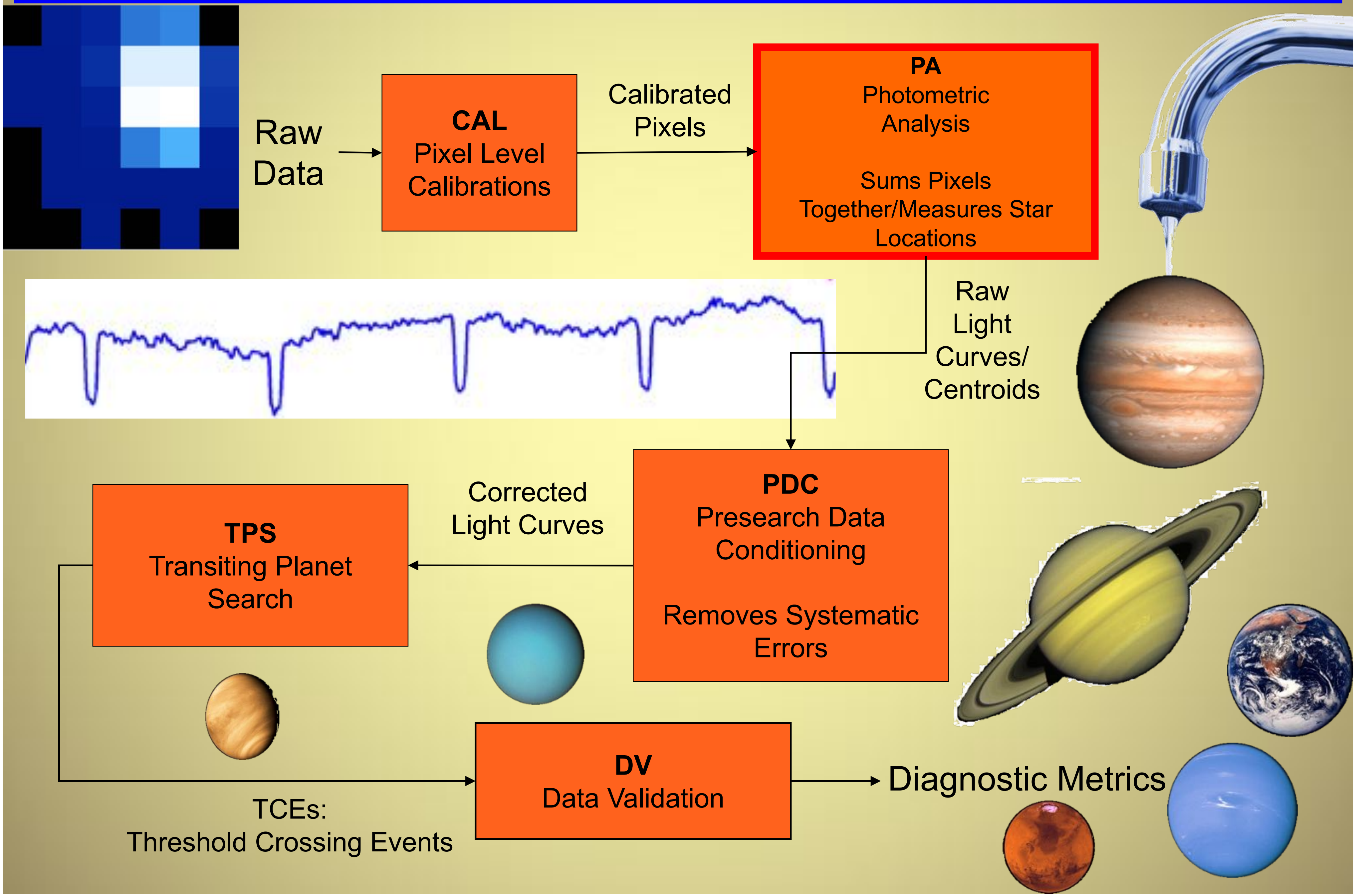
A Search for Earth-size Planets



The Kepler Science Pipeline: From Pixels To Planets

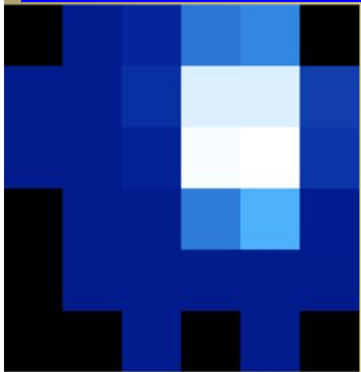
Kepler

A Search for Earth-size Planets



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

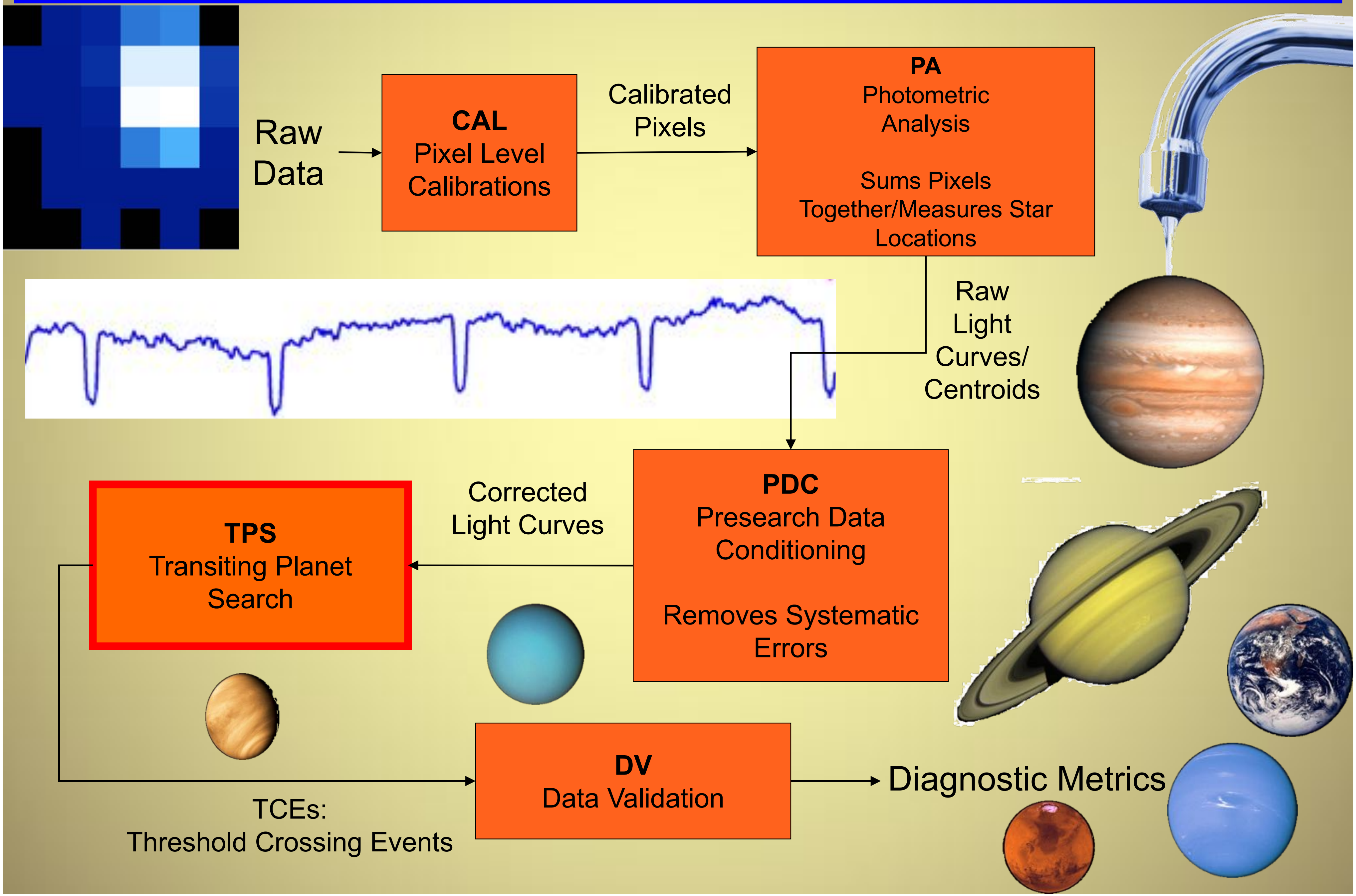
TCEs:
Threshold Crossing Events

DV
Data Validation

Diagnostic Metrics

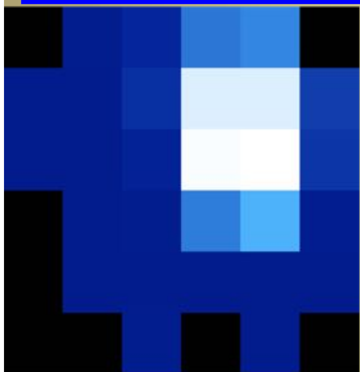
The Kepler Science Pipeline: From Pixels To Planets

Kepler
A Search for Earth-size Planets



The Kepler Science Pipeline: From Pixels To Planets

A Search for Earth-size Planets



Raw Data →

CAL
Pixel Level Calibrations

→ Calibrated Pixels →

PA
Photometric Analysis

Sums Pixels Together/Measures Star Locations

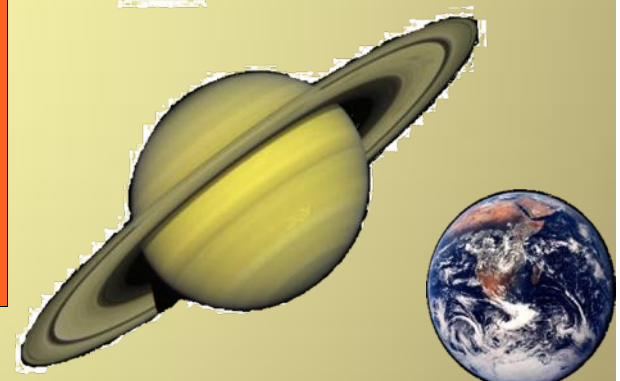
→ Raw Light Curves/ Centroids →

PDC
Presearch Data Conditioning

Removes Systematic Errors

← Corrected Light Curves ←

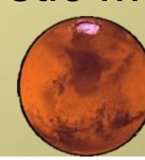
TPS
Transiting Planet Search



→ TCEs: Threshold Crossing Events →

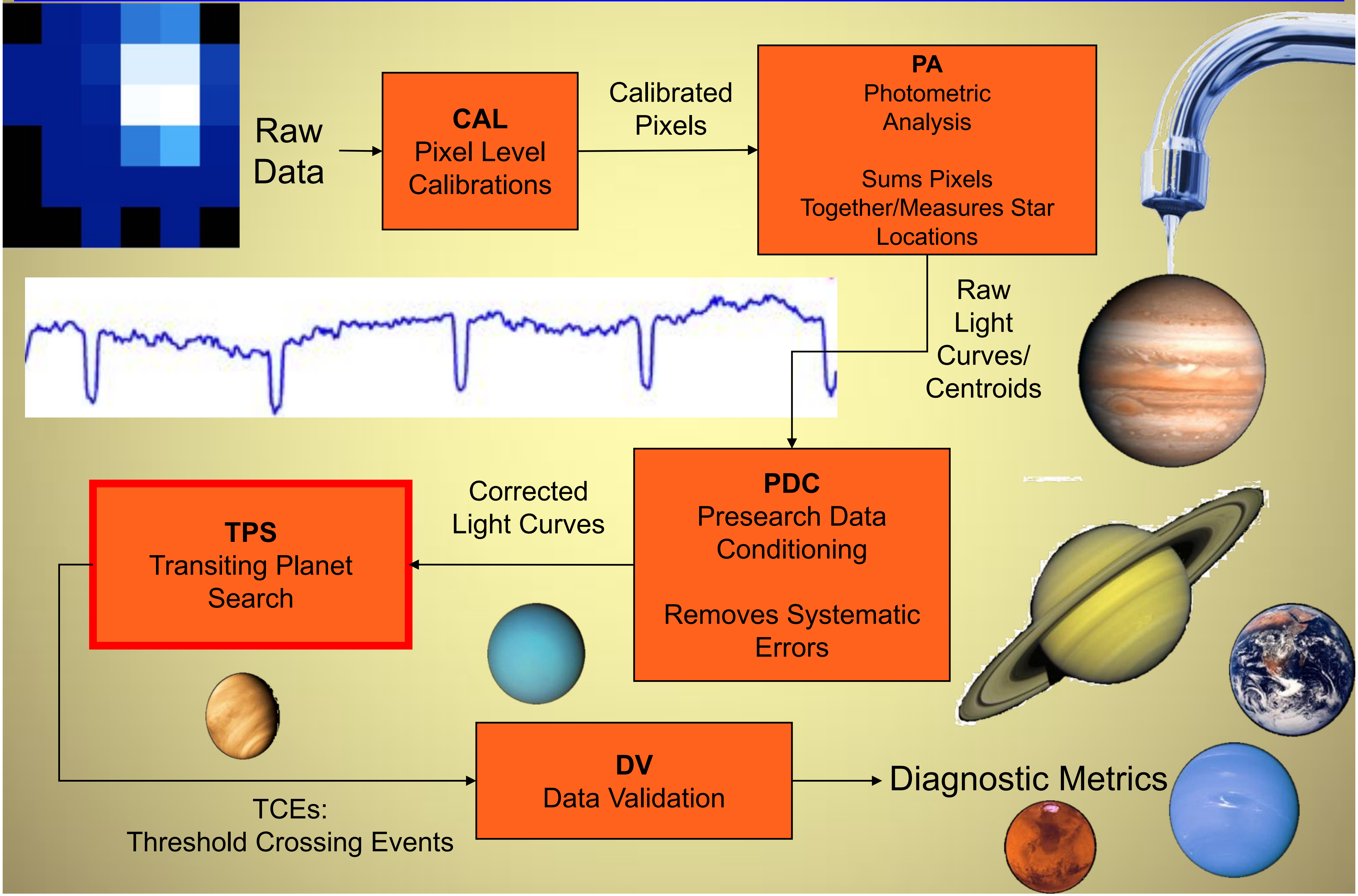
DV
Data Validation

→ Diagnostic Metrics



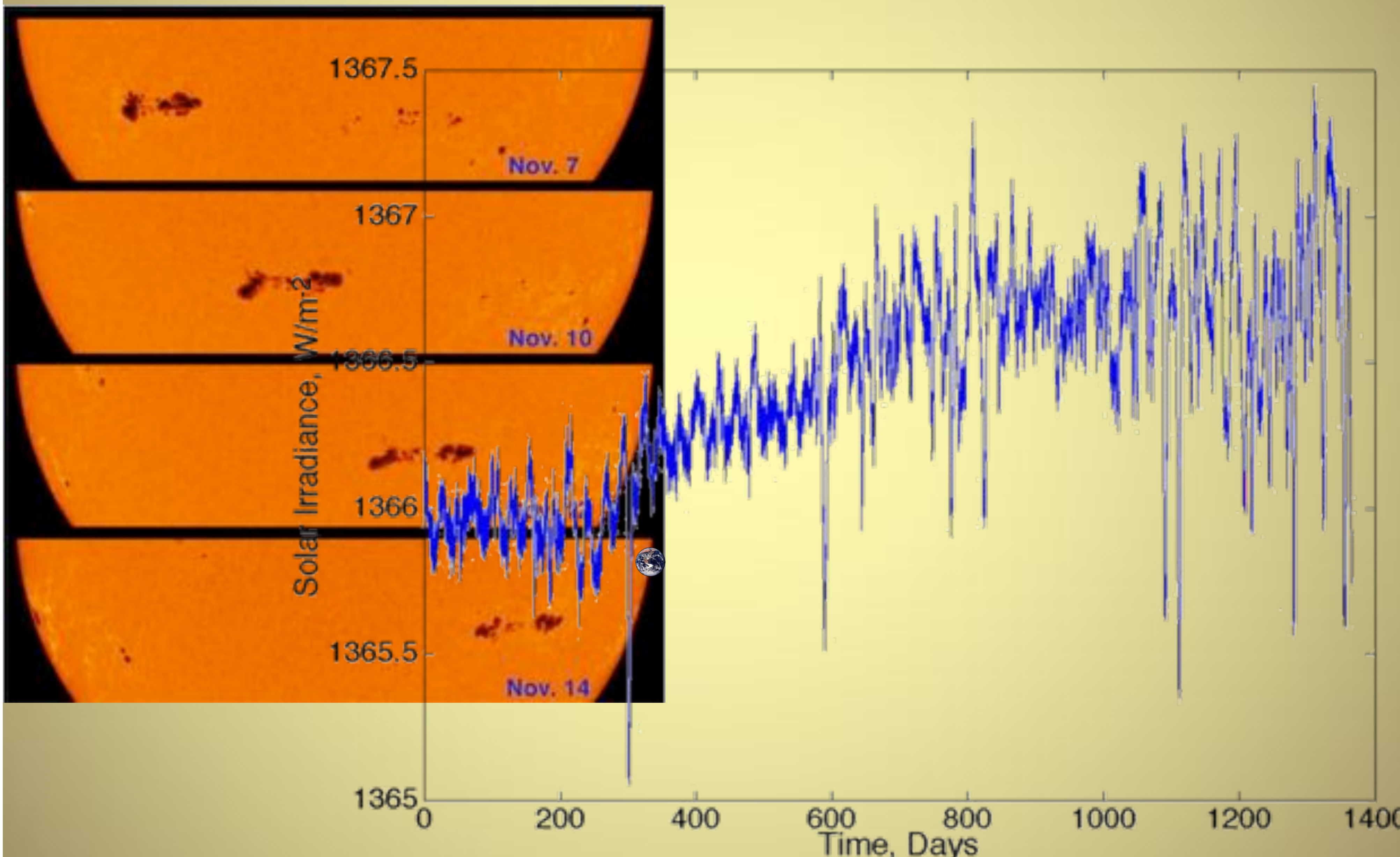
Detecting Transiting Planets

A Search for Earth-size Planets



Solar Variability

A Search for Earth-size Planets



Detecting Deterministic Signals

The problem:

- H0: *x(n)*= *w(n)* or
- H1: *x(n)*= *s(n)* + *w(n)*

s(n) is the signal of interest

x(n) is the time series we observe

w(n) is the observation noise (Gaussian)

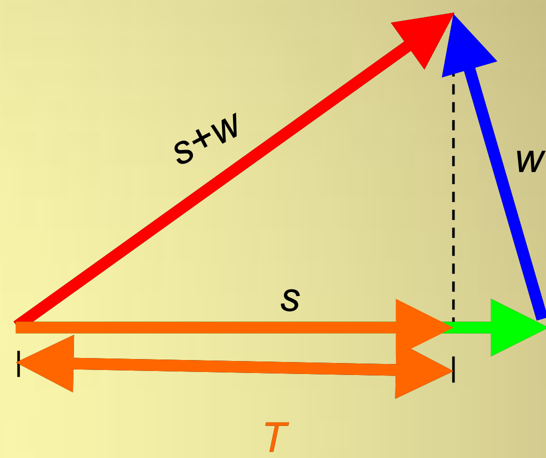
The best method for detecting a known signal in additive Gaussian noise is a matched filter

A matched filter measures the correlation between the data and the signal, normalized by the rms variation of the observation noise

Detection Statistics

A Search for Earth-size Planets

Define $$T = \frac{x^T s}{\sigma_w \sqrt{s^T s}}$$



Under H0: $$\langle T \rangle = 0, \quad \sigma_T^2 = 1$$

Under H1: $$\langle T \rangle = \frac{1}{\sigma_w}\sqrt{s^T s}, \quad \sigma_T^2 = 1$$

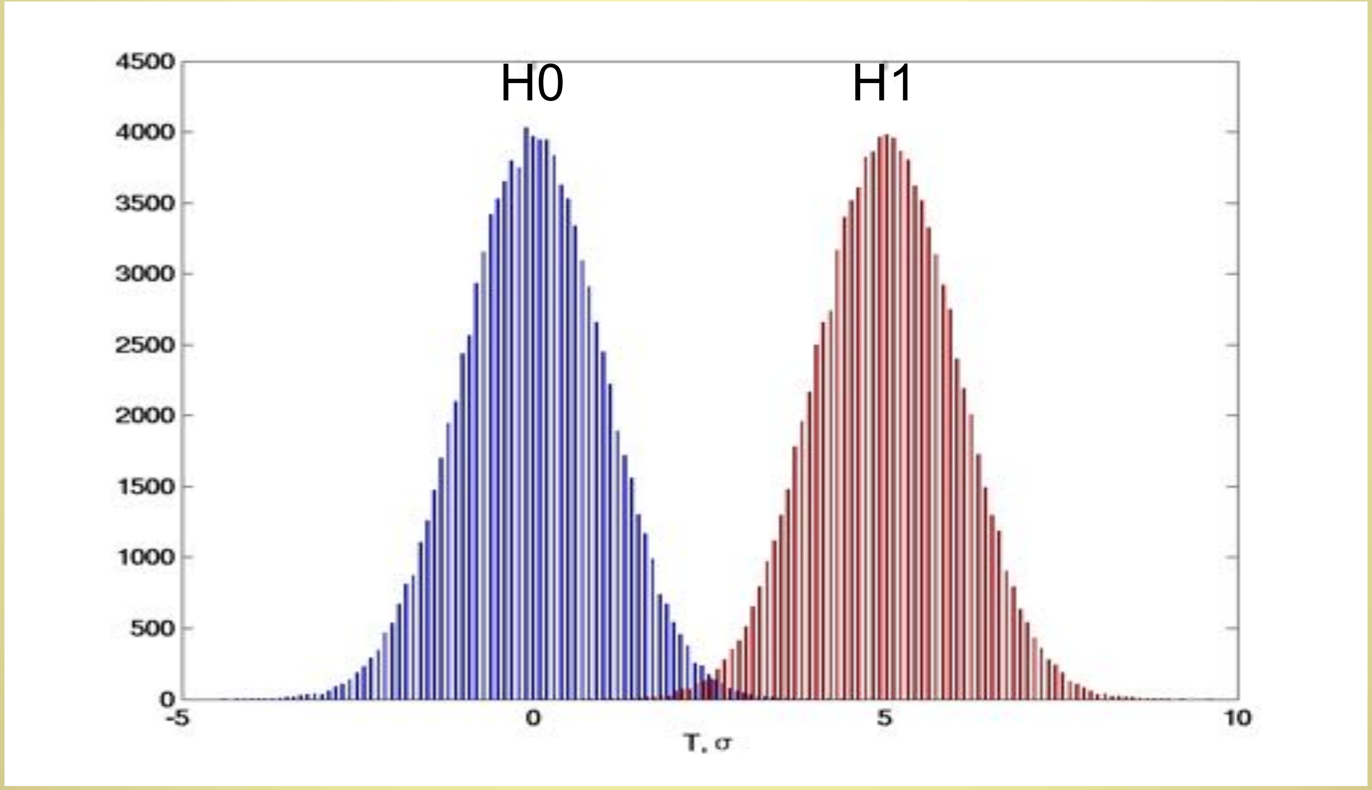
If $T < \gamma$, then choose H0, if $T > \gamma$, then choose H1

Receiver Operating Curves

A Search for Earth-size Planets

T is a Gaussian random variable

$$P_F = \frac{1}{\sqrt{2\pi}} \int_{\gamma}^{\infty} \exp\left(-\tfrac{1}{2} y^2\right) dy \qquad P_D = \frac{1}{\sqrt{2\pi}} \int_{\gamma-\langle T \rangle}^{\infty} \exp\left(-\tfrac{1}{2} y^2\right) dy$$



How do we choose the threshold, γ?

If amplitude of *s* not known, we generally set γ to control P_F.
(Neyman-Pearson Criterion)

Detection Statistics For Colored Noise

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 $$T = \frac{x^T R^{-1} s}{\sqrt{s^T R^{-1} s}} = \frac{(Hx)^T (Hs)}{\sqrt{(Hs)^T (Hs)}} = \frac{\tilde{x}^T \tilde{s}}{\sqrt{\tilde{s}^T \tilde{s}}} > \gamma$$

How do we determine R?

Looks like a simple matched filter!

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

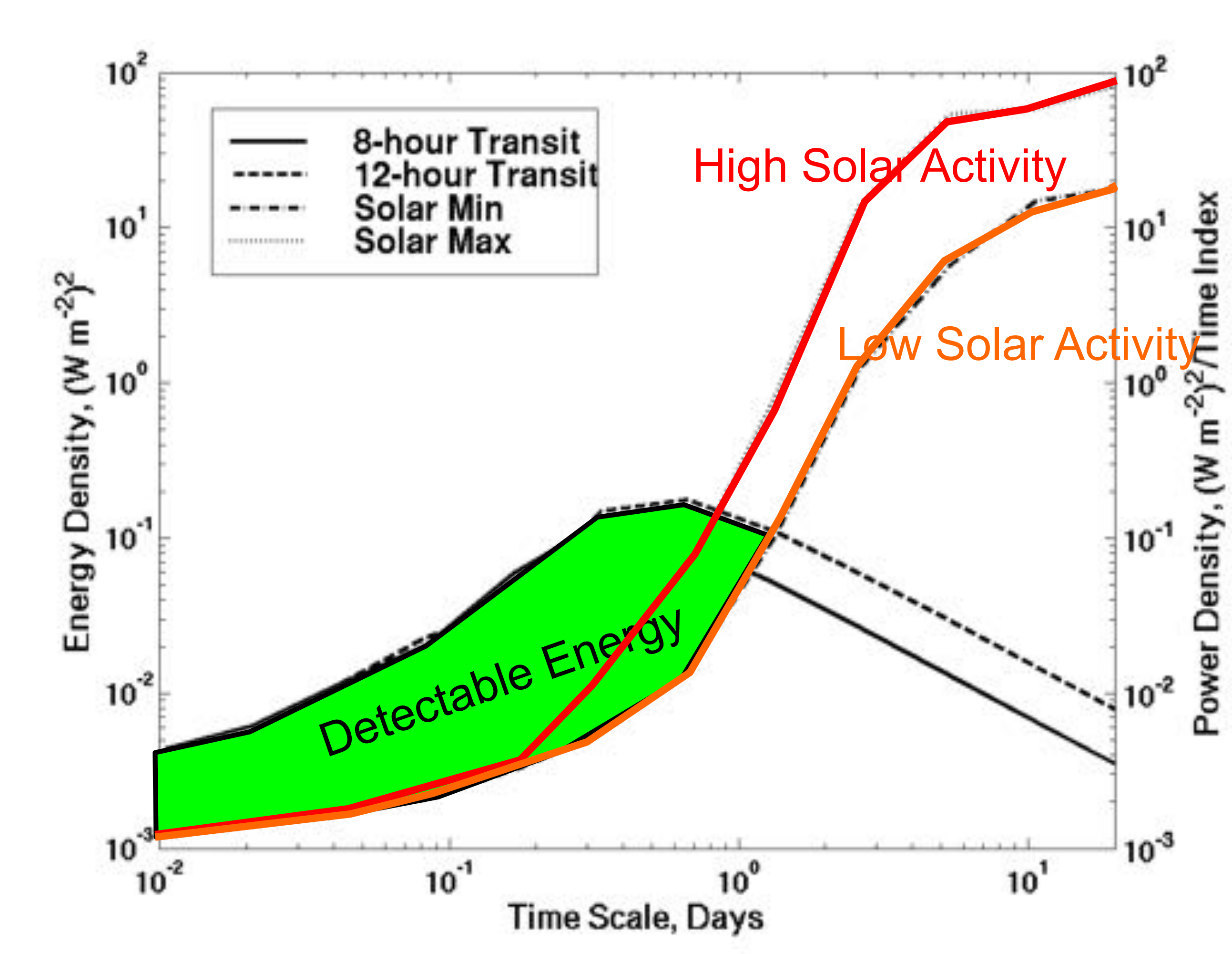
A Search for Earth-size Planets

Is stellar variability stationary?

No!

We must work in a joint time-frequency domain

Wavelets are a natural choice

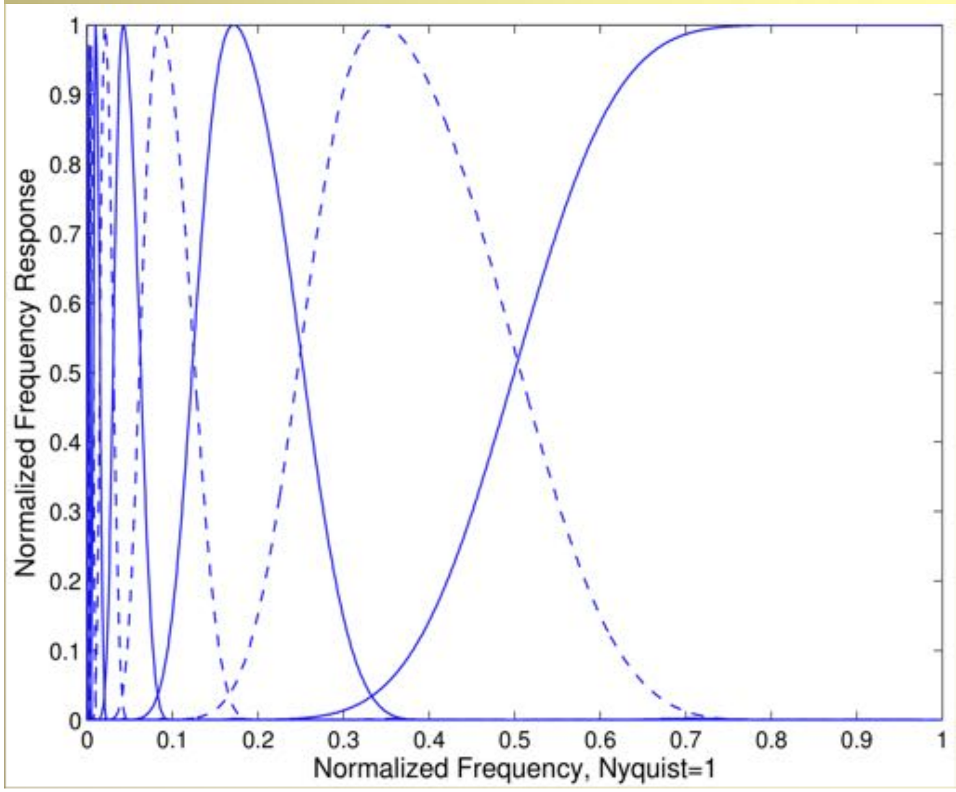


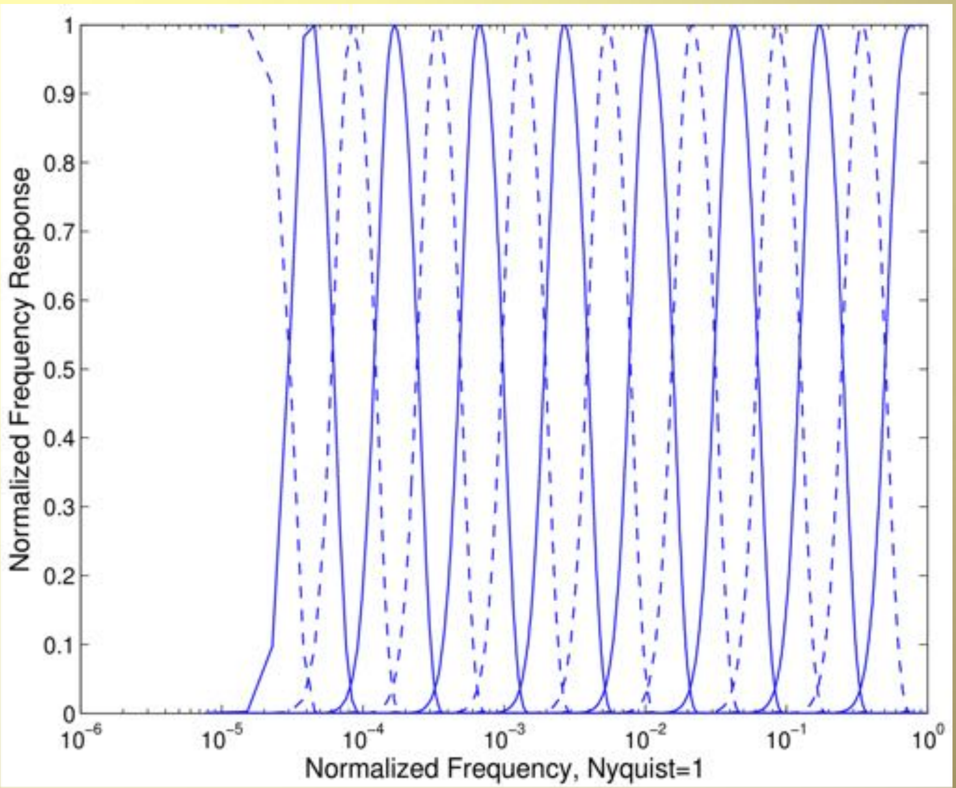
A Wavelet-Based Approach

Filter-Bank Implementation of an Overcomplete Wavelet Transform

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

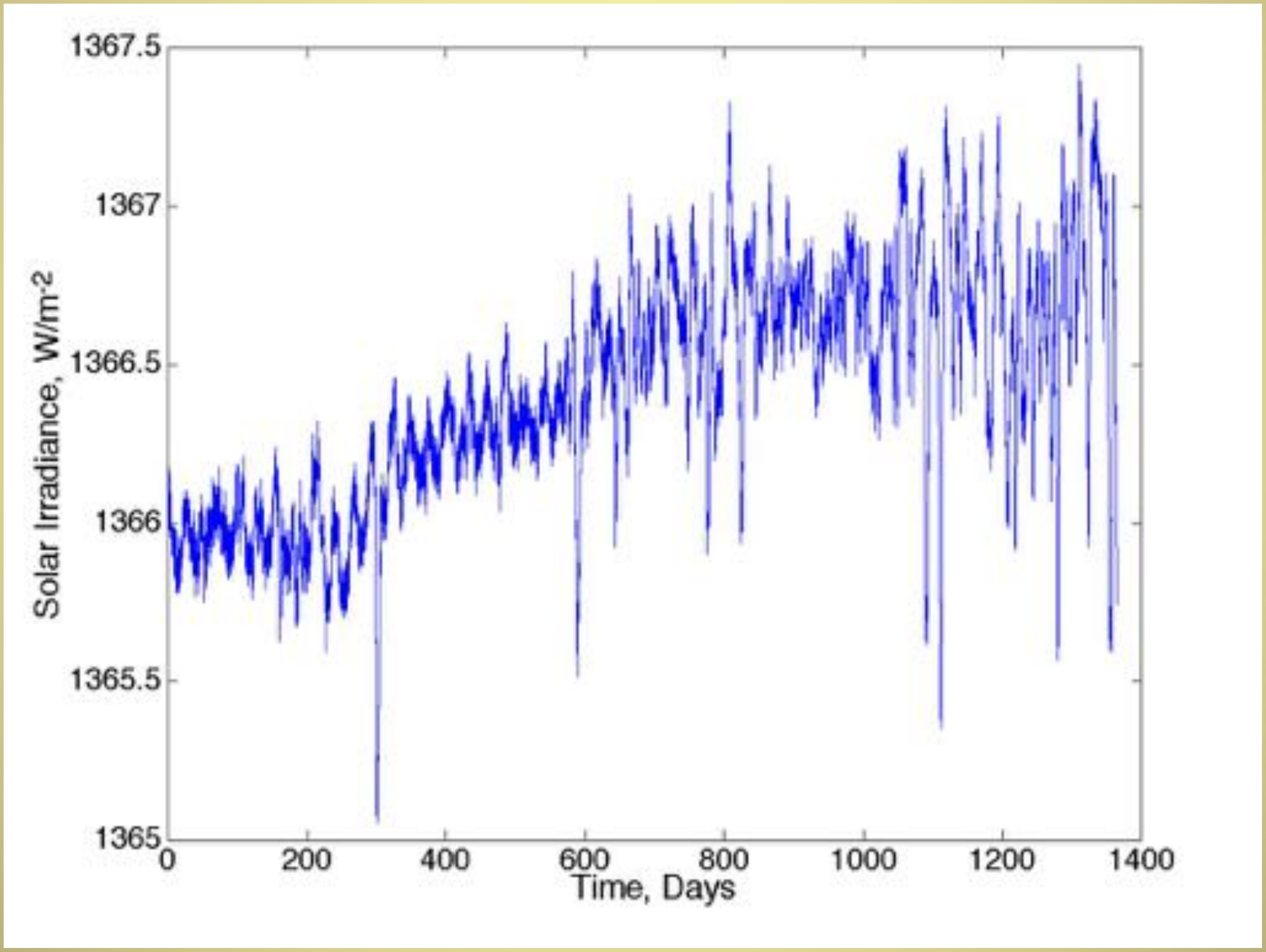
$$W_X(i,n) = \{h_1(n) * x(n), h_2(n) * x(n), \ldots, h_M(n) * x(n)\} = \{x_1(n), x_2(n), \ldots, x_m(n)\}$$





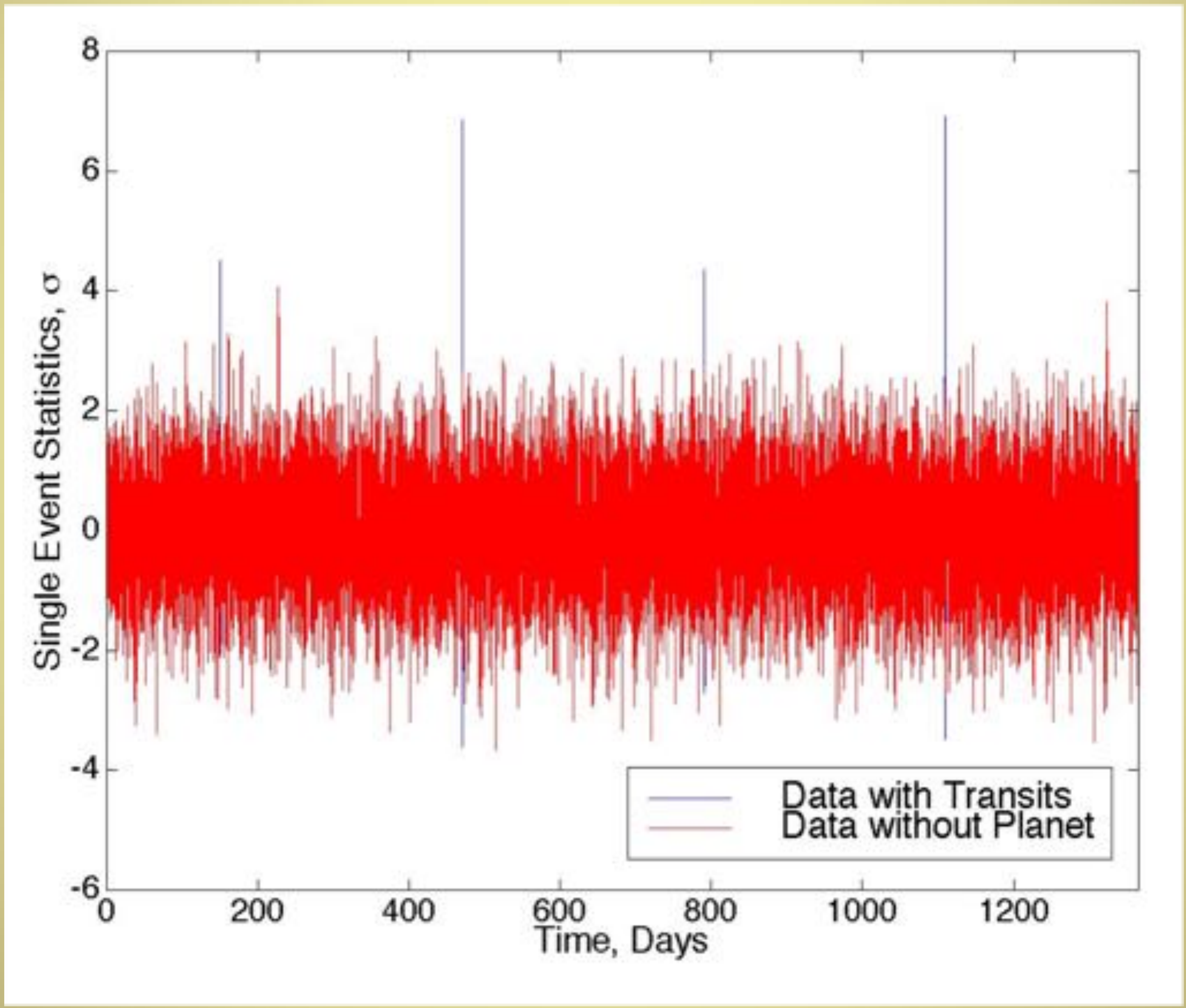
Kepler-like Noise + Transits

Kepler
A Search for Earth-size Planets



Single Transit Statistics

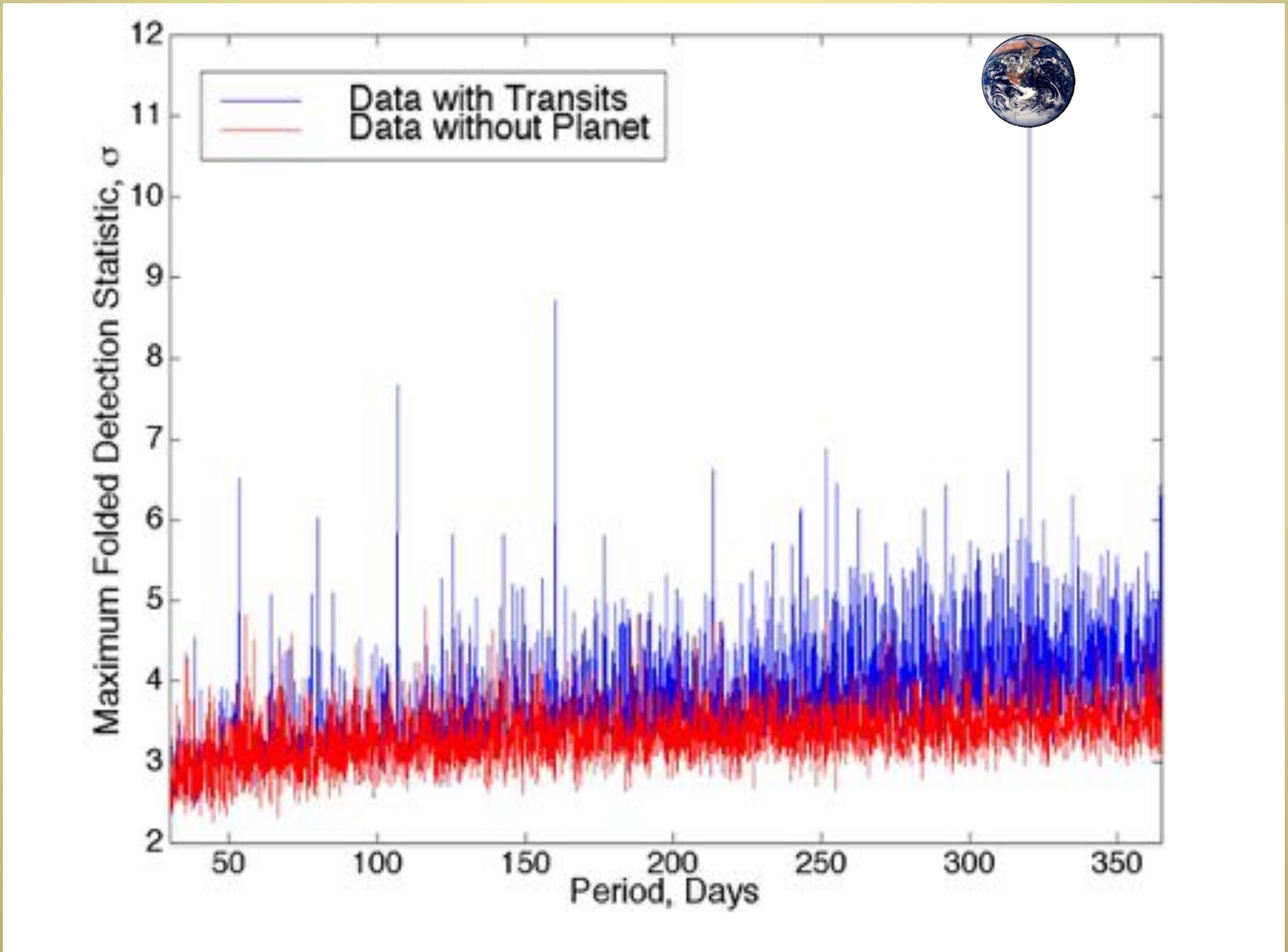
A Search for Earth-size Planets



Folded Transit Statistics

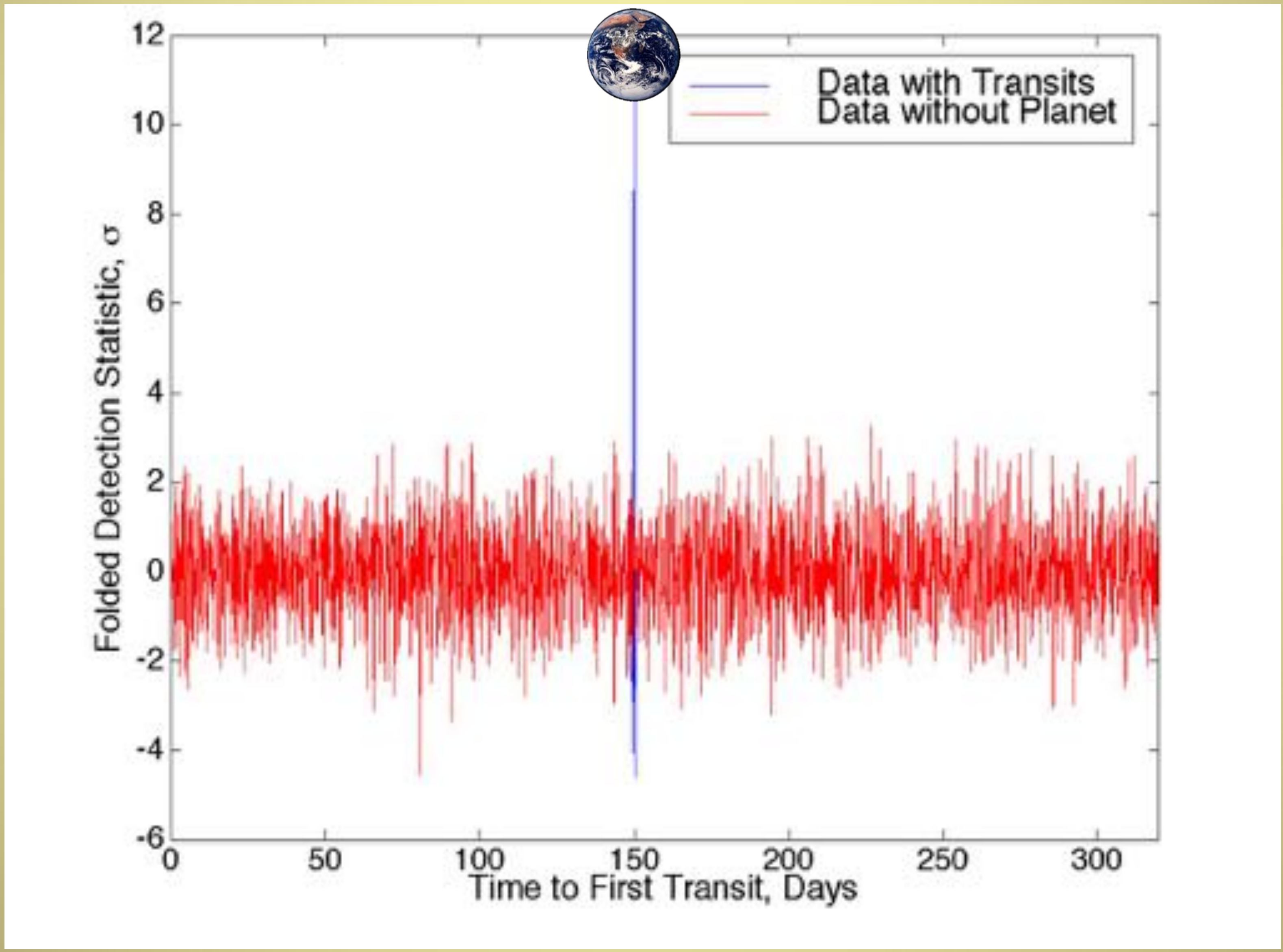
Kepler

A Search for Earth-size Planets



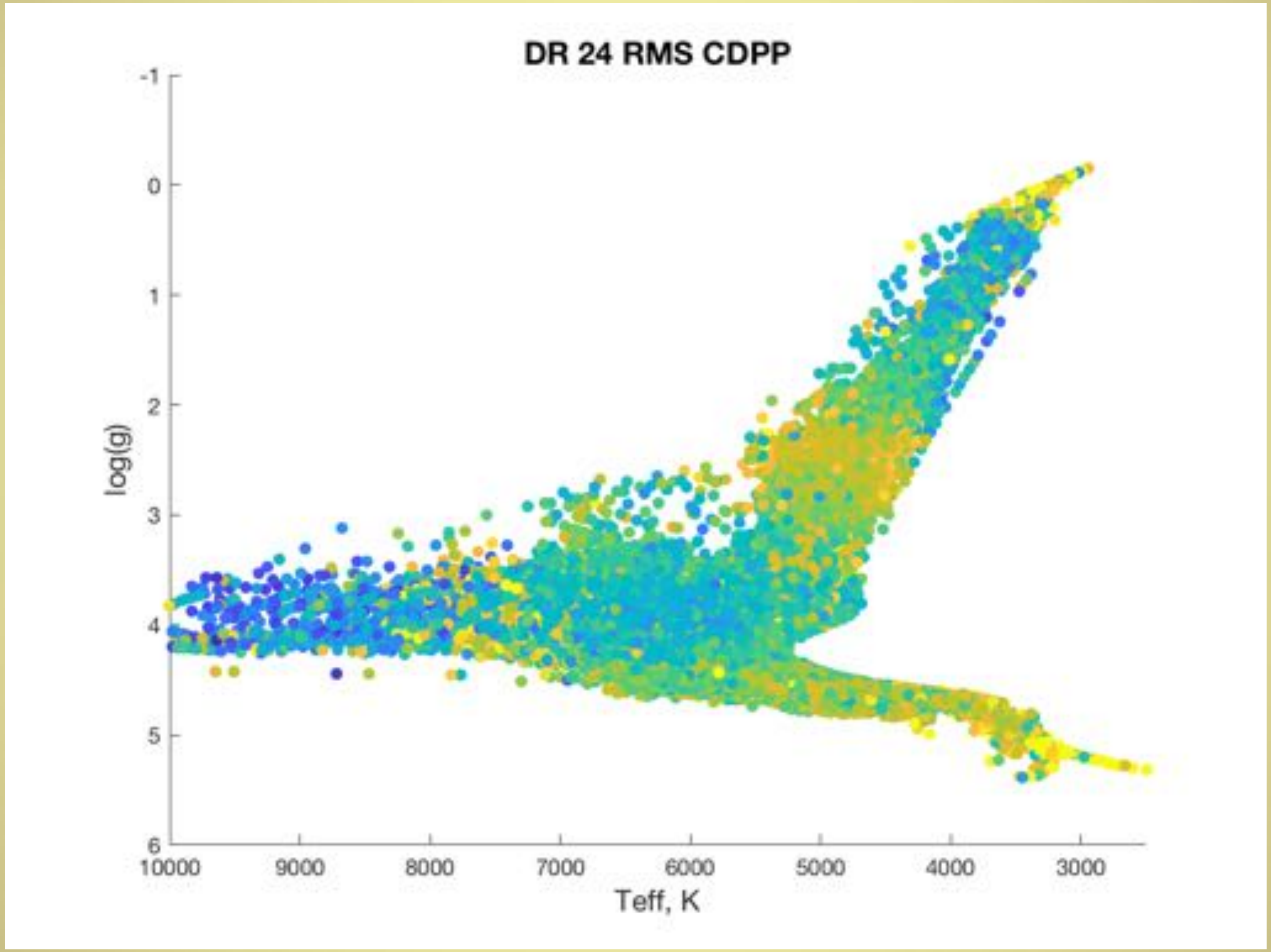
Folded Statistics at Best-Matched Period

A Search for Earth-size Planets



Photometric Precision

A Search for Earth-size Planets



G dwarfs appear to be quiet, and M dwarfs appear to be much noisier

Excess Stellar Variability

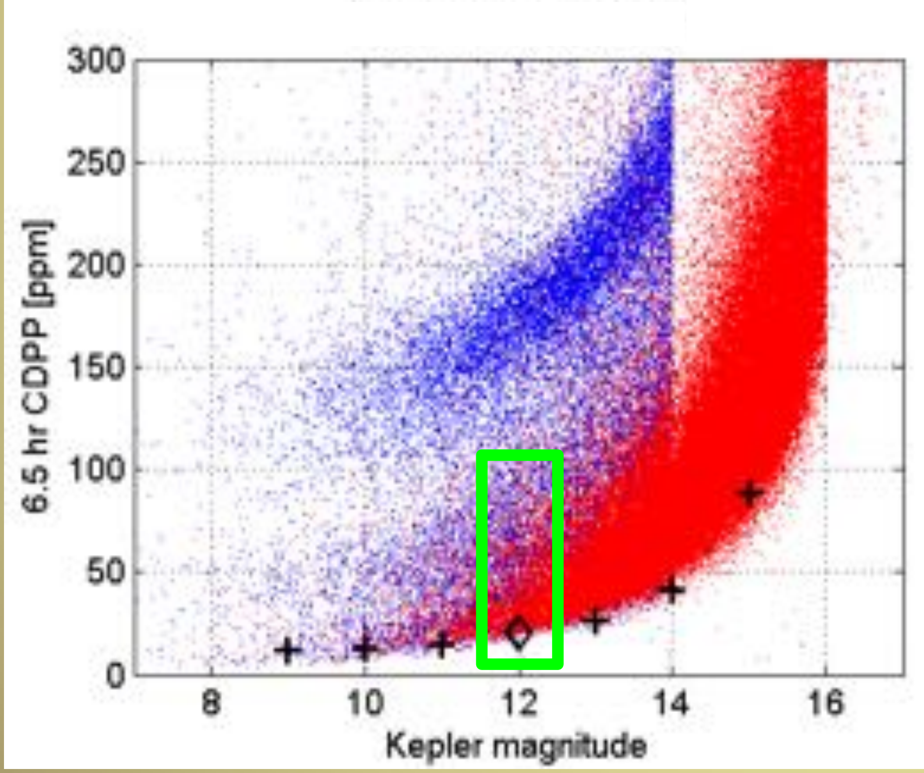
Kepler
A Search for Earth-size Planets

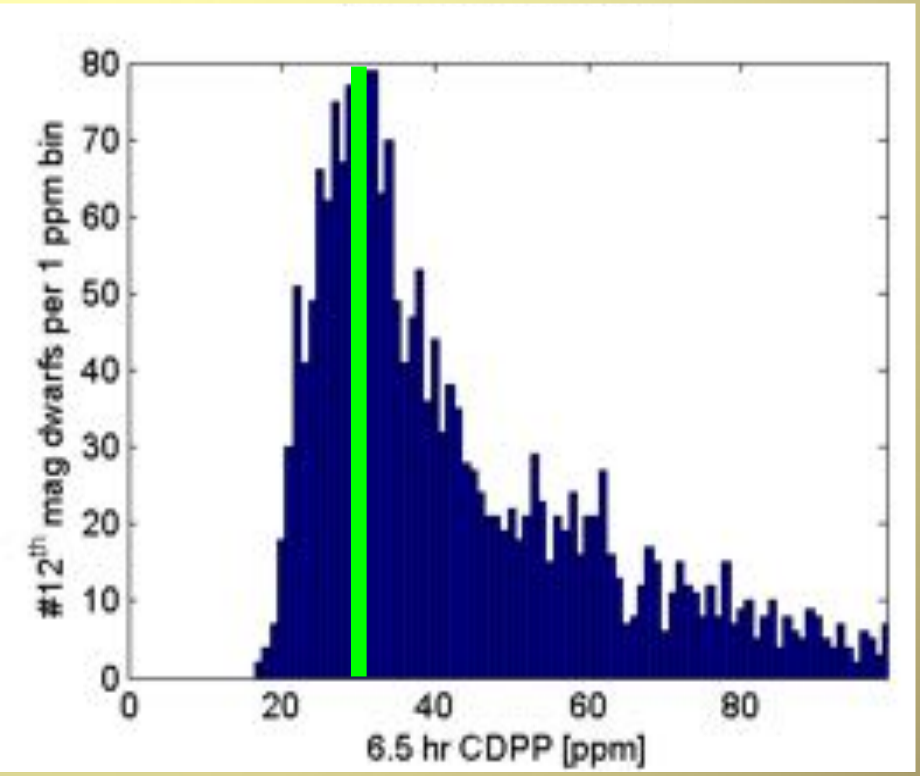
Original Noise Budget (Kp=12):
- 14 ppm Shot Noise
- 10 ppm Instrument Noise
- 10 ppm Stellar Variability

=> 20 ppm Total Noise

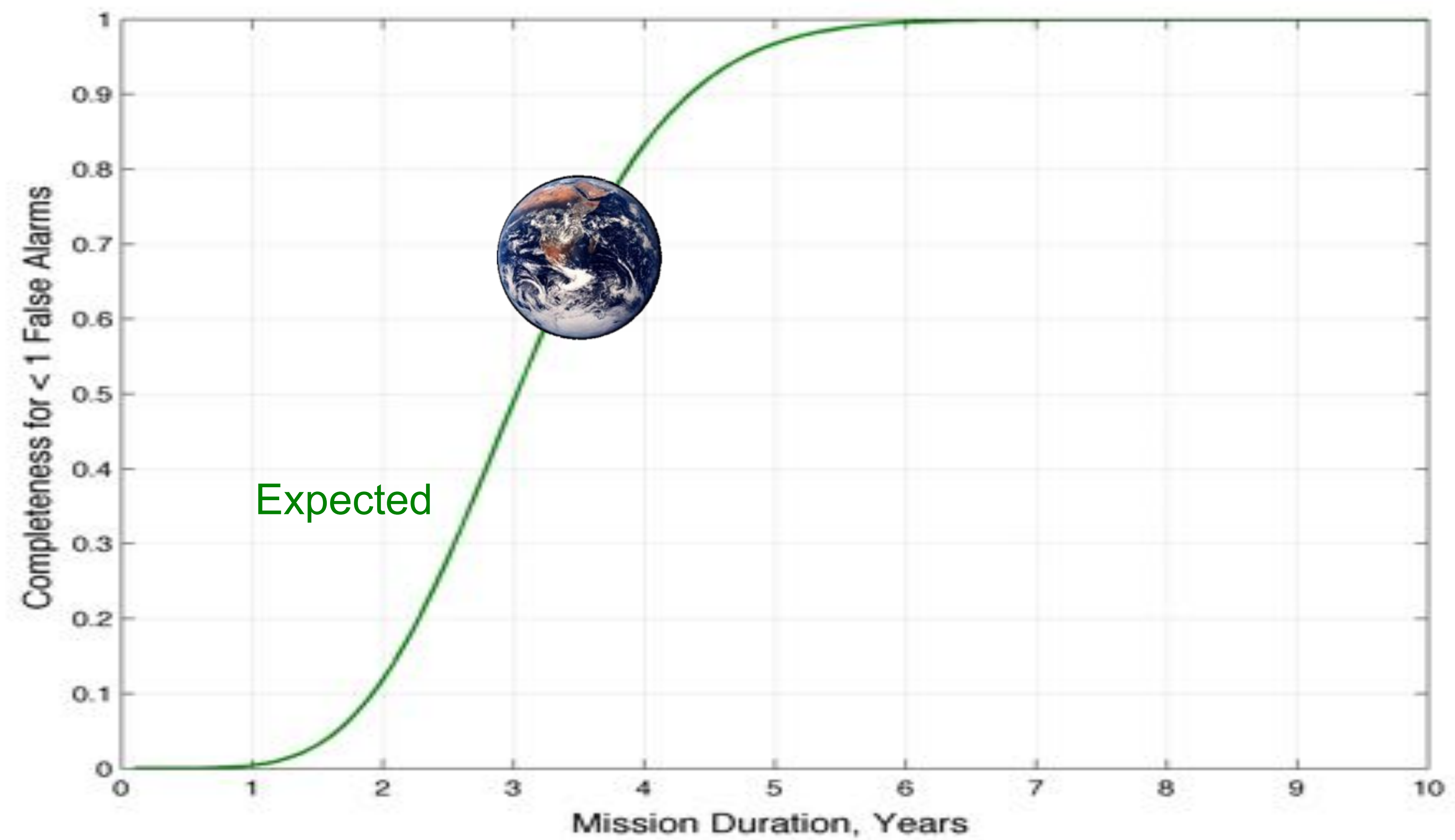
Reality (11.5 ≤ Kp ≤ 12.5)
- 17 ppm Shot Noise
- 13 ppm Instrument Noise
- 20 ppm Stellar Variability

=> ~29 ppm Total Noise





Completeness Vs. Time



Original expectations yielded ~70% completeness for Earth analogs at 3.5 years

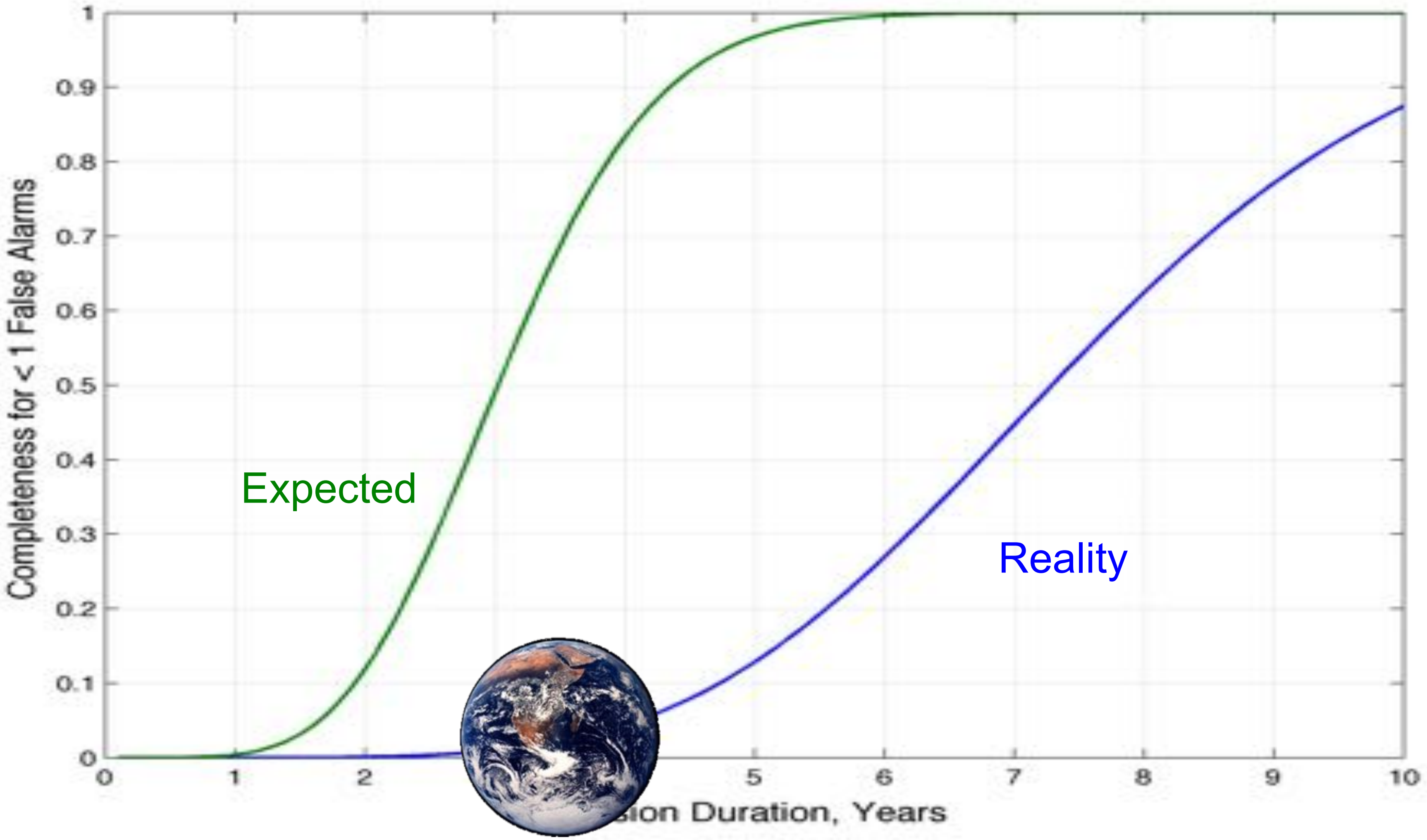
Completeness Vs. Time

A Search for Earth-size Planets

Current expectations yield <5% completeness for Earth analogs at 3.5 years

Completeness Vs. Time

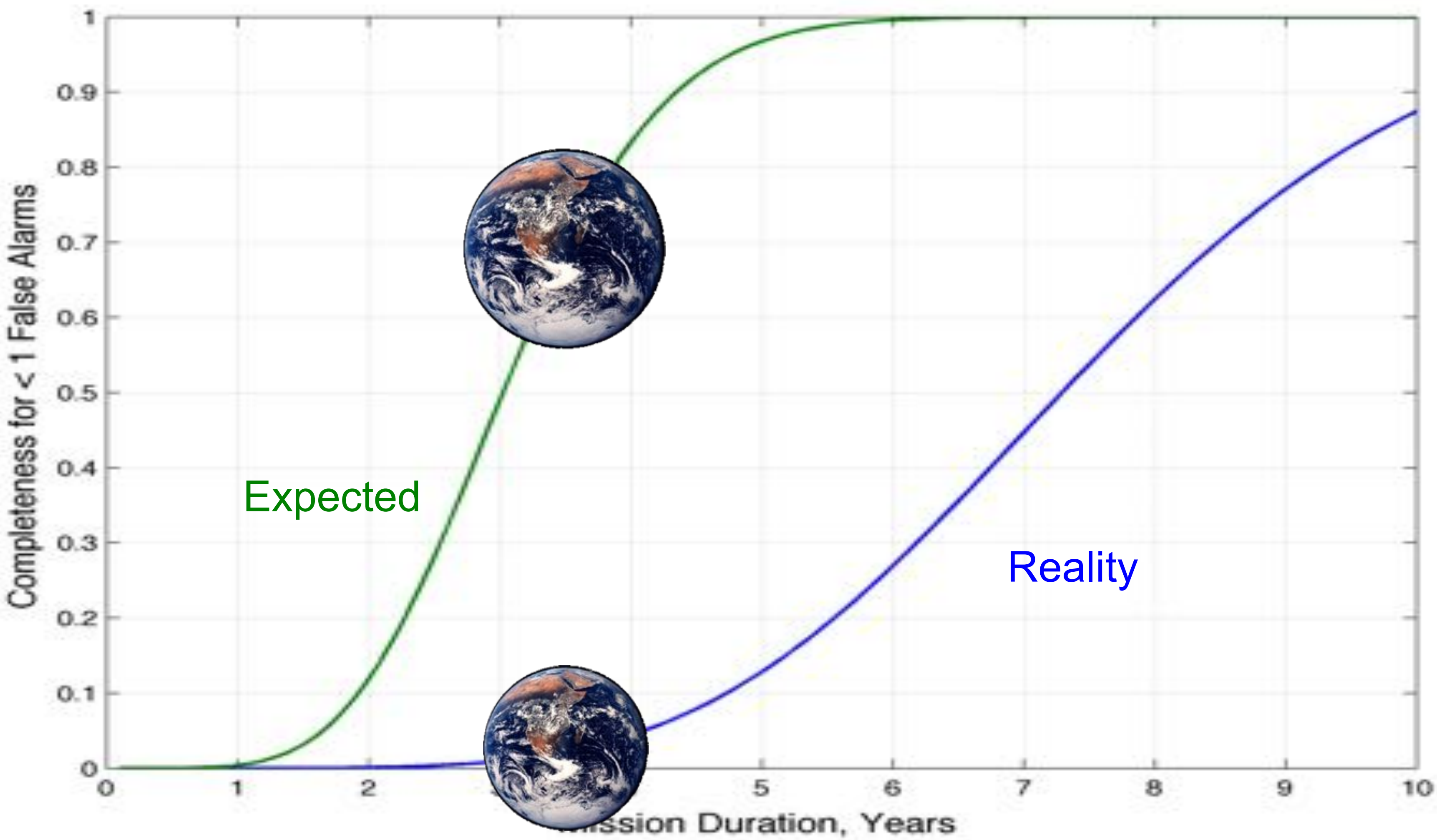
A Search for Earth-size Planets



~70% completeness for 1.2-R_e planets in same orbits at 3.5 years

Completeness Vs. Time

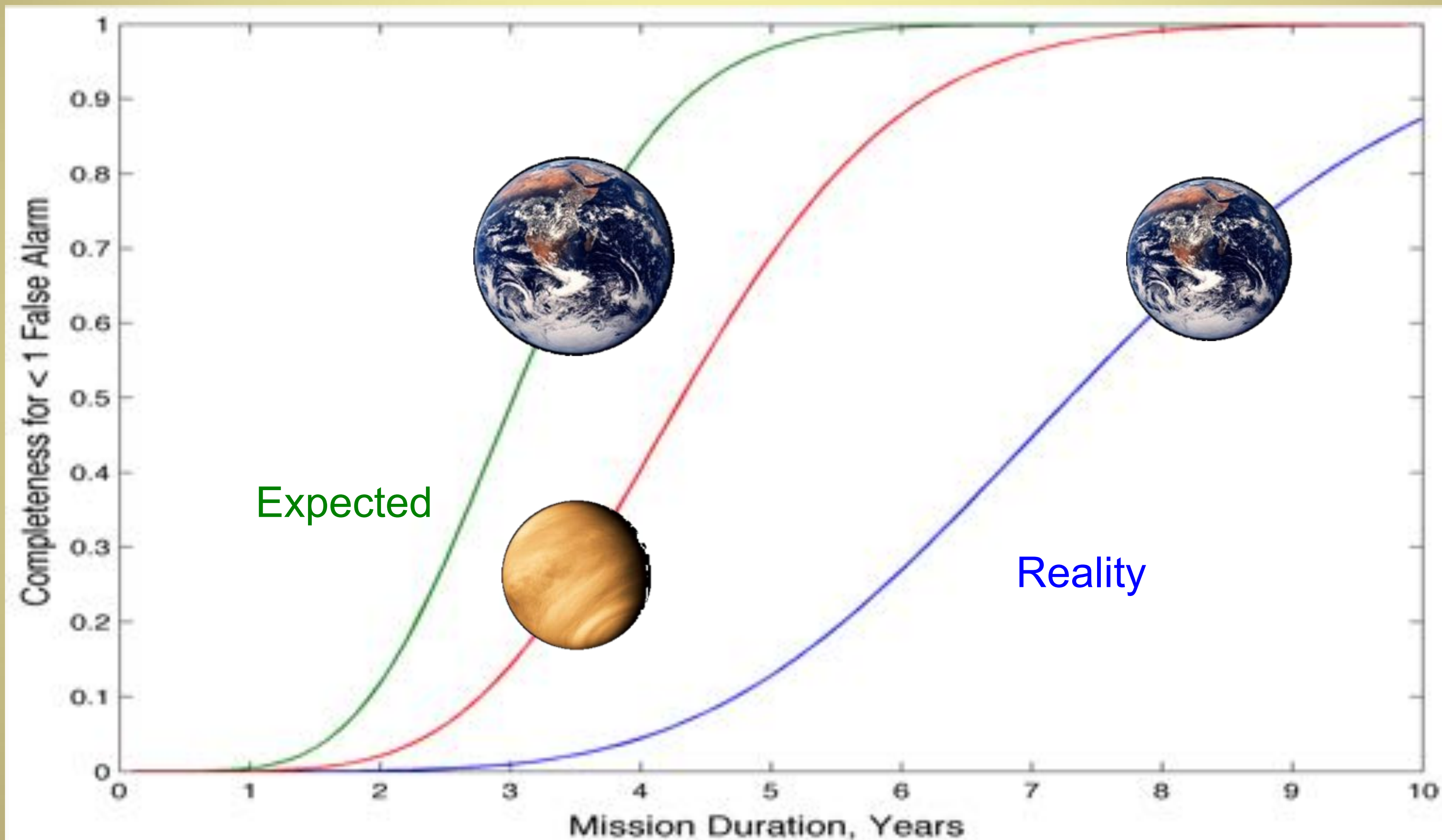
A Search for Earth-size Planets



Kepler will recover ~70% completeness for Earth analogs after 8 years

Completeness Vs. Time

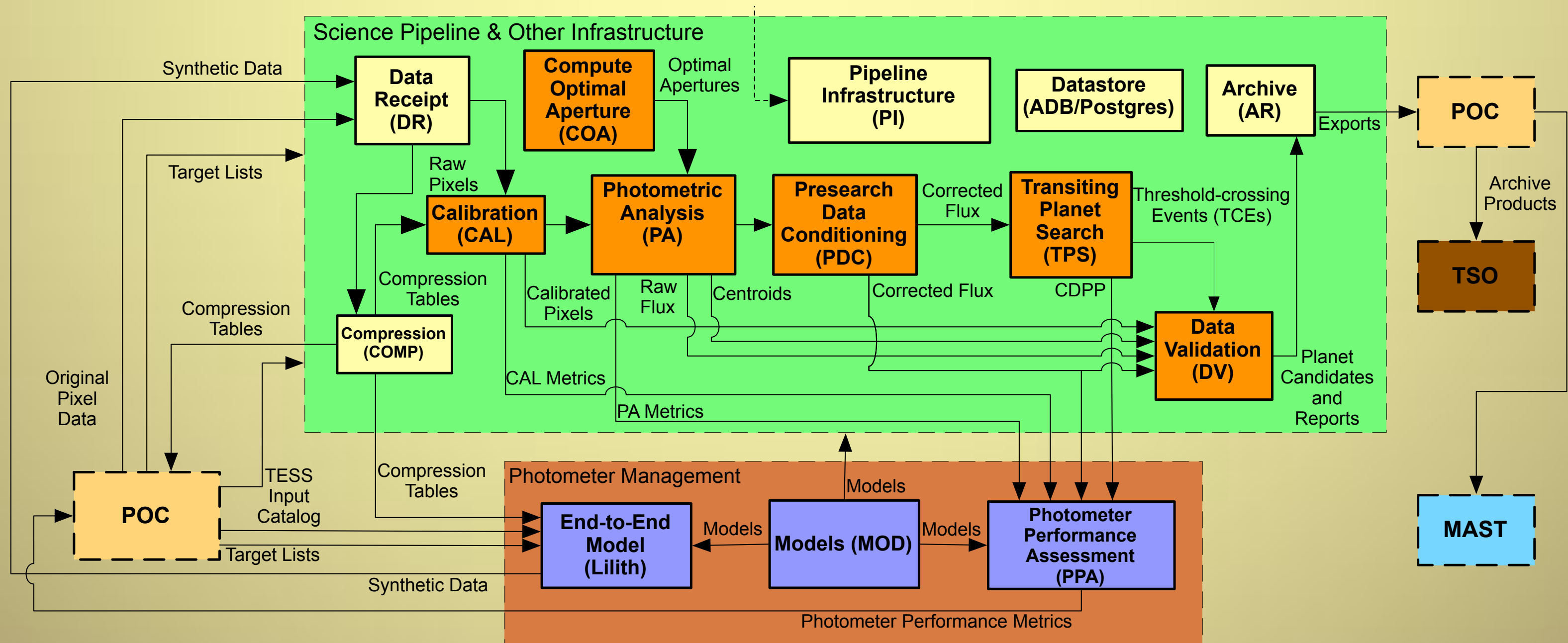
A Search for Earth-size Planets



Kepler will detect virtually all Venus analogs within 8 years

Developing the TESS Pipeline

- ~13X pixel data rate over Kepler
- Leveraged heritage from Kepler pipeline
- Significantly lower cost (~46 FTEs over project lifetime)
- Significant speed improvements:
    - Colocated servers and storage with NAS Pleiades supercomputer
    - Moved pixel-level calibrations to C++
    - Sped up Presearch Data Conditioning by 10X
    - Originally projected 20+ days to process one sector
    - Complete pipeline requires ~5 days to process one sector



Summary

Kepler
A Search for Earth-size Planets

- Stellar variability presents a fundamental limit on the detectability of transiting Earth-like planets
- Adaptive matched filters can provide near-optimal detection of Earth-size transits and characterize the observation noise
- Larger than expected stellar variability can be compensated for by increasing the duration of the campaign
- Controlling instrumental noise and systematics is also very important as shot noise, instrument noise and stellar variability should be comparable in a well designed mission