Planet Detection: The *Kepler* Mission

Jon M. Jenkins
SETI Institute/NASA Ames Research Center

Friday February 25, 2011

GISS CESS 2011, NYC
Outline

• 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
A Search for Earth-size Planets

The Kepler spacecraft surveys a part of the Orion arm of our Milky Way galaxy. This neighborhood of the galaxy has enough potential targets and is far enough from the ecliptic so as not to be obscured by the Sun.

Mission Overview

Goal: to determine the frequency of Earth-size planets in Earth-like orbits about sun-sized stars

Borucki et al. 2010

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

• 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

Transit Depth: 0.00015

\( R = 1.4 \text{ R}_{\text{Earth}} \)

Period = 0.83 days
Kepler Candidates as of June 2010

- Size Relative to Earth
- Orbital Period in days

The graph shows the relationship between size and orbital period for Kepler candidates.
Kepler Candidates as of February 1, 2011
Kepler Candidates as of February 1, 2011

The graph shows the relationship between the size of potential planets relative to Earth and their candidate equilibrium temperature in °F. There are two distinct clusters of data points, with one in blue and the other in yellow. The size is represented along the vertical axis, while the temperature is along the horizontal axis. The graph includes a scale for both axes, with numbers ranging from 0 to 4000 for the temperature and 1 to 20 for the size.
The Kepler Science Pipeline: From Pixels To Planets

- **Raw Data**
  - **CAL**: Pixel Level Calibrations
    - Sums Pixels Together/Measures Star Locations

- **TPS**: Transiting Planet Search
  - **PDC**: Presearch Data Conditioning
    - Removes Systematic Errors
  - **DV**: Data Validation
    - Diagnostic Metrics

- **PA**: Photometric Analysis
  - Corrected Light Curves/Centroids
  - Raw Light Curves/Centroids

**TCEs**: Threshold Crossing Events
Pixel-Level Data From Kepler

Module 17 Output 2

Zoomed image near planet HAT-P-7b

0.09x0.09 degrees
80x80 pixels
6400 pixels total

1.13 (h) x 1.22 (w) degrees

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

Scaled to show faint detail
The Kepler Science Pipeline: From Pixels To Planets

- Raw Data → CAL (Pixel Level Calibrations) → Calibrated Pixels → PA (Photometric Analysis)
  - Sums Pixels Together/Measures Star Locations
  - Raw Light Curves/Centroids

- Raw Data → TPS (Transiting Planet Search) → Corrected Light Curves → PDC (Presearch Data Conditioning)
  - Removes Systematic Errors

- TCEs (Threshold Crossing Events) → DV (Data Validation) → Diagnostic Metrics
Solar Variability

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:

$$T = \left[ \int \frac{X(f)S^*(f)}{P(f)} df \right] \sqrt{\int \frac{S(f)S^*(f)}{P(f)} df}$$
PSDs for Solar-Like Variability

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

$$W_X(i,n) = \{h_1(n) \ast x(n), h_2(n) \ast x(n), \ldots, h_M(n) \ast x(n)\}$$

$$= \{x_1(n), x_2(n), \ldots, x_m(n)\}$$
Kepler-like Noise + Transits
Single Transit Statistics

Data with Transits
Data without Planet
A Search for Earth-size Planets

Folded Statistics at Best – Matched Period

![Graph showing Folded Detection Statistic vs. Time to First Transit, Days. The graph compares Data with Transits and Data without Planet.](image)
Analyzing a Light Curve

Flux Time Series

Correlation Time Series

Normalization Time Series

CDPP Time Series

Single Event Statistics Series

Time, Days

A Search for Earth-size Planets
Folding Detection Statistics

A Search for Earth-size Planets
The Kepler Science Pipeline: From Pixels To Planets

**CAL**
Pixel-level calibrations
Calibrated Pixels

**PA**
Photometric Analysis
Sums pixels/measures star locations

**TPS**
Transiting Planet Search

**PDC**
Presearch Data Conditioning
Removes systematic errors

**DV**
Data Validation

**Raw Data** → **CAL** → **PA** → **TPS** → **PDC** → **DV** → **Diagnostic Metrics**

**TCEs:** Threshold Crossing Events

**Raw Light Curves/Centroids**
PDC Often Does a Good Job

A Search for Earth-size Planets
A star fitted with systematics (green) extracted by SVD from 1400 “quiet” stars from Q1 on channel 2.1.
PDC Is Fundamentally Flawed

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

\[ \hat{y} = H\theta \]

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

For a Gaussian prior:

$$\hat{\theta}_{\text{MAP}} = \left(H^T C_w^{-1} H + C_\theta^{-1}\right)^{-1} \left(H^T C_w^{-1} y + C_\theta^{-1} \mu_\theta\right)$$

For uniform white noise

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

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

While as $\sigma^2 \rightarrow \infty$, $\hat{\theta}_{\text{MAP}} \rightarrow \left(C_\theta^{-1}\right)^{-1} \left(C_\theta^{-1} \mu_\theta\right) = \mu_\theta$
A Search for Earth-size Planets

SVD decomposition of Quiet Stars

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
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

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