TRANSIT DETECTION
in the presence of stellar noise

SUZANNE AIGRAIN (Oxford)
Steve Roberts, Hannu Parviainen, Vinesh Rajpaul, Ben Pope
TIMESCALES

TOTAL SOLAR IRRADIANCE (TSI) POWER SPECTRUM

LOG POWER DENSITY

ACTIVE REGIONS

GRANULATION

OSCILLATION

PLANETARY ORBITS

TRANSITS

YEARS WEEKS DAYS HOURS MINUTES

S

LOG FREQUENCY
THE TRADITIONAL WAY

1. Correct Instrumental Effects
2. Filter Stellar Variability
3. Detect Transits
CoRoT EXAMPLE

ITERATIVE NON-LINEAR FILTER (Aigrain & Irwin 2004)
CoRoT EXAMPLE

BOX-LEAST SQUARES TRANSIT SEARCH (Kovacs, Zucker & Mazeh 2002)

\[ P \sim 20h, \]
\[ \text{depth } 0.0003, \]
\[ R_{\text{planet}} \sim 2 R_{\text{Earth}} \]
\[ \text{CoRoT-7b} \]
\[ (\text{Leger+2009}) \]
Sun-like activity doesn’t matter for warm Neptunes, hot Super-Earths, etc…

IT DOES FOR TEMPERATE EARTHS
NEITHER WHITE NOR STATIONARY

FREQUENCY CONTENT of transits vs Total Solar Irradiance (TSI) variations

Jenkins (2002)
CHANGE IN SINGLE-TRANSIT SNR over solar activity cycle

Jenkins (2002)

ADAPTIVE WAVELET-BASED MATCHED FILTER TRANSIT
SETTING THE THRESHOLD

WANT <= 1 FALSE ALARMS over entire mission (1/600,000):

- threshold MES = 7.1
- expected sensitivity
  - ~80% for MES = 7.1
  - ~84% for MES = 8

KEPLER CATALOG GENERATION

- run transit search
- record all threshold crossing events (TCEs) - 100000’s!!!
- vet them to weed out astrophysical and instrumental false alarms
- community follow-up and/or statistical validation
IS THE SUN TYPICAL?

Most Sun-like stars are MORE VARIABLE than the Sun.

RESIDUAL NOISE (not from known instrumental sources) ON TRANSIT TIMESCALES for bright Kepler main sequence stars (Gilliland+2011)

THE SUN
**KEPLER TRANSIT INJECTION TESTS**

Christiansen+(2013,2015a,b,2016,2017)

where $p$ is the probability of detection, $\Gamma$ is the gamma function, $x$ is the expected MES, and $c$ is a scaling factor, for $\text{MES} \leq 15$.

A fit of this function to the histograms gives coefficients $a = 30.87$, $b = 0.271$, $c = 0.940$. This means that a 50% detection efficiency is not achieved until a MES of 8.41, as compared to the idealized 7.1 sigma.

As the MES increases, the detection efficiency flattens out at $\approx 94\%$, an improvement over the SOC 9.2 pipeline for which the transit injection experiment (Christiansen et al. 2015b, 2016) recovered 92% for short-period injections (<100 days) and 81% for long-period injections (>100 days).

**Figure 2:** The fraction of simulated transits recovered as a function of the expected multiple event statistic (MES) by the Kepler SOC 9.3 pipeline using the Q1-Q17 DR25 pixel-level injected light curves. The black dashed line is $\text{MES} = 7.1$. The red dashed line is the hypothetical performance of the detector on perfectly whitened noise, which is an error function centered at $\text{MES} = 7.1$. The solid blue line is the gamma CDF fit to the histogram.
KEPLER “FINAL” CATALOG
Thompson+(2017)

ENHANCED VARIABILITY lowers the measured SNR of Earth-like transits

TO CATCH MORE EARTHS: lower the threshold
  • means more false alarms too
  • only feasible with automated vetting (Robovetter)
The latest Kepler catalog will yield improved estimates of $\eta_{\text{Earth}}$

STILL BASED ON VERY SMALL NUMBER OF DETECTIONS

PLATO will survey brighter, more varied sample of stars
ACTIVE STAR

ACTIVE REGIONS

PLANETARY ORBITS

TRANSITS

YEARS WEEKS DAYS HOURS MINUTES

LOG POWER DENSITY

LOG FREQUENCY

GRANULATION

OSCILLATIONS
EVOLVED STAR

LOG POWER DENSITY

ACTIVE REGIONS

GRANULATION

OSCILLATIONS

PLANETARY ORBITS

TRANSITS

YEARS  WEEKS  DAYS  HOURS  MINUTES

LOG FREQUENCY
ITS NOT JUST STELLAR NOISE
STOCHASTIC PROCESSES

STELLAR VARIABILITY = INTERPLAY OF MAGNETISM & CONVECTION

- too complex / ill understood to predict analytically
- intrinsically STOCHASTIC

MODEL STOCHASTIC PROCESS EXPLICITLY

- parametrise statistical properties of data (mean, covariance)
- build any physical knowledge into model
- Bayesian framework - marginalise over nuisance parameters
- easily combined with deterministic phenomena (planets)
GAUSSIAN PROCESSES

LIKELIHOOD: \( p(y|x, \text{model}) = \mathcal{N}(y|m, K) \)

MEAN FUNCTION: \( m=f(x, \Theta) \), COVARIANCE MATRIX: \( K_{ij} = k(x_i, x_j, \Phi) \)
TYPES OF GP

white noise: \[ K = \sigma^2 I \]
GPs ARE ALREADY USED in transit modelling

- CORRECT INSTRUMENTAL EFFECTS
- MODEL STELLAR VARIABILITY
- MODEL TRANSITS
Kepler-91

KOI 2133.01 (Bathala et al. 2013)
  • red giant host star

CONFLICTING STUDIES
  • Esteves et al. (2013): phase curve indicates self-luminous object
  • Lillo-Box et al. (2014b) RV + phase curve confirm planet
  • Sliski & Kipping (2014): stellar density from transit and asteroseismology inconsistent
GPs ARE ALREADY USED in systematics correction

MODEL INSTRUMENTAL EFFECTS
MODEL STELLAR VARIABILITY

DETECT TRANSITS
K2SC
Aigrain+(2016)

K2 MISSION: pointing variations + intrapixel variations → systematics

SYSTEMATICS: unknown 2-D function of star position

 STELLAR VARIABILITY: unknown 1-D function of time

NEED FLEXIBLE, JOINT MODEL
K2SC
Aigrain+(2016)

K2 MISSION: pointing variations + intrapixel variations —> systematics

SYSTEMATICS: unknown 2-D function of star position

 STELLAR VARIABILITY: unknown 1-D function of time

NEED FLEXIBLE, JOINT MODEL

80 DAYS

SYSTEMATICS CORRECTED

SYSTEMATICS ONLY
K2SC
Aigrain+(2016)
ADVERTISMENT
BREAK

K2SC version 2 light curves now available at MAST
Campaigns 3-10
GPs for DETECTION?

CHALLENGING

- large datasets - GP regression is $O(N^3)$

POSSIBLE WITH FAST MATRIX INVERSION?

- e.g.: CELERITE (Foreman-Mackey+2017)
IMPLICATIONS FOR PLATO

DON’T FILTER, MODEL!

- correct known, predictable instrumental effects to best ability
- measure/model the rest, and the stars
- understand better relationship between existing algorithms/GPs
- PLATO pipeline is being defined now…

TEST ON KEPLER LIGHT CURVES

- we know much more about variability properties of target stars now

CAREFUL TREATMENT OF ACTIVITY WILL BE CRITICAL

- to maximise detection efficiency
- to UNDERSTAND detection efficiency and reliability
- for RADIAL VELOCITY follow-up