Forward Modeling the Architectures of Exoplanetary Systems: A Clustered Model using Kepler Data

Matthias Y. He
Eric B. Ford (Advisor)
Pennsylvania State University
Darin Ragozzine
Brigham Young University

ExoPAG 21: January 4, 2020
Kepler’s multi-transiting systems are extremely informative to study. Many of these planets are in multi-transiting systems!
How do we make sense of all these planets?

- How do we disentangle observational biases from real trends?
- What do these observed planets suggest about the underlying systems?
- What are the trends in their architectures? What do they suggest about planet formation processes?
Framework for forward modeling planetary systems and the Kepler mission

1. **Step 1**: Define a statistical model for the intrinsic distribution of exoplanetary systems.

   \[ f(P, R) \]

2. **Step 2**: Generate an underlying population of exoplanetary systems (*physical catalog*) from a given model.

3. **Step 3**: Generate an observed population of exoplanetary systems (*observed catalog*) from the *physical catalog*.

4. **Step 4**: Compare the simulated *observed catalog* with the Kepler data using a distance function.

5. **Step 5**: Optimize the distance function to find the best-fit model parameters.

6. **Step 6**: Explore the posterior distribution of model parameters using a Gaussian Process (GP) model.
Framework for forward modeling planetary systems and the Kepler mission

We have a **full forward model** for simulating the Kepler mission!

**Define model**

\[
\lambda_c, \lambda_p, \\
\alpha_P, \alpha_{R1}, \alpha_{R2}, \\
f_{\sigma_{i,\text{high}}}, f_{\sigma_{i,\text{high}}}, \\
\sigma_{i,\text{low}}, \sigma_e, \\
\sigma_P, \sigma_R
\]

**Simulate a physical catalog**

Intrinsic systems

**Simulate an observed catalog**

Kepler

**Compare with Kepler data**

- Updated stellar radii from Gaia DR 2 (and \(b_p-r_p\) colors)
- Detailed Kepler detection efficiency
  - Hsu et al. (2018, 2019)
  - He, Ford, & Ragozzine (2019)
- Kepler DR 25 catalog (uniform vetting)
  - Fit to all observed marginal distributions simultaneously

Key differences:
Many previous studies assume that planets are independent in period and in size. We test a model where the period and radius of each planet are drawn independently.
Models assuming independent planets fail to reproduce the observed population

The number of multi-transiting systems is significantly under-produced

The period distribution appears well modeled with a single power-law (between 3 and 300 days)

The period ratio distribution is poorly modeled

The transit depth ratios are not as peaked
There are significant intra-system correlations: planets are clustered in periods and sizes.

Planets are drawn as a clustered point process, where each cluster has a period scale and radius scale.
Our clustered model provides a significantly improved description of planetary systems!

Observed multiplicities are fit extremely well

Both the period and period ratio distributions are well reproduced

Fit to transit depth ratios appear better, but distances not improved
Planetary systems have low eccentricities and consist of two populations of mutual inclinations. We find (from our clustered models):

- low eccentricities ($e \sim 0.02$)
- two populations of mutual inclinations
  - “hot”: isotropic $i_m$ for $\sim 40\%$ of systems to explain the excess of single-transit systems
  - “cold”: $i_m \sim 1.3^\circ$ for remaining $\sim 60\%$ of systems

The transit duration ratio distribution is very well matched.
Our clustered models provide a great fit to the Kepler data!

Access our model catalogs (or simulate your own) at: https://github.com/ExoJulia/SysSimExClusters
What about correlations between planetary systems and their host stars?

We split the stellar sample (79935 FGK stars) in half by Gaia b_p-r_p color —→ The exoplanet counts are also roughly split in half

Assuming the same distribution of planets around all stars produces more detected planets around the bluer stars!
Planetary systems are more common around late type stars than early type stars. The fraction of stars with planets is higher for redder (later type) stars compared to bluer (earlier type) stars. This trend is supported by studies by Howard et al. (2012), Mulders, Pascucci, & Apai (2015a), and He, Ford, & Ragozzine (in prep.).
A linear dependence between the fraction of stars with planets (FSWP) and the host star color fits the multiplicity distribution.

Both clustered models fit the overall multiplicity distribution well, but...

The linear function significantly improves the fit to the “bluer” and “redder” halves.

The architectures of planetary systems across FGK stars are similar, aside from the FSWP.
Summary

- **Multi-planet systems are clustered in periods and planet sizes**
  - The non-clustered model cannot fit the multiplicities, period ratios, and radius ratios
  - A clustered model better reproduces the multiplicities and period ratios

- **There are two populations of orbital architectures: low and high inclinations**
  - $e \sim 0.02, i_m \sim 1.3^\circ$ for 60% of systems
  - $i_m > 10^\circ$ for 40% of systems (Kepler dichotomy)

- **Planetary systems are more common around late type (cooler) stars**
  - The *overall* fraction of FGK stars with planets (FSWP) between 3 and 300 days is $\sim 60%$
  - FSWP increases from $\sim 30%$ (early F) to $\sim 100%$ (late K)
Feel free to use our models and simulated catalogs!

- Serves as a point of comparison for planet formation simulations
- Can be used to predict additional planets given already detected planets

Hsu et al. (2018, 2019)
He, Ford, & Ragozzine (2019)
He, Ford, & Ragozzine (in prep.)

Download or simulate model catalogs: https://github.com/ExoJulia/SysSim
ExClusters
Questions?

Kepler systems with 3+ planets

Simulated systems with 3+ planets
Extra slides
Framework for forward modeling planetary systems and the Kepler mission

**Step 1:** Define a statistical model for the intrinsic distribution of exoplanetary systems.

**Step 2:** Generate an underlying population of exoplanetary systems (*physical catalog*) from a given model.

- Draw stars from Kepler catalog
- Populate each system with planets
- Assign an orbit to each planet

- Kepler DR 25 (uniform vetting)
- Gaia DR 2 (updated radii, and $b_p-r_p$ colors)
- Draw periods, radii, masses
- Allow for two populations: Dynamically “hot” and “cold”
Framework for forward modeling planetary systems and the Kepler mission

Step 1: Define a statistical model for the intrinsic distribution of exoplanetary systems.

Step 2: Generate an underlying population of exoplanetary systems (physical catalog) from a given model.

Step 3: Generate an observed population of exoplanetary systems (observed catalog) from the physical catalog.

- Calculate which planets transit
- Add a transit noise and detection efficiency model
- Compute observed properties of planets

Christiansen (2017), Hsu et al. (2018, 2019)

Multiplicity, periods, depths, durations
Framework for forward modeling planetary systems and the Kepler mission

1. Define a statistical model for the intrinsic distribution of exoplanetary systems.
2. Generate an underlying population of exoplanetary systems (*physical catalog*) from a given model.
3. Generate an observed population of exoplanetary systems (*observed catalog*) from the *physical catalog*.
4. Compare the simulated *observed catalog* with the Kepler data using a distance function.
5. Optimize the distance function to find the best-fit model parameters.
6. Explore the posterior distribution of model parameters using a Gaussian Process (GP) model.
Model 1: planets are drawn independently in period and size

Planets are drawn as a Poisson point process with independent periods and sizes.

\[ N_p \sim \text{Poisson}(\lambda_p) \]

\[ P \sim f(P) = C_1 P^{\alpha_P} \]

\[ R_p \sim f(R_p) \propto \begin{cases} R_p^{\alpha_{R1}}, & R_{p,\text{min}} \leq R_p \leq R_{p,\text{break}} \\ R_p^{\alpha_{R2}}, & R_{p,\text{break}} < R_p \leq R_{p,\text{max}} \end{cases} \]
Models assuming independent planets fail to reproduce the observed population.
Model 2: planets are clustered in periods

Planets are drawn as a clustered point process better reproduce the observed period ratios

\[ N_c \sim \text{Poisson}(\lambda_c) \]
\[ N_p \sim \text{ZTP}(\lambda_p) \]
\[ P_c \sim f(P_c) = C_1 P_c^{\alpha P} \]
\[ P' \sim \text{Lognormal}(0, N_p \sigma_p) \]
\[ P = P' P_c \]
The clustered periods model fits the multiplicity, period, and period ratio distributions
Model 3: planets are clustered in periods and sizes

\[ R_{p,c} \sim f(R_{p,c}) = \begin{cases} C_3 R_{p,c}^{\alpha R_1}, & R_{p,c} \leq R_{p,\text{break}} \\ C_4 R_{p,c}^{\alpha R_2}, & R_{p,c} > R_{p,\text{break}} \end{cases} \]

\[ R_p \sim \text{Lognormal}(R_{p,c}, \sigma_R) \]
Our fully clustered model provides a significantly improved description of planetary systems!
The transit duration ratio distribution encodes population orbital properties

\[ \xi = \left( \frac{t_{\text{dur},\text{in}}}{t_{\text{dur},\text{out}}} \right) \left( \frac{P_{\text{out}}}{P_{\text{in}}} \right)^{1/3} \]

- Mutual inclinations randomize impact parameters
- Orbital eccentricities randomize velocity during transit

\( i_m = 0, e = 0 \) (coplanar + circular)

Kepler distribution
Differences between our models in their predictions for the underlying populations
We train a fast emulator for our models to quantify model uncertainties.

**Forward model** (slow, ~10s)
- Input model parameters
  - $\lambda_c, \lambda_p$
  - $\alpha_p, \alpha_{R1}, \alpha_{R2}$
  - $f_{\sigma_i, \text{high}}, \sigma_{i, \text{high}}$
  - $\sigma_{i, \text{low}}, \sigma_e$
  - $\sigma_p, \sigma_R$
- Simulate a physical catalog
- Simulate an observed catalog
- Compute distance with Kepler data

**Emulator** (fast, <0.01s)
- Input model parameters
  - $\lambda_c, \lambda_p$
  - $\alpha_p, \alpha_{R1}, \alpha_{R2}$
  - $f_{\sigma_i, \text{high}}, \sigma_{i, \text{high}}$
  - $\sigma_{i, \text{low}}, \sigma_e$
  - $\sigma_p, \sigma_R$
- Emulator
  - (Gaussian Processes, advanced statistical methods)
- Emulated distance with uncertainties

$D_W$
What about correlations between planetary systems and their host stars?

We split the stellar sample (79935 FGK stars) in half by Gaia b_p-r_p color

—> The exoplanet counts are also roughly split in half

Assuming the same distribution of planets around all stars produces more detected planets around the bluer stars!