Uniform Modeling of KOIs:

MCMC Notes for Data Release 25

KSCI-19113-001

Kelsey L. Hoffman and Jason F. Rowe

4 April 2017

NASA Ames Research Center
Moffett Field, CA 94035

Prepared by: Kelsey Hoffman ________________________ Date 4/4/17
Kelsey L. Hoffman, Support Scientist

Prepared by: ________________________ Date 4/4/17
Jason F. Rowe, Transit Scientist

Approved by: Michael R. Haas ________________________ Date 4/4/17
Michael R. Haas, Science Office Director

Approved by: ________________________ Date 4/4/17
Natalie M. Batalha, Project Scientist

Document Control

Ownership

This document is part of the *Kepler* Project Documentation that is controlled by the *Kepler* Project Office, NASA/Ames Research Center, Moffett Field, California.

Control Level

This document will be controlled under KPO @ Ames Configuration Management system. Changes to this document shall be controlled.

Physical Location

The physical location of this document will be in the KPO @ Ames Data Center.

Distribution Requests

To be placed on the distribution list for additional revisions of this document, please address your request to the *Kepler* Science Office:

Michael R. Haas
Kepler Science Office Director
MS 244-30
NASA Ames Research Center
Moffett Field, CA 94035-1000

or

Michael.R.Haas@nasa.gov

DOCUMENT CHANGE LOG

CHANGE DATE	PAGES AFFECTED	CHANGES/NOTES
04/04/2017	all	DR 25 Revision of DR24 Document (KSI-19084-002)

Table of Contents

1. Introduction

This document describes data products related to the reported planetary parameters and uncertainties for the *Kepler* Objects of Interest (KOIs) based on a Markov-Chain-Monte-Carlo (MCMC) analysis. Reported parameters, uncertainties and data products can be found at the NASA Exoplanet Archive[1]. The codes used for this data analysis are available on the Github website[2] (Rowe 2016). The relevant paper for details of the calculations is Rowe et al. (2015). The main differences between the model fits discussed here and those in the DR24 catalogue are that the DR25 light curves were used in the analysis, our processing of the MAST[3] light curves took into account different data flags, the number of chains calculated was doubled to 200 000, and the parameters which are reported are based on a damped least-squares fit, instead of the median value from the Markov chain or the chain with the lowest χ^2 as reported in the past.

The *Kepler* Mission (Borucki et al. 2010) used a 0.95-m aperture space-based telescope to continuously observe more than 150 000 stars for 4 years (Koch et al. 2010). We modeled and analyzed most KOIs listed at the Exoplanet Archive using *Kepler* observations. KOIs are not modeled with MCMC analysis when the transit event does not have sufficient S/N for proper modeling (S/N >~ 7), the KOI does not correspond to a transit event (e.g., KOI-54), or the Markov-Chain did not converge. As shown in Figure 1, the planet candidates for DR25 have orbital periods ranging from less than a day to greater than a year and radii ranging from that of the Moon to larger than Jupiter. Note, that there is an over-abundance of planetary candidates at periods close to 372 days that likely is due to instrumental noise modulated by the orbital period of *Kepler* around the Sun. Transit model and inferred planet parameters, such as radius, were obtained by fitting a model (Mandel & Agol 2002) to the photometric time-series produced from *Kepler* observations and convolving those parameters with measured stellar parameters from Mathur et al. (2017). The calculated models assume that all transit-like events are produced by planets that emit zero light in the *Kepler* bandpass. This assumption means that fitted parameters for stellar binaries and stellar blends will have significant systematic errors.

When a planet is observed to transit a star we see a drop in the observed flux that is proportional to the ratio of projected surface areas of the planet and star. The duration of the transit is dictated by the orbital motion of the planet and the shape is sculpted by the tilt of the orbital plane relative to the observer and the brightness profile of the host star. Using a transit model parameterized by the mean stellar density ($\rho*$), the ratio of the planet and star radii ($r/R*$), impact parameter (b), orbital period (P) and transit epoch (T0) we determine the depth, duration and shape of the transit. Our adopted parameterization assumes that the combined mass of all transiting planets is much less than the mass of the host star. Orbits are assumed to be non-interacting circular orbits. For eccentric orbits the planet-star separation during time of transit can be significantly different than the semi-major axis. Thus, a consequence of using circular

[1] http://exoplanetarchive.ipac.caltech.edu

[2] https://github.com/jasonfrowe/Kepler

[3] https://archive.stsci.edu/kepler/

orbits is that the model derived values of ρ^* can be systematically different from the true stellar values.

Details of the transit model can be found in §4 of Rowe et al. (2014). The procedure for determining the model parameters can be summarized as follows:

1. Processed the DR25 Q1-Q17 PDC-MAP *Kepler* photometry retrieved from MAST, which included detrending the data,
2. Starting with an initial guess model based on either the DR 24 model, or the TCE period, epoch, depth, and duration with limb darkening parameters matched to the updated stellar properties, we obtained a model fit with least-squares using a Levenberg-Marquardt routine (More et al. 1980),
3. Inspected the model fits with visual and numerical diagnostics and updated the problematic models,
4. Measured transit timing variations (TTVs), and if significant TTVs were found, then the TTVs were included in the model fit,
5. Ran MCMC routines using the model fit as a seed,
6. Inspected MCMC results and updated the problematic cases, and
7. Generated adopted model parameters and posteriors.

The details for these tasks can be found in §4 of Rowe et al. (2014) and §5 of Rowe et al. (2015). Here we describe the products produced in steps 2, 4 and 7.

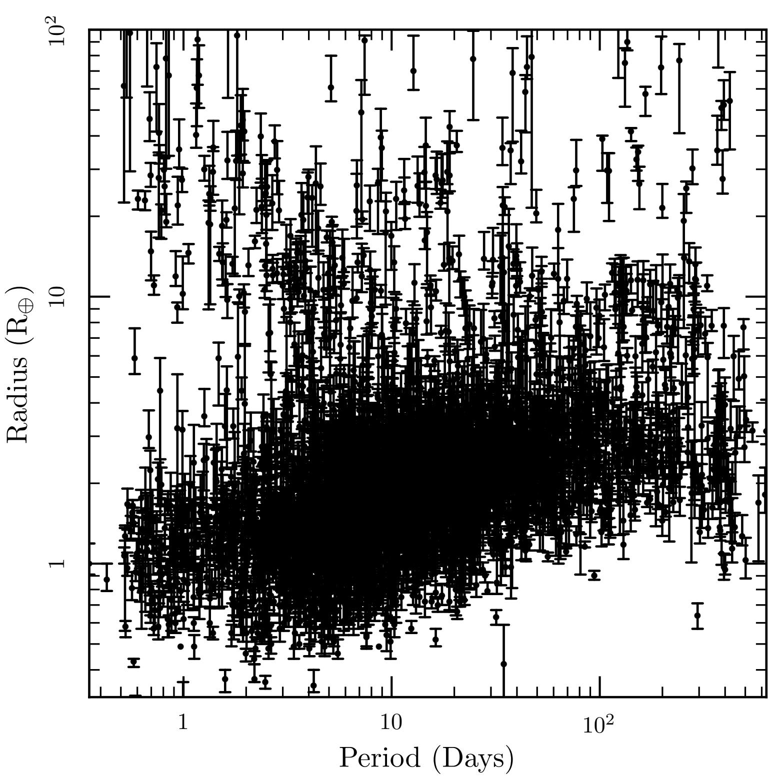

Figure 1: A plot of radius vs period for the planet candidates, including the errors on the radius calculation, from the DR25 catalogue. Note the overabundance of planet candidates at periods close to 372 days.

2. Directory Structure

For each KOI system there is a single directory as shown in Figure 2. The directory has the naming convention: "koiXXXX.n" where XXXX is the integer KOI number starting with 1. All KOIs around the same star will be found in this one directory. For example the six-planet system KOI157, with planets 157.01, 157.02, … , 157.06, is all contained in a single directory named "koi157.n". In this directory you will find data files containing light curves used for fitting, model fit parameters, transit-timing variations and calculated Markov-Chains as indicated in Figure 2. The next four sections explain the format of each file type.

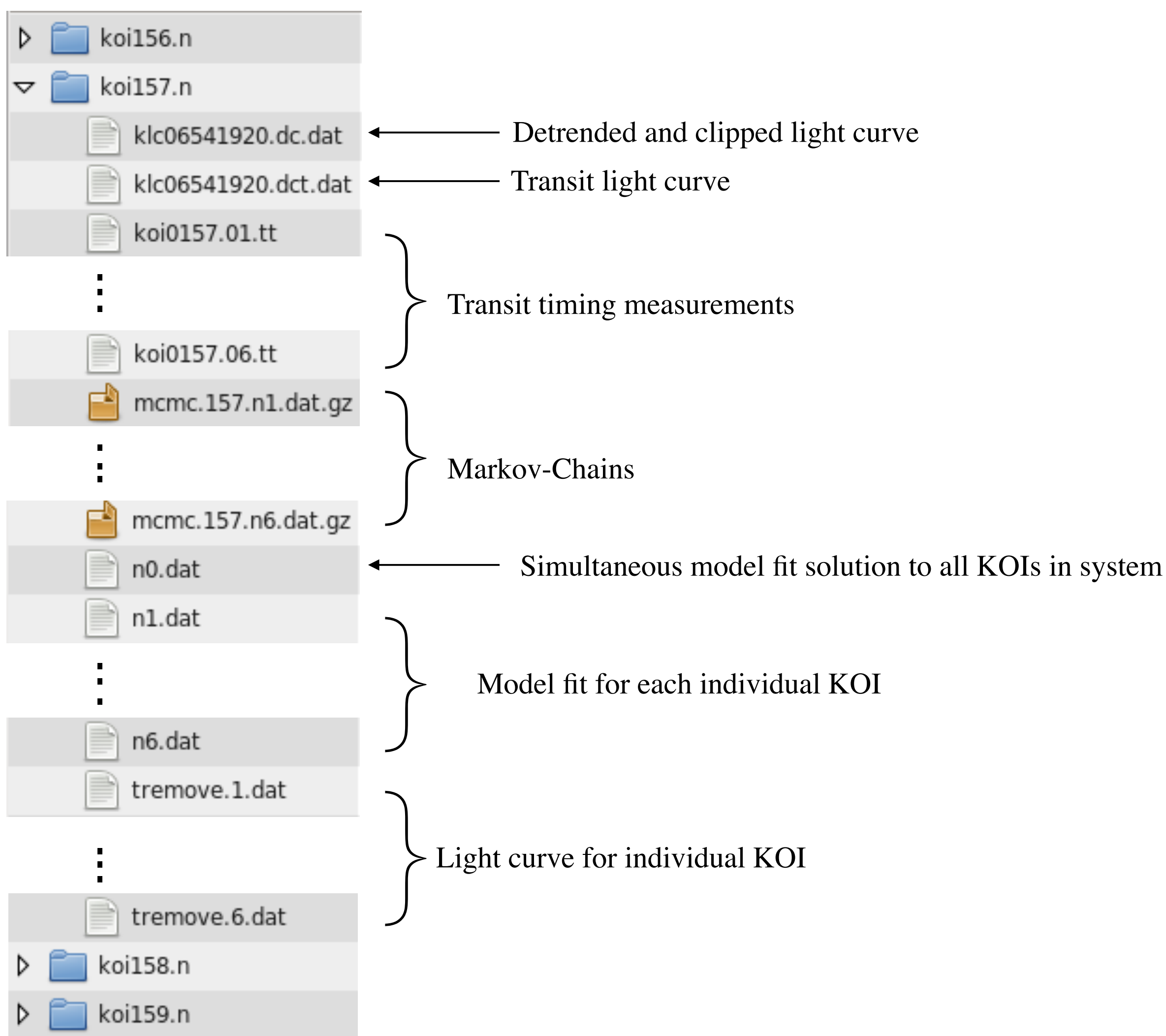

Figure 2: Example directory structure for the KOI-157 system.

3. Light Curve Files

For each KOI system we include several light curves that were used in generating the model fits:

1. Detrended and clipped light curve – the PDC light curve available from the MAST with the data further detrended after protecting transits, clipping outliers, and rejecting cadences flagged with bit values: 1, 2, 3, 4, 6, 7, 9, 13, 15, 16, and 17 (see Table 2-3 in Thompson et al. 2016). These files are named klc<kepid>.dc.dat, where <kepid> is the *Kepler* ID number of the host star corresponding to the KOI in the directory name.
2. Transit light curve – a version of the detrended and clipped light curve that only includes data within +/- one transit duration of the transit-time centers for all the planet candidates associated with a specific KOI. These light curves are especially useful for analyzing multi-planet systems. They were used to obtain the simultaneous model fits for all KOIs in a system as found in the n0.dat file (see Figure 2). These light curves are named klc<kepid>.dct.dat, where <kepid> is the *Kepler* ID number of the host star corresponding to the KOI in the directory name.
3. Light curve for individual KOI – a version of the detrended and clipped light curve that has all planets removed based on a transit model except for planet X. These files are named tremove.X.dat (see Figure 2), where X is the planet that remains.

4. Model Fit Parameters

Each KOI system has two types of best fits: simultaneous and independent. A file named n0.dat contains a global fit for all planets associated with the target star. In the case of KOI157, the parameters are a simultaneous fit of all six known transiting planets to the *Kepler* observations. There are also files named n1.dat, n2.dat, etc., that contain fit parameters for each planet fitted independently. An example is shown below of the contents of n1.dat, the model fit for KOI157.01:

```
RHO  1.2411307580E+00  0.0000000000E+00  3.0000000261E-03  0.0000000000E+00
NL1  4.2130000000E-01  0.0000000000E+00  0.0000000000E+00  0.0000000000E+00
NL2  2.5330000000E-01  0.0000000000E+00  0.0000000000E+00  0.0000000000E+00
NL3  0.0000000000E+00  0.0000000000E+00  0.0000000000E+00  0.0000000000E+00
NL4  0.0000000000E+00  0.0000000000E+00  0.0000000000E+00  0.0000000000E+00
DIL  0.0000000000E+00  0.0000000000E+00  0.0000000000E+00  0.0000000000E+00
VOF  0.0000000000E+00  0.0000000000E+00  0.0000000000E+00  0.0000000000E+00
ZPT -8.5928215914E-09  0.0000000000E+00  9.9999999748E-07  0.0000000000E+00
EP1  7.1176156554E+01  0.0000000000E+00  9.9999997474E-06  0.0000000000E+00
PE1  1.3024933066E+01  0.0000000000E+00  7.6951636284E-06  0.0000000000E+00
BB1  4.5426641109E-03  0.0000000000E+00  3.0000000261E-03  0.0000000000E+00
RD1  2.5234613596E-02  0.0000000000E+00  8.9658970080E-05  0.0000000000E+00
EC1  0.0000000000E+00  0.0000000000E+00  0.0000000000E+00  0.0000000000E+00
ES1  0.0000000000E+00  0.0000000000E+00  0.0000000000E+00  0.0000000000E+00
KR1  0.0000000000E+00  0.0000000000E+00  0.0000000000E+00  0.0000000000E+00
TE1  0.0000000000E+00  0.0000000000E+00  0.0000000000E+00  0.0000000000E+00
EL1  0.0000000000E+00  0.0000000000E+00  0.0000000000E+00  0.0000000000E+00
AL1  0.0000000000E+00  0.0000000000E+00  0.0000000000E+00  0.0000000000E+00
```

There are 5 columns. The first column gives the name of the fitted parameter, the second column gives the value of the fitted parameter and the fourth column determines whether that parameter was fitted or held fixed. The third and fifth columns were not used. The fourth column is zero if the parameter was held fixed during the fitting procedure, any other value indicates that the parameter was fitted. The fitted parameters are:

RHO - ρ^*, mean stellar density (g/cm^3).

NL1-4 - limb-darkening parameters. If NL3=NL4=0, then a quadratic law was adopted, otherwise a non-linear law (Claret & Bloemen 2011) was used.

DIL - fraction of light from additional stars in the aperture that diluted the observed transit. 0 - no dilution is present, 1 - additional source corresponds to 100% of flux.

VOF - radial velocity zero point (m/s). We did not include radial velocities in our fits.

ZPT - photometric zero point (relative). Detrending aims to have ZPT ~ 0.

EPy - T0, time of first transit for each planet y in units of BJD-2454900. For a multi-planet fit, there will be an entry for each planet: EP1, EP2, EP3,…

PEy - P, orbital period for each planet y (days).

BBy - b, impact parameter for each planet y.

RDy - r/R*, ratio of planet radius and star radius for each planet y.

ECy, ESy - eccentricity vector $e^{1/2}\cos(\omega)$, $e^{1/2}\sin(\omega)$ for each planet y.

KRy - radial velocity amplitude for each planet y. Doppler beaming is included (m/s).

TEy - secondary eclipse depth for each planet y (ppm).

ELy - amplitude of ellipsoidal variations for each planet y (ppm).

ALy - amplitude of phase-curve variations from albedo for each planet y (ppm).

Table 1 gives the matching parameter names as listed in the NASA Exoplanet Archive[4]. Figure 3 shows a model fit to the *Kepler* observations of KOI-18.01 (*Kepler*-5b).

Parameter	NASA Exoplanet Archive	Description
RHO	koi_srho	fitted mean stellar density (g/cm^3)
NL1-4	koi_ldm_coeff1,2,3,4	limb-darkening coefficients
ZPT	N/A	photometric zero point
EPy	koi_time0bk	transit epoch (BKJD)
PEy	koi_period	orbital period (days)
BBy	koi_impact	impact parameter
RDy	koi_ror	planet-star radius ratio

Table 1: Parameter names and corresponding labels from the NASA Exoplanet Archive.

[4] http://exoplanetarchive.ipac.caltech.edu/docs/API_kepcandidate_columns.html

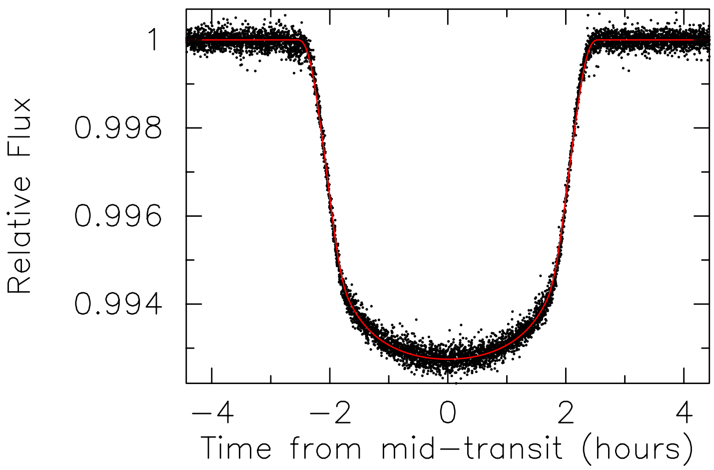

Figure 3: The transit of KOI-18.01 (*Kepler*-5b). The black dots show the long-cadence photometry from *Kepler* phased to the orbital period of *Kepler*-5b. The red line shows the model fit.

5. Transit Timing Variations

When transit-timing variations (TTVs) are detected, there is a file in the KOI directory with a ".tt" extension. The naming convention is "koiXXXX.0Y.tt" where XXXX is the KOI number and starts at 0001 and "Y" starts at 1 and is the planet number. Using the first few lines from the koi0137.01.tt file, an example of the contents is as follows:

```
60.7659403214000       3.518791081489780E-003  1.046973550741307E-003
68.4075085610000       6.613473893963828E-004  1.283974447298764E-003
76.0490768006000       2.166293968514310E-003  1.318249410890999E-003
83.6906450402000       2.341750138128873E-003  1.216472535053347E-003
91.3322132798000       5.052430408710507E-003  1.388923766915780E-003
106.615349759000      -4.452892475143244E-003  1.568101399638258E-003
114.256917998600      -1.614370642769813E-003  1.469243962292597E-003
121.898486238200       1.557455529706431E-003  1.786336376262667E-003
```

The first column is the calculated (expected) time (BJD-2454900) of transit based on the orbital period of the best-fit model. The second column gives the observed minus calculated (O-C) transit time (days) based on a fit of the transit model to the individually observed transit events. The third column is the uncertainty on the O-C time (days). If a ".tt" file is present, then transit-timing variations are included in the transit model. Figure 4 shows the measured TTVs for KOI-137.01 along with the Fourier transform of the TTVs.

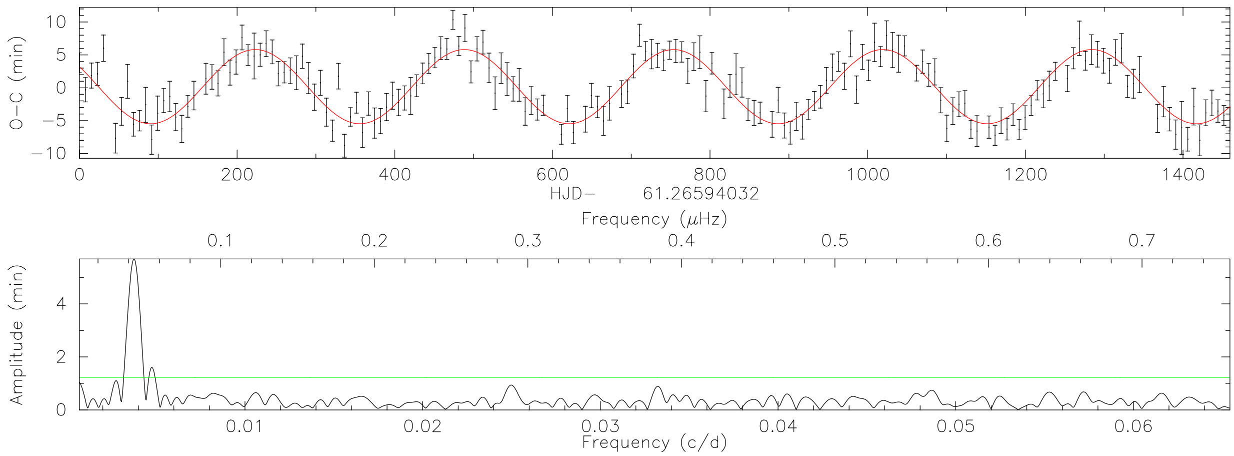

Figure 4: The top panel shows the O-C diagram of TTVs measured from KOI-137.01 (*Kepler*-18c). The bottom panel shows a Fourier transform of the TTVs. The green line is a 3-σ detection threshold.

6. Markov-Chains

Using the model fit and if necessary, TTVs, a Markov-Chain-Monte-Carlo routine was run to calculate a series of model parameter sets that were then used to estimate posterior distributions for each model parameter. The files are compressed with gzip and have the naming convention, koiXXXX.nY.dat.gz, where XXXX is the KOI number, which starts at 1, and Y is the planet number, which starts at 1. An example of the contents showing the first 3 lines of the file (which are wrapped) is:

```
          18

 1.4914682798E+04  0 11  1.2533986362E+00  4.2010000000E-01  2.5400000000E-01
0.0000000000E+00  0.0000000000E+00  0.0000000000E+00  0.0000000000E+00
6.6148941153E-07  7.1176263245E+01  1.3024925322E+01  1.9651895472E-02
2.5704275054E-02  0.0000000000E+00  0.0000000000E+00  0.0000000000E+00  0.0000000000E
+00  0.0000000000E+00  0.0000000000E+00

 1.4914718902E+04  0  1  1.2545401112E+00  4.2010000000E-01  2.5400000000E-01
0.0000000000E+00  0.0000000000E+00  0.0000000000E+00  0.0000000000E+00
6.6148941153E-07  7.1176263245E+01  1.3024925322E+01  1.9651895472E-02
2.5704275054E-02  0.0000000000E+00  0.0000000000E+00  0.0000000000E+00  0.0000000000E
+00  0.0000000000E+00  0.0000000000E+00
```

The first line gives the number of parameters, Np, in the model. For a single-planet fit, there are Np=18 parameters, for 2 planets there are Np=28, and 10 additional parameters for each additional planet. After the first line, each subsequent line has 3+Np entries. The first column reports chi-square for the parameter set, the second column is a flag to indicate if the parameter set was accepted (0) or rejected (1), and the third column is a flag to indicate which parameter was varied (0 for a vector jump (Gregory 2011), otherwise 1 through Np). If a chain was rejected, the last accepted chain is listed. The model parameters are listed in the order matching the contents of the best-fit file as described in §4. Thus, the fourth column lists the mean-stellar density. An example of the posterior distributions are shown in Figure 5 for KOI-571.05.

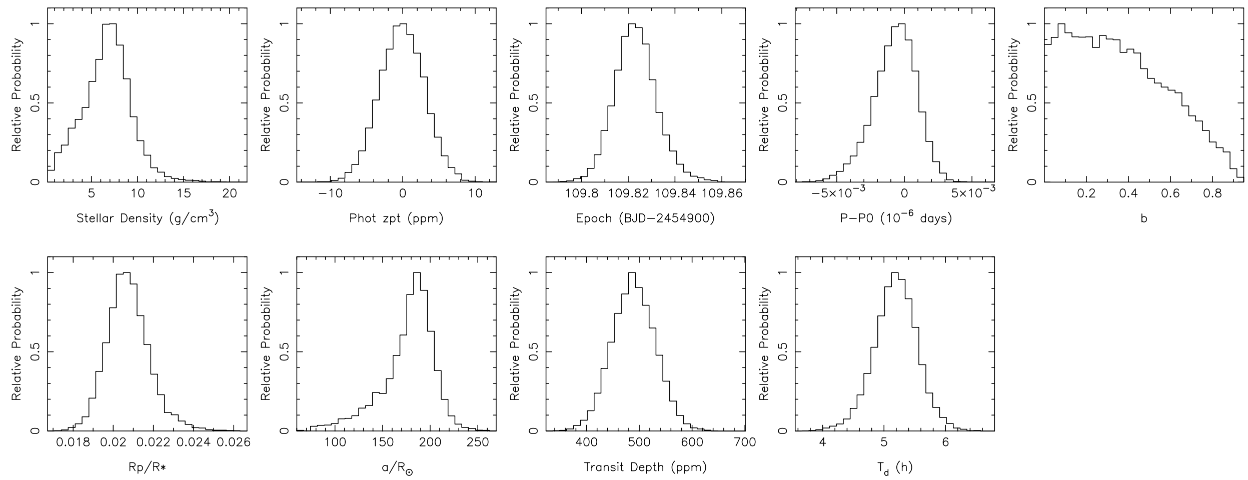

Figure 5: Posterior distributions for KOI-571.05 (*Kepler*-186f). Starting at the top and moving left to right we have: stellar density, photometric zero point, T0, P, b, r/R*, a/R*, transit depth and transit duration.

The Markov Chains were also calculated without any photometetric light curve data, using the same initial starting position as the real fit, to explore the impact of priors and ensure that there are no systematic biases in the MCMC routine that could bias the actual results. There are only 6 fitted parameters and the histograms of the posterior distributions without priors are shown in Figure 6. It was found that the distributions are flat for all fitted parameters, meaning that there are no strong biases in the MCMC routine.

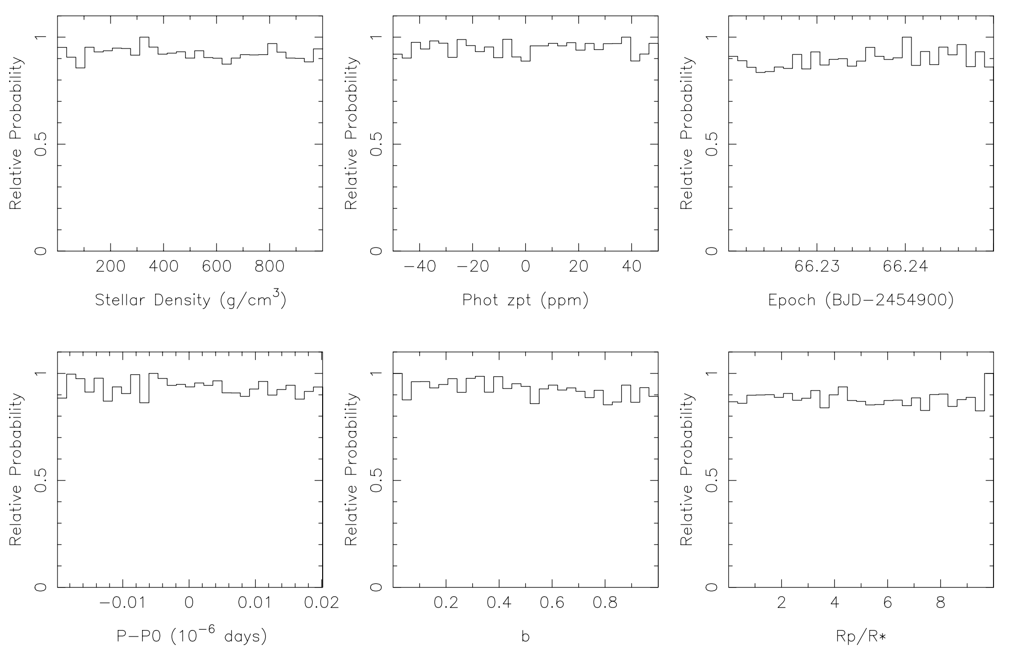

Figure 6: Posterior distributions of fitted parameters calculated without priors. Starting at the top and moving left to right there is: stellar density, photometric zero point, T0, P, b, and r/R*.

7. Using the Files to Estimate Your Own Posteriors

MCMC routines typically require what is known as 'burn-in'. To account for burn-in we recommend excluding at least the first ~20% of the provided chains.

To estimate any posteriors you must use both accepted and rejected flagged chains. For example, if you wish to estimate the average value of the modeled stellar density you would calculate the average based on every entry in the fourth column of the MCMC file after excluding the first ~20%. Figure 5 shows histograms based on an MCMC analysis of KOI-571.05 (*Kepler*-186f), which can be used to estimate posterior distributions.

As another example, the values for a/R* for DR25 were calculated via Kepler's 3rd law:

$$(a/R^*)^3 = \rho^* \, G \, P^2/(3\,\pi)\ .$$

We used the values for fitted mean stellar density (ρ^*) and period (P) to calculate a chain for the ratio a/R* for which we then estimated the median value and confidence interval.

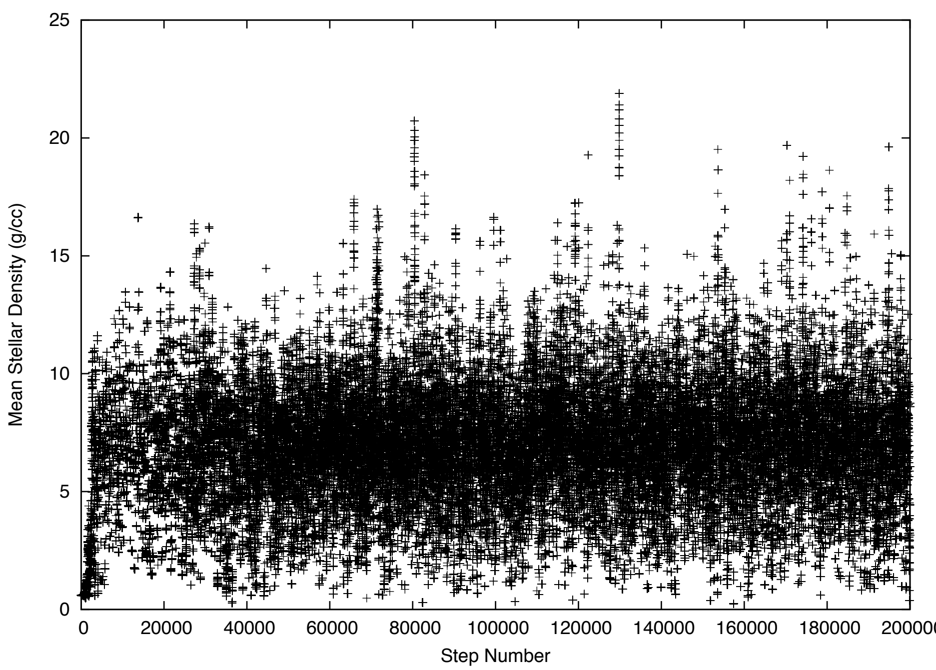

Figure 7: The value of the mean stellar density for 200 000 chains based on *Kepler* observations of KOI-571.05 (*Kepler*-186f). The first ~20% of the chains are not well mixed or behaved. This is due to burn-in of the MCMC routine and stabilization of the Gibbs factor to achieve an acceptance rate of 20-30%.

8. References

Borucki, W., et al. 2010, Science, 327, 977

Claret, A., & Bloemen, S. 2011, A&A, 529, 75

Gregory, P. C. 2011, MNRAS, 410, 94

Koch, D., et al. 2010, ApJ, 713, 79

Mandel, K., & Agol, E. 2002, ApJ, 580, L171

Mathur, S., Huber, E., Batalha, N. M., et al. 2017, ApJS, 229, 30

More, J., Garbow, B., & Hillstrom, K. 1980, Argonne National Laboratory Report ANL-80-74

Rowe, J. F., Bryson, S. T., Marcy, G. W., et al. 2014, ApJ, 784, 45

Rowe J.F., Coughlin, J.L., Antoci, V., et al. 2015, ApJS, 217, 16

Rowe, J. 2016, *Kepler*: *Kepler* Transit Model Codebase Release, https://doi.org/10.5281/zenodo.60297

Thompson, S. E., et al. 2016, *Kepler* Archive Manual, KDMC-19008-006