

## HEK VI: ON THE DEARTH OF GALILEAN ANALOGS IN *KEPLER*, AND THE EXOMOON CANDIDATE KEPLER-1625B I

A. TEACHEY<sup>1</sup>, D. M. KIPPING<sup>1</sup> & A. R. SCHMITT<sup>2</sup>  
ateachey@astro.columbia.edu

<sup>1</sup>Department of Astronomy, Columbia University, 550 W 120th St., New York, NY 10027

<sup>2</sup>Citizen Scientist

### ABSTRACT

Exomoons represent an outstanding challenge in modern astronomy, with the potential to provide rich insights into planet formation theory and habitability. In this work, we stack the phase-folded transits of 284 viable moon hosting *Kepler* planetary candidates, in order to search for satellites. These planets range from Earth-to-Jupiter sized and from  $\sim$ 0.1-to-1.0 AU in separation - so-called “warm” planets. Our data processing includes two-pass harmonic detrending, transit timing variations, model selection and careful data quality vetting to produce a grand light curve with a r.m.s. of 5.1 ppm. We find that the occurrence rate of Galilean-analog moon systems can be constrained to be  $\eta < 0.38$  to 95% confidence for the 284 KOIs considered, with a 68.3% confidence interval of  $\eta = 0.16^{+0.13}_{-0.10}$ . A single-moon model of variable size and separation locates a slight preference for a population of Super-Ios,  $\sim 0.5 R_{\oplus}$  moons orbiting at 5-10 planetary radii. However, we stress that the low Bayes factor of just 2 in this region means it should be treated as no more than a hint at this time. Splitting our data into various physically-motivated subsets reveals no strong signal. The dearth of Galilean-analogs around warm planets places the first strong constraint on exomoon formation models to date. Finally, we report evidence for an exomoon candidate Kepler-1625b-i, which we briefly describe ahead of scheduled observations of the target with the Hubble Space Telescope.

*Keywords:* planetary systems — techniques: photometric

### 1. INTRODUCTION

Moons present unique scientific opportunities. In our Solar System, they offer clues to the mechanisms driving early and late planet formation, and several of them are thought to be promising targets in the search for life, as several are rich in volatiles (e.g. Squyres et al. 1983; Hansen et al. 2006) and possess internal heating mechanisms (e.g. Morabito et al. 1979; Hansen et al. 2005; Sparks et al. 2016). The moons of our Solar System also demonstrate the great variety of geological features that may be found on other terrestrial worlds.

In this new era of exoplanetary science it stands to reason that moons in extrasolar systems, so-called exomoons, should tell us a great deal about the commonality of the processes that shaped our Solar System and may yield just as many surprises as their host planets before them. Just as the study of exoplanets has complicated our picture of planetary formation by revealing (for example) the existence of Hot Jupiters (Mayor & Queloz 1995) – worlds without Solar System analogs – so too might moons show us what else is possible and uproot conventional thinking about satellite formation

mechanisms.

Galilean-sized moons ( $\sim 0.2$ - $0.4 R_{\oplus}$ ) are generally thought to be able to form in a variety of ways. For the regular satellites of Jupiter, the Galilean moons are thought to have condensed out of a circumplanetary disk, akin to planet formation within a protoplanetary disk (Canup & Ward 2002). This process is expected to limit regular satellites to a cumulative mass of  $\mathcal{O}[10^{-4}]$  that of the primary (Canup & Ward 2006). Higher mass-ratio moons, such as the Earth’s Moon, are evidently viable too and may form from catastrophic collisions in the first few hundred million years of the solar system, coalescing from that collision’s debris (e.g. Ida et al. 1997). Finally, retrograde Triton is hypothesized to have originated from a captures via a binary exchange mechanism (Agnor & Hamilton 2006). Put together, Galilean-sized satellites appear to have formed via at least three independent pathways within the Solar System, and their existence around exoplanets can therefore be reasonably hypothesized.

Galilean-sized exomoons are challenging to detect for a number of reasons. First, the transit of  $0.2$ - $0.4 R_{\oplus}$

moon across a Sun-like star results in a depth of 3–13 ppm, below the typical sensitivity achievable with *Kepler* (Christiansen et al. 2012). Second, the moon signal will almost certainly be found at each epoch in a different location with respect to the host planet, sometimes occurring before the transit, sometimes after, and at a different projected distance from the planet (Kipping 2011). Third, multiple moons around a single planet may wash out any transit timing (Sartoretti & Schneider 1999) or duration variations (Kipping 2009a,b). The three-body motion combined with the potential for overlapping disks (syzygies), makes accurate modeling of exomoon transits computationally demanding (Kipping 2011). Fourth, the long-period nature of plausible moon hosts means that relatively few transits are usually available. In this regime, a planet+moon light curve model has sufficient flexibility to almost always provide a superior fit to the limited data (Kipping 2013), thereby necessitating rigorous Bayesian approaches to model selection. Despite these challenges, the “Hunt for Exomoons with Kepler” (HEK) project has performed Bayesian photodynamical fits of ∼60 exoplanets to date (Kipping et al. 2012, 2013a,b, 2014, 2015b), with no unambiguous detections and limits typically hovering around an Earth-mass.

In this work, our project pursues a different approach to searching for evidence of moons which focuses on seeking a population of moons around a population of planets. Rather than pursuing individual limits which are then combined to constrain the population, we here approach the problem from a broad statistical perspective in order to directly measure the occurrence rate of moons,  $\eta_{\mathcal{C}}$ .

## 2. STACKING EXOMOOONS

### 2.1. Phase-Folding

The work presented here aims to exploit the power of stacking in order to search for exomoons. Stacking is a familiar technique to those studying exoplanet transits, who typically phase-fold a light curve upon the period. For a linear ephemeris, the transits align leading to a coherent signal. It is important to stress that this act does not improve the signal-to-noise ratio (SNR). The amount of data before and after stacking is the same, with the only difference being that stacking assumes the ephemeris of the planet is known to infinite precision. Modeling the full unstacked light curve with a model conditioned upon the same ephemeris assumption would result in identical posteriors and thus no improvement is achieved for the inference itself. Despite this, stacking is attractive because the signal’s coherence means that full light curve modeling is unnecessary in the context of signal detection. Specifically, one may simply evaluate

the weighted mean centered around the pivot point of the fold to achieve a detection, which is why the popular Box Least Squares (BLS) algorithm (Kovács et al. 2002) is a computationally efficient yet sensitive tool in conventional transit detection.

Stacking light curves in pursuit of exomoons is complicated by the fact simply phase-folding light curves upon a linear ephemeris will lead to the moon appearing at different phases in each epoch. Despite the fact the moon is not perfectly coherent, it is constrained to lie within the Hill sphere and this imparts some quasi-coherent properties into the phase-folded light curve. Simon et al. (2012) were the first to describe this possibility, where they argued that this quasi-coherence will lead to an increase in the photometric scatter in the temporal region surrounding the planetary event – an effect they dubbed “scatter peak”. A similar idea is discussed in Heller (2014), who instead considered looking for a slight photometric decrease in this temporal region. By considering the probability density of the moon’s sky-projected position, an effect dubbed as the “orbital sampling effect” (OSE), Heller (2014) derives a simple formula for predicting phase-folded light curve shape in the presence of moons, enabling a simple approach to seeking exomoon shadows.

As with the case of a planetary transit, or indeed any kind of stacking, this approach does not boost SNR in anyway. The data volume and quality are the same before and after the stacking. However, unlike the planet case, the shape of the phase-folded moon signal represents a washed-out depiction of the individual signals. Accordingly, the subtle individual variations in durations, positions and shapes are lost, meaning that stacking imposes a fundamental loss of information content, and therefore sensitivity.

A similar situation occurs when observing planetary transits with long exposures, such as the 30 minute cadence (LC) mode of *Kepler*, causing the shape of the transit to be slightly washed-out, thereby degrading the information content. It is for this reason that short-cadence (SC) *Kepler* data provides tighter constraints on transit times, despite the fact both see the same SNR transit depth (e.g. see Kipping et al. 2013a and Kipping et al. 2014).

Accordingly, searching for exomoons in phase-folded light curves will always be less sensitive than full photodynamic fits – although precisely how much has not been formally evaluated and would be sensitive to the specific planet-moon parameters. Despite this, phase-folding is attractive for its simplicity and as a pure detection tool – analogous to BLS for planet hunting.

One particularly attractive feature of phase-folding exomoons is that a sizable fraction of the quasi-coherent signal appears exterior to the planetary transit. As

suming the planetary transits are well-aligned and the duration is well-known, one may simply crop the planetary transit leaving behind a pure exomoon signal. This greatly simplifies the analysis, since the planet properties are not covariant with this signal. For these reasons, we identify this out-of-transit phase-folded moon signal as the target signature in this work.

## 2.2. Planet-Stacking

Unlike a simple planetary phase-fold, the quasi-coherent nature of the phase-folded moon light curve means that a large number of transits are needed to produce a predictable signal. At its core, the phase-folded moon signal depends upon the law of averages and thus co-adding relatively few transits can lead to a phase-folded moon signal which is highly irregular and erratic. Without a characteristic and predictable shape, it is very difficult to convincingly argue the signal is genuinely a moon, rather than some peculiarity of the data in those limited co-added events. Indeed, Heller (2014) argue that at least a dozen events are generally needed, such that  $N \gg 1$  and the averaging effect can become noticeable.

Unfortunately, *Kepler*'s primary mission lifetime of 4.35 years means that the long-period planets, where moons are most *a-priori* expected to be viable, were only observed to transit a few times. Only in a small number of cases are there *Kepler* planets for which their period is long enough such that moons are dynamically stable for Gyr and we possess  $N \gg 1$  transit events within the 4.35 years of *Kepler* observations. This point seemingly excludes phase-folding as a viable exomoon approach, except for a few rare cases.

We devised an approach to solve this problem, inspired by the work of Sheets & Deming (2014). In that work, the authors not only phase-folded each planetary light curve but also co-added different planets together. This allowed them to greatly increase the number of ensemble phase folded signals, which in their case was used to boost the overall SNR. In many ways, this approach is reminiscent of a hierarchical Bayesian model (HBM; Hogg et al. 2010) but by stacking the objects the identities of each object are sacrificed. While an HBM approach would be better suited, in general, direct planet stacking is attractive for its simplicity, particularly if the objective is purely to test whether an ensemble signal even exists rather than attempting to perform detailed characterization of said signal. Therefore, in the same vein, we decided to try stacking different phase-folded planet signals together, to solve the  $N \gg 1$  problem. We highlight that Hippke (2015) independently arrived at the same idea and published before our effort, although many differences exist in our actual implementations, as will become clear throughout this

paper.

We will refer to the out-of-transit phase-folded planet-stacked photometric light curve as the “grand light curve” in what follows, for notational convenience.

Co-adding different planets decreases the overall noise, since we are extending the data volume upon which our inferences are conditioned. However, this approach is not guaranteed to increase the SNR, since many of the objects co-added may not even possess moons and thus their inclusion only dilutes the overall signal, rather than co-adding to it. Nevertheless, we can quantify the overall signal amplitude as being a combination of the occurrence rate and the moon radii. Even so, selecting a sample of planets which are expected, *a priori*, to be plausible hosts for large moons will be crucial for maximizing our chances of a successful detection and correspondingly deriving meaningful, physically-constraining upper limits.

## 3. TARGET SELECTION

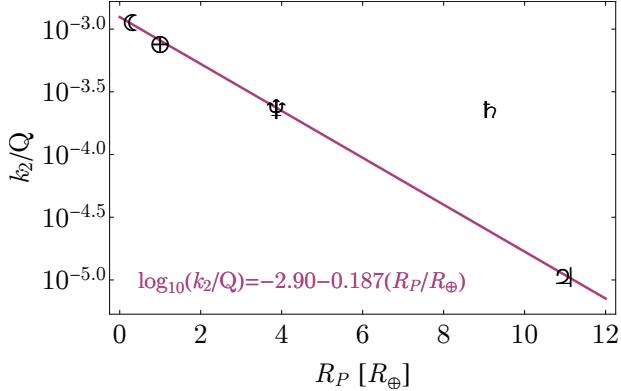
### 3.1. Automated Target Selection

Not all exoplanets are equally likely, *a priori*, to yield an exomoon detection. At the most basic level, two questions guide our target selection process: 1) what is the largest stable moon plausible around a given planet 2) would this moon be detectable, given the current data in hand?

In this work, we estimate whether a detectable moon is plausible following a similar approach to that adopted in previous HEK papers. Specifically, we employ the *Target Selection Automatic* (TSA) algorithm described in Kipping et al. (2013a).

To summarize, the algorithm first estimates a mass for each Kepler Object of Interest (KOI) using the maximum likelihood radius reported on the [NASA Exoplanet Archive](#) (NEA; Akeson et al. 2013) and the mass-radius relation defined in Kipping et al. (2013a). This is used to further estimate the extent of the planet's Hill sphere. Moons are expected to have their lifetime limited by the time it takes to tidally spin out from the Roche limit to some critical fraction of the Hill sphere,  $f R_H$ . Using the expressions of Barnes & O'Brien (2002), this logic may be inverted to compute the maximum allowed moon mass,  $M_{S,\max}$ , which can survive for a fiducial age of  $t_* = 5$  Gyr. For TSA, we set  $f = 0.9309$  for the optimistic case of a retrograde moon (Domingos et al. 2006).

In order to compute  $M_{S,\max}$ , we must adopt a value for the ratio  $(k_2/Q)$ , which represents the efficiency of tidal dissipation. Whereas in previous HEK papers we simply adopted  $(k_2/Q) = 0.5/10^5$ , here we use an empirical relation based on the System System. As shown in Figure 1, empirical estimates of  $(k_2/Q)$  for the Moon



**Figure 1.** Relationship between the tidal property  $k_2/Q$  and planet size for the Solar System bodies. When excluding Saturn, the remaining four points closely follow a power-law used in this work.

(Dickey et al. 1994), the Earth (Kozai 1968), Neptune (Trafton 1974), Jupiter (Lainey et al. 2009) and Saturn (Lainey et al. 2012) follow a power-law distribution versus planetary radii, except for Saturn. In this work, we invert this empirical relation (ignoring Saturn) to estimate  $(k_2/Q)$  for a given  $R_P$ . While we don't claim this to be a fundamentally general law, it at least provides a somewhat more reasonable estimate than the blanket fixed value assumed previously.

Once the maximum moon mass has been computed, it is then converted into a moon radius using the same mass-radius relation as before. Since target selection was conducted early on in the two-year duration of this research comprising this paper, it predated the more robust probabilistic mass-radius relation of Chen & Kipping (2017a), which is why that latter relation was not used for these calculations. We query the combined differential photometric precision (CDPP) of each host star (Christiansen et al. 2012), which along with the maximum moon radius, allows us to estimate the signal-to-noise ratio (SNR) expected due to the moon.

For the SNR calculation, we estimate the phase-averaged signal amplitude using the so-called “orbital sampling effect” (OSE) described in Heller (2014). While these expressions formally assume a large number of transits, which is rarely true, they work well as an approximation for the signal strength marginalized over the moon's phase, which is of course unknown to us *a priori*. The expected OSE flux decrease for the out-of-transit data is given in Heller (2014) (and stated later in Equation 9), from which we may integrate over the signal to compute the signal strength of the out-of-transit portion to be

$$S = \left(\frac{R_S}{R_\star}\right)^2 \frac{\sqrt{(a_{SP}/R_P)^2 - 4} - 2 \cos^{-1}(2R_P/a_{SP})}{\pi((a_{SP}/R_P) - 2)}. \quad (1)$$

The fraction on the right-hand side varies from about a quarter to a third for  $(a_{SP}/R_P)$  in the range of 5 to 100 i.e. it is a relatively gentle function. We therefore take the limit of large  $(a_{SP}/R_P)$ , giving

$$S = \frac{1}{\pi} \left(\frac{R_S}{R_\star}\right)^2. \quad (2)$$

The SNR may now be calculated by dividing this by the noise expected

$$\text{SNR} = \frac{1}{\pi} \left(\frac{R_S}{R_\star}\right)^2 \frac{\sqrt{T_{\text{Hill}}/0.25}}{\text{CDPP}_6} \sqrt{\frac{B}{P}} \quad (3)$$

where  $B$  is the baseline of observations, optimistically assumed to be the full Q1-17 baseline for these calculations and  $\text{CDPP}_6$  is the combined differential photometric precision on a 6 hour timescale. We may express the Hill time, assuming a simple circular orbit approximation, as

$$T_{\text{Hill}} = f \left(\frac{P}{2\pi}\right) \left(\frac{M_P}{3M_\star}\right)^{1/3}, \quad (4)$$

which when substituted in leads to the  $P$  terms cancelling out, such that

$$\text{SNR} = \frac{(R_S/R_P)^2}{\pi} \left[ \frac{\sqrt{2Bf}}{3^{1/6} \sqrt{\pi} \text{CDPP}_6} \left(\frac{M_P}{3M_\star}\right)^{1/6} \right]. \quad (5)$$

As discussed later in Section 6, we find that the OSE model overestimates the signal-to-noise for large  $a_{SP}$ , with numerical experiments showing it is around a factor of three too high by the time we hit  $(a_{SP}/R_P) = 100$ . We therefore correct the SNR quoted above by dividing by a factor of  $\sim 3$ , which together with the  $\pi$  denominator we simply approximate to a factor of  $\sim 10$  denominator, yielding

$$\text{SNR} \simeq \frac{(R_S/R_P)^2}{10} \left[ \frac{\sqrt{2Bf}}{3^{1/6} \sqrt{\pi} \text{CDPP}_6} \left(\frac{M_P}{3M_\star}\right)^{1/6} \right]. \quad (6)$$

### 3.2. Applying to the Kepler Planetary Candidates

TSA was first run for this project in November 2014, at which time 7305 KOIs were listed on NEA. However, 27 were removed due to having some incomplete column entries. Of these, 4109 were not classified as a “false-positive” by NEA and thus were considered further. In order to calculate SNR, basic stellar properties are required and so we cross-referenced this list with the

[Huber et al. \(2014\)](#) catalog, for we were unable to find a match for 11 KOIs and thus they were removed giving us a total of 4098 KOIs which were then put through the TSA algorithm.

Due to the ensemble nature of our analysis, the total SNR is expected to be much greater than that of individual objects and thus we can afford to use a relatively generous SNR cut. Accordingly, we elected to use  $\text{SNR} > 0.1$  and apply the criteria that  $P < B/4 = 397.39$  days (to give three transits), leading to a sample of 966 KOIs outputted from the TSA procedure. These targets are visualized in Figure 2.

Since this work has taken several years to complete, some of the objects that were considered exoplanet candidates when we began are now considered false-positives. In December 2016, we elected to remove all KOIs if they were either classified as “false-positives” at [NEA](#) or the probability of any false-positive scenario was in excess of 1%, as given by the values listed on [NEA](#). This filter removed 292 ( $\sim 30\%$ ) of the objects originally considered, reducing the 966 KOIs to 674 for the remainder of the survey.

## 4. DATA PROCESSING REQUIREMENTS

### 4.1. Overview

The objective of this work is to create a phase-folded planet-stacked out-of-transit light curve, which may be used to search for evidence of exomoons. For the sake of brevity, we will refer to this light curve as the grand light curve, or simply GLC, in what follows.

We identify four unique and critical requirements to realizing this objective, specifically:

1. removal of TTVs,
2. very high quality light curves,
3. temporal rescaling,
4. two-pass data processing.

We explain and discuss these requirements in what follows.

### 4.2. TTVs

In order to create an accurate phase-folded light curve of a sequence of planetary transits, it is necessary to ensure the transits accurately phase up. In the absence of transit timing variations (TTVs), this is straightforward and is a simple linear ephemeris fold. However, the signal we seek, an exomoon, will always introduce a small TTV signal into the data ([Sartoretti & Schneider 1999](#)). Moreover, TTVs can be caused by other effects, notably perturbing planets ([Holman & Murray 2005](#); [Agol et al.](#)

2005) and thus TTVs are observed to be fairly common ( $\gtrsim 10\%$ ) in *Kepler* planetary systems ([Holczer et al. 2016](#)). Carefully removing these TTVs is crucial in creating an accurate phase-folded transit signal.

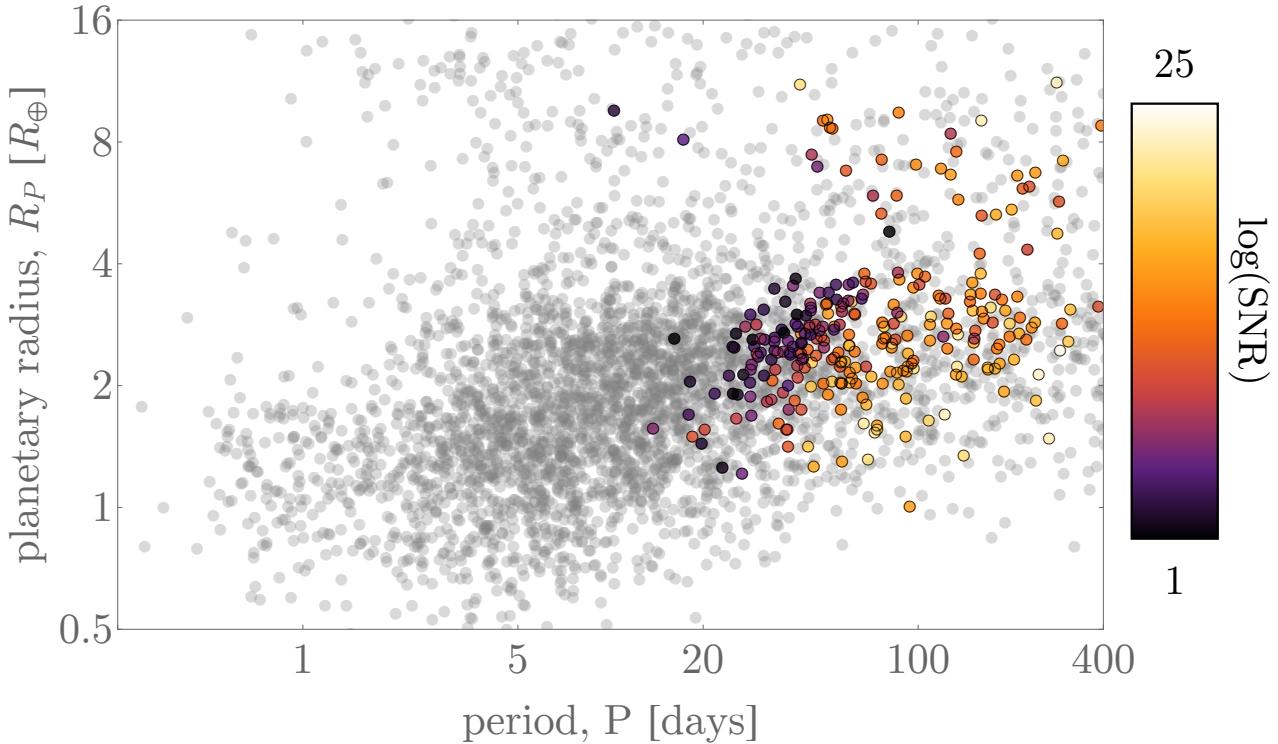
One approach might be to take the catalog of known TTVs from [Holczer et al. \(2016\)](#) and use these for corrections. There are several reasons why this is unsatisfactory for this work though. First, in order to assess robust confidence limits, we require covariant, joint posteriors distributions of the transit times and basic planet parameters, which were not derived in [Holczer et al. \(2016\)](#). Second, whenever possible, accurate phase stacking is aided by first conducting model selection between the TTV and linear ephemeris models, which itself formally requires computation of the Bayesian evidence - again something not derived in [Holczer et al. \(2016\)](#). Third, the TTVs derived in [Holczer et al. \(2016\)](#) were conditioned on a different data set to that used in this work. More specifically, although both [Holczer et al. \(2016\)](#) and this work are based on *Kepler* photometry, our data detrending methods are distinct meaning that these differences should be expected to affect the TTV measurements to some degree. When one is ultimately seeking the discovery of a few parts-per-million signal, these conventionally minor issues cannot be ignored and should be expected to influence the results.

For these reasons, we concluded that creating an accurate grand light curve was not possible without first deriving TTV posteriors ourselves for every system considered.

### 4.3. Data Quality

There is a unique property of the phase-folded moon signal that has strong implications for the data quality requirements, which is not conventionally an issue for planet analyses. The GLC signal is a phase-fold of the planetary transit, after removing TTVs, and thus at any given instant in phase, the moon actually only induces a transit-dip for some fraction,  $F$ , of the co-added light curves. Geometry demands that this fraction must always be less than one-half (i.e.  $F < 0.5$ ) for all phase points occurring outside of the planetary transit signal. This is a key point which has a major implication - median binning kills the GLC signal.

This is extremely important, because median binning is a robust point estimate. The forgiving nature of median binning means that one can actually do a bad job of detrending some small fraction of your light curves (which represent outliers) yet still recover an accurate phase-folded signal. However, if one cannot use median binning, then one is forced to use mean-based estimates which are sensitive to each and every transit co-added. In this case, even a single inaccurately detrended transit light curve will contribute to the phase-stacked signal.



**Figure 2.** Location of the KOIs selected to search for exomoons around. The 966 colored points, color-coded by the SNR given in Equation 6, tend to be at relatively long orbital periods, where Hill spheres are larger. Points with a solid, black circle around it are the 347 KOIs found to pass our data quality vetting.

Once again, since we seek the detection of a signal with an amplitude of a few parts per million, this cannot be ignored and demands the highest levels of scrutiny and data quality.

We therefore establish that each and every transit used in our grand light curve must be verified to be of very high data quality, which of course greatly increases the time demands needed to complete such an analysis.

#### 4.4. Temporal Rescaling

When we finally arrive at an accurate phase-folded light curve for each planet, they must be combined into a single grand light curve. This is similar to the co-addition performed for occultations by [Sheets & Deming \(2014\)](#). In their case, each occultation has a distinct duration and thus simply co-adding the occultations would cause the signal to smear out and produce an averaged signal distorted from the true morphology. To overcome this, [Sheets & Deming \(2014\)](#) re-scaled each event by the known duration and then co-added, producing a more coherent signal. Just as with the occultations, each GLC signal will have a different velocity and impact parameter and thus cause a different duration. However, the problem is actually worse since unlike [Sheets & Deming \(2014\)](#), we don't know what the true duration of each event should be, since the duration is highly sensitive to the semi-major axis of the moon(s), which are of course

not yet discovered.

Ultimately, re-scaling will always be flawed since we can't know the semi-major axis of the moon prior to discovering it. A full hierarchical Bayesian model (HBM) would be an appealing way of approaching this problem, allowing each object to have a unique semi-major axis. However, since each planet would not satisfy  $N \gg 1$  transits, the OSE approximation would break down and thus each system would require modeling with a full photodynamic simulation, such as that from LUNA ([Kipping 2011](#)). For five years, we in the HEK project have been conducting Bayesian regression of individual systems with LUNA, and the computational demands for even individual systems are formidable ( $\sim 30,000$  CPU hours per planet). Linking this into a full HBM would be computationally extremely challenging and was not a strategy we elected to pursue here.

Moreover, in this work, we ultimately hoped to find a signal which was visually evident in the final grand light curve and thus not conditional upon the inferences of an HBM. While this does not maximize the information content of the final data product, we are motivated to follow this philosophy on the basis that the discovery of any novel phenomenon, which exomoons would represent, requires a much higher confidence than routine discoveries ([Gould et al. 2004](#)).

Accordingly, we seek a method of re-scaling which is “least-bad” and will maximize the expected signal coherence even when marginalizing out our ignorance of the moon’s semi-major axis. One approach is that of Hippke (2015), who re-scaled by the duration of the planetary transit. The advantage of this is mostly simplicity, the duration is well-constrained and easy to understand. One downside of this is that even if all the moons had the same semi-major axes, they would still lead to the grand light curve having a smeared out OSE signal, since each system has different barycentric velocity and impact parameter across the star. Another approach, re-scaling by the Hill radii, is not possible since the exoplanet masses are unknown.

Instead, in this work, we argue a better approach is to re-scale the time axis into distance from the planet, in units of planetary radii. This can be accomplished by considering the original Seager & Mallén-Ornelas (2003) equation for the duration of a planet, under the assumption of circular orbits:

$$T_{23}^{14} = \frac{P}{\pi} \sin^{-1} \left( \sqrt{\frac{(1 \pm p)^2 - b^2}{a_R^2 - b^2}} \right). \quad (7)$$

If we let  $(1 \pm p) \rightarrow 1$  in the above, we recover the transit duration as defined when the center of the planet overlaps with the stellar limb,  $\tilde{T}$  (Kipping 2010a). Thus, at contact point 1 & 4, we can think of this instant in time as when a shell of radius  $p$  centered on the planet first starts to induce transit-dip features. By extension, we could adapt  $(1 \pm p) \rightarrow (1 + t'p)$  in the above, which would equal the duration of a shell of radius  $t'p$ , centered on the planet, to start/end creating transit-dip features. In this way, we can think  $t'p$  as being the orbital distance of the moon at the instant in time when the transit begins/ends. Accordingly,  $t'$  represents the planet-moon distance in units of the planetary radius. This convenient form allows us to use the transit observables directly to convert from time into a physically motivated dependent variable via:

$$t' = \left[ \sqrt{b^2 + (a_R^2 - b^2) \sin^2 \left( \frac{2\pi}{P}(t - \tau) \right)} - 1 \right] / p. \quad (8)$$

If all of the moons shared the same  $(a_{SP}/R_P)$ , this would produce a coherent signal. In reality, we do not expect this statement to be true but moons do appear in the Solar System to be distributed log-uniformly with respect to this term (Kane et al. 2013). This approach means that we could model the resulting grand light curve assuming exomoons followed a formulaic distribution for  $(a_{SP}/R_P)$ , such as a log-uniform.

In order to convert from  $t \rightarrow t'$ , we need estimates

for the impact parameter and scaled semi-major axis. Since our data is not strictly the same as that used for the inferences quoted elsewhere, a self-consistent analysis demands we derive these estimates ourselves, which forms another requirement for our work.

While our conversion equation assumes a circular orbit, if we fit the data under the same assumption, the relative estimate is at least self-consistent. Further, eccentric planets have smaller regions of stability for exomoons (Domingos et al. 2006) and have likely experienced scattering which decreases the chances for moons further (Gong et al. 2013). Thus, if the planet is eccentric, the incorrect conversion is likely irrelevant since such planets likely do not contribute OSE-signals into the grand light curve anyway.

#### 4.5. Two-Pass Detrending

In this work, we use the CoFiAM algorithm to detrend the *Kepler* light curves. We direct the reader to Kipping et al. (2013a) for details on the algorithm, but essentially its goal is to remove long-term trends without introducing any power, in a Fourier sense, at periodicities less than the transit duration. This requirement ensures that any signals with a time-scale approximately equal to or less than this duration will not be distorted by the detrending process itself, since a transit can be thought of as a Fourier series with the lowest frequency being that of the duration (Waldmann 2012). Accordingly, both the planet and moon transits are preserved, in contrast to polynomial-based methods which introduce power at all frequencies.

High frequency noise is not even attempted to be removed, but is monitored by measuring the autocorrelation at the cadence-lag and used to optimize both the harmonic filtering and subsequent identification of “bad transits”, which are typically rejected.

A disadvantage of CoFiAM is the requirement for a precise estimate of the time and duration of all transits in the time series. For this reason, it generally is not useful for blind searches for exoplanet transits. However, when seeking exomoons this requirement is generally true and indeed CoFiAM should be really thought of as an exomoon optimized detrending method more than anything else.

As before, transit times and durations are often available for the planets under consideration but those times and durations were conditioned on a different data detrending. In order to make our analysis self-consistent, we must derive these times and durations ourselves. Here-in lies a chicken-and-egg problem though, since to derive these times we first need detrended data, which itself first requires the times. To tackle this, we use two passes to iterate onto the best solution. The first pass uses the literature values for the times and duration

of the transits and then performs CoFiAM. The second pass takes the times and durations inferred using the first-pass data product, and then performs a new round of detrending with CoFiAM. This approach ensures both self-consistency and reliability in our estimates, and provides several opportunities to vet the data quality ensuring only the highest quality light curves are used in the final analysis.

## 5. DATA PROCESSING PIPELINE

### 5.1. Overview

As motivated in Section 4, we require a joint posterior distribution for the times and basic transit parameters of all planets used for the final analysis. Using a two-pass approach to detrending-fitting ensures that our inferences are self-consistent and conditioned upon the actual data product used in this work. In this section, we provide a detailed explanation of the data processing steps comprising each pass, which are tailored to the specific and unique goals of the two. A illustrative overview is provided in Figure 3 for reference.

### 5.2. Pass A

#### 5.2.1. PASS A GOALS

We first re-assert that the primary goals of pass A are to: 1) derive transit times for each transit 2) derive the transit duration for each planetary candidate. These products may then be used to conduct a second detrending later in pass B, since our detrending procedure (CoFiAM) requires the times and durations for execution.

These two objectives necessitate detrending of the data first, since we do a covariant detrending plus inference model which require an HBM, which is beyond the scope of this work for reasons described earlier.

For pass A, we make all decisions regarding data quality based upon whether we conclude that these two goals can be achieved. For example, in cases where there is plenty of out-of-transit data, but no in-transit, these would be rejected in pass A (but can be picked up later in pass B since such signals may still contain exomoon transits). Ultimately, defining a clear and independent objective for each pass allows us to optimize required steps.

In what follows, we describe the different stages of data analysis which are performed in pass A.

#### 5.2.2. STAGE A1

The first step is simply to download the *Kepler* SAP photometry for each target. We wrote a shell script to step through the target list (see Section 3), creating a local directory for each KOI. Using a WGET script, any and all of the target's LC data is then downloaded from MAST (from

[archive.stsci.edu/pub/kepler/lightcurves/](http://archive.stsci.edu/pub/kepler/lightcurves/)) and saved in the appropriate directory. As the data was downloaded in batches, the *Kepler* SOC data processing pipeline used for these files varied from 9.0.3 (2013-04-18 creation date) to 9.2.23 (2014-11-18 creation date).

Pre-prepared template detrending scripts are copied into each target directory. As with previous HEK papers, we use CoFiAM to perform our detrending. We direct the reader to Kipping et al. (2013a) and Kipping et al. (2013b) for details on the algorithm. We stress here that the algorithm is designed specifically for the exomoon problem and requires detailed initial information such as planetary periods, transit times and durations in order to work, which ultimately again explains why the two-pass data processing strategy is used.

We also use the *Kepler* Simple Aperture Photometry (SAP) time series throughout, since this time series has less chance of having unintended artificial signals present than the Pre-search Data Condition (PDC) time series (Stumpe et al. 2012; Smith et al. 2012), which has been subject to data massaging techniques already.

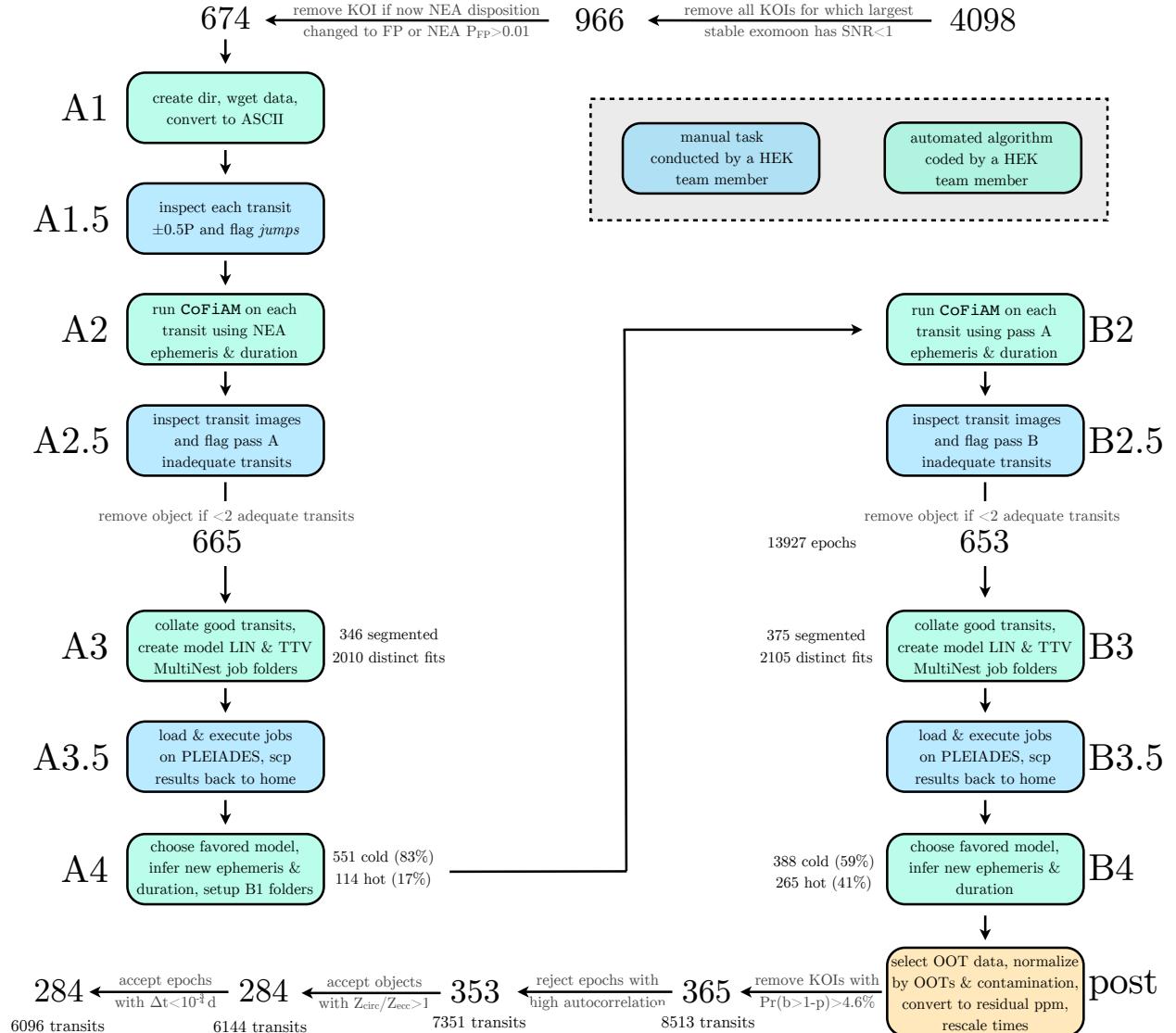
#### 5.2.3. STAGE A1.5

We manually inspect an image of each and every transit light curve epoch, centred upon the time of transit minimum expected from a linear ephemeris and including  $\pm 0.5$  orbital periods of data either side. We never attempt to stitch different quarters together and instead simply reject any data which occurs in a different quarter to that of the transit epoch under consideration. As before, ephemeris parameters are taken from the NASA Exoplanet Archive for this task.

At each epoch, we identify if any sharp jumps, exponential flux variations or any other anomalous light curve feature exists in the data. This process is performed by one of us with the perspective of whether CoFiAM would be able to fit the light curve variations or not. The aim is to keep a sufficiently long series of data to perform a robust detrending, but clip out patterns which may degrade the performance of CoFiAM, with an appreciation for the basis set which CoFiAM employs.

We initially pursued a variety of automated metrics for this purpose, such as standard deviation, autocorrelation and linear trends. However, we found that a wide variety of anomalous features survived and thus deceived these simple metrics. Rather than creating an ever larger battery of metrics, for which still no guarantee of completeness could be assured, we instead acknowledged that the human eye remains an unparalleled tool in quickly identifying anomalies in time series data.

Anomalous features are flagged by saving the instant in time just before/after the feature, depending on whether the feature occurs after/before the time of transit minimum. This process required approximately



**Figure 3.** Schematic diagram of the pipeline used to process the *Kepler* SAP photometry to ultimately construct the grand light curve. We color code the manual steps in blue and the automated steps in green. The left column broadly describes pass A, and the right column pass B. For details on each step, we direct the reader to the relevant subsection in Section 5.

60 hours of human labor in total.

#### 5.2.4. STAGE A2

Stage A2 involves the first detrending of the light curves. As mentioned earlier, this is performed automatically using CoFiAM and the list of anomalous features to ignore (found manually in stage A1.5). While details of CoFiAM can be found in Kipping et al. (2013a,b), we point out some general options selected for the execution in this paper.

The outlier threshold was set to  $3\sigma$  from a 20-point moving median. Before detrending, all planetary transits are removed with an exclusion window of  $\pm 0.6T_{14}$  for all events, including the object of interest.

Each transit is detrended separately using  $\pm 0.5P$  of

the data surrounding each event. If the transit epoch has any associated anomalous flux changes, as discussed in stage A1.5, then points beyond these times are also cropped.

Periodic functions described by a sum of harmonic cosines are explored from twice the data baseline down to twice the transit duration, with a cap of 30 harmonics (beyond which we tend to encounter numerical instabilities). This choice ensures that CoFiAM does not disturb the shape of the planetary transit in a Fourier sense (Waldmann 2012).

Each model is regressed to the data then ranked by the local autocorrelation, as computed using the Durbin & Watson (1950) statistic (DW) on the timescale of the

LC cadence, with the lowest autocorrelation being favored. The favoured model is then applied to original time series, re-including the planetary transit of interest. The final light curve is saved with a  $\pm 6.66T_{14}$  gap either side (which is the local timescale used for the DW calculation).

Finally, an image of every detrended transit is stored along with the best DW statistic.

### 5.2.5. STAGE A2.5

This is the second manual stage in pass A, where we manually identify “bad” transits. In some rare cases, CoFiAM fails and produces a light curve which cannot be used for fitting, due to visually evident trends remaining in the data. For example, if we missed a location of an anomalous flux variation in stage A1.5, CoFiAM may be trying to detrend sharp features with a smooth cosine function, producing a poorly detrended light curve. By manually going through the light curves in this way, it is essentially a second-check of the data quality, catching any missed anomalies from earlier.

In general, these assessments are made by one of us searching for any visually evident trends which would significantly impede our ability to fit the light curves to determine  $T_{14}$ . A bad transit does not necessarily have a poor DW statistic, although that tends to be a common scenario. Because we anticipate a second pass, we can be generous in considering acceptable data qualities at this stage. This acceptance level is non-constant, since we try to allow KOIs displaying very frequent bad transits to have at least a few transits which can be used for fitting in stage A3. Vice versa, if a KOI has many clean transits, we apply more stringent conditions in assessing data quality. Finally, we note that assessments are generally based on SNR of the transit, not the raw wiggles in the data, but the relative size of those wiggles compared to the transit. In cases of very low SNR, where the transit is not visible in a single epoch, we work under the assumption that sometimes data will wiggle up and sometimes down, but we must trust that on the average there is power (thus we try to allow for almost anti-transit like features in the interests of being unbiased and balanced). In such cases, our criteria switches from trying to make a good measure of the transit duration to simply avoiding “catastrophic” detrendings.

In addition to a “bad-transit flag”, we also use a “sparse flag” for transits where there is insufficient in-transit data (or none at all). In some rare cases, a third type of flag was used, “missing flag”, where the data are well-detrended, we have good temporal coverage, but a transit which should be visually obvious (given the transit depth) is missing in the data. We consider that these cases are most likely due to an erroneous transit ephemeris on the [NEA](#).

If fewer than two good transits remain for a KOI, the object is removed from our sample as being a useful object. In total, this removes 9 KOIs dropping our sample down from 674 to 665 KOIs.

### 5.2.6. STAGE A3

The third-stage is an automated shell script which begins by stitching the good transits together for each KOI into a single file. Our script then creates two directories for a linear ephemeris model fit,  $\mathcal{H}_{\text{LIN}}$ , and a transit timing variation model fit,  $\mathcal{H}_{\text{TTV}}$ , to be fed into [MULTINEST](#) (Feroz & Hobson 2008; Feroz et al. 2009). The script then queries the orbital period,  $P$ , and time of transit minimum,  $\tau$ , from the [NEA](#) database to construct priors for these terms. In the case of model  $\mathcal{H}_{\text{LIN}}$ , the prior on both terms is uniform centered on reported [NEA](#) value with  $\pm 1.0$  days window. For model  $\mathcal{H}_{\text{TTV}}$ , the period is treated as a fixed parameter with the individual transit times following a uniform prior centered on the expected time of transit for a linear ephemeris with a  $\pm 1.0$  day window again.

For model  $\mathcal{H}_{\text{LIN}}$ , the basic free parameters are  $P$  &  $\tau$ , two quadratic limb darkening terms (we use the  $q_1$  &  $q_2$  prescription from Kipping 2013), the ratio-of-radii,  $p$ , the impact parameter,  $b$ , and the stellar density,  $\rho_\star$ . Uniform priors are adopted for all except  $\rho_\star$  which follows a log-uniform prior. This gives a total of  $d = 7$  free parameters, which is easily handled by [MULTINEST](#). For model  $\mathcal{H}_{\text{TTV}}$ , the period is treated as fixed, removing one degree of freedom, but then each transit epoch has a unique  $\tau$  parameter (same prior as  $\mathcal{H}_{\text{LIN}}$ ), giving us  $d = 6 + N$  free parameters. For  $d \gtrsim 20$ , the performance of [MULTINEST](#) is severely impeded and global fits are not possible.

A common approach is “templating”, where one folds the transits, creates a template which is then regressed to the individual epochs (e.g. Holczer et al. 2016). This approach underestimates measurement uncertainties since it ignores the covariance between the transit shape parameters and the individual transit times. Rather than underestimating errors, we prefer to end-up overestimating them and so adopt a different strategy. Instead, we split up our light curves in segments of  $\sim 10$  epochs each, providing a manageable number of free parameters for each. The downside is that each segment is not able to utilize the information about the global transit shape learnt from other segments, and so the uncertainties will be larger (but more robust) than templating. Accordingly, in stage A3 our script automatically segments the data up where necessary.

We find a total of 346 out of our sample of 665 KOIs require segmenting, whereas the rest are treated in a single fit due to the tractable number of epochs available.

### 5.2.7. STAGE A3.5

Stage A3.5 was primarily performed using NASA’s PLEIADES cluster, and essentially involved loading, compiling, executing and then retrieving the over two thousands light curve regression jobs required. In total, we estimate that approximately  $\sim 100,000$  CPU hours were used during this phase of the analysis and spanned several months of wall time.

### 5.2.8. STAGE A4

In the fourth stage, we segue into pass B by laying the ground-work needed using the results from our light curve fits. The first task is to decide for each KOI whether it is dynamically “hot” or “cold”, by which we mean whether  $\mathcal{H}_{\text{LIN}}$  model (cold) or  $\mathcal{H}_{\text{TTV}}$  model (hot) is preferred. In cases where the fits were completed using a single segment, the evidences from MULTINEST can be directly used to compute the Bayes factor and rigorously assign the preferred model. For segmented models, direct evidence calculation is not possible since the TTV model has multiple copies of the same parameters for the transit shape. Instead, we use weighted linear regression to find the maximum likelihood linear ephemeris through the posterior transit times and then inspect the residuals for evidence of TTVs. This is simply done using a  $p$ -value test searching for a excessive  $\chi^2$  (cut off used was  $3\sigma$ ).

Formally, model assessments using a  $p$ -value are incorrect since they never actually compared to another model. More precisely, the  $p$ -value test is likely to frequently infer significant evidence for the alternative hypothesis even in cases where it should not. For example, excess noise from other sources or a single poor measurement could lead to the  $p$ -value test favoring the TTV model erroneously. Let us consider the effect of this by imagining fitting a linear ephemeris set of transits with the TTV model. The times of transit and basic transit parameters will all still come out formally correct, just with inflated uncertainties. Giving each epoch a free transit time is still able to recover the original linear ephemeris solution. Therefore, despite the  $p$ -value’s tendency to overestimate significance, this merely acts to conservatively inflate our uncertainties and does not formally invalidate our inferences.

In total, we find 551 of the 665 KOIs are “cold”, with the remaining 114 being dubbed hot. For comparison to later, we point out that one might expect pass B to increase the hot fraction due to the improved detrending and thus greater sensitivity to even small TTVs.

Stage 4 ends by duplicating all of the KOI folders into a new directory with a small text file recording the favored model and ephemeris parameters. The transit duration is also recorded in this file, where for  $\mathcal{H}_{\text{LIN}}$ -favored KOIs is computed directly from the joint poste-

riors and for  $\mathcal{H}_{\text{TTV}}$ -favored KOIs from a weighted sum of each segment’s marginalized credible interval for the duration.

## 5.3. Pass B

### 5.3.1. PASS B GOALS

Before describing each data processing step for pass B, we first re-assert the objectives, which represents the backdrop against which all decisions in pass B are framed.

Ultimately, the data product from pass B should be high-quality, cleaned light curves with accurate estimates of the transit times and basic parameters needed for stacking and re-scaling. Our tolerance for poor-quality light curve is necessarily more stringent here, since unlike pass A, there is second-chance for these light curves and they have to be of sufficient quality for stacking by the time pass B is complete.

### 5.3.2. STAGE B2 & 2.5

Downloading the data (stage 1A) and removing jumps (stage A1.5) does not need to be repeated since the raw data product is unchanged. Accordingly, we skip straight to stage B2. Mirroring stage A2, we detrend the SAP light curves using CoFiAM but now using the duration and transit times determined earlier in stage A4 (specifically we use the maximum *a posteriori* values).

In stage 2.5 we again inspect these light curves visually for poorly detrended examples and find 12 KOIs end up with fewer than two usable transits after this process. Removing this objects reduces our sample from 665 to 653 KOIs in what follows.

### 5.3.3. STAGE B3 & 3.5

As with stage A3, stage B3 collates the good transits and sets up folders for models  $\mathcal{H}_{\text{LIN}}$  and  $\mathcal{H}_{\text{TTV}}$  ready for fitting with MULTINEST. Once again this results in just over two thousand distinct jobs to run, with 375 of the KOIs being segmented.

Unlike stage A, we here allow the light curve model to account for any known blending from nearby sources. These are collated from Everett et al. (2015); Kolbl et al. (2015); Adams et al. (2012, 2013); Dressing et al. (2014); Law et al. (2014), and we use *Kepler*-converted magnitudes to estimate the appropriate contamination factor for each band. These blending factors are treated as Gaussian priors, with a standard deviation set by the uncertainty on the converted magnitude. In total, 39 of the targets required a blending term to be included.

In stage 3.5, we again load, compile, execute and retrieve these runs on PLEIADES, requiring another round of 100,000 CPU hours and several months of wall time.

### 5.3.4. STAGE B4

Finally stage B4 repeats stage A4 with the new light curve fits, performing model selection using the same framework described earlier. After completion, we found that the fraction of cold KOIs indeed decreased as expected from 83% to 59%, giving 265 hot KOIs in our sample. In this work, these TTVs represent purely a nuisance but we acknowledge that this data set represents a rich and interesting catalog for others in the community. We therefore make all of the transit times, both for the hot and cold samples, publicly available at [github.com/alextachey/TTV\\_posteriors](https://github.com/alextachey/TTV_posteriors).

For each KOI, we export the maximum *a posteriori* transit fit and the corresponding vector of out-of-transit baseline fluxes (OOTs), which are found by linear minimization of the maximum *a posteriori* model and the data (this is actually done on the fly during every regression step, following the approach using by [Kundurthy et al. 2011](#)). These OOTs will be useful later for stacking the final light curves.

## 5.4. Post-Processing

In post-processing, we aim to export a single file for each KOI which contains a phase-folded light curve suitable for planet-stacking. First, the transit times are removed using the favored model and the maximum *a posteriori* parameter vector. Next, the planetary transit is removed by excluding all data which falls within the  $2\sigma$  upper limit of the derived full duration,  $T_{14}$ . Each epoch is then corrected for any residual DC power detected by the OOT vector regressed during Stage B4. Global blending factors, as well as quarter-to-quarter aperture flux contamination factors are corrected for following the approach described in [Kipping & Tinetti \(2010\)](#).

Next, we subtract unity from the normalized fluxes and multiply by a million to create a ppm residual light curve for each object. Finally, the time column is converted to  $t'$  using Equation (8) and the maximum *a posteriori* transit parameters derived from the preferred model regressed in Stage B4.

## 5.5. Filtering

Before stacking these planets together, we first remove KOIs and individual transits which fail to satisfy several criteria. First, we require a  $2\sigma$  confidence that the planet is not a grazing event, which would mean that  $b > (1 - p)$ . Grazing events have degenerate planetary radii and could be far larger in size, potentially even a false-positive. Erring on the side of caution, we remove any such KOI which filters out 288 KOIs, leaving us with 365 targets.

We next test for excessive autocorrelation using the DW metric. For each transit, we generate 10,000 mock

realizations where the data are drawn from perfect normal distributions at the exact same sampling observed in the data and the reported uncertainties. We use these synthetic transits to generate an expected distribution for the DW metric and flag any transits for which the real DW metric is more than two standard deviations away from our synthetic population. If a KOI has 50% or more of its transits flagged in this way, the entire KOI is dropped from the sample. This filtering removes a further 12 KOIs.

Next, we compare the transit derived stellar density (which assumes a circular orbit) to an independent estimate, in order to exploit the photoeccentric effect ([Kipping et al. 2012](#); [Dawson & Johnson 2012](#)) to infer a minimum eccentricity,  $e_{\min}$ , of each KOI ([Kipping 2014](#)). We draw a random sample from the transit derived posterior found in stage B4 and divide it by a random sample drawn from the KOI's corresponding stellar density posterior derived in [Mathur et al. \(2017\)](#). This ratio is then converted into a minimum eccentricity using Equation (39) of [Kipping \(2014\)](#), and the process is repeated until we have derived 40,000 posterior samples for  $e_{\min}$  for each KOI. For each KOI, we also construct a prior for  $e_{\min}$  based off the prior used in the transit fits and [Mathur et al. \(2017\)](#) distribution.

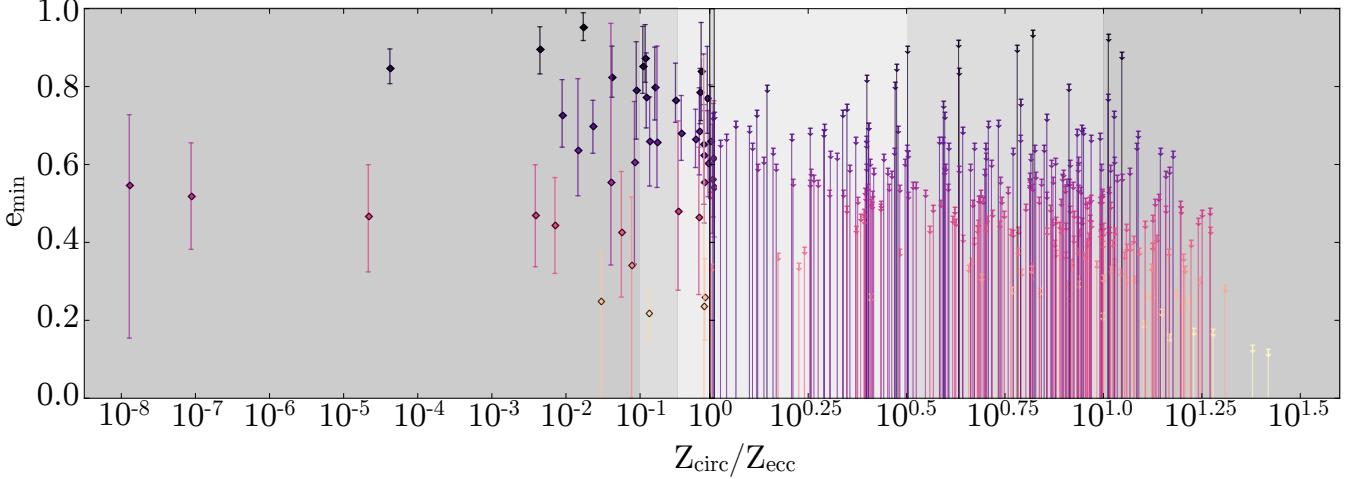
We next evaluate the Savage-Dickey ratio between the posterior and the prior to estimate the Bayes factor,  $Z_{\text{circ}}/Z_{\text{ecc}}$ . We find that 284 KOIs have  $Z_{\text{circ}}/Z_{\text{ecc}} > 1$ , implying a near-circular orbit, whereas the other 69 KOIs are rejected for further analysis, on the basis that eccentric planets likely result from scattering which would disrupt moon systems ([Gong et al. 2013](#)).

Finally, we elected to remove transits for which we are unable to measure the transit time to within a precision of  $10^{-0.75}$  days, which is necessary to ensure we are able to reasonably phase-fold transits together. This did not change the number of KOIs from 284, but did reduce the number of transits in our sample from 6144 to 6096.

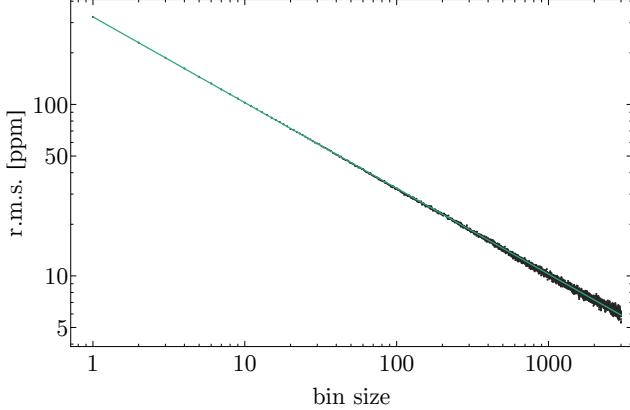
## 5.6. Constructing a grand light curve

With each target now having a fully processed phase-folded light curve, we are ready to finally stack different targets together to create a grand light curve. This stacking can be across all 284 surviving targets, or a subset of them, as explored later. Although we here describe the case for the complete ensemble, the planet-stacking procedure follows the methodology described here when dealing with subsets.

Across the 284 KOIs, we have 6096 unique transits comprising of 364059 photometric measurements. The re-scaled times are well-described by a half-normal distribution with a standard deviation of 113. We elect to remove any points which fall outside of the range  $t' > 150$ , leaving us with 309750 points.



**Figure 4.** Minimum eccentricity of 353 KOIs derived using the photoeccentric effect as a function of the Bayes factor for a circular vs eccentric orbit. KOIs to right of unity are depicted as upper limits on eccentricity, whereas we plot  $1\sigma$  credible intervals for the others. The 284 KOIs favoring a circular orbit are considered further as viable exomoon candidates in this work.



**Figure 5.** RMS of the grand light curve as a function of bin size, demonstrating the expected  $N^{-1/2}$  scaling of white noise.

The grand light curve photometry shows no evidence for correlated noise structure, as expected from averaging so many independent data sets together. This is verified in Figure 5, where we plot the root mean square (r.m.s.) of the time series as a function of a bin size, which displays excellent agreement with the expected root  $N$  scaling.

Dividing the fluxes by their formal reported uncertainties, we find that the robust r.m.s. (given by 1.4286 multiplied by the median absolute deviation) equals 1.09, indicating only a small amount of extra noise above the reported uncertainties. We re-scaled the errors by this factor and then performed  $3\sigma$  clipping, removing just under one percent of the points. The final time series is found to have a standard deviation of 5.1 ppm when binned to a scale of  $\Delta t' = 0.5$ .

## 6. MODELING

### 6.1. Overview

In addition to manipulating the *Kepler* data to construct the grand light curve, we also require the ability to model its shape, as a function of various exomoon parameters of interest.

Initially, we operated under the assumption that the OSE expressions of Heller (2014) would be suitable for this purpose. Indeed, this was the model used in Hippke (2015). However, detailed testing revealed that the OSE expressions were not suitable for our purposes, with severe disagreement being found between the Heller (2014) predictions and photodynamical tests executed in this work in certain regimes. We attribute the differences due to at least three effects, which we describe in what follows:

- The “Callisto effect”,
- Inter-population variation,
- Heteroscedastic weighting.

### 6.2. Photodynamical Tests of OSE

Heller (2014) derives an analytic formula for the shape of the phase-folded planet-stacked light curve of an exomoon, which for the out-of-transit component considered in our work may be expressed as:

$$f_{OSE}(t') = \begin{cases} -\left(\frac{(R_S/R_\star)^2}{\pi}\right) \cos^{-1}\left(\frac{t'}{a_{SP}/R_P}\right) & \text{if } t' > \frac{a_{SP}}{R_P}, \\ 0 & \text{otherwise.} \end{cases} \quad (9)$$

where  $t'$  represents our re-scaled time axis, equal to the separation of the moon from the planet in planetary radii. We note that Heller et al. (2016) provide a slightly modified version accounting for limb darkening, although this does not impact the following discussion.

A different light curve model (called LUNA) was previously presented in Kipping (2011), which provides an analytic model for the light curve morphology of each individual planet-moon transit computed “photodynamically”. Photodynamics, a test first coined in Carter et al. (2011), refers to a light curve simulator which evolves a planetary system at each time step and computes the corresponding shadows cast onto the sky-projected stellar disk. Unlike the OSE model of Heller (2014), this model is not specific to phase-folded events but of course can be easily used to simulate such a case by simply folding the final predicted light curve. We also highlight that LUNA has the ability to simulate arbitrary eccentricities and three-dimensional orbital configurations, as might be expected for real moon systems.

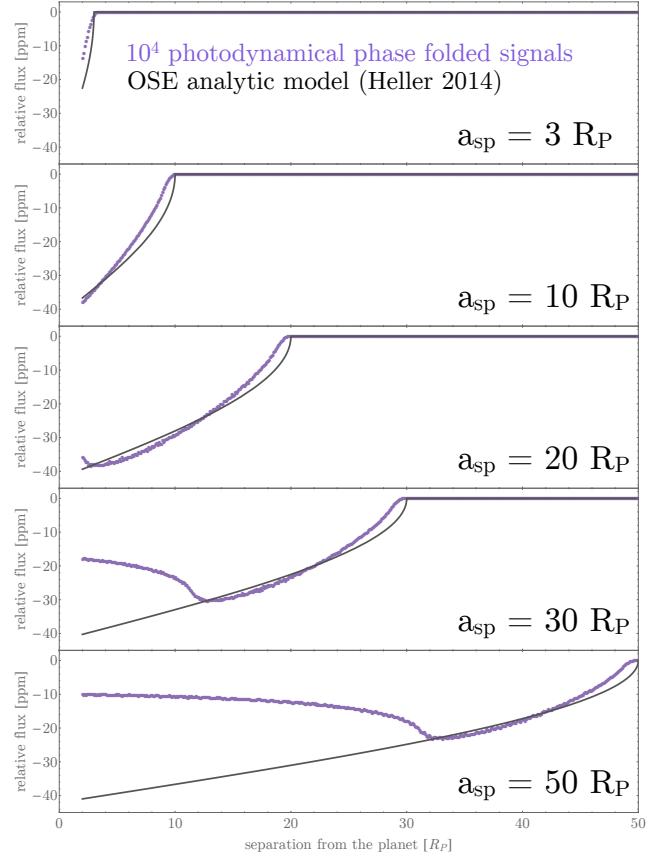
In general, LUNA provides a physically detailed light curve simulation, but comes at the expense of greater computational cost than the simple closed-form expressions of OSE. For these reasons, if the accuracy of OSE is validated, it would be far simpler and thus preferable to employ the OSE formalism for our model fits of the grand light curve. However, for several reasons described in what follows, we came to conclude that OSE would not be an accurate modeling tool for our data product.

### 6.2.1. The Callisto Effect

We first consider the simple case case of a  $P = 100$  d Jupiter-like planet transiting a Sun-like star with an Earth-like moon. The Jupiter-like planet has a impact parameter of  $b = 0.458$  (the expectation value of the impact parameter prior; Kipping & Sandford 2016), all orbits are perfectly circular, the moon is perfectly coplanar and the star has quadratic limb darkening coefficients of  $q_1 = 0.4$  and  $q_2 = 0.3$  (see Kipping 2013).

We simulate  $N = 10,000$  consecutive transits of the planet observed by *Kepler*, which would correspond to a time series of 2738 years. The data is assumed to be long-cadence, uniformly sampled, contains no gaps and the time series has zero noise. The simulations are then phase-folded upon a linear ephemeris (the moon’s mass is set to zero to prevent TTVs), the planetary transit is removed, the temporal axis is re-scaled and then the remaining data is binned to a window size of  $t' = 0.1$ .

This represents a completely ideal case for OSE, since the data are noise-free, perfectly sampled and we certainly satisfy  $N \gg 1$  transits. In this highly controlled regime, we should expect OSE to provide excellent agreement to the photodynamical simulations.



**Figure 6.** Illustration of the Callisto effect, where even under ideal conditions we find that moons with  $(a_{SP}/R_P) \gtrsim 20$ , such as Callisto, show significant disagreement between the predictions of OSE (Heller 2014) and the photodynamical simulator LUNA (Kipping 2011).

We repeated the simulation 393 times, consecutively increasing the semi-major axis of the moon from  $(a_{SP}/R_P) = 2$  to  $(a_{SP}/R_P) = 100$  in 0.25 steps. For  $(a_{SP}/R_P) \lesssim 20$ , we find excellent agreement between OSE and the folded LUNA simulations, as expected, but beyond this point significant disagreement appears. As this work is primarily interested in Galilean-analog moons, we note that Io, Europa and Ganymede all fall interior to this point and thus are well-described by OSE. However, Callisto with  $(a_{SP}/R_P) = 27$ , is an example where significant tension exists between photodynamics and OSE, which is why we dub this the “Callisto effect”. Figure 6 depicts the large discrepancies which emerge beyond 20 planetary radii.

### 6.2.2. Inter-population variation

The OSE model is derived assuming one co-adds many transits of the same planet-moon system i.e. that the basic parameters of the system are not changing. However, in our case we co-add different systems together which have distinct planet and moon parameters. For example, in our Galilean moon fits, we assume that the

moons have inclinations representative of Io, Europa, Ganymede & Callisto, but each planet's moon system will have unique moon inclinations, despite being drawn from a common underlying distribution. While we could co-add many OSE signals together modeled individually with the corresponding parameters, each OSE curve would be modeling only a small number of transits and thus would be formally invalid - since it is by definition an ensemble model. Without detailed investigation, it was unclear that one could simply co-add across a population in this way and recover the correct phase stacked signal and thus we preferred to use LUNA which provided an accurate model of the individual events.

#### 6.2.3. Laplace Resonances

A subtle and minor point of concern was dealing with the Laplace resonance in the OSE framework when modeling Galilean analogs. In LUNA each individual moon transit is generated and thus we are able to assign relative phases between the satellites such that they reside in not only the correct mean motion resonance but also share the Laplace resonance in terms of their mutual phases:  $\pi = \lambda_{\text{Io}} + 3\lambda_{\text{Europa}} + 2\lambda_{\text{Ganymede}}$ . In contrast, the OSE framework never models the individual events, rather just the ensemble, and we were unable to demonstrate that OSE was correctly accounting for such a phase lock.

#### 6.2.4. Heteroscedastic weighting

Finally, OSE is an average of light curves, which by definition means each light curve is given precisely the same weight. Second, each light curve contribution is assumed to uniformly and densely sampled. Our real data products do not satisfy such constraints, since first we co-add the different planets together using weights based off the root mean square of the photometric residuals. Second, light curves are not uniformly sampled, featuring data gaps and removed outliers, as well as being non-uniformly transformed in time via our temporal re-scaling. Since LUNA models individual events, data gaps, integration time effects (Kipping 2010b) and re-scaling is easily accounted before applying any co-addition, enabling us to ensure our model is representative of the data.

### 6.3. Photodynamic Look-Up Tables (LUTs)

As a result of the myriad of complicated effects influencing the final model yet the relative low-dimensionality of the model itself, we elected to build a grid of pre-computed models which accounted for all of the effects described above. For each KOI, we took the maximum *a posteriori* transit parameters from the planet-only fits (using the favored model) and generated a planet+moon light curve using LUNA with the same

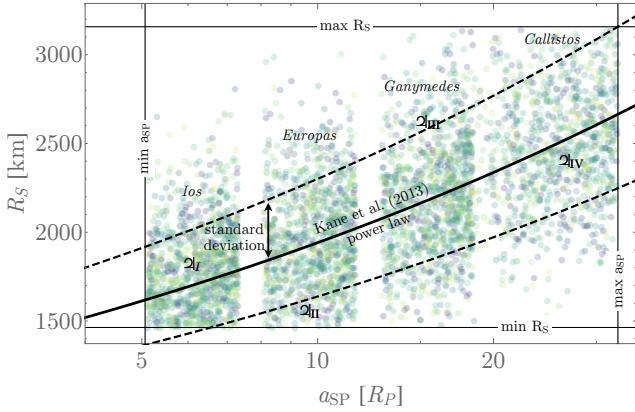
planet parameters but adding in one or more moons. The model curve is generated at precisely the same cadence as the data used for the planet-stack and accounts for the long-cadence integration time using 1 minute numerical re-sampling (Kipping 2010b). After all of KOI model light curves have been computed for a specific choice of underlying moon population, they are co-added with the same weighting used for the real data. In other words, we inflict precisely the same transformations to the model as we do to the data, to ensure a like-for-like comparison at the end.

#### 6.3.1. Galilean Analogs

The moons are generated in two ways. The first case was for a Galilean analog. Here, we assume that four moons orbit each KOI with properties resembling those of Io, Europa, Ganymede & Callisto. To inject some stochastic variation between each moon, yet maintain the 1:2:4 resonance of the inner three, we randomly place each Io-analog to have a semi-major axis of  $(a_{SP}/R_P) \sim \mathcal{U}[0.8 \times 6.1, 1.2 \times 6.1]$ , where 6.1 is the actual value for Io around Jupiter and  $\mathcal{U}[a, b]$  is a uniform distribution from  $a \rightarrow b$ . The next moon along is then assumed to lie in a 2:1 resonance, such that  $(a_{SP}/R_P)_{\text{Europa}} = 2^{2/3}(a_{SP}/R_P)_{\text{Io}}$ , and similarly for the Ganymede-analog with respect to the Europa-analog. The semi-major axis of the Callisto-analog, which does not reside in the resonance chain, is then randomly drawn as  $(a_{SP}/R_P)_{\text{Callisto}} \sim \mathcal{U}[3 + a_{SP}/R_P]_{\text{Ganymede}}, 1.2 \times 27.2]$ , where 27.2 is the actual value for Callisto around Jupiter. Any moon systems generated where two moons have semi-major axes within 3 planetary radii of each other are rejected.

To generate moon radii which were stochastic yet representative, we adopted the radius-separation power-law model of Kane et al. (2013). The authors note that the radii of moons tend to increase with respect to semi-major axis following a power-law model. We took the four Galilean moons in isolation and calibrated a least squares power-law model to it, giving  $\log(R/R_\oplus) \sim \mathcal{N}(6.95 + 0.27 \log(a_{SP}/R_P), 0.17)$ , where the standard deviation quoted is that resulting from the residuals of the best-fitting line. To protect against peculiar draws, we required that the quadrature sum of the radii was within 20% of the actual sum for the Galilean system, that the minimum radius moon was at least 80% the radius of Io and the maximum radius moon was no more than 120% the radius of Ganymede. The resulting covariant distribution is illustrated in Figure 7

To go from physical radii,  $R_S$ , to  $R_S/R_P$  needed for the modeling (or equivalently  $R_S/R_\star = (R_S/R_P)(R_P/R_\star)$ ), we used planetary radii derived in Chen & Kipping (2017b) to make the conversion. We also included the slight TTV and TDV effects induced



**Figure 7.** One thousand randomly generated Galilean-moon system analogs using the method described in the main text. Each color represents a unique system comprising of four moons.

by the Galilean moons by including  $(M_S/M_P)$ , which was computed by using `forecaster` (Chen & Kipping 2017a) applied to the moon radius to predict a mass, and then using the physical planetary masses predicted in Chen & Kipping (2017b).

The mean longitude of the inner two moons are randomly generated uniformly but the third is enforced to satisfy a Laplacian resonance. The inclination of the moons are randomly drawn from a Von Mises distribution with  $\kappa = 42637$ , which we found to maximize the likelihood of a Von Mises distribution conditioned on the real Galilean moons. Eccentricities were kept fixed at zero and the moons all follow perfect keplerian orbits i.e. we do not model gravitational interactions between the moons.

After stacking the resulting model light curves with correct cadence and weightings, we tried varying the fraction of systems which harbor moons. Treating each system as a Bernoulli experiment with a probability of having a moon system given by  $\eta$ , we found varying  $\eta$  was equivalent to simply scaling the  $\eta = 1$  resulting light curve by the same factor. Having demonstrated this, we were able to exploit it to aid in later fits.

In cases where a subset of systems were modeled, the process described above was repeated creating unique model light curves for each specific subset. Generally, the shape of each resulting light curve were very similar, but ended up with different amplitudes as a result of the differing weights, data gaps and stellar radii for the host stars. As noted these factors represent data specific properties for each subset and were saved for later use with the single moon simulations.

### 6.3.2. Single Moons

While a single moon is four times quicker to generate than four Galilean moons, the Galilean moon follow an

expected distribution in terms of their sizes and orbital semi-major axes. For a single hypothetical moon, we have no idea what these properties are *a priori* and thus our grid cannot simply span  $\eta$ , as before, but now must also span  $R_S$  and  $a_{SP}$  leading to a three-dimensional look-up table. Fortunately, the effect of  $\eta$  is a simple scaling and thus can be applied easily during the fits themselves, yet this still means we need to generate a two-dimensional grid of models, rather than just a single look-up example in the case of the Galilean-analogs.

We setup a logarithmically-spaced grid from  $R_S = 0.2 R_\oplus$  to  $R_S = 2.0 R_\oplus$ , with 16 unique grid points. For  $a_{SP}$ , our grid is again logarithmic, defined as  $a_{SP} = 2^x$  where  $x$  is stepped through from 1 to 7 in 0.1 steps, leading to a total grid size of 976 elements. The moon is treated as being exactly coplanar and circular with random phase and as before we generate unique light curves for each KOI and then co-add with the appropriate weightings to create our final models. For these simulations we set  $(M_S/M_P) = 0$ , since some simulations permit very massive moons which would cause noticeable TTVs, which would then be subsequently removed anyway by our data processing pipeline described earlier.

When dealing with subsets, we apply the scaling factors found earlier with the Galilean-analog experiments, since the computation time to create the grid required many weeks. During the actual fits, we used a bi-linear interpolation of every unique binned photometric data point, conditioned upon the LUT. We also added an extra grid point of  $R_S = 0.0 R_\oplus$ , corresponding to a flat line, to provide numerical stability if fits attempted to compute the likelihood of radii below our  $R_S = 0.2 R_\oplus$  limit.

## 7. ANALYSIS

### 7.1. Galilean Global Fits

We first discuss our results from regressing our photodynamical phase-folded planet-stacked planet+moon light curve models (see Section 6) to all 284 KOIs deemed to be of suitable quality for our analysis (see Section 3). As discussed, the fits are conducted for two different light curve models, a Galilean-analog and a single moon.

For the Galilean-analog, the only parameter directly affecting the light curve model is  $\eta$ , the fraction of KOIs which harbor a Galilean-analog. In addition, we added two other free parameters into our fits. The first was to account for excess photometric scatter,  $\sigma$ , and was simply added in quadrature to our derived uncertainties in the planet-stacked light curve. The second was an offset term,  $\gamma$ , to allow for a re-normalization of the data set. Whilst  $\eta$  and  $\gamma$  were assigned uniform priors,  $\sigma$  was as-

sumed to follow a log-uniform prior from 0.1 to 10 ppm. The regression was performed using MULTINEST with 1000 live points and typically completed within a few minutes on a laptop. The fits were repeated ten times each, from which the posteriors were combined. Since MULTINEST estimates the marginal likelihood, we repeated our fits with  $\eta$  fixed to zero and removed as a free parameter, giving us a direct estimate of the Bayes factor,  $B_{SP}$ , for the moon model.

We find that the null model is slightly favored, with  $\log B_{SP} = -0.84$ , or an odds ratio of 2.3-to-1 preference for the null model. The resulting light curve and model fitting lines are shown in Figure 8, and the associated posterior distribution is plotted in Figure 9. Our results imply that  $\eta < 0.38\%$  to 95% confidence for the 284 KOIs considered, with a 68.3% confidence interval of  $\eta = 0.16^{+0.13}_{-0.10}$ .

### 7.2. Single Moon Global Fits

As discussed in Section 6.3.2, the single moon case required on-the-fly interpolation of a look-up table for the likelihood calls. For this reason, we found it more practical to use a Markov Chain Monte Carlo (MCMC) algorithm instead of MULTINEST. Our regression was performed using a simple MCMC that we wrote, which used the Metropolis rule for sampling and normal proposal functions tuned by hand to give a  $\sim 50\%$  acceptance rate. Ten independent chains were seeded from random locations within the prior volume, all of which converged within 100 steps, and were then allowed to propagate for  $10^5$  accepted steps, giving  $\sim 10^6$  posterior samples in total.

In addition to the three free parameters used in the Galilean fit, we added in a parameter controlling the semi-major axis of the moon, ( $a_{SP}/R_P$ ), and the moon size in Earth radii, ( $R_S/R_\oplus$ ). Both were assumed to follow log-uniform priors spanning the limits in our LUT. The  $\eta$  term was allowed to follow a log-uniform prior spanning 0.01 to 100, since technically it represents an effective moon in this case and thus should be interpreted as the average number of moons per system.

The fits converged to a solution of  $R_S = 0.51^{+0.59}_{-0.23} R_\oplus$  and  $(a_{SP}/R_P) = 6.3^{+7.6}_{-3.1}$  for  $\eta = 0.43^{+0.33}_{-0.28}$ . This fit does not directly return a marginal likelihood, since MCMC was used, nor is the Savage-Dickey ratio suitable given that three extra covariant free parameters have been added. However, the kernel-approach shown later reveals that the evidence favoring the single moon fit in this region is modest at  $B_{SP} \simeq 2$ .

### 7.3. Single Moon Kernel

While the single moon fit is useful for identifying the maximum *a posteriori* region of parameter space, it does not provide a clear view of the overall likelihood trends

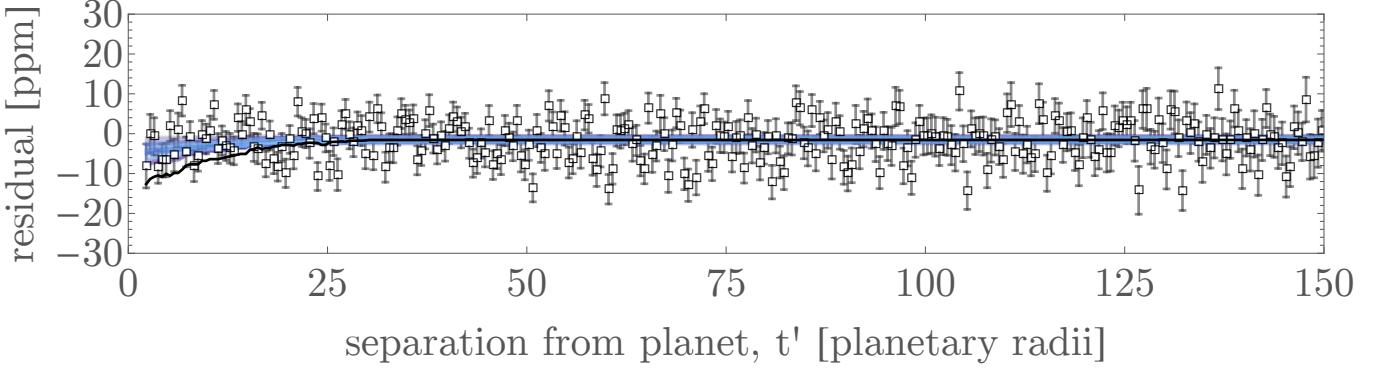
occurring within the prior volume. To address this, we repeated our fits for a single moon but fixed  $a_{SP}/R_P$  and  $R_S$  to a specific choice and just regressed  $\eta$ ,  $\sigma$  and  $\gamma$ . Since no interpolation was necessary, it was straightforward to use MULTINEST with  $\eta$ -rescaling on a single interpolated model each time.  $R_S$  was varied across a grid from 0.02 to 2 Earth radii in 100 log-evenly spaced steps. Similarly,  $a_{SP}/R_P$  was varied from 2 to 100 in 100 log-even steps.

At each point, we derived a three-dimensional joint posterior distribution and marginalize over  $\sigma$  and  $\gamma$  to directly measure the occurrence rate of exomoons at each location. The posterior derived is mathematically equivalent to

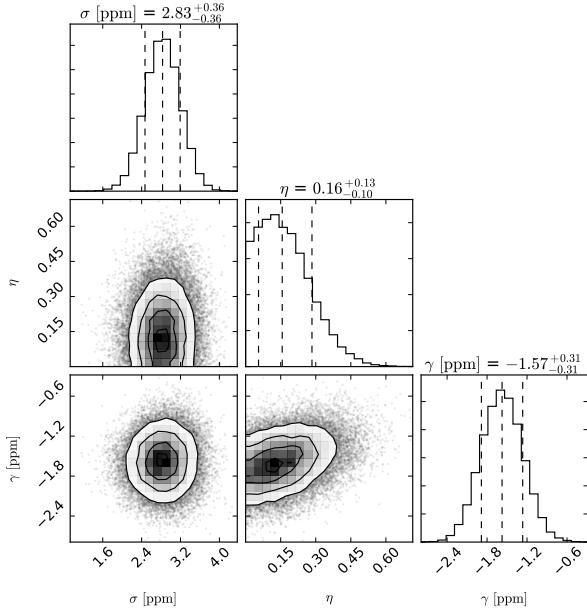
$$P(\eta|R_S, a_{SP}/R_P) = \int \int P(\eta, \sigma, \gamma|R_S, a_{SP}/R_P) P(\sigma) P(\gamma) d\sigma d\gamma \quad (10)$$

In addition to deriving a posterior at each grid point, we also estimate the evidence against the null model, allowing us to compute the Bayes factor. In Figure 10 we plot the Bayes factor and exomoon frequency for the ensemble as a function of effective moon radius  $R_S$  and semi-major axis  $a_{SP}$ . The Bayes factor (left) indicates whether the moon model is favored over the model without a moon. Red represents regions of parameter space where the moon model is disfavored, while green regions are areas where the moon model is favored, and intensity represents our degree of confidence in that model selection. We emphasize to the reader that paying attention to the contours in this plot is essential for an accurate interpretation; while much of the plot appears green, the moon model is in fact only weakly favored ( $B_{SP} \simeq 2$ ) on a small island in parameter space. By contrast, large values of  $R_S$  and  $a_{SP}$  are strongly ruled out ( $B_{SP}$  around 0.01). A value of 1 in this plot means we can make no statement about one model being a better fit to the data than the other.

The right side of Figure 10 should then be read in the context of the left side. The similarity between the contours on both sides is readily apparent. For large values of  $R_S$  and  $a_{SP}$  we find an exceptionally low occurrence rate, while at the lower end of these variables the occurrence rate shoots up. It is tempting to read this as a moon signal, but in the context of the Bayes factor on left it is clear these occurrence rates are not at all constrained or well supported by the evidence. Only in regions of high confidence ( $B_{SP}$  much greater or much less than 1) should the exomoon frequency values be given much credence. It is worth point out, perhaps, that there is little qualitative difference between a very low exomoon occurrence rate and a very low value for  $B_{SP}$ . Both are consistent with virtually no signal in this



**Figure 8.** Phase-folded planet-stacked light curve of all 284 KOIs deemed to be of acceptable quality. Temporal axis has been re-scaled and binned, with uncertainties shown given by the standard deviations within each bin. Black solid line represents the expected signature if  $\eta = 100\%$  of the KOIs had a Galilean-analog moon system. Blue lines show 100 posterior samples from our fits, giving  $\eta = 0.16^{+0.13}_{-0.10}$ .



**Figure 9.** Corner plot of the three-parameter joint posterior distribution from our Galilean-analog moon fit. This fit was disfavored over a null fit with a Bayes factor of  $B_{SP} = 0.43$ .

region of parameter space.

#### 7.4. Evidence for a Population of Super-Ios?

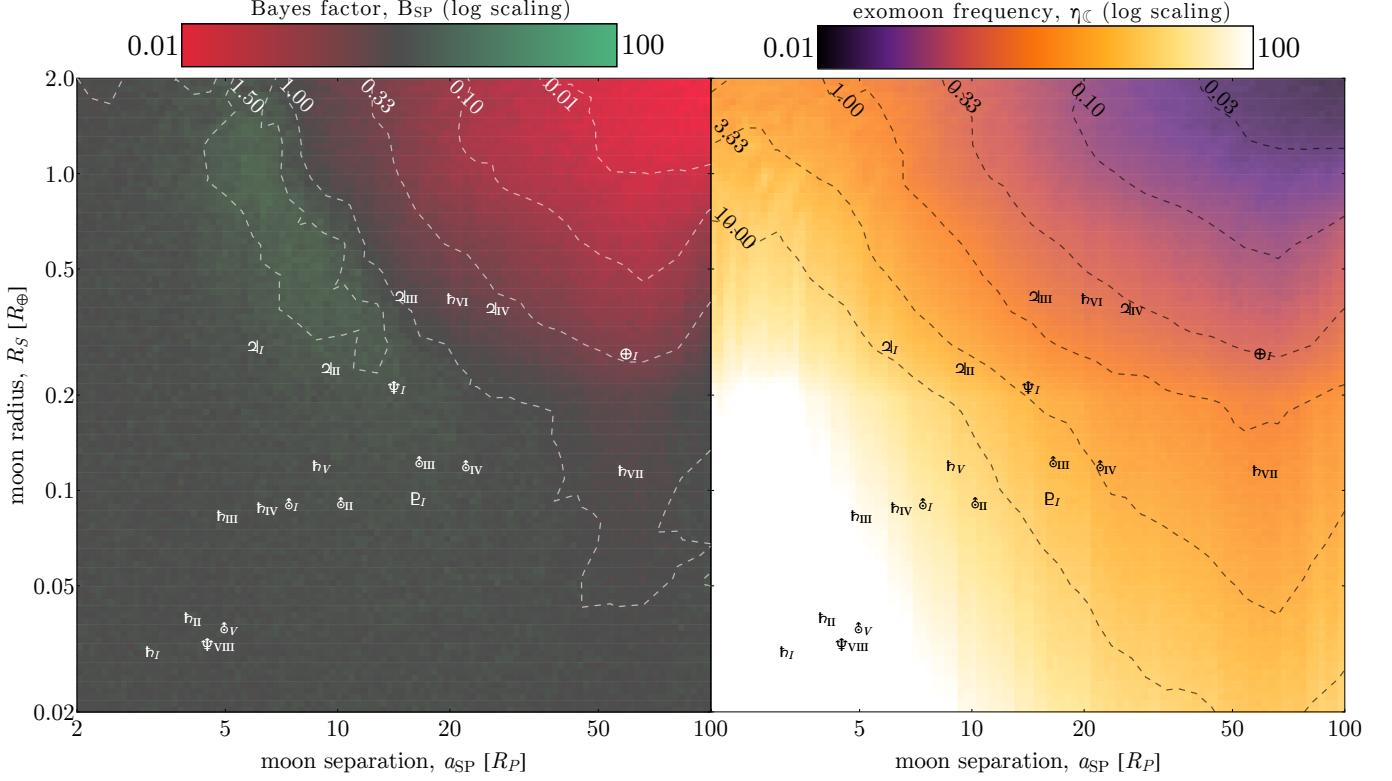
The island on the left side of Figure 10 where  $B_{SP} > 1.50$  is intriguing, if only marginally significant. We hesitate to make any strong statement about this region of parameter space, but it is worth pointing out that theoretical modeling (e.g. [Namouni 2010](#)) suggests that while planets migrating inward will tend to lose their moons in the process due to a shrinking Hill sphere, they are more likely to retain moons orbiting closer to the host planet (that is, close-in moons tend to survive planetary migration to smaller  $a_P$ ). Recall that we are probing planets

within roughly 1 AU of their host star, suggesting that a large fraction of these planets may have migrated from beyond the snow line. The island of modest moon signal could therefore be evidence (albeit inconclusive) of a population of moons that have survived migration by virtue of their separation from their host planet. Note, however, that more recent theoretical work ([Spalding et al. 2016](#)) suggests by contrast that satellites closer in to their planet ( $a_{SP} \lesssim 10R_P$ ) are also vulnerable to dynamical moon-loss during migration. It is unclear, then, whether one or both of these mechanisms could be at play here, and indeed, the strength of these mechanisms rely in part of the evolutionary history of each system which will of course be unique. In any case, our results suggesting a dearth of exomoons at small  $a_P$  appear to provide observational support for the findings of both [Namouni \(2010\)](#) and [Spalding et al. \(2016\)](#), and more broadly, could be evidence of giant planet migration.

#### 7.5. Subset Fits

In addition to fitting the entire sample with effective moon and Galilean analog GLC curves, we also performed GLC fits on a number of physically motivated subsets. The aim here was to identify whether a certain class of planets in the sample preferentially hosts moons over another. As such we divided the sample into several equally-populated pairs and fit the GLC curve to these subsets. These pairs were small/large planets, cold/hot planets, cold/hot stars, and inner/outer planets. We also split the sample into single/multi-planet systems, and habitable zone/non-habitable zone planets (most of the latter residing inside the innermost edge of the habitable zone). These last two categories, of course, are not equally populated.

The results for Galilean moon fits can be seen in Figure 11 and Table 1. The thick black line in the figure represents the peak posterior value, while the light red lines



**Figure 10.** *Left:* Heatmap of the Bayes factor  $B_{SP}$  as a function of single effective moon radius  $R_S$  and semi-major axis  $a_{SP}$  for the ensemble. Red indicates regions of parameter space where the moon model is disfavored, while green represents regions where the moon model is favored. Greater color intensity corresponds to greater confidence in the selected model. *Right:* Exomoon frequency in the ensemble as a function of  $R_S$  and  $a_{SP}$ . A collection of solar system moons are plotted for context.

Galilean Analog Subset Fits							
Group	$\eta_{\mathbb{C}} [1\sigma]$	$\eta_{\mathbb{C}}^{\max} [2\sigma]$	$B_{SP}$	Group	$\eta_{\mathbb{C}} [1\sigma]$	$\eta_{\mathbb{C}}^{\max} [2\sigma]$	$B_{SP}$
Smaller Planets	[0.06, 0.35]	0.48	$0.353 \pm 0.035$	Larger Planets	[0.07, 0.44]	0.61	$0.448 \pm 0.035$
Colder Planets	[0.15, 0.68]	0.86	$1.003 \pm 0.070$	Hotter Planets	[0.07, 0.39]	0.53	$0.426 \pm 0.043$
Colder Stars	[0.07, 0.33]	0.44	$0.639 \pm 0.063$	Hotter Stars	[0.06, 0.47]	0.66	$0.411 \pm 0.034$
Inner Planets	[0.21, 0.64]	0.80	$2.564 \pm 0.250$	Outer Planets	[0.03, 0.28]	0.42	$0.224 \pm 0.018$
Single-Planet Systems	[0.08, 0.50]	0.68	$0.689 \pm 0.066$	Multi-Planet Systems	[0.07, 0.33]	0.44	$0.420 \pm 0.038$
Habitable Zone	[0.23, 0.88]	0.97	$1.679 \pm 0.083$	Non-habitable Zone	[0.08, 0.37]	0.50	$0.653 \pm 0.066$

**Table 1.** Table of occurrence rates  $\eta$  for various subsets of the 284 planets examined in this work. Here  $\eta$  represents  $1\sigma$  credible interval values from the posterior distributions while 95 pct is the 95<sup>th</sup> percentile upper limit.  $B_{SP}$  is the Bayesian evidence computed by MULTINEST.

represent 50 fair draws from the posterior. From these plots we can make a number of comparisons. Dividing the sample in two by size, we see a marginally higher occurrence rate for the larger planets. For planet temperature we see a higher occurrence rate on the colder end. We also see a higher exomoon occurrence rate for colder stars, which we can take to mean later-type or evolved stars. All of these observations are in line with what we might expect *a priori*.

We see very little difference in the occurrence rate for single- and multi-planet systems. Interestingly, though,

the inner 50% of planets (those closest to their star) show a significantly higher exomoon occurrence rate than the those farther away, where the maximum *a posteriori* value indicates a total non-detection. This is unexpected, since the Hill sphere shrinks with smaller semi-major axis. Finally, and perhaps most intriguingly, the maximum *a posteriori* values for habitable zone planets ran away to the maximum, while non-habitable zone planets have a much lower occurrence rate. This should be read with caution, however, considering the size of the error bars in the habitable zone planet case.

Single Effective Moon Subset Fits							
Group	$R_S [1\sigma]$	$R_S^{\max} [2\sigma]$	$B_{SP}$	Group	$R_S [1\sigma]$	$R_S^{\max} [2\sigma]$	$B_{SP}$
Smaller Planets	[0.02, 0.36]	0.90	0.75	Larger Planets	[0.02, 0.41]	1.18	0.76
Colder Planets	[0.02, 0.41]	1.12	0.81	Hotter Planets	[0.02, 0.42]	1.03	0.79
Colder Stars	[0.02, 0.35]	0.83	0.79	Hotter Stars	[0.02, 0.36]	1.07	0.74
Inner Planets	[0.03, 0.81]	1.40	1.04	Outer Planets	[0.02, 0.29]	0.83	0.69
Single-Planet Systems	[0.02, 0.42]	1.05	0.79	Multi-Planet Systems	[0.02, 0.34]	0.96	0.72
Habitable Zone	[0.03, 1.10]	1.66	1.12	Non-habitable Zone	[0.02, 0.33]	0.92	0.74

**Table 2.** Table of effective moon sizes  $R_S$  for various subsets of the 284 planets examined in this work, in units of Earth radii. We present  $1\sigma$  credible interval values from the posterior distributions while 95 pct is the 95<sup>th</sup> percentile upper limit. Here  $B_{SP}$  is the Savage-Dickey ratio computed from the  $\log(R_S)$  posteriors.

While the comparisons above are made between equally populated subsets, there are far more non-habitable zone planets than there are planets in the habitable zone, making the results in the latter case far more uncertain.

Single, effective moon fits were also performed for these same OSE subsets. Results can be found in Table 2. Unlike the case for Galilean analog fits, we cannot meaningfully quote an occurrence rate in this case because the depth of the moon signal is controlled by the size of the effective moon. There is a degeneracy between moon size and occurrence rate, so we model the size of the moon as a proxy for occurrence rate. In essence, a smaller effective moon can mean either a) we have a high occurrence rate, with small moons, or b) we have a low occurrence rate of larger moons. The truth, of course, is probably somewhere in the middle.

To characterize whether our fits exclude a null detection to high confidence we compute the Savage-Dickey ratio between a uniform prior and the posterior distribution for  $R_S^2 = 0$ . We treat this number as our Bayes factor. For the colder stars we find  $B_{SP} = 1.53$ , for outer planets  $B_{SP} = 0.7$ , large planets  $B_{SP} = 1.29$ , multi-planet systems  $B_{SP} = 0.77$ , and habitable zone planets  $B_{SP} = 3.11$ . These values suggest that there is only marginal evidence in support an effective moon signal in the cold star, large planet, and habitable zone planet subsets, while the null hypothesis is favored for the outer planet and multi-planet subsets.

## 8. EXOMOON CANDIDATE KEPLER-1625B I

### 8.1. Individual Fits

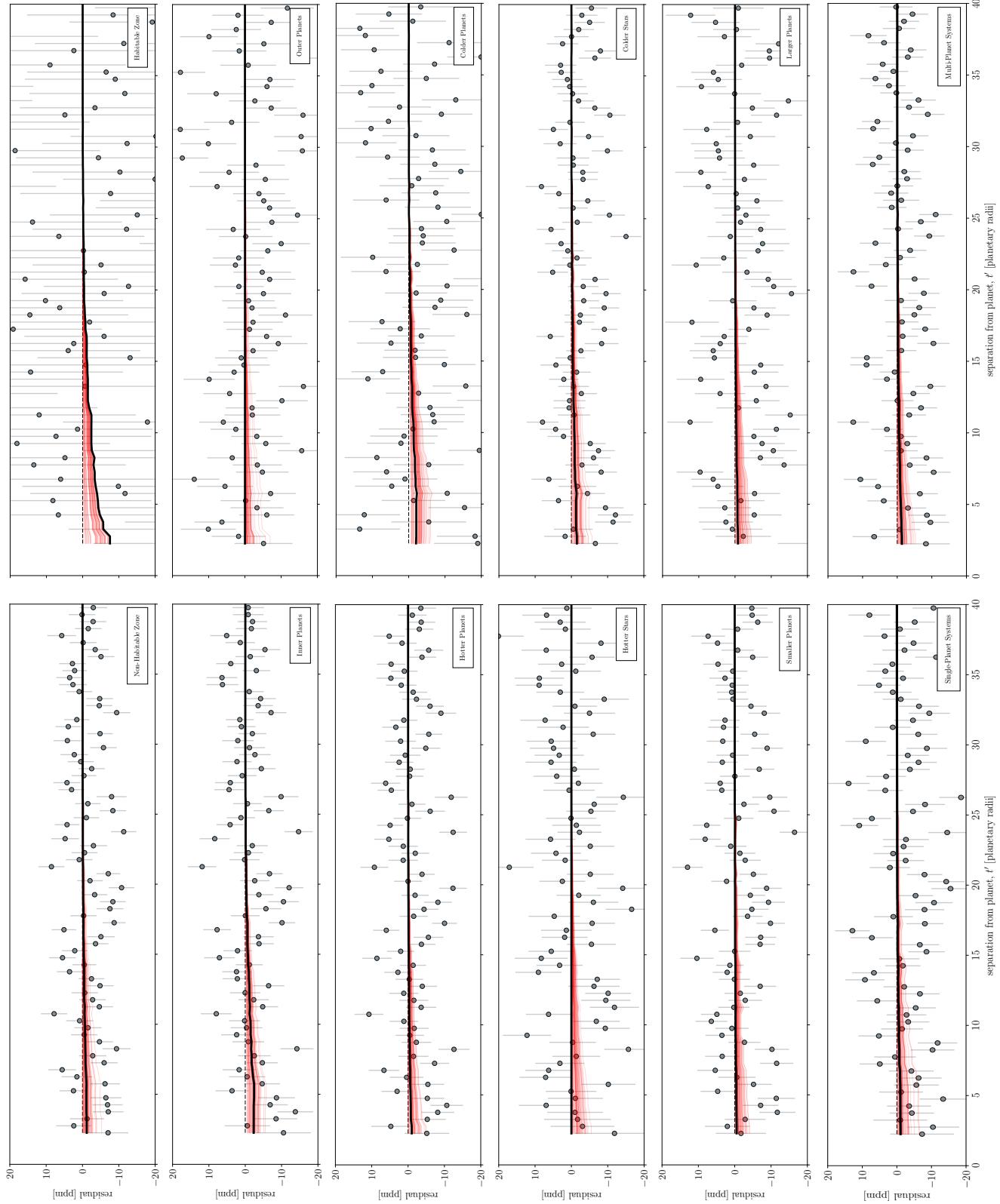
Thus far, our analysis has taken a population-based approach to seeking exomoons, in contrast to the conventional adopted in previous HEK papers where candidates were interrogated individually. Although a full suite of photodynamical Bayesian fits to each planetary candidate is beyond the scope of this work, representing a formidable computational challenge, we do here investigate systems individually with a simpler model.

For each KOI, we took the final phase-folded light curves and used MULTINEST to fit the Heller (2014)

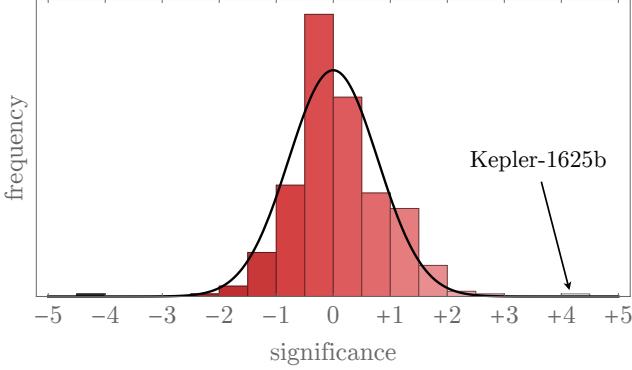
OSE model through the data. As discussed earlier, OSE does have drawbacks, such as the Callisto effect (see Section 6.1), yet it remains a useful and quick tool for checking for any significant flux decreases surrounding the phase-folded transits. For each KOI, we fit for an offset term, a photometric jitter, the moon size, and a semi-major axis,  $(a_{SP}/R_P)$  with MULTINEST using 1000 live points. In the earlier ensemble fits, we used a log-uniform prior on the moon-size (see Section 7.1), which means negative moon radii cannot be sampled. In these fits, we wished to allow for negative radius moon, which correspond to inverted transit signals, to provide insights into possible biases affecting the results. Accordingly, we modified our moon parameter to be the satellite-to-star ratio-of-radii,  $s$ , squared, to account for the fact transit detection bias is approximately proportional to  $s^2$ .

For each KOI, we evaluated the median  $s^2$  value from the posterior chain and divided by the lower quantile bounding a 68.3% confidence interval, a metric which we loosely refer to as “significance” in what follows (although it is best interpreted as a power). In Figure 12, we show a histogram of the resulting significances for 353 KOIs, where we have also included the likely eccentric KOIs to provide a wider sampler to assess distribution properties. Inspection of the results reveals a sizable spread centered around unity, as might be expected. However, we note that KOI-5084.01 (Kepler-1625b) appears quite deviant from the bulk population with a  $+4.4\sigma$  significance. Similarly, KOI-4202.01 (Kepler-1567b) shows is an outlier on the negative scale with a  $-4.4\sigma$  significance.

The negative outlier clearly cannot be a genuine exomoon and thus we do not consider it further it what follows. The positive outlier though could be an isolated candidate missed by the ensemble analyses. Excluding the two outliers, the histogram shown in Figure 12 is well-described by a normal distribution centered on zero with a standard deviation of 0.8. If we generate a list of 353 random variates from that distribution and take the maximum value,  $+4.4$  is highly improbable - specifically



**Figure 11.** Galilean analog GLC plots for a variety of sample subsets.



**Figure 12.** Histogram of the “significance” of an OSE detection for several hundred KOIs, the test which revealed the presence of a possible candidate with Kepler-1625b.

we used Monte Carlo simulations to estimate the probability is  $4 \times 10^{-6}$ . On this basis, Kepler-1625b appears quite unexpected and thus worthy of more detailed investigation.

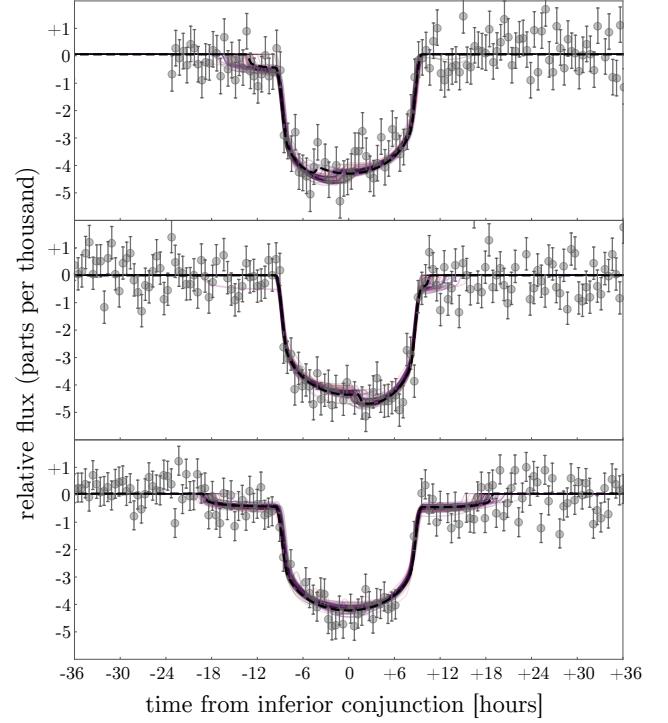
### 8.2. Detailed Investigation of Kepler-1625b

To investigate further, we performed detailed photodynamical fits of Kepler-1625b using the LUNA model and MULTINEST, in the same manner as that conducted in our previous series of HEK papers (e.g. see Kipping et al. 2013b). This enables a rigorous Bayesian model selection to ensure not only a physically plausible model can explain the photometry, but that the moon parameters are justified given the extra complexity they introduce.

Comparing the evidences between a planet-only and planet+moon model revealed that the moon model was favored with  $\log B_{SP} = 10.2$ , or a  $4.1\sigma$  preference, consistent with the level found previously. The light curve fits are illustrated in Figure 13.

Given the limited number of transits available, just three, the conventional HEK approach of cross-validating by removing a transit was not tenable and thus we adapted this strategy somewhat in what follows. Instead, we performed  $k$ -fold cross-validations, where we omit one-half of a transit (centered about the time of inferior conjunction) and re-fit the remaining data blindly each time. Since there are three transits in the *Kepler* data, this yields six unique ways of conducting the cross-validation. In each case, we performed a new blind fit and found a positive detection consistent with the original signal in all cases, with Bayes factors indicating 3.9, 2.4, 4.6, 4.0, 3.1, and  $2.7\sigma$  preference for the moon model (comparable to the original  $4.1\sigma$  detection when using all of the data).

As with all previous moon candidates we attempted to rule out all other possible explanations for the signal. For any candidate moon signal there will be one of two possible explanations: the signal is either an instrumental artifact or it has a true astrophysical ori-



**Figure 13.** The three transits of Kepler-1625b observed with *Kepler*, overlaid with 100 draws from the model posteriors. The black line is the maximum *a-posteriori* model.

gin (be it an exomoon or something else). To test the possibility of an instrumental aberration we performed an independent and manual detrending on the *Kepler* DR25 data (our OSE survey used DR24) using CoFiAM on the PA and PDC data. In both cases the planet-moon model remained the favored hypothesis over the planet-only model. In contrast, polynomial-based detrending was found to remove the signal, likely due to very long-timescale nature of the driving event occurring in epoch three.

We also examined data from individual *Kepler* pixels to determine whether there might have been anomalous behavior in the vicinity of the transit events (as occurred for false-positive Kepler-90g.01; Kipping et al. 2015a), but no unusual spikes or drop outs were detected in any relevant pixel. We also verified that there were no bad data flags at the time of the transits, which was a source of false-positive in the moon candidate around PH-2b (Kipping et al. 2015b).

If the signal were astrophysical in origin, then there are several possible hypotheses, including a moon. A ring is not likely since it should produce a coherent signal in all events, which is not seen, unless the ring precession rate is very fast. Rotating spots on the surface of the star also affect the light curve, primarily producing a long-term undulation in the data. This long-term trend is removed via our detrending (repeated several times independently) and thus the only remaining pos-

sible starspot-induced signal would be crossing events. However, starspot crossings (when a transiting planet occults a dark spot; e.g. see [Rabus et al. 2009](#)) cannot be responsible for the observed moon-like dips. This is because spot crossings can only occur inside the main planetary transit and can never produce out-of-transit flux decreases, as seen for Kepler-1625b, purely from a geometrical argument. If the signal were confirmed then, this would leave the exomoon hypothesis as the leading explanation based on current information.

The quoted stellar properties of Kepler-1625 in the NASA Exoplanet Archive (NEA) changed significantly from DR24 to DR25, owing to the addition of updated information in the latter data release ([Mathur et al. 2017](#)). This update pushed the star from a sub-Solar to a super-Solar radius ( $R_\star = 0.838^{+0.366}_{-0.079} \rightarrow 1.793^{+0.263}_{-0.488} R_\odot$ ), enhanced the metallicity from sub- to super-Solar abundances, and lowered the density substantially ( $\rho_\star = 2.059^{+0.4626}_{-1.306} \rightarrow 0.2636^{+0.3257}_{-0.0768} \text{ g/cm}^3$ ), indicating that this star is likely climbing the giant branch. Critically, our planet-only and planet+moon fits favor a low stellar density of  $\rho_\star = 0.387^{+0.034}_{-0.083} \text{ g/cm}^3$  and  $\rho_\star = 0.405^{+0.028}_{-0.054} \text{ g/cm}^3$  and if the true density were much higher, then Kepler-1625b would need to be either highly eccentric or blended ([Kipping 2014](#)), both of which would be severely detrimental to the exomoon hypothesis. Determining the true nature of this star is critical as it will also dictate the sizes of the planet and moon derived from the transit depth (which we describe shortly).

We also attempted to recover a rotation period for the star but the amplitude of variability is far too small for an accurate description, which adds further weight against an activity-induced moon-like feature.

Our photodynamical fits combined with the DR25 stellar properties indicate that Kepler-1625b is likely a Jupiter-sized planet with approximately ten times Jupiter's mass, orbited by a moon roughly the size of Neptune. We calculate the radii of the planet and moon by measuring the depth of the flux dip ( $\Delta F/F = (R_o/R_\star)^2$ , where  $R_o$  is the radius of the object in question) and we are able to derive a mass based on the photodynamical model fit. We note that both the planet and the moon show good agreement between mass and radius estimates and lie in physically reasonable parameter space based on the mass-radius forecaster model of [Chen & Kipping \(2017a\)](#). We find the semi-major axis of the moon  $a_S = 19.1^{+2.1}_{-1.9} R_P$ , which is well outside the Roche limit and comfortably within the Hill sphere for this planet. It is dynamically stable and should not have spun out / escaped over 5 Gyr ([Barnes & O'Brien 2002](#)).

While the existence of a Neptune-sized moon has not been anticipated in the literature, we cannot readily rule

out its existence on these grounds. Indeed, the existence of Hot-Jupiters was also wholly unexpected prior to their discovery in the mid-1990s. It seems clear that a moon of this type could not have formed in a circumplanetary accretion disk akin to that which is thought to have formed the regular moons of Jupiter and Saturn. It is conceivable however that the moon could have been captured by the planet, a kind of intermediate process between typical capture scenarios (e.g. Neptune and Triton; see [Agnor & Hamilton 2006](#)) and the cataclysmic impact event that is believed to have formed Earth's Moon ([Čuk & Stewart 2012](#)). In this scenario a grazing impact might be experienced as a kind of extreme atmospheric drag sufficient to capture the passing body. Observation of this system might therefore not only produce the first unambiguous detection of an exomoon, but could also go a long way in demonstrating once again that what we observe in our Solar System is not all that is possible.

### 8.3. Validating the Exomoon Candidate Kepler-1625b I

At this time, we remain cautious about the reality of this signal, given the relatively small number of transits available. This is particularly true because the third transit appears to be crucial to the exomoon interpretation and can be removed using polynomial-based detrending approaches. Detrending the photometric time series of long-period transits is more challenging than their shorter-period counterparts and it remains wholly plausible that the signal observed is nothing more than an artifact of our detrending process. We strongly emphasize these points and encourage the community to not treat this signal as genuine until it can be confirmed.

Fortunately, our photodynamic moon fit yields a testable prediction for the morphology of the next transit event occurring October 2017. We have secured HST observations to validate the signal in due course and we strongly advocate treating this object as no more than a candidate at this time, similar to previous moon candidates discussed in earlier HEK papers.

## 9. CONCLUSION

In this work we have examined 284 *Kepler* exoplanets (from an original sample of 4098 KOIs) in search of an exomoon signal in the ensemble. We performed a rigorous multi-stage analysis to select only the highest quality data, measure and correct for TTVs, and stack a total of 6096 transit events to characterize the exomoon population. As a byproduct of our work we present new TTV posterior distributions, along with a handful of stellar properties, and make them available online to the community.

Our results place new upper limits on the exomoon population for planets orbiting within about 1 AU of their host star, upper limits that are remarkably low.

We have also analyzed subsets of the ensemble to test the effect of various data cuts, and we have identified the regime in which the OSE model presented in Heller (2014) breaks down, which we call the “Callisto Effect”. Our analysis suggests that exomoons may be quite rare around planets at small semi-major axes, a finding that supports theoretical work suggesting moons may be lost as planets migrate inward. On the other hand, if the dearth of exomoons can be read as a reliable indicator of migration, our results suggest a large fraction of the planets in the ensemble have migrated to their present location.

Finally, we have briefly highlighted our identification of an exomoon candidate in the Kepler-1625 system, for which we have secured a follow-up observation with HST. This candidate has passed a thorough preliminary inspection, but we emphasize again our position that the *Kepler* data are insufficient to make a conclusive statement about the existence of this moon. Only after the HST observation is made should any claim about this moon’s existence be given much credence.

This paper includes data collected by the Kepler mission. Funding for the Kepler mission is provided by the NASA Science Mission directorate.

Resources supporting this work were provided by the NASA High-End Computing (HEC) Program through the NASA Advanced Supercomputing (NAS) Division

at Ames Research Center.

This research has made use of the NASA Exoplanet Archive, which is operated by the California Institute of Technology, under contract with the National Aeronautics and Space Administration under the Exoplanet Exploration Program.

This research has made use of the `corner.py` code by Dan Foreman-Mackey at [github.com/dfm/corner.py](https://github.com/dfm/corner.py).

This work made use of the Michael Dodds Computing Facility, for which we are grateful to Michael Dodds, Carl Allegretti, David Van Buren, Anthony Grange, Cameron Lehman, Ivan Longland, Dell Lunceford, Gregor Rothfuss, Matt Salzberg, Richard Sundvall, Graham Symmonds, Kenneth Takigawa, Marion Adam, Dour High Arch, Mike Barrett, Greg Cole, Sheena Dean, Steven Delong, Robert Goodman, Mark Greene, Stephen Kitt, Robert Leyland, Matthias Meier, Roy Mitsuoka, David Nicholson, Nicole Papas, Steven Purcell, Austen Redman, Michael Sheldon, Ronald Sonenthal, Nicholas Steinbrecher, Corbin Sydney, John Vajgrt, Louise Valmoria, Hunter Williams, Troy Winarski and Nigel Wright.

We thank members of the Cool Worlds Lab for helpful conversations in preparing this manuscript. DMK acknowledges support from NASA grant NNX15AF09G (NASA ADAP Program). AT acknowledges support from the NSF GRFP grant DGE 16-44869.

## REFERENCES

- Adams, E. R., Ciardi, D. R., Dupree, A. K., Gautier, T. N., III, Kulesa, C. & McCarthy, D., 2012, AJ, 144, 42
- Adams, E. R., Dupree, A. K., Kulesa, C. & McCarthy, D., 2013, AJ, 146, 9
- Agnor, C. & Hamilton, D. P., 2006, Nature 441, 192
- Agol, E., Steffen, J., Sari, R. & Clarkson, W., 2005, MNRAS, 359, 567
- Akeson, R. L., Chen, X., Ciardi, D., et al., 2013, PASP, 125, 989
- Barnes, J. W. & O’Brien, D. P. 2002, ApJ, 575, 1087
- Canup, R. M. & Ward, W. R., 2002, ApJ 124, 3404
- Canup, R. M. & Ward, W. R., 2006, Nature 441, 834
- Carter, J. A., Fabrycky, D. C.; Ragozzine, D. et al., 2011, Science, 331, 562
- Chen, J. & Kipping, D. M., 2017a, ApJ, 834, 17
- Chen, J. & Kipping, D. M., 2017b, ApJ, submitted (astro-ph:1706.01522)
- Christiansen, J. L., Jenkins, J. M., Caldwell, D. A., et al. 2012, PASP, 124, 1279
- Ćuk, M. & Stewart, S. T. 2012, Science, 338, 1047
- Dawson, R. I. & Johnson, J. A., 2012, ApJ, 756, 122
- Dickey, J. O., Bender, P. L., Faller, J. E., et al., 1994, Science, 265, 482
- Domingos, R. C., Winter, O. C. & Yokoyama, T., 2006, MNRAS, 373, 1227
- Dressing, C. D., Adams, E. R., Dupree, A. K., Kulesa, C. & McCarthy, D., 2014, AJ, 148, 78
- Durbin, J. & Watson, G. S., 1950, Biometrika, 37, 409
- Everett, M. E., Barclay, T., Ciardi, D. R., Horch, E. P., Howell, S. B., Crepp, J. R., Silva, D. R., 2015, AJ, 149, 55
- Gong, Y.-X., Zhou, J.-L., Wie, J.-W. & Wu, X.-M., 2013, ApJL, 769, L14
- Gould, A., Gaudi, S. B. & Han, C., 2004, arXiv e-prints:0405217
- Feroz, F. & Hobson, M. P., 2008, MNRAS, 384, 449
- Feroz, F., Hobson, M. P. & Bridges, M. 2009, MNRAS, 398, 1601
- , Hendrix, A., Esposito, L., et al. 2005, AGU Fall Meeting Abstracts,
- Hansen, C. J., Esposito, L., Stewart, A. I. F., et al. 2006, Science, 311, 1422
- Heller, R., 2014, ApJ, 787, 14
- Heller, R., Hippke, M. & Jackson, B., 2016, ApJ, 820, 88
- Holman, M. J. & Murray, N. W., 2005, Science, 307, 1288
- Huber, D., Silva A. V., Matthews, J. M. et al., 2014, ApJS, 211, 2
- Hippke, M., 2015, ApJ, 806, 51
- Hogg, D. W., Myers, A. D. & Bovy, J., 2010, ApJ, 725, 2166
- Holczer, T., Mazeh, T., Nachmani, G., et al., 2016, ApJS, 225, 9
- Ida, S., Canup, R. M., & Stewart, G. R. 1997, Nature 389, 353
- Kane, S. R., Hinkel, N. R. & Raymond, S. N., 2013, AJ, 146, 122
- Kipping, D. M., 2009a, MNRAS, 392, 181
- Kipping, D. M., 2009b, MNRAS, 396, 1797
- Kipping, D. M., 2010a, MNRAS, 407, 301
- Kipping, D. M., 2010b, MNRAS, 408, 1758
- Kipping, D. M. & Tinetti, G., 2010, MNRAS, 407, 2589
- Kipping, D. M., 2011, MNRAS, 416, 689
- Kipping, D. M., Bakos, G. Á., Buchhave, L. A., Nesvorný, D. & Schmitt, A. R., 2012, ApJ, 750, 115

- Kipping, D. M., Dunn, W. R., Jasinski, J. M. & Manthri, V. P., 2012, MNRAS, 421, 1166
- Kipping, D. M., 2013, MNRAS, 435, 2152
- Kipping, D. M., Hartman, J., Buchhave, L. A., Schmitt, A. R., Bakos, G. Á., Nesvorný, D., 2013, ApJ, 770, 101
- Kipping, D. M., Forgan, D., Hartman, J., Nesvorný, D., Bakos, G. Á., Schmitt, A. R., Buchhave, L. A., 2013, ApJ, 777, 17
- Kipping, D. M., 2013, Proc. for the Frank N. Bash Symp. 2013: New Horizons in Astronomy, held October 68, 2013 in Austin, TX
- Kipping, D. M., 2014, MNRAS, 440, 2164
- Kipping, D. M., Nesvorný, D., Buchhave, L. A., Hartman, J., Bakos, G. Á., Schmitt, A. R., 2014, ApJ, 784, 28
- Kipping, D. M., Schmitt, A. R., Huang, X., Torres, G., Nesvorný, D., Buchhave, L. A., Hartman, J., Bakos, G. Á., 2015, ApJ, 813, 14
- Kipping, D. M., Huang, X., Nesvorný, D., et al. 2015, ApJL, 799, L14
- Kipping, D. M. & Sandford, E., 2016, MNRAS, 463, 1323
- Kipping, D. M., Cameron, C., Hartman, J. D. et al., 2016, AJ, 153, 93
- Kolbl, R., Marcy, G. W., Isaacson, H. & Howard, A. W., 2015, AJ, 149, 18
- Kovács, G., Zucker, S., Mazeh, T., 2002, A&A, 391, 369
- Kozai, Y., 1968, PASJ, 20, 24
- Kundurthy, P., Agol, E., Becker, A. C., Barnes, R., Williams, B. & Mukadam, A., 2011, ApJ, 731, 123
- Lainey, V., Arlot, J.-E., Karatekin, Ö., et al., 2009, Nature, 459, 957
- Lainey, V., Karatekin, Ö., Desmars, J., et al., 2012, ApJ, 752, 14
- Law, N. M. Morton, T., Baranec, C., Riddle, R., Ravichandran, G., Ziegler, C., Johnson, J. A., Tendulkar, S. P., Bui, K., Burse, M. P., Das, H. K., Dekany, R. G., Kulkarni, S., Punnadi, S., Ramaprabaksh, A. N., 2014, ApJ, 791, 35
- Mayor, M. & Queloz, D., 1995, Nature, 378, 355
- Mathur, S., Huber, D., Batalha, N. M., Ciardi, D. R., Bastien, F. A., Bieryla, A., Buchhave, L. A., Cochran, W. D., Endl, M., Esquerdo, G. A., Furlan, E., Howard, A., Howell, S. B., Isaacson, H., Latham, D. W., MacQueen, P. J. & Silva, D. R., 2017, ApJS, 229, 30
- Morabito, L. A., Synnott, S. P., Kupferman, P. N., & Collins, S. A. 1979, Science, 204, 972
- Namouni, F. 2010, ApJL, 719, L145
- Rabus, M., Alonso, R., Belmonte, J. A., Deeg, H. J., Gilliland, R. L., Almenara, J. M., Brown, T. M., Charbonneau, D., Mandushev, G. 2009, A&A, 494, 391
- Sartoretti, P. & Schneider, J., 1999, A&AS, 134, 553
- Seager, S., Mallén-Ornelas, G., 2003, ApJ, 585, 1038
- Sheets, H. A. & Deming, D., 2014, ApJ, 794, 133
- Simon, A. E., Szabó, Gy. M., Kiss, L. L. & Szatmáry, K., 2012, MNRAS, 419, 164
- Smith, J. C., Stumpe, M. C., Van Cleve, J. E., Jenkins, J. M., Barclay, T. S., Fanelli, M. N., Girouard, F. R., Kolodziejczak, J. J., McCauliff, S. D., Morris, R. L. & Twicken, J. D., 2012, PASP, 124, 1000
- Spalding, C., Batygin, K., & Adams, F. C. 2016, ApJ, 817, 18
- Squyres, S. W., Reynolds, R. T., & Cassen, P. M. 1983, Nature, 301, 225
- Sparks, W. B., Hand, K. P., McGrath, M. A., et al. 2016, ApJ, 829, 121
- Stumpe, M. C., Smith, J. C., Van Cleve, J. E., Twicken, J. D., Barclay, T. S., Fanelli, M. N., Girouard, F. R., Jenkins, J. M., Kolodziejczak, J. J., McCauliff, S. D. & Morris, R. L., 2012, PASP, 124, 985
- Trafton, L., 1974, ApJ, 193, 477
- Waldmann, I. P., Tinetti, G., Drossart, P., Swain, M. R., Deroo, P. & Griffith, C. A., 2012, ApJ, 744, 35