The Astronomical Journal, 159:142 (16pp), 2020 April https://doi.org/10.3847/1538-3881/ab7001

Loose Ends for the Exomoon Candidate Host Kepler-1625b

Alex Teachey[1], David Kipping[1], Christopher J. Burke[2], Ruth Angus[3,4], and Andrew W. Howard[5]
[1] Department of Astronomy, Columbia University, 550 W. 120th Street, NY, USA; ateachey@astro.columbia.edu
[2] Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, MA, USA
[3] Department of Astrophysics, American Museum of Natural History, 79th Street at Central Park West, NY, USA
[4] Center for Computational Astrophysics, Flatiron Institute, 162 Fifth Avenue, NY, USA
[5] California Institute of Technology, Pasadena, CA, USA
Received 2019 April 30; revised 2019 December 9; accepted 2020 January 23; published 2020 March 5

Abstract

The claim of an exomoon candidate in the Kepler-1625b system has generated substantial discussion regarding possible alternative explanations for the purported signal. In this work, we examine these possibilities in detail. First, the effect of more flexible trend models is explored, and we show that sufficiently flexible models are capable of attenuating the signal—although this is an expected byproduct of invoking such models. We also explore trend models using x- and y-centroid positions, and show that there is no data-driven impetus to adopt such models over temporal ones. We quantify the probability that the 500 ppm moon-like dip could be caused by a Neptune-sized transiting planet to be $<0.75\%$. We show that neither autocorrelation, Gaussian processes, nor a Lomb–Scargle periodogram are able to recover a stellar rotation period, demonstrating that K1625 is a quiet star with periodic behavior <200 ppm. Through injection and recovery tests, we find that the star does not exhibit a tendency to introduce false-positive dip-like features above that of pure Gaussian noise. Finally, we address a recent reanalysis by Kreidberg et al. and show that the difference in conclusions is not from differing systematics models but rather the reduction itself. We show that their reduction exhibits, in comparison to the original analysis: (i) slightly higher intraorbit and post-fit residual scatter, (ii) $\simeq$900 ppm larger flux offset at the visit change, (iii) $\simeq$2 times larger y-centroid variations, and (iv) $\simeq$3.5 times stronger flux-centroid correlation coefficient. These points could be explained by larger systematics in their reduction, potentially impacting their conclusions.

Unified Astronomy Thesaurus concepts: Natural satellites (Extrasolar) (483); Irregular natural satellites (Extrasolar) (2025); Exoplanet detection methods (489)

1. Introduction

Last year, Teachey & Kipping (2018) (TK18 hereafter) presented evidence for a large exomoon orbiting the gas giant Kepler-1625b. That work was based on a joint analysis of three transits of the planet observed with *Kepler*, and a fourth transit observed with the *Hubble Space Telescope* (*HST*) in 2017 October (GO-15149; PI Teachey). The conclusion was based on the presence of significant transiting timing variations (TTVs) in the system, as well as a sustained dip in the star's brightness following planetary egress. These two lines of evidence were interpreted as self-consistent indications that a large moon is present in the system. A number of alternative explanations for these two signals were explored, and the likelihoods of these alternatives were considered. Taken together, the exomoon hypothesis emerged as the best explanation for the data in hand.

Since the publication of TK18, discussions with and among colleagues have highlighted open questions and unresolved issues emerging from the analysis. In this work, we take the opportunity to address some of these points and present an update on the prospects of confirming or rejecting the exomoon hypothesis for Kepler-1625b.

This paper is structured as follows. In Section 2, we explore other systematic models to account for the long-term trend seen in the TK18 light curve and the effects they have on the interpretation. In Section 3, we address differences between our work and that of another group (Kreidberg et al. 2019), whose independent reduction and analysis we became aware of during the course of writing this paper. In Section 4, we discuss the possibility that the moon-like dip is in fact caused by a second, previously undetected transiting planet in the system. In Section 5, we provide a more detailed assessment of the host star's activity and investigate the possibility that it could be responsible for the moon-like dip. In Section 6, we use forward propagation of the TK18 solution to determine the location and probability of seeing exomoon transits in future epochs. Conclusions are summarized in Section 7.

2. Other Systematic Models

2.1. Overview

TK18 employed three different models to account for the long-term trend seen in their data. These were broadly motivated to follow, as closely as possible, the most standard approaches in the literature for previous WFC3 analyses; see Wakeford et al. (2016) for an overview of WFC3 systematics. Most authors have previously elected to use a simple linear trend for this correction, of the form $a_0+a_t\ (t-t_0)$ (e.g., see Huitson et al. 2013; Knutson et al. 2014; Ranjan et al. 2014). In some rarer cases, a quadratic model has been invoked (Stevenson et al. 2014a, 2014b), and thus both of these models were attempted. A third exponential model was also attempted, giving three trend models in total, all with time as the independent variable.

Gaussian processes (GPs) have also been utilized in, for example, Evans et al. (2018) to handle WFC3 systematics. In general, however, GPs are not obviously appropriate for the moon search unless there is reason to suspect the data are not drawn from a sequence of independent Gaussians. As we show later, we see no evidence of time-correlated noise structure.

However, the flexibility of GPs means they will inevitably fit out a moon-like dip, and insomuch as less flexible detrending models explored here are also capable of attenuating or removing the moon signal, invoking GPs here is neither well-motivated nor particularly illuminating.

TK18 argued that it was crucial to perform this detrending simultaneous to the transit fits, repeating for each model (planet/moon/TTV), to account for the fact that the trend model appeared highly covariant with the moon-like dip. In comparing these models, we re-emphasize here that full Bayesian evidences should be used. As a nonlinear model (Kipping 2011), the number of degrees of freedom cannot be estimated, and thus reduced χ^2 comparisons are certainly invalid (Andrae et al. 2010). Another popular alternative to computing evidences is the Bayesian Information Criterion (BIC; Schwarz 1978). This was used, for example, in Kreidberg et al. (2019). Here again, there are serious concerns about its use for this problem. By invoking a Laplacian approximation on the posterior, one approximates the posterior to a Gaussian centered on the maximum likelihood estimator, which is inappropriate for highly multimodal posteriors such as those resulting from exomoon fits (Kipping et al. 2012). Further, the BIC is not guaranteed to yield a Bayes factor that is close to one computed using priors an observer would consider appropriate, as it assumes the unit information prior on the model parameters (Weaklieam 1999). For these reasons, model comparisons are performed using the Bayesian evidence in what follows.

The three trend models considered by TK18 allowed for an offset between the two visits, which was most apparent in the extreme channels and to a lesser degree in the white light curve. Clearly, the models explored by TK18 are a small subset of an essentially infinite number of possible models one could try. In general, the more flexible the model, the easier it is to fit out the moon-like dip when assuming no moon to be present. It was for this reason that more flexible trend models were not explored by TK18, since any sufficiently flexible instrument model can fit out interesting astrophysics.

Nevertheless, this was neither demonstrated nor investigated in detail in that work. For this reason, we revisit the trend modeling here, exploring: (a) the effect of going to higher-order polynomials, (b) the effect of allowing for discontinuous trend models, and (c) the effect of changing the dependent variable.

2.2. Higher-order Polynomials

Although an infinite number of polynomials exist beyond a quadratic trend, we here perform a cubic model as a simple extension to illustrate the effects. We reran the moon (M) and zero-radius moon (Z) models on the TK18 using MULTINEST (Feroz & Hobson 2008; Feroz et al. 2009) and LUNA (Kipping 2011), as was done in TK18, except we add an additional cubic term to the quadratic trend model.

The resulting maximum a posteriori light curve is shown in the second row of Figure 1. The shape closely matches that of the exponential model shown above it (and indeed, the quadratic model from TK18), reflecting the fact that the cubic term is almost zero. In fact, we can quantify this statement by evaluating the 1σ credible interval of the marginal posterior of the cubic coefficient to be 630^{+430}_{-450} ppm, reflecting how the posterior is only 1.4σ from zero.

It is therefore not surprising that: (a) the cubic fit returns a similar Bayes factor in favor of the exomoon to the original models of TK18, and (b) the cubic fit has an overall lower evidence than the other models ($\log \mathcal{Z}_M = 6311.34 \pm 0.16$), since it includes effectively wasteful parameter volume. In other words, the model has been penalized for additional complexity.

2.3. Discontinuous Polynomials

TK18 only considered models where the function is continuous across the visit change, except for a flux offset. This means that the higher-order polynomial coefficients, a_t and a_{t2} in the case of a quadratic model, are the same on both sides of the visit change. The motivation for this was that: (a) the small trends either side of the visit change did not correlate with centroid position and thus did not appear to be instrumental in nature, and (b) the star's intrinsic variability should not change dramatically on either side of the visit change.

We discuss here the effect of relaxing this assumption. As in the case of higher-order polynomials, this essentially represents a more flexible model. The simplest discontinuous polynomial is two independent straight lines ("bilinear" in what follows). As in TK18, all detrending choices explored in this work are implemented after the hook correction has been applied. Because the hook correction minimizes intraorbit rms independent of the model under consideration, there is no covariance between them, and therefore the evidences are not impacted. We refer readers to TK18 for a more thorough discussion of this choice.

We find that the fits favor a very pronounced reversal in the systematic gradient located at the visit change, as can be seen from the third row of Figure 1. It is unclear how this behavior could manifest physically, since flux is apparently uncorrelated with centroids in each individual visit for the comparison star KIC 4760469; see Figure S10 and Section 1.2.10 of Teachey & Kipping (2018). That is, while there is no known explanation for the visit-long trends, we would a priori expect the target star and the comparison star to display similar systematic morphologies. The fact that they do not leads us to question whether a downturn in the target star trend can be attributed to instrumental systematics. At the same time, we see no reason to expect the star to exhibit a pronounced reversal coincidental with *HST*'s visit change. The second visit slope has a negative gradient that absorbs the decrease in brightness caused by the moon-like dip—and for this reason, the evidence is significantly attenuated for the exomoon model. Indeed, the moon solution is quite distinct from the original paper, and can be immediately dismissed as suspicious because the moon ingress is nearly coincident with the flux offset associated with the visit change.

The bilinear model has two free parameters per visit, giving four in total—the same as the number of free parameters describing the TK18 quadratic and exponential models. Despite having the same number of free parameters, it does not necessarily have the same degrees of freedom. These two concepts are distinct if the underlying model is nonlinear (Andrae et al. 2010), which is true here because of the nonlinear step function occurring at the visit change. In fact, it can be seen that the bilinear model in fact has more freedom, because it does not require a continuous gradient across the boundary, unlike the case of the TK18 quadratic model.

It is therefore perhaps not surprising that this more flexible model is able to fit out the moon-like dip sufficiently well to

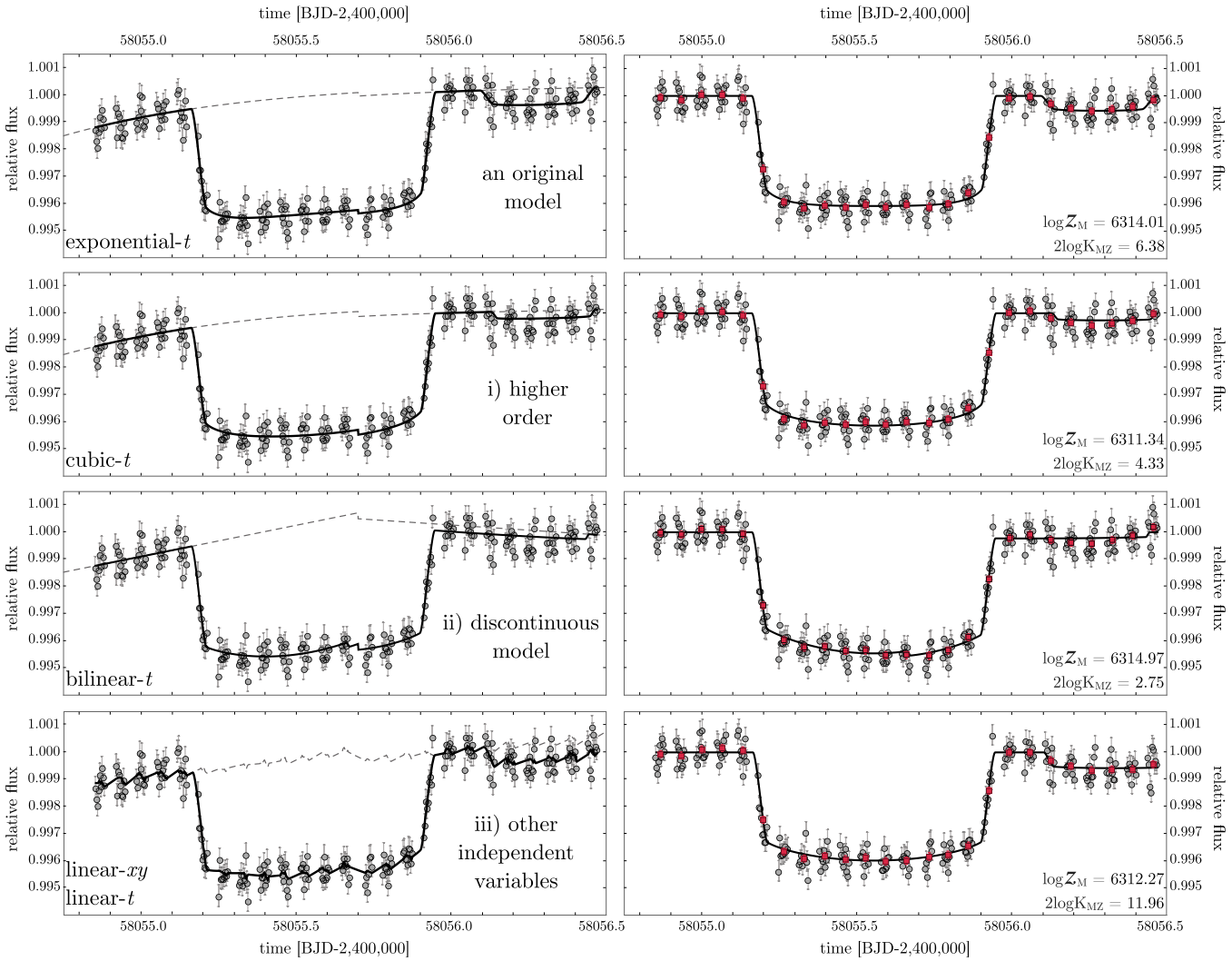

Figure 1. Comparison of three different long-term trend models applied to the TK18 data. The left column shows the uncorrected data with the trend model overlaid, while the right column shows the post-correction data. Systematic models more flexible than those considered by TK18 attenuate the formal exomoon evidence, and also find alternative modes that are inconsistent with the TK18 candidate signal.

find no evidence for the putative moon. This analysis does not particularly add to or detract from the exomoon hypothesis, since the behavior can be understood as a byproduct of employing more flexible trend models. Clearly, the attenuation of the moon-like dip does not necessarily imply that the trend model is incorrect; strictly speaking the conclusion is simply that the moon+flexible model is not supported by the data. Put another way, the moon model and the planet-only model are essentially equiprobable with this trend model, and therefore we would not claim evidence for a moon. We find no physical or data-driven motivation for adopting the bilinear model, nor is there precedent for doing so in the literature. On the other hand, the observation is unprecedented in several ways (e.g., the faintness of the target, the 40 hr duration, and the objective itself), so we cannot rule out the possibility that we are observing unique or heretofore only marginally important systematic effects.

2.4. Changing the Independent Variable

We now consider a third and final modification to the systematic model—namely, modeling the systematics as a function of both time and centroid position, rather than simply a function of time. We start by taking our simplest model, the linear-t model, given by

$$S(t) = a_0 + a_{t1}(t - t_0), \tag{1}$$

and extending it to include a linear dependency on x- and y-centroid positions:

$$S(t, x, y) = a_0 + a_{t1}(t - t_0) + a_{x1}(x - x_0) + a_{y1}(y - y_0), \tag{2}$$

where a_i are coefficients to fit for and the subscript 0 variables represent the median time/centroid positions. Using the same photodynamical MULTINEST fitting software from TK18, the resulting maximum a posteriori light curve is shown in the fourth row of Figure 1.

The figure, as well as the evidences quoted in the panel, show that the same moon is again detected. We therefore conclude that adding x and y as linear independent variables to the systematic model does not significantly affect the conclusions of TK18.

Table 1
The Effect of Enforcing Coplanarity on TK18 Results

reduction	$2\log(\mathcal{Z}_{M,\mathrm{coplanar}} - \mathcal{Z}_Z)$
linear-t	1.05 ± 0.32
quadratic-t	-0.28 ± 0.33
exponential-t	-0.01 ± 0.33

Note. A repeat of the model fits undertaken in TK18, now forcing the moon to be coplanar, i.e., $i_s = 90°$.

2.5. Fixing Orbits to Coplanar

The inclined solution for the exomoon candidate K1625b-i is particularly curious. The fact the posteriors favor an inclined solution suggests that it should be very difficult to fit the same moon to the existing data (both *Kepler* and *HST*) if one imposes coplanarity. To investigate this, we repeat the three trend model fits of TK18 for the M models but fix $i = 90°$. Comparing the resulting evidences to the original z models of that work ($2\log\mathcal{Z}_{MZ} = 17.77 \pm 0.33$, 3.61 ± 0.33, and 6.38 ± 0.34 for the linear, quadratic, and exponential models, respectively) indeed show the case for an exomoon is removed (see Table 1). This highlights the importance of including inclination in such fits.

2.6. Using the Comparison Star as a Model Benchmark

Since the comparison star is expected to be stable (TK18), it provides a useful control test for comparing the different possible systematic trend models. Expanding to quadratic order in x, y, and t, we fitted nine different models to the comparison star (assuming an intrinsically flat baseline) using MULTINEST. The various models and resulting evidences are listed in Table 2.

In TK18, only three of these models were considered, but it turns out that none of the other six models proposed here yield an evidence superior to the simple time models. We conclude that analysis of the comparison star indicates that systematic models using x- and y-centroid positions are not supported by the current data.

3. Comparison to Kreidberg et al. (2019)

During the final preparations of this paper, it came to our attention that Kreidberg et al. (2019) (henceforth KLB19) had conducted an independent reduction of the *HST* WFC3 observations of Kepler-1625 and concluded that there was no evidence for an exomoon based on the apparent lack of a moon-like dip following planetary egress. We will compare the KLB19 reduction and results to that of TK18 in what follows.

3.1. Raw Photometry Comparison

It is instructive to make a side-by-side comparison of the raw photometry presented in TK18 and the new reduction by KLB19 before any hook or trend corrections have been applied, which is shown in the top panel of Figure 2.

Before any systematic effects have been corrected, the photometry from both groups is apparently quite similar, yet important differences are evident. Quick inspection of both light curves reveals a much more pronounced offset in flux for KLB19 occurring at the instant of the visit change between orbits 14 and 15 (marked by the vertical dashed lines in that figure). As described in TK18, the full guide star acquisition performed at the beginning of orbit 15 was responsible for the introduction of this offset. Detailed modeling (described in Section 3.6) finds that the amplitude of the offset increases in every case from the TK18 reduction to that found in KLB19:

1. From (20 ± 110) ppm in TK18 to (-900 ± 120) ppm in KLB19, for the linear-t Z model;
2. From (-140 ± 120) ppm in TK18 to (-850 ± 130) ppm in KLB19, for the quadratic-t Z model; and
3. From (-20 ± 110) ppm in TK18 to (-880 ± 110) ppm in KLB19, for the exponential-t Z model.

There is, of course, only one ground truth in terms of the motion of the telescope and astrophysical variation. Because we can reasonably assume the star itself is not exhibiting a sudden change in flux after the 14th *HST* orbit, the discontinuity there must be systematic. Thus, a larger discontinuity could be viewed as being farther from the star's ground truth, requiring a more substantial correction that could impact the results of KLB19.

3.2. Hook Correction

We next applied the exact same hook correction algorithm described in TK18 to the KLB19 reduction. KLB19 also uses the nonparametric approach of TK18, thereby providing a fair comparison of the two, and this is shown in the bottom panel of Figure 2).

The mean intraorbit photometric rms from KLB19 is somewhat smaller at 360.7 ppm, versus 374.8 ppm for TK18.[6] At first glance, this appears to indicate that the KLB19 reduction is more precise. However, inspection of Figure 2 reveals that the 22nd *HST* orbit appears to display an anomalously low scatter of just 85 ppm. While TK18 also find that this orbit has the lowest scatter, the rms is much more consistent with the other orbits, at 210 ppm.

For normally distributed data, the standard deviation of sample standard deviations equals $\sigma/\sqrt{2(n-1)}$. Given that the mean rms for TK18 is $\sigma = 374.8$ ppm and the average number of points per orbit is 8.8, one should expect rms values with a standard deviation of 94.9 ppm. The actual standard deviation of rms values is less than one percent larger at 95.5 ppm. This, in turn, means that the 210 ppm smallest rms value is only 1.7σ from the mean.

For the KLB19 orbits, the expected standard deviation in rms values is 91.3 ppm and the observed value is 8.6% higher at 99.2 ppm. Critically, the 22nd orbit now appears to be a 3σ outlier. Strictly speaking, the formula above is only valid for $n \gg 1$, so we are at the limit of applicability in the present case. Thus, a better estimate for how surprising this orbit is can be obtained by masking the orbit, taking the mean of the other rms values, and then generating fake Gaussian data for all orbits and Monte Carlo evaluating the expected distribution of rms values. This reveals that the 22nd orbit from KLB19 is anomalously low at the 4.0σ level.

This seems highly implausible from a statistical perspective, and would make the 22nd orbit intraorbit rms an outlier by most definitions. In the presence of outliers, a more robust summary statistic to compare the precision of each reduction is the *median* intraorbit rms, rather than the mean. On this basis,

[6] TK18 quote 375.5 ppm, but that value is the mean intraorbit rms after 10 rounds of hook correction iterations, whereas the final light curve actually uses 100 rounds.

Table 2
Bayesian Evidences from Applying Various Systematic Models to the Comparison Star KIC 4760469

Label	Systematics Model	$\log \mathcal{Z}$
linear-t^*	$a_0 + a_{t1}(t - t_0) + (a_0' - a_0)\mathcal{H}[t - t_{II}]$	-0.60 ± 0.06
quadratic-t^*	$a_0 + a_{t1}(t - t_0) + a_{t2}(t - t_0)^2 + (a_0' - a_0)\mathcal{H}[t - t_{II}]$	0.00 ± 0.06
exponential-t^*	$a_0 + a_{e1}\exp\left(\frac{t - t_0}{a_{e2}}\right) + (a_0' - a_0)\mathcal{H}[t - t_{II}]$	-0.38 ± 0.06
linear-$xy^\dagger$	$a_0 + a_{x1}(x - x_0) + a_{y1}(y - y_0)$	-0.61 ± 0.06
linear-xy linear-t	$a_0 + a_{t1}(t - t_0) + a_{x1}(x - x_0) + a_{y1}(y - y_0)$	-4.29 ± 0.06
linear-xy quadratic-t	$a_0 + a_{t1}(t - t_0) + a_{t2}(t - t_0)^2 + a_{x1}(x - x_0) + a_{y1}(y - y_0)$	-3.96 ± 0.07
quadratic-xy	$a_0 + a_{x1}(x - x_0) + a_{x2}(x - x_0)^2 + a_{y1}(y - y_0) + a_{y1}(y - y_0)^2$	-1.27 ± 0.06
quadratic-xy linear-t	$a_0 + a_{t1}(t - t_0) + a_{x1}(x - x_0) + a_{x2}(x - x_0)^2 + a_{y1}(y - y_0) + a_{y1}(y - y_0)^2$	-4.97 ± 0.07
quadratic-xy quadratic-t	$a_0 + a_{t1}(t - t_0) + a_{t2}(t - t_0)^2 + a_{x1}(x - x_0) + a_{x2}(x - x_0)^2 + a_{y1}(y - y_0) + a_{y1}(y - y_0)^2$	-3.43 ± 0.07

Note. All evidences are quoted with 1334.18 subtracted—the absolute value obtained for the second model listed—"quadratic-t." Models with an * indicate that this is one of the original models used by TK18. Models with a † are those used by Kreidberg et al. (2019). The model considered by Kreidberg et al. (2019) is formally indistinguishable from the systematic models used in TK18, and therefore not favored over those used in that work.

the original TK18 reduction is marginally more precise at 360.6 ppm, versus 362.4 ppm for KLB19. That is, they are essentially indistinguishable from the perspective of their noise profile. The source of this improbably low scatter in the 22nd orbit is unknown, but it appears to be present in the data before the application of the hook correction.

3.3. Centroids

Centroids deserve special attention because KLB19 use the target's position on the detector as the basis for their systematics model correction. TK18 presented their centroid variability in Figure S10 of that work, for both Kepler-1625 and the comparison star KIC 4760469. Figure 3 directly compares the centroids of TK18 to those of KLB19 for Kepler-1625, where morphological similarities are apparent.

As was found in TK18, there is substantial variation of the apparent centroid position within an orbit, which we attribute to the hook and/or breathing effects rather than a real variation. For this reason, long-term behavior (associated with pointing drift) is best tracked using the orbit median centroid positions, shown in black in Figure 3.

We find that the range in interorbit y-centroid position is 10% higher for the first visit and 30% higher for the second visit in the KLB19 reduction than that of TK18, and 2.2 to 2.5 times higher for the x-centroid position. Similarly, the change in centroid position after the visit change is 2.6 times higher in x for KLB19 than for TK18.

We tried offsetting the median centroid of each orbit and then orbit folding (see lowest row of Figure 3) to look at intraorbit centroid variations, rather than interorbit. As before, we find higher intraorbit centroid variability for the KLB19 reduction, increased by similar levels.

The origin of these centroid discrepancies is unclear. Systematic effects such as the hook and breathing effects likely play a role in the calculated position for each image, as suggested by the intraorbit centroid variations, which do not appear to be associated with pointing drift. Different handling of these systematics could therefore reasonably explain the discrepancy. Due to the image rotation performed in TK18, it is not possible to apply our centroid corrections to the KLB19 reduction, nor is it possible to use the KLB19 centroids for detrending our extracted light curve.

We also point out that the calculation of these centroids is handled differently in TK18 and KLB19. TK18 simply calculates the flux weighted centroid of the optimal aperture at every time step as

$$X_{\rm centroid} = \frac{\sum_i^N x_i f_i}{\sum_i^N f_i}, \tag{3}$$

where x_i and f_i are the pixel coordinate and flux of pixel i, respectively. Calculation of the centroid in y is identical. TK18 performed this operation for both the target star and the comparison star KIC4760469, which showed good agreement.

By contrast, KLB19 perform a more complicated analysis to compute the motion, with different methodologies for the x and y directions. For the y (or "spatial") direction, KLB19 sum the flux in each column of the image at each time step, perform a four-pixel Gaussian convolution of the resulting array, and then an interpolation to compute a best-fitting offset from a template at each time step. It is not obvious how spatial information is recovered from this algorithm as described in KLB19, nor whether comparison to a template could introduce biases. KLB19 perform a similar operation for the x (or "spectral") direction, though now only summing up along the target spectrum instead of along each row in the image. The result, as shown in Figure 3 is morphologically similar to TK18, though with larger systematics.

3.4. Systematic Trend Comparison

KLB19 use systematic models that decorrelate against x and y centroid position rather than just time. Because the previous subsection has conducted a like-for-like comparison of this model, it is instructive to inspect the systematic parameter posteriors that result. We list these values in Table 3.

As can seen be seen from the table, KLB19 find an overall stronger dependency between flux and centroid position than TK18, with almost all of the variability coming in via the y-direction. It is also worth noting that the sign of a_{y1} reverses for the KLB19 reduction.

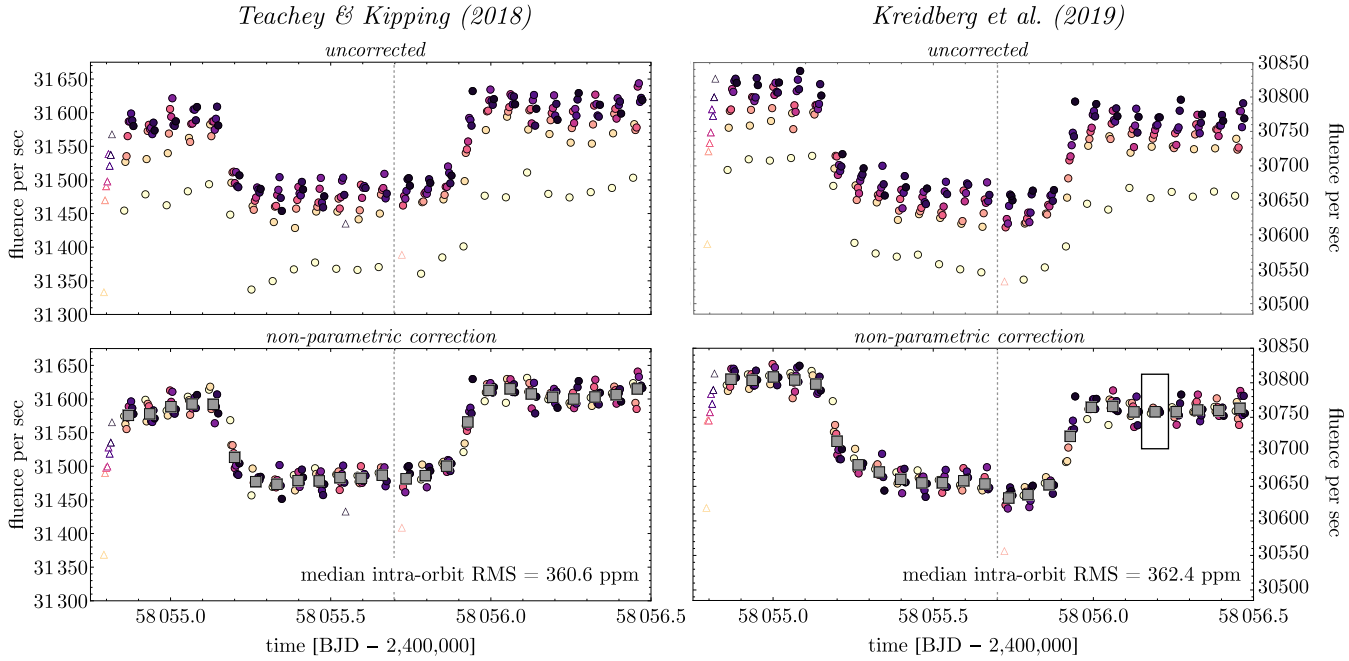

Figure 2. A comparison of light curves before and after the hook correction from TK18 and KLB19. Compare to Figure 2 in TK18. As in that work, the data are color-coded by the observation number within each *HST* orbit (light yellow for the first observation, dark purple for the last). The grey squares in the bottom panel represent the binned flux for each orbit. Triangles indicate observations from the first orbit, which are left out of the hook correction normalization. The anomalously low scatter in the 22nd orbit of the KLB19 analysis is highlighted with a rectangle.

We also remind the reader that our earlier comparison of different trend models applied to the comparison star found that models including x–y pixel position were disfavored (Section 2.6) over temporal models. In any case, it is difficult to determine the degree to which these discrepancies arise from the different centroiding approach, versus how much is due to differences in the raw fluxes owing to the reduction itself.

3.5. Model Evidences Comparison

TK18 perform model comparison using the Bayes factor calculated using Bayesian evidences (marginal likelihoods). In contrast, KLB19 perform their model selection using reduced χ^2 and the BIC. As discussed earlier in Section 2.1, model comparison using the reduced χ^2 is invalid for nonlinear models, and it is thus not appropriate for transit light curve fits. The BIC is also inappropriate due to the multimodality of the posterior, which is poorly described by the Laplacian approximation used by BIC. Further, it is generally not guaranteed to produce an approximation of the Bayes factor (Weaklieam 1999), and indeed it has been argued to not even represent an approximation to any exact Bayesian solution—including the Bayes factor (Gelman & Rubin 1995). Accordingly, we strongly urge the avoidance of these tools for exomoon model selection.

To perform a full comparison between the two reductions, it is instructive to repeat the full photodynamical MULTINEST fits conducted by TK18 on the KLB19 reduction. This allows us to evaluate what the Bayes factor would be for the exomoon had we used this data set instead.

We fit the hook-corrected light curves of KLB19 using the same three models used by TK18—linear in time, quadratic in time, and exponential in time—all of which also include a flux offset parameter at the visit change. Further, we ran the photodynamical MULTINEST fits for both TK18 and KLB19 adopting a fourth systematics model—one motivated by the choice of KLB19 to decorrelate against centroid position. Specifically, this model is linear in time as well as in x- and y-centroid position, i.e., an example of changing the independent variable. The results of these fits are summarized in Table 4.

As can be seen from the table, the KLB19 reduction consistently yields lower Bayes factors for the moon solution versus that found by TK18. Although a moon dip is favored in all cases (contrasting with the BIC and reduced χ^2 testing of KLB19), the strength of the evidence is diminished to such a degree that we would not consider it justifiable to claim evidence for an exomoon. Combined with the investigation described earlier in Section 2.4, this strongly suggests that the differing conclusions between TK18 and KLB19 is not due to the choice of systematic model, but rather due to the reduction itself. This is the same conclusion reached by KLB19. However, we do not agree that the moon's existence has been ruled out, particularly in light of a second independent reduction and analysis carried out by Heller et al. (2019), which also finds evidence for the moon-like dip following the planet's transit.

We should point out that Nelson et al. (2020) found, through an extensive comparison of approaches to computing model evidences, that uncertainties are likely to be underestimated. As such, the uncertainties quoted in Tables 2 and 4 may be too low. For each run, we used 4000 live points, which is twice the recommended number for accurate evidence uncertainties (Feroz & Hobson 2008). In any case, artificially low uncertainties would not invalidate the salient features of our argument here, namely that (1) we see no strong impetus to adopt a detrending model based on centroids (see Section 3.3) and (2) evidence for the moon is considerably weaker based on the KLB19 light curve.

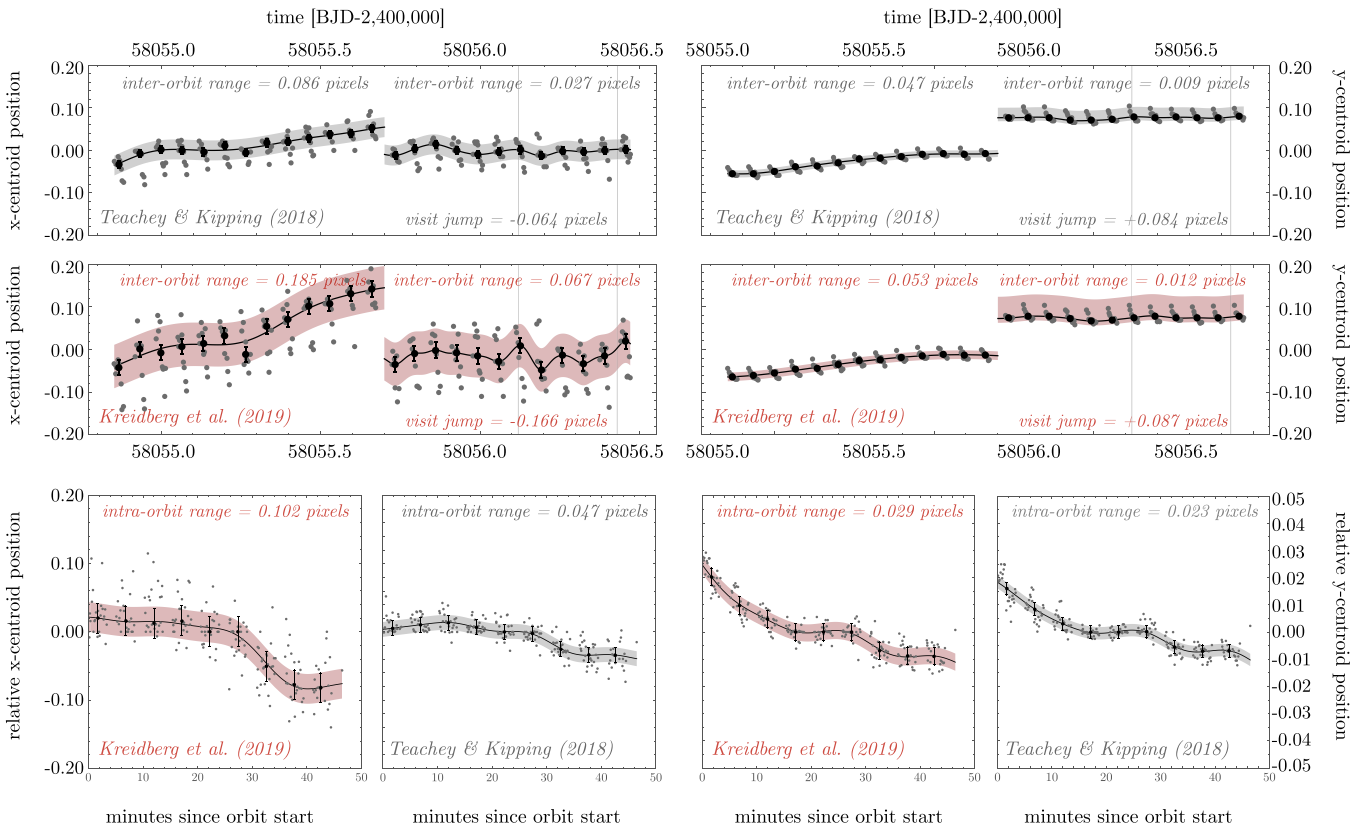

Figure 3. Comparison of the centroids reported by TK18 (gray) and those of KLB19 (red). In all cases, we find that the KLB19 centroids exhibit larger variations.

Table 3
Comparison of Centroid Gradient Terms

reduction	a_{x1}	a_{y1}
TK18, Kepler-1625	$+2700^{+910}_{-900}$	$+4780^{+440}_{-450}$
TK18, KIC 4760469	$+580^{+970}_{-960}$	$+3050^{+610}_{-620}$
KLB19, Kepler-1625	$+830^{+400}_{-400}$	-10170^{+410}_{-410}

Note. Using the systematic model "linear-*xy*" only (which assumes no part of the trend is dependent on time), we compare here the parameters a_{x1} and a_{y1} (i.e., the centroid gradient terms), which result from the three different reductions. Elements list median and 68.3% credible intervals in units of parts per million per day.

Table 4
Bayesian Model Evidence Comparisons for Various Systematic Models

model	TK18	Kreidberg et al. (2019)
linear-t^*	17.77 ± 0.33	1.08 ± 0.32
quadratic-t^*	3.61 ± 0.33	1.38 ± 0.32
exponential-t^*	6.38 ± 0.34	1.88 ± 0.33
linear-*xy* linear-*t*	11.96 ± 0.34	0.56 ± 0.34

Note. Comparing evidences for the four systematic model detrendings, applied to the TK18 and KLB19 reductions. Each element represents $2\log(\mathcal{Z}_M - \mathcal{Z}_M)$—the Bayes factor for the exomoon. *= original fits from TK18.

3.6. Model Residuals Comparison

The null hypothesis is that no moon is present around K1625, and so the obvious place to conduct a residual analysis is on the no-moon models (model *Z*).

The original residual analysis conducted by TK18 (see Figure S17) shows that, without a moon there appears to be high time-correlated noise when inspecting simple rms versus bin-size diagrams. However, as shown in that same figure, the origin of the time-correlated noise excess is apparently localized in time to the specific point where TK18 claim evidence for a moon-like dip.

A fairer test of residual noise is then to continue using the null hypothesis but mask out the region where TK18 claim a photometric anomaly associated with a possible moon. To accomplish this, we compute the maximum a posteriori model residuals for the exponential-*t* model (since this is the model used for light curve comparison by KLB19 in their Figures 3 and 4), for both the TK18 and KLB19 reductions, and then mask out the region $t > 2458056.1$ BJD. This also conveniently removes orbit 22 of KLB19, which is argued to show anomalously low scatter in Section 3.2.

We then compute an rms versus bin-size diagram, as shown in Figure 4. We find that both reductions display Gaussian-like behavior with no clear indications of excess noise. Without any binning, the rms values are 369.0 ppm for TK18 and 370.3 ppm for KLB19, i.e., effectively identical.

3.7. Presence of TTVs

The sustained moon-like dip in the *HST* observation observed by TK18 is one important element of the case for the exomoon. However, another critical, self-consistent component of the case for the exomoon presented in TK18 is the presence of TTVs in the system. A large moon like the one described in TK18 is expected to exert a significant

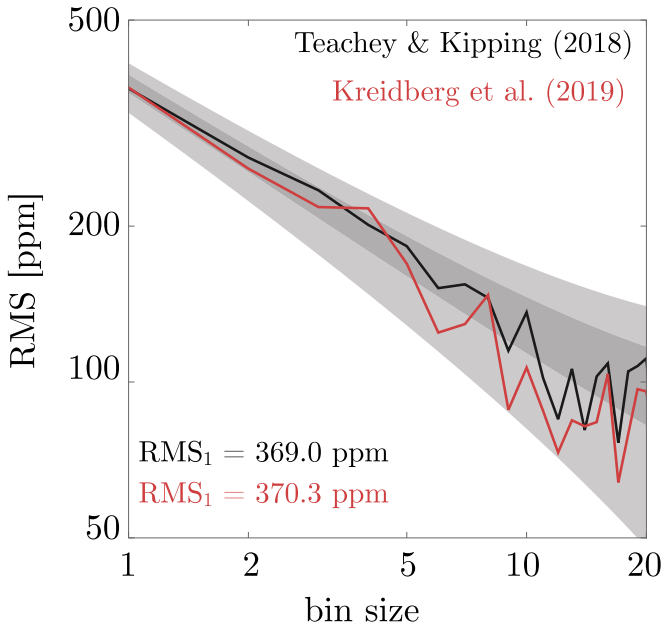

Figure 4. The rms vs. bin-size diagrams for the exponential-t model regressed to the TK18 reduction (black) and the KLB19 reduction (red). In both cases, the model assumes no moon but masks the region $t > 2458056.1$ BJD where the moon-like dip is seen by TK18. Both reductions appear consistent with Gaussian noise properties (the gray 1 and 2σ regions shown).

gravitational influence on the planet, detectable in the photometry in the form of timing variations.

As described in TK18, the transit of Kepler-1625b in the *HST* observation occurred a full 78 minutes earlier than anticipated based on a linear ephemeris calculated from the three transits of the planet observed by *Kepler*. This indicates the presence of TTVs to $\sim 3\sigma$ confidence.

We fit the transit timings for the KLB19 light curve and find $\tau = 58055.5539^{+0.0013}_{-0.0012}$, $58055.5538^{+0.0013}_{-0.0012}$, and $58055.5539^{+0.0013}_{-0.0012}$ for the linear-t, quadratic-t, and exponential-t models, respectively. Comparing this to the value in TK18 of $\tau = 58055.5563^{+0013}_{-0014}$, we consider the presence of TTVs to be validated by the new analysis, and it is worth noting that the new reduction actually suggests the *HST* transit occurred even earlier than was found in TK18.

3.8. Summary

We have executed a detailed comparison of the KLB19 reduction and that of TK18. KLB19 argue that there is, in fact, no evidence for a moon in their light curve, based on the absence of the moon-like dip. We also find that their reduction does not strongly support the presence of the moon-like dip (although it is still formally favored using a Bayesian model comparison), after applying the same hook correction and full Bayesian photodynamical model selection methods used by TK18 (see Section 3.6). We note, however, that the KLB19 reduction validates the presence of TTVs in the system, though TTVs alone do not constitute sufficient evidence for a moon.

The question naturally arises as to why the two studies yield different results—and which one is ultimately correct. We have argued that there are two major differences between the TK18 analysis and that of KLB19, and so presumably one (or both) of these is responsible for the discrepancy. The first is the choice of systematic model and the second is the independent reduction itself.

3.8.1. Systematic Model?

The first major difference between TK18 and KLB19, as explored in Section 2.6, is that KLB19 use a systematic model to correct for the long-term trend correlating flux with x- and y-centroid position, while TK18 only decorrelate against time. TK18 found no correlation between flux and centroid for the comparison star KIC 4760469, and in this work we have shown that among a broad suite of possible models, some with and some without such correlations, models including x and y correlations are consistently disfavored (see Section 2.6).

Even so, this does not address whether this different choice in detrending is ultimately responsible for the overall differing conclusions. We conclude that the detrending choice is unlikely to be the underlying cause, since refitting the original TK18 data including centroid correlation terms still recovers the same exomoon signal to comparable confidence as before (see Section 2.4).

3.8.2. Reduction?

With the detrending choice shown to be an unlikely explanation for the discrepant conclusions, we turn our attention to the reduction itself. There are certainly differences between the two reductions, both with respect to the methodologies (described in their respective papers) and the results.

With regard to the methodology, the KLB19 pipeline clearly has a track record that the TK18 reduction does not. Even so, the present observation is unprecedented in several ways. The star is significantly fainter than previous *HST* targets, the duration of the observation is far longer than typical transmission spectroscopy observations, and the nature of the pursued signal is fundamentally different. Therefore, it is reasonable to ask whether the KLB19 pipeline is guaranteed to be better than the one we have developed.

We note also that the procedure for selecting an optimal aperture as described in KLB19 is potentially problematic for the moon search. Their approach is to explore various apertures until they find the one that minimizes scatter with respect to the transit model. This differs from our approach, which does not assume a model. We can only guess that, given the computational expense of running a full exploration of parameter space with an MCMC simultaneous with the selection of an aperture, a static planet-only model was assumed and the scatter was minimized with respect to it. This approach could inadvertently incentivize the selection of an aperture for which the moon signal is attenuated. Nevertheless, the final aperture selected by KLB19 is quite similar to that of TK18, the primary difference being a 13% smaller aperture for KLB19, which probes slightly farther into the blue and a bit less into the red than the aperture of TK18.

We also identified anomalous behavior with orbit 22 of the *HST* observation as produced by the KLB19 pipeline, which shows suspiciously low photometric scatter. We are unable to determine the source of that anomaly, however.

KLB19 states that the moon-like signal presented in TK18 is "likely an artifact of the data reduction." However, no faults with the original reduction pipeline were found, nor was any step in the reduction pipeline identified as being the source of

the moon-like dip. Therefore, it is perhaps more accurate to conclude (as we do here) simply that the different pipelines have produced different results. It is noteworthy that a recent analysis by Heller et al. (2019), using their own independent reduction pipeline, also recovered a moon-like signal very similar to that presented in TK18. As such, the original interpretation of the data presented in TK18 has now been both validated and called into question in the literature. We thus argue that the existence of the moon remains an open question and additional observations are warranted.

To summarize Sections 3.1–3.3, we find that the product of the KLB19 reduction:

1. Exhibits marginally higher median intra-orbit rms (362.4 ppm versus 360.6 ppm) after correcting for the hooks;
2. Has a $\simeq$900 ppm larger flux offset at the visit change;
3. Has $\simeq$2 times larger variations in the y-centroid positions;
4. Has an x-centroid flux correlation coefficient $\simeq$3.5 times greater, and with opposite sign to KIC 4760469;
5. Exhibits a marginally higher residual rms (370.3 ppm versus 369.0 ppm) after fitting out a "no-moon model" and masking the claimed moon region in both reductions.

Accordingly, we argue that the KLB19 reduction is not obviously superior in any measurable way.

4. Second Transiting Planet?

4.1. Overview

One possible false-positive scenario for the moon-like dip that was not discussed in TK18 was the possibility that the dip is real but caused by a second transiting planet, not a moon. This scenario was not investigated in the original paper, because of the location of the dip with respect to the TTV offset —indicating a strong case for the exomoon hypothesis—as well as the inherently unlikely possibility that a planet could have evaded detection by *Kepler* but appear in this small segment of *HST* data. Nevertheless, this is certainly a valid concern, and the probability of this scenario was not quantified in the original paper, so we address it here.

We express the probability that the moon-like dip was caused by a second (hypothetical) transiting planet, K1625c, with orbital period P_c, as

$$\mathbb{P}_c = \Pr(\mathcal{T}, \bar{\mathcal{D}}_{\rm Kep}, \mathcal{D}_{\rm HST}|P_c), \tag{4}$$

where $\mathcal{T}$ is shorthand for the probability that $b < 1$ (i.e., that planet c has the correct geometry to transit), $\mathcal{D}_X$ denotes "detected by X." Via Bayes' theorem, we can express the probability as

$$\begin{aligned}\mathbb{P}_c &= \Pr(\mathcal{T}|P_c)\Pr(\bar{\mathcal{D}}_{\rm Kep}|\mathcal{T}, P_c)\Pr(\bar{\mathcal{D}}_{\rm HST}|\mathcal{T}, \bar{\mathcal{D}}_{\rm Kep}, P_c),\\ &= \Pr(\mathcal{T}|P_c)\Pr(\bar{\mathcal{D}}_{\rm Kep}|\mathcal{T}, P_c)\Pr(\bar{\mathcal{D}}_{\rm HST}|\mathcal{T}, P_c),\end{aligned} \tag{5}$$

where on the second line we remove the conditional $\bar{\mathcal{D}}_{\rm Kep}$ since there is no causal dependency.

To simplify the analysis, we will assume that any other planets in the system are coplanar with Kepler-1625b, whose low impact parameter essentially guarantees that these planets will be transiting as well. Accordingly, we assume $\Pr(\mathcal{T}|P_c) \simeq 1\ \forall P_c \in \{P_{\rm min}, P_{\rm max}\}$, where $P_{\rm min}$ and $P_{\rm max}$ are some yet-to-be-determined minimum/maximum limits on the period of planet c.

This optimistic assumption of coplanarity means that we will tend to overestimate the chance that the moon-like dip is caused by a second planet—which is the conservative option—and reduces the overall complexity of the problem.

4.2. Basic Properties of a Hypothetical K1625c

The depth of the moon-like dip varies between the three different long-term trend models adopted by TK18. In all three cases, the radius is approximately Neptune-like, yielding $4.90^{+0.79}_{-0.72}\ R_\oplus$ for the linear model, $3.09^{+1.71}_{-1.19}\ R_\oplus$ for the quadratic model, and $4.05^{+0.86}_{-1.01}\ R_\oplus$ for the exponential model. The last value is not only the median of the three but also represents the favored model by TK18. For this reason, we will assume here that the hypothetical second transiting planet has a radius of $4\ R_\oplus$ in what follows.

The moon-like dip is approximately flat-bottomed, indicating that if it were due to a transiting planet, the impact parameter is small, i.e., the planet must be nongrazing. This means the inferred radius from the depth is a fair estimator of the true radius.

The duration of the moon-like dip varies between the models from 8.5 hr for the linear and quadratic models to 7.8 hr for the exponential (using the $\tilde{T}$ transit duration definition of Kipping 2010). This therefore establishes that the duration of the hypothetical K1625c must exceed 7.8 hr. This is still a relatively long transit duration and implies that the orbital period is not short.

For any given orbital period, the longest possible duration corresponds to a zero impact parameter. Therefore, for any given duration, the shortest allowed orbital period corresponds to a zero-impact parameter. We can therefore take this duration value and convert it into a minimum period. Assuming a circular orbit, one may solve the Kipping (2010) $\tilde{T}$ duration equation for P in the limit of $b \to 0$, and also transform $a/R_\star$ into stellar is density, $\rho_\star$, using Kepler's Third Law. Since $\rho_\star$ is well-constrained from *Gaia* and isochrone modeling to be $0.29^{+0.13}_{-0.09}$ g cm^{-3} (TK18), we can solve for the minimum period numerically to find $P > 16$ days, as shown in Figure 5.

4.3. Probability of a Missed TCE

Here, $\Pr(\bar{\mathcal{D}}_{\rm Kep}|\mathcal{T})$ denotes the probability that a Neptune-sized transiting planet was undetected by the *Kepler* pipeline—i.e., a missed threshold-crossing event (TCE). There are no detected TCEs for Kepler-1625 aside from Kepler-1625b in DR25 (Thompson et al. 2017), but this fact alone does not provide a probability that one was missed by the *Kepler* pipeline.

The probability of missed TCEs is most directly computed by using the per target detection contours for DR25 reported by Burke & Catanzarite (2017). The `KeplerPORTs` software, first discussed in Burke et al. (2015), computes detection completeness contours for a given *Kepler* target through transit injection and recovery tests, and provides the most realistic estimate of completeness available. The stellar parameters used by Burke & Catanzarite (2017) are the DR25 Mathur et al. (2017) values, for which Kepler-1625 is reported as a $1.79^{+0.26}_{-0.49}$ $R_\odot$—which is approximately the same as the *Gaia*-based value found by TK18 of $1.73^{+0.24}_{-0.22}\ R_\odot$. This therefore demonstrates that the `KeplerPORTs` detection contours for a given planetary size do not require any significant update since the minor revision of TK18.

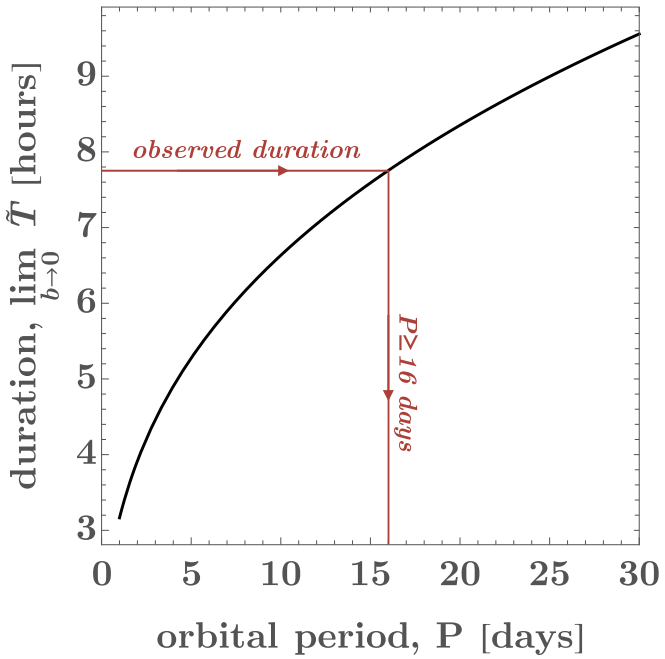

Figure 5. The moon-like dip reported by TK18 has a duration of at least 7.8 hr. Plotting the maximum transit duration (for a circular orbit) as a function of period for a planet around Kepler-1625, one can see that the period cannot be smaller than 16 days to explain the dip.

After running `KeplerPORTs` on our target (see Figure 6), we extracted a slice along the radius axis of $4\,R_\oplus$, corresponding to the dip seen in the *HST* data by TK18. This is shown in the leftmost panel of Figure 7.

`KeplerPORTs` natively computes completeness only out to 500 days, and indeed, by this point the probability of missing a $4\,R_\oplus$ exceeds 99% and is effectively unity—meaning there is little point in extending past this period.

4.4. Probability of K1625c Transiting in the HST Window

If the moon-like dip were due to another planet, then within the *HST* window of $W = 38.8$ hr, we would have observed a single transit of our hypothetical planet K1625c. The *HST* photometry is approximately four times superior to that of *Kepler*, ergo *HST* is effectively complete to a Neptune-sized transit of the observed duration. Thus, $\Pr(\mathcal{D}_{\rm HST}|\mathcal{T})$ then simply reduces to the probability that the planet will have the correct phasing to transit within the 38.8 hour observing window.

Consider the possibility that K1625c has an orbital period of 100 yr. The chance of seeing this world transit in a fixed window of observations is clearly going to be very low. Indeed, the chance of seeing a planet with period P transit at least once in a window is $\propto 1/P$. This is known as the window effect. It is described in detail by Kipping (2018), who shows that

$$\begin{aligned}&\Pr(n = 1|P_c, W, \mathcal{T})\\ &= \Pr(n \geqslant 1|P_c, W, \mathcal{T})(1 - \Pr(n \geqslant 2|P_c, W, \mathcal{T})),\end{aligned} \tag{6}$$

where n is the number of transits observed in the window of duration W and the components probabilities are

$$\Pr(n \geqslant 1|P_c, W, \mathcal{T}) = \begin{cases} 1 & \text{if } P_c \leqslant W, \\ \dfrac{W}{P_c} & \text{if } P_c > W, \end{cases} \tag{7}$$

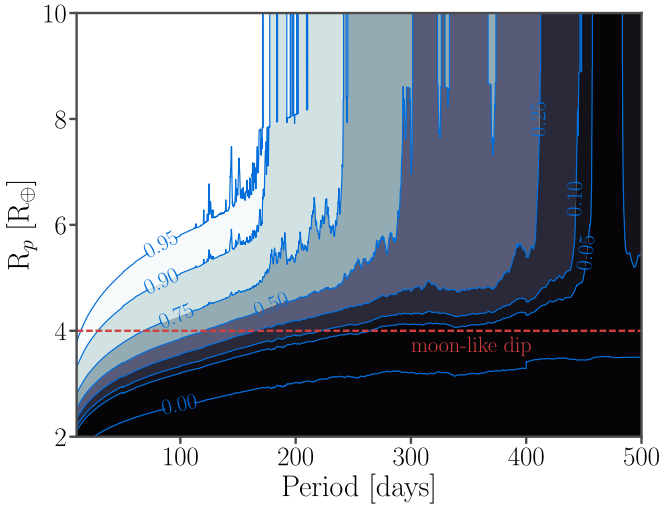

Figure 6. Detection completeness contours plot generated by `KeplerPORTs` (Burke et al. 2015; Burke & Catanzarite 2017) for the target Kepler-1625.

and

$$\Pr(n \geqslant 2|P_c, W, \mathcal{T}) = \begin{cases} 1 & \text{if } P_c \leqslant \frac{W}{2}, \\ \dfrac{W - P_c}{P_c} & \text{if } \frac{W}{2} < P_c \leqslant W, \\ 0 & \text{if } P_c > W. \end{cases} \tag{8}$$

Kipping (2018) shows how a lower limit on the period can be derived from the relative phase of the transit within the window, but in our case, a far more constraining lower limit on the period comes from the duration argument earlier in Section 4.2. Imposing this as a hard limit simplifies Equation (6) to

$$\Pr(n = 1|P_c, W, \mathcal{T}) = \frac{W}{P}. \tag{9}$$

This is shown in the middle panel of Figure 7. Finally, we may write that $\Pr(\mathcal{D}_{\rm HST}|\mathcal{T}) = \Pr(n = 1|P_c, W, \mathcal{T})$, since we treat *HST* as effectively complete to Neptune-sized transits.

4.5. Combining the Constraints

The final step is to combine the probabilities from above using Equation (5), which is shown in the rightmost panel of Figure 7. The probability peaks at $P_c = 133.3$ days with $\mathbb{P}_c = 0.74\%$, and decreases monotonically either side. Given the presence of TTVs in the system (TK18), the most plausible scenario to explain both the dip and the TTVs would be an interior transiting planet close to a mean motion resonance (e.g., 2:1 would lead to $P_c \simeq 144$ days).

The probability computed above suggests that the existence of another transiting planet causing the moon-like dip is quite unlikely, which might lend additional credence to the exomoon hypothesis. At the same time, this probability should be weighed against the probability of observing such a large exomoon. This comparison unfortunately eludes us for the time being, as there are currently no other verified exomoons in the literature, and the occurrence rate of such an unanticipated object cannot be quantified at this time.

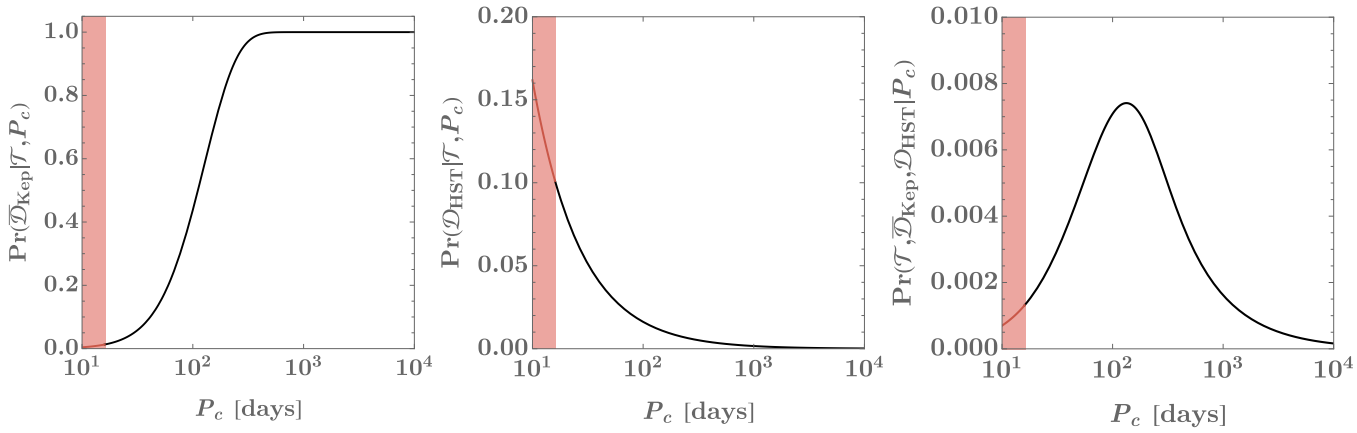

Figure 7. Probability that a Neptune-sized transiting planet evaded detection by *Kepler* (left), was seen to transit in the *HST* window of TK18 (middle), and the probability of both of these statements being true—as a function of the planet's orbital period.

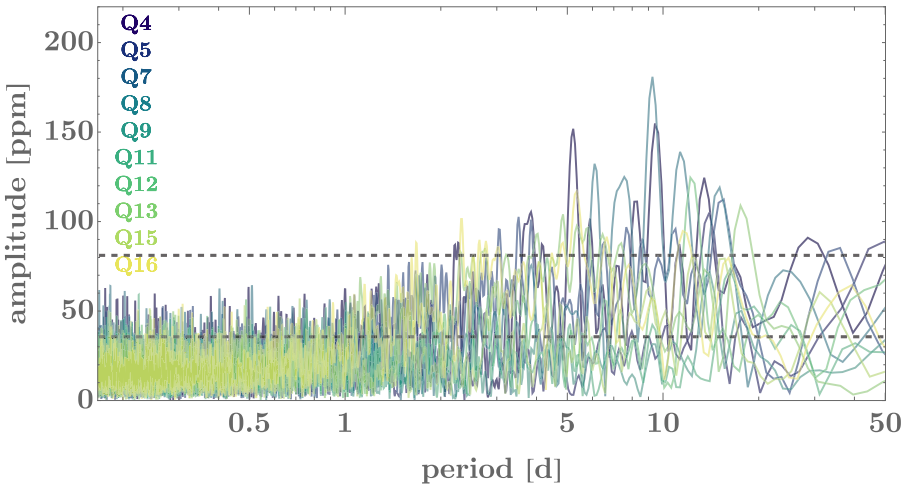

Figure 8. LS periodogram for the PDC data of Kepler-1625 for 10 quarters. Activity appears lower than 200 ppm for periods <50 days. Dashed lines represent the *p*-values of 0.05 for the most active and most quiescent quarters based on bootstrapping. Activity above these limits may be considered real, but a clear rotation period across all quarters is not detected.

5. Stellar Activity

5.1. Rotation

There is no known rotation period for Kepler-1625 at the time of writing. The star is included within the autocorrelation function (ACF) catalog of McQuillan et al. (2014), but no clear rotation period was found in that work.

We attempted to search for the rotation period using a Lomb–Scargle (LS) periodogram, applying the algorithm to each *Kepler* quarter (PDC data) independently. Since each quarter is treated independently, and each quarter has a duration of ≃90 days, it is not possible to detect periods longer than approximately half this value. The results out to 50 days are therefore shown in Figure 8.

Consistent with the analysis of McQuillan et al. (2014), we are unable to identify any clear rotation period from the LS periodogram. We find that the maximum amplitude of a periodic signal <50 days must be less than 200 ppm, significantly lower than the amplitude of the moon-like dip reported by TK18.

We also attempted to recover a rotation period using Gaussian process regression. We used the `celerite` software package (Foreman-Mackey et al. 2017) to model the light curve, with the kernel function consisting of a mixture of two simple harmonic oscillators with periods separated by a factor of two.[7] Exploring the posterior PDF of the star's rotation period using `PyMC3` (Salvatier et al. 2016) we infer a rotation period of $12.9^{+0.7}_{-0.6}$ days. However, for this period, the natural log of the Q factor (or damping ratio) was -3.1 ± 0.3. This means that the light curve, when modeled as a damped harmonic oscillator, is overdamped ($Q \sim 0.05$), indicating that the stellar brightness variations are incoherent, which suggests that the star spot lifetimes are shorter than the rotation period of the star on average. This, combined with the inability of both the LS periodogram and the ACF to recover a reliable period, implies that the signal is aperiodic and non-sinusoidal. Taken together with the low rms of the light curve (∼200 ppm), this indicates that Kepler-1625 is an inactive star and that there is little evidence of short-timescale (sub-hour) variability that could mimic the ingress of a moon.

Although not a direct measure of the rotation period, the v sin i can also provide some useful information on rotation. We obtained two Keck High Resolution Echelle Spectrometer spectra without iodine in 2018 October and November, in order to attempt to measure the velocity broadening. Using the `SpecMatch` pipeline described in Petigura et al. (2017), we

[7] See https://exoplanet.readthedocs.io/en/latest/tutorials/stellar-variability/.

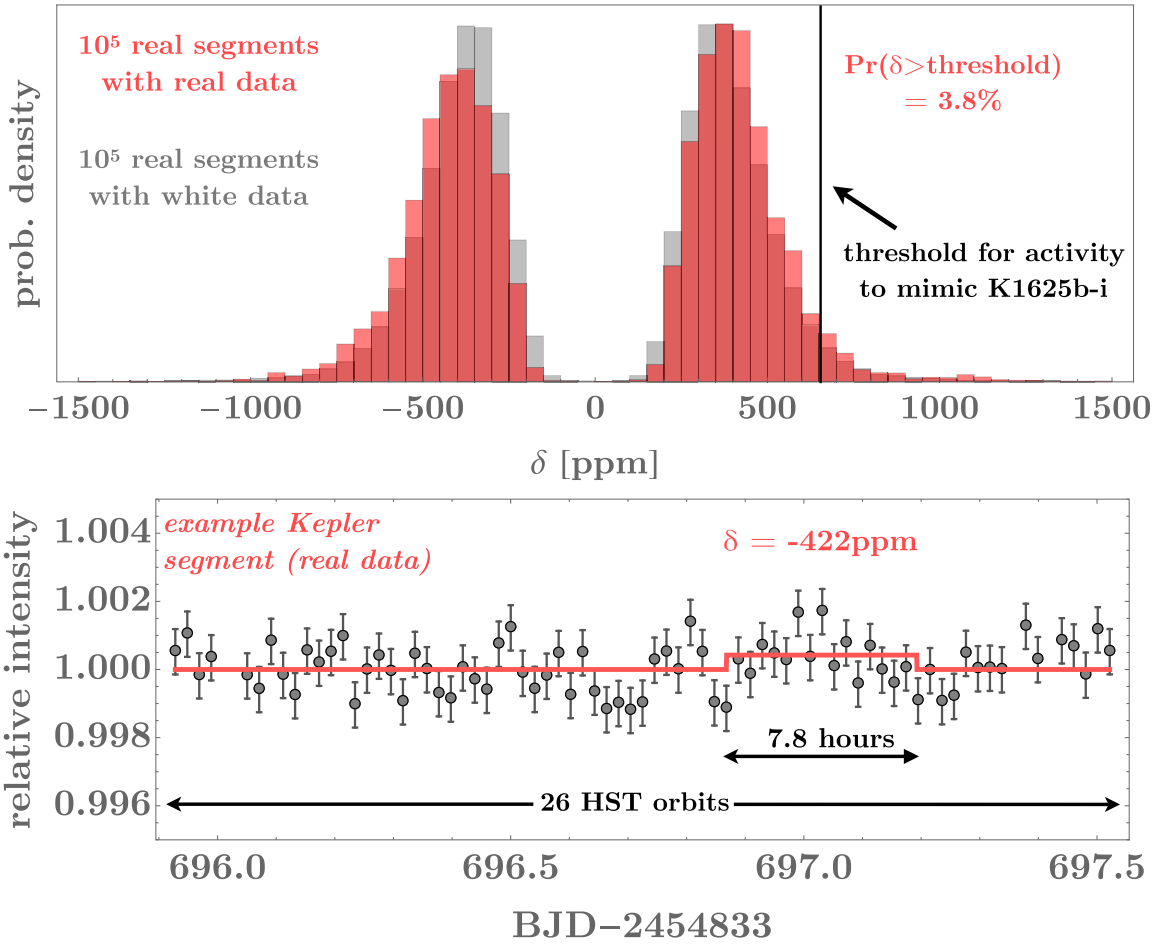

Figure 9. Bottom: example of a random *Kepler* segment of Kepler-1625 with a duration equal to that of the TK18 *HST* observing window. Regressing the best-fitting box with a duration of 7.8 hr (the same as for the moon-like dip) finds an inverted transit in this case of 422 ppm depth. Top: repeating this exercise on 10^5 random segments, we obtain a nearly symmetric distribution of best-fitting box amplitudes (red histogram). For comparison, we repeated the simulations assuming pure Gaussian noise only (gray histogram), which is nearly identical.

obtain a marginal detection of $v \sin i = (1.9 \pm 1.0)\,\mathrm{km\,s^{-1}}$. Combining this with the isochrone posteriors from TK18 yields a minimum rotation period of 45^{+44}_{-15} days (a minimum because we do not know sin *i*). It is therefore probable that the rotation period falls within the 50 day range that *Kepler* is sensitive to, but that the amplitude of rotational modulations is simply too small to reliably recover.

5.2. Activity-induced Dips

Stellar activity can produce complex morphological signatures in photometric time series (Dumusque et al. 2014). Although the photometric periodic behavior of Kepler-1625 appears limited to <200 ppm (see Section 5.1)—too small to induce an effect comparable to the moon-like transit reported by TK18—shorter, nonperiodic variations deserve our attention.

The moon-like dip is characterized by a transit depth of ≃500 ppm in the integrated light (white) bandpass of WFC3 (TK18). As a near-infrared instrument, stellar activity is generally expected to be suppressed by WFC3 versus an optical bandpass like that of *Kepler*. To estimate the magnitude of this effect, we took the isochrone posterior chains for Kepler-1625 (TK18) and extracted the median effective temperature of the star, $T_{\rm eff} = 5563$ K. We then assume spots on the surface with a temperature approximately 2000 K cooler than the photosphere, typical of sunspots. Integrating a *Planck* function multiplied by the bandpass response function for *Kepler* and WFC3 reveals that spots would appear 1.3 times larger in amplitude, as viewed by *Kepler* than WFC3. Accordingly, if the moon-like dip is due to spots, then one should expect to see frequent dips of amplitudes of ≃650 ppm (500 × 1.3) in the 11 quarters of *Kepler* data for Kepler-1625.

To test this hypothesis, we extracted a random segment from a random quarter of the available *Kepler* PDC time series, with the segment duration being equal to 26 *HST* orbits (which was the time window observed by TK18). Each quarter was first detrended using a median filter of window size equal to five times the minimum duration of the moon-like dip, approximately 7.8 hr (TK18). We then performed a blind search for the best-fitting box-like transit within this segment, forcing the box to have a duration equal to that of the TK18 moon-like dip. The central time and depth were optimized for in a least-squares sense. An example of this is shown in lower panel of Figure 9.

The best-fitting box was saved and then a new random segment was picked; this process was repeated 10^5 times. A histogram of the best-fitting depths is shown in the upper panel of Figure 9. Because this is the best-fitting depth within a segment, these depths always deviate from zero, as the regression routine is allowed to try many different possible central times. There is symmetry about zero, with just as many

inverted transits as positive transits being recovered. We find that 3.8% of the experiments run on the *Kepler* data are able to produce a best-fitting transit of depth exceeding 650 ppm (and 3.5% produce depths <-650 ppm). Naively, one might interpret this as indicating that the moon-like dip reported by TK18 is only 2.1σ significant (3.8%). However, these simulations were conducted for the 0.95 m *Kepler* telescope data—and not the 2.4 m *HST* data set in which the dip is actually observed. To interpret this 3.8% number, one must consider the plausible origin of these spurious (i.e., moon-mimicking) events.

If indeed the signals are spurious, there are two possible causes for these random quasi-dips. Either (1) time-correlated noise structure caused by intrinsic stellar activity is able to produce >650 ppm dips, or (2) the noise is not significantly correlated (i.e., white noise) but the noise budget of the *Kepler* photometry is sufficiently large that the best-fitting boxes can infrequently exceed 650 ppm.

If the former were true, then the dip observed by *HST* could be explained as one of these 3.8% instances of an activity-driven false-positive. If the latter were true, then one could expect it to be highly improbable for the *HST* moon-like dip to be a product of Gaussian noise, as the measurement uncertainties are 3.8 times smaller than that of *Kepler*.

Clearly this is an important distinction. To distinguish between them, we can set up another experiment where we repeat our previously described Monte Carlo experiments—except we replace the real *Kepler* photometric fluxes with artificial fluxes computed assuming pure Gaussian (white) noise. The artificial data are drawn from a normal distribution with a mean of unity and standard deviation equal to the standard deviation of a randomly picked real *Kepler* segment.

After drawing 10^5 segments and replacing the photometric fluxes with white noise, we make a histogram of the best-fitting box depths as before and find a very similar distribution (shown in Figure 9). The 650 ppm threshold is exceeded in a similar number of trials, 2.4%. We interpret the similarity between these two distributions as evidence favoring the hypothesis that the spurious, moon-mimicking detections are simply a product of Gaussian-like noise controlled by photon-counting statistics, rather than being due to intrinsic stellar activity. Since the *HST* data is much more precise, the probability of a white noise–driven box is far smaller, and is in fact accurately accounted for in our evidence calculations since we assumed a normal likelihood function in TK18. We therefore conclude that there is no evidence from the *Kepler* analysis that activity is a plausible explanation for the moon-like dip reported by TK18 in the *HST* data.

6. Follow-up

6.1. Photometric Follow-up

The best way to confirm the presence of the exomoon candidate would be to see it transit again. To this end, we have explored various avenues to observe future transits of Kepler-1625b. Unfortunately, this is a very challenging target for transit observations because of its faintness ($K_P = 15.756$) and the very long duration of the planet's transit (~19 hr). These challenges are exacerbated by the fact that the exact location of the exomoon cannot be known ahead of time for any given transit; as we project into the future, our predictions are naturally degraded as the uncertainties in our posterior samples propagate. A wide range of times before planetary ingress and after planetary egress must therefore be monitored in order to cover as many geometries as possible.

These limitations generally restrict any efforts to detect the exomoon transit to space-based telescopes. However, targeted observations of this sort clearly require considerable resources to be dedicated, to the exclusion of other priorities. While *Spitzer* may be a suitable alternative to observing with *HST*, the former can only observe ~35% of the sky at any given time due to pointing restrictions, and cannot observe the 2019 May transit as the target falls within the zone of avoidance. Future observations carried out with a survey (nontargeted) spacecraft could potentially bear fruit, though we note that the *Transiting Exoplanet Survey Satellite* observation of the *Kepler* field will occur in 2019 July, missing the 2019 May transit of Kepler-1625b.

On the other hand, transit timings could potentially be measured from the ground more easily, and continued monitoring of the TTVs and (to the extent possible) transit duration variations would be valuable. A single instrument may be able to monitor in its entirety either planetary ingress or egress, but likely not both, due to the time separation of these two events. Of course, this requires the target to be up at night long enough for the observation to be made, and the telescope must be located at a longitude where the event can be observed in its entirety without sunset or sunrise encroaching. Latitude is also a consideration; while northern latitudes place the target above the horizon for longer durations, they also experience a greater range of night lengths.

Radial velocity (RV) measurements of the system may also provide additional evidence for or against the moon hypothesis. On the one hand, RVs could potentially yield evidence for a second planet in the system, in which case the observed TTVs might be attributable to that planet and the case for the moon would be weakened. On the other hand, if an additional massive planet can be ruled out, or strong constraints can be placed on the mass and location of an undetected planet, the moon could emerge as a stronger candidate—insomuch as an alternative mechanism for the timing variations is weakened or removed altogether.

Of course, RV measurements should also provide a reliable measurement of the target planet's mass. If Kepler-1625b is revealed to be significantly less massive than anticipated, this could also weaken the moon case, as it would be more difficult to support such a large moon. Conversely, a mass measurement consistent with the inferred mass presented in TK18 could lend additional credence to the moon hypothesis. Figure 10 presents our best estimate for the planet's mass, which may be tested by the acquisition of RVs.

An additional complication for photometric confirmation through transit observations arises from the exomoon candidate's inferred large inclination with respect to the planet's orbital plane (TK18 found $i_S = 42^{+15}_{-18}$, 49^{+21}_{-22}, and 43^{+15}_{-19} degrees for the linear, quadratic, and exponential detrendings, respectively). This has the effect of sending the moon high above or below the disk of the star for a significant fraction of its orbit, precluding the possibility of a transit when the moon is in these positions. Coupled with the uncertainty in the moon's true anomaly, this means that there is no guarantee of seeing the moon transit at all, for any given transit observation. Thus, a null detection of the moon for any given epoch cannot be interpreted as definitive evidence that the moon does not exist.

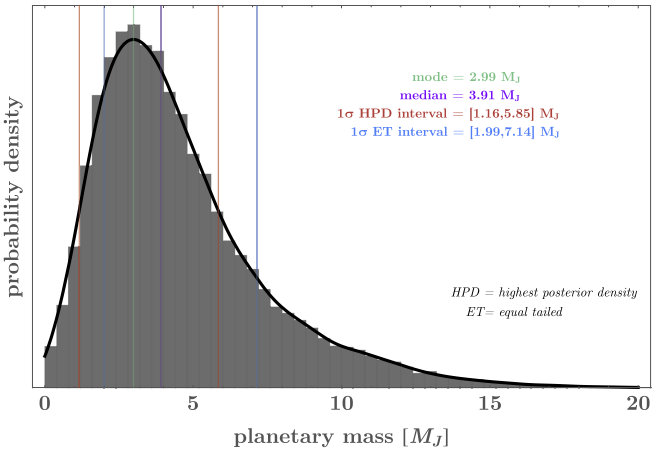

Figure 10. Combined mass posterior distribution for Kepler-1625b.

Only with many repeated observations, all lacking evidence for the moon, could the moon truly be ruled out to high confidence (Martin et al. 2019). Clearly, this has a multiplicative impact on the telescope requirements, and naturally leads to the conclusion that follow-up transit observations are only worthwhile to the extent that they are not excessively expensive.

Using the posterior samples from Teachey & Kipping (2018), it is possible to predict the morphology of the combined planet and moon system for future transit events—though as noted, these predictions naturally deteriorate with every epoch. We elected to consider 10 epochs, including the original observed epoch for comparison, and calculate 1000 projections of the transit light curve for Kepler-1625b. For this purpose, we used model $\mathcal{M}$ and repeated for each of the three trend models used by Teachey & Kipping (2018). The light curves can be viewed in Figure 11.

For each epoch, we find the time of transit minimum for the planet component only and use these times to compute a median mid-transit time and an associated standard deviation, which is quoted in the panels of Figure 11. We also consider the moon component in isolation and count how often the moon presents any deviation in flux away from unity—the probability that the moon will transit at all in each epoch. These probabilities are again added to the panels of Figure 11. Finally, we use the moon component only, to compute a probability distribution for the most likely location one should expect to observe the exomoon (assuming it transits at all). These curves are shown in gray in Figure 11.

As expected, we find that the uncertainty in the time of transit grows as we project further into the future. Regressing a power law to the uncertainties, we find that uncertainty grows as $t^{2/13}$, to an excellent approximation. The probability of a moon transiting oscillates for the first few epochs but then tends toward slightly less than 50%, which is broadly consistent with the findings of Martin et al. (2019). Observers can therefore treat the chances of seeing the moon in a given future transit as approximately 1 in 2.

We also highlight that epoch 9 (2019 May) appears to be the most favorable for follow-up. It has the highest probability of seeing a moon transit out of any future epoch and a clean prediction for the location of said transit (before planetary ingress). A proposed observation of this transit with *HST* was not awarded—and as pointed out, there are no other viable space-based options for this event.

In light of these challenges, photometric confirmation of the exomoon candidate may remain elusive for some time, until repeated observations may be performed at relatively low cost. Of course, if the moon is real, eventual confirmation is probably inevitable, but in the near-term, it will likely remain merely a candidate.

7. Conclusions

In this work, we have examined a number of alternative hypotheses put forth by the community to explain the two critical pieces of the exomoon case for Kepler-1625b: the presence of significant TTVs, and a sustained flux reduction in the *HST* light curve following planetary egress. We have explored various additional detrending models, employing more degrees of freedom, and found that while some of these approaches are able to attenuate the purported moon signal, this is to be expected given their flexibility. That is, from the standpoint of their Bayesian evidences, more flexible detrending models paired with planet-only transit models are in some cases indistinguishable from simpler detrending models combined with system models that include a moon. While we cannot rule out the presence of unprecedented systematic effects, we also see no evidence for them, and therefore the adoption of more flexible detrending models that attenuate the moon signal are not particularly well-motivated.

We have investigated the differences between the light curve presented in TK18 and a new reduction from Kreidberg et al. (2019), and find that while the source of the discrepancy is not readily identifiable, our light curve displays effectively identical noise properties. Therefore, the KLB19 light curve is not demonstrably superior. We also highlight once again the work of Heller et al. (2019); through their own independent reduction and analysis, they also found evidence of a moon-like dip following planetary egress.

In terms of a possible additional transiting planet in the system, we have calculated the probability that such a planet could have gone undetected in the *Kepler* data and transit in the short time window of the *HST* observation. We find the maximum probability of this scenario to be $<0.75\%$.

To determine whether the dip in brightness measured with *HST* could be due to stellar activity, we have attempted to measure a rotation period for the star in the *Kepler* data using a variety of standard methods. We are unable to recover it, indicating that the star exhibits negligible periodic variability. We have also searched for photometric dips that might be associated with (nonperiodic) star spot crossings. We find that such dips, while possible to find in the *Kepler* data, are consistent with Gaussian noise.

Finally, we discussed the outlook for confirming the presence of the exomoon using space-based transit monitoring, RV observations, and ground-based measurement of transit timings. We find that the system poses a number of substantial challenges to observational confirmation in the near-term, and conclude that while modest ground-based observations may be worthwhile for (1) constraining the mass of the planet, (2) quantifying the probability of an unseen perturber in the system, and (3) measuring TTVs, additional targeted observations from space likely fail a reasonable cost-benefit analysis. Confirming or refuting the moon to high confidence may therefore require many years and the advent of additional space-based time-domain survey data that can be acquired at minimal cost.

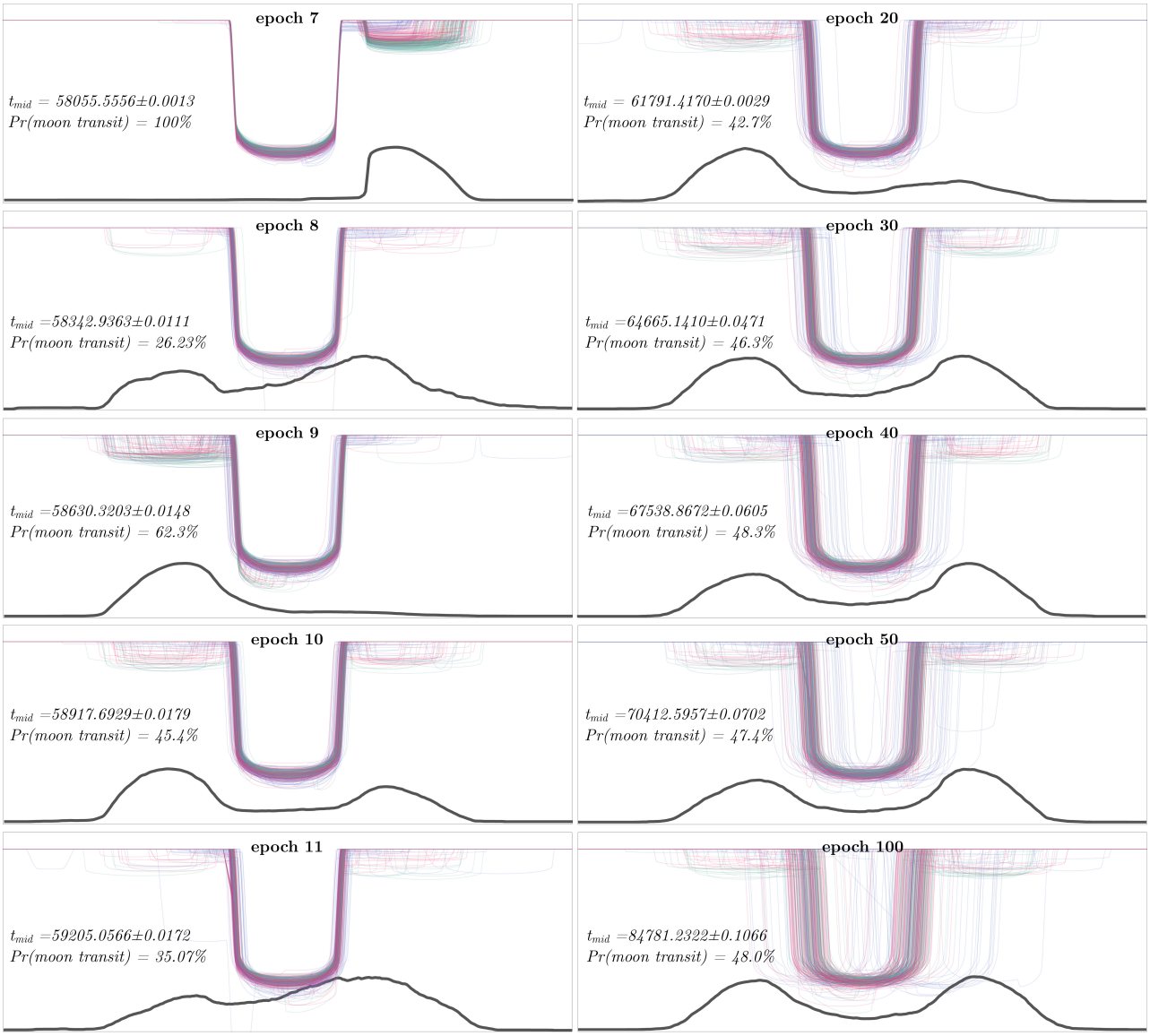

Figure 11. Projections of the light curve of Kepler-1625b and its candidate moon into the future, using the posterior samples of TK18. We show 100 light curves from each trend model (green = linear, blue = quadratic, and pink = exponential). Uncertainties can be seen to visually grow over time. The gray distributions reflect the most likely location for a moon transit.

We would like to thank Laura Kreidberg for useful discussions and providing data products in advance of her paper, which allowed us to investigate in greater detail the source and degree of the discrepancies between her findings and ours. We also thank Erik Petigura for performing the v sin i measurement. We thank the anonymous reviewer for thorough comments, which strengthened this paper. Finally, we wish to thank past and present *HST* and *Kepler* scientists and engineers, mission support personnel, and the crews of STS-31, 61, 82, 103, 109, and 125, who through their dedication have been jointly responsible for making this work possible.

Analysis was carried out in part on the NASA Supercomputer PLEIADES (Grant #HEC-SMD-17-1386). A.T. is supported through the NSF Graduate Research Fellowship (DGE-1644869). D.K. is supported by the Alfred P. Sloan Foundation Fellowship. This work is based in part on observations made with the NASA/ESA *Hubble Space Telescope*, obtained at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-26555. These observations are associated with program #GO-15149. Support for program #GO-15149 was provided by NASA through a grant from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-26555." This paper includes data collected by the *Kepler* mission. Funding for the *Kepler* mission is provided by the NASA Science Mission directorate.

ORCID iDs

Alex Teachey https://orcid.org/0000-0003-2331-5606
David Kipping https://orcid.org/0000-0002-4365-7366
Christopher J. Burke https://orcid.org/0000-0002-7754-9486
Ruth Angus https://orcid.org/0000-0003-4540-5661
Andrew W. Howard https://orcid.org/0000-0001-8638-0320

References

Andrae, R., Schulze-Hartung, T., & Melchior, P. 2010, arXiv:1012.3754
Burke, C. J., & Catanzarite, J. 2017, Planet Detection Metrics: Per-Target Detection Contours for Data Release 25, Kepler Science Document, KSCI-19111-001
Burke, C. J., Christiansen, J. L., Mullally, F., et al. 2015, ApJ, 809, 8
Dumusque, X., Boisse, I., & Santos, N. C. 2014, ApJ, 796, 132
Evans, T. M., Sing, D. K., Goyal, J. M., et al. 2018, AJ, 156, 283
Feroz, F., & Hobson, M. P. 2008, MNRAS, 384, 449
Feroz, F., Hobson, M. P., & Bridges, M. 2009, MNRAS, 398, 1601
Foreman-Mackey, D., Agol, E., Angus, R., & Ambikasaran, S. 2017, AJ, 154, 220
Gelman, R., & Rubin, D. 1995, Sociological Methodology, 25, 165
Heller, R., Rodenbeck, K., & Bruno, G. 2019, A&A, 624, A95
Huitson, C. M., Sing, D. K., Pont, F., et al. 2013, MNRAS, 434, 3252
Kipping, D. M. 2010, MNRAS, 407, 301
Kipping, D. M. 2011, MNRAS, 416, 689
Kipping, D. M. 2018, RNAAS, 2, 223
Kipping, D. M., Bakos, G. Á., Buchhave, L., Nesvorný, D., & Schmitt, A. 2012, ApJ, 750, 115
Knutson, H. A., Dragomir, D., Kreidberg, L., et al. 2014, ApJ, 794, 155
Kreidberg, L., Luger, R., & Bedell, M. 2019, ApJL, 877, L15, (KLB19)
Martin, D., Fabrycky, D., & Montet, B. 2019, ApJL, 875, L25
Mathur, S., Huber, D., Batalha, N. M., et al. 2017, ApJS, 229, 30
McQuillan, A., Mazeh, T., & Aigrain, S. 2014, ApJS, 211, 24
Nelson, B. E., Ford, E. B., Buchner, J., et al. 2020, AJ, 159, 73
Petigura, E. A., Howard, A. W., Marcy, G., et al. 2017, AJ, 154, 107
Ranjan, S., Charbonneau, D., Désert, J.-M., et al. 2014, ApJ, 785, 148
Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. 2016, PeerJ Computer Science, 2, e55
Schwarz, G. 1978, AnSta, 6, 461
Stevenson, K. B., Bean, J. L., Fabrycky, D., & Kreidberg, L. 2014b, ApJ, 796, 32
Stevenson, K. B., Bean, J. L., Seifahrt, A., et al. 2014a, AJ, 147, 161
Teachey, A., & Kipping, D. M. 2018, SciA, 4, 1784, (TK18)
Thompson, S. E., Coughlin, J. L., Hoffman, K., et al. 2017, ApJS, 235, 38
Wakeford, H. R., Sing, D. K., Evans, T., et al. 2016, ApJ, 819, 10
Weaklieam, D. L. 1999, Sociological Methods & Research, 27, 359