FALSE POSITIVE PROBABILITIES FOR ALL *KEPLER* OBJECTS OF INTEREST: 1284 NEWLY VALIDATED PLANETS AND 428 LIKELY FALSE POSITIVES

Timothy D. Morton[1], Stephen T. Bryson[2], Jeffrey L. Coughlin[2,3], Jason F. Rowe[4], Ganesh Ravichandran[5],
Erik A. Petigura[6,7], Michael R. Haas[2], and Natalie M. Batalha[2]

[1] Department of Astrophysical Sciences, Princeton University, 4 Ivy Lane, Princeton, NJ 08544, USA; tdm@astro.princeton.edu
[2] NASA Ames Research Center, M/S 244-30, Moffett Field, CA 94035, USA
[3] SETI Institute, 189 Bernardo Ave, Mountain View, CA 94043, USA
[4] Département de Physique, Université de Montréal, Montréal, QC H3T 1J4, Canada
[5] Department of Computer Science, Columbia University, 1214 Amsterdam Ave, New York, NY 10027, USA
[6] California Institute of Technology, Pasadena, CA 91125, USA

ABSTRACT

We present astrophysical false positive probability calculations for every *Kepler* Object of Interest (KOI)—the first large-scale demonstration of a fully automated transiting planet validation procedure. Out of 7056 KOIs, we determine that 1935 have probabilities <1% of being astrophysical false positives, and thus may be considered validated planets. Of these, 1284 have not yet been validated or confirmed by other methods. In addition, we identify 428 KOIs that are likely to be false positives, but have not yet been identified as such, though some of these may be a result of unidentified transit timing variations. A side product of these calculations is full stellar property posterior samplings for every host star, modeled as single, binary, and triple systems. These calculations use vespa, a publicly available Python package that is able to be easily applied to any transiting exoplanet candidate.

Key words: methods: statistical – planetary systems

Supporting material: machine-readable tables

1. INTRODUCTION

The *Kepler* mission has revolutionized our understanding of exoplanets. Among many other important discoveries, *Kepler* has identified several previously unsuspected features of planetary systems, such as the prevalence of planets between the size of Earth and Neptune, and a population of very compact multiple-planet systems. And perhaps most notably, it has enabled for the first time estimates of the occurrence rates of small planets ($\gtrsim 1\ R_{\oplus}$) out to orbits of about one year (e.g., Petigura et al. 2013; Foreman-Mackey et al. 2014; Burke et al. 2015). It is important to remember, however, that these revolutionary discoveries depend intimately on another revolution—how to interpret transiting planet *candidate* signals in the absence of unambiguous positive confirmation of their veracity.

Before *Kepler*, every survey searching for transiting exoplanets demanded that a candidate signal be verified as a true planet via radial velocity (RV) measurement of its mass. This would involve a series of follow-up observations in order to weed out astrophysical false positive scenarios—typically stellar eclipsing binaries in various configurations. However, following this model has been largely impossible for *Kepler* because of the quantity and character of the planet candidates (thousands of mostly small-planet candidates around relatively faint stars). There have been a small number of *Kepler* planets with masses measured by RVs (e.g., Marcy et al. 2014; Santerne et al. 2016), and significantly more that have been confirmed as planets by measurement of transit timing variations (TTVs) in multi-planet systems (e.g., Fabrycky et al. 2012; Ford et al. 2012; Steffen et al. 2012, 2013; Jontof-Hutter et al. 2016), but this still leaves the vast majority of candidates inaccessible to dynamical confirmation.

This situation has inspired the development of *probabilistic validation* as a new approach to evaluating transit candidates. The principle of probabilistic validation is to demonstrate that all conceivable astrophysical false positive scenarios are negligibly likely to be the cause of a transit candidate signal compared to the explanation of a planet transiting the presumed target star. The BLENDER method pioneered this approach and has validated many *Kepler* candidates (e.g., Borucki et al. 2012; Kipping et al. 2014; Torres et al. 2015). More recently, the PASTIS analysis suite has been introduced (Díaz et al. 2014) and used to validate both *Kepler* and *CoRoT* candidates (e.g., Moutou et al. 2014; Santerne et al. 2014). An alternative validation approach for candidates in multiple-planet systems has also been applied to a large number of *Kepler* systems based on the general argument that it is unlikely to see multiple false-positive signals in the same *Kepler* light curve (Lissauer et al. 2012), resulting in validations of over 800 planets with 99% confidence (Lissauer et al. 2014; Rowe et al. 2014). This methodology differs from the BLENDER/PASTIS approach in two significant ways: (a) it is applicable only to planets in multi-planet systems, and (b) it relies on broad-brush general arguments rather than analyzing the details of candidate signals individually.

While they have both proven useful for the purposes of validating individual candidates of particular interest, neither BLENDER nor PASTIS is designed for fully automated batch processing of large numbers of candidates. Morton (2012) describes a computationally simpler planet validation procedure designed for exactly such a purpose, based on the idea of describing eclipse light curves as simple trapezoids and simulating realistic populations of astrophysical false positives. This procedure has also been used in the literature to validate a number of *Kepler* planets (e.g., Muirhead et al. 2012; Dawson et al. 2012; Swift et al. 2013), and has also been applied to a

---

[7] Hubble Fellow.

number of candidates found by the *K2* mission (Becker et al. 2015; Montet et al. 2015). The code that implements this procedure is publicly available in the Python module `vespa`[8] (Morton 2015b).

This work presents results from applying `vespa` *en masse* to the entire *Kepler* catalog. This is both the first time that most *Kepler* candidates have been individually analyzed to assess false positive probability and the first time that a detailed automated planet validation calculation has been applied on such a large scale. Section 2 describes the methods used, Section 3 describes the data set, Section 4 presents the results, Section 5 compares these results with observational studies, and Section 6 contains concluding remarks.

2. METHODS

In this work, we apply the fully automated FPP-computing procedure described in Morton (2012, hereafter M12) to 7056 *Kepler* Objects of Interest (KOIs; see Section 3 for details). While we refer the reader to M12 for a detailed description of the method, we outline it briefly in this section.

2.1. False Positive Probabilities

The basic idea of `vespa` is to assign probabilities to different hypotheses that might describe a transiting planet candidate signal. If $\{H_i\}$ is the set of all considered hypotheses, the probability for any given model i is

$$\Pr(H_i) = \frac{\pi_i \mathcal{L}_i}{\sum_j \pi_j \mathcal{L}_j}, \qquad (1)$$

where π_i is the "hypothesis prior" and $\mathcal{L}_i$ is the "hypothesis likelihood."[9] The prior represents how intrinsically probable the hypothesized scenario is to exist, and the likelihood represents how closely the shape of the observed transit signal matches with the expected shape of a signal produced by the hypothesis.

The `vespa` procedure models an eclipse signal as a simple trapezoid, parametrized by depth δ, total duration T, and shape parameter T/τ, where τ is the "ingress/egress" duration (such that a completely V-shaped transit has $T/\tau = 2$). For the transit signal being evaluated, the joint posterior probability density function (PDF) of these shape parameters is sampled with Markov Chain Monte Carlo (MCMC), using the `emcee` sampler (Foreman-Mackey et al. 2013). This allows the likelihood for each hypothesis to be determined by simulating a physically realistic population of the hypothesized astrophysical scenario and using this population to define the PDF for the trapezoidal parameters under the hypothesis. The likelihood is then

$$\mathcal{L}_i = \int p_{\mathrm{sig}}(\boldsymbol{\theta}) \, p_i(\boldsymbol{\theta}) \, d\boldsymbol{\theta}, \qquad (2)$$

where $\boldsymbol{\theta}$ is the vector of trapezoidal shape parameters, p_{sig} is the posterior PDF of the signal, and p_i is the PDF for the parameters under hypothesis i.[10] The hypotheses supported by

`vespa` are the following: unblended eclipsing binary (EB), hierarchical-triple eclipsing binary (HEB), chance-aligned background/foreground eclipsing binary (BEB), and transiting planet (Pl).[11] In this work, we also implement "double-period" versions of each of the stellar false positive scenarios, acknowledging the possibility that if an EB has similar primary and secondary eclipse depths, then it might be mischaracterized as a primary-only transiting planet signal at twice the orbital period (especially if diluted). We note that the determination of the diluted eclipse depth of all these blended scenarios assumes that the light from the blended system is *fully contained* within the target's photometric aperture. That is, these scenarios do not account for the possibility that only a small fraction of the light from a nearby contaminating star might be in the aperture, many of which have already been identified via other methods (Bryson et al. 2013; Coughlin et al. 2014).

Observational constraints are incorporated in two different ways. First, photometric (or spectroscopic/asteroseismic) measurements of the target star are folded into the population simulations of each hypothesis (see Section 2.2). All other constraints are applied to narrow down which simulated instances of each scenario may be counted in the final prior and likelihood evaluations; for example, only blended eclipsing binaries with secondary eclipse depths shallower than the observed limits contribute to the construction of the p_i trapezoidal shape parameter PDF. For the "double-period" scenarios, we require the primary and secondary eclipse signals to have depths within 3σ of each other, where σ is defined as the fitted uncertainty in the trapezoid-model depth of the candidate signal.

The steps `vespa` takes to calculate the FPP of a transit signal are thus as follows.

1. Generate posterior samples for the transit signal under the trapezoid model, using MCMC.
2. Generate population simulations for each hypothesis scenario being considered (conditioned on available observations of the target star; see Section 2.2).
3. Fit each simulated eclipse in each scenario with a trapezoid model (using least-squares optimization).
4. Evaluate priors and likelihoods for each hypothesis, taking into account all available observational constraints.
5. Use Equation (1) to calculate the posterior probability for each scenario.

To quantify uncertainty due to the Monte Carlo nature of this procedure, `vespa` is also able to repeat these calculations any desired number of times by bootstrap resampling (with replacement) of the simulated populations, and recalculating the likelihoods based on the resampled populations. This mitigates the chances for rare outliers in a simulation to significantly affect the calculated FPP.

We note that a built-in weakness of model selection is that it assumes that the set of models being considered is comprehensive. This could in principle lead to a situation where one model is strongly preferred over all other models, but even that model is a poor explanation of the data—in this work, this could lead to improperly validated planets. There are two

---

[8] https://github.com/timothydmorton/vespa

[9] This factor is more widely known as the "Bayesian evidence" or "marginalized likelihood"; Morton (2014) argues for the term "hypothesis likelihood," as it can be clarifying to think of it that way.

[10] $\mathcal{L}_i$ may be seen to be the "evidence" or "marginalized likelihood" of the trapezoidal model under hypothesis i, with p_{sig} being the likelihood and p_i being the prior, integrated over the $\boldsymbol{\theta}$ parameter space. But for clarity, and for continuity with previous publications, we continue to call $\mathcal{L}_i$ the "likelihood" for hypothesis i.

[11] We note that we do not consider "blended transiting planet" false positive scenarios, either due to physically associated or chance-aligned companions. See Section 4.6 for more discussion.

general strategies to try to address this issue. The first is to somehow quantify the absolute goodness-of-fit of the models, and require that a validated planet pass some threshold test. The other strategy, which we adopt here, is to expand the set of models to be more comprehensive. In order to do this, we introduce two artificial models: "boxy" and "long." The "boxy" likelihood function is a step function at some minimum value of T/τ (zero below this threshold, and constant above), and constant throughout the space of the other trapezoidal parameters. Similarly, the likelihood of the "long" model is a step function at some minimum threshold value of duration T. These thresholds are set relative to the simulated trapezoidal shapes of the planet model: the T/τ threshold is the maximum value from the simulated planet population, and the T threshold is the 99% percentile of simulated planet population values. We also choose the model priors for these artificial models to be low, reflecting that we expect only a small number of signals to be unexplained by any of the astrophysical scenarios: the number that we choose for each of these models is 5×10^{-5}, corresponding to an expectation that there may be ∼10 such signals among the ∼200,000 *Kepler* targets.

2.2. Stellar Properties

The most substantial difference between the current implementation of `vespa` and the procedure documented in M12 is how stellar properties are treated. Previously, either the target star's mass and radius were explicitly provided, or they were randomly generated according to the stellar population expected along the line of sight by the TRIdimensional modeL of thE GALaxy (TRILEGAL) Galactic stellar population synthesis tool, but constrained to agree with some observed color(s) of the star (e.g., $J - K$), to within some specified tolerance. This strategy was used both to generate the host stars for the transiting planet model and the binary and triple stars for the EB and HEB false positive models.

The new method now used by `vespa` uses the `isochrones` Python module (Morton 2015a) to fold in observational constraints on the host star. At its core, `isochrones` performs 3D linear interpolation in mass–[Fe/H]–age parameter space for a given stellar model grid. This method of stellar modeling for FPP calculation debuted in Montet et al. (2015) and is explained there in more detail. Instead of randomly generating stars (or binary or triple systems of stars) from a predefined distribution and culling them to approximately agree with observed colors, *all* available constraints on the target star are used to condition a direct fit of either a single-, binary-, or triple-star model to the Dartmouth grid of stellar models (Dotter et al. 2008; Feiden et al. 2011). This fit is done using multi-modal nested sampling, implemented with `MultiNest` (Feroz et al. 2009, 2011, 2013), via the `PyMultiNest` wrapper (Buchner et al. 2014). Monte Carlo samples of stellar properties for the population simulations are then drawn directly from these posterior samples.

As a result, `vespa` creates full posterior samplings of the physical properties of the host star, modeled as a single-, binary-, and triple-star system, as a by-product of the FPP calculation. Parameters directly fitted for in this process are stellar mass, age, [Fe/H], A_V extinction, and distance. For binary and triple fits, secondary and/or tertiary mass parameters are added, with all other parameters assumed to be the same among all components. Photometric observations upon which these fits are conditioned are assumed to be the

Table 1
Priors Used in Stellar Property Fits

Parameter	Prior
Primary mass M_A	$\propto M_A^{-2.35}$, $M_A > 0.1$
Secondary mass M_B	$\propto (M_B/M_A)^{0.3}$, $0.1 <= M_B < M_A$
Tertiary mass M_C	$\propto (M_C/M_B)^{0.3}$, $0.1 <= M_C < M_B$
Age [Gyr]	$\mathcal{U}(1, 15)$ [a]
[Fe/H]	$\frac{0.8}{0.15}\mathcal{N}(0.016, 0.15) + \frac{0.2}{0.22}\mathcal{N}(-0.15, 0.22)$ [b]
A_V [mag]	$\mathcal{U}(0, A_{V,\mathrm{max}})$ [c]
Distance d	$\propto d^2$

Notes.
[a] The age range for the Dartmouth stellar model grids used.
[b] Double-Gaussian fit to the measured local stellar metallicity distribution (Casagrande et al. 2011; Hayden et al. 2015).
[c] Maximum allowed value is the Galactic extinction at infinity calculated along the star's line of sight, according to Schlegel et al. (1998).

sum of all components. If spectroscopic and/or asteroseismic measurements are used (e.g., constraints on effective temperature or stellar surface gravity), they are assumed to relate to only the primary star. Priors used in these fits are listed in Table 1—notably, we use a prior on [Fe/H] based on a double-Gaussian fit to the local metallicity distribution (Casagrande et al. 2011; Hayden et al. 2015). Posterior chains of all other stellar parameters of interest (e.g., temperature, surface gravity, radius, etc.) are derived from the chains of fitting parameters by evaluting the stellar models using `isochrones`.

3. DATA AND CONSTRAINTS

The goal of this work is to calculate the FPP for every KOI, regardless of classification, as CONFIRMED, CANDIDATE, or FALSE POSITIVE. As such, we begin with a list of 7470 KOIs from the Q1–Q17 DR24 table at the NASA Exoplanet Science Institute (NExScI) Exoplanet Archive (the most recent available uniform catalog). We then gather ancillary data and constraints from various sources in order to enable the `vespa` calculation:

1. The R.A./decl. coordinates of each star from the *Kepler* Input Catalog (KIC).
2. *grizJHK* photometry from the KIC, with *griz* bands corrected to the Sloan Digital Sky Survey (SDSS) photometric scale according to Pinsonneault et al. (2012).
3. Stellar T_{eff}, [Fe/H], and $\log g$ values and uncertainties from the Huber et al. (2014) stellar properties catalog, if the provenance of these values is from spectroscopy or asteroseismology.
4. Detrended *Kepler* photometry used for the MCMC modeling of Rowe et al. (2015), along with information about individually fitted transit times, where available.
5. Best-fit $R_p/R_\star$ from the Rowe et al. (2015) MCMC analysis.
6. Centroid uncertainty information from the NExScI Exoplanet Archive: we assume that the allowed "exclusion" radius for a blend scenario is $3\times$ the uncertainty in the fitted centroid position (the `koi_dicco_msky_err` column in the Archive table). We floor this value at $0\rlap{.}''5$, to prevent unrealistically small exclusion radii. If this quantity is not available from the archive we set a default exclusion radius of $4''$.
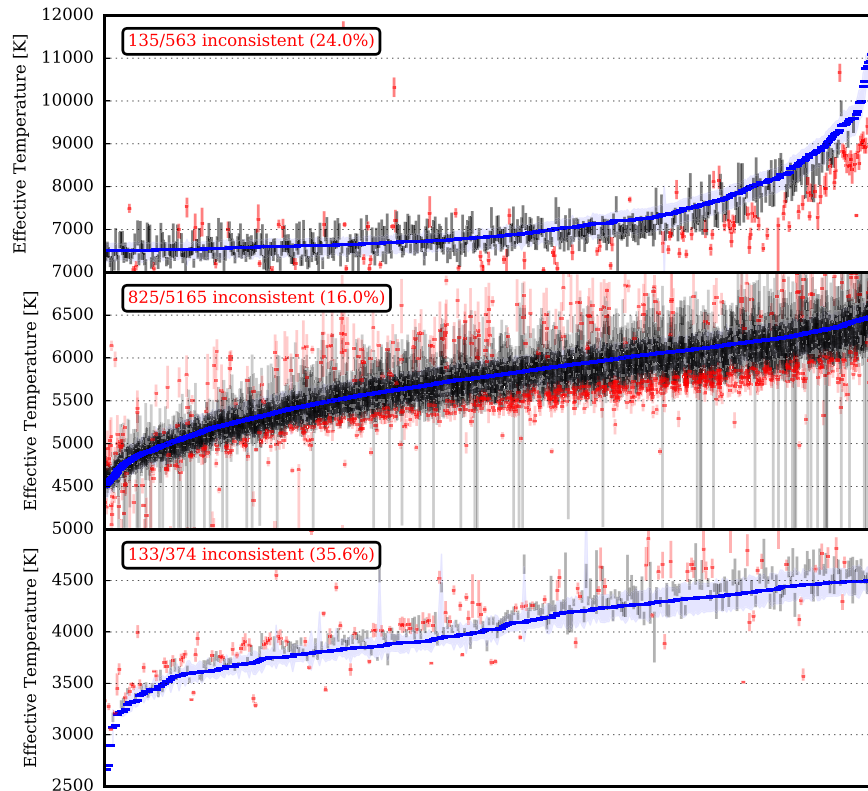
Figure 1. Comparison between effective temperatures estimated from the `isochrones` analysis in this work and those from the *Kepler* stellar parameters catalog (Huber et al. 2014, hereafter H14). The bottom panel shows stars for which H14 predicts $T_{eff} < 4500$ K, the middle spans 4500 K $< T_{eff} < 6500$ K, and the top has $T_{eff} > 6500$ K. Blue horizontal bold lines are the H14 values in sorted order; blue shading represents the error bars from H14. Vertical lines span the 1σ credible region of the `isochrones` fits; these lines are gray if they overlap with the H14 1σ region and red (with the median marked by a point) if they are inconsistent. This comparison shows that the stellar parameters estimated in this work are broadly consistent with H14, though less so for the coolest and hottest stars.

7. The maximum secondary eclipse depth allowed by the *Kepler* photometry. This quantity is derived by searching the phased-folded KOI light curve for the deepest signal at any other phase other than that of the primary transit. This "model-shift uniqueness test" is described in both Section 3.2.2 of Rowe et al. (2015) and Coughlin et al. (2015a), and the values of these metrics for the Q1–Q17 DR24 release will soon be published (Coughlin et al. 2015b). The maximum secondary depth we use is

$$\delta_{max} = \delta_{sec} + 3\sigma_{sec}, \qquad (3)$$

where δ_{sec} is the fitted depth and σ_{sec} is the uncertainty on that depth (including red noise). As the DR24 pipeline uses two different detrending methods, we calculate δ_{max} and σ_{max} using both methods and take the maximum between the two. When these metrics are not available for a particular KOI, we default to $10\times$ the uncertainty in the *Kepler* pipeline measured transit depth (`koi_depth_err1`).

As explained in the `vespa` documentation, we first specify this ancillary data in a `star.ini` and `fpp.ini` file for each KOI, and then for each we run the command-line script `calcfpp`. This end-to-end calculation (which includes the `isochrones` fits for single-, double-, and triple-star models) takes approximately 30 minutes per KOI on a single core, allowing the entire set to be calculated in approximately one day on the Princeton Univeristy "Tigress" computing cluster, using 200 cores.

4. RESULTS

The results of the `vespa` calculations are presented in Tables 5 and 6, and are discussed in the following subsections.

4.1. Stellar Properties

As discussed in Section 2.2, `vespa` fits for stellar properties as part of its FPP-calculating procedure, using the `isochrones` package. Thus, we obtain posterior samplings of the physical properties of each KOI as a side effect of this batch calculation, a result of general interest independent of FPP. Table 5 presents summarized results of these single-star fits. While `vespa` also fits double- and triple-star models for each KOI, these are of less general interest and so we do not present them separately.

Figures 1 and 2 compare the estimated effective temperatures, metallicities, and radii derived in this work to those independently determined for (or compiled by) the official *Kepler* stellar properties catalog (Huber et al. 2014 hereafter H14). While there is largely general agreement, there are also some discrepancies, highlighting some difficulties of estimating physical stellar properties.

In particular, we note that for stars which H14 list as $T_{eff} < 4000$ K, `isochrones` predicts systematically hotter temperatures. Many of the H14 properties for these stars are taken from Dressing & Charbonneau (2013, D13). Those properties were determined by trying to match the *grizJHK* photometry of a grid of model stars from the Dartmouth models, supplemented by some interpolation. D13 also
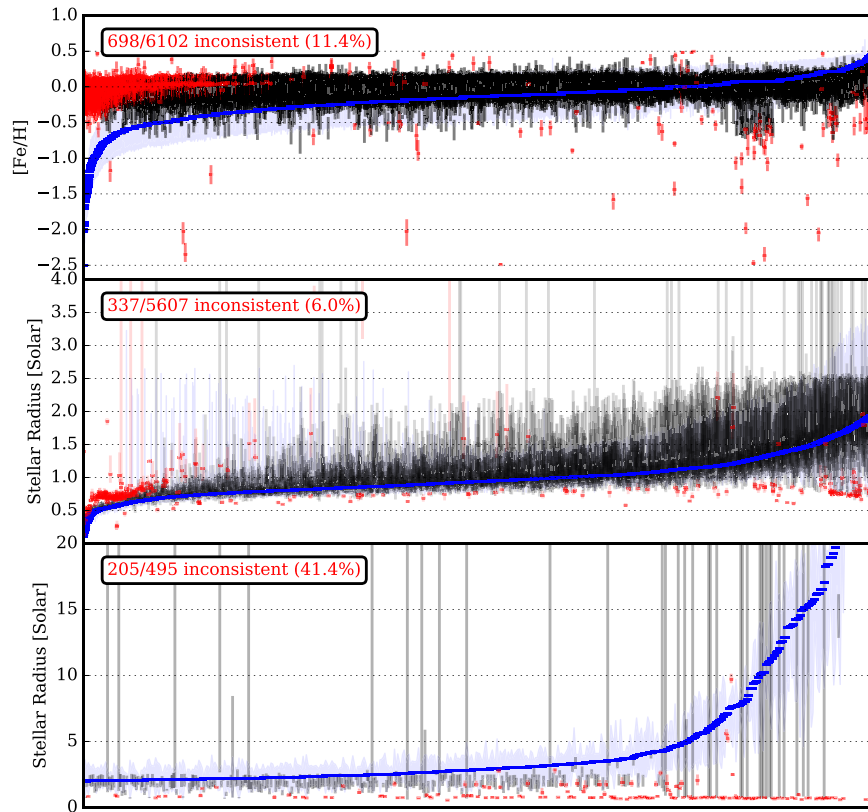
Figure 2. Comparison between metallicities and radii estimated from the `isochrones` analysis in this work and those from the *Kepler* stellar parameters catalog (Huber et al. 2014, hereafter H14). The top panel shows metallicity for all stars in the sample. The middle panel shows stars for which H14 estimates $R_\star < 2\,R_\odot$, and the bottom shows $R_\star > 2\,R_\odot$. Blue horizontal bold lines are the H14 values in sorted order; blue shading represents the error bars from H14. Vertical lines span the 1σ credible region of the `isochrones` fits; these lines are gray if they overlap with the H14 1σ region and red (with the median marked by a point) if they are inconsistent. This comparison shows that the stellar parameters estimated in this work are broadly consistent with H14, though less so for the more evolved stars. The metallicity estimates of the `isochrones` calculations are driven by the use of the local metallicity prior (Casagrande et al. 2011; Hayden et al. 2015, Table 1).

imposed priors on [Fe/H] and the height of stars above the plane of the Galaxy. To validate their methodology, D13 compare their results for 26 nearby stars to the masses predicted for those stars by combining parallax measurements with the Delfosse et al. (2000) relation between mass and absolute *K*-band magnitude. While they find general good agreement, D13 does note that their masses are on average about 5% lower than the Delfosse-predicted masses. Our estimated masses for these stars are typically ~10%–15% higher than those estimated by D13.

The same data (*grizJHK* photometry) and the same stellar models (Dartmouth) were used for both this work and D13, raising the question of the origin of the systematic differences between these methods. The primary origin of this discrepancy appears to be the fact that `isochrones` performs a full multi-modal posterior exploration of the stellar parameter space, marginalizing over the unknown A_V extinction in the process, while D13 uses a fixed 1 mag of *V*-band extinction per 1000 pc and selects the maximum-likelihood match to the grid of models. As we allow for a maximum extinction up to the measured A_V extinction at infinity, not explicitly tied to distance, this typically allows for slightly hotter stars with slightly more extinction than was permitted by Dressing & Charbonneau (2013).

The other significant discrepancy between the `iso-chrones` results and H14 is among evolved stars, as seen in the lower panel of Figure 2. Many of these stars have densities measured via asteroseismology, and `isochrones` does not

unambiguously identify all of them as evolved. However, it should be noted that we do in fact identify over half of them as probably significantly evolved—this is made possible by the multi-modal posterior sampling of `MultiNest` used by `isochrones`. In addition, as the middle panel of Figure 2 shows, even for stars not positively identified as evolved by H14, isochrones often allows for a significant range of stellar radius—also desirable behavior, as H14 estimates the properties for many of these using only broadband photometry as well, which means their true nature is not securely known. The need for caution when estimating the radii of KOI host stars has also been emphasized by Bastien et al. (2014), who find from photometric "flicker" measurements that a significant number of FGK KOI hosts may be slightly evolved.

We emphasize that stellar parameter estimation is not the central goal of this work, nor are the FPP results very sensitive to the exact estimated stellar properties. The exception to this would be if the stellar density estimate is significantly mis-estimated, which would be the case if a star is not properly identified as evolved. However, we note that of the 730 KOIs with host stars $>2\,R_\odot$ at the NExScI Archive, 502 of them already have FALSE POSITIVE designations; additionally, Sliski & Kipping (2014) find a large false positive rate for KOIs with evolved host stars. Given all of these considerations, we believe that potential systematic issues with stellar property determinations in various corners of parameter space do not strongly affect the main results, which are the astrophysical false positive probabilities of thousands of KOIs.
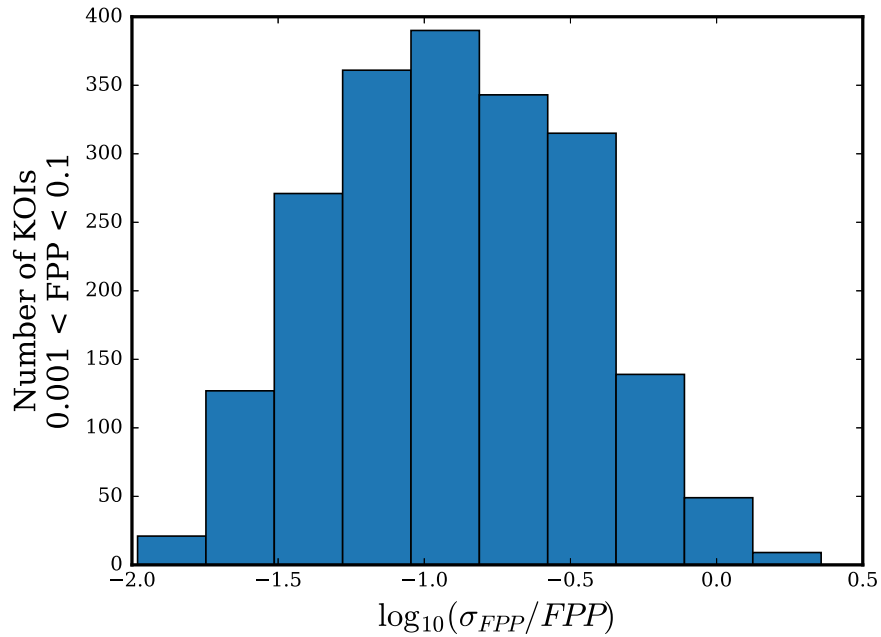
Figure 3. Fractional uncertainties for KOIs with FPPs between 0.001 and 0.1; that is, within an order of magnitude of the validation threshold. FPP values and uncertainties are determined by the mean and standard deviations of vespa calculations based on 10 bootstrap resamplings (see Section 2.1) of a single set of simulated populations for each KOI.

4.2. False Positive Probabilities

Of the 7470 KOIs in the Q1–Q17 DR24 table at the NExScI Exoplanet Archive, vespa successfully calculates the FPP for 7056. Section 4.7 contains detailed explanations of the failure reasons for the 414 KOIs for which we do not present vespa results. FPPs and their uncertainties are determined as the mean and standard deviations of 10 bootstrap recalculations of the initially simulated populations for each KOI (as described in Section 2.1). These results are listed in Table 6. The median fractional FPP uncertainty for KOIs with $0.001 < $ FPP $ < 0.1$ (within an order of magnitude of the validation threshold) is about 12%, and this distribution is shown in Figure 3.

In order to properly interpret these results, it is necessary to understand the range of applicability of the vespa calculation. First of all, this method selects between different specific explanations for the transit-like signal, and cannot comment on whether the signal might be caused by stellar variability or an instrumental false alarm. Thus, vespa results on low signal-to-noise ratio (S/N) candidates that are not clearly transit-like must be viewed with caution. This being said, the reason for including the artificial "boxy" and "long" models in the model selection calculation (Section 2.1) is to flag signals that do not fit well with any of the astrophysical eclipse models—in fact, the "long" model is preferred ($>50\%$) by 526 KOIs that are already dispositioned as FALSE POSITIVE.

Additionally, an important constraint used in the FPP calculation is the allowed sky area inside which a blended false positive may live. As described in Section 3, this value is taken to be three times the uncertainty on the fitted centroid position from the pixel-level data. However, many KOIs have already been identified to be blended binary false positives displaced from the target star. Some of these are found by detecting significant centroid offsets in the pixel-level data—in these cases, vespa treating the confusion radius simply as the uncertainty in the centroid position will clearly give a misleading result. Other displaced false positives have been

identified as originating from displaced stars by finding KOIs with matching periods and epochs (Coughlin et al. 2014), and often the "parents" of these signals are outside the pixel masks, and so are unable to be detected via centroid-measuring methods. In these cases as well, the vespa assumptions break down, and the FPP calculations will not be valid.

To summarize, the results presented in Table 6 are strictly reliable only for KOIs that have already strongly passed the other *Kepler* vetting tests, and that are not indicated to be clearly poor fits to all the proposed hypotheses. The first cut for this is the KOI disposition: FALSE POSITIVE indicates failure of one or more of these tests (and thus probable invalidity of the vespa calculation). However, because of the generally permissive philosophy of the *Kepler* dispositioning, not all CANDIDATE KOIs have the same level of reliability in their disposition, due an "innocent until proven guilty" philosophy. That is, something is not identified as a FALSE POSITIVE *unless there is positive confirmation of false positive status*. In particular, when pixel-level analysis fails to determine whether the signal is indeed coming from the target star because of low S/N, such a KOI will still receive a CANDIDATE disposition. Therefore, the vespa-calculated FPP may be considered most reliable only when a KOI is designated a CANDIDATE (or CONFIRMED) and has a large enough S/N to enable secure positional determination. In addition, for the greatest reliability we also require a signal's multiple-event statistic (MES; equivalend to S/N) to be greater than 10, in order to avoid low-S/N signals that might be caused by light-curve systematics.

To enable interpretation, Table 6 thus contains the current KOI disposition, MES, and the results of the positional probability calculations of Bryson & Morton (2015). Figure 4 shows FPPs for all KOIs passing the following criteria:

1. Dispositioned CANDIDATE or CONFIRMED at the NExScI archive,
2. MES > 10, indicating the signal is unlikely to be caused by systematic noise in the light curve, and
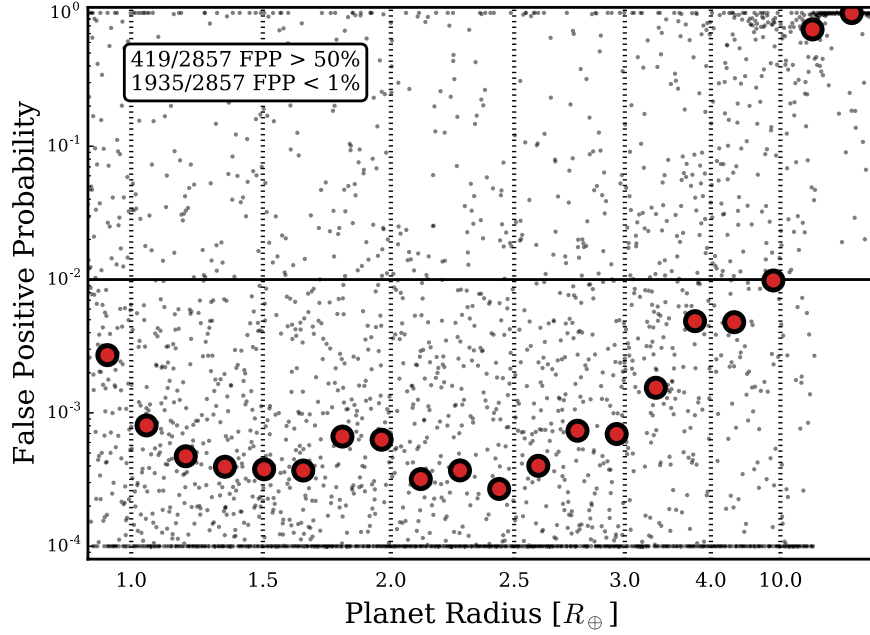
Figure 4. False positive probabilities of all CANDIDATE or CONFIRMED KOIs for which we consider the `vespa` calculations to be reliable (floored at 10^{-4} for visualization purposes), meaning they are considered to be reliably located on the target star (Pr > 0.99, with "score" >0.3) according to pixel-level analysis (Bryson & Morton 2015), and have MES >10. Of these 2857 KOIs, 1935 have FPPs less than 1% (1284 of which have not yet been dispositioned as CONFIRMED). Noteably, 419 are likely false positives (FPP > 0.5), consistent with the Morton & Johnson (2011) and Fressin et al. (2013) a priori estimates of the overall *Kepler* candidate false positive rate. Red circles correspond to median FPP values in equal-sized bins.
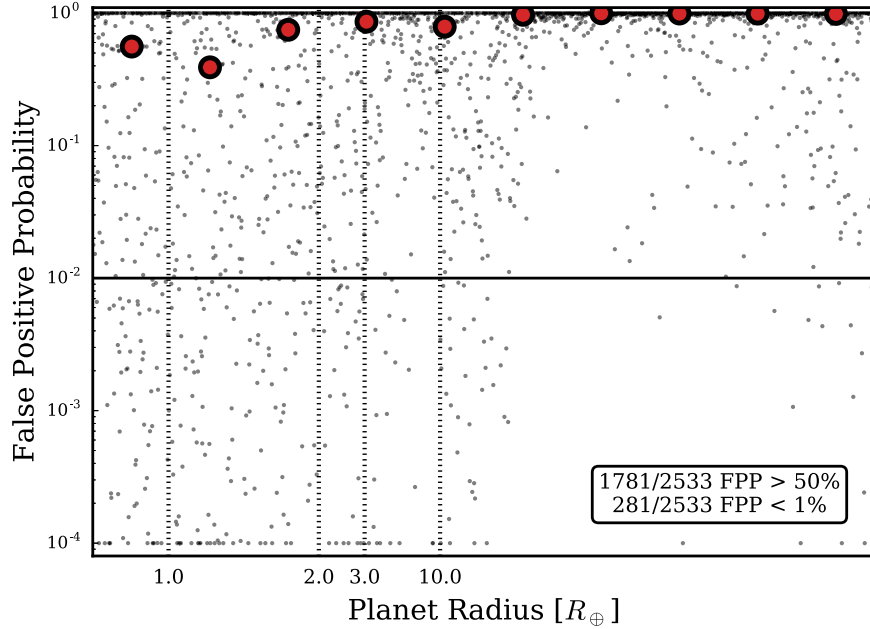
Figure 5. Same as Figure 4, but for KOIs currently dispositioned as FALSE POSITIVE. The vast majority are also identified by `vespa` as likely false positives. There are also some that have low FPPs, but this can be explained by the fact that many of the reasons for dispositioning a KOI as a FALSE POSITIVE also invalidate assumptions made by `vespa`; for example, that the signal is spatially coincident with the target star (see Section 4). This figure illustrates that `vespa` is effective (though not 100% efficient) at recovering known false positives.

3. Probability >0.99 of being on the target star, according to Bryson & Morton (2015), along with a positional probability "score" >0.3 (indicating a reliable result).

These selections leave 2857 KOIs for which the `vespa` results can be considered the most reliable; Table 2 presents the mean FPPs for different subsets, showing several notable features. Single KOIs are about 2.5× more likely to be false positives than KOIs in multiple-KOI systems, in qualitative agreement with Lissauer et al. (2012)—and this is true even without giving any "multiplicity boost" to multi-KOI systems for the increased transit probability of subsequent planets once one planet transits in a coplanar system. Also, large candidates typically have high FPPs, in agreement with Santerne et al.

Table 2
Mean FPPs of Candidate KOIs with Reliable `vespa` Calculations

Selection	Number	Mean FPP
All	2857	0.155
Singles	1688	0.206
Multis	1169	0.082
$R_p > 15\,R_\oplus$	256	0.837
$10\,R_\oplus < R_p < 15\,R_\oplus$	91	0.220
$4\,R_\oplus < R_p < 10\,R_\oplus$	252	0.218
$2\,R_\oplus < R_p < 4\,R_\oplus$	1160	0.066
$R_p < 2\,R_\oplus$	1098	0.071

(2012, 2016). And finally, the mean FPPs are very consistent with the a priori predictions of Morton & Johnson (2011) and with the analysis of Fressin et al. (2013).

4.3. Unidentified Ephemeris Matches?

One potential concern worth addressing in some detail, before deciding which planets to validate based on the `vespa` calculations, is the possibility of false positives caused by distant contamination but not identified through the "period-epoch match" (PEM) technique used by Coughlin et al. (2014), due to the fact that not all stars in the *Kepler* field were monitored by the mission. We thus estimate here the probability that a CANDIDATE KOI to which `vespa` assigns a low FPP might still be caused by such a scenario.

In the Q1–Q17 DR24 KOI table used as the basis for this work, 980 KOIs were identified as PEMs. Of these, 187 were not identified as false positives by any other method. Only 15 of these 187 survived all the quality cuts described in Section 4.2. And of these 15, only 3 have FPP < 0.01. (See Section 5.3 for an example of a KOI caused by a "column anomaly" effect that went unidentified by the *Kepler* pipeline but to which `vespa` assigned a high FPP.) Thus, we expect only about 0.3% (3/980) of as-yet unidentified distant-contamination FPs to end with FPP < 0.01 according to `vespa`. As the fraction of false positives from pixel contamination but unidentified as PEMs among the entire KOI sample is estimated to be something around 23% (Coughlin et al. 2014), we estimate that there remains a small (~0.06%) residual FPP for all KOIs, even when the `vespa`-calculated FPP is negligibly tiny.

4.4. Validation of 1284 New Planets

While the FPP below which to claim planet validation is clearly an arbitrary choice, there is precedent to using FPP < 0.01 as the threshold—Rowe et al. (2014) validated over 800 multi-planet KOIs using this number, and Montet et al. (2015) used it to validate a sample of K2-Campaign 1 planets. Adopting this same threshold and adding the 0.06% residual FPP estimated in Section 4.3 to the `vespa`-calculated value, we find that of the KOIs with reliable `vespa` FPPs, 1935 have FPP < 0.01, and are thus validated at the 99% level. These KOIs are labeled as such in Table 6. These are not all new validations, however, since a number of them are already CONFIRMED. Figure 6 shows a different kind of summary, grouping KOIs by disposition and splitting up the CANDI-DATES according to the `vespa` results, showing that 1284 KOIs are newly validated at the 99% level. Figure 7 shows the radii and periods of the CONFIRMED and CANDIDATE KOIs, with the transparency of the points representing the

`vespa`-calculated FPP. Figure 8 compares the temperature and radii of the stars hosting these planets between the H14 and `isochrones` analysis.

As there is significant interest in identifying potentially habitable planets, Table 3 lists the properties of 9 CANDI-DATE KOIs that are newly validated by this work that *may* fall within the optimistic habitable zones of their host stars (Kopparapu et al. 2013), according to the stellar properties reported by the DR24 table at the NExScI Archive. We note that while more detailed follow-up observations (imaging and spectroscopy) have been taken for each of these targets, we make no attempt here to characterize these systems in detail. However, we do note that high-resolution imaging of KOI-2418 and KOI-3010 reveal that these two host stars have close companions that may or may not change the habitable nature of the planets (e.g., if the planets happen to transit the secondary star instead of the primary). Additionally, analysis of high-resolution spectroscopy will solidify the properties of all these host stars, affecting the habitable zone boundaries. We thus emphasize that this list is neither complete nor final, serving only to draw attention to new validations of interest rather than to be a definitive statement on potential habitability. We also note that based solely on the properties in the DR24 table, KOI-5475.01 (listed as having a 448 day orbital period) should also be included in this list of potentially habitable-zone planets. However, as explained in Coughlin et al. (2015b, Section 5.5.4), this particular KOI actually has a 224 day period, making its insolation too high to be within the habitable zone; we have thus excluded it from Table 3.

4.5. Likely False Positives

In addition to the confident validations, we identify 428 KOIs that currently have a CANDIDATE disposition but for which `vespa` calculates FPP > 0.9; these KOIs are likely to be false positives. As Figure 7 shows, many of these newly identified false positives have large radii—this is again because of the dispositioning philosophy adopted by the *Kepler* team, which does not use any cut in transit depth or inferred "planet" size to identify FALSE POSITIVES. To identify these likely false positives from the `vespa` calculations, we do not require CANDIDATES to obey the same selections we used to ensure a clean sample for validation. This is because even if a signal has characteristics such that a low `vespa` FPP would not be sufficient to validate it, a high `vespa` FPP is still sufficient to cast doubt on its planetary nature. As a demonstration of the ability of `vespa` to identify false positives, Figure 5 shows the FPPs for KOIs that are dispositioned as FALSE POSITIVE and have MES > 10. The vast majority of these have large FPPs.

One final note about these calculations is that unidentified TTVs will increase the FPP of a transit signal, as the shape of the folded light curve will be distorted, typically resulting in a longer signal which a trapezoid model will also identify as more V-shaped. While we have analyzed light curves correcting for known TTVs when available, we have also undoubtedly missed many systems with as-yet-unspecified TTV signals. This means that a large FPP may be simply an indication of unidentified TTVs rather than an astrophysical false positive, especially in multi-planet systems, which should overall have a very low false positive rate (Lissauer et al. 2014; Rowe et al. 2014). It is thus probable that despite multi-KOIs having lower FPPs than singles (Table 2), even these relatively low FPPs are inflated by the effect
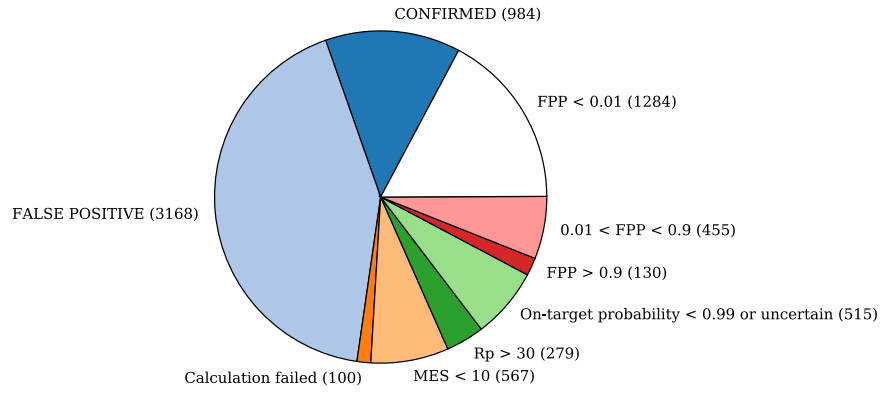
Figure 6. A summary of how the the calculations presented in this paper advance our understanding of the true nature of KOIs. More than half of all KOIs to date have already been dispositioned FALSE POSITIVE or CONFIRMED. For those dispositioned CANDIDATE we further categorize them according to their reliability. The "no reliable calculation" category means that the vespa calculation was not successful. "On-target probability uncertain" indicates that the positional probability calculations of Bryson & Morton (2015) are not reliable (score < 0.3). "On-target probability <0.99" means that the positional probability calculations indicate that there is a non-negligible chance that the source of the transit signal is not at the position of the KOI. The remaining three categories are all CANDIDATE KOIs reliably confirmed to be located at the presumed target star, grouped by false positive probability. One thousand two hundred eighty four of these are new planet validations.
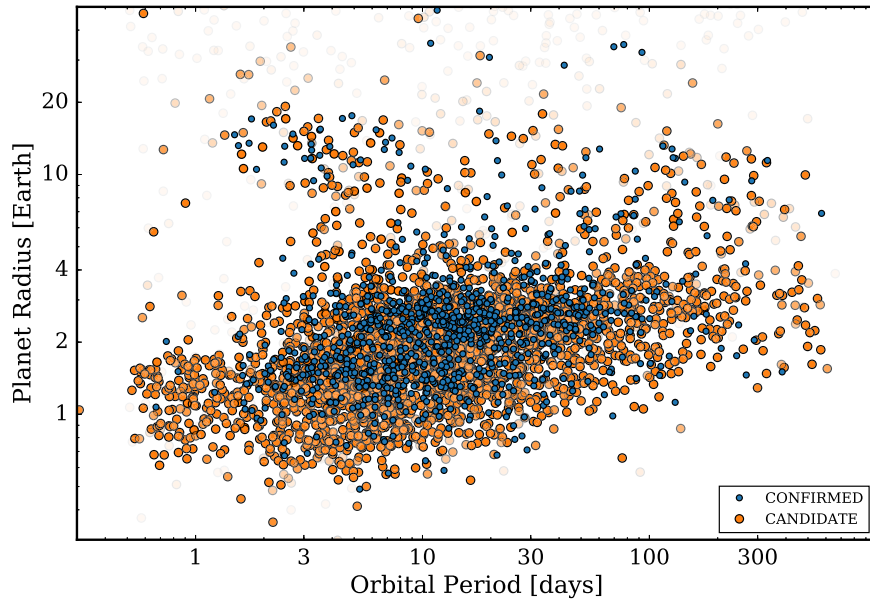
Figure 7. Periods and radii of KOIs with CANDIDATE and CONFIRMED dispositions. Blue circles have previously been identified as CONFIRMED. Candidates are orange circles, shaded by false positive probability, with a transparent circle representing a high FPP.

of TTVs. Table 6 lists whether known TTVs were accounted for when constructing the folded transit light curve for each KOI.

4.6. Blended Transiting Planets?

In this work, the only astrophysical false positive scenarios we consider are eclipsing binary stars (EB, HEB, BEB, and the double-period versions thereof). Previous work studying *Kepler* false positive rates (e.g., Fressin et al. 2013) has also considered the "blended transiting planet" to be a false positive—i.e., a fainter companion star hosting a transiting planet larger than what would be inferred if it were transiting the target star. Because we do not consider such a scenario to be a false positive, the vespa analysis presented here does not quantify its probability. As a result, we are *not* able to unambiguously determine the radii of the planets we validate—all the planet radii listed in Table 6 are based on the assumption that the planet transits the target star and that the target star is unblended. If the target star is actually a member of an unresolved binary system

(as a significant fraction of KOIs undoubtedly are), then the true planet radius will be larger (significantly larger if transiting a fainter companion). This was indeed the motivation for Fressin et al. (2013) to consider the "blended transiting planet" as a false positive; part of the goal of that work was to compute the planet occurrence rates in different radius bins. However, trying to distinguish between single and binary target star configurations is beyond the scope of this work; we thus follow the precedent of Lissauer et al. (2014, especially Section 5) by acknowledging the potential for substantial radius uncertainties among the validated planet sample, while nevertheless defending the validations themselves as robust.

4.7. Failure Modes

We were not able to successfully run vespa on all of the KOIs. The various reasons for these occasional failures are detailed below.
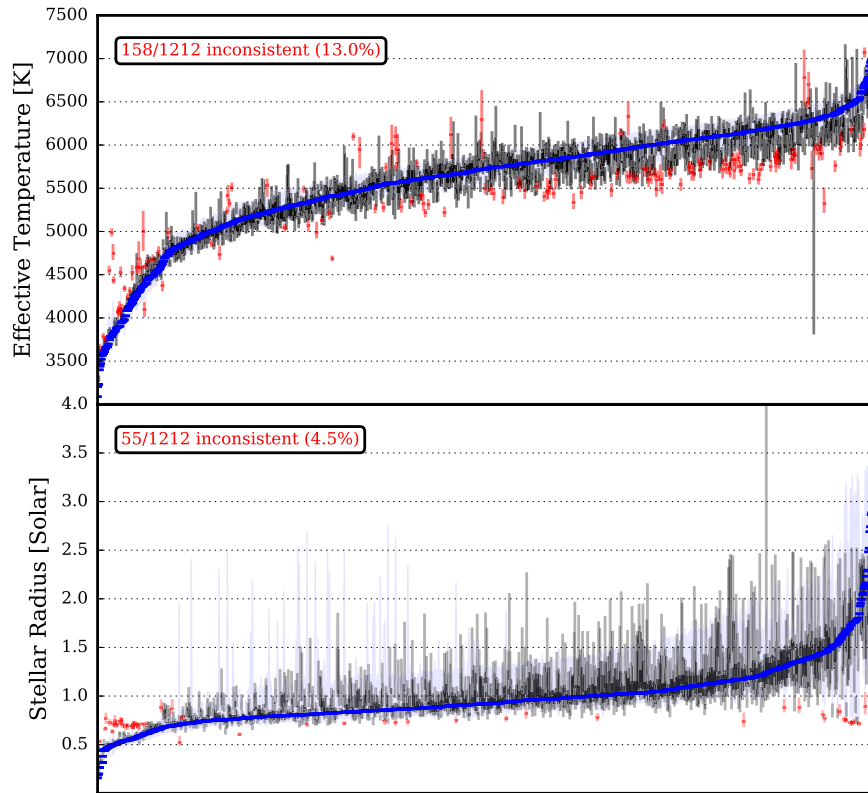
Figure 8. Comparison between temperatures and radii estimated from the `isochrones` analysis in this work and those from the *Kepler* stellar parameters catalog H14, for the sample of stars hosting planets validated in this study. As seen in previous figures, the `isochrones` analysis tends to overpredict the stellar radii compared to H14 for the coolest stars. A small number of stars are also estimated by the photometric analysis of H14 to be evolved, but not by `isochrones`.

Table 3
Newly Validated Planets in the Optimistic Habitable Zone

KOI	Kepler Name	Period (days)	R_P $R_\oplus$	F_p F_e	T_{eff} (K)	$\log g$	$R_\star$ $R_\odot$	$M_\star$ $M_\odot$
463.01	Kepler-560 b	18.478	$1.57^{+0.26}_{-0.26}$	$1.26^{+0.58}_{-0.43}$	3387^{+59}_{-50}	$4.96^{+0.10}_{-0.10}$	$0.30^{+0.05}_{-0.05}$	$0.30^{+0.05}_{-0.05}$
854.01	Kepler-705 b	56.056	$1.96^{+0.25}_{-0.25}$	$0.64^{+0.23}_{-0.18}$	3593^{+58}_{-66}	$4.78^{+0.08}_{-0.08}$	$0.47^{+0.06}_{-0.06}$	$0.49^{+0.06}_{-0.06}$
2418.01	Kepler-1229 b	86.829	$1.12^{+0.13}_{-0.22}$	$0.35^{+0.12}_{-0.14}$	3724^{+60}_{-74}	$4.84^{+0.09}_{-0.08}$	$0.41^{+0.05}_{-0.08}$	$0.43^{+0.05}_{-0.07}$
3010.01	Kepler-1410 b	60.866	$1.56^{+0.15}_{-0.15}$	$0.93^{+0.25}_{-0.21}$	3903^{+50}_{-60}	$4.74^{+0.06}_{-0.06}$	$0.52^{+0.05}_{-0.05}$	$0.54^{+0.05}_{-0.05}$
3282.01	Kepler-1455 b	49.277	$1.97^{+0.25}_{-0.19}$	$1.30^{+0.50}_{-0.34}$	3894^{+83}_{-101}	$4.71^{+0.06}_{-0.07}$	$0.54^{+0.07}_{-0.06}$	$0.55^{+0.06}_{-0.06}$
4036.01	Kepler-1544 b	168.811	$1.83^{+4.73}_{-0.17}$	$1.02^{+13.70}_{-0.25}$	4893^{+141}_{-110}	$4.54^{+0.06}_{-0.98}$	$0.76^{+1.97}_{-0.07}$	$0.73^{+0.28}_{-0.06}$
4356.01	Kepler-1593 b	174.510	$1.91^{+0.16}_{-0.21}$	$0.29^{+0.09}_{-0.09}$	4366^{+131}_{-166}	$4.82^{+0.06}_{-0.05}$	$0.46^{+0.04}_{-0.05}$	$0.49^{+0.03}_{-0.03}$
4450.01	Kepler-1606 b	196.435	$1.98^{+0.72}_{-0.15}$	$1.38^{+1.48}_{-0.32}$	5536^{+161}_{-140}	$4.57^{+0.03}_{-0.24}$	$0.82^{+0.29}_{-0.06}$	$0.90^{+0.09}_{-0.09}$
5856.01	Kepler-1638 b	259.337	$1.70^{+0.76}_{-0.21}$	$1.47^{+2.03}_{-0.44}$	5906^{+183}_{-147}	$4.47^{+0.10}_{-0.29}$	$0.85^{+0.38}_{-0.10}$	$0.77^{+0.10}_{-0.04}$

Note. This table lists CANDIDATE KOIs validated in this work that may lie within the optimistic habitable zones of their host stars. The stellar and planetary properties for this table are taken from the DR24 table at the NExScI Exoplanet Archive. Further individual study of each of these systems using detailed follow-up observations will either solidify or amend their potentially habitable nature. In particular, we note that high-resolution imaging observations on the CFOP archive reveal both KOI-2418 and KOI-3010 to have close companions which may or may not affect their habitable nature. http://cfop.ipac.caltech.edu.

Table 4
`vespa`-calculated FPPs of the Santerne (2015) RV Sample

RV-based Nature	Number	Mean FPP	Median FPP
Planet	43	0.1	3.6e-05
Brown dwarf	3	0.012	0.0026
Eclipsing binary (EB)	43	0.75	0.97
Contaminating EB	13	0.78	0.99
Undetermined	18	0.31	0.01

1. 28 KOIs did not receive MCMC modeling. Most of these have been are already designated FALSE POSITIVE at the archive.
2. 233 KOIs did receive MCMC modeling but had unphysical fit results; e.g., negative $R_p/R_\star$ or best-fit impact parameter greater than $(1 + R_p/R_\star)$. These KOIs were left out of the `vespa` calculations.
3. The host stars of 74 KOIs are not included in the H14 stellar property catalog, and thus were not included in this analysis.

Table 5
Stellar Properties

Star	$M_\star$ ($M_\odot$)	$R_\star$ ($R_\odot$)	$T_{\rm eff}$ (K)	$\log g$ (cgs)	[Fe/H] (dex)	Age (Gyr)	d (pc)	A_V (mag)	$\pi(T_{\rm eff})$	$\pi(\log g)$	$\pi([{\rm Fe/H}])$
K01122	$0.93^{+0.09}_{-0.06}$	$0.91^{+0.15}_{-0.08}$	5582^{+282}_{-206}	$4.49^{+0.06}_{-0.11}$	$-0.03^{+0.15}_{-0.17}$	$9.61^{+0.34}_{-0.38}$	1002^{+182}_{-100}	$0.34^{+0.23}_{-0.20}$	⋯	⋯	⋯
K01123	$0.99^{+0.05}_{-0.05}$	$0.97^{+0.18}_{-0.08}$	5706^{+76}_{-115}	$4.46^{+0.06}_{-0.15}$	$0.05^{+0.14}_{-0.16}$	$9.64^{+0.28}_{-0.39}$	1114^{+213}_{-96}	$0.26^{+0.05}_{-0.10}$	⋯	⋯	⋯
K01124	$1.06^{+0.09}_{-0.08}$	$1.09^{+0.23}_{-0.13}$	5975^{+139}_{-201}	$4.39^{+0.09}_{-0.15}$	$0.02^{+0.14}_{-0.18}$	$9.56^{+0.21}_{-0.31}$	1126^{+240}_{-145}	$0.38^{+0.10}_{-0.16}$	⋯	⋯	⋯
K01125	$1.07^{+0.08}_{-0.06}$	$1.09^{+0.27}_{-0.11}$	5884^{+95}_{-136}	$4.39^{+0.08}_{-0.17}$	$0.12^{+0.14}_{-0.14}$	$9.59^{+0.19}_{-0.35}$	1422^{+355}_{-155}	$0.29^{+0.07}_{-0.11}$	⋯	⋯	⋯
K01126	$0.93^{+0.10}_{-0.08}$	$0.91^{+0.15}_{-0.09}$	5618^{+307}_{-247}	$4.49^{+0.07}_{-0.11}$	$-0.06^{+0.15}_{-0.18}$	$9.60^{+0.34}_{-0.38}$	906^{+168}_{-105}	$0.51^{+0.26}_{-0.23}$	⋯	⋯	⋯
K01127	$0.92^{+0.04}_{-0.05}$	$0.89^{+0.11}_{-0.06}$	5509^{+64}_{-85}	$4.51^{+0.04}_{-0.11}$	$0.02^{+0.15}_{-0.18}$	$9.66^{+0.35}_{-0.43}$	1193^{+142}_{-83}	$0.21^{+0.04}_{-0.08}$	⋯	⋯	⋯
K01128	$0.92^{+0.03}_{-0.03}$	$0.84^{+0.02}_{-0.02}$	5468^{+68}_{-70}	$4.56^{+0.01}_{-0.01}$	$-0.04^{+0.10}_{-0.10}$	$9.27^{+0.20}_{-0.16}$	400^{+12}_{-12}	$0.13^{+0.07}_{-0.07}$	(5487, 100)	(4.56, 0.01)	(0.00, 0.14)
K01129	$0.83^{+0.04}_{-0.04}$	$0.78^{+0.04}_{-0.03}$	5071^{+57}_{-87}	$4.58^{+0.02}_{-0.05}$	$0.05^{+0.16}_{-0.16}$	$9.64^{+0.38}_{-0.40}$	727^{+39}_{-37}	$0.29^{+0.05}_{-0.11}$	⋯	⋯	⋯
K01130	$0.86^{+0.05}_{-0.06}$	$0.83^{+0.07}_{-0.06}$	5404^{+66}_{-55}	$4.55^{+0.04}_{-0.08}$	$-0.12^{+0.15}_{-0.18}$	$9.64^{+0.39}_{-0.41}$	886^{+74}_{-73}	$0.06^{+0.06}_{-0.04}$	⋯	⋯	⋯
K01132	$1.39^{+0.23}_{-0.18}$	$1.60^{+0.66}_{-0.33}$	6564^{+316}_{-303}	$4.17^{+0.14}_{-0.23}$	$0.06^{+0.14}_{-0.13}$	$9.26^{+0.16}_{-0.13}$	1525^{+611}_{-329}	$0.48^{+0.20}_{-0.19}$	⋯	⋯	⋯
K01133	$0.97^{+0.06}_{-0.05}$	$0.97^{+0.26}_{-0.09}$	5551^{+58}_{-76}	$4.46^{+0.07}_{-0.20}$	$0.17^{+0.15}_{-0.13}$	$9.75^{+0.26}_{-0.49}$	1253^{+344}_{-116}	$0.36^{+0.03}_{-0.06}$	⋯	⋯	⋯
K01134	$0.86^{+0.03}_{-0.04}$	$0.81^{+0.04}_{-0.03}$	5097^{+54}_{-58}	$4.56^{+0.03}_{-0.06}$	$0.15^{+0.15}_{-0.17}$	$9.64^{+0.40}_{-0.40}$	937^{+53}_{-42}	$0.13^{+0.04}_{-0.07}$	⋯	⋯	⋯
K01135	$1.22^{+0.24}_{-0.17}$	$1.34^{+0.59}_{-0.28}$	6291^{+350}_{-301}	$4.26^{+0.15}_{-0.24}$	$0.01^{+0.13}_{-0.18}$	$9.39^{+0.20}_{-0.19}$	1454^{+645}_{-329}	$0.63^{+0.22}_{-0.21}$	⋯	⋯	⋯
K01137	$0.87^{+0.04}_{-0.04}$	$0.84^{+0.05}_{-0.04}$	5323^{+97}_{-81}	$4.53^{+0.04}_{-0.05}$	$-0.02^{+0.11}_{-0.11}$	$9.80^{+0.24}_{-0.40}$	489^{+27}_{-24}	$0.15^{+0.10}_{-0.09}$	(5328, 115)	(4.50, 0.06)	(0.00, 0.16)
K01138	$0.97^{+0.04}_{-0.05}$	$0.95^{+0.15}_{-0.07}$	5668^{+82}_{-87}	$4.47^{+0.05}_{-0.14}$	$0.02^{+0.12}_{-0.15}$	$9.64^{+0.30}_{-0.39}$	1289^{+207}_{-99}	$0.10^{+0.06}_{-0.07}$	⋯	⋯	⋯
K01139	$1.51^{+0.22}_{-0.16}$	$1.86^{+0.67}_{-0.37}$	6799^{+155}_{-240}	$4.07^{+0.16}_{-0.21}$	$0.06^{+0.13}_{-0.15}$	$9.19^{+0.12}_{-0.09}$	977^{+348}_{-196}	$0.47^{+0.09}_{-0.15}$	⋯	⋯	⋯
K01140	$1.07^{+0.17}_{-0.11}$	$1.11^{+0.37}_{-0.18}$	5980^{+301}_{-286}	$4.37^{+0.12}_{-0.21}$	$0.01^{+0.13}_{-0.17}$	$9.51^{+0.27}_{-0.28}$	517^{+177}_{-92}	$0.74^{+0.20}_{-0.23}$	⋯	⋯	⋯
K01141	$0.58^{+0.02}_{-0.03}$	$0.56^{+0.02}_{-0.02}$	4043^{+36}_{-34}	$4.71^{+0.01}_{-0.01}$	$-0.17^{+0.09}_{-0.11}$	$9.50^{+0.39}_{-0.30}$	383^{+18}_{-20}	$0.05^{+0.06}_{-0.03}$	(3961, 82)	(4.71, 0.02)	(−0.16, 0.16)
K01142	$0.81^{+0.04}_{-0.05}$	$0.77^{+0.04}_{-0.05}$	5102^{+72}_{-62}	$4.58^{+0.03}_{-0.05}$	$-0.09^{+0.15}_{-0.19}$	$9.72^{+0.32}_{-0.42}$	899^{+50}_{-63}	$0.09^{+0.08}_{-0.05}$	⋯	⋯	⋯
K01143	$0.97^{+0.09}_{-0.07}$	$0.95^{+0.14}_{-0.09}$	5724^{+207}_{-195}	$4.47^{+0.07}_{-0.12}$	$-0.03^{+0.14}_{-0.20}$	$9.59^{+0.29}_{-0.37}$	1061^{+172}_{-109}	$0.29^{+0.17}_{-0.17}$	⋯	⋯	⋯
K01144	$0.94^{+0.06}_{-0.06}$	$0.92^{+0.12}_{-0.08}$	5687^{+157}_{-133}	$4.49^{+0.06}_{-0.11}$	$-0.08^{+0.16}_{-0.19}$	$9.60^{+0.31}_{-0.36}$	1071^{+143}_{-101}	$0.18^{+0.13}_{-0.12}$	⋯	⋯	⋯
K01145	$1.07^{+0.04}_{-0.04}$	$1.09^{+0.09}_{-0.07}$	5974^{+96}_{-95}	$4.40^{+0.05}_{-0.06}$	$0.02^{+0.09}_{-0.09}$	$9.53^{+0.18}_{-0.27}$	830^{+68}_{-57}	$0.18^{+0.08}_{-0.08}$	(5944, 109)	(4.38, 0.07)	(0.00, 0.12)
K01146	$0.46^{+0.03}_{-0.03}$	$0.44^{+0.03}_{-0.03}$	3804^{+36}_{-34}	$4.81^{+0.03}_{-0.03}$	$-0.43^{+0.08}_{-0.09}$	$9.59^{+0.40}_{-0.40}$	224^{+18}_{-17}	$0.05^{+0.06}_{-0.04}$	(3784, 78)	(4.82, 0.06)	(−0.42, 0.16)
K01147	$1.09^{+0.11}_{-0.08}$	$1.15^{+0.30}_{-0.15}$	6067^{+119}_{-172}	$4.36^{+0.10}_{-0.18}$	$0.01^{+0.13}_{-0.18}$	$9.52^{+0.19}_{-0.27}$	650^{+170}_{-88}	$0.23^{+0.08}_{-0.12}$	⋯	⋯	⋯
K01148	$1.17^{+0.08}_{-0.05}$	$1.35^{+0.26}_{-0.18}$	6127^{+57}_{-59}	$4.24^{+0.11}_{-0.13}$	$0.07^{+0.08}_{-0.07}$	$9.57^{+0.09}_{-0.14}$	1019^{+196}_{-132}	$0.04^{+0.04}_{-0.03}$	(6071, 95)	(4.20, 0.13)	(0.10, 0.11)
K01150	$1.01^{+0.03}_{-0.03}$	$1.02^{+0.10}_{-0.06}$	5769^{+63}_{-57}	$4.42^{+0.06}_{-0.08}$	$0.05^{+0.09}_{-0.09}$	$9.69^{+0.18}_{-0.35}$	520^{+50}_{-33}	$0.07^{+0.05}_{-0.05}$	(5718, 109)	(4.39, 0.08)	(0.08, 0.13)
K01151	$0.90^{+0.04}_{-0.05}$	$0.87^{+0.05}_{-0.04}$	5555^{+60}_{-59}	$4.52^{+0.04}_{-0.05}$	$-0.13^{+0.12}_{-0.12}$	$9.66^{+0.24}_{-0.36}$	418^{+23}_{-21}	$0.08^{+0.05}_{-0.05}$	(5528, 113)	(4.50, 0.05)	(−0.20, 0.18)
K01152	$0.63^{+0.00}_{-0.06}$	$0.59^{+0.00}_{-0.05}$	4183^{+20}_{-32}	$4.69^{+0.03}_{-0.00}$	$-0.16^{+0.03}_{-0.23}$	$9.02^{+0.01}_{-0.01}$	325^{+4}_{-7}	$0.20^{+0.03}_{-0.02}$	(3806, 50)	(4.77, 0.07)	(−0.13, 0.10)
K01153	$0.99^{+0.07}_{-0.07}$	$1.00^{+0.17}_{-0.10}$	5889^{+142}_{-123}	$4.44^{+0.07}_{-0.13}$	$-0.12^{+0.17}_{-0.21}$	$9.60^{+0.24}_{-0.36}$	904^{+159}_{-100}	$0.14^{+0.10}_{-0.09}$	⋯	⋯	⋯
K01154	$1.58^{+0.18}_{-0.17}$	$1.97^{+0.53}_{-0.41}$	7066^{+315}_{-265}	$4.05^{+0.15}_{-0.16}$	$-0.03^{+0.16}_{-0.18}$	$9.14^{+0.10}_{-0.07}$	982^{+262}_{-199}	$0.26^{+0.16}_{-0.15}$	⋯	⋯	⋯
K01156	$0.83^{+0.04}_{-0.04}$	$0.79^{+0.04}_{-0.04}$	5144^{+51}_{-54}	$4.57^{+0.03}_{-0.05}$	$-0.04^{+0.14}_{-0.16}$	$9.72^{+0.32}_{-0.42}$	748^{+43}_{-42}	$0.10^{+0.04}_{-0.06}$	⋯	⋯	⋯
K01158	$1.05^{+0.07}_{-0.07}$	$1.06^{+0.21}_{-0.10}$	5951^{+111}_{-121}	$4.41^{+0.07}_{-0.15}$	$-0.01^{+0.14}_{-0.17}$	$9.58^{+0.20}_{-0.33}$	1448^{+293}_{-148}	$0.15^{+0.08}_{-0.09}$	⋯	⋯	⋯
K01159	$0.78^{+0.04}_{-0.04}$	$0.74^{+0.03}_{-0.04}$	4910^{+76}_{-62}	$4.60^{+0.02}_{-0.04}$	$-0.05^{+0.16}_{-0.16}$	$9.69^{+0.34}_{-0.40}$	647^{+30}_{-35}	$0.11^{+0.09}_{-0.07}$	⋯	⋯	⋯
K01160	$0.81^{+0.04}_{-0.05}$	$0.78^{+0.04}_{-0.06}$	5162^{+65}_{-64}	$4.58^{+0.03}_{-0.06}$	$-0.10^{+0.15}_{-0.20}$	$9.69^{+0.35}_{-0.42}$	1036^{+64}_{-80}	$0.10^{+0.06}_{-0.07}$	⋯	⋯	⋯
K01161	$0.97^{+0.06}_{-0.05}$	$0.95^{+0.17}_{-0.07}$	5605^{+67}_{-88}	$4.47^{+0.06}_{-0.15}$	$0.10^{+0.15}_{-0.15}$	$9.69^{+0.27}_{-0.45}$	699^{+129}_{-55}	$0.57^{+0.04}_{-0.07}$	⋯	⋯	⋯
K01162	$0.98^{+0.05}_{-0.05}$	$1.01^{+0.13}_{-0.09}$	5867^{+70}_{-62}	$4.42^{+0.08}_{-0.11}$	$-0.14^{+0.13}_{-0.13}$	$9.72^{+0.17}_{-0.37}$	412^{+52}_{-37}	$0.05^{+0.05}_{-0.04}$	(5751, 108)	(4.34, 0.13)	(−0.28, 0.16)
K01163	$0.94^{+0.04}_{-0.05}$	$0.92^{+0.12}_{-0.06}$	5584^{+78}_{-106}	$4.50^{+0.05}_{-0.12}$	$0.01^{+0.14}_{-0.17}$	$9.67^{+0.31}_{-0.42}$	874^{+115}_{-66}	$0.22^{+0.06}_{-0.10}$	⋯	⋯	⋯
K01164	$0.54^{+0.02}_{-0.03}$	$0.51^{+0.02}_{-0.02}$	3775^{+32}_{-33}	$4.75^{+0.02}_{-0.02}$	$-0.04^{+0.07}_{-0.08}$	$9.58^{+0.40}_{-0.35}$	180^{+10}_{-11}	$0.04^{+0.05}_{-0.03}$	(3684, 55)	(4.78, 0.07)	(0.07, 0.10)
K01165	$0.93^{+0.03}_{-0.04}$	$0.89^{+0.04}_{-0.03}$	5521^{+67}_{-53}	$4.51^{+0.03}_{-0.05}$	$0.01^{+0.10}_{-0.10}$	$9.59^{+0.27}_{-0.35}$	544^{+28}_{-22}	$0.05^{+0.06}_{-0.04}$	(5357, 113)	(4.50, 0.05)	(0.00, 0.16)
K01166	$1.03^{+0.07}_{-0.06}$	$1.05^{+0.31}_{-0.11}$	5826^{+86}_{-118}	$4.41^{+0.08}_{-0.20}$	$0.06^{+0.15}_{-0.15}$	$9.65^{+0.21}_{-0.37}$	1368^{+401}_{-148}	$0.23^{+0.05}_{-0.10}$	⋯	⋯	⋯
K01167	$0.88^{+0.04}_{-0.05}$	$0.84^{+0.07}_{-0.05}$	5353^{+78}_{-101}	$4.54^{+0.04}_{-0.08}$	$-0.02^{+0.15}_{-0.16}$	$9.65^{+0.37}_{-0.42}$	586^{+50}_{-38}	$0.21^{+0.06}_{-0.11}$	⋯	⋯	⋯
K01169	$1.04^{+0.03}_{-0.02}$	$0.96^{+0.03}_{-0.02}$	5684^{+52}_{-44}	$4.49^{+0.01}_{-0.02}$	$0.23^{+0.06}_{-0.05}$	$9.22^{+0.20}_{-0.15}$	464^{+14}_{-13}	$0.04^{+0.05}_{-0.03}$	(5640, 85)	(4.50, 0.02)	(0.36, 0.08)
K01170	$0.95^{+0.05}_{-0.06}$	$0.93^{+0.13}_{-0.08}$	5687^{+116}_{-107}	$4.48^{+0.06}_{-0.12}$	$-0.07^{+0.15}_{-0.19}$	$9.65^{+0.29}_{-0.41}$	826^{+120}_{-71}	$0.14^{+0.09}_{-0.09}$	⋯	⋯	⋯
K01171	$1.07^{+0.07}_{-0.07}$	$1.11^{+0.28}_{-0.13}$	5957^{+108}_{-142}	$4.38^{+0.09}_{-0.17}$	$0.04^{+0.14}_{-0.15}$	$9.58^{+0.19}_{-0.33}$	1450^{+363}_{-176}	$0.21^{+0.06}_{-0.12}$	⋯	⋯	⋯
K01172	$1.17^{+0.15}_{-0.10}$	$1.26^{+0.46}_{-0.17}$	6202^{+136}_{-165}	$4.30^{+0.11}_{-0.22}$	$0.03^{+0.13}_{-0.16}$	$9.46^{+0.15}_{-0.15}$	625^{+227}_{-89}	$0.21^{+0.08}_{-0.12}$	⋯	⋯	⋯
K01173	$1.16^{+0.16}_{-0.09}$	$1.24^{+0.48}_{-0.17}$	6160^{+102}_{-149}	$4.31^{+0.11}_{-0.22}$	$0.04^{+0.13}_{-0.15}$	$9.48^{+0.15}_{-0.23}$	1003^{+385}_{-145}	$0.25^{+0.05}_{-0.11}$	⋯	⋯	⋯
K01175	$1.00^{+0.07}_{-0.05}$	$1.32^{+0.28}_{-0.26}$	5674^{+75}_{-84}	$4.19^{+0.17}_{-0.17}$	$0.07^{+0.09}_{-0.14}$	$9.96^{+0.08}_{-0.11}$	604^{+128}_{-117}	$0.22^{+0.06}_{-0.08}$	(5676, 110)	(4.10, 0.16)	(0.10, 0.14)
K01176	$0.77^{+0.03}_{-0.04}$	$0.71^{+0.02}_{-0.03}$	4739^{+115}_{-70}	$4.62^{+0.02}_{-0.02}$	$-0.02^{+0.13}_{-0.13}$	$9.45^{+0.43}_{-0.26}$	666^{+29}_{-30}	$0.13^{+0.16}_{-0.09}$	(3806, 388)	(4.76, 0.13)	(−0.06, 0.19)
K01178	$1.54^{+0.20}_{-0.17}$	$1.89^{+0.60}_{-0.38}$	6871^{+237}_{-271}	$4.07^{+0.15}_{-0.19}$	$0.04^{+0.14}_{-0.16}$	$9.17^{+0.12}_{-0.08}$	1193^{+371}_{-239}	$0.36^{+0.14}_{-0.17}$	⋯	⋯	⋯
K01179	$1.11^{+0.12}_{-0.10}$	$1.16^{+0.35}_{-0.15}$	6079^{+152}_{-233}	$4.36^{+0.09}_{-0.20}$	$0.03^{+0.13}_{-0.16}$	$9.51^{+0.19}_{-0.27}$	715^{+216}_{-104}	$0.47^{+0.10}_{-0.18}$	⋯	⋯	⋯

(This table is available in its entirety in machine-readable form.)

Table 6
False Positive Probability Results

KOI	P (day)	TTV?	R_p ($R_\oplus$)	S/N	δ_{sec}[a] (ppm)	r_{excl}[b] ('')	Pr_{EB}[c]	Pr_{EB2}[c]	Pr_{HEB}[c]	Pr_{HEB2}[c]	Pr_{BEB}[c]	Pr_{BEB2}[c]	Pr_{boxy}[d]	Pr_{long}[d]	f_p[e]	p_{pos}[f]	s_{pos}[g]	Disp.[h]	FPP[i]	σ_{FPP}[j]	Failure[k]	Kep[l]
K02360.01	2.304	N	6.64	22.6	42	0.50	0	0	0	0	0	0.0053	0	0.99	0.052	0.00	0.12	FP	1	0	⋯	⋯
K02361.01	5.784	N	2.49	16.7	156	2.19	0.025	0.0016	0.0016	0.00026	0.0095	0.0054	0	0	0.201	1.00	0.25	CA	0.043	0.0026	⋯	⋯
K02362.01	2.237	N	1.90	22.3	152	0.63	0	1.5e-05	0	0	0.0027	0.013	0	0	0.189	0.72	1.00	FP	0.016	0.00047	⋯	⋯
K02362.02	11.085	N	2.32	15.8	316	0.87	0.0018	0.00013	8.8e-05	1.6e-05	0.00091	0.00018	0	0	0.215	1.00	0.64	CA	0.0031	0.00054	⋯	1208b
K02363.01	3.139	N	1.03	19.6	51	0.90	0	0.00025	0	2.7e-05	0.029	0.015	0	0	0.101	1.00	0.88	CA	0.044	0.0016	⋯	⋯
K02364.01	5.242	N	1.50	18.1	73	0.54	0	0	0	0	0	0	0.067	0.12	0.146	1.00	1.00	CA	0.19	0.057	⋯	⋯
K02365.01	35.968	N	1.98	19.4	81	1.44	0.02	0.0004	0.00072	4.1e-05	0.0023	0.0012	1.2e-06	0	0.208	1.00	0.88	PL	0.024	0.0021	⋯	⋯
K02365.02	110.974	N	1.45	10.6	65	3.30	0	0	0	0	0.00031	4.1e-05	0.001	0	0.145	1.00	0.59	PL	0.0013	0.0001	⋯	⋯
K02366.01	25.369	N	1.51	18.4	32	1.32	0.0016	0	4.1e-06	0	0	0	0.005	0	0.152	1.00	1.00	CA	0.0066	0.00055	⋯	1209b
K02367.01	6.892	N	1.40	16.4	25	2.16	0.066	0.022	0.015	0.003	0.42	0.41	0	0	0.137	1.00	0.99	CA	0.94	0.0029	⋯	⋯
K02368.01	8.071	N	1.70	16.0	111	1.08	6.3e-05	1.4e-05	0	1.3e-06	0.0002	0.00011	2.3e-06	0	0.177	1.00	0.62	CA	0.0004	4.7e-05	⋯	1210b
K02369.01	11.018	N	2.55	18.3	138	0.78	0	0	0	0	1.3e-05	0	8.2e-06	0	0.202	1.00	0.99	CA	2.2e-05	4.8e-06	⋯	1211b
K02369.03	7.227	N	1.32	7.5	116	2.22	0	0	0	0	2.6e-05	0.00044	0.00015	1e-05	0.130	0.86	0.33	CA	0.00062	0.00012	⋯	⋯
K02370.01	78.732	N	5.39	17.7	188	0.54	0.44	0.0027	0.074	0.0017	0.026	0.0012	0	0	0.051	1.00	1.00	CA	0.54	0.034	⋯	⋯
K02371.01	12.941	N	2.14	19.4	85	1.74	8.1e-05	0	0	0	2.3e-05	0	1.1e-05	0	0.212	1.00	0.99	CA	0.00012	2.7e-05	⋯	1212b
K02372.01	5.350	N	1.19	20.2	18	2.10	0	0	0	0	1.1e-06	0	2.8e-05	0	0.108	1.00	0.78	CA	3e-05	1.6e-06	⋯	1213b
K02373.01	147.281	N	2.19	13.5	176	2.88	0.00042	3.4e-05	3.5e-06	3.3e-06	0.0017	0.004	0.00019	0	0.216	0.97	1.00	CA	0.0064	0.00052	⋯	⋯
K02374.01	5.262	N	1.30	18.5	56	0.75	0	0	0	0	0.00012	0.00016	1.1e-05	1.3e-06	0.132	1.00	0.88	PL	0.00029	5.8e-05	⋯	⋯
K02374.02	12.163	N	1.48	14.8	138	2.52	0.0012	0.004	6.2e-06	0.00058	0.3	0.18	0	0	0.153	1.00	0.56	PL	0.49	0.0055	⋯	⋯
K02375.01	40.879	N	2.00	18.4	109	2.49	0	0.00026	0	6.9e-06	0.4	0.6	0	0	0.209	0.01	0.06	FP	0.99	0.0024	⋯	⋯
K02376.01	18.826	N	2.36	17.1	229	1.23	0.0021	4.1e-06	9.7e-06	0	8.3e-05	4.3e-06	1.1e-05	0	0.214	1.00	0.90	CA	0.0023	0.00063	⋯	1214b
K02377.01	13.903	N	1.50	15.0	131	0.90	2e-05	2.2e-06	0	1.5e-06	0.0025	0.00065	2e-06	0	0.151	0.07	0.88	CA	0.0031	0.00012	⋯	⋯
K02378.01	4.767	N	1.22	12.8	65	1.23	0	6.4e-05	0	2.3e-05	0.0024	0.00051	3.6e-06	0	0.121	1.00	0.64	CA	0.0031	0.00016	⋯	1215b
K02379.01	40.009	N	80.54	17.8	553	1.29	0.34	0.061	0.026	0.0089	0.37	0.2	0	0	0.001	1.00	0.98	FP	1	0	⋯	⋯
K02380.01	6.357	N	1.93	15.6	102	0.78	0.033	0.014	0.0027	0.0019	0.014	0.014	1.4e-06	0	0.219	1.00	0.97	CA	0.081	0.0057	⋯	⋯

Notes.

[a] Maximum secondary eclipse depth allowed.

[b] "Exclusion radius" inside of which false positive scenarios are allowed.

[c] Probabilities for different astrophysical false positive scenarios: unblended eclipsing binary (EB), hierarchical eclipsing binary (HEB), and background/foreground eclipsing binary (BEB). "2" indicates double-period scenario.

[d] Artificial models to identify signals that are poorly described by any of the astrophysical scenarios.

[e] Assumed "specific planet occurrence rate" for this planet.

[f] Probability of signal to be on the target star, according to S. T. Bryson et al. (2015, in preparation).

[g] Positional probability score, from S. T. Bryson et al. (2015, in preparation).

[h] Exoplanet Archive disposition: false positive (FP), candidate (CA), or confirmed (PL).

[i] False positive probability; mean of 10 bootstrap recalculations.

[j] False positive probability uncertainty; standard deviation of 10 bootstrap recalculations.

[k] Reason for failure: (1) No MCMC modeling available from Rowe et al. (2015); (2) Unphysical MCMC fit from Rowe et al. (2015); (3) No stellar parameters available from Huber et al. (2014); (4) No weak secondary data available; (5) MCMC trapezoid fit did not converge; (6) Period too short for implied star (orbit within star); (7) Other unspecified vespa error.

[l] Kepler number assigned, if validated.

(This table is available in its entirety in machine-readable form.)

4. Eight KOIs have no `koi_depth_err1` value on the archive, and thus had no weak secondary constraint and were left out of the calculations.

5. For 38 KOIs, the trapezoid MCMC fit did not converge. The convergence criterion was for the autocorrelation time of the chain for each parameter to be shorter than 10% of the total chain length.

6. For 39 KOIs, the orbital period and stellar properties of the candidate imply the orbit to be within its host star's Roche limit. This usually happens when the host star is estimated to be a giant, and these situations are nearly always false positives.

The numbers in this list correspond to the numbers in the "failure" column in Table 6.

5. COMPARISON WITH FOLLOW-UP OBSERVATIONAL STUDIES

One of the difficulties with probabilistic validation is that, by necessity, it is typically invoked when no other method of follow-up confirmation is possible. It can therefore be difficult to find ways to compare the results of a calculation such as vespa to any known observational ground-truth, or to "validate the validations." However, because so much follow-up observational effort has gone into *Kepler* candidates over the last few years, there actually are two different data sets that do provide such information for relatively small subsets of candidates. Here, we discuss the results of vespa calculations for these candidates and the implications for the reliability of the vespa framework.

5.1. Spitzer Photometry

The first of these data sets is from Désert et al. (2015), who observe the transits of 50 *Kepler* candidates with the *Spitzer Space Telescope* in order to observationally constrain the FP rate. The idea behind these observations is that a blended EB false positive will often show a color-dependent transit depth, and so comparing the candidate depths measured by *Kepler* to those measured in the infrared by *Spitzer* would give an idea of how likely a signal is to be a false positive. The results of their analysis suggest that fewer than 8% of their observed candidates are likely to be false positives. Of these 50 candidates, vespa on its own calculates FPP > 0.1 for four. Two of these (KOI-103.01 and KOI-248.02) are systems with known significant TTVs; the other two (KOI-247.01 and KOI-555.02) are likely false positives, with FPPs of 0.90 and 0.81, respectively. For all but two of the remaining candidates, vespa gives FPP < 0.01. The false positive rates calculated in Désert et al. (2015) are also based on probabilistic arguments very similar to Morton & Johnson (2011) and thus are not quite candidates for "ground-truth" comparison, but the lack of transit chromaticity and the agreements between these two independent studies are certainly supportive of the vespa results.

5.2. RV Monitoring of Large Candidates

A much more independent and powerful test data set has recently become available in the work of Santerne et al. (2016), which presented the results of a long-term RV-monitoring campaign targeting 129 *Kepler* giant-planet candidates. Of these, they confirm 45 to be planets and identify 48 as eclipsing

binaries, 15 as contaminating EBs (CEBs), and 3 as brown dwarfs. They are unable to determine the nature of the remaining 18. These results imply a relatively high FP rate among giant-planet candidates, possibly near 50%, which sounds potentially concerning, although we reiterate that KOIs are not ruled FALSE POSITIVES based on their transit depth or inferred size alone. Additionally, when we look at the vespa results on this sample, we see very good agreement between our results and the RV- detected "ground truth" (Table 4). Notably, confirmed planets show a mean FPP of about 10% and a median of much less than 1%, while confirmed EBs (CEBs) show a mean FPP of 75% (78%) and a median of 97% (99%). The three brown dwarfs also show low FPPs, which is understandable because vespa doesn't pretend to predict anything about the mass of the companion, and BDs are essentially the same size as giant planets. It is also instructive to investigate the four cases where vespa computes high FPPs (>0.5) for confirmed planets. Three of these (KOI-377.01, KOI-1426.02, and KOI-1474.01) have significant TTVs, and one is a grazing eclipse (KOI-614.01). KOI-1474.01 is also on a highly eccentric orbit (Dawson et al. 2012), in addition to its TTVs, which also contributes to its high FPP. Overall, this comparison powerfully demonstrates the reliability of the vespa calculation, showing that even in a population of candidates that include many false positives, it is able to effectively identify which are true planets and which are not.

5.3. The Enigmatic Case of KOI-6705.01

We also briefly discuss another case of individual interest. KOI-6705.01, a 0.99 day signal around a mid M-dwarf star, was identified as a KOI of possible interest by Gaidos et al. (2016). After significant follow-up observations and detailed full-frame image analysis, they concluded that the signal was most likely due to a charge-transfer effect from a 1.99 day EB located on the same CCD column. For this KOI, vespa calculates an FPP of 1, with the "long" model preferred by far (of the astrophysical models, the double-period BEB—the true scenario—is preferred). This is an excellent example of what was discussed in Section 4.3: that even effects like column anomalies that were not identified by the *Kepler* team as ephemeris matches will typically be identified as false positives by vespa.

6. CONCLUSIONS

In this work, we have calculated the astrophysical false positive probability (FPP) for every *Kepler* object of interest (KOI) in the Q1–Q17 DR24 table, using the publicly available Python module vespa, which implements the procedure introduced in Morton (2012), with improvements in stellar parameter modeling and the inclusion of new "double-period" false positive scenarios and artificial models to identify KOIs that are not explained well by any of the astrophysical models. We have also for the first time estimated uncertainties in the vespa calculation, through a bootstrap resampling procedure.

While the assumptions behind this calculation are not necessarily valid for every KOI (see Section 4.2), we have identified 1284 KOIs that have reliable FPPs of <0.01, resulting in validation of their planetary nature at the 99% confidence level, more than doubling the number of confirmed *Kepler* exoplanets. Among this set of newly validated planets

are nine that are consistent with being in the habitable zones of their host stars. We also identify 428 new likely false positive KOIs, although we note that some of these may be due to unidentified or miscorrected transit timing variations.

The reliability of these calculations depends significantly on the results of Bryson & Morton (2015), which quantify the probability that the eclipse/transit signal is spatially coincident on the sky with the presumed target star. Without confirmation that the transit signal is not coming from a significantly displaced source, the sky area used as part of the prior for the false positive scenarios would need to be significantly larger than the positional uncertainty value assumed in this work (described in Section 2.1). Additionally, the blended false positive scenarios that vespa considers are assumed to be fully contained within the photometric aperture; this assumption would also be broken if the source of the eclipse were significantly displaced. We estimate that perhaps 0.06% of the planets we validate could be signals coming from significantly displaced sources, similar to those identified as period-epoch matches by Coughlin et al. (2014) but unidentified by that analysis.

While previous a priori false positive rate estimates (Morton & Johnson 2011; Fressin et al. 2013) have made clear that the *Kepler* planet candidate catalogs are generally low enough to ignore for the purposes of planet occurrence rate calculations, any more detailed study of any small subset of individual KOIs should understand in more detail the FPPs of those specific candidates. This type of small-sample candidate culling using individually calculated FPPs has already been done in the literature (Morton & Swift 2014; Morton & Winn 2014); the publication of this full catalog allows the community to do the same. In particular, several other studies have shown that specific samples of KOIs tend to have larger FPPs than the global average, so studies involving giant-planet KOIs (Santerne et al. 2016) or evolved stars (Sliski & Kipping 2014) are in even greater need of the individual FPP analysis presented here.

The *Kepler* mission has demonstrated that space-based transiting planet surveys identify planet candidates at a rate much faster than traditional follow-up techniques can confirm them. As a result, false positive probability quantification techniques are now an integral part of the landscape of exoplanet science. While the present work is the first large-scale demonstration of a fully automated validation procedure, there is much progress still to be made. For example, there is currently no support within vespa to calculate the FPP for a candidate which has a *specifically identified* but previously unknown close companion. Future development plans for the vespa package include support for this scenario, as well as other improvements. One of the most important of these will be to allow for a contaminating EB to not be fully contained within the target photometric aperture; that is, modeling the probability for further-away stars to be EBs contributing only a small amount of their flux to the target photometry, such as is the case for many of the false positives identified via ephemeris matching by Coughlin et al. (2014). Full inclusion of this effect will allow for even low-S/N candidates to receive confident vespa analysis, which is now limited only to KOIs for which confident pixel-level positional analysis is possible.

Beyond *Kepler*, future transit missions such as TESS and PLATO will require automated false positive analysis in order to efficiently sift through the large numbers of candidates that

they will find. This work demonstrates that vespa will be a valuable tool toward this purpose. Instructions for how to reproduce the calculations upon which this work is based may be found at http://github.com/timothydmorton/koi-fpp. Summary plots for all the test results presented in this paper may be found at http://kepler-fpp.space.

REFERENCES

Bastien, F. A., Stassun, K. G., & Pepper, J. 2014, ApJL, 788, L9
Becker, J. C., Vanderburg, A., Adams, F. C., Rappaport, S. A., & Schwengeler, H. M. 2015, ApJL, 812, L18
Borucki, W. J., Koch, D. G., Batalha, N., et al. 2012, ApJ, 745, 120
Bryson, S. T., Jenkins, J. M., Gilliland, R. L., et al. 2013, PASP, 125, 889
Bryson, S. T., & Morton, T. D. 2015, KSCI-19092-001
Buchner, J., Georgakakis, A., Nandra, K., et al. 2014, A&A, 564, A125
Burke, C. J., Christiansen, J. L., Mullally, F., et al. 2015, ApJ, 809, 8
Casagrande, L., Schönrich, R., Asplund, M., et al. 2011, A&A, 530, A138
Coughlin, J. L., Bryson, S. T., Burke, C. J., et al. 2015a, KSCI-19104-001
Coughlin, J. L., Mullally, F., Thompson, S. E., et al. 2015b, ApJS, in press (arXiv:1512.06149)
Coughlin, J. L., Thompson, S. E., Bryson, S. T., et al. 2014, AJ, 147, 119
Dawson, R. I., Johnson, J. A., Morton, T. D., et al. 2012, ApJ, 761, 163
Delfosse, X., Forveille, T., Ségransan, D., et al. 2000, A&A, 364, 217
Désert, J.-M., Charbonneau, D., Torres, G., et al. 2015, ApJ, 804, 59
Díaz, R. F., Almenara, J. M., Santerne, A., et al. 2014, MNRAS, 441, 983
Dotter, A., Chaboyer, B., Jevremović, D., et al. 2008, ApJS, 178, 89
Dressing, C. D., & Charbonneau, D. 2013, ApJ, 767, 95
Fabrycky, D. C., Ford, E. B., Steffen, J. H., et al. 2012, ApJ, 750, 114
Feiden, G. A., Chaboyer, B., & Dotter, A. 2011, ApJL, 740, L25
Feroz, F., Hobson, M. P., & Bridges, M. 2009, MNRAS, 398, 1601
Feroz, F., Hobson, M. P., & Bridges, M. 2011, MultiNest: Efficient and Robust Bayesian Inference, Astrophysics Source Code Library, ascl:1109.006
Feroz, F., Hobson, M. P., Cameron, E., & Pettitt, A. N. 2013, arXiv:1306.2144
Ford, E. B., Fabrycky, D. C., Steffen, J. H., et al. 2012, ApJ, 750, 113
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Foreman-Mackey, D., Hogg, D. W., & Morton, T. D. 2014, ApJ, 795, 64
Fressin, F., Torres, G., Charbonneau, D., et al. 2013, ApJ, 766, 81
Gaidos, E., Mann, A. W., & Ansdell, M. 2016, ApJ, 817, 50
Hayden, M. R., Bovy, J., Holtzman, J. A., et al. 2015, ApJ, 808, 132
Huber, D., Silva Aguirre, V., Matthews, J. M., et al. 2014, ApJS, 211, 2
Jontof-Hutter, D., Ford, E. B., Rowe, J. F., et al. 2016, ApJ, 820, 39
Kipping, D. M., Torres, G., Buchhave, L. A., et al. 2014, ApJ, 795, 25
Kopparapu, R. K., Ramirez, R., Kasting, J. F., et al. 2013, ApJ, 765, 131
Lissauer, J. J., Marcy, G. W., Bryson, S. T., et al. 2014, ApJ, 784, 44
Lissauer, J. J., Marcy, G. W., Rowe, J. F., et al. 2012, ApJ, 750, 112
Marcy, G. W., Isaacson, H., Howard, A. W., et al. 2014, ApJS, 210, 20
Montet, B. T., Morton, T. D., Foreman-Mackey, D., et al. 2015, ApJ, 809, 25

Morton, T. 2014, PhD thesis, California Institute of Technology
Morton, T. D. 2012, ApJ, 761, 6
Morton, T. D. 2015a, Isochrones: Stellar model grid package, Astrophysics Source Code Library, ascl:1503.010
Morton, T. D. 2015b, VESPA: False positive probabilities calculator, Astrophysics Source Code Library, ascl:1503.011
Morton, T. D., & Johnson, J. A. 2011, ApJ, 738, 170
Morton, T. D., & Swift, J. 2014, ApJ, 791, 10
Morton, T. D., & Winn, J. N. 2014, ApJ, 796, 47
Moutou, C., Almenara, J. M., Díaz, R. F., et al. 2014, MNRAS, 444, 2783
Muirhead, P. S., Johnson, J. A., Apps, K., et al. 2012, ApJ, 747, 144
Petigura, E. A., Howard, A. W., & Marcy, G. W. 2013, PNAS, 110, 19273

Pinsonneault, M. H., An, D., Molenda-Żakowicz, J., et al. 2012, ApJS, 199, 30
Rowe, J. F., Bryson, S. T., Marcy, G. W., et al. 2014, ApJ, 784, 45
Rowe, J. F., Coughlin, J. L., Antoci, V., et al. 2015, ApJS, 217, 16
Santerne, A., Díaz, R. F., Moutou, C., et al. 2012, A&A, 545, A76
Santerne, A., Hébrard, G., Deleuil, M., et al. 2014, A&A, 571, A37
Santerne, A., Moutou, C., Tsantaki, M., et al. 2016, A&A, 587, 64
Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525
Sliski, D. H., & Kipping, D. M. 2014, ApJ, 788, 148
Steffen, J. H., Fabrycky, D. C., Agol, E., et al. 2013, MNRAS, 428, 1077
Steffen, J. H., Fabrycky, D. C., Ford, E. B., et al. 2012, MNRAS, 421, 2342
Swift, J. J., Johnson, J. A., Morton, T. D., et al. 2013, ApJ, 764, 105
Torres, G., Kipping, D. M., Fressin, F., et al. 2015, ApJ, 800, 99