

# Resolving period aliases for TESS monotransits recovered during the extended mission

Benjamin F. Cooke<sup>1,2\*</sup>, Don Pollacco<sup>1,2</sup>, David R. Anderson<sup>1,2</sup>, Daniel Bayliss<sup>1,2</sup>, François Bouchy<sup>3</sup>, Samuel Gill<sup>1,2</sup>, Nolan Grieves<sup>3</sup>, Monika Lendl<sup>1,3</sup>, Louise D. Nielsen<sup>1,3</sup>, Stéphane Udry<sup>3</sup> and Peter J. Wheatley<sup>1,2</sup>

<sup>1</sup>Department of Physics, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK

<sup>2</sup>Centre for Exoplanets and Habitability, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK

<sup>3</sup>Observatoire de Genève, University of Geneva, 51 ch des Maillettes, CH-1290 Veroy, Switzerland

Accepted 2020 November 9. Received 2020 November 2; in original form 2020 September 14

## ABSTRACT

We set out to explore how best to mitigate the number of period aliases for a transiting Transiting Exoplanet Survey Satellite (TESS) system with two identified transits separated by a large time period on the order of years. We simulate a realistic population of doubly transiting planets based on the observing strategy of the TESS primary and extended missions. We next simulate additional observations using photometry (NGTS) and spectroscopy (HARPS and CORALIE) and assess its impact on the period aliases of systems with two TESS transits. We find that TESS will detect around 400 exoplanets that exhibit one transit in each of the primary and extended missions. Based on the temporal coverage, each of these systems will have an average of 38 period aliases. We find that, assuming a combination of NGTS and CORALIE over observing campaigns spanning 50 d, we can find the true alias, and thus solve the period, for up to 207 of these systems with even more being solved if the observing campaigns are extended or we upgrade to HARPS over CORALIE.

**Key words:** surveys – planets and satellites: detection – planetary systems.

## 1 INTRODUCTION

The Transiting Exoplanet Survey Satellite (TESS; Ricker et al. 2015) has recently completed its observations of the northern ecliptic hemisphere during the second half of its primary mission. As of 5 July 2020 TESS has begun its extended mission in which it will first carry out a re-observation of the southern ecliptic hemisphere, approximately 2 yr after first observing it (for extended mission details see NASA 2019 senior review<sup>1</sup> and NASA response<sup>2</sup>).

TESS has already produced numerous discoveries of planets around bright stars (e.g. Gandolfi et al. 2018; Huang et al. 2018; Vanderspek et al. 2019; Gilbert et al. 2020) as well as discoveries of systems for which we observe only a single transit (e.g. Gill et al. 2019, 2020b; Lendl et al. 2020), including the first monotransiting planet from TESS (Gill et al. 2020a). The detection of monotransits is a natural consequence of the limited temporal coverage of TESS when compared to other surveys such as the Wide Angle Search for Planets (WASP; Pollacco et al. 2006) or Kepler/K2 (Borucki et al. 2010; Howell et al. 2014). These monotransits have been explored via simulations (Cooke et al. 2018; Villanueva, Dragomir & Gaudi 2019), which show we can expect a large number of monotransit candidates in the TESS primary mission data. These predictions have been verified by recent monotransit searches including Montalto et al.

(2020) who find 15 Southern hemisphere candidates and the NGTS monotransit team (see Gill et al. 2020a, 2020b for a description of the pipeline used) who have identified over 50 candidates from the first few TESS sectors alone. Many of these candidates exhibit additional extended mission transit events as discussed later in this paper. Because of the close dependence of monotransit detections on survey coverage the determination of orbital periods for single-transit candidates will be heavily influenced by the continuation of TESS into its extended mission (Cooke, Pollacco & Bayliss 2019).

With the data from the first year of the extended TESS mission (i.e. Year 3 of TESS), most southern ecliptic TESS targets will be observed for about twice as long. However, these additional observations will be carried out approximately 2 yr after the initial observations. Due to this, there will be significant gaps in the photometric baseline that may hide additional transits. This effect is most pronounced for those systems that only transited once during the primary mission and will only transit once more in the first year of the extended mission. Based on only two transits we cannot infer the true orbital period, we know only that it is bounded. The maximum period is given by the separation of the two observed transits which is on the order of  $\sim 2$  years. The minimum period is given by the constraint that only one transit is seen in each TESS observing run, approximately 10 d for a single sector of observations. Therefore, we are left with a discrete set of period aliases (Cooke et al. 2019).

Period aliases are an important consideration when analysing non-continuous observations. This is a common problem for ground-based observations due to the daytime and weather interruptions in the data; any signal short enough to fit within a day can be missed and

\* E-mail: B.Cooke@warwick.ac.uk

<sup>1</sup>2019 Senior Review Subcommittee Report.

<sup>2</sup>NASA Response to the 2019 Senior Review of Operating Missions.

it requires sufficient night time observations to rule out these potential periods. However, for longer period systems, this is a significant issue for space-based observations as well (Dawson & Fabrycky 2010; Becker et al. 2019). Some attempts have been made to use period priors to prioritize aliases to search (Dholakia et al. 2020), but this is far from foolproof. To actually rule out aliases requires additional data in the form of photometry and/or spectroscopy. This paper looks at exactly what effect additional observations have on the number of period aliases of a given system and how the effect changes depending on the amount and quality of these observations. We explore both photometry and spectroscopy by simulating each method, and investigating how this reduces the set of period aliases for TESS discoveries. For a recent comparison of photometric and spectroscopic follow-up methods, see Cooke & Pollacco (2020).

This paper is laid out in the following manner. Section 2 discusses the stellar and planetary population used as well as the simulated TESS, photometric and spectroscopic data. Section 3 details our attempts to use the data to identify and analyse the period aliases. Section 4 presents the results of our simulation and analysis, while Section 5 outlines our conclusions.

## 2 OBSERVATIONS

### 2.1 Simulation population

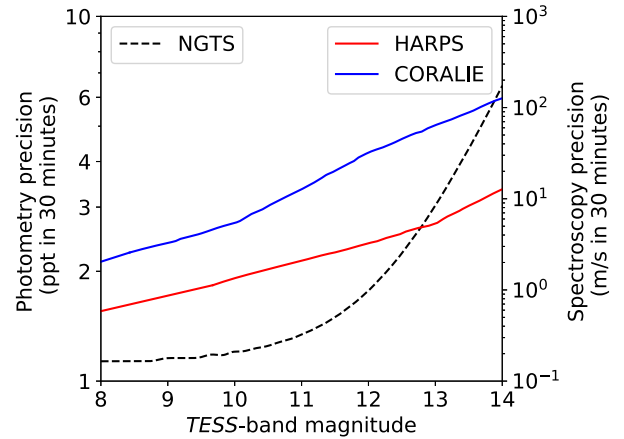
To predict the number of planets for which TESS will observe two individual transits, we first create a planetary population. TESS observations of this population are then simulated. To do this, we proceed as in Cooke et al. (2018) that we summarize below.

We use the TESS Input Catalogue Candidate Target List version 8 (Stassun et al. 2019) accessible via the Mikulski Archive for Space Telescopes (MAST<sup>3</sup>) as our stellar population, which includes *Gaia* DR2 parameters (Gaia Collaboration et al. 2018). We perform the following simple cuts on this population based on TESS-band magnitude ( $3.0 \leq m_T \leq 17.0$ ) and effective stellar temperature ( $2285 \leq T_{\text{eff}} \leq 10050 \text{ K}$ ) leaving 4789372 targets in the southern ecliptic hemisphere.

We generate planets around these stars by drawing from period-radius bins under particular occurrence rates given by Dressing & Charbonneau (2015) for M-stars and Hsu et al. (2019) for FGK stars. Note that this is an updated set of occurrence rates compared to the procedure laid out in Cooke et al. (2018). In particular, Hsu et al. (2019) uses the final Kepler Data Release (DR25; Thompson et al. 2016) and the improved stellar parameters made available via *Gaia* DR2 (Gaia Collaboration et al. 2018). More specific transit parameters are then calculated, using equations and distributions presented in Winn (2010) and Barclay, Pepper & Quintana (2018). We include a uniform distribution in  $\cos i$  for inclination used to calculate transit probability, but assume all planets around the same star share an inclination value.

### 2.2 TESS observations

To identify which systems would be observed to transit once in both the primary and extended TESS missions, we simulate TESS observations. Since it will be the first hemisphere to be re-observed in the extended mission, and due to our use of southern ecliptic hemisphere based follow-up observatories, we only simulate southern ecliptic hemisphere TESS targets. Our methods follow that



**Figure 1.** Instrumental precision as a function of target TESS magnitude for NGTS, HARPS, and CORALIE (adapted from fig. 14 of Wheatley et al. 2018 and fig. 10 of Günther et al. 2017).

described in Cooke et al. (2019). We simulate observations only at times for which there are data points not affected by systematics from the TESS primary mission. This has the effect of producing a more realistic coverage fraction per sector and accounts for systematic effects such as momentum dumps. We assume that the primary mission coverage can be replicated during the extended mission, noting that in the extended mission full-frame images will be taken with an improved cadence of 10 min.

Based on our simulated TESS coverage, we can now determine which planet transits are detectable with TESS. Using our planet parameters and a fifth-order polynomial TESS noise approximation (Stassun et al. 2018), we calculate a signal-to-noise ratio ( $S/N$ ) for each planet. Those planets which exhibit a transit during a region of TESS observations that manage to achieve the required  $S/N \geq 7.3$  are recorded as detectable. For more specific details regarding the simulated TESS observations, see Cooke et al. (2019).

Once the simulation is complete, we select all those planets with one detectable transit during the TESS primary mission and one detectable transit during the first year of the extended mission as our double transit sample.

### 2.3 Simulated photometry

We simulate additional photometric data, using the Next Generation Transit Survey (NGTS; Wheatley et al. 2018). For this additional photometry, we follow the procedure being employed by NGTS (see Gill et al. 2019; Lendl et al. 2020). Observations are simulated from sunset to sunrise using a cadence of 12 s. We use the NGTS noise model as a function of host magnitude shown in Fig. 1 and compare this to the signal size of the planet,  $(R_p/R_*)^2$ . Additionally, we allow the use of multiple NGTS telescopes if required to reach the necessary  $S/N$  level. To calculate the  $S/N$  level for multiple telescope, we use a  $1/\sqrt{N}$  relation as has been shown to be a good approximation by Bryant et al. (2020). Where NGTS can achieve an  $S/N \geq 3.0$ , we use the simulated photometry. Below this  $S/N$  threshold we assume that photometry is of insufficient quality to help constrain the period of the system and do not simulate follow-up photometry.

To best utilize photometric time, we only simulate observations on nights when we predict a transit may occur (see Section 4.5 for details including accounting for transit time uncertainties). Since this paper explores those systems with transits in both the primary and extended TESS missions, we can use the TESS transit times and

<sup>3</sup>[http://archive.stsci.edu/tess/tic\\_ctl.html](http://archive.stsci.edu/tess/tic_ctl.html)

coverage to create a set of discrete period aliases and we therefore only simulate NGTS observations on nights that at least one alias predicts will contain a transit. This is a realistic strategy that avoids wasted time that would occur by observing on nights that provide no period constraining information.

We can then run this strategy for a fixed number of nights, observing some subset of those. In the main body of this paper, we consider an overall time-span of 50 nights. The Appendix then includes the results for a range of different maximum time-spans (A1). As an additional measure, we include telescope downtime from instrumental and weather effects. This is included by assuming that a random 20 per cent of nights are unavailable for observations (for more details, see Cooke et al. 2020).

## 2.4 Simulated spectroscopy

Spectroscopic data are simulated using the High-Accuracy Radial-velocity Planetary Search (HARPS; Mayor et al. 2003) on the ESO 3.6 m telescope and CORALIE (Queloz et al. 2000) on the Euler 1.2 m telescope. We simulate an observing strategy of one 30-min exposure every 7 d for each candidate, including a noise value based on the noise models of the instruments as a function of host magnitude shown in Fig. 1. The amplitude of the radial velocity signal is a function of the true period (which is known), the stellar mass (which is known), and the planetary mass (which is unknown). Therefore, we first predict the planetary mass using the planetary radius and the MRExo mass–radius relation<sup>4</sup> from Kanodia et al. (2019). For each radius value, we draw a random value from a mass distribution generated using MRExo.

As with photometry, this procedure can be run for any time-span, the difference being that now the number of data points is one seventh of the number of nights. In this paper, we again consider an overall time-span of 50 nights using CORALIE with alternative results (including using HARPS) presented in the Appendix. As with the simulated photometry in Section 2.3, we account for telescope downtime from instrumental and weather effects as well as scheduling conflicts. This is included by assuming that a random 20 per cent of nights are unavailable for observations (for more details, see Cooke et al. 2020).

## 3 ALIAS DETERMINATION

### 3.1 Period aliases from TESS

Based on our simulated TESS observations in Years 1 and 3 (primary and extended mission), we select the subset of planets that exhibit one transit during each year. The separation of these transit times gives an upper limit to the period of the planet as well as the knowledge that the true period must be equal to this separation divided by an integer. Assuming that the minimum period that a planet could have and still only be observed once in a sector is  $\sim 10$  d we can make a list of all possible period aliases. Some of these periods however can be ruled out by the TESS coverage we already have. We test each period alias by comparing with the TESS coverage. For each alias, we create a list of all times at which the system would transit (including times before the Year 1 transit and after the Year 3 transit) and check if there are any TESS observations within half a transit duration of any of these. If there are TESS observations that overlap with any transit time (excluding the two discovery transits) we can rule out that alias

as one that would result in an additional transit which has not been seen. In this way, we reduce the list of period aliases down to only those allowed based on the TESS photometry.

### 3.2 Period aliases using additional data

With the newly simulated photometry and spectroscopy, we can now attempt to better constrain the true period of each monotransit and rule out more aliases. We do this first using the TESS and follow-up photometric data, then the TESS and spectroscopic data, then finally using the TESS, photometric, and spectroscopic data.

To combine the NGTS data, we proceed on a nightly basis. For each night, we check if any of the possible period aliases predict a transit to occur. If so, we simulate observations. If no transit is seen it shows that the true period is not one of those that predicted a transit this night and we therefore remove all periods that predicted a transit that night from our list of aliases. If a transit is seen we can use this third transit to remove any aliases that do not predict this additional detection. This process repeats each night until the period is solved or we reach our observing limit.

Ruling out aliases using spectroscopic data requires a different approach. First, we require a minimum of five radial velocity measurements before we begin to rule out aliases. Once we have reached this threshold every time, we simulate a new spectroscopic data point we fold the radial velocity data on each alias and plot in phase space. For each allowed period, we model the radial velocity variation assuming a circular orbit and fitting for the amplitude, phase, and gamma velocity. We then calculate the residuals between our simulated measurements and this fit. If the average of the absolute residuals is greater than 1.5 times the noise threshold at the chosen magnitude we reject the alias. Otherwise, the alias is still a valid option for the system. Since we require at least 5 data points, the average of the residuals is unlikely to reach this threshold for the true period. In multiple runs, the true period was never rejected. This process is repeated until the period is solved or we reach our observing limit.

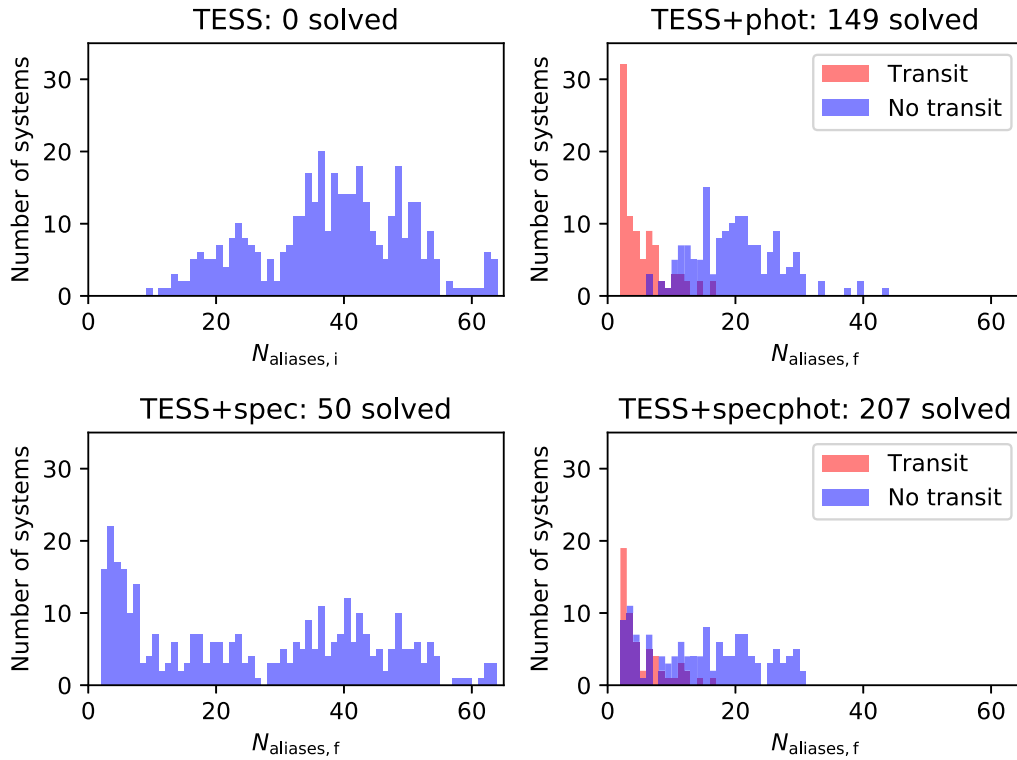
To find the viable aliases using a combination of TESS photometry, photometric follow-up, and spectroscopic follow-up, we simply examine the alias sets from TESS and photometry and from TESS and spectroscopy and allow only those aliases present in both sets. This is carried out on a nightly basis to determine the earliest point at which a system is solved.

## 4 RESULTS AND DISCUSSION

### 4.1 TESS extended mission monotransits

Based on our simulated observations, we expect 759 monotransits will be found in the TESS southern ecliptic primary mission. Of these, approximately 65 per cent will transit again in the first year of the extended mission with most only transiting once more. The result is that slightly over 50 per cent of the primary mission southern monotransits will transit once more in the first year of the extended mission, giving a total of 395 planets that will exhibit one detectable transit in both the primary and extended TESS missions for the southern ecliptic hemisphere. This number exceeds those presented in Cooke et al. (2018, 2019) by a factor of 2–3. This is due to the updated occurrence rate used in this study. The occurrence rate used here comes from Hsu et al. (2019) that uses a more up to date and larger sample based on the Kepler mission. Therefore, these results will be more robust than similar studies based on outdated occurrence rates.

<sup>4</sup><https://github.com/shbhuk/mrexo>



**Figure 2.** Distributions of the number of period aliases from each combination of observational data. The four plots show  $N_{\text{aliases},i}$  and the three  $N_{\text{aliases},f}$  values for the four combinations of observations. For the right hand column plots (those including additional photometry), the data is split into those for which an additional transit is seen (red) and those for which no additional transit is seen (blue). Each plot tile denotes the data used and how many systems have been solved with this data (these systems are not shown in the plots to avoid overwhelming the histograms). Here, we assume a time-span of 50 d and use CORALIE for spectroscopy.

#### 4.2 Number of aliases, $N_{\text{aliases}}$

For each southern ecliptic hemisphere planet that TESS will observe to transit once in each year we have produced a list of period aliases. From this list, we determine the number of aliases,  $N_{\text{aliases}}$ , for each system based on different subsets of observational data. For each system, we then calculate  $N_{\text{aliases}}$  using four different combinations of observational data; TESS data only, TESS and additional photometry, TESS and additional spectroscopy and finally TESS, and additional photometry and spectroscopy. We define the number of aliases found using TESS data alone as  $N_{\text{aliases},i}$  and the updated value found by taking into account different combinations of follow-up data as  $N_{\text{aliases},f}$ .

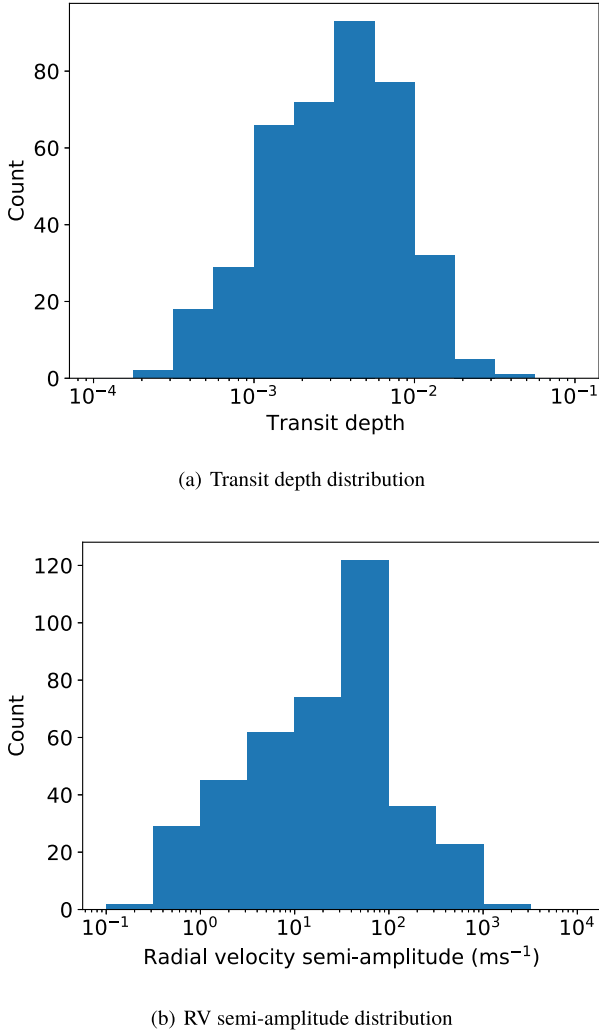
Fig. 2 shows the distribution of the number of aliases from TESS,  $N_{\text{aliases},i}$ , and the number of aliases from the three combinations of follow-up data,  $N_{\text{aliases},f}$ .

For each plot in Fig. 2, we show the value of  $N_{\text{aliases},i}$  or  $N_{\text{aliases},f}$  for all unsolved systems when accounting for the four regimes of observational data. The solved systems (those for which the period is known, i.e.  $N_{\text{aliases},f} = 1$ ) are not shown in the plots, as they show large spikes at 1 overwhelming the rest of the distributions. The number of solved systems, and therefore the number of systems not shown, is in the title of each plot. The top left plot shows the distribution of the number of aliases per system using TESS data alone,  $N_{\text{aliases},i}$ . This gives an average value of 38 aliases per system; however, it should be noted that there is a double peak in the distribution. This corresponds to those systems that are observed in multiple TESS sectors. Systems with more TESS coverage have correspondingly fewer aliases leading to a secondary peak around 20. Systems

with a single TESS sector per year have more aliases and peak around 40.

The top right and bottom plots show the distributions of  $N_{\text{aliases},f}$  based on three combinations of follow-up data. As can be seen photometry solves significantly more systems than spectroscopy (149 and 50 systems solved, respectively) but combining the two methods leads to a number of solved systems greater than simply adding the uniquely solved systems together (207 solved systems). Figures including additional photometry, those in the right-hand column, are split depending on whether the photometry caught additional (one or more) transits. It is seen that catching a transit leads to a large reduction in  $N_{\text{aliases},f}$  (see Gill et al. 2020a for an example of this in practice) and all of the systems solved from additional photometry alone (149) had at least one additional transit detected. Included in the plots are those systems for which the photometry is of insufficient S/N. Transits may occur during observations of these systems but would not be detected with sufficient confidence. For these systems additional photometry cannot help to constrain the period. Additionally, some of the systems result in spectroscopic signals smaller than the respective noise level. These points are present in the plot but additional spectroscopy would not be able to rule out any periods. Fig. 3 shows the distributions of both transit depth and radial velocity semi-amplitude for the monotransits considered. For both photometry and spectroscopy, we require  $S/N \geq 3$  to simulate observations.

As well as the number of solved systems proved by each combination of observational data these distributions show the number of aliases for those unsolved systems. For TESS photometry only, all 395 systems are unsolved and the average (mean) value of  $N_{\text{aliases},i}$  is

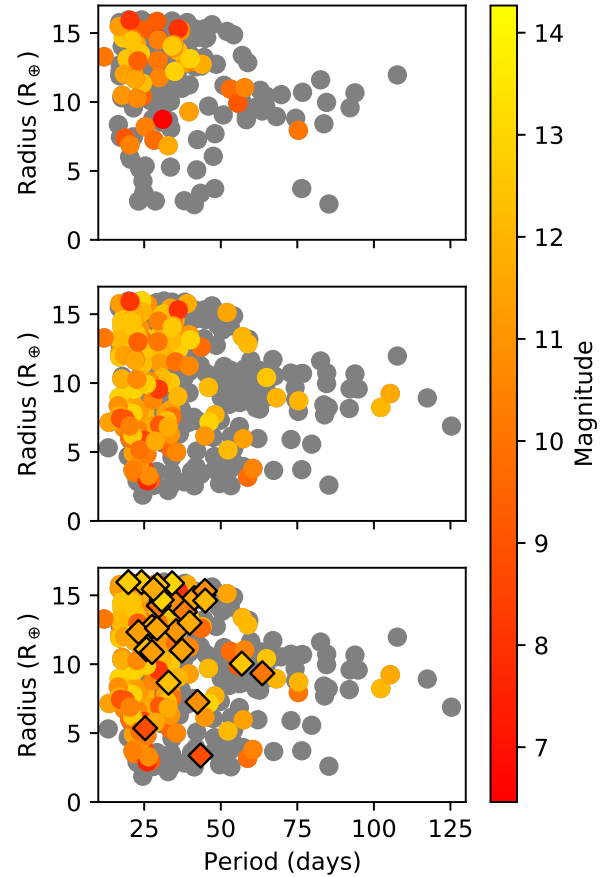


**Figure 3.** Distribution of signal size for photometry and spectroscopy for all considered monotransits.

$\sim 38$ . When photometry is included we have 246 unsolved systems with an average  $N_{\text{aliases}, f}$  value of  $\sim 14$ . Using spectroscopy instead gives 345 unsolved systems with an average  $N_{\text{aliases}, f}$  value of  $\sim 26$ . Finally, including both photometric and spectroscopic data on top of the TESS photometry gives 188 unsolved systems with an average  $N_{\text{aliases}, f}$  value of only  $\sim 12$ .

From Fig. 2, we see that the number of systems solved when combining the follow-up photometry and spectroscopy is larger than the sum of the systems solved using each method individually. Further analysing this, we find that of the systems solved by spectroscopy, 19, or 38 per cent, are also solved using photometry. This absolute value is obviously then the same for the photometrically solved systems also solved by spectroscopy but in this case it is only 13 per cent of the total.

Taking this analysis forwards to the systems solved when combining the two methods we find that 27 of these (13 per cent) are not solved by either photometry or spectroscopy individually. These 27 systems are therefore systems for which each method manages to rule out different subsets of aliases but is incapable of completely solving the period. Taking the common aliases left from both follow-up methods results in these systems being solved. It should also be noted that photometric and spectroscopic data provide complementary information about the radius and mass of

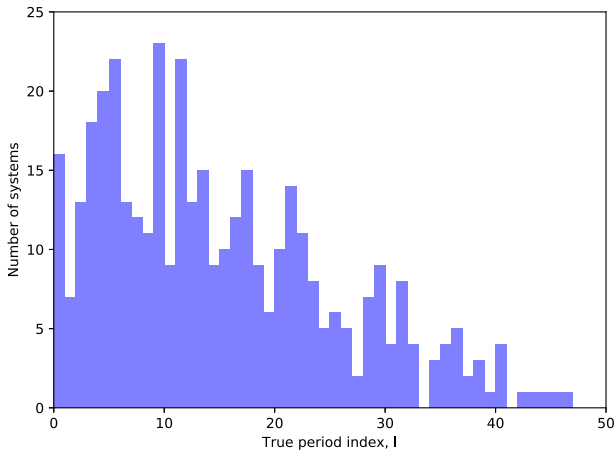


**Figure 4.** Period and radius distribution of solved systems. The top plot shows the distribution for spectroscopically solved systems, the middle is that for photometrically solved systems and the bottom shows the combination of both methods. Diamond points denote systems only solved when both methods are combined. The colour of each point denotes its TESS-band magnitude.

the planet, another reason that a combination of follow-up data is highly desirable.

Fig. 4 shows the distribution of solved systems with orbital period and planetary radius. The top panel shows the results for spectroscopy with systems solved being coloured by magnitude and unsolved systems being grey. Systems with insufficient S/N are omitted from the plot. The middle panel shows the corresponding plot for photometry and the bottom panel shows the combination of the two methods. In this panel, all systems are plotted regardless of either photometric or spectroscopic S/N and systems that are only solved when the methods are combined are shown as diamonds.

From this plot, we see that the majority of systems solved by spectroscopy have radii between 10 and  $15R_{\oplus}$ . This is as expected, since predictions and observations suggest this radii range corresponds to the most massive planets that are the ones that spectroscopy is most sensitive to. The photometric planets are more spread in radius since they tend to be insensitive to radius (and mass) as long as the transit depth is above the sensitivity threshold. Both methods favour shorter periods since these are obviously easier to confirm within a given observing programme. It also seems that the subset of planets solved only when both data sets are combined (diamonds in Fig. 4) have a broader period distribution. Once again, this is unsurprising as the additional data help to confirm these more challenging systems.



**Figure 5.** True period index,  $I$  for all 395 TESS systems that show a single transit in each of the Years 1 and 3 data sets. A value of  $I = 0$  means the true period is the smallest allowed alias.

### 4.3 True period index, $I$

Based on the increased likelihood of detecting short period planets (transit probability scales as  $P^{-\frac{2}{3}}$ ), it might be natural to assume that, even for unsolved systems, the true period is most commonly the shortest of the period aliases. However, this is frequently untrue. Fig. 5 shows the distribution of the true period index  $I$  for all 395 systems that transit once in each of the TESS primary and extended missions.  $I$  is the index of the true period in the list of all period aliases based on the TESS data alone. A value of  $I = 0$  means that the true period is the shortest allowed alias.

Fig. 5 shows that the average value of  $I$  is  $\sim 15$  and only 16 of 395 systems have  $I = 0$ . This strongly supports the need for the additional photometric and/or spectroscopic observations discussed in this paper. Even working up from the shortest periods (i.e. lowest  $I$  values), it still requires an average of  $\sim 14$  aliases to be ruled out before the true period is found. Using additional observations to reduce  $N_{\text{aliases}, i}$  and therefore reduce  $I$  is vital. There have been some attempts at predicting the true period of a systems from a set of aliases using statistical methods (Dholakia et al. 2020), but due to the statistical nature, the reliability for individual system is poor thus an attempt to reduce  $N_{\text{aliases}, i}$  is still needed.

Another method for identifying the true period is to use the transit shape. Some attempts have been made to use a Bayesian inference based on the transit shape and an assumption of zero eccentricity to constrain orbital period to  $\sim 10$  per cent from a single K2 transit (Osborn et al. 2016). Longer period systems however are more likely to have non-zero eccentricity (Halbwachs, Mayor & Udry 2005; Pont et al. 2011); therefore, this method is not robust. Additionally, many of these systems will be discovered in the TESS 30 min cadence images (Cooke et al. 2018, 2019), where the transit shape is less well defined.

### 4.4 Improvements in number of aliases, $\Delta N_{\text{aliases}}$

As well as exploring the absolute number of aliases as a result of different amounts and methods of observations it is useful to explore how these values change. This is shown in Fig. 6 which plots  $\Delta N_{\text{aliases}}$ , where  $\Delta N_{\text{aliases}}$  is defined as

$$\Delta N_{\text{aliases}} = \frac{N_{\text{aliases}, f} - N_{\text{aliases}, i}}{N_{\text{aliases}, i} - 1}. \quad (1)$$

In this formalism, a value of  $\Delta N_{\text{aliases}} = 0$  means that  $N_{\text{aliases}, i}$  has not been improved by additional follow-up data, whereas a value of  $\Delta N_{\text{aliases}} = 1$  means all incorrect aliases have been ruled out and the period of the system has been solved. The value can be interpreted as the fraction of aliases that each method can rule out.

To better show the change in the number of allowed period aliases these plots show all 395 systems, including both solved and unsolved systems. This leads to 2 peaks in each plot at 0.0 and 1.0. The peak at 1.0 corresponds to the solved systems, i.e.  $N_{\text{aliases}, f} = 1$ . The peak at 0.0 corresponds to those systems, where  $N_{\text{aliases}, f} = N_{\text{aliases}, i}$ . For photometry, these are the systems for which NGTS cannot reach sufficient S/N meaning the data cannot be used to detect transits and for spectroscopy these systems are those for which the spectrograph noise (CORALIE in these plots) is comparable to, or greater than, the RV signal of the planets meaning the period cannot be seen under the scatter. These peaks overwhelm the rest of the data and are truncated in the plots but are described fully below.

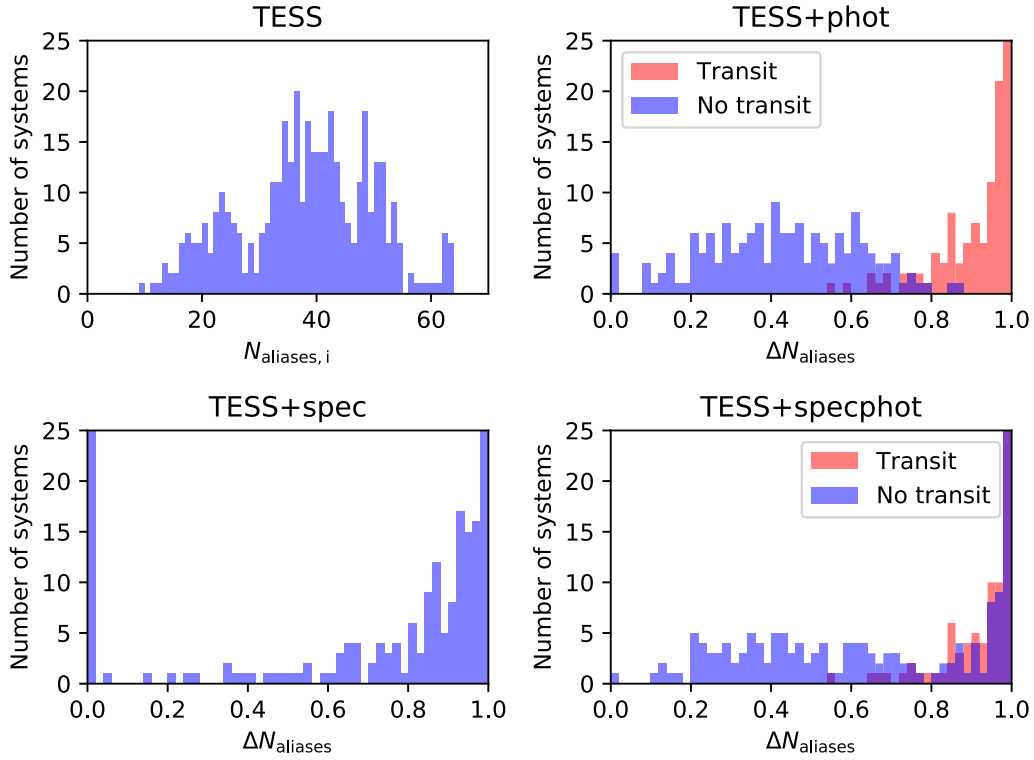
Once again we see a clear separation in the right-hand column plots based on the detection of an additional transit. Those systems seen to transit again are almost all above 0.8 (meaning over 80 per cent of the aliases are eliminated) with 149 being solved fully. For systems without an additional transit the fraction lies mainly between 0.1 and 0.7, the full distribution gives 297 systems ( $\sim 75$  per cent) that have more than half their aliases eliminated. Ignoring the unsolvable systems (insufficient S/N), this is  $\sim 76$  per cent.

Spectroscopy shows a slightly different distribution. The peak at 1.0 is much smaller than photometry (50 compared to 149), but the peak at 0.0 is larger (210 compared to 4). Exterior to the two peaks, the spectroscopy distribution is then geared towards higher fractions with a larger clump of systems between 0.9 and 1.0. In fact, we see that almost all systems with sufficient S/N have more than half of their aliases ruled out (93 per cent). This is as expected since spectroscopy data continues to add information on each subsequent night, whereas photometry data are less useful until catching an additional transit. The reason that comparatively few spectroscopy systems are able to go from only a few remaining aliases to completely solved ( $N_{\text{aliases}, f} = 1$ ) is that by this point only similar period aliases are left. This method of spectroscopic observations is efficient at ruling out significantly different periods but finds it very difficult to identify the correct period between two similar values. Photometry is less affected by this feature, as an additional transit is, in general, very good as discriminating between two similar periods.

Once again, combining the two methods gives the best result with the overall improvement in fraction of aliases eliminated by spectroscopy being further improved by systems with additional transit from photometry. When using both methods only 1 out of 395 systems are not improved and 335 systems ( $\sim 85$  per cent) have more than half of their period aliases eliminated.

### 4.5 Time taken versus nights observed

An additional aspect of this simulation that must be considered is that though our simulation is run for 50 d, no system is observed for that many nights. First, each system is only observed by each method until that method has managed to solve the system after which further observations are not simulated. Secondly, for spectroscopy, a system is only observed once every 7 d, allowing for weather. This means that the number of nights for which a system is observed spectroscopically is approximately equal to  $\frac{T}{7} \times 0.8$ , where  $T$  is either the time until the system is solved or the time-span of the simulation, whichever is smaller. For photometry, a system is only observed when one of its period aliases predicts a transit will occur. This means that the



**Figure 6.** Improvements in number of period aliases found when additional observational data is accounted for. Shown are distributions of  $\Delta N_{\text{aliases}}$  for the three combinations of additional observations. The top left plot repeats the top left plot in Fig. 2 to show the values of  $N_{\text{aliases}, i}$  by which the other three subplots are divided. A value of  $\Delta N_{\text{aliases}} = 0$  means no improvement has been made, whereas  $\Delta N_{\text{aliases}} = 1$  means the system has been solved. As in Fig. 2 the right-hand column plots (those including additional photometry) are split into those systems for which an additional transit is seen (red) and those for which no additional transit is seen (blue). Here, we assume a time-span of 50 d and use CORALIE for spectroscopy. The spikes at 0.0 and 1.0 overwhelm the rest of the data and are truncated here, the values are given in the text.

actual number of nights observed can be a small fraction of the time over which the simulation is run. Additionally, we do not necessarily observe for entire nights at a time. For spectroscopy, we use a noise model that assumes half-hour exposures so each simulated night of observing is only half an hour of spectroscopy time. For photometry, we observe each night for the full length of any transit that is predicted to occur that night i.e. if only a partial transit is predicted we only simulate observing time during that window. As such the total number of telescope hours is different from the number of nights observed and we show the values here. Fig. 7 shows plots of the number of hours that systems are observed for over the 50 d for which this simulation is run.

Additionally, for photometry we build in a buffer time based on the propagation of transit timing uncertainties from the TESS transits (as a bonus this buffer time will allow us to obtain out of transit baseline photometry, helping to improve transit characterization). For each period alias the uncertainty is equal to  $2\sigma_t/N$ , where  $\sigma_t$  is the uncertainty on the transit centre time of each of the TESS transits and  $N$  is the integer value used to calculate each period alias. We choose  $\sigma_t = 1\text{hr}$  for this calculation, twice the cadence for the TESS full-frame images that are the source of the majority of our targets. Propagating this uncertainty forward we find the uncertainty on successive transit times,  $\sigma_T$ , to be given by

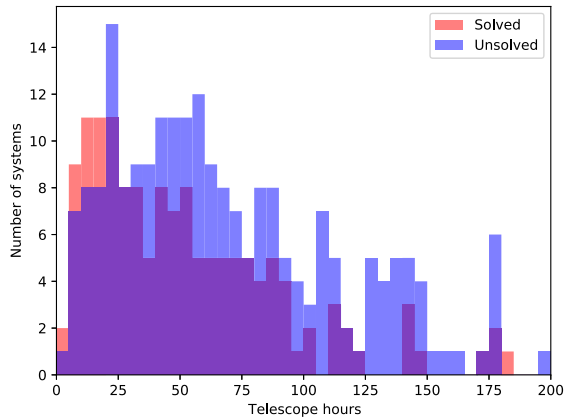
$$\sigma_T = 2\sigma_t \left( \frac{1}{2} + \frac{n}{N} \right), \quad (2)$$

where  $n$  is the number of periods after the second TESS transit. For this, we assume that our simulated observations begin approximately

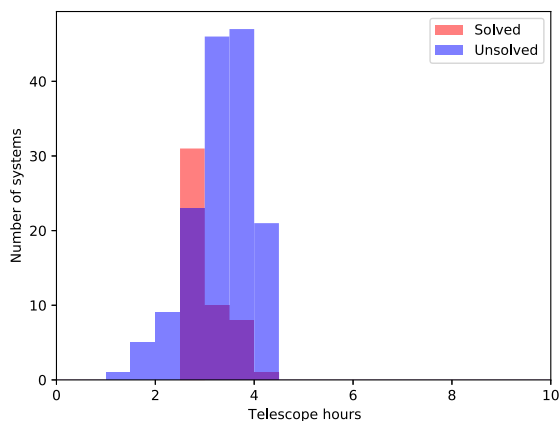
100 d after the end of the TESS extended mission for the southern ecliptic hemisphere. This is likely an overestimation, since we should be able to start following up targets from early TESS sectors while the extended mission is ongoing. For each estimated transit time for which we observe, our observation window is extended by this value to ensure transits are not missed due to inaccuracies in the TESS transit timing measurements.

In these figures, we see that each system is observed for a significantly smaller number of hours than are available within the observing span of 50 nights. Each method is split into two categories: those which are solved ( $N_{\text{aliases}, f} = 1$ , red) and those which remain unsolved ( $N_{\text{aliases}, f} > 1$ , blue). Systems with insufficient  $S/N$  are not shown. For photometry, the average number of hours observed is 62.8 and 114.2 for the solved and unsolved systems, respectively (ignoring those systems for which we have insufficient  $S/N$  that are not observed at all). The use of multiple telescopes is accounted for when calculating telescope hours. This highlights the efficacy of a targeted photometric campaign as opposed to a stare campaign, where we simply wait for an additional transit. For spectroscopy, we see the number of hours observed is 2.8 and 3.1 for solved and unsolved systems, respectively. Based on the strategy of one half-hour every 7 d, combined with the weather constraints we expect an average unsolved value of approximately  $\sim 3.3$  that matches what is seen.

Looking at these plots combined shows us that the majority of solved systems are solved with the first 5 h of spectroscopy time or the first 80 h of photometry time. Therefore, if attempting to most efficiently solve the largest number of systems, the ideal strategy may



(a) Photometric hours observed



(b) Spectroscopic hours observed

**Figure 7.** Telescope hours utilised by each system for photometry and spectroscopy. Systems are coloured for solved (red) and unsolved (blue).

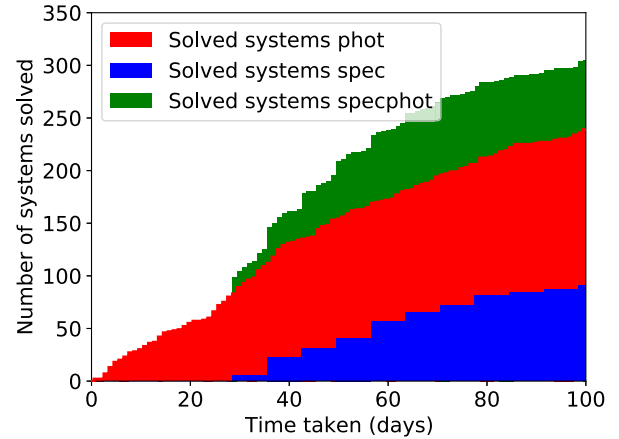
be to observe each system for 10 nights (assuming 0.5 h per night for spectroscopy and an average of 8 h per night for photometry) and if not solved move on to a different system. However, this neglects the fact that some systems are of greater scientific interest than others. Therefore, it actually requires a more careful consideration of individual system parameters to determine which systems may be worth a larger fraction of available telescope time.

#### 4.6 Effect of increasing follow-up time

The plots and analysis so far have used a time-span of 50 d and CORALIE for spectroscopy. Table A1 shows how the  $N_{\text{aliases}, f}$  values change for different combinations of time-span and spectroscopy instrument. However, it is also interesting to look in more detail at how the number of solved systems increases as the time-span is allowed to increase. This is shown in Fig. 8.

Fig. 8 shows that the fraction of systems that can be solved using either photometry or spectroscopy is a function that will asymptote before reaching unity. The reason for this is as explained above, systems with sufficiently small transit or RV signals will become lost in the noise of their respective instrumental precision's and will thus never be solved.

Initially, we see that the number of systems solved through photometry rises quite slowly. This is because, in the early stages



**Figure 8.** Cumulative histogram of solved systems as a function of additional photometric or spectroscopic time. The distribution runs for 100 d with photometry being carried out by NGTS and spectroscopy consisting of one CORALIE point every 7 d. Only systems that reach  $N_{\text{aliases}, f} = 1$  within 100 d are shown.

of the follow-up campaign, the coverage is sparse enough that only catching an additional transit will solve a system fully. Also, because there has not been time for other aliases to be ruled out by non-detections, only systems for which the observed transits produce no aliases will solve a system. As the observations continue the rate of solving begins to increase. This is due to the fact that at this stage, we are beginning to see enough aliases ruled out through nights of non-detections that catching a third transit is more likely to remove all aliases that are left. Additionally, the likelihood of a fourth transit being seen increases that dramatically reduces the number of aliases.

Spectroscopically, solved systems only begin to show up around 28 d, since we require at least 5 data points, taken once every 7 d (including a data point at time zero). As we see in Fig. 7(b), there is a spike in solved systems upon reaching this threshold; however, due to the randomness of weather effects different systems reach this threshold on different nights with 28 simply being the earliest possible time. The slow increase from here on is then the result of the difficulty of using spectroscopy to fully solve a system. Were we to focus on systems for which spectroscopy can rule out 80 per cent of aliases this peak would be much higher, but it is challenging for spectroscopy to definitively decide between similar periods.

The number of systems solved by combining the two methods is always above those solved using either method alone as expected. However, as time increases the fractional addition in number of solved systems caused by this combination above those solved by photometry alone is reduced. As seen above, at 50 d the total number of systems is increased from 149 to 207, 39 per cent, but by 100 d the increase is from 240 to 305, only 27 per cent. This seems to be due to the fact that by this time, the number of systems solved by spectroscopy is increasing only very slowly; therefore, spectroscopy is adding little to the number of overall solved systems whereas the systems solved via photometry are still increasing, albeit slower than they were originally, due to the shorter and easier systems having been solved already.

It is also of interest to note how many transits are required for a photometric solution. It is seen that for the majority of systems (122/149) only one additional transit is required to fully solve the period. The remaining solved systems are done so with two additional



transits. This matches what we see for photometry in Fig. 8. The initial quick increase in solved systems is a result of those systems that can be solved by catching only one more transit. This then levels off as we approach the longer period systems which require a larger number of additional transits to solve.

## 5 CONCLUSIONS

Catching two widely separated transits of a long period transiting system is insufficient for identifying the true period of the orbit. We have taken the TESS primary and extended mission observations of the southern ecliptic hemisphere as an example and showed that there are approximately 395 exoplanets with a single transit detected in each year. Based on the TESS data alone, we show this leaves an average of  $\sim 38$  period aliases per system with the true alias being, on average, the 15th smallest possibility. Building on this, we have simulated attempts to mitigate these aliases using additional photometric and spectroscopic data.

We find that the amount of additional photometric and/or spectroscopic observations taken can have a significant impact on both the number of fully resolved systems (defined here as systems with  $N_{\text{aliases}, f} = 1$ ) and the mean number of aliases for the unsolved systems. The results presented in Figs 2 through 8 and Table A1 can be summed up by the following points:

(i) For all simulation lengths photometry fully solves the period of significantly more systems than spectroscopy; roughly three times more for a 50 d simulation (Figs 2 and 8).

(ii) Combining both methods results in the highest total number of solved systems. At 50 d, 13 per cent of these were not solved by either method independently (Fig. 2).

(iii) When selecting only those systems which reach the required S/N threshold for each method photometry rules out at least half of the period aliases for 76 per cent, whereas the corresponding value for spectroscopy is 93 per cent (Fig. 6).

(iv) Spectroscopically solved systems favour higher mass planets that leads to a peak between 10 and  $15R_{\oplus}$  which corresponds to the largest mass planets. Additionally, it is easier to solve shorter period systems (Fig. 4).

(v) Photometrically solved systems are insensitive to mass or radius, assuming they exceed the S/N threshold. As for spectroscopy, shorter period systems are easier to solve (Fig. 4).

(vi) Systems solved through a combination of the two methods tend to have a broader period distribution, extending to longer period systems (Fig. 4).

(vii) For an observing period of 50 d, the majority of systems solved by photometry are done so in under 80 h of telescope time. All spectroscopically solved systems are solved in under 5 h (Fig. 7).

(viii) We suggest that, with the singular goal of confirming the period for the greatest number of systems, a 50 d follow-up campaign is preferred when accounting for the cost-benefit ratio. However, if attempting to characterize more scientifically interesting systems different campaign lengths may be more efficient.

(ix) Longer observing periods lead to more solved systems but both methods asymptote before solving all systems (Fig. 8).

(x) Using HARPS instead of CORALIE leads to an increase in the number of spectroscopically solved systems by a factor of 3–5 that decreases as the observing period increases (Table A1).

## ACKNOWLEDGEMENTS

BFC acknowledges a departmental scholarship from the University of Warwick. DLP acknowledges support from the Royal Society during this period.

The contributions at the University of Warwick by PJW, DRA, and SG have been supported by STFC through consolidated grants ST/L000733/1 and ST/P000495/1.

We thank the referee for their helpful comments.

## DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

## REFERENCES

- Barclay T., Pepper J., Quintana E. V., 2018, *ApJS*, 239, 2  
 Becker J. C. et al., 2019, *AJ*, 157, 19  
 Borucki W. J. et al., 2010, *Science*, 327, 977  
 Bryant E. M. et al., 2020, *MNRAS*, 494, 5872  
 Cooke B. F., Pollacco D., 2020, *MNRAS*, 495, 734  
 Cooke B. F., Pollacco D., West R., McCormac J., Wheatley P. J., 2018, *A&A*, 619, A175  
 Cooke B. F., Pollacco D., Bayliss D., 2019, *A&A*, 631, A83  
 Cooke B. F., Pollacco D., Lendl M., Kuntzer T., Fortier A., 2020, *MNRAS*, 494, 736  
 Dawson R. I., Fabrycky D. C., 2010, *ApJ*, 722, 937  
 Dholakia S., Dholakia S., Mayo A. W., Dressing C. D., 2020, *AJ*, 159, 93  
 Dressing C. D., Charbonneau D., 2015, *ApJ*, 807, 45  
 Gaia Collaboration et al., 2018, *A&A*, 616, A10  
 Gandolfi D. et al., 2018, *A&A*, 619, L10  
 Gilbert E. A. et al., 2020, *AJ*, 160, 116  
 Gill S. et al., 2019, *MNRAS*, 491, 2805  
 Gill S. et al., 2020a, *ApJL*, 898, L11  
 Gill S. et al., 2020b, *MNRAS*, 495, 2713  
 Günther M. N., Queloz D., Demory B.-O., Bouchy F., 2017, *MNRAS*, 465, 3379  
 Halbwachs J. L., Mayor M., Udry S., 2005, *A&A*, 431, 1129  
 Howell S. B. et al., 2014, *PASP*, 126, 398  
 Hsu D. C., Ford E. B., Ragozzine D., Ashby K., 2019, *AJ*, 158, 109  
 Huang C. X. et al., 2018, *ApJ*, 868, L39  
 Kanodia S., Wolfgang A., Stefansson G. K., Ning B., Mahadevan S., 2019, *ApJ*, 882, 38  
 Lendl M. et al., 2020, *MNRAS*, 492, 1761  
 Mayor M. et al., 2003, *Messenger*, 114, 20  
 Montalto M. et al., 2020, *MNRAS*, 498, 1726  
 Osborn H. P. et al., 2016, *MNRAS*, 457, 2273  
 Pollacco D. L. et al., 2006, *PASP*, 118, 1407  
 Pont F., Husnoo N., Mazeh T., Fabrycky D., 2011, *MNRAS*, 414, 1278  
 Queloz D. et al., 2000, *A&A*, 354, 99  
 Ricker G. R. et al., 2015, *J. Astron. Telesc. Instr. Syst.*, 1, 014003  
 Stassun K. G. et al., 2018, *AJ*, 156, 102  
 Stassun K. G. et al., 2019, *AJ*, 158, 138  
 Thompson S. E. et al., 2016, Kepler Data Release 25 Notes Kepler Science Document KSCI-19065-002  
 Vanderspek R. et al., 2019, *ApJ*, 871, L24  
 Villanueva Steven J., Dragomir D., Gaudi B. S., 2019, *AJ*, 157, 84  
 Wheatley P. J. et al., 2018, *MNRAS*, 475, 4476  
 Winn J. N., 2010, preprint ([arXiv:1001.2010](https://arxiv.org/abs/1001.2010))

## APPENDIX: NUMBER OF ALIASES

**Table A1.**  $N_{\text{aliases, f}}$  results from 10 combinations of time-span and spectroscopic instrument. For each combination of observations, we show the number of solved systems.  $N_{\text{aliases, f}}$  then refers to the mean number of aliases for all unsolved systems. The bold row matches the data plotted in Figs 2, 6, and 7. Data for the TESS only simulation are not shown as they are unchanged by follow-up strategies.

Time-span (d)	Spectroscopy instrument	Solved (phot)	$N_{\text{aliases, f}}$ (phot)	Solved (spec)	$N_{\text{aliases, f}}$ (spec)	Solved (specphot)	$N_{\text{aliases, f}}$ (specphot)
30	CORALIE	86	19.2	9	33.7	104	18.4
30	HARPS	90	18.5	41	31.9	129	16.6
40	CORALIE	118	16.8	21	29.4	156	15.8
40	HARPS	131	16.7	94	23.4	220	14.1
<b>50</b>	<b>CORALIE</b>	<b>149</b>	<b>14.2</b>	<b>50</b>	<b>25.7</b>	<b>207</b>	<b>11.6</b>
50	HARPS	150	13.7	191	17.9	298	8.8
60	CORALIE	170	11.9	57	22.5	242	9.6
60	HARPS	180	12.8	232	17.4	323	7.9
70	CORALIE	187	10.8	82	24.9	254	10.2
70	HARPS	198	10.9	243	17.5	340	7.5

This paper has been typeset from a  $\text{\TeX}/\text{\LaTeX}$  file prepared by the author.