MNRAS **000**, 1–9 (2018) Preprint 26 March 2018 Compiled using MNRAS LATEX style file v3.0

TESS Asteroseismic Predictions for Red Giants using Machine Learning

M. Schofield[1,2]★, G. R. Davies[1,2], W. J. Chaplin[1,2], A. Miglio[1,2], M. F. Randrianandrasana
[1]*Department of Physics and Astronomy, the University of Birmingham, Birmingham B15 2TT, UK*
[2]*Stellar Astrophysics Centre (SAC), Department of Physics and Astronomy, Aarhus University, Ny Munkegade 120, DK-8000 Aarhus C, Denmark*

Last updated 2017 May 31; in original form 2017 May 31

ABSTRACT

Summary: This paper presents a method to identify the detectable solar-like modes in Red Giants with TESS using a Machine Learning Classifier. It requires only the global parameters $\log(g)$, π, [M/H], $T_{\rm eff}$ and the stellar magnitude. This can be used as an asteroseismic target-selection function for future space missions.

Method: Lightcurves for *Kepler* stars with fitted radial mode frequencies and amplitudes from Davies & Miglio (2016) were used. These lightcurves were degraded to generate equivalent TESS lightcurves. The length of observation was reduced, *Kepler* white noise was removed, the bandpass was adjusted, and TESS white noise was added. A detection test was run on 3 observed modes in these 'TESS-like' lightcurves. Lastly, a Classifier was used to predict mode detectability with TESS based upon global photometric and spectroscopic parameters. The Classifier successfully predicted mode detectability in the original *Kepler* stars, and for 1 year of TESS-like data.

Application: By changing only the length of dataset, instrumental noise level and bandpass of observation, this tool can make solar-like detection predictions for future missions such as K2, PLATO and CoRoT. It can make predictions and select targets much faster than traditional detection tests. It is especially useful in an age where extremely large datasets from MAST and Gaia are available (Eisenstein et al. (2011), Gaia Collaboration et al. (2016)).

1 INTRODUCTION

Satellites such as *Kepler* have allowed asteroseismology of solar-like and Red Giant stars to advance rapidly since the last century Chaplin & Miglio (2013). Power spectra can now be resolved to detect individual modes of oscillation in Main Sequence and Red Giant stars (Lund et al. (2017), Davies & Miglio (2016)).

Future space missions such as TESS (Ricker et al. 2014), K2 (Howell et al. 2014), CoRoT (Baglin et al. 2006) and PLATO (Rauer et al. 2014) will add to our understanding of stellar structure and evolution. These missions will provide a large amount of high-precision data. More than ever, the field of stellar astrophysics will require tools to perform big-data analysis (Kremer et al. 2017).

One of the tools than can be used to handle the large amount of data from future missions is Machine Learning. In this work, Machine Learning was used to create a TESS target selection function using the set of *Kepler* Red Giant Branch, Red Clump and Secondary Clump stars from Davies & Miglio (2016). The algorithm made in this work is publicly-available[1]. It can be downloaded and used as a target selection function for any future space mission.

In most situations, Machine Learning is used to solve problems in one of two ways: either by using Supervised or Unsupervised Learning. Unsupervised Learning involves problems where there are no known labels or results. Conversely, Supervised Learning involves problems where there is a known result.

In Unsupervised Learning, there are no known results/labels. The aim of Machine Learning in this case would be to find trends between variables. This has been be used to identify the evolutionary stage of stars by analysing their lightcurves without using previously labelled data. The evolutionary stage of variable stars has been determined in this way using an Unsupervised Classifier (Valenzuela & Pichara 2018). Similarly, a neural network has been used to identify the evolutionary stage of Red Giant stars (Hon et al. (2017), Hon et al. (2018)).

Supervised Learning has also been used to classify the evolutionary stage of variable stars (Nun et al. (2014), Elorrieta et al. (2016)). In these papers, classifying evolutionary stage was treated as a Supervised problem because previously labelled data were used. This previously labelled data is known as training data: this is used to train the Machine Learning algorithm. In the problem of variable stars, lightcurves that had already been classified were used to train the algorithm (this is the training dataset). This algorithm was then used to classify the lightcurves of unidentified stars (this is known as the testing dataset). In this way, the authors were able to identify the evolutionary stage of every star in their sample.

In this work, a classifier was *not* used to identify the evolutionary stage of stars. For the first time, a Supervised Classifier was used here to identify which solar-like modes could be detected inside Red Giant stars with TESS. These modes were given to the

[1] https://github.com/MathewSchofield/TRG

Classifier as known labels, so this is a Supervised Learning problem[2].

A Classifier was used here to identify which solar-like modes could be detected inside Red Giant stars with TESS. As a result, this algorithm can be used as the target selection function for TESS. Furthermore, this code has been written specifically for easy adaption to other satellites. It can therefore also be used as the target selection function for K2, CoRoT and PLATO.

Within Supervised Learning, two common algorithms that are used are Classification and Regression. In Regression, the relationship between variables is interpreted using a measure of uncertainty (such as using χ^2 tests). Models are fitted using the independent data, and uncertainty is measured. The models are then improved by reducing this uncertainty. Note that regression is used when the label is continuous. For example, predicting the magnitude of a star is a problem suited to regression, as a star can have any magnitude (Steinhardt & Jermyn 2018).

Conversely, Classification algorithms work by assessing similarity[3]. In Classification, the training set is separated into groups based on the similarity of the data. The more information that was gained by splitting the data, the better. For example, if the problem were to separate Red Giant stars from Main-Sequence stars, a star could be classified as either a Red Giant (1), or not a Red Giant (0). In this example, having a Luminosity above $\sim 10L_{\odot}$ would be a strong indicator that the star was a Red Giant so the data could be separated into groups here. The Classifier would continue to separate the dataset until the Red Giant and Main Sequence samples were distinct. Examples of Classifiers being used to identify the evolutionary stages of stars cab be found in Ness et al. (2015) and Wu et al. (2017).

In this work, a Classifier was not used to identify the evolutionary stage of stars. Instead, it was used to identify which solar-like modes could be detected by TESS, and which could not. individual fitted modes from Davies & Miglio (2016) were degraded to look like TESS observations. The detectable solar-like modes from this degraded set were then identified by a Classifier. By separating the stars into those with detected modes and those without, a Classifier was used to select the optimal targets for TESS to observe.

Firstly, Section 2 describes how the size of the datasets were increased to improve the predictive ability of the Classifier. Section 3 then outlines how the timeseries of every star were treated before transforming them into power spectra. Section 4 goes on to explain the detection test that was run on the solar-like oscillations. This returned a probability of detection $P_{\rm det}$ for every mode. Each mode was grouped into a discrete class depending upon it's detection probability; each mode was either very likely to be observed (2), quite likely (1), or unlikely (0).

Lastly, Section 5 illustrates how the stars were classified. The stars were put into a group with detected modes, and a group without. This was done by giving a Supervised Classifier photometric and spectroscopic information on every target from APOKASC (Pinsonneault et al. 2014). 70% of the stars were used to train the Classifier; 30% of the sample was kept to test the algorithm. The Classifier was given global asteroseismic and spectroscopic information, as well as mode detection probabilities of the stars in the training set. It then made predictions about mode detectability on the testing set.

The Classifier recognised patterns between the variables in the training set, and successfully made predictions about mode detectability with a 0.98% precision for the original *Kepler* data. It achieved a precision of 0.86 for 1 year of TESS-like observation, and 0.74 for 27 days of TESS-like observation.

[2] https://machinelearningmastery.com
[3] http://www.simafore.com

2 THE DATASET

Machine Learning performs best when a large dataset is available to train the algorithm on. In order to increase the size of the dataset from Davies & Miglio (2016) above 1000, the magnitude of each *Kepler* star was perturbed. Each star had it's magnitude perturbed 100 times before the lightcurves were transformed to TESS-like power spectra.

The noise functions of *Kepler* and TESS were used as PDFs to draw magnitudes from. These were used because they provide realistic distributions of the number of stars at different magnitudes that the satellites observed/will observe. Many more fainter stars are observed than brighter stars because the volume of space that contains stars increases as the distance of observation increases.

The noise function of *Kepler* depends on the *Kepler* magnitude of the star, K_p. This noise function is from Chaplin et al. (2011). It is given by

$$\sigma = \frac{10^6}{c} \times \sqrt{c + 9.5 \times 10^5 \left(\frac{14}{K_p}\right)^5}, \quad (1)$$

where

$$c = 1.28 \times 10^{0.4(12-K_p)+7}. \quad (2)$$

Similarly, the noise function of TESS was estimated using the 'calc noise' IDL procedure (from William Chaplin, private communication), which depends on the I_c-band magnitude of the star and the number of pixels in the photometric aperture used when observing the star. This is given by

$$N_{\rm aper} = 10 \times (n + 10), \quad (3)$$

where n is

$$n = 10^{-5.0} \times 10^{0.4\times(20-I_{\rm mag})}. \quad (4)$$

These noise functions were used as the PDFs to draw stellar magnitudes from. 100 magnitudes were drawn for every *Kepler* Red Giant. After removing gaps in the data, this left 60,000 samples. The lightcurves of this much larger *Kepler* dataset were then degraded to look like TESS observations.

3 MODIFYING *Kepler* DATA TO LOOK LIKE TESS

Before a Classifier could be used on the large Red Giant sample, the timeseries data from *Kepler* needed to be adjusted for a different satellite and mission. This could either be done in the time or frequency domains. Adjustments were made in both domains, and the results were compared.

Several different adjustments needed to be made to the *Kepler* data. One difference between the missions is the length of observation. The *Kepler* mission observed for 4 years, while TESS' nominal 2 year mission will observe stars for between 27 days to 1 year, according to the star's ecliptic latitude (Figure 1). This distinction can be clearly seen when comparing the timeseries', see Figure 2.

As well as reducing the dataset length, the bandpass of observation needed to be adjusted. TESS will observe in a much redder

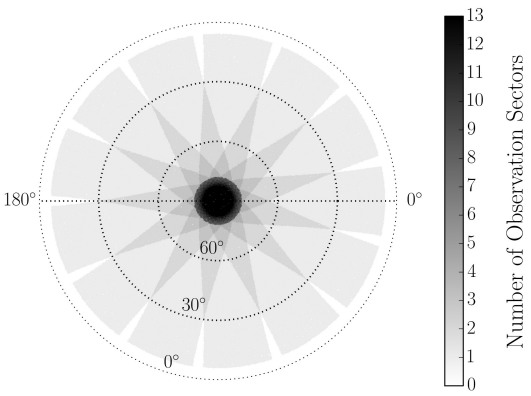

Figure 1. TESS field of view, centred around the ecliptic pole. Each strip will be observed for 27 days before the satellite rotates. Image taken from Campante et al. (2016).

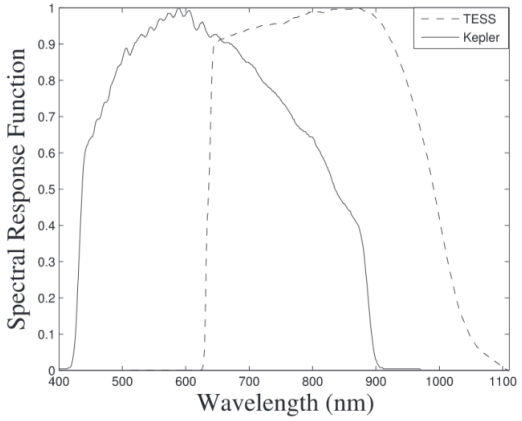

Figure 3. The bandpass' of the *Kepler* and TESS missions. TESS will observe at longer (i.e redder) wavelengths than *Kepler*. This will reduce the amplitude of oscillations and granulation, whilst the white noise level will by unaffected. Image taken from Placek et al. (2016).

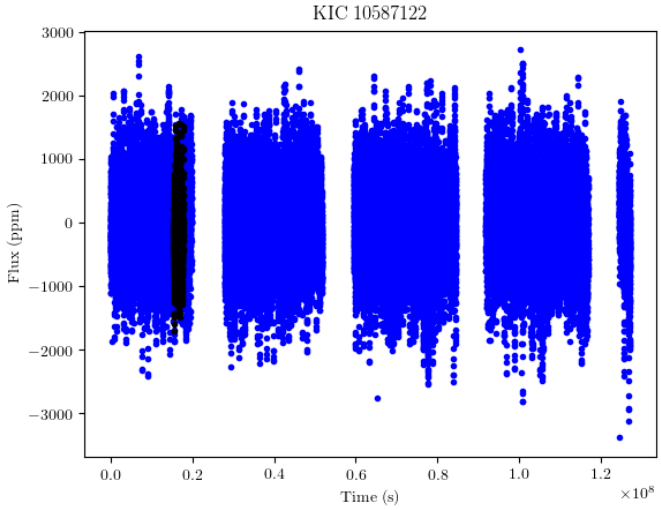

Figure 2. The 4-year long Power Spectra of KIC 10587122 is plotted in blue. Overplotted is the 27-day time segment with most coverage (the period with fewest gaps in the data). Reducing the length of observation this drastically will badly hamper mode detectability.

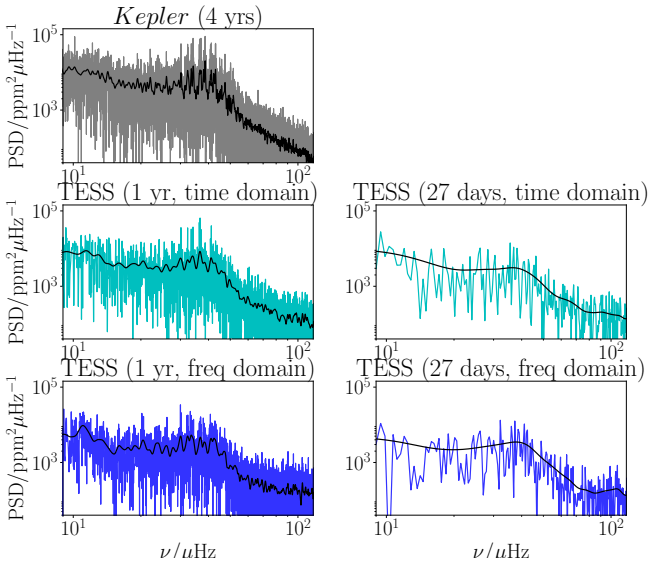

Figure 4. The Power spectra of KIC 10587122 is plotted five times with moving medians in black. The original Power Spectra is plotted in grey. The power spectra after making the data look like TESS are plotted in blue. The data was transformed in the time domain (light blue) or frequency domain (dark blue). The left column shows 1 year of TESS observation (the maximum). The right column shows 27-days (the minimum). Based on this, the time domain was chosen to transform the data in as the background noise level appears lower.

bandpass than that of *Kepler* , Figure 3. This has the effect of reducing the amplitude of stellar signals (i.e the signals due to stellar granulation and oscillation), while leaving the white noise component unaltered (Ballot et al. 2011). Campante et al. (2016) found this bandpass correction to be 0.85 for TESS.

Thirdly, the instrumental noise level in *Kepler* is different to the noise level in TESS. The noise level for the *Kepler* satellite depends on the *Kepler* magnitude of the star, equation 1.

Instrumental noise from TESS needed to be added to the time-series'. This noise level was estimated using equation 4, along with the 'calc noise' IDL procedure (from William Chaplin, private communication).

These three adjustments - the length of observation, the bandpass, and the noise level - were performed in both the time and frequency domains for comparison. The methods used are described in Figures 6 and 7. The resulting Power Spectra are compared in Figures 4 and 5. After comparing the results, the decision was made to perform adjustments in the time domain before transforming to the frequency domain because the background noise level is lower when the time domain method is used to transform the lightcurves to TESS-like power spectra.

After adjustments were made in the time domain, power spectra were generated for every star. Power spectra were generated for the original 4-year *Kepler* sample, 1 year of TESS observations and 27 days of TESS observations. A detection test was then run on the radial modes of every star in these 3 datasets.

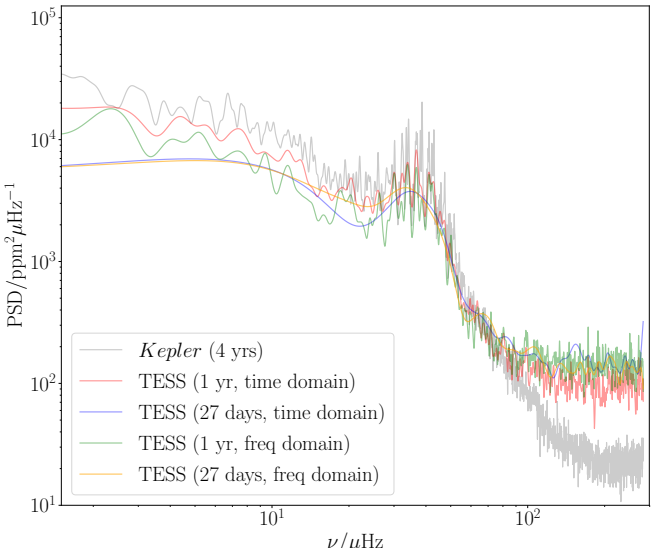

Figure 5. The Power Spectra of KIC 10587122. The original power spectra is in grey. The data was transformed into TESS observation and overplotted. The transformation was done in the time and frequency domains for comparison. Based on this, the time domain was chosen to transform the data in as the background noise level appears lower.

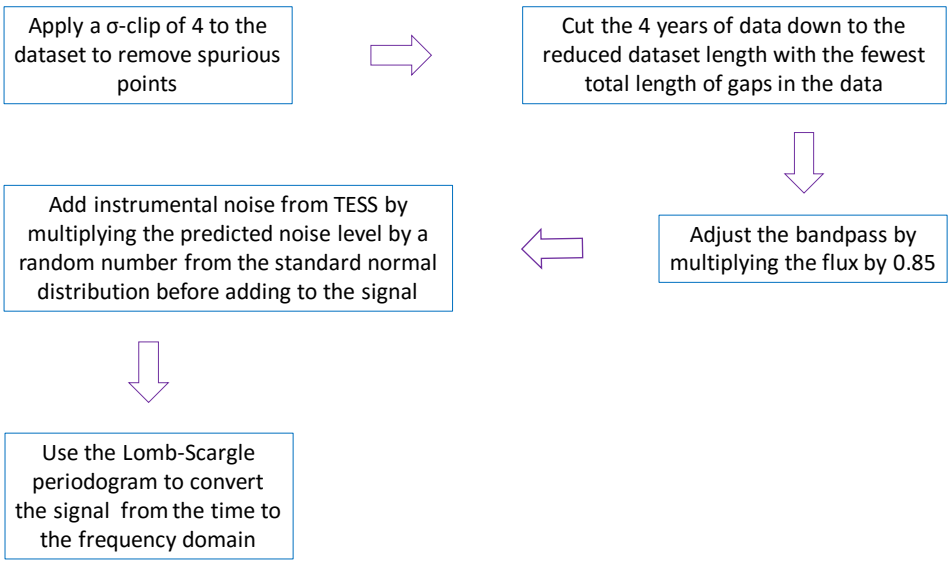

Figure 6. Flow chart of the method to convert the data from *Kepler* to TESS observations in the time domain.

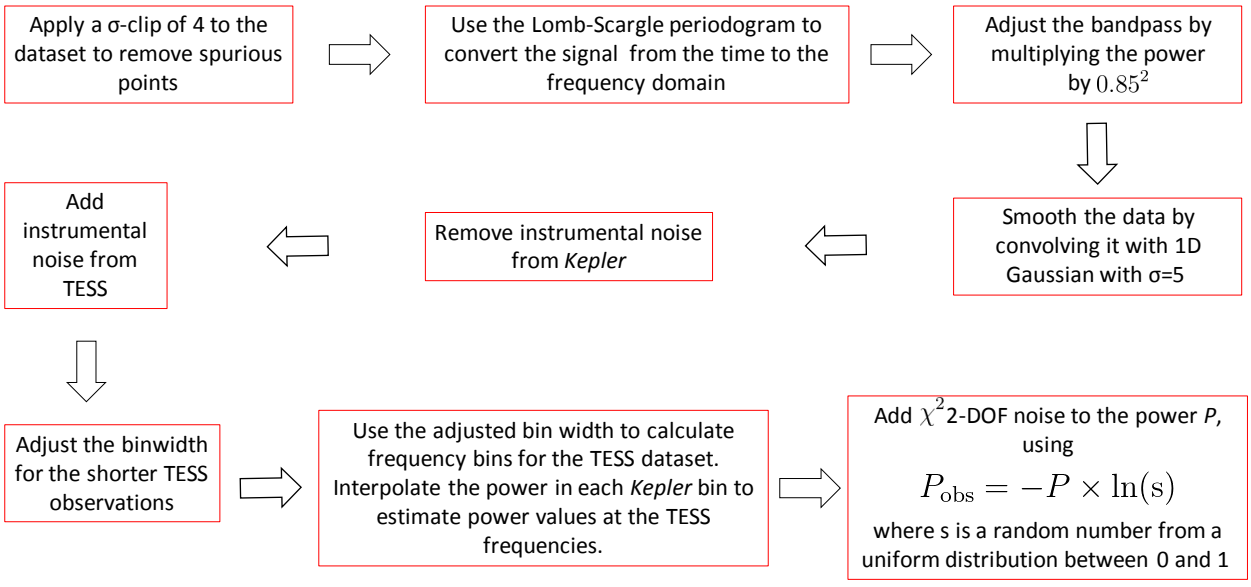

Figure 7. Flow chart of the method to convert the data from *Kepler* to TESS observations in the frequency domain.

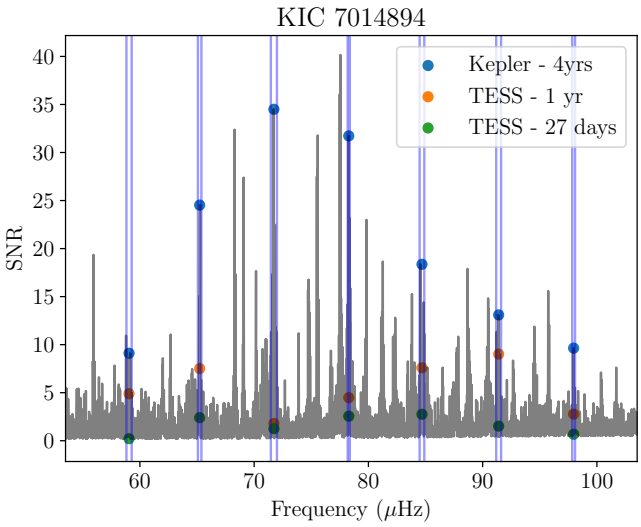

Figure 8. The SNR spectrum of KIC 7014894 after background subtraction. The SNR values of the radial modes in the star were extracted from this spectrum. The highest SNR value within the linewidth of each mode is taken to be the SNR value of that mode. The mode linewidths are shown as blue lines. The values of every mode in the original Kepler spectrum are plotted in blue. The overplotted orange points are the SNR values after degrading the signal to 1 year of TESS observation. Similarly, the green points are the SNR values of 27 days of TESS observations. The white noise level and reduced observation time severely reduce the SNR of TESS observations compared to *Kepler* .

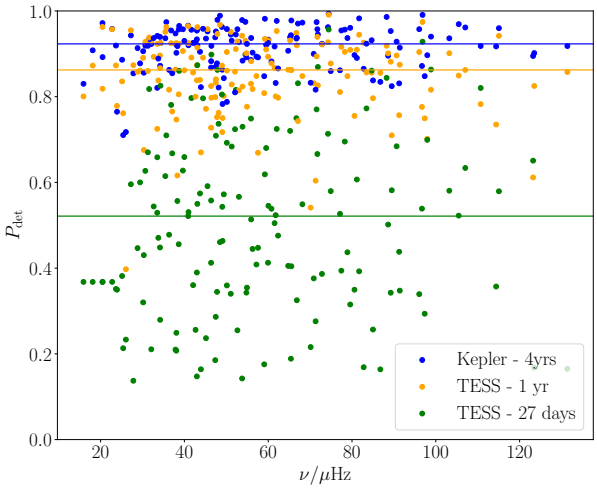

Figure 9. A plot showing the result of the detection test, after running on every mode in 20 stars. The results of the original power spectra are plotted in blue. The results of 1 year of TESS observation are in orange. 27 days of TESS observation is in green. At this short an observation, detecting individual modes will be extremely difficult.

4 DETECTION TEST

Section 3 described the method to transform the *Kepler* lightcurves into TESS-like power spectra. A detection test was then run on the stars to determine which modes were visible after observation by TESS, and which were not.

First, a moving median from Davies & Miglio (2016) was used to estimate the proportion of the signal that was due to white noise and stellar granulation. The solar-like mode envelope width was used as the frequency range of the moving median. This envelope width was calculated as

$$\Gamma_{\rm env} = 0.66\,\nu_{\rm max}^{0.88}, \tag{5}$$

from Mosser et al. (2012). The moving median was used to interpolate between frequencies in the power spectrum. It provided an estimate of the background B in the power spectrum (ppm$^2\mu$Hz^{-1}). This background was divided out of the power P (ppm$^2\mu$Hz^{-1}) in the power spectrum to the get Signal-to-Noise ratio of the spectrum,

$$\mathrm{SNR} = P/B. \tag{6}$$

Once the SNR spectrum for the star was recovered, the SNR values at the mode frequencies were extracted. To ensure the correct SNR values of every mode were used, a window was fitted around each peak-bagged mode frequency. The size of the window was given as the linewidth of the mode. The highest value in the window was taken as the SNR of the mode. An example of this for KIC 7014894 is shown in Figure 8.

Once all mode SNR values for the star were calculated, a detection test was run on each mode (Chaplin et al. 2011), (Campante et al. 2016).

The probability P that the SNR value of a solar-like mode of oscillation lies above some threshold $\mathrm{SNR}_{\rm thresh}$ is

$$P(\mathrm{SNR} \geqslant \mathrm{SNR}_{\rm thresh}) = p. \tag{7}$$

A false-alarm probability p of 5% was set; there is a 95% chance that the signal is due to solar-like oscillations, rather than noise. Equation 7 is solved for $\mathrm{SNR}_{\rm thresh}$ by substituting P with

$$P = \int_x^\infty \frac{e^{-x}}{\Gamma(N)} x^{N-1} dx. \tag{8}$$

N is the number of frequency bins that the mode occupies. In Davies & Miglio (2016), the Full-Width at Half Maximum linewidth of every peak-bagged mode in the Red Giants is given. The linewidth of each mode is used as the value of N in equation 8.

$\Gamma(N)$ is the Gamma function. The lower bound of Equation 8 is set to $x = 1 + \mathrm{SNR}_{\rm thresh}$. The noise in these bins is assumed to follow χ_2 $2n_{\rm bins}$ d.o.f statistics.

Once $\mathrm{SNR}_{\rm thresh}$ is found, Equation 8 is solved again. This time it is solved for P by setting $x = (1 + \mathrm{SNR}_{\rm thresh})/(1 + \mathrm{SNR})$. Here, SNR is the observed Signal-to-noise Ratio calculated from Equation 6. This calculates the probability that the frequency spike was due to stochastic excitation in the convective envelope of the star.

This detection test was applied to every mode of every star in the sample. Figure 9 shows the mode detection probabilities from this test for the original *Kepler* dataset, for 1-year of TESS observation and for 27 days of TESS observation. After the $P_{\rm det}$ values for the 3 datasets were calculated, a Classifier was used to determine if these results could be predicted rather than calculated. This is described in Section 5.

5 CLASSIFICATION

Section 3 describes how the lightcurves of every star were treated before a detection test was run on the oscillations in Section 4. After this, Classification was applied to the stars to separate them into

KIC	Iteration	log(g)	π	$T_{\rm eff}$	[M/H]	$I_{\rm mag}$
9205705	1	2.758	0.688	4685	-0.39	9.89
9205705	2	2.758	0.688	4685	-0.39	9.19
9205705	3	2.758	0.688	4685	-0.39	9.79
9205705	4	2.758	0.688	4685	-0.39	11.30
...						
9205705	100	2.758	0.688	4685	-0.39	7.81
2554924	1	2.799	0.969	4594	0.27	8.46
2554924	2	2.799	0.969	4594	0.27	9.26
...						

Table 1. An example of the X-dataset for 1 year of TESS-like observations. Every star has it's magnitude perturbed 100 times. See Table 2 for the equivalent Y-dataset.

KIC	Iteration	$P_{\rm det}$ (1)	$P_{\rm det}$ (2)	$P_{\rm det}$ (3)
9205705	1	1	2	2
9205705	2	1	2	2
9205705	3	1	2	2
9205705	4	1	2	1
...				
9205705	100	1	2	2
2554924	1	2	2	2
2554924	2	2	2	2
...				

Table 2. An example of the Y-dataset for 1 year of TESS-like observations. Every star has it's magnitude perturbed 100 times. White noise is then added to the timeseries and mode detection probabilities are calculated for 3 radial modes centred around $\nu_{\rm max}$. Lastly, these probabilities are put into discrete classes [0, 1 or 2]. The radial mode closest to $\nu_{\rm max}$is labelled $P_{\rm det}$ (2). See Table 1 for the equivalent X-dataset.

a suitable target list, and a list of stars that are not suitable for observation with TESS.

5.1 Preparing the data

Firstly, the detection probabilities of every mode were taken from Section 4. Each probability was put into a discrete bin (or *class*) depending on how likely the mode is to be detected. These discrete classes are given in equation 9.

$$P_{\rm det} = \begin{cases} 2 & \text{if } 1.0 \geqslant P_{\rm det} > 0.9 \\ 1 & \text{if } 0.9 \geqslant P_{\rm det} > 0.5 \\ 0 & \text{if } 0.5 \geqslant P_{\rm det} > 0.0 \end{cases}, \quad (9)$$

Using equation 9, every mode was assigned a discrete class [0, 1 or 2], depending on how high the probability of detection was for that mode. The same three radial modes (*3 labels*) were used for every star: the mode closest to the centre of the power-excess due to solar-like oscillations $\nu_{\rm max;n}$, the radial mode one overtone below that $\nu_{\rm n-1}$, and one overtone above that, $\nu_{\rm n+1}$ [$P_{\rm det}$ (1), $P_{\rm det}$ (2), $P_{\rm det}$ (3)]. It was important to use the same information for every star so that the algorithm could be trained on the patterns between the variables.

A classifier is an algorithm that can learn a relationship between variables. The classifier will map from some initial information about the star (the X data), to some unknown information (the Y data). In this work, the X data were magnitude (K_p or $I_{\rm mag}$), log(g), π, $T_{\rm eff}$ and [M/H]. The Y data labels were the $P_{\rm det}$ values of 3 radial modes, centred around $\nu_{\rm max}$. An example of the final dataset for 1 year of TESS-like observations are shown in Tables 1 and 2.

5.2 Target selection using a Classifier

The 60,000 samples were separated into a training dataset, and a testing set. 70% of the samples were used to train the classifier (46,410 stars); 30% of the stars were used to test the algorithm (19,890 stars). To train the Classifier, the X and Y data in the training set was given to the algorithm ($\rm X_{train}$ and $\rm Y_{train}$). Once the Classifier had been trained, the X data from the testing set was given to it ($\rm X_{test}$). The Classifier then predicted a set of Y data for the testing set ($\rm Y_{pred}$). This was compared to the actual Y data for the testing set ($\rm Y_{test}$). The more similar $\rm Y_{pred}$ is to $\rm Y_{test}$, the better the Classifier replicated the data.

Two metrics were used to measure the performance of the algorithm. The first was the precision of the Classifier, weighted across the classes [0, 1, 2] and the detection probability labels

Satellite	$T_{\rm obs}$	Precision	Hamming loss
Kepler	4 years	0.98	0.02
TESS	1 year	0.86	0.07
TESS	27 days	0.74	0.25

Table 3. Results of the Classifier on the original *Kepler* dataset, and the 1-year and 27-day TESS datasets. The 'Precision' column gives the average weighted precision of the Classifier across the 3 classes [0, 1, 2] and 3 labels [$P_{\rm det}$ (1), $P_{\rm det}$ (2), $P_{\rm det}$ (3)].

$P_{\rm det}$ (1), $P_{\rm det}$ (2) and $P_{\rm det}$ (3). The second was the Hamming loss[4] (Wegner 1960) of the algorithm. This was used to give a measure of similarity between the predicted $P_{\rm det}$ values $\rm Y_{pred}$, and the testing $P_{\rm det}$ values $\rm Y_{test}$;

$$H_{\rm loss}(Y_{\rm test}, Y_{\rm pred}) = \frac{1}{n_{\rm labels}} \sum_{j=0}^{n_{\rm labels}-1} 1(Y_{\rm pred} \neq Y_{\rm test}) \, . \quad (10)$$

A Hamming loss score of 0.0 means that $\rm Y_{pred}$ is identical to $\rm Y_{test}$. A score of 1.0 means that there are no similar values between $\rm Y_{pred}$ and $\rm Y_{test}$. The precision and Hamming loss of the Classifier on the *Kepler* and TESS datasets are shown in Table 3.

The classifier was adjusted in order to improve the predictions made by the classifier on the *Kepler* and TESS datasets. This was done by varying Equation 9. It was varied by changing the number of classes and range of $P_{\rm det}$ values for each class. The number of classes was varied from 2 (i.e the mode was detected (1) or it was not (0)) to 6. The width of each bin was also varied to ensure that bins were not underpopulated. It was found that the 3 classes and $P_{\rm det}$ ranges given in equation 9 gave the best predictions for the 3 *Kepler* and TESS datasets.

The results for the original *Kepler* dataset and for 1-year of TESS data are very good; the Classifier was able to replicate the mode detection predictions of the stars in these cases. This means that a Classifier can be used as a tool for target selection for future missions. Once the Classifier predicted $P_{\rm det}$ values, the stars could be ranked from those with many detected modes, to those with the fewest. In this way, the Classifier could be used as the target selection function of solar-like oscillators for TESS.

For the 27-day TESS dataset, the Classifier was only able to correctly predict the $P_{\rm det}$ values of 74% of the modes. This is likely

[4] http://scikit-learn.org

Label	Feature Importance
$\log(g)$	0.33
π	0.08
$T_{\rm eff}$	0.12
[M/H]	0.08
$I_{\rm mag}$	0.39

Table 4. The feature importance of each x-data label. The higher the feature importance, the more important the label was in predicting the detectability of modes inside the Red Giants.

to be because with these TESS targets, the dataset is too short and the white noise level is too high for individual modes to be reliably detected. In this case, the Classifier could not be expected to perform better. It is arguably no less reliable than a mode detection test, although it is much faster for large datasets.

5.3 Feature Importance

Here, a Classifier was used to predict the detection probabilities of individual modes in *Kepler* Red Giant stars. As well as being much faster than a conventional mode detection test, an added bonus of the Classifier method is that it returns the 'feature importance' of each label in the x-data.

As Table 1 shows, the x-data labels that are given to the Classifier are `[log(g), π, T_eff , [M/H], I_mag ]`. Some of these labels are more important than others when predicting the detection probability of solar-like modes `[P_det (1), P_det (2), P_det (3)]`. The relative importance of these labels is known as the feature importance. This feature importance sums to 1.

The feature importances of the x-data labels are given in Table 4. The values show that the label with the highest influence on mode detectability is stellar magnitude $I_{\rm mag}$. This is not surprising: a higher value of $I_{\rm mag}$ will result in more χ^2 2 Degrees-of-Freedom white noise in the observation (Equations 1 & 4). This will make oscillations less likely to be detected above the white noise background.

After stellar magnitude, surface gravity $\log(g)$ also has a heavy influence on $P_{\rm det}$ value. This is because $\log(g)$is a proxy for the central frequency of the solar-like mode envelope $\nu_{\rm max}$ ($\nu_{\rm max} \propto \log(g)$). As a star evolves from the Main-Sequence and up the Red Giant Branch, the radius of the star increases. This decreases the $\log(g)$ and $\nu_{\rm max}$values of the star. Kjeldsen & Bedding (1995) showed that oscillation amplitude is proportional to bolometric luminosity divided by mass ($A_{\rm osc} \propto L/M$). As a star evolves 'up' the Hertzsprung-Russell diagram, it's bolometric luminosity increases while it's mass roughly stays the same. This leads to larger oscillation amplitudes which are more likely to be detected. Mathur et al. (2011) also gives a good explanation of how the granulation properties (and hence the oscillation profile) change as a star evolves.

Comparatively, effective temperature $T_{\rm eff}$ is less useful than $\log(g)$ when predicting detection probability `[P_det (1), P_det (2), P_det (3)]`. It is less closely tied to the oscillation properties of the star than $\log(g)$, although it is a property used to estimate $\nu_{\rm max}$with the scaling relations (Chaplin et al. 2011). Lastly, parallax π and metallicity [M/H] are the least strongly tied to the oscillation properties of these RG stars.

Evolution	Satellite	$T_{\rm obs}$	Precision	Hamming loss
RGB	*Kepler*	4 years	0.98	0.03
RGB	TESS	1 year	0.83	0.16
RGB	TESS	27 days	0.75	0.25
RC	*Kepler*	4 years	0.96	0.04
RC	TESS	1 year	0.90	0.09
RC	TESS	27 days	0.77	0.24
2CL	*Kepler*	4 years	0.97	0.03
2CL	TESS	1 year	0.83	0.14
2CL	TESS	27 days	0.69	0.30

Table 5. Results of the Classifier when RGB, RC and 2CL stars are separated. Results are shown when the data are treated like *Kepler* stars, and when they are degraded to look like 1-year and 27-day TESS observation. The 'Precision' column gives the average weighted precision of the Classifier across the 3 classes [0, 1, 2] and 3 labels [$P_{\rm det}$ (1), $P_{\rm det}$ (2), $P_{\rm det}$ (3)].

5.4 Comparing results between different evolutionary states

So far in this paper, the the evolutionary state of the Red Giant population from Davies & Miglio (2016) has been ignored when predicting mode detection probability $P_{\rm det}$. In reality, the 1000 *Kepler* Red Giants used in this work are a mixture of Red Giant Branch (RGB), Red Clump (RC) and Secondary Clump (2CL) stars. This Section investigates whether there is any difference in the predictions if the stars are first separated into RGB, RC and 2CL groups. The evolutionary states of the stars in the sample were taken from Elsworth et al. (2017).

If a Red Giant Branch is massive enough, it will undergo the Helium flash and become a Red Clump star. Red Clump stars are Helium-core burning stars, and all have very similar core masses to each other. These stars have very different $g-$mode period spacings to RGB stars, but are otherwise difficult to differentiate (Chaplin & Miglio (2013), Bedding (2011), Beck et al. (2011)). If a star is more massive than $\geqslant 1.8 M_{\odot}$, it will instead become a Secondary Clump star. This means that it will undergo Helium-core burning without the Helium flash.

The Red Giant dataset was separated into 3 subsets; the RGB, RC and 2CL stars. Each subset was treated in the same way as in Section 5.2 Table 3. Table 5 gives the full list of results when the RGB, RC and 2CL stars are separated.

Table 5 shows that there is a negligible difference in predictions between stars undergoing Helium-core burning (RC and 2CL stars) and those that are not (RGB stars). This leads to the conclusion that a Classifier can predict the mode detectability of Red Giants at different evolutionary states equally well. It is therefore not necessary to separate out these stars into different evolutionary states before predicting $P_{\rm det}$.

6 CONCLUSION

1000 peak-bagged *Kepler* Red Giant stars from Davies & Miglio (2016) were used to determine whether asteroseismic target selection could be done using a Classifier. Rather than using Machine Learning to determine the evolutionary state of these stars, a Classifier was instead used to predict which individual solar-like oscillations inside the Red Giants could be detected with a future space mission. The mission in question here was TESS, although the same technique can be easily applied to K2, PLATO or CoRoT.

The number of samples from Davies & Miglio (2016) was first increased by perturbing stellar magnitudes 100 times for each

star. These perturbed magnitudes were drawn from a PDF of the noise function (equations 1 and 4). After removing stars where global parameters or fitted modes were unavailable, this left 60,000 *Kepler* samples.

Once the number of samples was increased, the timeseries' were degraded to transform them into TESS-like observations. The dataset length was reduced, white noise was added to the signal, and the bandpass of observation was reddened. A moving median was calculated for the power spectra of these Red Giants to estimate the total background in the signal. This was divided out of the spectra, leaving a Signal-to-Noise ratio at every frequency bin (equation 6).

A detection test was then run on the SNR values at every mode frequency given in Davies & Miglio (2016). This gave a detection probability $P_{\rm det}$ between 0.0 and 1.0 for every mode. In order to prepare the detection probabilities before Classification, each continuous $P_{\rm det}$ value was assigned a discrete class ([0,1 or 2]; equation 9).

A Classifier was then given the global photometric and spectroscopic properties of the Red Giant sample, along with mode detection probabilities for each star. The parameters $[\log(g),\ \pi,\ T_{\rm eff}\ ,\ [{\rm M/H}],\ I_{\rm mag}\ ]$ from APOKASC (Pinsonneault et al. 2014) were used as the 5 X-data labels. The $P_{\rm det}$ values of 3 radial modes centred around $\nu_{\rm max}$ were used as 3 Y-data labels. The classifier used the global stellar parameters (the X-data) to make predictions about mode detectability (the Y-data). The stars with the largest number of detected modes could then be selected as the Red Giants for observation by TESS.

The Classifier successfully made predictions about the original 4 years of Kepler data; the algorithm had a weighted precision 0.98 across the 3 $P_{\rm det}$ labels. This confirms the proof of concept that Classifiers can be used as a way to select solar-like asteroseismic targets before future missions. Classification vastly reduces the computation time required to produce a target selection function, especially when large datasets are involved ($\geqslant 50,000$ stars).

Degrading the Red Giant data to make predictions for 1 year of TESS observations was also successful. The predicted mode detections scored a weighted precision of 0.86 across the 3 $P_{\rm det}$ labels. This illustrates that Classification is a valid target selection method for TESS targets in the Continuous Viewing Zone (CVZ, (Ricker et al. 2014)).

Using the Classifier on 27 days of TESS data returned detection predictions with a precision of 0.74. This is too low for the Classifier to be used to select solar-like oscillators for 27 days of TESS observation. The precision is lower when stars are observed for 27 days by TESS because the white noise level is too high and the length of observation is too short to make robust predictions of individual solar-like oscillations. It may be that individual solar-like oscillations cannot be detected in 27 days of TESS data. If this is the case, then the Classifier should not be expected to make robust detection predictions for these stars.

REFERENCES

Baglin A., et al., 2006. p. 3749, http://adsabs.harvard.edu/abs/2006cosp...36.3749B

Ballot J., Barban C., van't Veer-Menneret C., 2011, Astronomy and Astrophysics, 531, A124

Beck P. G., et al., 2011, Science, 332, 205

Bedding T. R., 2011, arXiv:1107.1723 [astro-ph]

Campante T. L., et al., 2016, preprint, 1608, arXiv:1608.01138

Chaplin W. J., Miglio A., 2013, Annual Review of Astronomy and Astrophysics, 51, 353

Chaplin W. J., et al., 2011, The Astrophysical Journal, 732, 54

Davies G. R., Miglio A., 2016, arXiv:1601.02802 [astro-ph]

Eisenstein D. J., et al., 2011, The Astronomical Journal, 142, 72

Elorrieta F., et al., 2016, Astronomy and Astrophysics, 595, A82

Elsworth Y., Hekker S., Basu S., R. Davies G., 2017, Mon Not R Astron Soc, 466, 3344

Gaia Collaboration et al., 2016, Astronomy and Astrophysics, 595, A2

Hon M., Stello D., Yu J., 2017, Monthly Notices of the Royal Astronomical Society, 469, 4578

Hon M., Stello D., Yu J., 2018, Monthly Notices of the Royal Astronomical Society

Howell S. B., et al., 2014, Publications of the Astronomical Society of the Pacific, 126, 398

Kjeldsen H., Bedding T. R., 1995, Astronomy and Astrophysics, 293, 87

Kremer J., Stensbo-Smidt K., Gieseke F., Steenstrup Pedersen K., Igel C., 2017, preprint, 1704, arXiv:1704.04650

Lund M. N., et al., 2017, The Astrophysical Journal, 835, 172

Mathur S., et al., 2011, The Astrophysical Journal, 741, 119

Mosser B., et al., 2012, A&A, 537, A30

Ness M., Hogg D. W., Rix H.-W., Ho A. Y. Q., Zasowski G., 2015, The Astrophysical Journal, 808, 16

Nun I., Pichara K., Protopapas P., Kim D.-W., 2014, The Astrophysical Journal, 793, 23

Pinsonneault M. H., et al., 2014, The Astrophysical Journal Supplement Series, 215, 19

Placek B., Knuth K. H., Angerhausen D., 2016, Publications of the Astronomical Society of the Pacific, 128, 074503

Rauer H., et al., 2014, Experimental Astronomy, 38, 249

Ricker G. R., et al., 2014, Journal of Astronomical Telescopes, Instruments, and Systems, 1, 014003

Steinhardt C. L., Jermyn A. S., 2018, Publications of the Astronomical Society of the Pacific, 130, 023001

Valenzuela L., Pichara K., 2018, Monthly Notices of the Royal Astronomical Society, 474, 3259

Wegner P., 1960, Commun. ACM, 3, 322

Wu Y., et al., 2017, preprint, 1712, arXiv:1712.09779

This paper has been typeset from a TEX/LATEX file prepared by the author.