



# Periodic stellar variability from almost a million NGTS light curves

Joshua T. Briegal<sup>1</sup>★, Edward Gillen<sup>1,2</sup>†, Didier Queloz<sup>1</sup>, Simon Hodgkin<sup>3</sup>, Jack S. Acton<sup>4</sup>, David R. Anderson<sup>5,6</sup>, David J. Armstrong<sup>5,6</sup>, Matthew P. Battley<sup>5,6</sup>, Daniel Bayliss<sup>5,6</sup>, Matthew R. Burleigh<sup>4</sup>, Edward M. Bryant<sup>5,6</sup>, Sarah L. Casewell<sup>4</sup>, Jean C. Costes<sup>7</sup>, Philipp Eigmüller<sup>8</sup>, Samuel Gill<sup>5,6</sup>, Michael R. Goad<sup>4</sup>, Maximilian N. Günther<sup>9,10</sup>‡, Beth A. Henderson<sup>4</sup>, James A. G. Jackman<sup>5,6,11</sup>, James S. Jenkins<sup>12,13</sup>, Lars T. Kreutzer<sup>1,14,15</sup>, Maximiliano Moyano<sup>16</sup>, Monika Lendl<sup>17</sup>, Gareth D. Smith<sup>1</sup>, Rosanna H. Tilbrook<sup>4</sup>, Christopher A. Watson<sup>7</sup>, Richard G. West<sup>5,6</sup> and Peter J. Wheatley<sup>5,6</sup>

*Affiliations are listed at the end of the paper*

Accepted 2022 March 29. Received 2022 March 29; in original form 2021 July 19

## ABSTRACT

We analyse 829 481 stars from the Next Generation Transit Survey (NGTS) to extract variability periods. We utilize a generalization of the autocorrelation function (the G-ACF), which applies to irregularly sampled time series data. We extract variability periods for 16 880 stars from late-A through to mid-M spectral types and periods between  $\sim 0.1$  and 130 d with no assumed variability model. We find variable signals associated with a number of astrophysical phenomena, including stellar rotation, pulsations, and multiple-star systems. The extracted variability periods are compared with stellar parameters taken from Gaia DR2, which allows us to identify distinct regions of variability in the Hertzsprung–Russell Diagram. We explore a sample of rotational main-sequence objects in period–colour space, in which we observe a dearth of rotation periods between 15 and 25 d. This ‘bi-modality’ was previously only seen in space-based data. We demonstrate that stars in sub-samples above and below the period gap appear to arise from a stellar population not significantly contaminated by excess multiple systems. We also observe a small population of long-period variable M-dwarfs, which highlight a departure from the predictions made by rotational evolution models fitted to solar-type main-sequence objects. The NGTS data spans a period and spectral type range that links previous rotation studies such as those using data from *Kepler*, *K2*, and *MEarth*.

**Key words:** methods: data-analysis – techniques: photometric – stars: activity – Hertzsprung–Russell and colour–magnitude diagrams – stars: rotation – stars: variables: general.

## 1 INTRODUCTION

Many of a star’s physical properties can be inferred from its brightness variations over time. This variability can arise from a number of mechanisms, either intrinsic to the star through changing physical properties of the star and its photosphere, or through external factors such as orbiting bodies and discs. The rotation of magnetically active stars will also cause visible brightness changes. Eyer & Mowlavi (2008) categorize a large number of distinct variability types which range in period from milliseconds to centuries and in amplitude from a few parts-per-million (ppm) to orders of magnitude in the most explosive forms of variability.

Stellar rotation can be measured through photometric observation, as magnetic surface activity such as spots and plages cause photometric brightness fluctuations over time that is modulated by

both the rotation of active regions across the star, as well as active region evolution. Constraining stellar rotation rates is important, as this provides insight into the angular momentum of the star. Skumanich (1972) first hypothesized that a star’s rotation rate could be age dependant, obtaining the empirical relation between rotation period  $P_{\text{rot}}$  and age  $t$ :  $P_{\text{rot}} \propto t^{0.5}$ . Knowing a star’s age is fundamental to fully understanding its evolutionary state, and so being able to infer this property from an observable quantity such as rotation would greatly improve our understanding of stars in the local neighbourhood. In Barnes (2003), a semi-empirical model for deriving stellar ages from colour and rotation period was suggested, and the term ‘gyrochronology’ was coined. This model was subject to further improvements in Barnes (2007), a model which is commonly still used to age Solar-type and late-type main-sequence stars. These models work especially well for stars older than the age of the Hyades cluster, by which time we expect the initial angular momentum of stars to have little effect on the rotation period, and the angular momentum evolution to follow a Skumanich law (Kawaler 1988). For low-mass stars, it is widely accepted that late-time angular momentum loss will be governed by magnetized stellar winds which

\* E-mail: [jtb34@cam.ac.uk](mailto:jtb34@cam.ac.uk)

† Winton Fellow.

‡ Juan Carlos Torres Fellow.

§ ESA Research Fellow.

depend on magnetic field topology and stellar mass (Booth et al. 2017). For young stars ( $< 10$  Myr), angular momentum evolution may be dependant on magnetic coupling between the star and disc. Studies of pre-main-sequence (PMS) stars in young clusters such as T-Tauri stars in the Taurus–Auriga molecular cloud (Hartmann & Stauffer 1989) or in NGC 2264 (Sousa et al. 2016) show high levels of short period ( $< 10$  d) photometric variability, but objects with circumstellar discs present appear to rotate slower than those without, highlighting the effect of star-disc coupling on angular momentum evolution.

Understanding a star’s activity is important for exoplanet surveys. Not only is stellar activity a large source of noise in both transit and RV surveys (e.g. Queloz et al. 2001; Haywood et al. 2014; Dumusque et al. 2017), but stellar activity may also influence the potential habitability of orbiting planets. Stars that rotate rapidly, for example, often display higher flare rates than their more slowly rotating cousins, and these flares can be important for potential exoplanet habitability. On the one hand, flares can erode exoplanet atmospheres and modify their chemistry (e.g. Segura et al. 2010; Seager 2013; Tilley et al. 2019), while on the other, they can help initiate prebiotic chemistry and seed the building blocks of life (Ranjan, Wordsworth & Sasselo 2017; Rimmer et al. 2018), which may be especially important for M-dwarf systems.

The angular momentum of a host star and its planets are intrinsically linked. Gallet et al. (2018) demonstrate that tidal interactions between a host star and a close-in planet can affect the surface rotation of the star. They observe a deviation in rotation period from the expected magnetic braking law during the early MS phase of low-mass stars in the Pleiades cluster, which the authors attribute to planetary engulfment events. Conversely, angular momentum transfer through tidal interactions must be considered in the context of stellar spin-down through magnetic braking. The analysis by Damiani & Lanza (2015) demonstrates that to accurately model tidal dissipation efficiency and orbital migration the stellar angular momentum loss through magnetized stellar winds must be accounted for.

Large-scale photometric variability studies have recently allowed for data-driven analysis of stellar variability in extremely large samples. Stellar clusters allow studies of groups of stars with similar formation epochs and evolutionary conditions, so historically have been targeted by systematic surveys. These observations have come from ground-based surveys such as Monitor (Hodgkin et al. 2006; Aigrain et al. 2007) with observations of NGC 2516 (Irwin et al. 2007), SuperWASP (Pollacco et al. 2006) with observations of the Coma Berenices open cluster (Collier Cameron et al. 2009) and HATNet (Bakos et al. 2004) with observations of FGK Pleiades stars (Hartman et al. 2010). Recently, NGTS (Wheatley et al. 2018) observed the  $\sim 115$ -Myr old cluster Blanco 1, and a study by Gillen et al. (2020a) demonstrated a well-defined single-star rotation sequence which was also observed by KELT (Pepper et al. 2012) and studied in Cargile et al. (2014). In both of these works, a similar sequence was observed for stars in the similarly aged Pleiades, indicating angular momentum evolution of mid-F to mid-K stars follows a well-defined pathway which is strongly imprinted by  $\sim 100$  Myr.

As part of the transient search conducted by the All-Sky Automated Survey for Supernovae (ASAS-SN; Shappee et al. 2014), a catalogue of observed variable stars has been compiled. This catalogue contains variability periods and classifications for 687 695 objects<sup>1</sup> taken from a series of publications entitled ‘The ASAS-SN catalogue of variable stars’ (e.g. Jayasinghe et al. 2018, 2021). Such

catalogues are not focused on specific clusters or stellar types, but provide a broad view of different forms of stellar variability.

Space missions have allowed wide-field photometric variability surveys of stars with high precision and continuous time coverage. *CoRoT* (Auvergne et al. 2009), *Kepler* (Borucki et al. 2010), the extended *Kepler* mission (*K2*; Howell et al. 2014) and TESS (Ricker et al. 2014) have provided a wealth of stellar photometric data, which in turn has been the subject of extensive rotation studies (Basri et al. 2011; Ciardi et al. 2011; Affer et al. 2012; McQuillan, Mazeh & Aigrain 2014; Davenport & Covey 2018; Canto Martins et al. 2020; Gordon et al. 2021), revealing large scale trends in stellar variability periods. In particular, studies by McQuillan et al. (2014) and Davenport & Covey (2018) demonstrated a distinct bimodal structure in the rotation periods of main-sequence stars with respect to colour. Gordon et al. (2021) followed up these studies with analysis of data from the *K2* mission, hypothesising the bimodal structure arises from a broken spin-down law, caused by an internal angular momentum transfer between the core and convective envelope. Further details of this model are discussed in Section 5.

The Next Generation Transit Survey (NGTS; Wheatley et al. 2018) is a ground-based wide-field photometric survey achieving routinely milli-magnitude range photometric precision with 12-s sampling cadence and long observation baselines, typically 250 nights of data per target field. The primary science goal of NGTS is to extend the wide-field ground-based detection of transiting exoplanets to at least the Neptune size range. Such high precision photometry lends itself well to ancillary stellar physics such as cluster rotation analysis (Gillen et al. 2020a) or stellar-flare detection and characterization (Jackman et al. 2019). Ground-based observation adds extra layers of difficulty in variability studies when compared to space telescope data, as we must consider irregular sampling and telluric effects. In this study, we employ a generalization of the autocorrelation function (the G-ACF) which applies to this irregular sampling. We elect to use an autocorrelation function to extract variability as this has proven to be successful for extracting stellar variability by McQuillan, Aigrain & Mazeh (2013), McQuillan et al. (2014) & Angus et al. (2018) and for NGTS data in Gillen et al. (2020a). An Autocorrelation Function (ACF) also allows better detection of pseudo-periodic and phase-shifting variability often seen in young, active stars in comparison to more rigid variability extraction techniques such as Lomb–Scargle periodograms (Lomb 1976; Scargle 1982).

The paper is organized as follows. In Section 2, we discuss the NGTS and the data used, and in Section 3 we outline the methods used in this study to extract rotation periods. Our results are summarized in Section 4, with a discussion of these results in Section 5 and a brief summary of our findings in Section 6.

## 2 THE NGTS

NGTS is an array of twelve 20-cm telescopes based at ESO’s Paranal Observatory in Chile. Each telescope is coupled to a  $2K \times 2K$  e2V deep-depleted Andor iKon-L CCD camera with  $13.5\text{-}\mu\text{m}$  pixels, corresponding to an on-sky size of  $4.97$  arcsec. The data for this study were taken with the array in *survey mode*, where each telescope observes a sequence of survey fields (generally 2 per night), each field having an on-sky size of  $\sim 8\text{ deg}^2$ . These fields are spaced such as one field rises above  $30^\circ$  the previous field sets below  $30^\circ$ . This typically results in approximately 500 h coverage per field spread over 250 nights.

Fields are selected based on the density of stars, the proportion of dwarf stars, the ecliptic latitude and proximity to any bright or extended objects. Fields are typically selected with  $\leq 15\,000$  stars

<sup>1</sup> Accessed on 09/11/2021

brighter than an *I* band magnitude of 16, of which  $\geq 70$  per cent are dwarf stars. These fields will be more than  $20^\circ$  from the Galactic Plane. Fields within  $30^\circ$  of the ecliptic plane are also avoided due to the Moon affecting readings during about three nights per month.

In this study, we use the final light curves, associated metadata and quality flags of the standard NGTS pipeline as described in Wheatley et al. (2018). The data for each field is reduced and photometric measurements are made on source apertures to assemble a light curve per target star. As a part of the pipeline, these light curves are passed through a custom implementation of the SysRem algorithm (Tamuz, Mazeh & Zucker 2005) which removes signals common to multiple stars arising from various sources including the instrument, reduction software and the atmosphere.

## 2.1 Light-curve extraction

Photometric light curves are extracted for all sources detected within each NGTS field. Source detection is done using the IMCORE module in CASUTOOLS (Irwin et al. 2004) to generate an object list that is cross-matched against a number of catalogues. NGTS generates its own input source catalogue, as explained in section 5 of Wheatley et al. (2018). This source catalogue is cross-matched against a number of external catalogues. Cross-matching is done in position, as well as in colour and separation to limit spurious matches. This allows flagging of potential unresolved binaries in NGTS apertures.

A soft-edged circular aperture with a radius of 3 pixels (15 arcsec) is placed over each of these sources and placed in pixel coordinates using per-image astrometric solutions to account for radial distortion in addition to the autoguiding system. The sky background is estimated using bilinear interpolation of a grid of  $64 \times 64$  pixel regions for which the sky level is determined using a *k*-sigma clipped median. These raw light curves are then passed through the detrending pipeline described in Section 2.2 (section 6 of Wheatley et al. 2018).

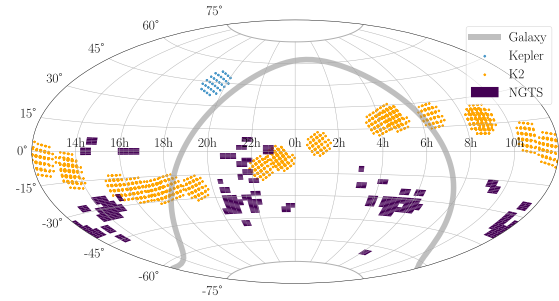
## 2.2 Systematics correction

In order to correct first-order offsets common to all light curves the detrending algorithm calculates a mean light curve for all objects to be used as an artificial ‘standard star’. This detrending algorithm is based on the SysRem algorithm first described in Tamuz et al. (2005) and adapted from the WASP project (Collier Cameron et al. 2006). Of note, this approach may not fully remove systematic trends which correlate with Moon phase and sidereal time. In particular, Moon-phased signals will show artefacts of imperfect background subtraction and any non-linearity in the detectors.

## 2.3 Data selection

The NGTS pipeline provides flags per image and per time stamp per object light curve which we use to pre-process light curves for variability analysis. These flags alert us to bad-quality data points as a result of pixel saturation, blooming spikes from nearby bright sources, cosmics, and other crossing events (including weather and laser guide stars) and any sky background changes. We removed any flagged data point from our light curves, and additionally checked if the majority of the light curve had been flagged ( $> 80$  per cent of data points). If this was the case, we removed the objects from processing.

We clipped our flux data to remove any points lying further than three median-absolute-deviations (MAD) from the median to remove any outliers not caught by the NGTS pipeline flags. This cut may remove some variability signals such as long-period eclipsing



**Figure 1.** An ICRS plot of the position of the 94 NGTS fields used in this study (solid dark blue squares). The *Kepler* and *K2* fields are included as blue and orange squares, respectively, as well as the Galactic Plane as a thick grey line.

binaries where the variability is a small fraction of the phase curve. Manual inspection of a single field confirmed this was not the case, however, this cannot be guaranteed for all fields processed. Finally, in order to speed up data processing, we binned our light curve into 20-min time bins. This reduced the number of data points to process per light curve from 200 000 to roughly 10 000. The G-ACF computation time scales as  $\mathcal{O}(n^2m)$  for *n* data points with *m* lag time steps, so reducing the number of timestamps in our light curve significantly improved processing time with a caveat that we will be unable to detect any periods below 40 min. For this study, that is focused on longer period variability, this limit is not of concern.

We removed six fields identified as containing large open cluster populations. This study will focus on stars in the field and this avoids contamination of large numbers of young variable stars in known open clusters. Removing these six fields left a total of 829 481 light curves to process. The positions of the 94 NGTS fields in RA and Dec. used in this study are shown in Fig. 1. In this figure, we plot the *Kepler* and *K2* field centre pointings, as well as the position of the Galactic Plane.

The 94 fields used in this study were observed for an average of 141 nights during different observation campaigns (lasting an average of 218 d) between September 2015 and November 2018. The shortest observational baseline for this data set was 84 d and the longest 272 d. In total, 73 of the 94 fields had observational baselines over 200 d. We detected periodic variability in light curves spanning  $8 < I_{\text{NGTS}} < 16$  mag with 50 per cent (90 per cent) of our detections being brighter than 13.5 (15.4) mag.

## 3 PERIOD DETECTION

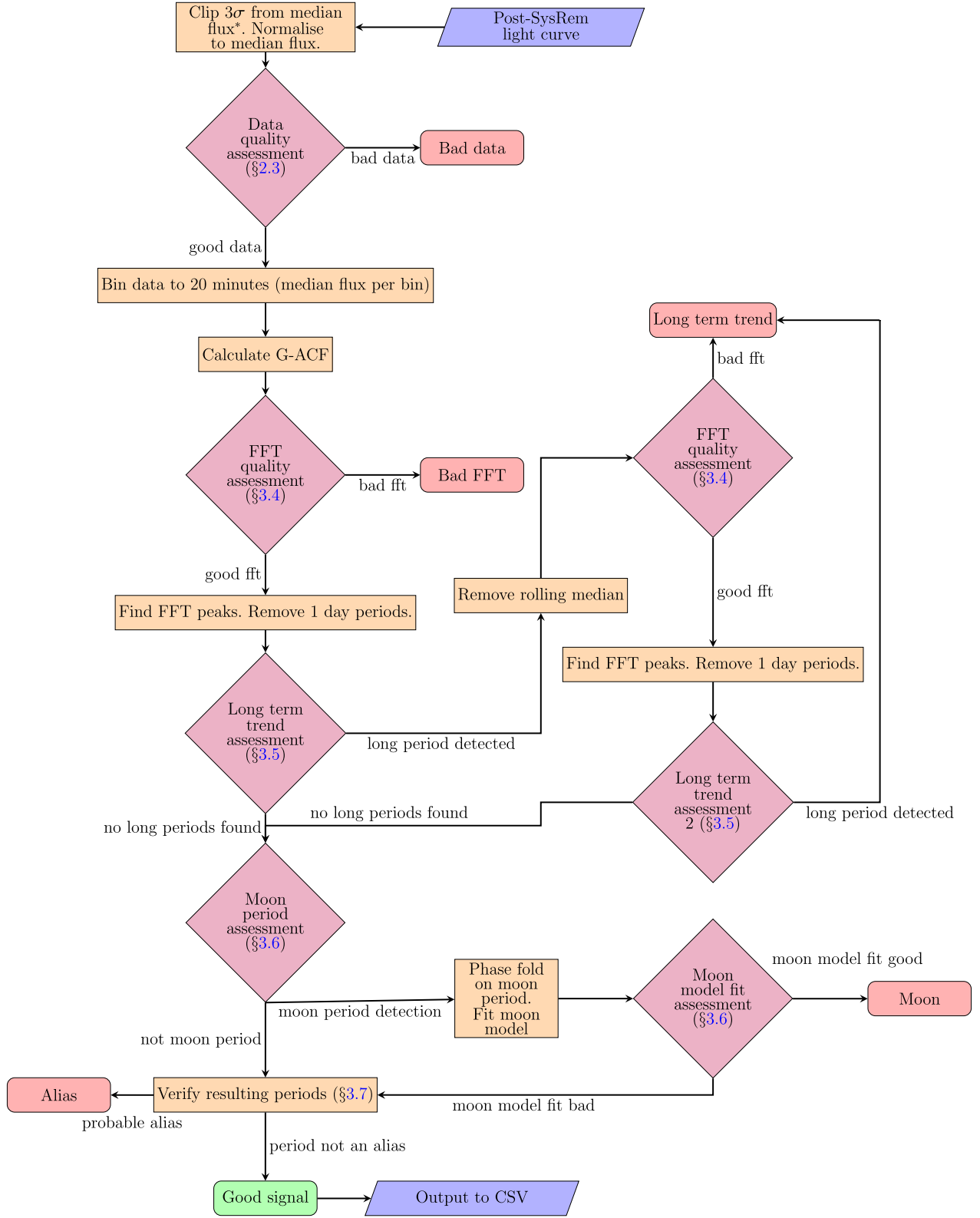
The period detection pipeline is outlined in the flowchart in Fig. 2. Further details of each step are given in the subsequent sections.

The open-source code of the periodicity detection pipeline can be found on GitHub.<sup>2</sup>

### 3.1 Generalised Autocorrelation Function (G-ACF)

The G-ACF is essentially an extended and generalized form of the standard ACF which can be applied to any time-series, regardless of sampling. A complete and detailed mathematical description of this algorithm is available in a separate paper by Kreutzer et al. [submitted]. This generalization is done by (a) generalizing the lag time *k* to a *generalized lag*  $\hat{k}$  which is a continuous variable within

<sup>2</sup>[https://github.com/joshbriegal/periodicity\\_detection](https://github.com/joshbriegal/periodicity_detection)



**Figure 2.** A schematic of the period detection pipeline, per NGTS light curve.  $\sigma$  refers to the median absolute deviation (MAD).



the range of our time series and (b) defining a *selection function*  $\hat{S}$  and a *weight function*  $\hat{W}$ .

Taking a standard definition of the ACF (e.g. Shumway & Stoffer 2006):

$$\rho(k) := \frac{1}{N} \sum_{i=1}^{i_{\max}-k} (X_i - \langle X_I \rangle) \times (X_{i+k} - \langle X_I \rangle), \quad (1)$$

where  $\langle X_I \rangle$  denotes the mean of the time-series values and the normalization  $N$  is the total sum of squares  $N := \sum_{i \in I} (X_i - \langle X_I \rangle)^2$ .

We can generalize this to

$$\begin{aligned} \hat{\rho}(\hat{k}; \hat{W}, \hat{S}) := & \frac{1}{N} \sum_{\substack{i \in I \\ t_i + \hat{k} \leq \max(T_I)}} \left[ (X(t_i) - \langle X_I \rangle) \right. \\ & \times (X(\hat{S}(t_i + \hat{k})) - \langle X_I \rangle) \\ & \left. \times \hat{W}(|\hat{S}(t_i + \hat{k}) - (t_i + \hat{k})|) \right], \end{aligned} \quad (2)$$

where  $N := \sum_{i \in I} (X_i - \langle X_I \rangle)^2$  denotes the total sum of squares and  $\langle X_I \rangle$  is the mean of the time-series values set.

### 3.2 The selection function, $\hat{S}$

The selection function  $\hat{S}$  provides a mapping between time labels within the original time series and the lagged time series at each lag time-step. The most natural choice of a selection function would be to select the point closest in time within the original time series for each point in the lagged time series. Further details of this selection function, including a cartoon outlining the method, are detailed in Kreutzer et al. (submitted).

### 3.3 The weight function, $\hat{W}$

The weight function,  $\hat{W}$ , should be a function  $\hat{W} : [0, \infty) \rightarrow [0, 1]$  with  $\hat{W}(0) \equiv 1$  to ensure that for a regular time series the G-ACF reduces back to the standard ACF. One such example is a rational weight function such as

$$\hat{W}(\delta t) = \frac{1}{1 + \alpha \delta t}, \quad \alpha > 0, \delta t \geq 0, \quad (3)$$

in which  $\delta t$  is the time difference between the time label in the original time series and the lagged time series mapped by the selection function  $\hat{S}$ . The parameter  $\alpha$  was taken as the median value of the time series (as a time difference from the first data point), as prescribed in Kreutzer et al. (submitted). We experimented with two different weight functions and elected to use the rational function as the final extracted periods were not dependant on this choice and this function is very simple. We used the minimum gap between time stamps as our lag resolution (time-steps in generalized lag,  $\hat{k}$ ); this was 20 min as we bin the data prior to analysis (Section 2.3).

### 3.4 Fast Fourier transform (FFT)

In order to extract a period from the G-ACF, we elected to use an FFT (Cooley, Lewis & Welch 1969). Extracting periods from an ACF can be done in a number of ways, most simply by selecting the first (or largest) peak in the ACF (e.g. as in McQuillan et al. 2014). This can lead to inaccuracies, in particular for weaker signals as this relies on the first peak being prominent in the ACF. We elected to use an extraction method that relies on the periodicity of the ACF, and the regular sampling of the G-ACF lends itself

to an FFT. Other more complex methods such as fitting a damped harmonic oscillator to the ACF have been used previously (Angus et al. 2018). This, in general, did not alter extracted periods enough to warrant the additional complexity for such an exploratory work. We also experimented with using fewer ACF peaks rather than the entire signal in order to refine the period, but again the additional complexity was deemed unnecessary for a large scale rotation study.

The FFT is a robust and well-documented method of extracting periodic signals. In this study we used the implementation in the `numpy.fft` package (Harris et al. 2020). We calculated the FFT with a padding factor of 32, to allow precise resolution of peaks in the Fourier transform. As phase information is lost in taking the ACF of the initial data, a real Fourier transform is sufficient.

To extract the most likely frequencies, we searched for peaks in the Fourier transform. A peak is defined as the central point in a contiguous sequence of five points which monotonically increases to the peak, followed by a monotonic decrease from the peak. Additionally, the amplitude of a peak must be greater than 20 per cent of the highest peak in the periodogram to be included. Here an automated cut was made – any Fourier transforms with more than six peaks were removed as noise. This threshold was selected based on a manual vetting process for one NGTS field (10 000 objects) which demonstrated that for these objects with ‘noisy’ Fourier transforms less than 1 per cent had genuine periodic signals. Removing these objects entirely greatly reduced the number of false positives extracted without removing many ‘real’ signals. In total, 63 per cent of processed objects were flagged as having no significant periodicity based on this FFT check.

### 3.5 Long-term trend assessment

A time baseline of  $\sim 250$  d allows for the extraction of periodic signals up to  $\sim 125$  d long. Signals longer than this may be present in the data, however, observing one or fewer complete variability cycles cannot definitively characterize a periodic signal. This variability may not be periodic, but rather a long term trend in the data arising from instrumental or telluric changes over these time-scales. These objects may still contain interesting periodic variability at a shorter time-scale, so by detecting and removing a long-term trend we can more accurately calculate the period and amplitude of this variability.

If the most significant peak in the FFT (see Section 3.4) was at a period greater than half the length of the signal baseline it was flagged as a long term trend. When this occurred we computed a high-pass filter for the signal by calculating the median flux at each time step in a rolling window which is 10 per cent of the time extent of the light curve. This captures any long term behaviour without removing any shorter period variability. We divided this median filter from our signal and re-ran the cleaned light curve back through the signal detection pipeline. If no signal of interest was detected at this stage (either we found noise or residuals of our median filter), the object was flagged as having a long-term trend and removed from processing.

### 3.6 Moon signal assessment

During initial testing of the period extraction algorithm, it was noted that a large number of periods between 27 and 30 d were identified by the period search algorithm. Upon closer inspection, these periods had very similar phases and could be split into two groups of signal shapes. The two signal shapes, when phase folded on a new Moon epoch, appeared as a slight increase or decrease in flux at 0.5 phase, i.e. full Moon. This was accompanied by an increase in scatter in the

flux measurements at full Moon. Examples of contaminated signals are shown in Fig. A1.

We fitted a model to these Moon correlated noise signals (‘Moon signals’) and flagged and removed any objects which fitted the expected trend. A detailed description of the model and removal process is given in Appendix A.

### 3.7 Alias checks

As we are using an FFT to extract periodicity from our G-ACF, we are prone to aliasing. Aliasing is a well known and well-described problem in signal processing, and if the true frequency of the signal and the sampling frequency are known it is trivial to calculate the frequency of aliases as

$$\nu_{\text{alias}} = \nu_{\text{true}} \pm n \cdot \nu_{\text{sampling}}, \quad (4)$$

where  $n$  is an integer. We define period as the inverse of frequency, i.e.  $P = \frac{1}{\nu}$ . In the case of ground-based observation, the most common sampling period will be 1 d. In addition, although the background correction should remove this, there will remain residuals of the brightness trend expected throughout the night’s observation. Although the sampling of the G-ACF is regular, the sampling of the inputted light curve will affect the shape of the G-ACF. We thus expect peaks in the FFT associated with 1-d systematic signals, as well as the true signal aliased with the 1-d sampling.

For each light curve, we first removed any periods arising from the 1-d sampling. We removed periods within 5 per cent of 1 d, as well as within 5 per cent of integer multiples of 1 d in period and integer multiples of 1/day in frequency. We then assessed whether groups of periods were aliases of one another with respect to common sampling periods using basic graph theory. We construct a graph of frequencies connected by the standard alias formula in equation (4), using sampling periods of one day, 365.256 36 d (one year), 27.321 58 d (Lunar sidereal period) and 29.530 49 d (Lunar synodic period). Each vertex in the graph represents an FFT peak frequency, with connections (edges) made if two frequencies can be related to one another through equation (4) given one of our sampling frequencies. Note we considered aliases arising from both the synodic and sidereal Lunar period, however, given the 5 per cent tolerance used for assessing similarity, these two-sampling frequencies connected the same frequencies in the majority of examples.

For each connected sub-graph (i.e. a group of frequencies connected by the same sampling aliases), we determined the frequency for which the phase folded light curve had the lowest spread in flux and took this to be the correct period. We calculated the 5th–95 percentile spread in flux within bins of 0.05 width in phase and then calculated the average of these values weighted by the number of points within each flux bin. In addition to the FFT peak periods, we also checked the RMS of twice and half the periods, as in some cases we found twice the FFT peak period was the correct period. This was assessed by-eye initially, and appeared to be much more common for short period objects due to aliasing from the 1-d sampling. This same approach was taken by McQuillan et al. (2013), however, we elected to automate the process rather than by-eye confirmation.

### 3.8 Further signal validation

Due to the ground-based nature of NGTS, some fields were not continuously observed for the entirety of the field time-baseline. As a result of bad weather and technical downtime, there were gaps in observations lasting several weeks for a number of the fields used in this study. In these cases, it is no longer correct to use the entire

time baseline as a cut-off for robust periods. Instead, we elected to find the longest period of continuous observation within these fields and remove any periods greater than half this time length. We define a period of continuous observation as a period in which there are no observation time gaps of greater than 20 per cent of the entire field baseline. For our 250-night observation baseline, this equates to gaps of 50 d or longer. This removed 907 detected periodic signals from 11 different fields, and manual inspection of the removed signals confirmed that many of the removed detections were systematic periods arising from the long sampling gaps, rather than astrophysical variability.

Additionally, a number of detected periodic signals with unphysically large amplitudes were detected. On inspection it appears these signals were incorrectly processed by the NGTS pipeline, resulting in non-physical flux values. In our final sample, we elected to remove any signals with a relative amplitude  $> 1.0$ . This removed 58 signals, and manual inspection of the removed signals confirmed the majority of signals removed were non-physical; especially for the largest amplitude signals. The cut-off was chosen empirically based on the signal amplitude distribution of our sample.

Our initial search resulted in 17 845 periodic detections. Removing 907 long-term trends left 16 938 detections. Finally, removing 58 unphysically large amplitude signals resulted in 16 880 detections.

### 3.9 Cross-matching with Gaia DR2 & TICv8

In order to assess our variability period sample within a meaningful scientific context, we elected to use Gaia Data Release 2 (DR2, Gaia Collaboration 2018a) for cross-matching and to identify the nature of corresponding objects and their stellar parameters. The NGTS database contains cross-matching information with many external catalogues, including Gaia DR2. Detail on how the cross-matches are found is given in section 5 of Wheatley et al. (2018) and briefly in Section 2.1 of this paper.

As an extension of the Gaia DR2 data, the most recent Tess Input Catalogue (TICv8, Stassun et al. 2019) contains Gaia DR2 data relevant to this study plus additional calculated values and cross-match data. These include more accurate calculated distances from Bailer-Jones et al. (2018) and reddening values which have been used to calculate absolute magnitudes.

More recently, the Gaia Early-DR3 (Gaia Collaboration 2021) contains improved precision on the astrometric fits to many objects from Gaia DR2, however, as we are using many derived parameters from external catalogues we elected to continue to use the DR2 parameters throughout this study.

### 3.10 Extinction correction

In the final data products, we assess variability in the context of the colour–magnitude diagram (CMD) which requires the calculation of absolute magnitudes. In order to be as accurate as possible, we combined Gaia  $G$  magnitudes ( $G$ ) with distance estimates and accounted for extinction. We used the per-object reddening values from TICv8, multiplied by a total-to-selective extinction ratio of 2.72 to account for the Gaia  $G$ -band extinction ( $A_G$ ). Further details on how the reddening values and the total-to-selective extinction ratio were calculated can be found in section 2.3.3 of Stassun et al. (2019). Our final value for absolute magnitude was calculated using the formula:

$$M_G = G - 5 \log_{10}(\text{distance}) + 5 - A_G. \quad (5)$$

**Table 1.** A sample of variability periods, amplitudes, positions and catalogue cross-match identifiers in the NGTS data set. A number of catalogue cross-match columns have been excluded for publication clarity. The full table will be available at CDS via anonymous ftp to [cdsarc.u-strasbg.fr](https://cdsarc.u-strasbg.fr) (130.79.128.5) or via <https://cdsarc.unistra.fr/viz-bin/cat/J/MNRAS>, or as supplementary material.

NGTS ID	NGTS RA	NGTS Dec	NGTS Mag	Period	Amplitude	Gaia DR2 ID	TICv8 ID
NG0613-3633.231	94.88721	−35.20762	14.77188	117.30427	0.07218	2885392740653834368	124854845
NG0613-3633.234	91.91176	−35.20084	15.86231	128.42220	0.18731	2885953869540806656	201389809
NG0613-3633.235	94.93884	−35.20675	12.91320	117.53460	0.04857	2885392878092780544	124854842
NG0613-3633.262	94.95269	−35.20598	14.51873	109.77205	0.11461	2885392225257749760	124854841
NG0613-3633.481	93.77213	−35.22205	13.69049	0.29365	0.13175	2885521658392050944	124689517
NG0613-3633.598	93.31896	−35.22787	11.48757	92.88398	0.00860	2885530999944081792	201530507
NG0613-3633.773	95.01907	−35.23832	12.55225	110.36974	0.07016	2885380160692365824	124855736
NG0613-3633.1101	95.06110	−35.25333	15.16207	128.42220	0.22943	2885381333220681216	124855723
NG0613-3633.1181	95.06864	−35.25766	13.60488	100.74969	0.08537	2885380577306436736	124855720
NG0613-3633.1479	95.12023	−35.27187	14.86311	100.46635	0.25487	2885380439867479040	124922604

**Table 2.** Variability periods, amplitudes, positions, and catalogue cross-match identifiers for all variable objects in the NGTS data set (table format).

Column	Format	Units	Label	Description
1	A18	–	NGTS_ID	NGTS source designation
2	F9.5	deg	NGTS_RA	Source right ascension (J2000)
3	F9.5	deg	NGTS_DEC	Source declination (J2000)
4	F8.5	mag	NGTS_MAG	NGTS I-band magnitude
5	F9.5	days	PERIOD	Extracted variability period
6	F7.5	–	AMPLITUDE	5–95 percentile relative flux
7	I19	–	GAIA_DR1_ID	Cross-matched Gaia DR1 identifier
8	I19	–	GAIA_DR2_ID	Cross-matched Gaia DR2 identifier
9	I10	–	TIC_ID	Cross-matched Tess Input Catalogue (v8) identifier
10	A16	–	TWOMASS_ID	Cross-matched 2MASS identifier
11	A19	–	WISE_ID	Cross-matched WISE identifier
12	A10	–	UCAC4_ID	Cross-matched UCAC4 identifier

## 4 RESULTS

Using the G-ACF period extraction pipeline, we derived variability periods for 16 880 stars observed with NGTS. A subset of these results is shown in Table 1, along with positions and cross-match data. The format of the results table is shown in Table 2.

### 4.1 Periodicity in colour–magnitude space

Fig. 3 shows our variability sample in colour–magnitude space, commonly known as a Hertzsprung–Russell (HR) Diagram or a CMD. Table 3 details the breakdown of outputs from the pipeline. Once cross-matched with TICv8, we were left with a total of 16 880 variable light curves from the initial sample of 829 481 light curves. This gives a final detection percentage of 2.04 per cent. The detection percentage varies in colour–magnitude space as shown in Fig. 3(a), highlighting potential regions of increased variability or increased sensitivity of NGTS and the signal detection pipeline.

All conversions between  $T_{\text{eff}}$ ,  $G_{BP} - G_{RP}$ , and  $G - G_{RP}$  in the following sections are calculated using relations defined in the ‘Modern Mean Dwarf Stellar Colour and Effective Temperature Sequence’ (Pecaut & Mamajek 2013),<sup>3</sup> interpolated using a univariate cubic spline. The isochrones in the HR diagrams are taken from PARSEC v1.2S (Bressan et al. 2012). We elected to use these isochrones as they have been proven to fit the Gaia DR2 main sequence

well in Gaia Collaboration (2018b). We produce isochrones using PARSEC v1.2S, selecting the Gaia DR2 passbands from Evans et al. (2018).<sup>4</sup> The isochrone at 1 Gyr gives a good indication of where the main sequence lies, with the earlier age isochrone at 10 Myr indicating locations on the HR diagram of potentially younger stellar populations. We note, as shown in Gillen et al. (2020b), that the PARSEC v1.2 models appear to be less reliable at PMS ages, but should be sufficient for their indicative use in this study.

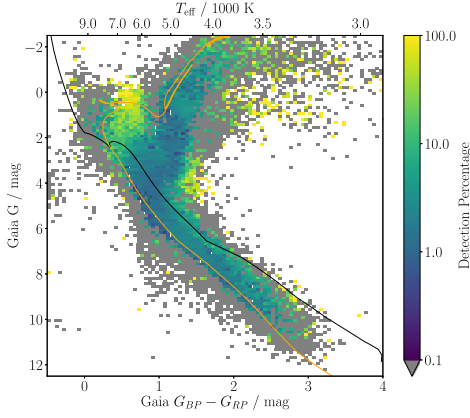
Fig. 3(a) highlights regions of interest in terms of detection percentage. Additionally, Fig. 3(b) shows the number of detections in each bin. Where detection percentage approaches 100 per cent this is often indicative of a single variable object falling in this colour–magnitude bin. As in Gaia Collaboration (2019), we identify distinct regions of variability within the HR diagram and suggest the types of variable objects which may lie at each location.

The region at the top of the main sequence ( $G_{BP} - G_{RP} \sim 0.4$ ,  $G \sim 1.0$ ) reveals a high proportion of variable objects. We also see a region of increased variability at the ‘elbow’ of the main sequence and the Red-Giant Branch (RGB) ( $G_{BP} - G_{RP} \sim 1.5$ ,  $G \sim 4$ ). These objects may be young, massive objects with high levels of activity.

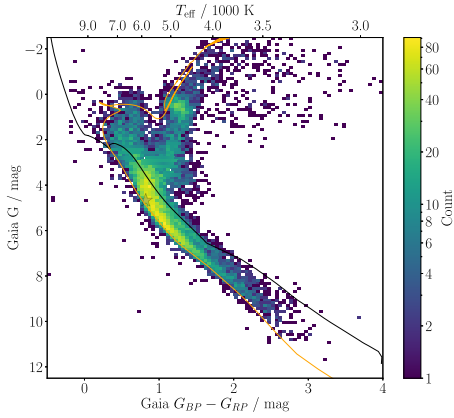
In Fig. 3(c), we plot the median period in each colour–magnitude bin. Of particular interest, we see distinct regions of different variability periods on the HR diagram. There is a region of short median period at the top of the main sequence ( $G_{BP} - G_{RP} \sim 0.4$ ,  $G \sim 1.0$ ). Typical spot-driven photometric modulation will not be present on these hotter, radiative stars. The majority of variability seen in this

<sup>3</sup>A more recent version of the table including Gaia DR2 colours is maintained at [http://www.pas.rochester.edu/~emamajek/EEM\\_dwarf\\_UBVIJHK\\_colors\\_Teff.txt](http://www.pas.rochester.edu/~emamajek/EEM_dwarf_UBVIJHK_colors_Teff.txt)

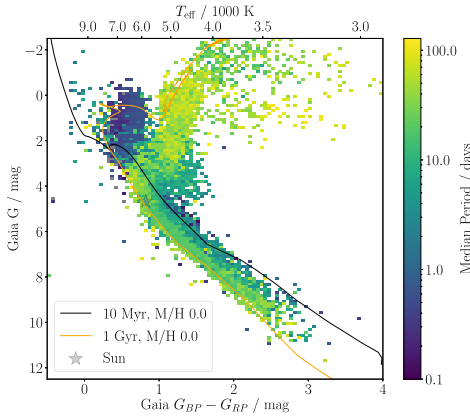
<sup>4</sup>using the CMD 3.4 input form at <http://stev.oapd.inaf.it/cgi-bin/cmd>



(a) The empirical detection percentage per bin. This is defined as the ratio of the number of detected periodic signals to all observed objects per bin. 0 detections within bins are coloured grey.



(b) The number of objects with detected variability within each colour-magnitude bin.

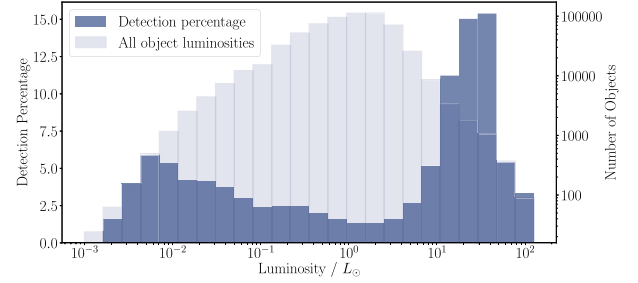


(c) The median variability period within each colour-magnitude bin.

**Figure 3.** Binned colour-magnitude (HR) diagrams of the NGTS variability sample. PARSEC v1.2 (Bressan et al. 2012). Solar metallicity isochrones of ages 10 Myr and 1 Gyr are included as solid black and orange lines, respectively.

**Table 3.** A table of the output states of the 829 481 NGTS objects analysed by the signal detection pipeline. Note a further 907 objects were removed due to large observation gaps in a number of fields, and an additional 58 with spuriously large amplitudes resulting in a final total of 16 880 variability periods (see Section 3.8).

Output state	Count	% of total	% of detections
Bad data	43 358	5.227	–
Noisy FFT	528 105	63.667	–
Moon	175 565	21.166	67.043
Alias	57	0.007	0.022
Long-term trend	64 551	7.782	25.018
Periodic signal	17 845	2.151	6.916



**Figure 4.** Histogram of the empirical detection percentage (left y-axis) for all sources against luminosity, as well as the luminosity distribution for all observations (right y-axis).

region likely arises from pulsations. There may also be a number of magnetic OBA or chemically peculiar Ap stars within this region. In these stars, photometric brightness fluctuations are seen as a result of fossil magnetic fields imprinting chemical abundance inhomogeneity on the stellar surface (David-Uraz et al. 2019; Sikora et al. 2019). These targets are prime candidates for future spectropolarimetric observations (e.g. Grunhut et al. 2017).

A large number of the longest period variability signals lie on the RGB ( $G_{BP} - G_{RP} \gtrsim 1.0$ ,  $G \lesssim 2.0$ ). These signals could indicate extremely slowly rotating large stars or other photometrically varying sources such as giant star pulsations.

We also see a clear trend of increasing period as we move perpendicular down towards the main sequence along the Hayashi tracks (Hayashi 1961). There are potentially a number of effects at play here:

- 1) We would expect a population of equal mass binary stars with short rotation periods to lie 0.75 in absolute magnitude above the main sequence, contributing to the shorter median period in this range.
- 2) We would also expect a population of young stars to lie in this region of colour-magnitude space. In particular, we see short period objects which lie between the 10 Myr and 1 Gyr isochrones.

In this region of the HR diagram potentially lie PMS Young Stellar Objects (YSO) such as T-Tauri stars with protostellar debris discs, which we expect to have shorter rotation periods than main-sequence stars of the same mass (colour). The median period observed for the bulk of main-sequence objects is 20 to 30 d, as expected.

We plot detection percentage versus luminosity in Fig. 4. Luminosity values are taken from TICv8 (Stassun et al. 2019), calculated using equation (6).

$$\frac{L}{L_{\odot}} = \left( \frac{R}{R_{\odot}} \right)^2 \cdot \left( \frac{T_{\text{eff}}}{5772} \right)^4. \quad (6)$$



We use the radii values provided by TICv8. These radii values are either taken from pre-existing dwarf catalogue values (from Muirhead et al. 2018), or when these are not available (as is the case for a large majority of the NGTS sources) they are calculated from distance, bolometric corrections, G magnitude and a preferred temperature. Full details of this calculation are given in Stassun et al. (2018).  $T_{\text{eff}}$  values come from spectroscopic catalogues where available, otherwise they are derived from the de-reddened  $G_{BP} - G_{RP}$  colour.

As expected, we recover a much higher fraction of variable signals from more luminous stars, with up to 15 per cent of the most luminous objects in our sample having detectable variability signals. These objects will correspond to luminous giant stars, where we would expect large-amplitude variability arising from pulsations. The lowest number of variable objects coincides with the peak in the number of objects (at  $1.5\text{--}2.5 L_{\odot}$ ), where we detect variability in  $< 2$  per cent of objects. We also observe an increase in detection percentage for the faintest objects. Here, we should expect to be observing cooler dwarf stars and young stars which generally have higher levels of magnetic activity and could lead to increased detection of photometric variability. Additionally, close binaries may appear more luminous than single stars and from their position above the main sequence in the HR diagram (Fig. 3a) appear to have a higher detection percentage than equivalent single stars. Given the width of the luminosity bins used is larger than the expected luminosity increase from a single star to an equal luminosity binary (0.2 dex, a factor of  $\sim 1.6$  in luminosity), this will not have a large effect on the plotted distribution.

We assessed the distribution of detection percentage against on-sky RA and Dec. for our population. The distribution of detection percentage for field stars did not appear to have any obvious correlation with on-sky position.

## 4.2 Example variability signals

We show six examples of variability signals in Fig. 5. A table of stellar parameters for each object is included for reference. We selected the included objects to demonstrate a small selection of the variability we are able to extract from NGTS light curves. The stars are selected to have a range of spectral types, and demonstrate variability with different periods, amplitudes and signal shapes. In particular, using the object numbering as in Fig. 5 (1 to 6, top to bottom panels):

- 1) An extremely short period, semi-detached eclipsing binary. This object lies above the main sequence, as expected for a near-equal mass binary system.
- 2) A typical short period pulsation signal from an RR-Lyrae object.
- 3) A candidate YSO. Objects above the main sequence with periods of 1 to 10 d are excellent YSO candidates, suitable for follow-up infrared and spectroscopic observations.
- 4) An example of a variable red-giant star. These are stars such as Cepheids, semi-regular variables, slow irregular variables or small-amplitude red-giants.
- 5) A main-sequence late-G dwarf star, with small amplitude 20- to 30-d variability.
- 6) A long period M-dwarf.

Within the observed G-ACF signals we see artefacts arising from 1-d sampling aliases. These aliases are particularly relevant for signals of period  $< 1$  d, where it was necessary to perform the additional verification steps outlined in Section 3.7.

## 4.3 Cross-matching with previous catalogues

We cross-matched our NGTS variability periods with photometric variability catalogues in the literature. The ASAS-SN variability catalogue is a large catalogue of photometric variability. We took the latest available data, containing 687,695 variable stars from Jayasinghe et al. (2018) through to Jayasinghe et al. (2021).<sup>5</sup> We cross-matched our catalogue with the ASAS-SN catalogue, matching on TICv8 ID and Gaia DR2 ID. We found 2,439 matches with periods in both catalogues. A period–period comparison is shown in the left panel of Fig. 6. The majority (about  $1 \setminus 500$  stars) had similar periods from both catalogues. For approximately 750 stars, the periods differed by a factor of 2. This was most common for eclipsing binary targets in which the primary and secondary eclipses were of similar depths, and either the NGTS or ASAS-SN period was half the correct period. Periods with large discrepancies appear to be long term trends within the NGTS or the ASAS-SN data masking any shorter-term variability, or period aliasing resulting from the 1-d sampling seen in both surveys. The NGTS period extraction pipeline will not return periods close to 1 d or multiples thereof to reduce the number of systematic false positive detections. We see a number of periods in the ASAS-SN catalogue falling on exact fractions of 1 d, resulting in the ‘stripes’ of periods seen in the lower right of the Figure. We see structures within the period-period diagram resulting from objects for which the NGTS and ASAS-SN detections are aliases of one another with respect to 1-d sampling. Equation (4) can be used to calculate these connections and relations of the form

$$P_{\text{ASASSN}} = \frac{1}{P_{\text{sampling}} \pm \frac{1}{P_{\text{NGTS}}}} \quad (7)$$

are shown in Fig. 6. Three obvious sets of aliased periods exist which trace these relations, accounting for approximately 114 matches. We see two sets of related periods arising from 1-d sampling, with the same double phase folding for eclipsing binaries resulting in the set of periods approaching 2 d. There is also a small group of periods connected by aliases arising from 2-d sampling, however, the form of the relation is not shown in the Figure.

We were able to find three cross-matches with the MEarth rotation catalogue from Newton et al. (2018). Of these, NGTS was able to extract a short 0.4-d rotation period for an object which not present in the MEarth catalogue (NG1444-2807.12982). The two other objects (NG1214-3922.6732 & NG0458-3916.13434), NGTS detected a near 100-d period, similar to MEarth. The length of these periods would require extended observation from either survey to improve the accuracy as both surveys were only able to observe two to three complete variability cycles.

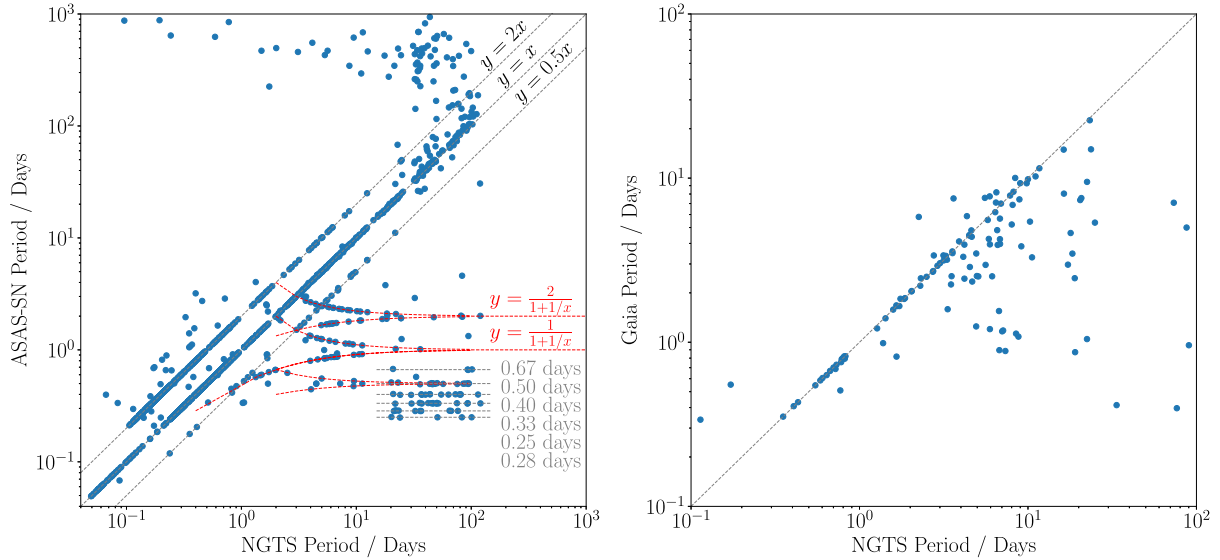
A variability study was conducted as part of the Gaia Data Release 2 (DR2, Gaia Collaboration 2018a), where photometric time series data were processed to detect and classify variable sources (as described in Holl et al. 2018). Photometric time series from Gaia are sparsely sampled and not optimized to detect photometric variability, so may produce an incorrect period. We cross-matched 126 objects against the rotation period database provided by the Gaia Collaboration on VizieR,<sup>6</sup> these period comparisons are shown in the right panel of Fig. 6. For 60 of the 126 periods that differed, we phase folded the NGTS data on both periods and manually inspected which phase fold appeared to be favourable. The NGTS period was favoured in the majority of cases through visual inspection. As expected for space-based data we do not see any aliasing artefacts in the Gaia

<sup>5</sup>The full catalogue is available at <https://asas-sn.osu.edu/variables>

<sup>6</sup><https://vizier.cds.unistra.fr/viz-bin/VizieR-3?-source=I/345/rm>



**Figure 5.** Example variable star signals across the HR diagram. From the left- to right-hand panels: a table of stellar parameters. The NGTS light curve, binned to 20 min. The G-ACF of the light curve. The light curve phase folded on the extracted period, each successive period is coloured according to a perceptually uniform sequential colourmap. The position of each star on the HR diagram is shown, the numbered labels 1–6 correspond to the stars top to bottom panels. Solar metallicity PARSEC isochrones of ages 10 Myr and 1 Gyr are included as solid black and orange line on the HR diagram, respectively.



**Figure 6.** NGTS variability periods from this study compared with ASAS-SN periods (left-hand panel) and Gaia (right-hand panel). Lines of equal period from both surveys are plotted in light grey, and for the ASAS-SN comparison lines showing periods differing due to incorrect phase folding by a factor of 2 shorter or longer are also plotted in light grey. The red dashed lines and associated equations indicate relations between periods arising from 1-d sampling. Light grey dotted horizontal lines in the left-hand figure and corresponding periods indicate where ASAS-SN has recovered periods corresponding to exact fractions of a day.

periods as in the cross-matching with ASAS-SN. This is a clear demonstration that the NGTS period recovery pipeline is well suited to deal with aliases arising from 1-d sampling

Finally, we cross-matched our sample with the variability catalogue from Canto Martins et al. (2020), which searched for rotation periods in 1000 TESS objects of interest. We found six objects in both catalogues by matching on TIC id. These come from three different results tables from Canto Martins et al. (2020): TIC 14165625 and 77951245 contain ‘unambiguous rotation periods’, TIC 100608026 and 1528696 contain ‘dubious rotation periods’ and TIC 150151262 and 306996324 contained no significant variability in the TESS data. Manual inspection of these objects confirmed the NGTS light curves contained variability at the reported period from this study. For TIC 14165625, the reported TESS period was approximately half the NGTS period, and for TIC 77951245 the reported periods were similar (5.8 and 5.4 d for NGTS and TESS, respectively), although the phase fold on the NGTS data were cleaner using the NGTS period.

Although a large number of photometric variable stars are known in the Kepler field, we are unable to cross-match with these catalogues as we do not observe this part of the sky. Additionally, we do not attempt to cross-match with small catalogues and papers reporting detections of individual variable objects. Two large variability catalogues we do not attempt cross-matches with are the Zwicky Transient Facility (ZTF) catalogue of periodic variable stars (Chen et al. 2020) or the catalogue of variable stars measured by the Asteroid Terrestrial-impact Last Alert System (ATLAS) (Heinze et al. 2018). The ZTF catalogue contains 4.7 million candidate variables and the ATLAS catalogue 621 702 candidate variables. Both surveys target much fainter objects than NGTS: the brightest candidates in both surveys are approximately as bright as the faintest objects observed by NGTS (Tonry et al. 2018; Masci et al. 2019). Due to the small overlap in brightness and a large number of candidates in each catalogue, we elected not to perform a cross-match. Further cross-matching with smaller catalogues is possible, as we provide the position in RA and Dec., as well as TICv8 and Gaia DR2 identifiers (where available) for all 16 880 variable sources.

#### 4.4 Period ranges of interest

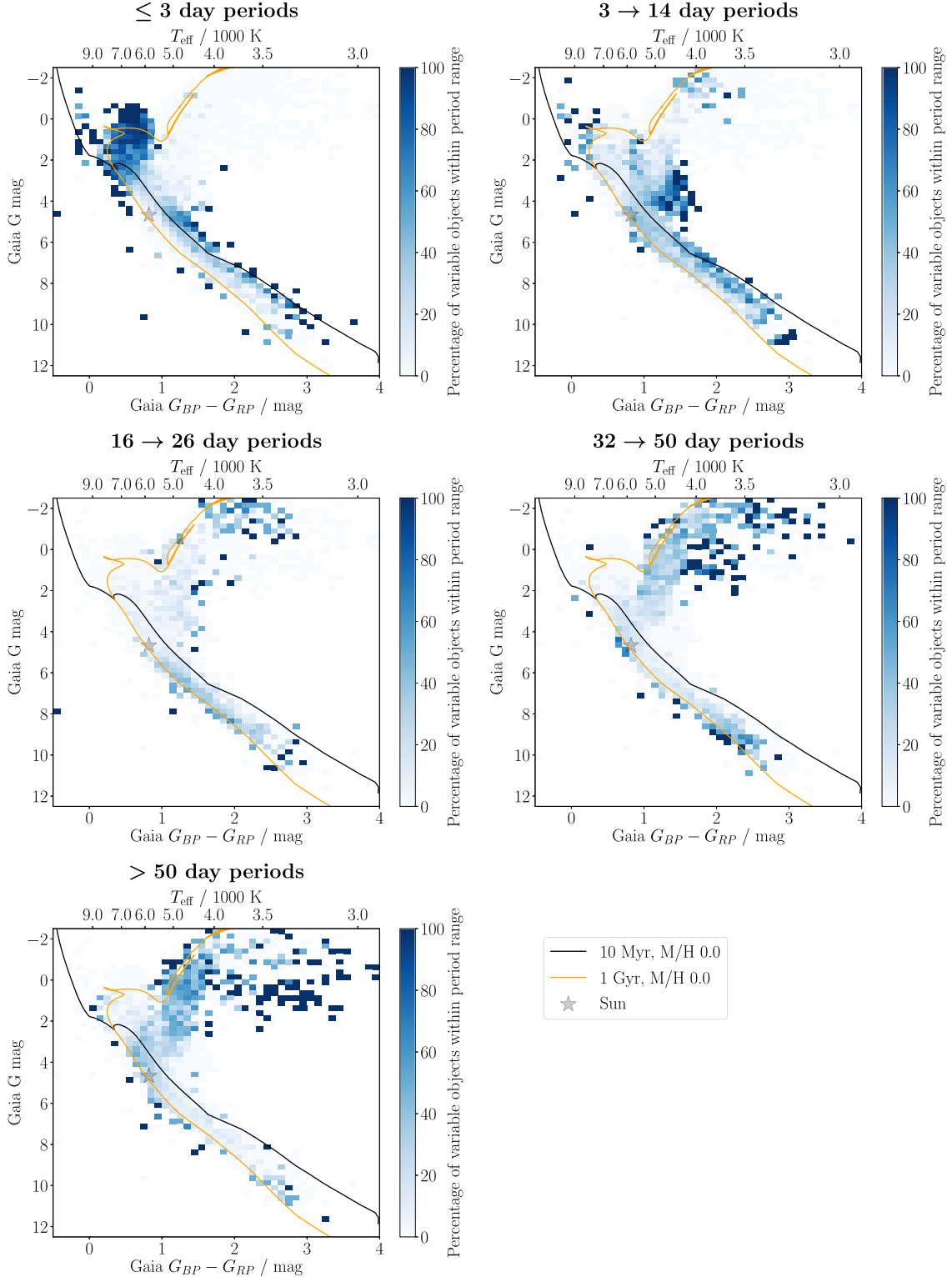
We break our results down into unevenly spaced intervals in variability period in order to assess how samples of similar variability periods are distributed in colour-magnitude space in Fig. 7. This reveals more information than Fig. 3 as we are able to probe into the high-density main sequence. We have selected the period ranges empirically taking into account the sampling gaps at 14 and 28 d arising from Moon contaminated signals.

The majority of the shortest period variability lies at the top of the main sequence. This could be indicative of  $\delta$ -Scuti, RR-Lyrae or rapidly oscillating Ap stars in the instability strip. Typically, RR-Lyrae type objects lie in this region at the lower end of the instability strip and pulsate with periods of less than 1 d. The peak density for less evolved stars is above the main sequence at this period range. Between 1 and 10 d, we would expect to observe the rotation of YSOs such as T-Tauri stars or young main-sequence stars (e.g. as seen in Gaia Collaboration 2019). We may also observe short-period binary star systems at this period range, which would also lie above the main sequence on the HR diagram. In the period range 3 to 14 d, we continue to see a peak density above the main sequence, though the bulk moves towards later spectral types compared to the very short periods.

Between 16 and 26 d, we see the peak density move towards the main sequence as well as a distinct lack of objects above the main sequence. At  $>30$  d, we start to see detections into the RGB as well as more M-type stars. We would expect giant, evolved stars to have longer period rotation or pulsations. Moving from between 32 and 50 d to  $>50$  day periods we see the bulk of objects move further up the RGB and down the main sequence towards cooler temperatures and redder colours.

#### 4.5 Period-colour distribution

We plot our variability periods against colour in Fig. 8 and see a number of prominent features. Most striking is the high density of stars known in the literature as e.g. the ‘I-Sequence’ (Barnes 2003)

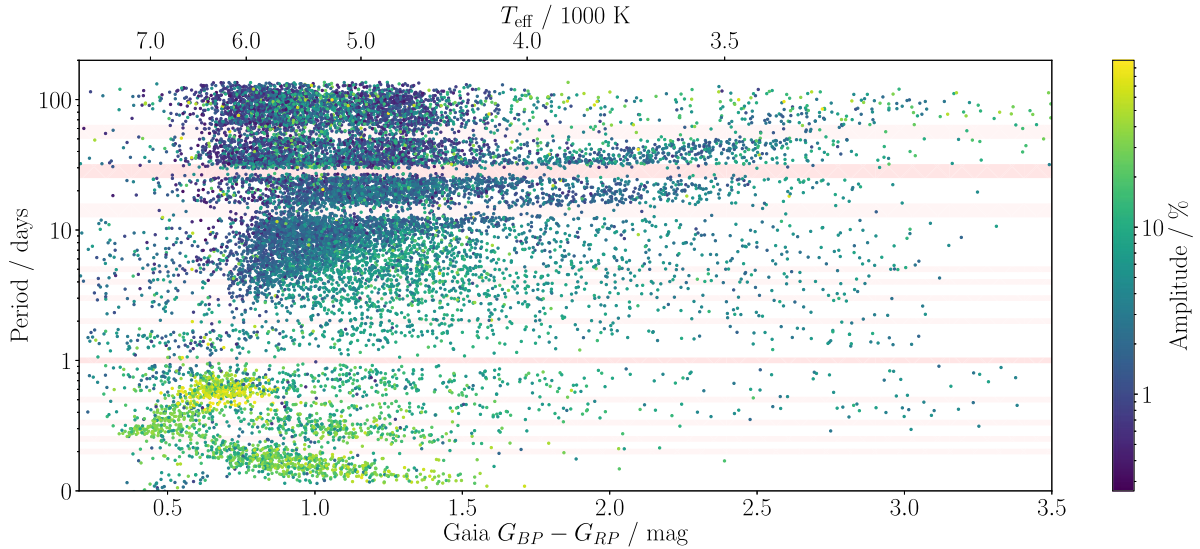


**Figure 7.** HR diagrams for the NGTS variability sample broken down into period ranges. Periods in the sample range from  $\sim 0.1$  to 130 d. The colour bar indicates the percentage of all variable objects across all period ranges which lie in this specific colour–magnitude–period bin. The sum of each bin across all five subplots will equal 100 per cent. Solar metallicity PARSEC isochrones of ages 10 Myr and 1 Gyr are included as solid black and orange lines, respectively.

or the ‘Ridge’ (Kovács 2015) spanning a period range from 4–40 d and  $G_{BP} - G_{RP}$  0.75 – 3.5. The shape of this envelope has been empirically defined by Angus et al. (2019), using a broken power-law gyrochronology model calibrated against the  $\sim 800$  Myr old Praesepe cluster.

We see a large number of long-period ( $> 40$  d) objects between  $G_{BP} - G_{RP}$  of  $\sim 0.7$  to  $\sim 1.4$ . We would expect a higher density of detections at this colour range due to the high-density main-sequence turnoff and red clump, as shown in Fig. 3(b). Older main-sequence stars at this colour range may exhibit long period rotational





**Figure 8.** Effective temperature and Gaia  $G_{BP} - G_{RP}$  colour against period for 16 880 stars. The colour indicates the 5th – 95th percentile spread of the signal in relative flux. To aid the eye, horizontal strips indicate regions of period space likely affected by systematics arising from the Moon or the 1-d sampling alias, with multiples of these periods more transparent.

modulation. The Cepheid instability strip lies within this colour range, and we would expect to see long-period oscillations from evolved stars driven by the  $\kappa$  mechanism (Saio 1993).

Far below the I-sequence we see a high density of much shorter period, high amplitude variability amongst hot objects at  $G_{BP} - G_{RP} \sim 0.5 \rightarrow 1.5$ , and period  $< 1$  d. This population corresponds to the top of the main sequence on an HR diagram.

We see two distinct groups of objects in period range shorter than 1 d, trending to short periods with increasing colour index ( $G_{BP} - G_{RP}$  0.75  $\rightarrow$  1.5). The two distinct groups are from the same region of the HR diagram – the equal-mass binary main sequence. The light curves showed distinct eclipsing binary signals (as seen in object 1 in Fig. 5), however, the longer period branch contained light curves phase folded on the correct period and in the shorter period branch light curves phase folded on half this period. This is an artefact of the RMS minimization step described in Section 3.7. For eclipsing binaries with slightly different primary and secondary eclipse depths, the full period will show a ‘cleaner’ phase folded light curve with separate primary and secondary eclipses. In comparison, for an equal depth binary the phase folded light curve will have a similar RMS if folded on the correct period or half the period, with the primary and secondary plotted over one another in phase space.

Finally, we observe an increasing upper period envelope with increasing colour for  $G_{BP} - G_{RP} > 1.5$ . We see a number of objects with  $G_{BP} - G_{RP} > 2.5$  having variability periods up to and exceeding 100 d. These objects are discussed in detail in Section 5.2.

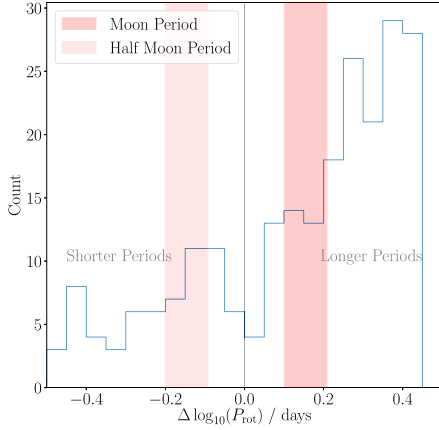
#### 4.6 Period Bi-modality

Within the I-sequence envelope, we see a hint of a region lacking in periodic signals between  $\sim 3500$  K and  $\sim 4500$  K ( $G_{BP} - G_{RP}$  2.5–1.5) and  $\sim 15$ –30 d. This gap has been the topic of extensive discussion in recent papers (such as McQuillan et al. 2014, Davenport & Covey 2018), and although faint, we do observe this gap in this ground-based data set. This gap has previously been fitted using a gyrochrone, roughly following a  $T_{\text{eff}}^{1/2}$  relation (Davenport & Covey 2018), as well as an empirical model using a similar  $T_{\text{eff}}^{1/2}$  relation (Gordon et al. 2021).

To demonstrate the gap is present in our data, we conduct the same analysis as in fig. 3 of Davenport & Covey (2018). We subtract model periods calculated with a 600 Myr gyrochrone defined in Meibom et al. (2011) from our periods. We constrain our data set to objects such that  $1.4 < G_{BP} - G_{RP} < 2.2$  to avoid the gyrochrone crossing the Moon signal sampling gaps. In Fig. 9, we observe a dearth of objects along the gyrochrone, demonstrating the same gap as in the *Kepler* field is present within the NGTS data.

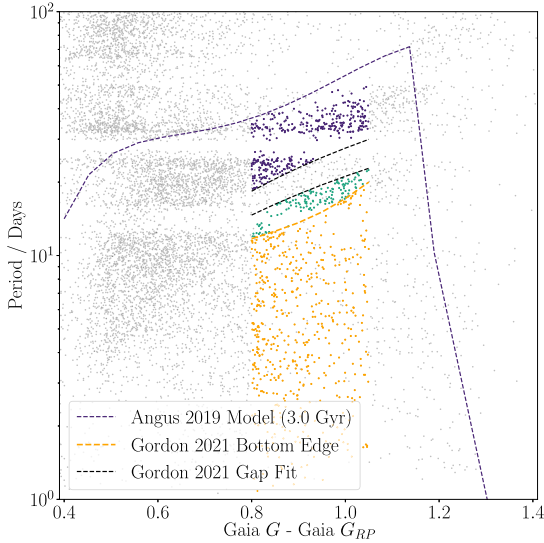
In Fig. 10, we separate our sample into three sub-samples based on a bi-modality gap model and empirical short-period lower limit from Gordon et al. (2021). We observe how far these objects lie in absolute magnitude from an approximate main-sequence isochrone defined at 1 Gyr with Solar metallicity ( $\Delta G$ ), as plotted in Fig. 3. We use this to assess where the three sub-samples lie on the CMD, to ascertain if they arise from distinct stellar populations in terms of colour and intrinsic brightness. We elect to remove potentially evolved stars, giants and sub-giants to ensure the models from Gordon et al. (2021) and Angus et al. (2019) which are fitted to main-sequence stars from *Kepler* and *K2* are applicable. We use the EVOLSTATE code described in Huber et al. (2017) and Berger et al. (2018). The code gives crude evolutionary states for stars based on temperature and radius, with the models derived from Solar-type stars. We remove objects with the ‘subgiant’ or ‘RBG’ flags.

We define our three sub-samples using a number of model constraints in period-colour space. We use the fifth-order polynomial model defined in Angus et al. (2019) to constrain the long-period upper envelope of stars, and the edge-detection based fit from Gordon et al. (2021) to constrain the short-period lower envelope. We calculate the upper and lower edge of the gap using the model defined in Gordon et al. (2021), and select stars from our I-sequence envelope on either side of this branch. This model was only defined for  $0.8 < G - G_{RP} < 1.05$ , so we only use objects within this bound to define the sub-samples. Our third sub-sample is defined as all objects below this boundary in period and will consist of stars not included in the *Kepler* and *K2* data sets which fall well below the well defined I-sequence in period. The model fits used in this section are detailed in Appendix B and plotted in Fig. 10(a).



**Figure 9.** Distribution of the distance from a 600-Myr gyrochrone of the log periods for stars  $1.4 < G_{BP} - G_{RP} < 2.2$ . We see two peaks in the distribution, with a reduced number of rotation periods along the model gyrochrone (grey vertical line). The range of distances from the model to the Moon and half Moon period are included to demonstrate the lower density of objects does not arise from a gap due to the Moon.

The histograms in  $\Delta G$  plotted in Fig. 10(b) show two similar single-peaked distributions from our two longer period sub-samples, and a distinct double-peak distribution for the shorter period sub-sample. We note that this second peak lies approximately 0.75 magnitudes above the peaks of the two longer period sub-samples which could indicate a population of binary objects which is not present in the upper two sub-samples. This confirms our previous observation from the HR diagram: a group of very short period



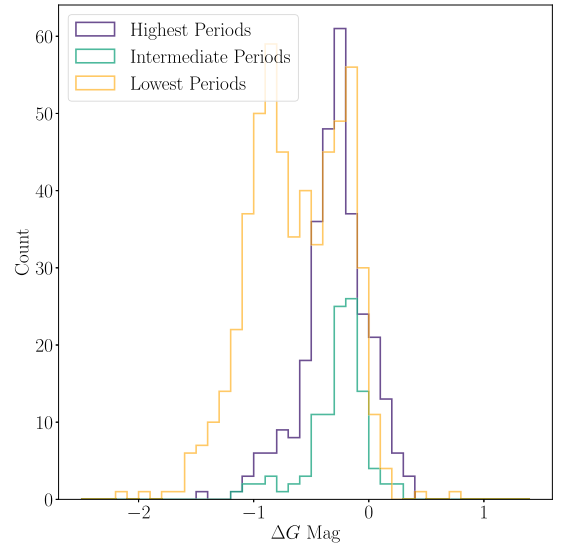
**(a)** Gaia  $G - G_{RP}$  colour vs. period. Three sub-samples spanning the observed period gap (above, below and significantly below the gap) are defined using models (see legend) and coloured blue, green and orange, respectively. Note we do not plot the large model uncertainties defined for the Angus et al. (2019) model for stars outside the range  $0.56 < G_{BP} - G_{RP} < 2.7$  ( $0.31 < G - G_{RP} < 1.17$ ).

objects just above the main sequence, which could correspond to a sample heavily contaminated by binary sources. The two longer period sub-samples appear to have by-eye similar distributions of  $\Delta G$ , which leads us to believe the two branches are drawn from similar stellar populations in terms of colour, intrinsic brightness, and multiplicity.

## 5 DISCUSSION

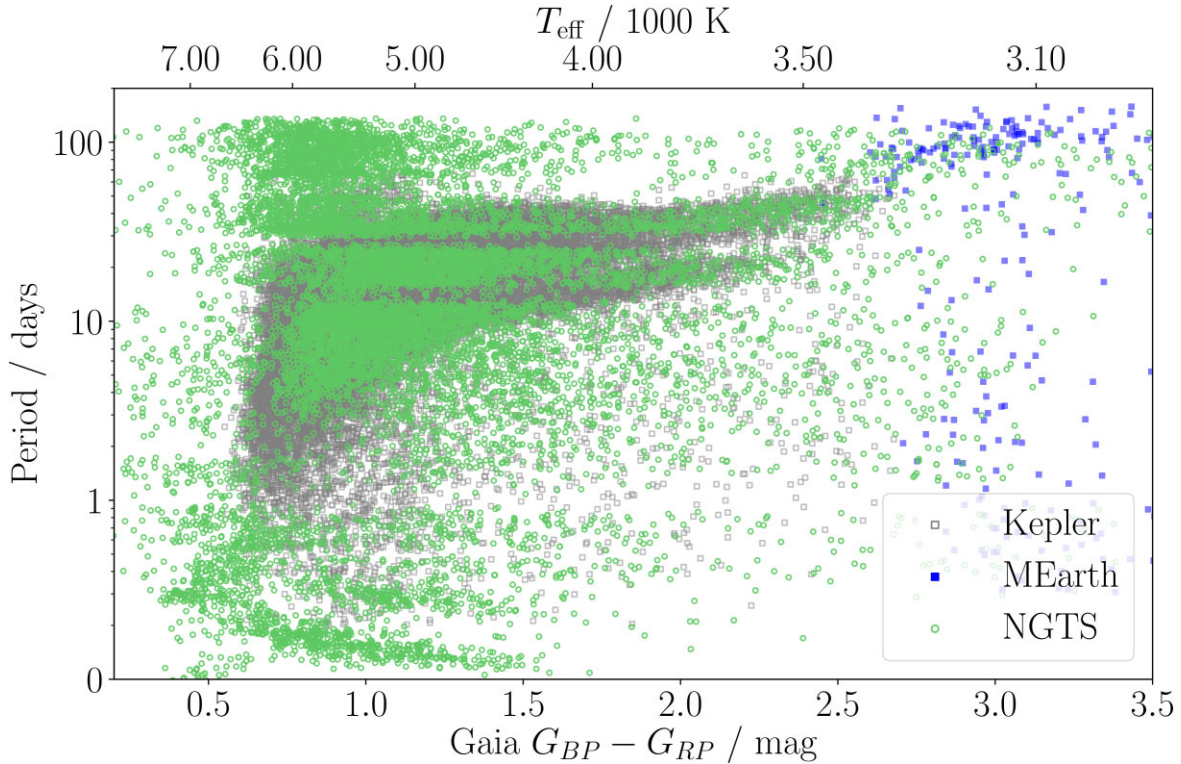
### 5.1 Comparison to similar studies

The NGTS data set demonstrates that we are able to use ground-based photometry to conduct stellar variability studies matching the scale of space-based data. In contrast to, for example, the *Kepler* data set used by McQuillan et al. (2014) and Davenport & Covey (2018), NGTS sources are not pre-selected. This provides a much more representative sample of field stars which is demonstrated in the much higher number of objects which lie away from the high-density I-sequence envelope of stars in period-colour space. Objects which lie within the I-sequence will encompass a selection of stars most likely to be main-sequence, single objects similar to the *Kepler* input catalogue. We overlay data from the *Kepler* rotation study by McQuillan et al. (2014) with our data in Fig. 11. In particular, we see a high density of objects at  $G_{BP} - G_{RP} \sim 1.0$  with periods longer than roughly 40 d not present in the *Kepler* data set. These objects lie in the RGB and AGB on the HR diagram, so will be giant objects which have not been removed from the NGTS study. We also see a large number of objects with much shorter periods than the I-Sequence envelope. These objects lie above the main sequence on the CMD and will be either short-period binary sources or potential YSOs.



**(b)** Histogram of distance in Gaia  $G$  from a main-sequence isochrone (1 Gyr, Solar metallicity) for our three sub-samples (coloured as in panel (a)).

**Figure 10.** Panel (a): Period-colour diagram of our sample, with three sub-samples defined by empirical models from Gordon et al. (2021) and Angus et al. (2019). Panel (b): Histograms of the magnitude difference in each of the three sub-samples from a main-sequence isochrone.



**Figure 11.** Effective temperature versus period data compared for this study (NGTS data, green circles), McQuillan et al. (2014) (*Kepler* data, grey squares), and Newton et al. (2018) (MEarth data, blue squares).

In addition to finding astrophysical signals of interest, we were also able to observe systematic periodicity down to amplitudes of 0.3 per cent.

This study highlights the power of ground-based photometric surveys in terms of the size and precision of the data set. We are able to extract a data set which rivals that of the *Kepler* and *K2* missions, with a much longer baseline (in the case of *K2*) and a much greater range of pointings (in the case of *Kepler*). As a corollary, this study also serves as an exercise that ground-based photometric data may prove more difficult to analyse systematically than space-based data due to increased sources of noise and aliasing. We note a lower recovery rate of periodic signals than other studies. McQuillan et al. (2014) found variability in 25.6 per cent of their  $\sim 130\,000$  objects, Gordon et al. (2021) found variability in almost 13 per cent of their 69 000 objects, and NGTS was able to find variability in about 2 per cent of 829 481 objects. We note that 21 per cent of all objects were flagged as having signals arising from Moon contamination, our largest source of systematic noise in the study.

The combination of a relatively long baseline ( $\sim 250$  d) and multiple pointings (94 used in this study) allows the NGTS data set to probe out to reasonably long period regimes (0.1–130 d) and across a range of spectral types (late-A to mid-M).

## 5.2 Long period M-Dwarfs

Previous studies such as Newton et al. (2018) have used targeted ground-based photometry to extract very long-period variability for M-dwarfs. We also observe these extremely long periods ( $> 100$  days) in our M-dwarf sample. Fig. 8 shows an upwards trend in period in the mid-M-dwarf sample at  $T_{\text{eff}} < 3500\text{K}$ . In order to provide a useful comparison to the MEarth rotation study, we also assessed this trend

for just dwarf stars (as defined by *evolstate*). Our sample contains 751 non-evolved, dwarf objects with variability periods with Gaia  $G_{BP} - G_{RP} > 2.21$ , which is the bluest limit of the MEarth rotation study catalogue.

In this study, the fields chosen had at most a 250-d time-series, which allows us to robustly extract periods up to roughly 125 d in length. Newton et al. (2018) observed periods up to 140-d long for some of these objects, hypothesising that an upper limit close to this period would occur through Skumanich-like angular momentum loss for stars of the ages observed in the local thick disc. Using the Skumanich  $t^{1/2}$  relation and taking the age of the local thick disc to be  $8.7 \pm 0.1$  Gyr (Kilic et al. 2017), we calculate the longest Skumanich relation period to be approximately 145 d. The NGTS rotation periods qualitatively agree with the distribution of rotation periods seen in M-dwarfs by Newton et al. (2018), however, we reach the period limit of the NGTS data just shy of the  $\sim 140$ -d limit in the MEarth detections. It is interesting to note the Skumanich relation still appears to hold from the longest period objects across samples, even into the fully convective M-dwarf population for which the physics of spin-down is not fully understood. Further observations of much older open clusters could shed light on this interesting long-period M-dwarf sample, and observations with much longer time baselines would allow us to probe into period regimes where spin-down could be more efficient than the Skumanich relation. We note that current photometric space missions such as TESS (Ricker et al. 2014) may be useful to shed light on this long term variability across the sky, but only at the ecliptic poles where objects will be observed for up to 1 yr continuously, with a one year gap before another year of continuous observation. Most of the sky will only be observed for 28 d at a time, meaning a maximum of 14-d periods could be reliably extracted.



This NGTS study overlaps both the *Kepler* rotation period data and the MEarth rotation period data, allowing more robust comparisons to be made between the two previously disjoint samples. The NGTS data set provides a broad view into stellar rotation, targeting similar Solar-type stars as observed by *Kepler*, as well as more diverse populations across the HR diagram and across a range of pointings.

### 5.3 Period bi-modality

We continue the ongoing discussion regarding the rotation period gap (McQuillan et al. 2014; Davenport & Covey 2018; Reinhold et al. 2019; Angus et al. 2020; Reinhold & Hekker 2020; Gordon et al. 2021), including the first ground-based data set to have observed this feature in period temperature space. Although the gap is not as clear as in the space-based data, we align models from a number of previous works to a region of lower density in the NGTS data, as shown in Fig. 9.

By utilizing empirical models from previous studies on *Kepler* and *K2* data, we separated our sample into three sub-samples: this is seen in Fig. 10. Within the two upper sub-samples, we see the highest period objects are on average further above the main sequence in *G* than the lower period objects. This effect has been previously observed, as Davenport & Covey (2018) saw a small increase in period as we move up in magnitude from the main sequence, but not as far as to be influenced by large numbers of binary objects. We note, similar to the Davenport & Covey (2018) study that we do not account for metallicity or age when considering the distance from a Solar metallicity defined main-sequence isochrone at 1 Gyr. Metallicity has been shown to affect the amplitude of variability signals and additionally may lead to observational biases whereby for a given mass, higher metallicity stars' variability is more easily detected (See et al. 2021). There is also the possibility of contamination by lower mass-ratio binary systems. Further observations of open clusters with defined stellar ages and a tight single-star main sequence may afford more conclusive evidence towards this period gradient across the main sequence. Such studies have been conducted on open clusters across a large range of ages such as Blanco 1 (~100 Myr) (Gillen et al. 2020a), Praesepe (~800 Myr) (Rebull et al. 2016, 2017), Ruprecht 147 (~3 Gyr) (Gruner & Barnes 2020), and M67 (~4 Gyr) (Barnes et al. 2016).

The two sub-samples do not appear to be significantly contaminated by multiple systems and arise from similar locations on the HR diagram. Combined with the knowledge that these objects are from a range of pointings, this supports the conclusion of Gordon et al. (2021) that these two sub-samples do not derive from two distinct star formation epochs.

A broken spin-down law as discussed in Gordon et al. (2021) would be explained well by our data, including the possibility that the (very few) objects observed within this gap are currently transitioning between the two longer period sub-samples. In this broken spin-down law, the angular momentum change of the star will deviate from the expected  $t^{1/2}$  relation proposed by Skumanich (1972) due to the transfer of angular momentum between the envelope and the core. Prior to this transfer of angular momentum, the core and envelope are decoupled, resulting in the expected  $t^{1/2}$  spin-down of the envelope but with a rapidly rotating core which will then reduce or even stop the spin-down once the core and envelope recouple. This model has been suggested to fit *Kepler* data in addition to *K2* data (Angus et al. 2020; Gordon et al. 2021), and theorists such as Lanzafame & Spada (2015) and later Spada & Lanzafame (2020) have incorporated these effects into stellar evolution models

which have been shown to fit observed cluster data of different ages. The proposed models include a two-zone model of internal stellar coupling, with a parameter describing the mass dependence of the coupling. The recent analysis of the ~3 Gyr old open cluster Ruprecht 147 by Gruner & Barnes (2020) demonstrates that the model from Spada & Lanzafame (2020) incorporating internal angular momentum transfer is best suited to model the rotational evolution of stars redder than K3 in comparison to more naive gyrochronology models.

Another suggestion for the origin of this gap comes from an analysis by Reinhold et al. (2019) and Reinhold & Hekker (2020) of *K2* data. In their proposed model, the gap arises from objects in which the photometric variability arising from spots and faculae is of similar magnitude, thus cancelling out resulting in lower amplitude variability that is correspondingly harder to detect. They observed a slight decrease in signal amplitude on either side of the gap in period, and hypothesised objects of this period could exhibit spot-faculae photometric cancellation. We do not observe such an obvious decrease in signal amplitude in our full sample, and when considering a smaller range of amplitudes more aligned with the *K2* sample we again did not see this amplitude gradient. This may be attributed to NGTS photometry being less precise than *Kepler*, and a small change on a signal of 1 per cent amplitude may not be detectable. To accurately determine the dominant surface feature of a star requires observations of spot-crossing events during planetary transits or Doppler images, neither of which are appropriate for follow-up from a large-scale photometric study.

## 6 CONCLUSIONS

In this study, we extract robust variability periods for 16 880 stars out of 829 481 stars observed with the NGTS, based in Paranal, Chile. This is the largest ground-based systematic photometric variability study conducted to date with such precise and high-cadence photometry and highlights both the advantages of such studies as well as the challenges. Using precise ground-based photometry, plus a generalization of the autocorrelation function to irregularly sampled data, we are able to detect variability amplitudes down to levels of 0.3 per cent. The contamination of signals by systematics demonstrates that using ground-based photometry requires further thought than using much cleaner space-based data in order to avoid false positives arising through aliases. The most common source of aliases arose from Moon contaminated signals as well as aliasing from the 1-d periodic sampling intrinsic to ground-based observations. We demonstrate we are able to overcome these limitations and produce robust variability signals across our sample.

In comparison to previous large-scale stellar variability studies, we note that with NGTS we are able to observe across the Southern sky (in comparison to *Kepler*'s single pointing, as in McQuillan et al. 2014 and Davenport & Covey 2018). We do not pre-select our targets as is the case for *Kepler* and *K2*, and so we are able to observe variability across a more varied stellar sample. In particular, we extract long-term variability periods for a population of cool dwarfs, similar to a population observed by Newton et al. (2018) using MEarth. This is made possible through our longer observation baseline than space-based missions such as *K2*. This large population, sampled across the sky over a long (250 d) baseline allows this study to connect previous space-based studies on main-sequence, predominantly Solar-type stars with ground-based M-dwarf studies, which were previously unconnected.



Within the bulk of our rotation period ‘I-Sequence’, we observe a gap between 15 and 25 d, first observed by McQuillan et al. (2014), and later studied in detail by Davenport & Covey (2018), Reinhold et al. (2019), Reinhold & Hekker (2020), Angus et al. (2020), and Gordon et al. (2021). Using models from Gordon et al. (2021), Angus et al. (2019), and Meibom et al. (2011) we are able to demonstrate that the gap is present in our data set, and also show that the two sub-samples of main-sequence objects above and below this gap appear to arise from similar stellar populations on the CMD which are not contaminated by high levels of binarity. This supports the hypothesis of a broken spin-down model as proposed by Lanzafame & Spada (2015) and Spada & Lanzafame (2020) rather than distinct populations of star formation.

We also conclude that although a large population study of field stars is useful for assessing trends in the wider stellar population, without well-defined ages of target stars it is difficult to confirm angular momentum models. We suggest that studies of open clusters with well-defined ages and tight rotation sequences such as the recent study by Gruner & Barnes (2020) will yield the most conclusive evidence towards how stellar angular momentum evolves over the lifetime of a star. Additionally, we observe a number of interesting non-main-sequence populations, including a small population of objects which lie well above the main sequence with short rotation periods. Follow-up observations of these targets would allow us to ascertain whether these stars are young, single stars such as T-Tauri objects, or multiple star systems. This data set presents a wealth of additional data with many avenues for follow-up science. These include both continued systematic variability analysis of the NGTS data and also more in-depth analysis of interesting sub-populations of variable objects not explored in this cardinal NGTS variability study.

## ACKNOWLEDGEMENTS

We thank the anonymous referee for constructive suggestions and comments that improved this paper. Based on data collected under the NGTS project at the ESO La Silla Paranal Observatory. The NGTS facility is operated by the consortium institutes with support from the UK Science and Technology Facilities Council (STFC) under projects ST/M001962/1 and ST/S002642/1.

This work was performed using resources provided by the Cambridge Service for Data Driven Discovery (CSD3) operated by the University of Cambridge Research Computing Service ([www.csd3.cam.ac.uk](http://www.csd3.cam.ac.uk)), provided by Dell EMC and Intel using Tier-2 funding from the Engineering and Physical Sciences Research Council (capital grant EP/P020259/1), and DiRAC funding from the Science and Technology Facilities Council ([www.dirac.ac.uk](http://www.dirac.ac.uk)).

This research has made use of the VizieR catalogue access tool, CDS, Strasbourg, France (DOI: 10.26093/cds/vizieR). The original description of the VizieR service was published in A&AS 143, 23.

JTB is supported by an STFC studentship as part of the CDT in Data Intensive Sciences. EG gratefully acknowledges support from the David and Claudia Harding Foundation in the form of a Winton Exoplanet Fellowship. MNG acknowledges support from MIT’s Kavli Institute as a Juan Carlos Torres Fellow and from the European Space Agency (ESA) as an ESA Research Fellow. JSJ gratefully acknowledges support by FONDECYT grant 1201371 and from the ANID BASAL projects ACE210002 and FB210003. JAGJ acknowledges support from grant HST-GO-15955.004-A from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-26555.

## DATA AVAILABILITY

The NGTS data underlying this article are publicly available through the ESO Archive in line with the NGTS data publication policy.

The variability period results produced in this article are available as supplementary material and at CDS via anonymous ftp to [cdsarc.u-strasbg.fr](mailto:cdsarc.u-strasbg.fr) (130.79.128.5) or via <https://cdsarc.unistra.fr/viz-bin/ca/J/MNRAS>.

## REFERENCES

- Affer L., Micela G., Favata F., Flaccomio E., 2012, *MNRAS*, 424, 11  
 Aigrain S., Hodgkin S., Irwin J., Hebb L., Irwin M., Favata F., Moraux E., Pont F., 2007, *MNRAS*, 375, 29  
 Angus R., Morton T., Aigrain S., Foreman-Mackey D., Rajpaul V., 2018, *MNRAS*, 474, 2094  
 Angus R. et al., 2019, *AJ*, 158, 173  
 Angus R. et al., 2020, *AJ*, 160, 90  
 Auvergne M. et al., 2009, *A&A*, 506, 411  
 Bailer-Jones C. A. L., Rybizki J., Foesneau M., Mantelet G., Andrae R., 2018, *AJ*, 156, 58  
 Bakos G., Noyes R. W., Kovács G., Stanek K. Z., Sasselov D. D., Domsa I., 2004, *Publ. Astron. Soc. Pac.*, 116, 266  
 Barnes S. A., 2003, *ApJ*, 586, 464  
 Barnes S. A., 2007, *ApJ*, 669, 1167  
 Barnes S. A., Weingrill J., Fritzewski D., Strassmeier K. G., Platais I., 2016, *ApJ*, 823, 16  
 Basri G. et al., 2011, *AJ*, 141, 20  
 Berger T. A., Huber D., Gaidos E., van Saders J. L., 2018, *ApJ*, 866, 99  
 Booth R. S., Poppenhaeger K., Watson C. A., Silva Aguirre V., Wolk S. J., 2017, *MNRAS*, 471, 1012  
 Borucki W. J. et al., 2010, *Science*, 327, 977  
 Bressan A., Marigo P., Girardi L., Salasnich B., Dal Cero C., Rubele S., Nanni A., 2012, *MNRAS*, 427, 127  
 Canto Martins B. L. et al., 2020, *ApJS*, 250, 20  
 Cargile P. A., James D. J., Pepper J., Kuhn R. B., Siverd R., Stassun K. G., 2014, *ApJ*, 782, 29  
 Chen X., Wang S., Deng L., de Grijs R., Yang M., Tian H., 2020, *ApJS*, 249, 18  
 Ciardi D. R. et al., 2011, *AJ*, 141, 108  
 Collier Cameron A. et al., 2006, *MNRAS*, 373, 799  
 Collier Cameron A. et al., 2009, *MNRAS*, 400, 451  
 Cooley J. W., Lewis P. A. W., Welch P. D., 1969, *IEEE Trans. Educ.*, 12, 27  
 Damiani C., Lanza A. F., 2015, *A&A*, 574, A39  
 Davenport J. R. A., Covey K. R., 2018, *ApJ*, 868, 151  
 David-Uraz A. et al., 2019, *MNRAS*, 487, 304  
 Dumusque X. et al., 2017, *A&A*, 598, A133  
 Evans D. W. et al., 2018, *A&A*, 616, A4  
 Eyer L., Mowlavi N., 2008, *J. Phys. Conf. Ser.*, 118, 012010  
 Gaia Collaboration, 2018a, *A&A*, 616, A1  
 Gaia Collaboration, 2018b, *A&A*, 616, A10  
 Gaia Collaboration, 2019, *A&A*, 623, A110  
 Gaia Collaboration, 2021, *A&A*, 649, A1  
 Gallet F., Bolmont E., Bouvier J., Mathis S., Charbonnel C., 2018, *A&A*, 619, A80  
 Gillen E. et al., 2020a, *MNRAS*, 492, 1008  
 Gillen E., Hillenbrand L. A., Stauffer J., Aigrain S., Rebull L., Cody A. M., 2020b, *MNRAS*, 495, 1531  
 Gordon T. A., Davenport J. R. A., Angus R., Foreman-Mackey D., Agol E., Covey K. R., Agüeros M., Kipping D., 2021, *ApJ*, 913, 70  
 Gruner D., Barnes S. A., 2020, *A&A*, 644, A16  
 Grunhut J. H. et al., 2017, *MNRAS*, 465, 2432  
 Harris C. R. et al., 2020, *Nature*, 585, 357  
 Hartmann L., Stauffer J. R., 1989, *AJ*, 97, 873  
 Hartman J. D., Bakos G. Á., Kovács G., Noyes R. W., 2010, *MNRAS*, 408, 475

- Hayashi C., 1961, *PASJ*, 13, 450
- Haywood R. D. et al., 2014, *MNRAS*, 443, 2517
- Heinze A. N. et al., 2018, *AJ*, 156, 241
- Hodgkin S. T., Irwin J. M., Aigrain S., Hebb L., Moraux E., Irwin M. J., Monitor Collaboration, 2006, *Astron. Nachr.*, 327, 9
- Holl B. et al., 2018, *A&A*, 618, A30
- Howell S. B. et al., 2014, *Publ. Astron. Soc. Pac.*, 126, 398
- Huber D. et al., 2017, *ApJ*, 844, 102
- Irwin M. J. et al., 2004, in Quinn P. J., Bridger A., eds, *Proc. SPIE Conf. Ser. Vol. 5493, Optimizing Scientific Return for Astronomy through Information Technologies*. SPIE, Bellingham, p. 411
- Irwin J., Hodgkin S., Aigrain S., Hebb L., Bouvier J., Clarke C., Moraux E., Bramich D. M., 2007, *MNRAS*, 377, 741
- Jackman J. A. G. et al., 2019, *MNRAS*, 482, 5553
- Jayasinghe T. et al., 2018, *MNRAS*, 477, 3145
- Jayasinghe T. et al., 2021, *MNRAS*, 503, 200
- Kawaler S. D., 1988, *ApJ*, 333, 236
- Kilic M., Munn J. A., Harris H. C., von Hippel T., Liebert J. W., Williams K. A., Jeffery E., DeGennaro S., 2017, *ApJ*, 837, 162
- Kovács G., 2015, *A&A*, 581, A2
- Lanzafame A. C., Spada F., 2015, *A&A*, 584, A30
- Lomb N. R., 1976, *Ap&SS*, 39, 447
- McQuillan A., Aigrain S., Mazeh T., 2013, *MNRAS*, 432, 1203
- McQuillan A., Mazeh T., Aigrain S., 2014, *ApJS*, 211, 24
- Masci F. J. et al., 2019, *Publ. Astron. Soc. Pac.*, 131, 018003
- Meibom S. et al., 2011, *ApJ*, 733, L9
- Muirhead P. S., Dressing C. D., Mann A. W., Rojas-Ayala B., Lépine S., Paegert M., De Lee N., Oelkers R., 2018, *AJ*, 155, 180
- Newton E. R., Mondrik N., Irwin J., Winters J. G., Charbonneau D., 2018, *AJ*, 156, 217
- Pecaut M. J., Mamajek E. E., 2013, *ApJS*, 208, 9
- Pepper J., Kuhn R. B., Siverd R., James D., Stassun K., 2012, *Publ. Astron. Soc. Pac.*, 124, 230
- Pollacco D. L. et al., 2006, *Publ. Astron. Soc. Pac.*, 118, 1407
- Queloz D. et al., 2001, *A&A*, 379, 279
- Ranjan S., Wordsworth R., Sasselov D. D., 2017, *ApJ*, 843, 110
- Rebull L. M. et al., 2016, *AJ*, 152, 114
- Rebull L. M., Stauffer J. R., Hillenbrand L. A., Cody A. M., Bouvier J., Soderblom D. R., Pinsonneault M., Hebb L., 2017, *ApJ*, 839, 92
- Reinhold T., Hekker S., 2020, *A&A*, 635, A43
- Reinhold T., Bell K. J., Kuzlewicz J., Hekker S., Shapiro A. I., 2019, *A&A*, 621, A21
- Ricker G. R. et al., 2014, in Oschmann J. M. J., Clampin M., Fazio G. G., MacEwen H. A., eds, *Proc. SPIE Conf. Ser. Vol. 9143, Space Telescopes and Instrumentation 2014: Optical, Infrared, and Millimeter Wave*. SPIE, Bellingham, p. 914320
- Rimmer P. B., Xu J., Thompson S. J., Gillen E., Sutherland J. D., Queloz D., 2018, *Sci. Adv.*, 4, eaar3302
- Saio H., 1993, *Ap&SS*, 210, 61
- Scargle J. D., 1982, *ApJ*, 263, 835
- Seager S., 2013, *Science*, 340, 577
- See V., Roquette J., Amard L., Matt S. P., 2021, *ApJ*, 912, 127
- Segura A., Walkowicz L. M., Meadows V., Kasting J., Hawley S., 2010, *Astrobiology*, 10, 751
- Shappee B. J. et al., 2014, *ApJ*, 788, 48
- Shumway R., Stoffer D., 2006, *Time Series Analysis and Its Applications: With R Examples*. Springer, Cham
- Sikora J. et al., 2019, *MNRAS*, 487, 4695
- Skumanich A., 1972, *ApJ*, 171, 565
- Sousa A. P. et al., 2016, *A&A*, 586, A47
- Spada F., Lanzafame A. C., 2020, *A&A*, 636, A76
- Stassun K. G. et al., 2018, *AJ*, 156, 102
- Stassun K. G. et al., 2019, *AJ*, 158, 138
- Tamuz O., Mazeh T., Zucker S., 2005, *MNRAS*, 356, 1466
- Tilley M. A., Segura A., Meadows V., Hawley S., Davenport J., 2019, *Astrobiology*, 19, 64
- Tonry J. L. et al., 2018, *ApJ*, 867, 105
- Wheatley P. J. et al., 2018, *MNRAS*, 475, 4476

## SUPPORTING INFORMATION

Supplementary data are available at *MNRAS* online.

### ngtsrot.csv

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

## APPENDIX A: DETAILED MOON SIGNAL ANALYSIS

In order to systematically detect Moon contaminated signals (e.g. as shown in Fig. A1), we fit a model to the flux data, phase folded on the expected Moon period for each NGTS field. The expected Moon period is calculated from a scaled expected Moon brightness curve, calculated as a product of the on-sky separation of the field from the Moon and the Moon illumination fraction,  $I = (1 + \cos(\theta_{\text{phase}}))/2$ .  $\theta_{\text{phase}}$  is the Moon phase angle defined for a time and ephemeris. For most fields, this gave a period of approximately 28.5 d, between the synodic and sidereal periods as expected.

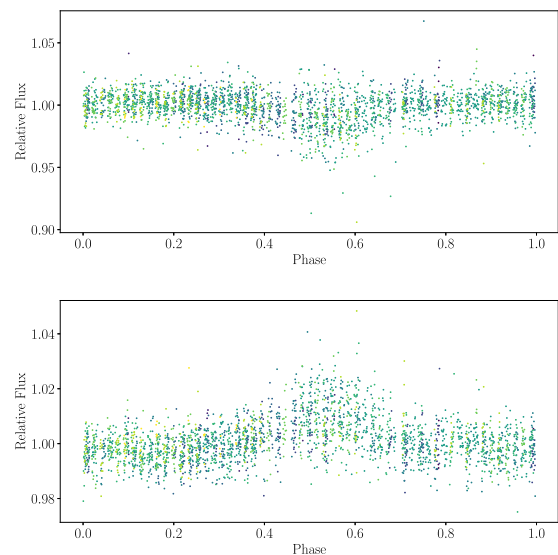
The model is a simple three-parameter, piecewise model described in equation (A1), where the parameter  $x$  is the location in half-phase  $x \in [0, 1]$ .

$$\begin{cases} \text{flux}_0 & 0 \leq x \leq \text{turnover} \\ mx + c & \text{turnover} < x \leq 1, \end{cases} \quad (\text{A1})$$

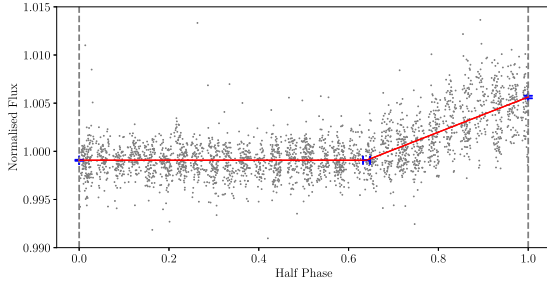
where

$$m = \frac{\text{flux}_1 - \text{flux}_0}{1 - \text{turnover}}$$

$$c = \text{flux}_1 - m.$$



**Figure A1.** Two examples of typical Moon tainted signals. For each object, the light curve is phase folded on the expected Moon period and epoch. 0.0 and 1.0 phase are at new Moon, 0.5 phase is at full Moon. We see an example of an overcorrected signal with a typical decrease in flux at full Moon (top panel). An undercorrected signal demonstrates the opposite trend (bottom panel). Both signals exhibit an increase in scatter at full Moon, with an otherwise fairly flat light curve.



**Figure A2.** The three-parameter Moon model fit is used to assess if a signal is contaminated by the Moon. The flux data is phase folded on the period of the Moon and then again in half such that 0.0 in phase corresponds to new Moon and 1.0 in phase corresponds to full Moon.

We fit for the three parameters  $\text{flux}_0$ ,  $\text{flux}_1$ , and turnover. This model fit is assessed by checking the following criteria, with an example shown in Fig. A2.

- (i) Is the model turnover point at the expected point in phase? (between 0.2 and 0.8 in half-phase).
- (ii) Is there a flux RMS increase after the model turnover point?
- (iii) Is there a noticeable (i.e.  $>1\sigma$ ) change in flux from new to full Moon?
- (iv) Is there any missing data at full Moon?

If three or more of these criteria are met, the object is flagged as Moon contaminated and removed from the processing.

## APPENDIX B: MODEL PARAMETERS

In Fig. 10(a), we use empirical models defined in Angus et al. (2019) and Gordon et al. (2021). In this section, we provide the model equations and the parameters used.

### B1 Angus model

We use the Praesepe-calibrated gyrochronology relation defined in Angus et al. (2019). The mathematical form of this fifth-order polynomial relationship is given in equations (B1) & (B2) below for two different  $G_{BP} - G_{RP}$  regimes:

$$\log_{10}(P_{\text{rot}}) = c_A \log_{10}(t) + \sum_{n=0}^4 c_n [\log_{10}(G_{BP} - G_{RP})]^n \quad (\text{B1})$$

for stars with  $G_{BP} - G_{RP} < 2.7$  and

$$\log_{10}(P_{\text{rot}}) = c_A \log_{10}(t) + \sum_{m=0}^1 b_m [\log_{10}(G_{BP} - G_{RP})]^m \quad (\text{B2})$$

for stars with  $G_{BP} - G_{RP} > 2.7$ . Here,  $P_{\text{rot}}$  is the rotation period in days, and  $t$  is age in years. We use the best-fitting coefficients from table 1 of Angus et al. (2019).

### B2 Gordon Model

We use the K2 calibrated model from Gordon et al. (2021) to define the upper and lower edges of the bi-modality gap seen in the I-sequence envelope. The gap edges are fitted using a function of the form:

$$P = A(G - G_{RP} - x_0) + B(G - G_{RP} - x_0)^{1/2}, \quad (\text{B3})$$

where  $P$  is the rotation period in days. This equation is defined empirically for K2 stars with  $0.8 < G - G_{RP} < 1.05$ . We use the best-fitting coefficients defined in table 7 of Gordon et al. (2021).

The lower edge of the K2 sample from Gordon et al. (2021) used an edge-detection method, and as such no parametric model form was given. We instead define our lower edge by eye, taking the edge-detection fit line from the Gordon et al. (2021) paper.

<sup>1</sup>*Astrophysics Group, Cavendish Laboratory, J.J. Thomson Avenue, Cambridge CB3 0HE, UK*

<sup>2</sup>*Astronomy Unit, Queen Mary University of London, Mile End Road, London E1 4NS, UK*

<sup>3</sup>*Institute of Astronomy, University of Cambridge, Madingley Rise, Cambridge CB3 0HA, UK*

<sup>4</sup>*School of Physics and Astronomy, University of Leicester, University Road, Leicester, LE1 7RH, UK*

<sup>5</sup>*Department of Physics, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK*

<sup>6</sup>*Centre for Exoplanets and Habitability, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK*

<sup>7</sup>*Astrophysics Research Centre, School of Mathematics and Physics, Queen's University Belfast, BT7 1NN Belfast, UK*

<sup>8</sup>*Institute of Planetary Research, German Aerospace Center, Rutherfordstrasse 2., D-12489 Berlin, Germany*

<sup>9</sup>*Department of Physics and Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

<sup>10</sup>*European Space Agency (ESA), European Space Research and Technology Centre (ESTEC), Keplerlaan 1, NL-2201 AZ Noordwijk, the Netherlands*

<sup>11</sup>*School of Earth and Space Exploration, Arizona State University, Tempe, AZ 85287, USA*

<sup>12</sup>*Núcleo de Astronomía, Facultad de Ingeniería y Ciencias, Universidad Diego Portales, Av. Ejército 441, Santiago, Chile*

<sup>13</sup>*Centro de Astrofísica y Tecnologías Afines (CATA), Casilla 36-D, Santiago, Chile*

<sup>14</sup>*Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, CB3 0WA, UK*

<sup>15</sup>*Max Planck Institute for Gravitational Physics (Albert Einstein Institute), Am Mühlenberg 1, D-14476 Golm, Germany*

<sup>16</sup>*Instituto de Astronomía, Universidad Católica del Norte, Angamos 0610, 1270709, Antofagasta, Chile*

<sup>17</sup>*Observatoire de Genève, Université de Genève, 51 Chemin Pegasi, CH-1290 Sauverny, Switzerland*

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