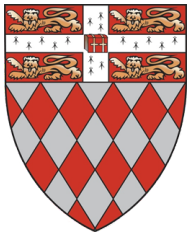
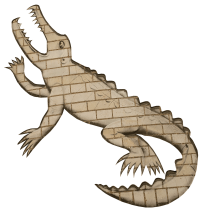
# Data-driven Approaches to Stellar Variability Detection and Characterisation

Joshua Thomas Briegal

Fitzwilliam College, Cambridge



Supervisors: Prof. Didier Queloz, Dr. Edward Gillen
Assessors: Prof. Andrew Collier Cameron, Dr. David Green



April 2022

This thesis is submitted for the degree of Doctor of Philosophy

# Summary

I present the work completed during my time as a PhD student within the Exoplanet Group of the Cavendish Laboratory, University of Cambridge, UK, as a part of the Centre for Doctoral Training in Data-Intensive Sciences. Much of this work has been conducted as a part of the NGTS consortium.

I implemented and tested a novel generalisation of the autocorrelation function (the G-ACF), which applies to irregularly sampled data, such as photometric light curves from ground-based telescopes. I demonstrated that this algorithm accurately estimated the standard ACF, even for poorly sampled astrophysical data, and produced accurate rotation periods that agreed with more complex and computationally expensive models.

I then applied the G-ACF to almost a million photometric light curves from NGTS, finding 16, 880 periodic variability signals from 829, 481 light curves. I combated the noise and aliasing associated with ground-based photometry to produce a stellar variability sample that rivals those from previous space-based photometric studies. I assessed how these variable objects were distributed within colour–magnitude and colour–period space, highlighting distinct populations of variable objects spanning late-A through to mid-M spectral types and with periods between $\sim$ 0.1and 130 days. Within colour–period space, I found a bi-modal structure previously observed in Kepler data and find samples of stars on either side of the gap appear to be from similar populations of stars in terms of colour, intrinsic brightness and multiplicity rather than distinct epochs of star formation.

Finally, I developed a comprehensive period extraction software package, `RoTo`, which uses multiple period extraction techniques to produce reliable period estimates from time-series data. I applied `RoTo` to NGTS observations of the $\sim$ 500 Myr old open cluster NGC 6633. I conducted a detailed study of the rotational variability of member stars, using a combination of literature and machine-learning methods to produce a robust membership list. I calculated distances and extinction values and produced a rotation period sample for the cluster. I compared the slow-rotator sequence of the cluster in colour–period space to similarly aged clusters. I conducted gyro- and isochrone fits to derive probabilistic age estimates for the cluster from rotation, which agreed with age estimates from other methods.

This work is firmly rooted within the principles of Data-Intensive Sciences; I applied performant algorithms to large photometric data sets to produce statistical results with minimal manual input. All of the software developed as a part of this PhD is open-source, and I have released two public `Python` packages.

# Contents

# Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee. Excluding tables, equations and the preamble, this report contains approximately 51,000 words.

# Acknowledgements

Thank you to Didier Queloz and Ed Gillen for their guidance and support throughout my PhD. Working with such experienced astronomers has allowed me to develop a broad skill-set whilst working on interesting and challenging problems within astrophysics. The opportunities given in terms of connections, travel and scientific knowledge have been unrivalled, even with the additional complexities of the COVID-19 pandemic. A further thank you to my assessors, Andrew Collier Cameron and Dave Green, for an engaging and insightful discussion on the topics of this thesis and beyond. Thanks again to Dave Green for providing his LaTeX template and his assistance throughout the creation of this document, as well as many fun afternoons spent demonstrating 1A physics labs.

I would like to thank my cohort of PhD students, both within the exoplanets group and the CDT for Data-Intensive Sciences. Thank you to Gareth Smith for his insight and code-sharing, as well as for proofreading this thesis. A special thanks to Richard Hall, Catriona Murray, Patrick Elwood, Kshitij Sabnis and Chris Desira for their help and friendship throughout the last four years.

Thank you to Jack and the rest of the team at Spherical Defence for the opportunities given to me throughout my CDT work placement. I'm particularly grateful to the ML development team: Merlin, Akbir and Fin. It was great fun working with the team, and it taught me a huge amount.

Thank you to the Exoplanet Group at the Cavendish Laboratory for assisting me where necessary from day one. I would also like to thank STFC for funding my PhD as part of the Centre for Doctoral Training in Data-Intensive Science and the NGTS consortium for providing access to their data and for such a welcoming group of astronomers to work with.

Finally, a thank you to my close friends and family for believing in (or putting up with) me over the last four years. It's not been easy for me either! To Melody, my wonderful fiancée, I could not have completed this work without your continuous support and patience, thank you for everything.

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), 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).

# Chapter 1

# Introduction

This introductory Chapter will paint a picture of the landscape in which the work discussed in this thesis is set. I will introduce the reader to concepts within time-domain astronomy, specifically the techniques used to generate the time-series-based light curves used throughout this work. I will walk through a brief history of the milestone photometric instruments and surveys used to map sources of varying brightness across the sky. These instruments can be broadly grouped into two classes: ground-based and space-based telescopes. I will discuss the advantages and disadvantages of both methods and compare the data taken. I will introduce the reader to the Next Generation Transit Survey, NGTS, whose data forms an integral part of this work. This thesis relies on over one million light curves taken by NGTS from 2016 onward.

I will discuss some of the astrophysical objects of interest that photometric surveys have targeted. Specifically, within the context of this thesis, I will focus on variable stars and open star clusters. Although many of these surveys' main scientific focus is on exoplanet detection and characterisation, stellar variability is both a nuisance in the form of noise as well as its own interesting and diverse scientific pursuit. A brief overview of the field of stellar variability and clusters will be given in this Chapter, with a more detailed scientific background given in Chapter 2.

Finally, I will introduce the reader to the concept of 'big data' and 'data-driven astronomy', which motivate a large portion of the work done in this thesis as a part of the Centre for Doctoral Training in Data-Intensive Science. As surveys become more powerful, the data volume produced increases enormously. Much resource is now devoted to processing, storing and analysing this huge volume of data, for which traditional, manual techniques are not viable.

## 1.1 Time-domain astronomy

Time-domain astronomy focuses on astrophysical objects and phenomena that cause detectable variation in an observable (for example, brightness, a spectrum or on-sky position). Generally, these variations will be on short timescales, extremely short when viewed in the context of astrophysical processes. Some of the shortest astrophysical variability we can observe falls under 'high-speed astrophysics', probing into physical processes taking place on milli-, micro- and even nanosecond scales. Objects such as rapidly rotating pulsars and small-scale magnetohydrodynamic instabilities require extremely fast observation cadences to capture such short-term variability (Dravins 1994). For longer timescales, a long observation baseline is required to collect enough data to observe such variability; for example, a small sample of long-period variable stars was observed for 45 months with the Kepler space telescope, and periodic signals with periods of 100 – 900 days were found (Hartig et al. 2014). Solar cycles observed since 1750 have enabled the study of long-term solar variability and activity changes (National Oceanic and Atmospheric Administration 2021).

Two important concepts within time-domain astronomy are *sampling cadence* and *observation baseline*. The sampling cadence is the spacing between observations. We will come across a wide variety of sampling cadences, as one cadence is not suitable for all observations. For a given observational setup a dim source will require a longer observation to combat photon noise than a brighter source, so we are, in general, unable to observe these dim objects on as short timescales as brighter objects. The positioning of the telescope will also affect sampling. One large difference between ground-based and space-based observations is that ground-based telescopes cannot observe during the day and may face additional periods of telescope downtime due to poor visibility or bad weather. Both space- and ground-based telescopes are vulnerable to technical downtime, for example, maintenance of ground-based surveys or satellite power outages in the case of space-based telescopes.

Ideally, astronomers would take observations on a set of evenly spaced points in time (i.e. *regularly sampled*); however, in practice, factors, as described above, will mean we end up with an irregularly sampled time series. Irregular sampling has been the topic of much frustration within signal processing, with Lomb (1976) and Scargle (1982) proposing the well-known Lomb–Scargle Periodogram method to deal with such gaps in the context of periodic signal detection. I will discuss further details of this method and other signal processing methods for time series in more detail in Section 2.2.

The observation baseline is the total time extent of observations of an astrophysical object. Often the observational baseline of a certain object is defined by celestial geometry: if we cannot point a telescope at an object as it is blocked by another body such as the Earth or the

Sun, we will be unable to observe it. Additionally, telescopes will have an operational lifetime, so for example, in the case of the space telescope Kepler (Borucki et al. 2010), two failed reaction wheels meant the telescope was unable to continue its ground-breaking long-baseline observations of a patch of the sky (now known as the Kepler field, which Kepler observed almost continuously from 2009–2013).

I will discuss three common observational methods in time-domain astronomy relevant to this thesis: photometry, spectroscopy and astrometry. The most detailed description will be of photometry, the method with which the data used in this work is taken. Spectroscopy is used widely within exoplanet and variability detection, providing different insights into variability through observations of shifting stellar spectra. Astrometry is an important technique used for mapping the positions and movements of objects in the sky and is discussed here in the context of the Gaia Mission (Gaia Collaboration et al. 2016). Gaia is an ongoing space-based mission to create an extraordinarily precise three-dimensional positional map of more than a thousand million stars throughout our Milky Way galaxy and beyond, mapping their positions, motions, luminosities, temperatures and compositions.

### 1.1.1 Photometry

Photometry is the science of measuring an object's brightness in a specific part of the electromagnetic spectrum. It is most often conducted by measuring the electromagnetic flux (i.e. photons) incident on an imaging device such as a Charge Coupled Device (CCD). The incident light is passed through a filter which allows only photons within a specific wavelength range to pass through, commonly this is within the visible spectrum due to the widespread availability of such CCDs. There are several standard astronomical filters used within the infrared, visible and ultraviolet parts of the electromagnetic spectrum, one example is the Johnson–Cousins UBVRI photometric system (for example as described in Landolt 2007). These wavelength bands split this part of the spectrum into **U**ltraviolet, **B**lue, **V**isual, **R**ed and **I**nfrared and are motivated by observations of standard stars from the Earth, accounting for absorption features in the Earth's atmospheric spectrum. By taking photometric measurements in multiple passbands, it is possible to calculate colour information about a source. This colour information can be used to infer temperature and aid in identifying specific classes of variable stars and binary systems.

By tracking the brightness of a source over time, we can produce a *photometric light curve* for a star. In general, to produce a light curve, we must track an object's movement across our CCD and account for many sources of photometric noise, including background brightness fluctuations. If we observe from the ground, there are many additional noise sources caused by Earth's atmosphere and the Moon. I will discuss these noise sources in more detail in Section 1.2.

Photometric light curves can reveal information about the source and any orbiting bodies. *Asteroseismology*, or the study of stellar oscillations, uses the frequency spectrum of a photometric light curve to observe the oscillation modes of a star and gain insight into its internal structure. Photometry also reveals information about the rotation and rotational evolution of a star. I will give a detailed description of different sources of stellar variability in Section 2.1. If objects pass in front of the star being observed, most interestingly companion stars or orbiting planets, a photometric light curve will also reveal these. Further details of the transit method will be outlined below.

#### 1.1.1.1 How do we measure brightness?

Astronomers traditionally use magnitude to express the brightness of astrophysical objects. Magnitude is a logarithmic brightness scale defined as:

$$m - m_{\rm ref} = -2.5 \log_{10} \frac{F}{F_{\rm ref}}, \tag{1.1}$$

where $m$ is the *apparent magnitude* of a source as observed from Earth and $m_{\rm ref}$ is the apparent magnitude of a suitable reference source. Here $F$ is the total incident flux of the detected source, and $F_{\rm ref}$ is the total incident flux of the reference source. For a given photometric filter, there exists a zero-point reference flux $F_{\rm ref}$ for reference magnitude $m_{\rm ref} = 0$. Although apparent magnitude is a measurable quantity and useful for comparing the brightness of objects as observed from Earth, the *absolute magnitude* allows the comparison of the intrinsic brightness of objects.

An object's absolute magnitude is defined to be equal to the apparent magnitude that the object would have if it were viewed from a distance of exactly 10 parsec, without extinction of its light due to absorption by interstellar matter and cosmic dust. If we know the distance to a source in parsec ($d_{\rm pc}$) (for example, from measuring the *parallax*), we can calculate the absolute magnitude, $M$, as

$$M = m - 5 \log_{10}(d_{\rm pc}) + 5 - A, \tag{1.2}$$

where $A$ is the extinction or reddening of the object in this band. As light scatters off dust and other matter in the interstellar medium, shorter-wavelength light from the emitted spectrum is preferentially absorbed or scattered due to the average size of interstellar dust grains, leaving a 'reddened' spectrum. Reddening is often measured as a 'colour-excess' by comparing photometric observations in multiple passbands of an object against model photometry. A commonly used colour excess is $E(B-V)$, which is related to the observed objects $B-V$ colour: $E(B-V) = (B-V)_{\rm observed} - (B-V)_{\rm intrinsic}$.

Although the true relationship between reddening and extinction has a complex wavelength dependence, Cardelli et al. (1989) demonstrated a simple relation which holds for a wavelength

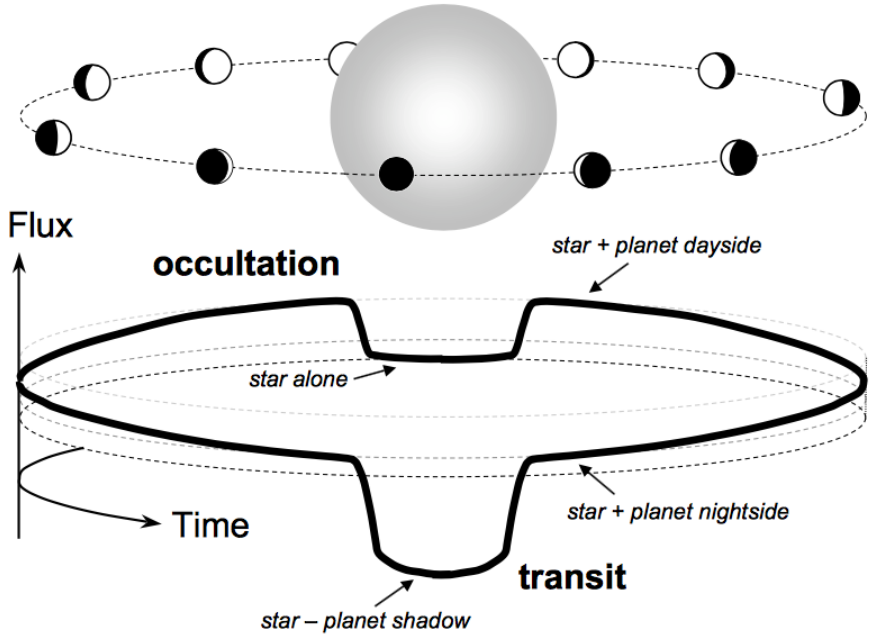



Figure 1.1: An illustration of transits and occultations. The flux drops as the planet blocks a fraction of the starlight during transit. The flux rises as the planet's dayside comes into view. The flux drops again when the planet is occulted by the star. Credit: Winn (2010).

range $0.125\mu m \leq \lambda \leq 3.5\mu m$ involving just one parameter, the total-to-selective extinction ratio: $R_V = A_V/E(B-V)$. Furthermore, Cardelli et al. (1989) showed that $R_V \sim 3.1$ characterises the mean extinction relation for stars in the Milky Way. This value is still widely accepted to model the Galaxy's extinction relation well.

#### 1.1.1.2 The transit method for exoplanet detection

By taking photometric light curves, it is possible to detect orbiting bodies which transit their parent. This is only possible when the system's orbital plane aligns with the observer's line of sight; however, it has been exceptionally successful in aiding the discovery of exoplanets and binary star systems. The first exoplanet to be discovered using the transit technique was HD 209458b by Charbonneau et al. (1999). Since then, the transit technique has been the most successful exoplanet discovery method, with almost 3,500 confirmed planets published using this technique (over 75% of all confirmed planets) (NExSci 2021). Of these, just over 2,400 were detected using the Kepler space telescope (Borucki et al. 2010), the most successful planet-hunting mission to date. In Section 1.2, I will discuss further details of Kepler and other large photometric survey missions.

The transit technique allows insight into the radius ratio of the two orbiting bodies, as

the expected 'dip' in the light curve during the transit can be modelled with a small number of parameters. To first order, transit depth (the fractional change in brightness) is equal to the radius ratio squared: $R_p^2/R_*^2$ for a planet of radius $R_p$ orbiting a star of radius $R_*$. The shape of this dip is shown in Figure 1.1. This technique, as mentioned, is also suitable for the detection and characterisation of binary star systems; Section 2.1.3 outlines further detail on the photometric detection of binary star systems.

### 1.1.2 Spectroscopy

Spectroscopy utilises the multi-wavelength nature of detected light. The light's wavelength spectrum is spread out and measured on a CCD by passing the light through an optical dispersion device such as a diffraction grating. By analysing measured spectra, it is possible to gain insights into the chemical composition of the observed target. In particular, when considering stars, their spectra can reveal chemical composition and aid in understanding temperature, density, mass, distance, and luminosity.

#### 1.1.2.1 Radial velocity (RV) spectroscopy

One notable and relevant use of spectroscopy is the detection of radial velocity (RV) signals. RV signals are measured by periodically monitoring spectra of a target star and tracking the change in the wavelength of spectral lines resulting from a Doppler shift. When considering observations of stars with orbiting bodies, an RV Doppler shift signal will appear as a regular, cyclical motion in wavelength as the source orbits the system's centre of mass. The light from the star is blue- and red-shifted as it moves towards and away from the observer, respectively. This motion implies the presence of a companion object exerting a gravitational pull on the star: either another star or a large or close-in exoplanet.

The RV technique allows us to extract information on the mass ratio of the two bodies (subject to an unknown inclination factor). Combined with the transit technique described in Section 1.1.1.2 enables an understanding of the two orbiting bodies' mass ratio and radius ratio. The discovery of the first exoplanet orbiting a main-sequence star 51 Pegasi b by Mayor & Queloz (1995) was made using the RV technique. Since then, RV has proven to be an extremely successful method of detecting exoplanets, with almost 900 confirmed planets published using this technique (about 20% of all confirmed planets) (NExSci 2021).

There have been vast improvements in the range and sensitivity of RV spectrographs over the last 20 years, with ELODIE (Baranne et al. 1996) and CORALIE (Queloz et al. 2000) beginning the era of high-resolution spectrographs. More recently, HARPS (Mayor et al. 2003) broke the precision barrier of 1 m $s^{-1}$, and the ESPRESSO spectrograph installed on the Very

Large Telescope (VLT) aims to further increase RV precision by an order of magnitude to 10 cm $s^{-1}$ (Pepe et al. 2021). Such instruments may enable RV detections of Earth-sized planets orbiting Sun-like stars.

The RV spectrometer installed on the space mission Gaia (Gaia Collaboration et al. 2016) has several purposes. The astrometry conducted with Gaia will be discussed in Section 1.1.3, and RV measurements provide additional insight into the kinematics of stars. RV measurements can also be used to detect unresolved binary systems, where two stars appear as one light source, but periodic RV measurements indicate the presence of a companion star. The instrument aboard Gaia is not as high-resolution as some of the ground-based spectrographs; however, this is not of concern for detecting similar-mass binary systems and large kinematic RVs.

### 1.1.3 Astrometry

Astrometry involves precise measurements of stars' locations in the sky. There are a few scientific objectives behind this. Firstly, to provide a stellar reference frame to which the motions of astrophysical objects may be referred. Secondly, to provide a fixed catalogue of astrophysical objects, including spatial distribution, motion and often basic stellar properties such as luminosity and mass. It is also possible to detect binary systems and even extremely large, wide-orbit exoplanets through astrometry, though astrometry is not a commonly used exoplanet detection method. As of October 2021, just one planet, DENIS-P J082303.1-491201, is listed on the exoplanet archive[a] (Sahlmann et al. 2013).

One of the most famous astrometric targets of interest is the binary star system 61-Cygni. As early as 1804, astronomers took notice of 61-Cygni's large proper motion. Not much later, in 1838, the first stellar parallax was measured for the system, giving a distance estimate of 3.2 pc, which is very close to the more recently measured value of ~ 3.5 pc (Hopkins 1916).

Astrometry, in principle, sounds straightforward. However, we must consider that due to the distances involved, we only observe a 2D map of 3D space and must infer distance and intrinsic brightness from the relative brightness and relative motion of the sources we observe.

The standard astrometric model contains six parameters, for example, as defined in Klioner (2003) and Lindegren et al. (2012):

- the right ascension $\alpha$ and declination $\delta$ define the position in some predefined coordinate system, measured as angles in degrees or radians;
- the components of the proper motion in right ascension $\mu_{\alpha^*}$ and in declination $\mu_\delta$ are the time derivatives of the barycentric coordinates, often measured in milli-arcseconds (mas) per year;

[a] https://exoplanetarchive.ipac.caltech.edu/overview/DENIS-P%20J082303.1-491201%20b#planet_DENIS-P-J082303-1-491201-b_collapsible. Accessed: 22/01/2022.

- the parallax $\varpi$, which is a proxy for the inverse of the distance to the source, measured in mas;
- and the radial velocity of the source $\mu_r$, which can be measured in mas per year or km/s.

In general astrometric surveys will only fit the first five parameters, as the radial velocity can also be measured through spectroscopy. Right ascension (RA, $\alpha$) and declination (dec, $\delta$) are tied to the coordinate system of choice, several of which will be detailed in Section 1.1.3.1.

Decoupling the motions of an object requires a fit to the positions of the object taken over time. The apparent position of this object will be affected by the motion of the object itself and the motion of the observer. The observer's motion can be useful here: to measure the parallax as a proxy for distance, observing an object from multiple angles and mapping its motion will allow a parallax measurement to be taken. I note here that, broadly, parallax is only a good proxy for distance for nearby objects. A series of papers by C.A. Bailer–Jones starting with Bailer-Jones (2015) demonstrate that for distant objects with large errors in their measured parallax, a prior assumption on the distribution of distances is required to infer a distance from a parallax measurement. Most recently, Bailer-Jones et al. (2018) estimates distances for 1.33 billion stars with parallax measurements from the Gaia survey's second data release (Gaia Collaboration et al. 2018c) and EDR3 (Bailer-Jones et al. 2021). I will use the catalogue from Bailer-Jones et al. (2018) to gauge the distances to stars in Chapter 5 of this work.

#### 1.1.3.1 A brief introduction to astronomical coordinate systems

Astronomical coordinate systems are well-defined customs for specifying positions of astrophysical objects relative to physical reference points available to an observer. All coordinate systems will include definitions of a fundamental plane (at 0° latitude) and a centre point from which a pole extends toward another reference point. There will be a primary direction to define 0° longitude, and a set of coordinates defines a point on a sphere in this spherical coordinate system.

*Equatorial coordinates* are defined with a centre point at the centre of the Earth in *geocentric* definitions and the centre of the Sun in *heliocentric* definitions. The celestial plane and the celestial poles define this system. The celestial equator is at the Earth's equator, and the celestial pole runs directly north-south through the centre of the Earth, as shown in Figure 1.2. Declination ($\delta$) and Right ascension ($\alpha$) or sometimes hour angle ($h$) are used in this coordinate system as latitude and longitude.

*Ecliptic coordinates* are fairly similar to equatorial but use the plane of Earth's orbit around the Sun as the fundamental plane. This system defines Ecliptic latitude ($\beta$) and ecliptic longitude ($\lambda$). In both of the above systems, the primary direction at 0° longitude (or RA) is defined when the Sun is at the March Equinox.

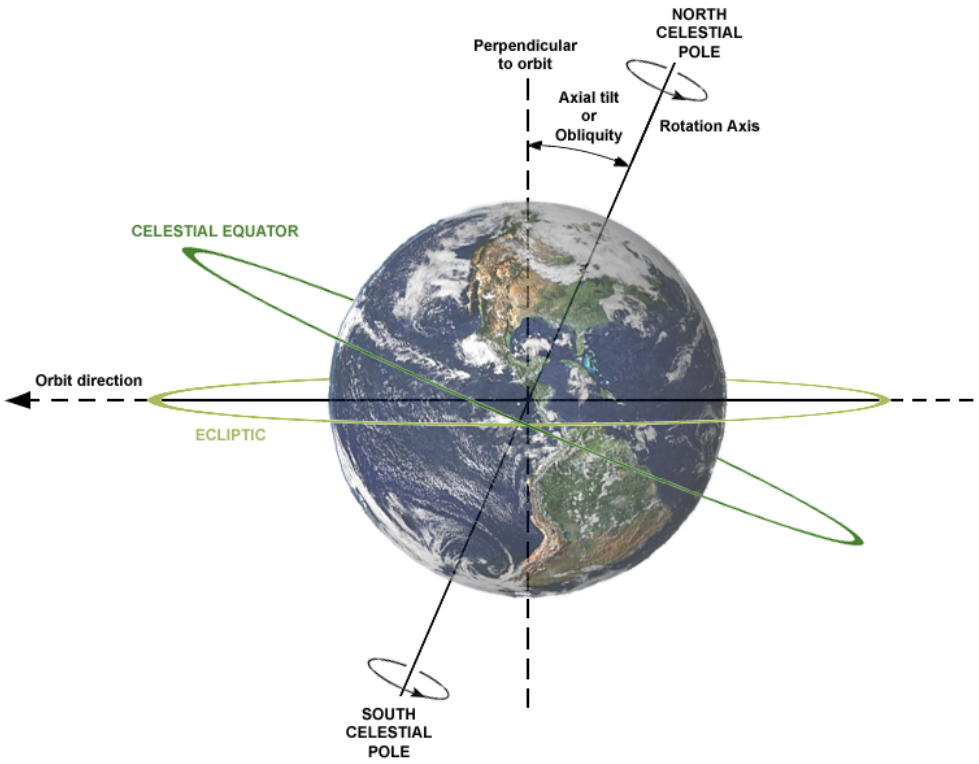



Figure 1.2: A visual description of the relations between the celestial and ecliptic equators. Credit: Dennis Nilsson, licensed under Creative Common Attribution 3.0 Unported license.

*Galactic coordinates* are always defined with the centre point at the centre of the Sun. The fundamental plane is the plane of the Milky Way, with the pole as the galactic pole. The aptly named galactic latitude and longitude $(b, l)$ are used in this system. The primary direction is towards the galactic centre.

A further refinement of these coordinate systems is the often used *International Celestial Reference System and Frame, ICRS* (Feissel & Mignard 1998). The origin lies at the barycentre of the Solar System, with an axis 'fixed' with respect to the stars. This definition allows the most appropriate coordinate system for defining reference positions and motions of astrophysical objects. In practice, the coordinate system is fairly similar to equatorial coordinates but uses positions of fixed quasars in the sky to define the pole and primary direction to provide a more static reference frame compared with astrophysical objects.

#### 1.1.3.2 Astrometric surveys

Astrometric catalogues and surveys date back to as early as 275 BC when the Greek astronomer Hipparchus measured and mapped the positions of the constellations and the celestial equator with the positions of solstices and equinoxes. More recently, international efforts have been

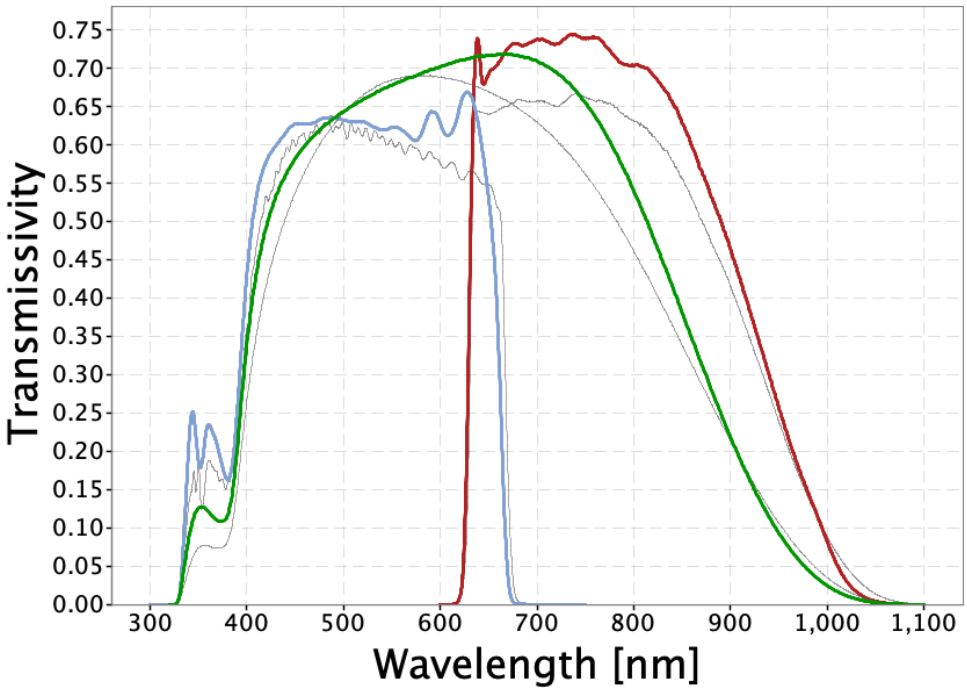



Figure 1.3: Gaia EDR3 passband transmissivities. The coloured lines in the figure show the Gaia $G$ (green), $G_{\rm BP}$ (blue) and $G_{\rm RP}$ (red) passbands, defining the Gaia EDR3 photometric system. The thin, grey lines show the nominal, pre-launch passbands used for Gaia DR1. Credits: ESA/Gaia/DPAC, P. Montegriffo, F. De Angeli, M. Bellazzini, E. Pancino, C. Cacciari, D. W. Evans, and CU5/PhotPipe team.

made to map the positions and motions of as many stars in the sky as possible. The first was the ESA Hipparcos Space Astrometry Mission and associated catalogue (ESA 1997; Perryman et al. 1997). The catalogue published positions, parallaxes and proper motions for about 100,000 stars with an accuracy of 0.7–0.9 mas for stars brighter than 9 mag, unprecedented at the time. The latest version of the catalogue, Tycho-2 (Høg et al. 2000), provides positions and proper motions for the 2.5 million brightest stars in the sky. Until the launch of Gaia in 2013, the Hipparcos / Tycho-2 catalogue was the de-facto reference catalogue for astrometric data.

Gaia (Gaia Collaboration et al. 2016) is a space telescope designed to produce astrometric solutions and two-colour photometry and radial-velocity spectrographs for as many sources as possible. Following its launch in 2013, there have been two full data releases (DR1 Brown et al. 2016) and (DR2 Brown et al. 2018), as well as an early data release 3 (EDR3) (Gaia Collaboration et al. 2021) with improved precision on DR2. The full Gaia DR3 is expected in 2022, with many new sources and derived properties, including radial velocities, an extended catalogue of variable stars and the first catalogue of binary stars.

Gaia takes photometric measurements of stars in three separate bandpasses: Gaia $G$, $G_{\rm BP}$ and $G_{\rm RP}$. The wavelength dependencies of these passbands are shown in Figure 1.3; roughly

speaking, the $G_{\rm BP}$ filter is bluer than the $G_{\rm RP}$ filter, and the $G$ filter spans both, although with less power at the extreme blue/red colours. It is important to note that the passbands are internally calibrated based on each data release cycle. So the magnitude of an object in the same nominal passband will be slightly different between data releases.

Most of the work in this thesis utilises the astrometric parameters from the second Gaia data release, with the final Chapter utilising the updated parameters from EDR3. This catalogue features five-parameter astrometric solutions for more than 1.3 billion sources with a magnitude range of $3 < G < 21$. Additionally, there are $G_{\rm BP}$ and $G_{\rm RP}$ colour magnitudes for more than 1.38 billion sources, allowing insight into the colour of these objects. Of interest to this work, 550,000 variable sources have been detected from the photometric telescope on Gaia. As the survey is firstly astrometric, the precision and cadence of the photometry are not optimised for detecting variability. Thanks to a large number of sources, a wide range of stellar variability signals have been detected: a number of these variable sources have also been observed and detected with NGTS; Chapter 5 will compare these detections.

## 1.2 Photometric surveys

I have introduced the science behind photometric brightness monitoring of stars and some scientific goals for doing so. This section will introduce some telescopes and surveys that utilise photometry to detect and characterise millions of astrophysical objects of interest, both on the ground and in space. I will give full details of the implementations and data structures used in generating photometric light curves for millions of objects from a set of images in Chapter 3 whilst introducing the Next Generation Transit Survey (NGTS) and the data used in this work.

A major improvement in the precision of photometric astronomy comes in the form of space-based telescopes. Observing from space comes with many improvements over ground-based observations, including the lack of atmospheric noise and no day-night cycle causing gaps in observation. Although we can improve precision and combat noise by observing from space, the size of the instrument is limited by rocket launch size constraints and budget.

### 1.2.1 Ground-based

The first generation of ground-based photometric surveys began operation in the early 2000s, with photometric precision of order 10 mmag (0.01 magnitude), which allowed exoplanet detections of Jupiter-sized planets orbiting bright stars (typically V-band magnitude < 12). These included the Transatlantic Exoplanet Survey (TrES, Alonso et al. 2007), XO (McCullough et al. 2005), the Hungarian-made Automated Telescope Network (HATNet, Bakos et al. 2002) and

the Wide-Angle Search for Planets (WASP Pollacco et al. 2006). WASP employed extremely wide-angle telescopes, allowing simultaneous monitoring of many photometric targets simultaneously. The SuperWASP system employs eight 0.11 m telescopes at two observatories (SuperWASP-North is in the Canary Islands and SuperWASP-South in South Africa). Each telescope array provides a field of view of 482 $\deg^2$. The first WASP data release (Butters et al. 2010) contained 3,631,972 raw images and 17,970,937 light curves taken across the two telescope arrays between 2004 and 2008.

The next generation of ground-based surveys includes the aptly-named Next Generation Transit Survey (NGTS, Wheatley et al. 2018), as well as a few other wide-angle photometric surveys, including the Qatar Exoplanet Survey (QES, Alsubai et al. 2013) and the Kilodegree Extremely Little Telescope (KELT, Pepper et al. 2007). This generation of wide-angle surveys boasts an order of magnitude increase in photometric precision, with NGTS able to reach 1 mmag precision across its 100 $\deg^2$ field-of-view with an extremely fast 12-second sampling cadence.

Further improvements to photometric instruments have given rise to targeted surveys such as MEarth (Irwin et al. 2014), as well as the TRAnsiting Planets and PlanetesImals Small Telescope (TRAPPIST, Gillon et al. 2011) and its successor the Search for habitable Planets EClipsing ULtra-cOOl Stars (SPECULOOS, Burdanov et al. 2018). These surveys select candidate objects to observe rather than a wide-field approach. This can often lead to higher precision, as well as being able to optimise the instrument for the type of object being observed. In particular, the three surveys mentioned above focus on M-dwarf stars, a class of stars much smaller and fainter than the Sun, around which it is possible to detect much smaller (and potentially habitable) planets.

### 1.2.2 Space-based

The first notable space-based photometric survey was COnvection, ROtation and planetary Transits or CoRoT (Auvergne et al. 2009), which launched in 2006 with the goals of detecting transiting rocky planets and performing asteroseismology. CoRoT ran until 2012, during which it observed thousands of target objects, with around 150 bright asteroseismic targets and 24 confirmed planet detections.

In 2009, NASA's Kepler (Borucki et al. 2010) was launched. Kepler's primary science mission was to take continuous photometric observations of approximately 150,000 main-sequence stars within a fixed field of view. This mission has led to one of the most widely used publicly available photometric data sets, with thousands of planet discoveries (and many more still to be confirmed) (NExSci 2021)[b] as well as large asteroseismological surveys (Yu et al.

[b] https://exoplanetarchive.ipac.caltech.edu/docs/counts_detail.html. Accessed: 22/01/2022.

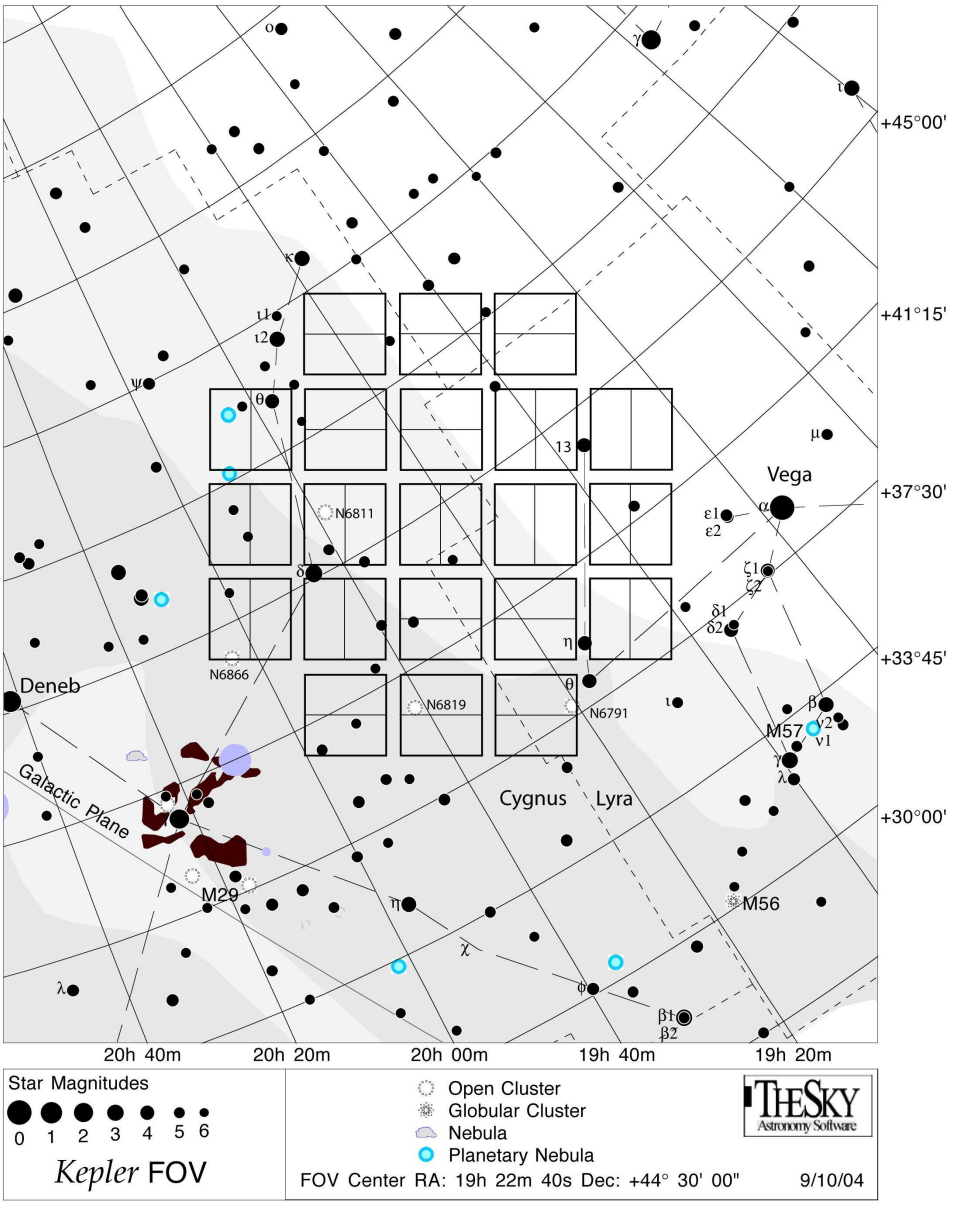



Figure 1.4: Kepler's Field of View is shown as a series of squares, each square maps to a set of two CCDs on Kepler's detector. Nearby astronomical objects of interest are shown and labelled. Credit: NASA, taken from https://www.nasa.gov/mission_pages/kepler/multimedia/images/fov-kepler-drawing.html. Accessed: 22/01/2022.

2018) and rotational analyses (McQuillan et al. 2014). I will further discuss the stellar rotation data aspect of the Kepler data set in Chapter 5.

Kepler's field of view was chosen based on its continuously observable position, away from the ecliptic plane and with an appropriate stellar density to accurately resolve as many sources as possible. Kepler looks towards The Cygnus–Lyra region in the northern sky, the FoV is centred on RA = $19^{h}22^{m}40^{s}$ and Dec = $+44°30'00''$ (J2000), and is around 15° across (115 $deg^2$). The FoV is shown in Figure 1.4. The telescope was rotated by 90° every 90 days to keep the solar panels pointing at the Sun; this means the Kepler data is divided into 90-day quarters, of which there were 17 between 2009 and 2013. The full data set includes up to 3.5 years of continuous observation comprised of short (1-minute) and long (30-minute) cadence light curves for ~160,000 stars. Kepler observed with a broadband filter (covering a wavelength range of ~ 400 to 865 nm), which maximises the sensitivity of the telescope and detector combination for detecting planets transiting main-sequence solar-type stars.

After the failure of a second reaction wheel (of four) in 2013, the telescope was unable to remain fixed on the aforementioned 'Kepler field', so a modified observation programme called K2 began (Howell et al. 2014). Although the observation baseline of these fields was much shorter than the original Kepler field, the mission was able to detect many short-period planets and provide useful insight into stellar variability at different pointings (Chaplin et al. 2015; Gordon et al. 2021).

NASA's current flagship space photometry mission is the Transiting Exoplanet Survey Satellite (TESS, Ricker et al. 2014) that launched in 2018. TESS took a different observational approach to Kepler, observing almost the entire sky over its primary 2-year mission whilst monitoring 200,000 of the nearest and brightest stars. TESS will observe most of its fields for 28 days; however, overlapping regions (as shown in Figure 1.5) will offer longer periods of continuous observation suitable for detecting longer period variability and shallower transits. In the extended mission, TESS has re-observed many fields observed in the first and second years of observation and has also observed several fields within the ecliptic plane that overlap with K2 campaign fields.

TESS provides data with different time cadences. 20-second cadence data is provided for a small number of targets (~ 600 per sector) in the extended mission. 'Postage stamp' regions defined around bright asteroseismological targets are downloaded at a 20-second cadence in both the primary and extended mission (approximately 16,000 targets per sector). The remaining pixels not included in these regions are downloaded as 30-minute/10-minute cadence full-frame images in the primary and extended missions, respectively[c]. TESS's bandpass has been constructed to be more sensitive to M-dwarfs than Kepler, which focussed on main-sequence

[c]Taken from https://heasarc.gsfc.nasa.gov/docs/tess/faq.html. Accessed: 22/01/2022

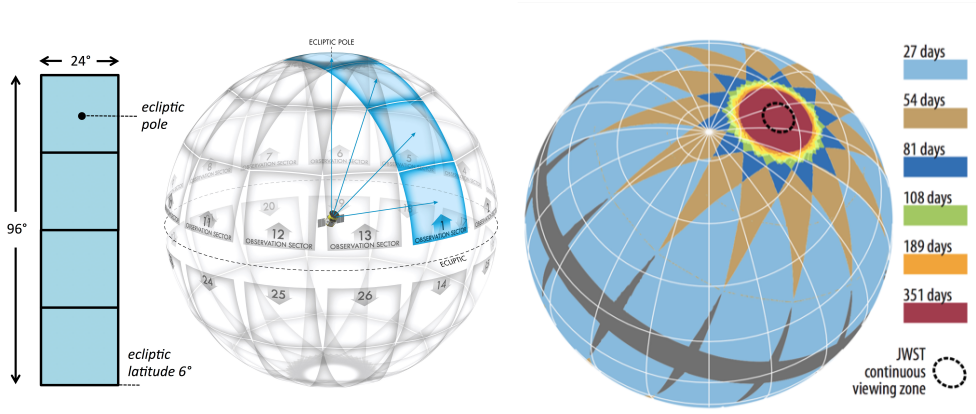



Figure 1.5: Schematic of the TESS observing pattern demonstrates the total observation length in each sector over the 2-year primary mission. Credit: (Ricker et al. 2014).

stars. The TESS bandpass is redder than that of Kepler, covering a wavelength range of 600–1000 nm, which allows much more precise measurements of cool stars.

The next generation of space-based wide-field photometric surveys will come in the form of the PLATO mission (Rauer et al. 2014), which is due to launch in 2026. The primary mission goal of PLATO is to detect and characterise planets transiting one million stars, focusing particularly on dwarf stars. Similar to some of the recent ground-based photometric surveys, PLATO employs an array of 26 cameras which allows a huge 1,100 deg$^2$ field of view with up to 2.5-second cadence on its fastest cameras.

### 1.2.3 Common noise sources in photometry

Errors and noise are an unfortunate part of any astronomical observations and can arise from astrophysical and non-astrophysical sources. Non-astrophysical sources can be driven by the physics of detectors or from engineering aspects of a telescope. Additionally, observations from the ground will add further noise in the form of atmospheric noise and background light not present in space-based observations. Astrophysical noise sources are nuisance signals arising from astrophysical origins and therefore depend on the scientific goal of your analysis. For example, in the case of exoplanet transit detection, stellar variability would be considered a source of noise, whereas, for stellar variability detection, transiting planets could be viewed as noise. Noise can be classified as correlated or uncorrelated; generally, uncorrelated or white noise is more easily dealt with through binning or averaging. Correlated, systematic or red noise requires complex treatment through statistical modelling; the covariance between light curve data points arising from noise sources can have timescales similar to signals of interest. For example, in the case of planetary transit detection, Pont et al. (2006) demonstrates correlated noise from ground-based surveys can drastically reduce the expected yield of transiting planets.

More recently, Gaussian process regression has been shown to accurately model astrophysical and correlated systematic noise from K2 mission light curves (Aigrain et al. 2015).

Here I will briefly outline some of the more common sources of noise. I will give details of how NGTS deals with several of these noise sources in Chapter 3.

With any form of astronomy, there exists a fundamental limit to the photometric precision in the form of photon noise. The quantum nature of photons drives photon noise: when a detector measures incident light, photons will hit the detector at time intervals dictated by a Poisson distribution. Photon noise is often referred to as Poisson or shot noise. A Poisson distribution has the property that variance is equal to expectation $E[N] = \mathrm{Var}[N] = \lambda t$, where $\lambda$ is the incident number of photons in a time interval $t$, and so the photon noise $\sigma_{\mathrm{photon}} = \sqrt{N}$. The signal-to-noise ratio, $N/\sigma = \sqrt{N}$, can be improved by receiving more photons, i.e. observing a brighter target or using a longer observation window. In the limit of large numbers of photons, photon noise is well modelled by a Normal distribution $N \sim \mathcal{N}(\lambda t, \lambda t)$, and hence is a white noise term.

Within the telescope, the detector can introduce systematics due to the nature of semiconductors. Charge-Coupled Devices (CCDs) are the most common detectors used in photometric instruments. The process of converting incident photons into scientific images provides a myriad of errors not limited to *quantisation noise*, *readout noise*, *dark current*, *pixel inhomogeneity* and *saturation overspill* (Gary 2007). Beyond the detector, the telescope aperture and dust may cause errors in the form of *flat-field errors*, and the *autoguiding* method used to track individual sources is also prone to causing errors through imperfect source tracking. Source tracking will require correction for Earth's rotation in ground-based missions. In contrast, for space-based missions, the correction can be extremely complex and will be dependent on the orbit and rotation of the satellite (Gary 2007).

When observing from the ground, the atmosphere interacts with the incoming light from astronomical sources. In particular, the atmosphere will cause *scintillation noise* due to varying refractive index through the thickness of the atmosphere. This causes stars in the sky to 'twinkle' as observed by eye. For a telescope, this will cause blurring of images and brightness fluctuations, which are both large sources of noise in photometric data. Additional complications come from scattering from molecules in the atmosphere, causing *differential extinction*, as different wavelengths of light are subject to different extinction levels. This causes sources with different colours to appear at different brightness depending on the scattering through the atmosphere. Finally, light pollution is a big problem for ground-based surveys, as diffuse light from populated areas and airglow and scattered light from the Moon will appear as noise to a detector. As discussed in Chapter 5, this sort of background light proves to be a large hurdle to overcome to conduct stellar variability studies with the NGTS data. It is possible to reduce these

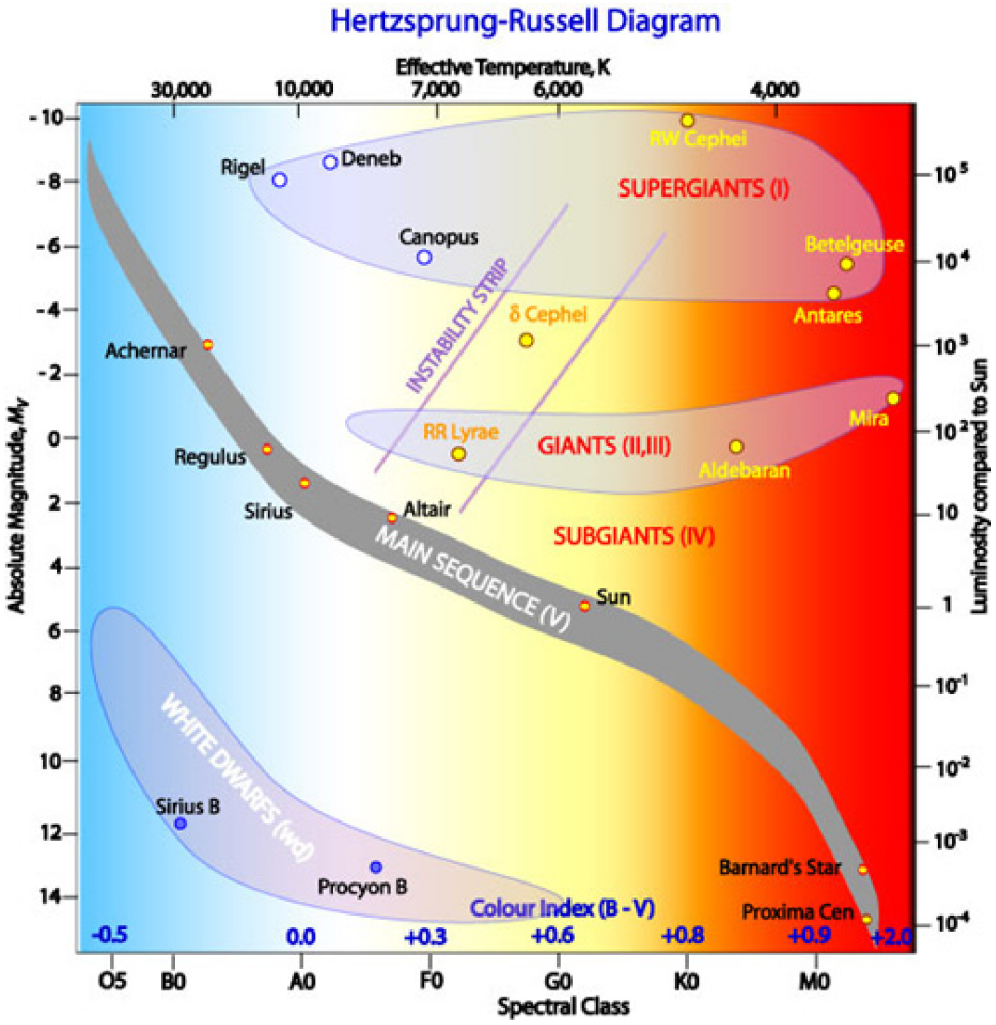



Figure 1.6: The Hertzsprung–Russell (HR) colour–magnitude diagram. Annotations for various stages of stellar evolution are included. A prominent feature is the main sequence (dark grey), which runs from the upper left (hot, luminous stars) to the lower right (cool, faint stars) of the diagram. Giant and supergiant stars lie above the main sequence, with the instability strip shown perpendicular to the main sequence. Credit: R. Hollow, CSIRO.

atmospheric effects by selecting a good telescope site. An ideal observing site will be far from any light pollution, with as little atmosphere as possible. Such sites are often in high, desert locations such as the ESO site at La Silla, Paranal in Chile, where NGTS and SPECULOOS are located. Further telescope design considerations such as large baffles can also limit stray Moonlight.

## 1.3 Stars and the HR diagram

The Hertzprung–Russell diagram (HR diagram) is an extremely important tool to aid in classifying stars and studying stellar evolution. The HR diagram is similar to a colour–magnitude diagram but plots theoretical quantities effective temperature $T_{\text{eff}}$ against luminosity $L$ rather than observable quantities colour (such as $B - V$ or $G_{BP} - G_{RP}$) against magnitude. The HR has a clear grouping of regions that map closely to the evolution throughout a star's life. This grouping – and thus ageing – is directly related to the physical processes stars of a given mass are undergoing. An example HR diagram is shown in Figure 1.6, with annotations indicating distinct regions of interest in colour–magnitude space. I will briefly introduce the evolution of stars of different masses and how their position on the HR diagram changes over their lifetimes. For a more detailed description of the formation and evolution of stars, I refer the reader to Stahler & Palla (2004).

A star is formed due to the gravitational collapse of gas and dust. As this matter collapses, it heats up, forming a protostar. Depending on the mass of the protostar, the evolution will follow a different track in HR space. Hayashi (1961) calculated these so-called 'Hayashi tracks' for protostars of different masses as the protostar evolves onto the Main Sequence. Low mass stars will evolve almost vertically down the HR diagram as they collapse isothermally towards the main sequence. As the protostar gains mass from the gas cloud surrounding it, it evolves into a pre-main-sequence (PMS) star. During the PMS, the star continues to collapse gravitationally and gains very little additional mass. Once the star reaches a certain threshold temperature and pressure, nuclear fusion begins at the centre of the star. Once hydrogen burning begins, the star joins the Main Sequence; hence, a star at this age is known as a Zero Age Main Sequence (ZAMS) star.

Around 90% of the stars in the universe are main sequence (MS) stars, fusing hydrogen to helium whilst maintaining hydrostatic equilibrium. The amount of time spent on the main sequence will depend entirely on the star's mass, with more massive stars burning through their hydrogen supplies much faster than smaller stars. Once a star has used its hydrogen supply, the star's mass will once again dictate how the star evolves from this point. For stars of solar mass and similar, once the hydrogen in its core is used up, fusion will cease, and the core will collapse under gravity. During this collapse, the star's outer layers may be pushed outwards due to the increased core temperature, and the star moves to the red-giant branch (RGB) on the HR diagram. Sub-solar stars, such as red-dwarf stars, burn hydrogen extremely slowly and can have main-sequence lifetimes longer than the age of the universe. As the core temperature rises, helium fusion may begin in the core, slowing the cooling of the outer layers of the star temporarily and moving the star to the Horizontal Branch. After this helium-burning phase

ceases, the star moves into the Asymptotic Giant Branch (AGB), which lies almost parallel to the RGB but at higher luminosity. Beyond this, the star will evolve along a post-AGB track to hotter temperatures with roughly constant luminosity. Eventually, the star will eject its outer layers, forming a white dwarf star at the bottom left of the HR diagram.

The extreme heat generated in the core of more massive stars will fuel the fusion of even higher mass elements, causing the star to swell in size. Due to the battling forces of gravity and radiation pressure, the star will pulsate. These stars lie along the instability strip; this strip contains many well-documented variable stars such as RR-Lyrae and Cepheid variables.

For extremely massive stars, the end of life produces cataclysmic variability such as core-collapse supernovae and can result in the formation of neutron stars and black holes in the case of the most massive stars.

### 1.3.1 Evolutionary tracks

As briefly discussed above, the evolution of a star is highly dependent on its mass, amongst other factors including chemical composition and the presence of circumstellar discs. Much work has been conducted on accurately modelling how stellar parameters evolve for stars of different masses and metallicities. Many of these models are a combination of several different underlying physics models, the details of which are beyond the scope of this thesis. These models offer insights into the behaviour of the interior and atmosphere of stars and consider both radiative and convective physics and the implications of magnetic activity (Feiden 2016) and rotation (Somers et al. 2020). Bressan et al. (2012) outline the input physics required for their well-used 'PARSEC' (PAdova and TRieste Stellar Evolution Code) models; I will briefly summarise some of the main points.

The chemical composition of the star affects a myriad of physical processes. One such measure of composition is metallicity, or the fraction of a star's mass not made up of hydrogen and helium. For example, the Sun comprises 73.81% hydrogen and 24.85% helium by mass, implying a solar metallicity of 0.0134 (Asplund et al. 2009). PARSEC models use more complex models of stellar chemical abundances, often taken from extremely precise modelling of the Sun, such as Caffau et al. (2010). The composition of a star's interior will affect the ability to absorb or transfer radiation (also known as the *opacity*). Once again, precise opacity models are employed within evolution codes to model internal radiative transfer. In conjunction with an equation of state (EOS), the opacity model effectively models the internal pressure and energy transport in the radiative interior. The models must also consider the nuclear reaction rates within the star. These are calculated from nuclear reaction networks, which consider the rates of many nuclear reactions possible within the star.

Convective regions of the star are often modelled using the mixing length theory (MLT) of convection (Cox & Giuli 1968), which describes convection as a simple local model. The gas is divided into rising and falling parcels of characteristic length $l$ (or sometimes $\alpha$), where $l$ defines the mixing length, or the length travelled by a parcel of gas before it is absorbed into the surrounding gas. This mixing length becomes the only free parameter in an MLT convection model. Hydrodynamical mixing instabilities or overshoots can occur at convective boundaries within the star (both in the core and the convective envelope). Internal overshoot from a convective core can have stark implications on the evolutionary properties of a star (Torres et al. 2014), and therefore is taken into consideration in most evolution models. Finally, evolution models must take macroscopic temperature gradients and microscopic diffusion of material into account.

The above physics is common to many evolutionary model codes, including the PARSEC models (Bressan et al. 2012), MESA (Modules for Experiments in Stellar Astrophysics, Paxton et al. 2010), the Dartmouth Stellar Evolution Database (Dotter et al. 2008), the $Y^2$ isochrones (Yi et al. 2001; Demarque et al. 2004) and the BCAH98 and BHAC15 models (Baraffe et al. 1998; Baraffe et al. 2015). Each model will use different underlying physics models and assumptions and a unique interpolation of masses and metallicities to produce tracks for input parameters. Hence, different models have different strengths in terms of the type of star they best approximate. In general, these models work best for solar-type stars, as many of the underlying physics models are calibrated to solar values, which are most easily observed. An example of a set of draws from the MESA model for stars of a range of sub-solar masses are shown in Figure 1.7. We can see the near-vertical Hayashi tracks as the isothermal collapse of stars onto the main sequence is modelled. By around 300 Myr, we see that the stars have converged onto the ZAMS. These models will also produce compositional information for the stars. For example, we see Deuterium and Lithium depletion occurring as the stars evolve towards the ZAMS; in Figure 1.7, the point at which these elements are depleted by a factor of 100 are shown along the tracks.

Complex magnetic fields within a star will cause additional changes to the star's evolution. These magnetic field effects are often ignored in stellar evolution models due to the high complexity and poor understanding of the physics involved due to limited observational capacity. Feiden (2016) adapts the Dartmouth stellar evolution models to include a magnetic field that interacts with the stellar plasma and is dependent on the internal structure. They demonstrate for a small selection of stars that the inclusion of this magnetic field resolves discrepancies in the reported ages and HR-diagram positions for K- and M-type stars in Upper Scorpius.

### 1.3.2 Isochrones

By generating multiple evolutionary tracks for different mass stars (generally with fixed metallicity), it is possible to generate a line of equal age across these tracks, an *isochrone*. Isochrones are an excellent tool for assessing where we would expect populations of similar ages to lie on the HR diagram. Isochrone fitting works especially well in the case of low-mass PMS stars as the isochrones lie almost parallel to the main-sequence but at higher luminosity. In Figure 1.7 (taken from Paxton et al. 2010), blue dashed lines join stars of different masses at the same evolutionary stage, i.e. isochrones. Such temporal snapshots can be seen in the CMD of young star clusters, which comprise populations of coeval stars with a range of masses.

## 1.4 Open star clusters

Clusters of stars within the Galactic disc are often referred to as open clusters. Open clusters are populations of stars that span a range of masses but possess essentially the same age and composition. Whilst observations of field stars provide insight into the wider stellar population, observations of populations of fixed age aid greatly in assessing how stellar properties vary with age. It is thought that stars within an open cluster were formed from the same giant molecular cloud and hence should have similar ages and metallicities. Several prominent open clusters have observations dating back thousands of years, such as the Pleiades, Hyades, or the Alpha Persei cluster, which are observable with the naked eye. As early as 1767, astronomers calculated that the observed star clusters must be physically related, as observing stars in such an alignment by chance was extremely small (Michell 1767). With more modern methods of astrometry and spectroscopy, it is possible to demonstrate that stars within a cluster have proper motions similar to the mean of the cluster and common radial velocities, such as within the Pleiades cluster (van Maanen 1945). Although most stars are observed in isolation, it is hypothesised that most stars form within clustered environments and spend parts of their early lives embedded in molecular clouds, gravitationally bound with other cluster members (for example Lada & Lada 2003; Zwart et al. 2010). Observable open clusters span a wide range of ages: from a few million years, with objects still exhibiting remnants of recent star formation, through to several gigayears (as old as the galactic disc). The co-eval stellar populations within open clusters are commonly used to provide snapshots of stellar evolution, and fitting them into an age-ranked succession and comparing with stellar evolution models affords an empirical understanding of the underlying phenomena. Observing clusters of similar age can also provide insight into inter-cluster variations of stellar properties. However, due to the imprecise nature of stellar age estimation, it is difficult to claim any significant differences in open cluster stars beyond the composition of the cluster (Fritzewski et al. 2020).

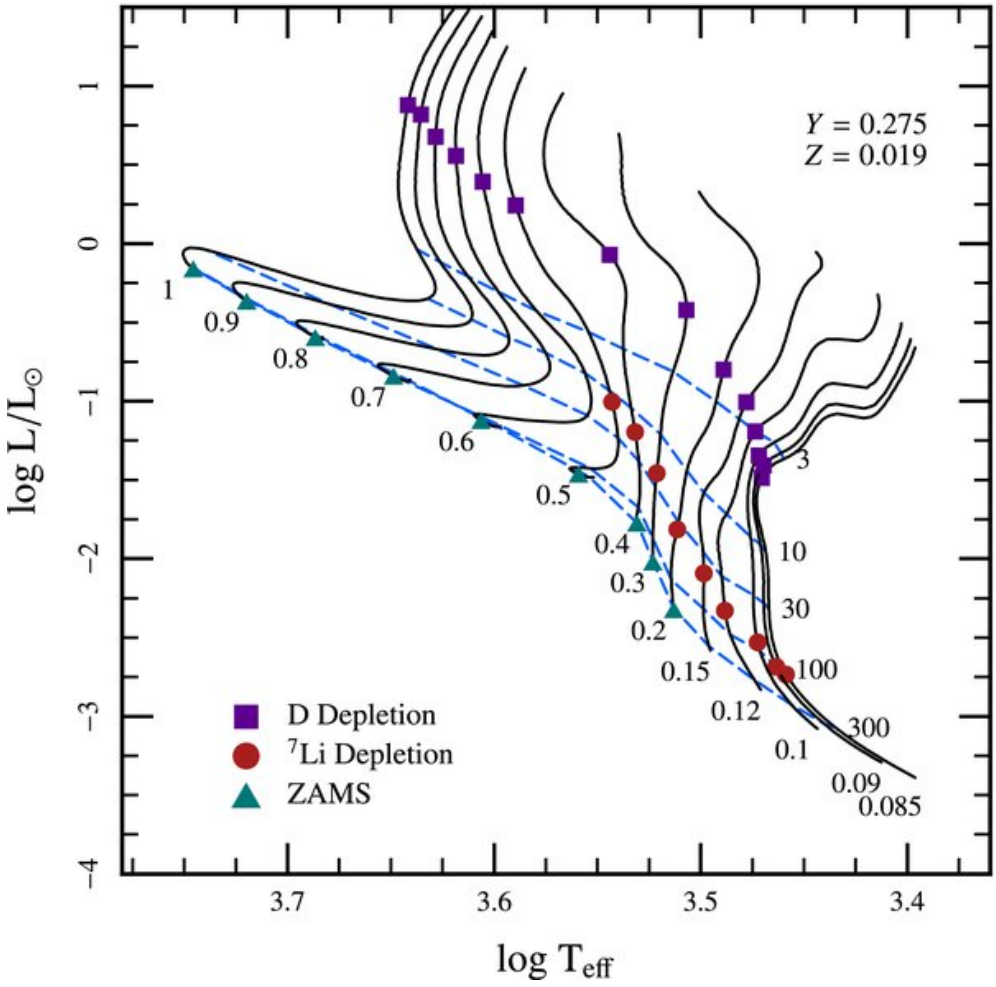



Figure 1.7: An example of some stellar evolution tracks from the Modules for Experiments in Stellar Astrophysics (MESA) stellar evolution code on the HR diagram. Each black line represents a stellar evolutionary track from PMS onto the ZAMS for stars of a different mass. The blue dashed lines correspond to points of equal age along each track (isochrones). The purple squares (red circles) show where D ($^7$Li) is depleted by a factor of 100 during the PMS. Credit: Paxton et al. (2010)

During star formation, rotational effects will influence a star's evolution in terms of structure, mixing and energy transport. Additionally, rotation gives rise to the stellar dynamo, which drives magnetic activity, including starspots and stellar winds (Henning et al. 2014). Although evolutionary models often do not account for the effects of rotation such as rotational mixing [d], empirical modelling and targeted observations of young objects and open clusters have aided with our understanding of the evolutionary effects of rotation.

### 1.4.1 Rotational evolution of stars

Young, PMS stars of solar mass and below have been shown to have rotation periods between 1 and 10 days long, with a bi-modal distribution of rotation periods with peaks at about 2 and 8 days (Herbst et al. 2001). This distribution has been confirmed through observations of dwarf stars in the Orion Nebula Cluster (ONC, Attridge & Herbst 1992; Choi et al. 1996; Herbst et al. 2001) and M dwarf members of the Pleiades cluster (Terndrup et al. 2000). Age estimates of these clusters give the ONC about 2 Myr (Palla & Stahler 1999), the Pleiades an age of 110–160 Myr and the Hyades $\sim$ 680 Myr (Gossage et al. 2018). The bi-modal distribution seen in very young stars has been attributed to the effects of star-disc interaction (Bouvier et al. 1997). Slower rotating stars are believed to have circumstellar discs still, whereas fast rotators have dissipated these discs and begin to spin up (i.e. decrease rotation period) to the ZAMS (Barnes 2003; Hennebelle et al. 2013). When the star has a circumstellar disc, the star's spin is regulated by angular momentum exchange with the disc, which locks the boundary of the magnetospheric cavity near the co-rotation radius, and hence maintains roughly constant angular velocity (Collier Cameron & Campbell 1993). As the star contracts gravitationally towards the ZAMS and the disc dissipates, the star's rotation rate will increase (Eggenberger 2013).

Studies of young open clusters have provided a wealth of information about the rotation periods of dwarf objects on the ZAMS. In particular, studies of the clusters NGC 2516 (Irwin et al. 2007; Fritzewski et al. 2020), M35 (Meibom et al. 2009), M50 (Irwin et al. 2009), the Pleiades (Hartman et al. 2010; Rebull et al. 2016b) and Blanco 1 (Cargile et al. 2014; Gillen et al. 2020) demonstrate these ZAMS clusters all show a universal rotation period distribution which appears to be age dependent. The existence of this age-dependent distribution lends strength to the idea of *gyrochronology*, accredited to Barnes (2003). Gyrochronology will be discussed in much detail throughout this thesis and is the theory that the rotation rate and the age of stars are linked: by observing the rotation rate of stars, one can infer the age.

[d] Some models do attempt to include the effects of rotation, such as Ekström et al. (2012) and recently Nguyen et al. (2022). Discussion of such models is beyond the scope of this work.

During the main-sequence, rotation rates of main-sequence stars have been observed to generally decrease with age. This spin-down appears to be mass dependent, with higher mass stars spinning down faster than lower mass M-dwarfs (Stauffer et al. 1987). This mass dependence agrees with the hypotheses of mass and angular momentum (AM) loss through magnetised stellar winds (Reiners & Mohanty 2012; Eggenberger et al. 2005) and the redistribution of AM throughout the stellar interior (Chaboyer et al. 1995; Lagarde et al. 2012; Charbonnel et al. 2013). Despite the (initial) scatter of rotation periods at the ZAMS, rotation rates appear to converge quickly. This is seen in observations of older star clusters such as M37 ($\sim$ 550 Myr; Hartman et al. 2009) and NGC 6811 ($\sim$ 1 Gyr; Meibom et al. 2011). Skumanich (1972) found a scaling relation $v \propto t^{-1/2}$ where $v$ is the average equatorial velocity, and $t$ is the age of the star, now known in the literature as the *Skumanich Law*. The Skumanich law is driven by the physics of the angular momentum of the solar wind, as described earlier by Weber & Davis (1967). Kawaler (1988) demonstrates that the Skumanich law can be explained by a surface angular momentum loss $\propto \omega^3$. Magnetised stellar winds are particularly prevalent in low-mass stars, where deep surface convection zones will drive magnetic fields and magnetised stellar winds. This appears to contradict the observation that higher mass stars spin down faster than low mass stars. However, Stauffer & Hartmann (1986), as well as Spada et al. (2016), hypothesise that the deep convective envelope in low-mass stars contains a higher proportion of the star's total AM than high-mass stars with shallow convective envelopes. This slows the spin-down rate in low mass stars, as the large AM of the convective zone resists the torque generated by the stellar wind. Additionally, this two-part model agrees with later observations of large numbers of rotating stars (e.g., McQuillan et al. 2014; Davenport & Covey 2018; Gordon et al. 2021; Briegal et al. 2022). If the convective envelope and the radiative core spin down separately, we expect the radiative core to continue to spin up due to contraction. Internal AM redistribution from the spun-up core to the spinning-down envelope could cause a reduction in this rotation period increase. This stalling of spin-down is a hypothesis for the observed rotation period gap seen in field stars by McQuillan et al. (2014), Davenport & Covey (2018), Gordon et al. (2021) and in the NGTS sample discussed later in this work.

Much less is known about the rotational evolution of post-MS stars into the RGB. Following AM loss through magnetised winds on the MS, the inflation of a star into the RGB and associated moment of inertia increase should result in a further slowing of the star's rotation. Previous spectroscopic studies show about 2% of observed giant stars exhibit rapid rotation (e.g., Fekel & Balachandran 1993; Massarotti et al. 2008; Carlberg et al. 2011) in disagreement with the expected slow down. These rapidly rotating stars may be due to two effects: companion star interactions or extremely massive stars without convective envelopes. A study by Ceillier et al. (2017) looked at the rotation rates of red giants observed with Kepler to aid understanding

of the discrepancy between theoretical and observed rotation rates for giant stars, as the two mechanisms described above should be more common than 2% of the population. This study found a similar rate (2.08%) of 'peculiar', i.e. rapidly rotating, giants. Whilst the authors claim this is in agreement with the expected rate of binary systems (e.g., Carlberg et al. 2011), for the most massive stars, they observe a lower rate of rotation detections than expected. They suggest that either the AM loss through winds is more than expected by Kawaler (1988), or that there is a substantial amount of radial differential rotation in these giants. Such complications sadly imply the lack of existence of a simple Skumanich-like scaling law for rotation periods in giant stars, which hinders the application of gyrochronology to giant stars.

### 1.4.2 Gyrochronology

Gyrochronology is a method for estimating the age of low-mass main-sequence stars from rotation periods. Empirical methods of stellar age estimation have existed for at least 20 years (such as Soderblom et al. 1991), however, the term 'gyrochronology' was coined in Barnes (2003). This theory draws on the observations of convergence of periods during the main sequence through AM loss by magnetised stellar winds for FGKM (solar-type) stars. Barnes further refined empirical gyrochronology relations in Barnes (2007), expanding upon the relation derived by Skumanich (1972) by adding a mass dependence (or colour correction) omitted from this simple Skumanich relation. The relation, which estimates the age of an FGKM star given its rotation period $P$ in days and $B - V$ colour, is

$$\log t_{\mathrm{gyro}} = \frac{1}{n}[\log P - \log a - b \log(B - V - 0.4)], \tag{1.3}$$

where t is in Myr and $n = 0.5189 \pm 0.007$, $a = 0.7725 \pm 0.011$, and $b = 0.601 \pm 0.024$. These coefficients were further updated empirically in Mamajek & Hillenbrand (2008) to better fit cluster data than the original Barnes (2007) values.

Following this work, a plethora of research into gyrochronology and its limits has been conducted. Using a sample of 24,124 stars observed by Kepler, Reinhold & Gizon (2015) calculated gyrochronological age estimates from photometric light curves. As shown in Figure 1.8, they found a broad agreement with the derived gyrochrones (lines of constant age in colour–period space) but highlight several assumptions without which gyrochronology becomes less valid. In particular, they note that gyrochronological estimates are most reliable for stars 500–2500 Myr old and that careful consideration of binary stars and sub-giant stars must be used when considering large samples. Furthermore, Gallet & Delorme (2019) demonstrates that large orbiting bodies and planetary engulfment events will also reduce the accuracy of gyrochronological estimates.

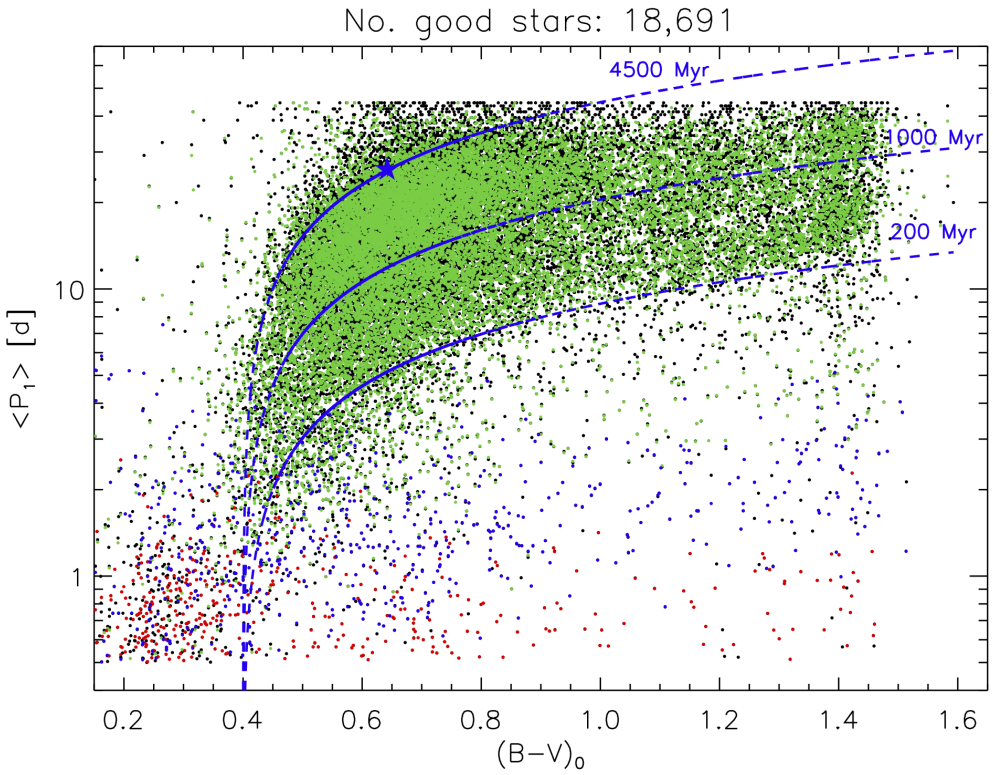



Figure 1.8: Rotation periods plotted against $B-V$ colour for 18,691 stars observed with Kepler. Black points indicate data using only one quarter of data, and green data points use an average rotation period from multiple quarters. Red and blue dots show stars with periods very stable in time. The blue dotted lines are gyrochrones of labelled ages. The blue star indicates the position of the Sun. Credit: Reinhold & Gizon (2015).

The work conducted in Chapter 5 assesses some aspects of the validity of gyrochronology when applied to a large sample of variable objects detected with NGTS. In particular, I will discuss a region of reduced rotational period detection first observed by McQuillan et al. (2014) and additionally by Reinhold & Gizon (2015); Davenport & Covey (2018); Gordon et al. (2021), which indicates a deviation from the expected Skumanich spin-down.

## 1.5 The Rise of 'big data' within astronomy

### 1.5.1 Big data sets

The use of wide-field cameras on large aperture telescopes and the use of multi-telescope arrays has led to larger data sets within astronomy. Probing into shorter timescales requires faster observations with more images taken per night, whilst long-baseline observations with space telescopes such as Kepler generate extremely long time series of single objects. The scale of data in astronomy is nicely summarised by the '3 V's of big data': *Volume, Velocity and Variety*.

I will explain these three concepts in the context of wide-field photometric surveys and similar.

#### 1.5.1.1 Volume

A 2048 × 2048 pixel CCD will produce a 16 MB image if stored at 32-bit integer precision. Additionally, the telescope will store data for each image taken including time, pointing, details on the filters used and how the measurement was taken (such as shutter speed and exposure time). Ground-based instruments will also store information on the weather conditions such as airmass, temperature and wind speed, and Sun and Moon positions. Over a night of observation, a facility may take thousands of images, resulting in tens of gigabytes of raw image data per night, not including additional data products such as sky background and flat-field images and image metadata or any processed photometric data. For ground-based surveys, it is acceptable to store this data volume on large hard-drive racks. However, the raw data volume for space telescopes is often much greater than the maximum satellite downlink limit requiring either reduced observation cadence or on-board pre-processing and data reduction. For example, in the case of Gaia, only a few dozen pixels around each source can be downlinked, which reduces the data output from Gbit/s to about 3 Mbit/s, in line with the downlink speed (Siddiqui et al. 2014). Future projects such as the Square Kilometre Array (SKA)[e] which will be the world's largest radio observatory, will produce approximately 0.5 to 1 TB of data per second for a 6-hour observation, which has resulted in the project focusing a huge amount of research effort into data handling in the form of a dedicated on-site processing facility known as the Science Data Processor. For further detail on the data considerations for the SKA, I refer the reader to Scaife (2020).

#### 1.5.1.2 Velocity

Velocity refers to the speed with which measurements are taken. When deciding on a suitable observation cadence for a facility, you should take into account the timescale of the astrophysical phenomena you wish to observe. For example, in the case of Gaia's astrometric measurements, observing the movement of stars can be done on a much-reduced cadence compared to a dedicated asteroseismology mission for which measurements on timescales of seconds or minutes are necessary to understand short timescale oscillations.

There are also hardware considerations to be made when deciding on an observation cadence. Shorter observations will have higher photon noise and limit the magnitude of targets able to be observed. This may be a positive for bright objects, as short observation windows will avoid CCD saturation. A CCD has a finite readout time, and telescope design must draw a

[e] https://www.skatelescope.org/. Accessed: 22/01/2022.

balance between readout speed and readout noise. In the case of NGTS, the CCDs are read at a speed of 3 MHz, reading an entire image in 1.5 seconds (Wheatley et al. 2018).

Should any data pre-processing be done, it will need to be optimised to reduce the lag between images being taken and processed data being available. For an NGTS field with an average of around 10,000 sources in each image taken at a 12-second cadence, this results in a high throughput of data and further contributes to the total volume of data. A typical NGTS field will contain around 10 GB of light curve data in the form of a time series, flux values, flux error values and any data processing flags. In the case of crowded fields around clusters, this can increase up to 200 GB per field.

#### 1.5.1.3 Variety

Each source within an image requires individual treatment to ascertain what the source is. Sources can be astrophysical or noise, and an astrophysical source could be a star, a galaxy, a Solar System object or a transient such as a meteor. A more detailed description of the NGTS source detection and photometric light curve generation pipeline will be given in Chapter 3, demonstrating such individual treatment of sources within an image.

Some sources even claim there are '10 V's of big data': *Volume, Velocity, Variety* as well as *Variability, Veracity, Validity, Vulnerability, Volatility, Visualisation and Value*[f]. I will leave it as an exercise to the reader to ascertain the value of the seven additional V's mentioned.

### 1.5.2 High-performance computing

The huge amounts of data generated by modern surveys have led to a need for suitable computing facilities and data storage to manage this data throughput and volume. High-Performance Computing (HPC) solutions have three main components: compute, network and storage. Often HPC clusters are used, which are servers networked together to allow programs to run simultaneously on individual servers. One way of measuring the performance of a computer is in floating-point operations per second (FLOPS). A typical single laptop core in 2021 has a power of approximately 5–6 GFLOPS, which provides ample computing power for everyday operation. Much more power is required for large-scale scientific data processing or for the training and evaluation of machine learning models. Launched in 2017, the Wilkes3 cluster based at the University of Cambridge can process approximately 6 PFLOPS ($6 \times 10^{15}$ FLOPS), i.e. 1 million times more powerful than a standard home computer[g].

The Wilkes3 cluster utilises Graphics Processing Units (GPUs) instead of the more traditional CPUs. GPUs harness the same advantages as HPC clusters: splitting operations into

[f] e.g., https://tdwi.org/articles/2017/02/08/10-vs-of-big-data.aspx. Accessed: 22/01/2022.
[g] https://www.hpc.cam.ac.uk/systems/wilkes-3

many small parallel computations increases the overall operating speed. Applications of GPUs to scientific computing have come largely in machine learning, where simple multiplicative operators are used in large numbers as opposed to more complex models that require a linear throughput of data.

The work conducted in Chapters 5 and 6 uses the Cambridge HPC facilities extensively. I developed the software used for variability extraction from NGTS light curves with HPC performance in mind.

### 1.5.3 Machine learning

As the scale of astronomical data grows, machine learning (ML) techniques such as artificial neural networks have been applied to many different problems and fields (Ball & Brunner 2010). Machine learning techniques differ from traditional model-fitting methods in that the same ML model applies to a wide variety of problems. In contrast, a parametric model is often predefined to best fit the application. ML models also allow the abstraction of complex non-linear behaviour, which would be difficult or expensive to calculate with a parametric model. This abstraction is a potential downside; machine learning is often referred to as a 'black box' as it is difficult to interpret why the ML model has made the decisions it has during training and evaluation. Machine learning algorithms are often divided into two groups: *supervised* and *unsupervised*.

Supervised machine learning algorithms are algorithms used to learn the relationship between a set of measurements (inputs) and a target variable or set of variables (outputs) given a set of provided examples (training data). Unsupervised models instead rely solely on the data to find best-fit model parameters and are often employed on data exploration problems such as clustering, dimensionality reduction and outlier detection. Details on implementing several successful ML algorithms used within time-domain astronomy are given in Section 2.3.

Two relevant machine learning applications are in stellar variability detection and characterisation and in candidate vetting and classification of transiting exoplanets from photometric light curves. The classification of variable star light curves is well-suited to machine learning but requires the selection of sensible input features. Namely, measurable quantities such as the period and the amplitude of the signal, stellar properties such as the colour index and magnitude and specific signal shape information such as the residual around the folded light-curve model (Dubath et al. 2011). This method of parameter extraction has been applied to light curves from the All-Sky Automated Survey (ASAS, Eyer & Blake 2005), Hipparcos (Dubath et al. 2011) and OGLE (Debosscher et al. 2007). A paper by Richards et al. (2011) provides a more in-depth description of how feature extraction should be considered to optimise the performance of classifiers such as a random forest (RF).

As the number of planetary transit candidates increases in modern photometric surveys, automated vetting of these candidates is preferable to the traditionally manual 'eyeballing' process. Within NGTS, the first example of this was the successful application of self-organising maps (SOMs) and random forests to candidate ranking. Armstrong et al. (2018) took a set of features from each transit fit including transit depth, period and stellar parameters. The algorithm generated a score for how likely the detection was to be a transit. The work demonstrated that this method performed exceptionally well on a set of NGTS light curves. In general, using such a tool to rank candidates could drastically reduce manual vetting time.

A notable advance in the application of ML to exoplanet detection came from Shallue & Vanderburg (2018) and their development of `Astronet`, which used an artificial neural network (ANN) to predict whether a given signal is a transiting exoplanet or a false positive caused by astrophysical or instrumental phenomena. Specifically, the work used a convolutional neural network (CNN), an architecture mainly used in image processing. The authors trained this on Kepler data and detected two new planets not previously found in the Kepler dataset with this method. Using a CNN does not rely on feature selection, as the CNN takes the light curve as an input. Instead, the authors optimised the 'views' of data given to the network, such as normalising by transit depth and providing zoomed 'local' views of the light curve around the transit. Some more recent studies have taken this approach following the success of `Astronet`. `Exonet` was proposed by Ansdell et al. (2018), which improved the performance of the CNN through the addition of scientific domain knowledge; `Exonet` was additionally given CCD pixel flux centroid information and stellar parameters to improve its classification. This model was later applied to light curves from TESS, where it performed well on previously-confirmed transit events and proposed a further 200 candidates (Osborn et al. 2020). Chaushev et al. (2019) was able to apply a similar CNN-based method to the ground-based, non-continuous data from NGTS. The authors claim that the time required for vetting can be reduced by half using a CNN while still recovering the vast majority of manually flagged candidates. In addition, the CNN was able to identify many new candidates with high probabilities which were not flagged by human vetter.

# Chapter 2

# Scientific Background

This section aims to provide a theoretical background on the physics used in this work. I will discuss the myriad forms of stellar variability that we can detect with NGTS and details of methods used within signal processing, time series analysis, and machine learning.

## 2.1 Stellar variability

I will discuss different forms of stellar variability, which can be sorted into three groups of periodic variables and eruptive or cataclysmic variables with no periodicity. The three groups mentioned are rotational, pulsational, and external variability in binary and higher-order star systems. Where a variability class is defined, the name often reflects the first or most significant star found to exhibit this form of variability and is highlighted in **bold** in the text. With the release of precise magnitude and colour information from Gaia, we can plot how variability classes are distributed in HR diagrams. Works such as Eyer & Mowlavi (2008) were able to create these plots prior to Gaia, however since then, a wealth of astrometric, photometric and asteroseismological data has been released, and the work of Eyer et al. (2019) produced HR diagrams using Gaia stellar parameters for known variables stars from published catalogues, which are included throughout this Section to aid the reader in placing variability classes in HR space (Figures 2.3, 2.7 and 2.9).

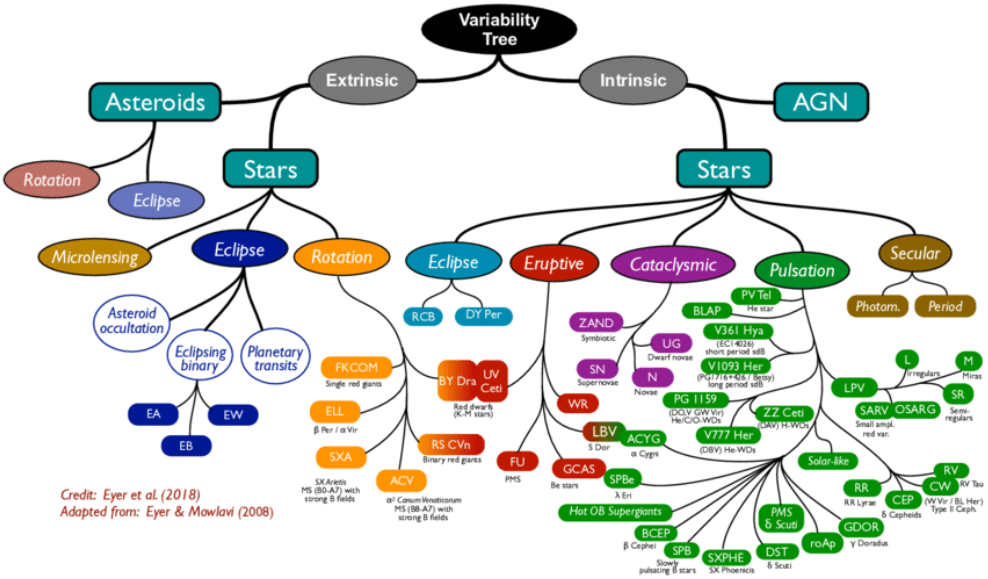



Figure 2.1: A 'Variability Tree', which organises variable objects according to the source of their variability. Credit: Eyer et al. (2019).

### 2.1.1 Rotation

Photometric variability will arise from rotating stars due to photometrically active regions on the stellar surface rotating into and out of view. These active regions are attributed to stellar spots, convective regions, and more explosive events like flares or other magnetic activity. These events can produce a photometrically variable signal with the same rotation period as the star, subject to any characteristic timescales of the processes themselves, which will affect the observed signal. Figure 2.3 places several distinct rotational variability classes on the HR diagram, some of which are discussed in the following sections.

Solar-like stars, FGK dwarfs, will often exhibit photometric variability in the form of stellar spot rotation. The periods for rotational variables vary from less than one day up to hundreds of days. Observed periods of spot modulation are generally skewed towards shorter periods in part because it is easier to observe faster rotators with short periods. Additionally, many stars observed with spot-based variability are young, magnetically active stars with many spots compared to less active but still rotating older stars (Strassmeier 2009). Figure 2.2 shows an example light curve of such a star, KIC 5110407, a K-type star observed with Kepler (Roettenbacher et al. 2013). We see an obvious sinusoidal variability signal evolve into a more complex 'double-dip' modulation pattern and back into a single sinusoid. This phase-shifting signal is typical of a spotted star; as groups of spots are formed, grow, shrink, and disappear on the stellar surface, we observe phase shifts in these signals and complex photometric variability. Over longer timescales, we know the Sun exhibits an 11-year-long magnetic cycle, causing an

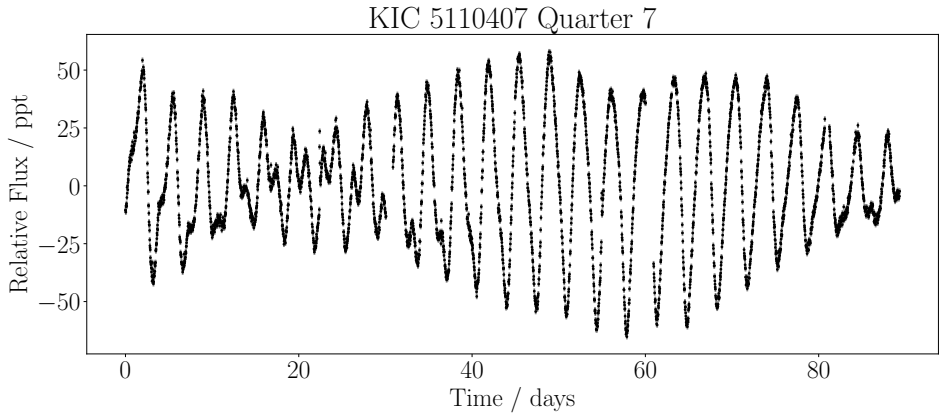



Figure 2.2: Kepler data from quarter 7 of the spotted star KIC 5110407. Relative flux (in parts per thousand, ppt) is plotted against time from the start of the quarter (in days).

ebb and swell in the magnetic activity. At solar maximum, we observe many more starspots than at solar minimum, and it has been shown that G–K dwarf stars exhibit these many-year long cycles as well (Oláh et al. 2016).

#### 2.1.1.1 Spots, faculae and plages

Local magnetic fields create starspots on the photosphere of stars, where fields are strong enough to suppress the regions of convective overturn and hence redirect the flow of energy outwards. This results in cool and, therefore, dark regions on the surface of a star (Strassmeier 2009). It is possible to detect the presence of starspots through rotation, and in recent years track their movement across the surface of the star through *Doppler imaging* (Vogt & Penrod 1983) or aperture synthesis imaging with interferometric telescopes (Parks et al. 2011). Despite these advances, most of our knowledge of spots comes from the Sun.

Sunspots have been observed to have lifecycles from a few days to a few months, with larger groupings of sunspots persisting for weeks to months. These sunspots expand and contract, as well as drift in latitude on the surface of the Sun. Sunspots have typical diameters of tens of thousands of km on the Sun, covering around 0.1% of the Solar surface. Extremely large spots can be observed on smaller or more magnetically active stars; a spot on the active K-giant star XX Tri covered about 22% of a hemisphere, estimated to be around 11 million km in diameter (approximately eight times the Solar diameter) (Strassmeier 1999). Fully convective and magnetically active M dwarf stars are predicted to be extremely spotty, with a relatively uniform distribution of spots in longitude and latitude (Barnes et al. 2017). Spotted, active K–M variable stars can be classed as **BY Draconis variables**; these stars have large numbers of starspots and hence exhibit photometric variability at the period of their rotation. Typically

these rotation periods are tens of days long, with amplitudes of 0.1–0.3mag in the visible (Percy 2007).

In addition to spots, there exist regions of increased brightness known as *faculae* and *plages*. Faculae exist in the photosphere of the Sun, between the small and short-lived convection cells known as solar granules. The concentration of magnetic field lines between these cells causes a very small region of increased brightness. Within the chromosphere, we see similar small regions of brightness known as plages, which appear to map closely to areas of increased activity. It has been shown that on timescales comparable to the 11-year solar magnetic activity cycle, the appearance of dark spots and bright plages are highly correlated, despite being formed in different layers of the stellar surface (Mandal et al. 2017).

On rotational timescales, the Sun's (and hence most likely FGK dwarfs) photometric brightness variation is dominated by spot activity within the optical regime. Stellar activity is not spot-dominated at shorter wavelengths ($\lesssim$ 400 nm) and during magnetic cycle minima when spot contribution is weak (Shapiro et al. 2016)[a].

Other main-sequence rotational variability classes have been identified, in which variability can be attributed to different processes than described above. **$\alpha^2$ Canum Venaticorum variables** are chemically peculiar late B to early F stars. Strong magnetic fields cause abundant heavy elements to move towards the stellar surface, producing associated brightness changes (Percy 2007). These brightness changes manifest as a sinusoidal signal in a photometric light curve with a period the same as the star's rotation period: typically a few days, and with amplitudes of 0.02–0.05 mag (Percy 2007).

#### 2.1.1.2 Pre-main-sequence stars

Extremely young PMS stars (<10 Myr) will still be contracting and will still be affected by the presence of circumstellar material. The link between the star and any circumstellar material will manifest in variability across a range of wavelengths and timescales. An important group of PMS variables are **T Tauri stars**. A T Tauri star is best defined by its spectrum, as the regions of low-density gas common to star-forming regions produce a well-defined set of spectral lines (Percy 2007), the details of which are beyond the scope of this work. This spectrum is typical of a star with an accretion disc (known as a classical T Tauri) or with the remains of the accretion disc in the case of a weak-line T Tauri (WTTS). T Tauri stars are of spectral type FGKM and, as PMS stars, will lie above the main sequence on an HR diagram.

[a]This is not the case for spectroscopic measurements, however. Meunier et al. (2010) show the convective blueshift suppression caused by plages can dominate the RV signal, even during magnetic cycle minima when spots may not be present.

Photometrically, T Tauri stars will exhibit fluctuations on time scales ranging from minutes to days, with amplitudes ranging from 0.01 mag through to multiple magnitudes. The variability timescales can be associated with the star's rotation and the orbit of any circumstellar disc present, the free-fall timescale and flaring of the star (Percy 2007). Of note for this study are timescales associated with rotation of the star and any obscuring by circumstellar material. These timescales are generally between 0.5 and 18 days. Herbst et al. (2001) demonstrates that the rotation periods of young stellar objects (YSOs) in the Orion Nebula Cluster (ONC) exhibit a bimodal distribution of rotation periods with peaks at about 2 and 8 days. This bimodality was attributed to disc-locking: contracting stars should spin up, but when magnetically linked to a circumstellar disc, this can prevent such a reduction in rotation period (as seen in the stars near the 8-day peak).

It is also possible for both main-sequence and PMS objects to exhibit non-periodic brightness variability. Although not studied in detail in this work, such sudden brightness changes will add noise to observed light curves. Variability arising from obscuration of the star by surrounding gas or dust can cause non-periodic variability. We have already seen systems that exhibit this effect, in the case of T Tauri variables. Obscuration can also occur due to material being separated from the star in the case of extremely rapidly rotating **$\gamma$ Cassiopeiae** B stars (Slettebak 1982). The magnitude of the obscuration can be extremely large: in the case of the star **R Coronae Borealis** a sudden reduction in brightness of 10 magnitudes occurred, thought to be attributed to a large dust cloud observed around the system with HST. This hypothesis was confirmed via IR excess (Jeffers et al. 2012).

**Flare stars** or **UV Ceti** variables are dwarf K and M stars which exhibit rapid increases in brightness of several magnitudes over short (seconds to minutes) timescales. This rapid increase is followed by a slower exponential decay back to quiescence. Stellar flares are thought to be driven by high-energy magnetic reconnection events like Solar flares. Flare occurrence appears to be randomly distributed in time; however, the distribution of flare energies appears to follow a well-defined power law as lower energy flares are much more likely than high energy flares. For particular magnetically active M dwarf stars, the presence of *nanoflares* (extremely low energy but frequent flares) may appear as additional noise in photometric measurements. These nanoflares are a subject of study with NGTS; the work by Dillon et al. (2020) utilises the high cadence sampling of NGTS photometry to correlate the presence of quasi-periodic oscillations in M dwarf brightness with the presence of nanoflare activity.

### 2.1.2 Oscillations and pulsations

Stellar pulsations arise from the expansion and contraction of the outer layers of a star as it attempts to maintain equilibrium. Radial pulsations are caused by a feedback mechanism known

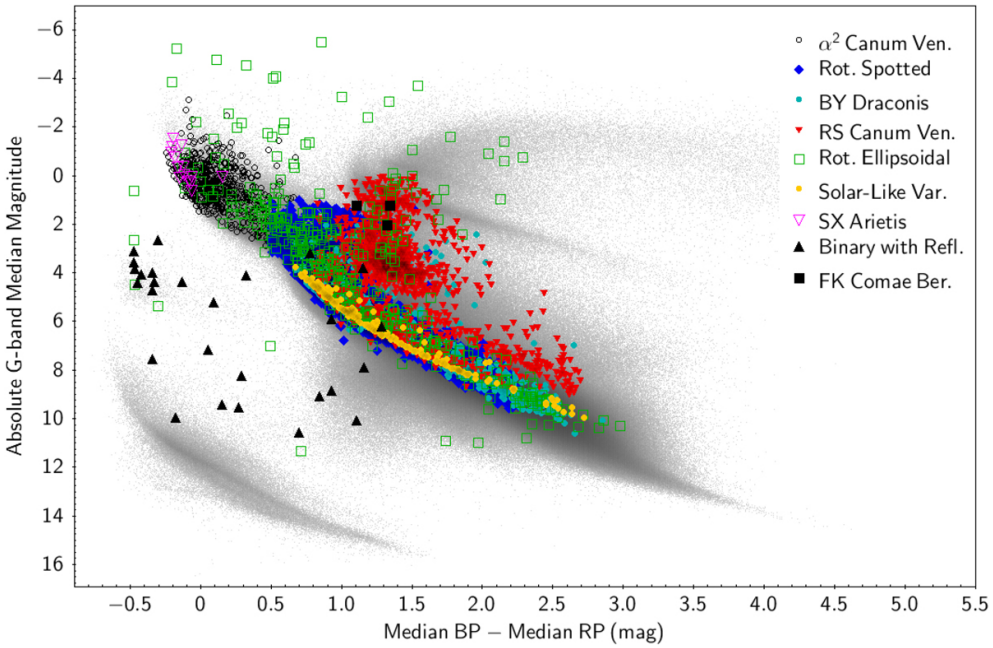
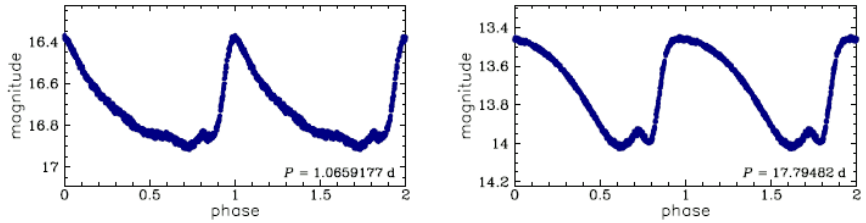



Figure 2.3: Rotating variable stars from published catalogues are placed in the observational colour–absolute magnitude diagram, with symbols and colours representing types as shown in the legend. The background points in grey show non-variable sources from Gaia. Credit: Eyer et al. (2019).

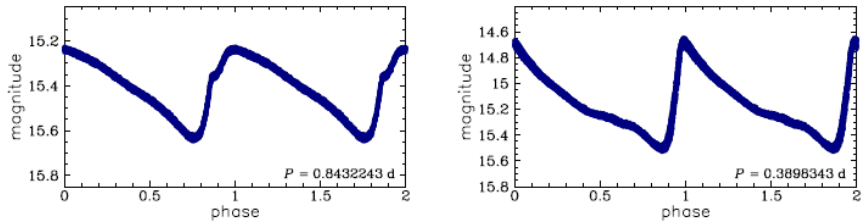
as the $\kappa$ mechanism in many variable stars. A rise in density within a partially-ionised layer of a star (as a result of gravitational pressure, for example) creates increased opacity ($\kappa$), causing increased energy absorption from the stellar interior. This increased absorption causes heating and expansion of the layer, reducing the opacity once again (Saio 1993). This mechanism is only effective within partially ionised gas layers, as ordinarily, the strong temperature dependence of opacity negates this feedback loop.[b] The position of stars on the HR diagram exhibiting $\kappa$ mechanism driven pulsations is along the *instability strip*, a line of increasing luminosity and decreasing temperature where stars have ionisation zones at the right depths to drive pulsations. The instability strip is labelled on the HR diagram in Figure 1.6, and several classes of pulsating variables are shown in Figure 2.7.

Radial pulsations may also trigger higher-order modes of oscillation. These modes are similar to spherical harmonic modes and can be modelled as such to infer their frequency relationship. Some stars may exhibit pulsations occurring in multiple modes simultaneously, such as multi-mode classical Cepheids, which exhibit multiple periods within their light curves, and provide useful tools for probing stellar interior structure and testing stellar opacity models (Moskalik & Dziembowski 2005).

[b]For non-ionised gases, this strong temperature dependence arises from metals being stripped of electrons at a higher temperature which reduces opacity at a much greater rate than the increase in pressure.




(a) Two example light curves of classical Cepheid variable stars from OGLE.




(b) Two example light curves of RR Lyrae (RRab) variable stars from OGLE.

Figure 2.4: Example light curves of $\kappa$-mechanism driven variable stars from the Optical Gravitational Lensing Experiment (OGLE) (Soszynski et al. 2015). The light curves are phase folded on the displayed period. Credit: OGLE Atlas of Variable Star Light Curves.

Smaller, non-radial pulsations in stars such as internal pressure (p) or gravity (g) waves like those seen within the ocean or Earth's atmosphere will also be present. The physics of these waves is much more complex than radial pulsations, as we have three degrees of freedom for these waves rather than one; this will not be discussed in detail. p- and g-waves are the basis of the signals observed in asteroseismological missions, which utilise these standing waves to probe stellar interiors in much the same way geophysicists use earthquake signals to probe the Earth's interior. An excellent review of Asteroseismology, including the physics of non-radial oscillations in stars, is given by Di Mauro (2016).

A common subdivide of classes of pulsating stars is by period. Broadly, this allows for a division of pulsating stars into main-sequence and more evolved giant stars. Younger, more compact stars will often have shorter period pulsations due to a higher sound speed internally, compared to evolved giants, which have a lower material density corresponding to longer period oscillations. I will outline several distinct variability classes that differ in period: firstly, stars in the instability strip with the same underlying physics but a range of pulsation periods, and secondly, stars with different pulsation mechanisms.

#### 2.1.2.1 $\kappa$ mechanism variables

**Cepheid variables**, named after the homonymous $\delta$ Cephei star first observed to display pulsation by John Goodricke in 1786 (Goodricke 1786), display a distinctive light curve shape: a

sharp increase in brightness followed by a slower dimming. Cepheid variables are separated into classic Cepheids/Type I Cepheids and Type II Cepheids. Despite displaying similar variability, these two groups exhibit markedly different ages, masses, and evolutionary histories. **Classical Cepheid Variables** are comparably young, massive ($> 1M_{\odot}$) G-type stars which have begun to move off the main sequence, beginning helium core burning (Percy 2007). Classical Cepheids are an example of pulsating stars driven by the $\kappa$ mechanism and hence exist in the instability strip of the HR diagram. Typically they exhibit periods of 1 to 70 days, with the distinctive shape described and shown in Figure 2.4a, in which two classic Cepheid variable light curves taken with the Optical Gravitational Lensing Experiment (OGLE) (Soszynski et al. 2015) are displayed. Most Cepheid variability amplitudes are of order 0.5 to 1 magnitude in the visible, though smaller amplitude Cepheid variables have been detected, and this small amplitude variability can be attributed to stars close to the border of the instability strip (Kovtyukh et al. 2012). Classical Cepheid variables follow a tight period–luminosity relation known as Leavitt's law. This empirical law was first calibrated by Hertzprung in 1913 (Hertzsprung 1913) and is still used to calculate the distance to classical Cepheids, albeit re-calibrated using measurements from the Hubble Space Telescope (Benedict et al. 2007). **Type II Cepheids** are similarly located on the instability strip on the HR diagram. However, they are low mass ($< 1M_{\odot}$), metal-poor giant stars, typically much older than their classical counterparts. They have higher effective temperatures than classical Cepheids, corresponding to late-F and G type stars (Percy 2007). Within the class of type II Cepheids, stars are further separated into three sub-classes based on variability period. **BL Herculis variables** vary over periods of 1 to 8 days, with amplitudes of about 0.1 mag. **W Virginis variables** have longer periods of 10 to 20 days, with larger amplitudes of around 1 mag. The longest period type II Cepheids are known as **RV Tauri variables**, exhibiting variability periods of more than 20 days with amplitudes of multiple magnitudes. These three sub-classes occupy different parts of the period–luminosity relationship and the instability strip. BL Herculis variables cross the instability strip post horizontal branch and towards the AGB. W Virginis variables are helium-burning stars within the AGB, and RV Tauri stars begin to evolve beyond the AGB towards white dwarfs (Percy 2007; Soszyński et al. 2018). The period–luminosity relations for classic Cepheids and type II Cepheids are displayed in Figure 2.5. This Figure highlights the tight relationships observed between period and luminosity. We see also the three sub-classes of type II Cepheids separated in period and luminosity.

**RR Lyrae stars** are the only stars within the group of short-period variables that are giant stars. RR Lyrae variables span spectral classes from A to late F and are older stars often found within globular clusters. They exhibit stable periods of 0.1 to 1 day, with visible amplitudes of up to 1.5 mag (Percy 2007). Despite having pulsations driven by the same underlying $\kappa$

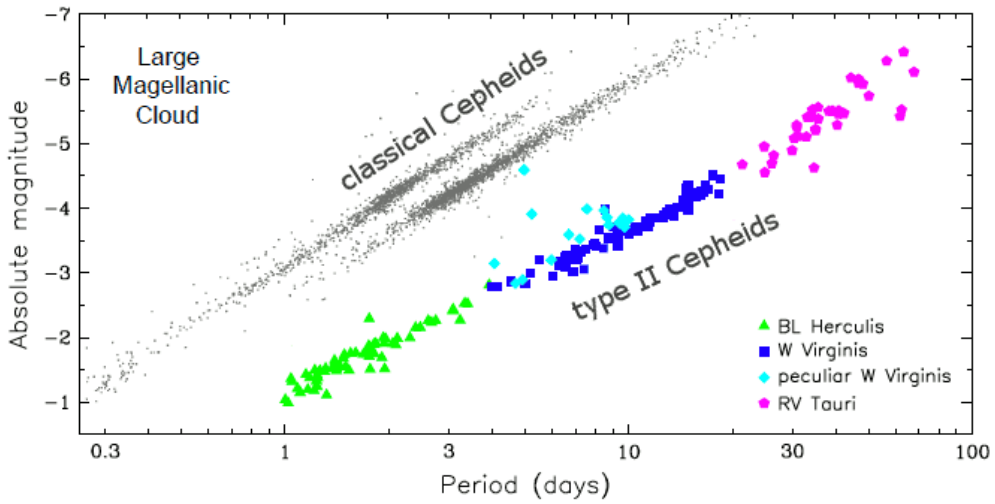



Figure 2.5: Period–Luminosity relations of classical Cepheids (grey points) and type II Cepheids (coloured points) within the Magellanic cloud. Credit: Soszyński et al. (2018), taken from the OGLE Atlas of Variable Star Light Curves.

process as Cepheid variables, they are given a separate variability classification due to their much shorter periods and chemical differences (Smith 2003). RR Lyrae stars are separated into sub-classes based on their pulsation modes. The fundamental mode pulsators are known as **RRab stars**, with first-overtone stars **RRc** and double-mode **RRd**. Fundamental mode RRab stars have asymmetric light curves with a steep rise in amplitude followed by a slow decrease of brightness after the maximum: for example, the two phase-folded light curves in Figure 2.4b. Approximately 50% (Jurcsik et al. 2009) of RRab stars display long-term modulations of the amplitudes and phases of their variability as shown in the OGLE light curve in Figure 2.6. This phenomenon has been termed *Blazkho* variability after Blažko (1907); however, the origin of this effect is unknown. Kolenberg (2008) provides an excellent review of the history and theories behind the Blazkho effect. RR Lyrae stars are on the horizontal branch of the HR diagram, burning helium in their core. Although all RR Lyrae have very similar luminosity, they span a large range in metallicity. The absolute magnitude and period of RR Lyrae stars are slightly dependent on the metal abundance, although this dependence is very small. As RR Lyrae stars are often found within dense globular clusters large distances from Earth, accurate parallax measurements and metallicities are difficult to obtain (Borissova et al. 2009).

Where the instability strip crosses the main sequence on the HR diagram, we find **$\delta$ Scuti variables**. $\delta$ Scuti variables are driven by the same $\kappa$ mechanism as Cepheids. They lie on the instability strip and follow a similar period–luminosity relation, hence their original name of dwarf Cepheids. They are stars of spectral types A to F with short-period regular variability of typical amplitudes a few hundredths of a magnitude (Percy 2007). Typical periods for $\delta$ Scuti

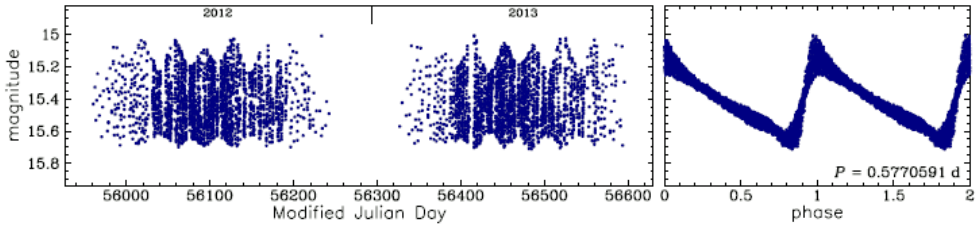



Figure 2.6: An example RRab star light curve exhibiting long-term modulation on the amplitude and phase of variability, known as Blazkho modulation. The light curve from OGLE (Soszyński et al. 2014) is displayed on the left and phase-folded on the displayed period on the right. Credit: OGLE Atlas of Variable Star Light Curves.

variables are in the range 0.02 to 0.3 days (0.5 to 7 hours). There are high-amplitude $\delta$ Scuti stars, with visible variability magnitudes larger than 0.3 mag. These HADS variables exhibit the 'saw-tooth' variability pattern seen in the more evolved Cepheid variables. A distinct class of HADS variables that appear much older with lower metal abundance are termed **SX Phoenicis stars** with short period variability of 0.03 and 0.08 days and amplitudes of order 0.1 mag. SX Phoenicis stars are a part of the so-called *blue straggler* population of stars. Where stars turn off the main sequence as they exhaust their hydrogen reserves, some stars are observed to have higher luminosity than expected from single-star evolutionary tracks. It is hypothesised that these stars have increased mass and luminosity due to an unresolved or even merged binary companion (Percy 2007; Sandage 1953).

#### 2.1.2.2 Other pulsating variables

Stars may exhibit variability driven by non-radial pulsations compared to the $\kappa$ mechanism pulsators. **$\gamma$ Doradus variable** stars lie just beyond the red edge of the $\delta$ Scuti instability strip and are typically stars of spectral type F0–F2. They exhibit variability periods ranging from 0.4 to 3 days and amplitudes up to 0.1 magnitudes in the visible. Due to the high-order, non-radial modes of pulsation, these stars often exhibit multi-period variability (Krisciunas 1993).

A-stars with 'peculiar' metal abundance of certain elements can exhibit extremely complex variability caused by strong magnetic field interactions and rotational effects interacting with $\delta$ Scuti type oscillations. These stars are known as **Rapidly Oscillating Ap (roAp) stars**. RoAp stars make excellent targets for asteroseismology studies. TESS has already found a number of these roAp stars, including a short 4.7-minute long pulsation period on one object (Cunha et al. 2019). The introduction of this paper by Cunha et al. (2019) provides an excellent summary of the status of modelling these complex stars, including the *oblique pulsator model* in which pulsations along a magnetic axis offset from the rotation axis cause modulation of the expected pulsation pattern.

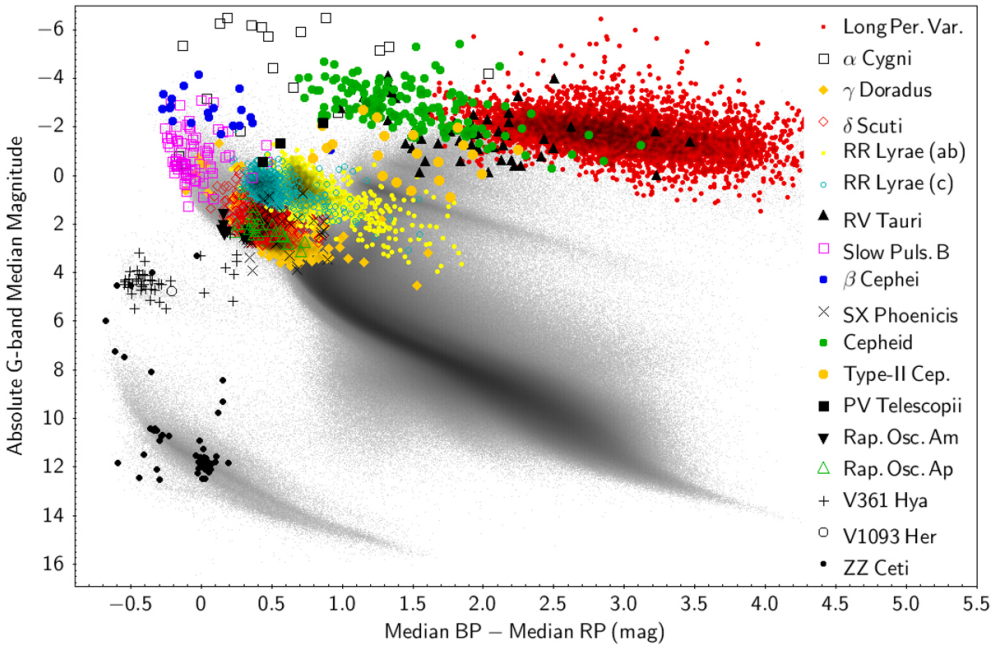



Figure 2.7: The same CMD as Figure 2.3, but with classes of pulsating stars labelled. Credit: Eyer et al. (2019).

There exist further classifications of short-period variable stars such as ***β* Cephei**, **ZZ Leporis**, **Slowly Pulsating B variables**. The book 'Understanding Variable Stars' by Percy (2007) provides detailed descriptions of these minor variability classes and the distinctions in stellar parameters and photometric variability observed by these stars.

### 2.1.3 Variable binaries

It is well known that multiple star systems are fairly common in the Milky Way. The PMS population shows at least 50% of systems with multiple stars, indicating that binary systems form early on in a system's life (Mathieu 1994). Similarly, for the MS population, estimates from Raghavan et al. (2010) suggest that around 44% of Solar-like stars have at least one companion. Of particular interest to this work are *close binary stars*, in which the interactions between two stars are large and the geometric probably of an eclipse being seen from Earth is larger. Many of these systems will not be resolvable even by the most powerful telescopes due to their extremely small on-sky angular separation. The binary systems' photometry, spectroscopy, and astrometry enable astronomers to detect multiple stars without fully resolving the system. Photometric variations occur in both eclipsing and non-eclipsing binaries; however, eclipsing binaries generally provide much more information. We refer to 'primary' and 'secondary' eclipses within an eclipsing binary system: the primary eclipse occurs when the fainter star eclipses the brighter star and vice-versa for the secondary. Two equal brightness stars will result

in the primary and second eclipses having equal depth. The exact geometry of the system and the stellar parameters of the two bodies will affect the shape of the light curve seen.

**Eclipsing Binary** star systems can be classified into three broad classes, based on the shape of the photometric light curve. **EA** binaries have light curves with almost flat out-of-eclipse light curves, with well-defined eclipses (although the secondary may be barely visible if the primary star is orders of magnitude brighter than the secondary). **EB** binaries have less well-defined eclipses, with more rounded light curves, and **EW** binaries display almost continuous variation. Although these classifications are based on observed photometry, the EA and EW binary classifications represent two extremes of a set of similar systems. The stars within an EA system are further apart than those within an EW system, with EB sitting somewhere in-between. The separation of the two stars leads to the physical definition of different binary systems in terms of star contact, formalised by Kopal (1955).

Within a binary system, the two bodies orbit a common centre of mass (the *barycentre*). If the two bodies are of comparable masses (e.g., two stars, rather than a star and a planet), the system's barycentre will lie well outside the interior of either body. Within this binary system, there exist surfaces of gravitational equipotential across which the gravitational potential from both bodies is equal. There also exist a set of *Lagrange points* at which the gravitational forces of the two bodies balance the centrifugal force arising from orbiting the barycentre. Within the rotating frame of the system, these are stationary points. Of particular note is the $L_1$ Lagrange point, which lies along the line between the two objects where the gravitational forces balance. In the case that $M_1 = M_2$, the $L_1$ point and the barycentre will coincide.

It is possible to draw a gravitational isopotential surface that intersects the $L_1$ Lagrange point, which appears as a 'figure-of-eight' shaped surface. The two lobes of this surface define the *Roche Lobes* of the two stars. Should one of the stars expand to fill its Roche lobe, material will fall from this star towards the other star, carrying mass and angular momentum away from the larger star and onto the smaller. The introduction of Roche lobes leads to the following definitions of binary systems: **Detached binaries** are systems in which both components lie within their Roche lobes. Tidal distortion of the stars is minimal, and the stars will be almost spherical. In **Semi-detached binaries** one star has filled its Roche lobe, but the other has not. The larger star will begin to distort, whilst the smaller remains spherical. The extreme of this definition are **Contact binaries**, in which both stars have filled their Roche lobes and are hence in contact. Within this designation of binary systems, we see that EA roughly correlates to detached systems, EB to semi-detached and EW to contact. Note that these designations are not exact; the prototype EA binary system, Algol, is a semi-detached system. The semi-detached nature of Algol was proven by noting that the less massive star appeared to have evolved faster than the massive one within the Algol system, in contradiction with basic stellar evolution

theories. This evolutionary mismatch could only be the case in the system as observed if mass transfer between the two stars had occurred, a scenario possible in a semi-detached binary system (Pustylnik 1998). Figure 2.8 provides three example light curves with a corresponding cartoon system. Typically periods of EA binaries are longer than those of EW binaries, as the contact between the two companions implies a close orbital radius. However, periods of binary systems range from as short as 17 minutes in the case of AM Canum Venaticorum (Roelofs et al. 2006) up to hundreds of thousands of years in the case of the Alpha Centauri system. Figure 2.9 shows the positions on the HR diagram of known eclipsing binary systems, categorised as EA, EB and EW, along with stars with known exoplanets which also vary in brightness due to planetary transits.

It is also possible to photometrically observe the effects of non-eclipsing binary systems: **Ellipsoidal Variables** are small amplitude fairly sinusoidal variables that oscillate in brightness due to changes in the shape of the stars within the system which are almost in contact. This small change in light emission area causes photometric brightness modulations. Such systems may not eclipse, and as such, must be followed up spectroscopically or astrometrically to confirm multiple stars, such as the spectroscopic binary and ellipsoidal variable system $\alpha$ Virginis (Spica) (Palate et al. 2013).

Close binary stars with strong magnetic activity, such as **RS Canum Venaticorum variables** can exhibit unusual photometric activity outside of eclipse that can be observed in systems that do not eclipse. Typically, this modulation is fairly sinusoidal, arising from spot groups on one star driven by powerful magnetic fields from its binary companion (Hall 1976). In general, the rotation period observed is similar to that of the binary system, and so periods of these systems range from less than a day up to multiple days in length. The stars within an RS Canum Venaticorum system can rotate significantly faster than a lone star of similar colour and luminosity, as the tidal forces between the two stars 'spin-up' the rotation of the individual stars (Percy 2007).

## 2.2 Signal processing & time series analysis

I will provide a theoretical background to some signal processing, time series analysis and machine learning methods used in this work. In particular, I will detail methods used to extract periodic variable signals from photometric light curves, including sinusoidal model fitting such as a Lomb–Scargle Periodogram (Lomb 1976; Scargle 1982), correlation-based methods such as the Autocorrelation Function, and phase folding methods including Phase Dispersion Minimisation (Stellingwerf 1978). Each of these classes of methods has pros and cons and performs better or worse depending on the characteristics of the underlying data. Such features

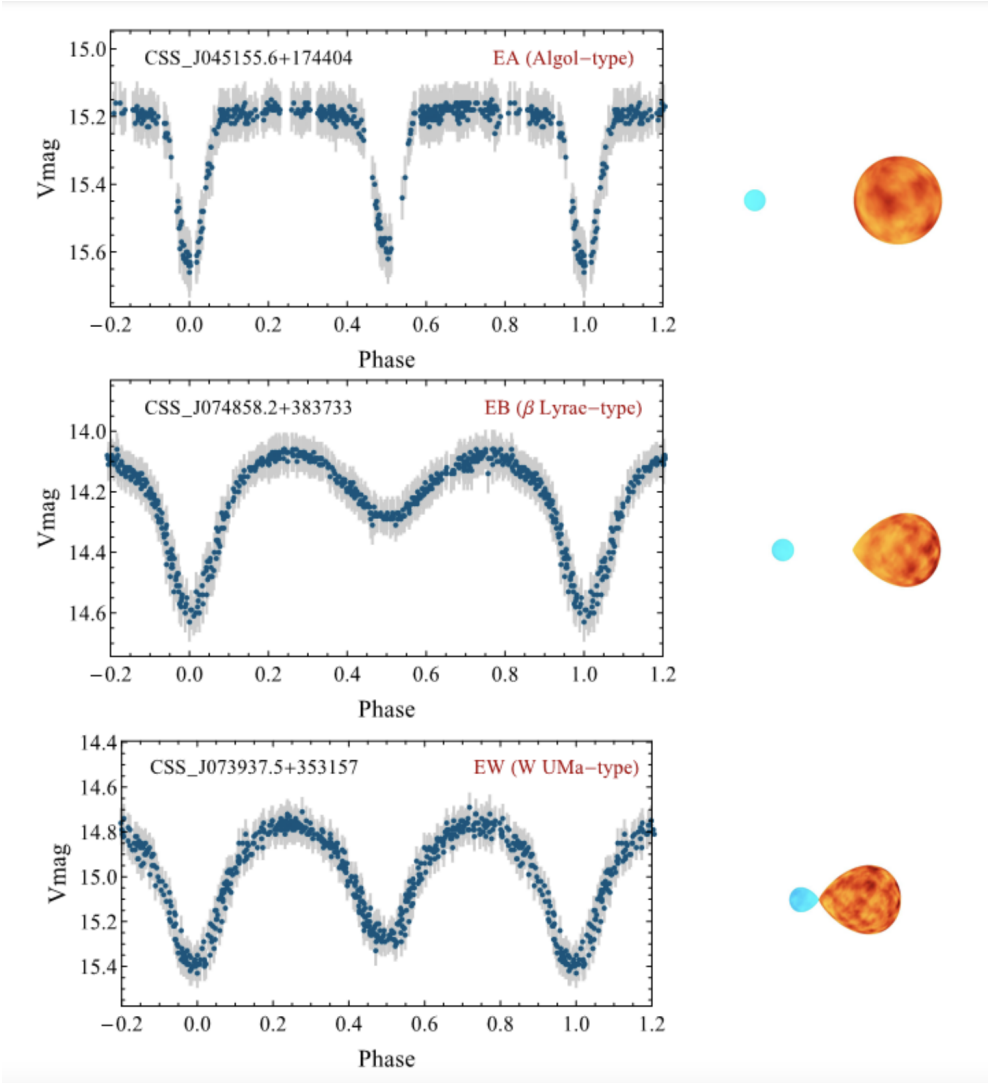



Figure 2.8: Typical eclipsing binary light curves are shown for an EA (top panel), EB (middle panel) and EW (bottom panel). On the right of each light curve is a cartoon of a binary system. In the case of the EB and EW systems, we see that stars begin to distort due to filling their Roche lobes. Credit: Jinbo Fu, Xiamen University.

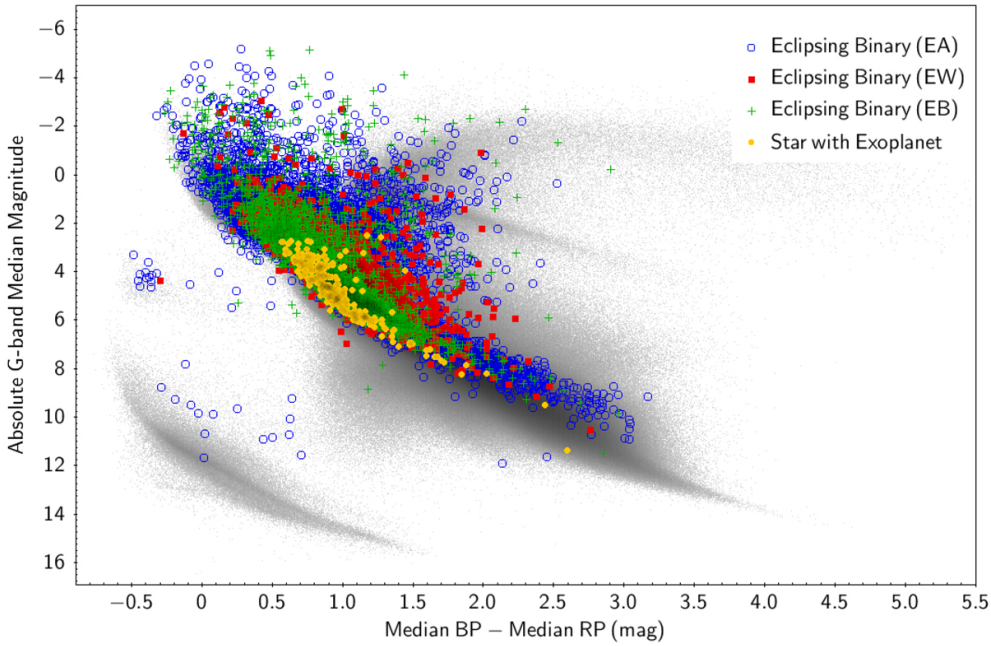



Figure 2.9: The same CMD as Figure 2.3, but with eclipsing binary systems highlighted. Credit: Eyer et al. (2019).

to consider are the sampling of the time series, the shape, period and amplitude of the signal and any additional noise in the light curve.

### 2.2.1 Fourier transforms

The broad study of Fourier Analysis involves approximating functions by sums of simpler trigonometric functions. One such example is a Fourier Transform, a mathematical transform that decomposes a function in space or time into a spatial or temporal frequency function. One definition of a complex Fourier transform of a continuous integrable function $f(t)$ is

$$\widehat{f}(\nu) = \int_{-\infty}^{\infty} f(t) e^{-2\pi i t \nu} d\nu \tag{2.1}$$

where $t \in \mathbb{R}$ and $\nu \in \mathbb{R}$ represent *time* and *frequency* respectively. (Rahman 2011). In the case of a finite sequence of equally-spaced samples of a function, rather than a continuous function, a **Discrete Fourier Transform (DFT)** must be used. This is the case for astronomical time series data, where we sample a star's brightness function at a series of discrete time points. For a set of data $t_1, ..., t_n$ we can define the discrete Fourier transform

$$d(\nu_j) = n^{-1/2} \sum_{i=1}^{n} t_i e^{-2\pi i \nu_j t} \tag{2.2}$$

for $j = 0, 1, ..., n-1$. The frequencies $\nu_j = j/n$ are the Fourier or fundamental frequencies (Shumway & Stoffer 2017). Such discrete functions are prone to *aliasing*: a finite sampling of functions can cause different signals to become indistinguishable when sampled. In the case of simple sinusoidal functions, sampling at a frequency $\nu_s$, the set of functions $\{\sin(2\pi(\nu + N\nu_s)t + \phi), N = 0, \pm1, \pm2, ...\}$ will generate identical samples. This will manifest in the frequency spectrum or a discrete Fourier transform as large responses at each of the frequencies in the set. The set of frequencies connected by aliasing with positive values will thus be

$$\nu_{\text{alias}} = \nu_{\text{true}} \pm n \cdot \nu_{\text{sampling}} \tag{2.3}$$

for a sinusoidal process of frequency $\nu_{\text{true}}$ sampled at frequency $\nu_{\text{sampling}}$.

An important concept within signal processing is the *Nyquist-Shannon sampling theorem* which states:

> If a function $f(t)$ contains no frequencies higher than B hertz, it is completely determined by giving its ordinates at a series of points spaced $1/(2B)$ seconds apart. (Shannon 1984)

If this signal is sampled at anything less than $2B$ samples per second, it is *undersampled* and missing information. Any reconstruction of this signal from its frequency spectrum will be prone to aliasing. Inversely, for a sampling frequency $\nu_s$, the *Nyquist frequency*, $\nu_s/2$, is the highest possible frequency for a signal to be perfectly sampled.

In this work, an implementation of the discrete Fourier transform known as the **Fast Fourier Transform (FFT)** is used. To speed up the calculation of the DFT, algorithms such as the FFT proposed by Cooley & Tukey (1965) take a divide-and-conquer approach to reduce the number of calculations required. Such optimisations rely on well-formed data. To improve the complexity from the $O(n^2)$ of the standard DFT to $O(n \log n)$, a time series must be approximately a power-of-two long. Steps such as padding the array with zeros at either end can help to improve the speed of the calculation.

Discrete Fourier transforms are mathematically only defined for a regular set of sampling points $t_{i+1} = t_i + \Delta t$, and hence other methods must be used. One way to combat this problem is to interpolate your signal onto a regular grid, but this will typically result in additional noise and unreliable frequency spectra (VanderPlas 2018).

### 2.2.2 Lomb–Scargle periodogram

The **Lomb–Scargle (LS) Periodogram** (Lomb 1976; Scargle 1982) attempts to solve the problems of approximating frequency spectra for unevenly sampled time series data. The LS

periodogram is effectively a least-squares optimisation of a sinusoidal model for each frequency. For a given frequency $\nu$, we propose a sinusoidal model of the form

$$y(t;\nu) = A_\nu \sin(2\pi\nu(t-\phi_\nu)) \tag{2.4}$$

where the amplitude $A_\nu$ and phase $\phi_\nu$ can vary with frequency. A least-squares optimisation is performed to fit the model, by minimising the $\chi^2$ statistic at each frequency:

$$\chi^2(\nu) \equiv \sum_{\mathrm{n}} (y_{\mathrm{n}} - y(t_{\mathrm{n}};\nu))^2 \tag{2.5}$$

where $y_{\mathrm{n}}$ is the value of the time series at the $n^{\mathrm{th}}$ data point. The work of Scargle (1982) was to combine these functions and additionally propose an arbitrary phase term $\tau$ to generalise the frequency dependant $\phi_\nu$ phase terms. This creates a periodogram of the form

$$P(\nu) = \frac{A^2}{2}\left(\sum_{\mathrm{n}} y_{\mathrm{n}} \cos(2\pi\nu(t_{\mathrm{n}}-\tau))\right)^2 + \frac{B^2}{2}\left(\sum_{\mathrm{n}} y_{\mathrm{n}} \sin(2\pi\nu(t_{\mathrm{n}}-\tau))\right)^2 \tag{2.6}$$

where $P(\nu)$ is the periodogram power and $A, B$ and $\tau$ are functions of the frequency $\nu$ and observing times $\{t_{\mathrm{i}}\}$. It can be shown that selecting functions for $A, B$ and $\tau$ such that

1) the periodogram reduces to the classic form in the case of equally-spaced observations,
2) the periodogram's statistics are analytically computable and
3) the period is insensitive to global time-shifts in the data,

will produce an LS periodogram of the form:

$$P_{\mathrm{LS}}(\nu) = \frac{1}{2}\left\{\frac{(\sum_{\mathrm{n}} y_{\mathrm{n}} \cos(2\pi\nu(t_{\mathrm{n}}-\tau)))^2}{\sum_{\mathrm{n}} \cos^2(2\pi\nu(t_{\mathrm{n}}-\tau))} + \frac{(\sum_{\mathrm{n}} y_{\mathrm{n}} \sin(2\pi\nu(t_{\mathrm{n}}-\tau)))^2}{\sum_{\mathrm{n}} \sin^2(2\pi\nu(t_{\mathrm{n}}-\tau))}\right\}, \tag{2.7}$$

$$\tau = \frac{1}{4\pi\nu}\tan^{-1}\left(\frac{\sum_{\mathrm{n}} \sin(4\pi\nu t_{\mathrm{n}})}{\sum_{\mathrm{n}} \cos(4\pi\nu t_{\mathrm{n}})}\right). \tag{2.8}$$

The LS periodogram provides an extremely useful and often accurate tool for extracting periodicity from time-series data, with the caveat that the underlying model is sinusoidal. Hence, the produced periodogram is a linear combination of sines that may not accurately model all signal shapes. Phase shifting signals, such as the brightness variations of rotating spotted stars, are one example of a signal shape not well modelled by a linear combination of sines. I also note that the statistical guarantees of the LS periodogram only hold in the case that observations contain only uncorrelated white noise. The LS periodogram is used widely in extracting periodicity from astrophysical time series. Large numbers of variable stars have been detected using such methods, such as The Zwicky Transient Facility (ZTF) catalogue of periodic variable stars (4.7 million stars, Chen et al. 2020), the Asteroid Terrestrial-impact

Last Alert System (ATLAS) catalogue (621,702 stars, Heinze et al. 2018) and the ASAS-AN variability catalogue (687,695 stars, Shappee et al. 2014, and Jayasinghe et al. 2019 through to Jayasinghe et al. 2020).

### 2.2.3 Autocorrelation function

The following section is taken from Briegal et al. (2022). Where the shape of the underlying signal is not known, and in particular not sinusoidal, self-similarity may be an appropriate method to extract periodicity. The **autocorrelation function (ACF)** is a measure of how similar a signal is to itself shifted by a time lag $k$. Shumway & Stoffer (2017) define the ACF of a regularly sampled time series $X_{\mathrm{I}}(t)$ as the function

$$\rho : \{0, 1, \ldots, i_{\max}\} \rightarrow [-1, 1] \tag{2.9}$$

$$\rho(k) := \frac{1}{N} \sum_{i=0}^{i_{\max}-k} (X_{\mathrm{i}} - \langle X_{\mathrm{I}} \rangle) \times (X_{i+k} - \langle X_{\mathrm{I}} \rangle) \tag{2.10}$$

where $\langle X_{\mathrm{I}} \rangle$ denotes the mean of the time series values and the normalisation $N$ is the total sum of squares $N := \sum_{i \in I} (X_{\mathrm{i}} - \langle X_{\mathrm{I}} \rangle)^2$. The choice of this normalisation implies that $\rho(0) \equiv 1$, i.e. a time series is maximally similar to itself when there is no lag.

While this is a standard way to introduce the ACF, it can be useful to think of the ACF as a function with a time domain instead of an integer lag domain. We can make this domain modification explicit by multiplying the argument by the sampling constant $\Delta t$:

$$\rho(k\Delta t) : \{0, \Delta t, \ldots, \Delta t \cdot i_{\max}\} \rightarrow [-1, 1]. \tag{2.11}$$

These descriptions are equivalent, but the latter view is more useful for the work presented in this thesis. Specifically, the ACF is only directly applicable to time series where the sampling is regular. However, it does not rely on underlying model assumptions, such as the LS periodogram.

The position of the peaks in the ACF must be measured to extract a period from the autocorrelation function, as the autocorrelation of a periodic function is itself periodic with the same period as the underlying function. The ACF has been used to extract periodic variability from photometric data taken with Kepler and K2 (McQuillan et al. 2014; Gordon et al. 2021) and forms the basis of the technique developed as a part of this thesis, the Generalised Autocorrelation Function (G-ACF, Chapter 4).

### 2.2.4 Phase folding and phase dispersion minimisation

**Phase Dispersion Minimisation (PDM)** is a method based on phase- or epoch-folding of data. Phase folding in this context transforms a function in time into a function in phase, where the phase of a time point $t_i$ ($\phi_i$) is defined as

$$\phi_i(P) \equiv \left| \frac{t_i - t_0}{P} \right| \tag{2.12}$$

for a period $P$ and an arbitrarily chosen epoch $t_0$. A light curve can be 'folded' onto itself by plotting the light curve in phase and overlaying successive periods of data. PDM (Stellingwerf 1978) selects the period for which the phase folded light curve has the least scatter. This scatter is formally defined as a variance over a set of observations such that

$$\sigma^2 = \frac{\sum (x_i - \bar{x})}{N - 1} \tag{2.13}$$

for a set of $N$ observations $\{(t_i, x_i)\}$ representing a time value and a measurement value respectively, where $\bar{x}$ represents the mean measurement value. The phase folded observations are binned for each candidate period, and a scatter is calculated in each bin. The total scatter is the sum across all bins, and the candidate period with the smallest total scatter is selected as the 'correct' period. PDM does not rely on model assumptions such as the LS periodogram and does not require uniform sampling. It can be computationally expensive to run without prior knowledge of the period, as for all candidate periods, a phase fold and set of scatters must be calculated.

### 2.2.5 Bayesian approaches

Bayes' theorem is used to calculate conditional probabilities: for two events, $A$ and $B$, Bayes' theorem is stated mathematically as

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}. \tag{2.14}$$

$P(A \mid B)$ is a conditional probability, the probability that A occurs given B is true. This is known as the *posterior* probability of $A$ given $B$. $P(B \mid A)$ is known as the *likelihood*, and $P(A)$ and $P(B)$ are *marginal* or *prior* probabilities, without any conditions. The power of Bayes' theorem is revealed in Bayesian inference when fitting a statistical model to data. In this context, Bayes' theorem is used to calculate the probability for a hypothesis given a set of data as evidence. Mathematically, for a set of observed parameters $\boldsymbol{X}$ and a model with a set of parameters $\boldsymbol{\theta}$ we can calculate the posterior probability of the parameters given the observations

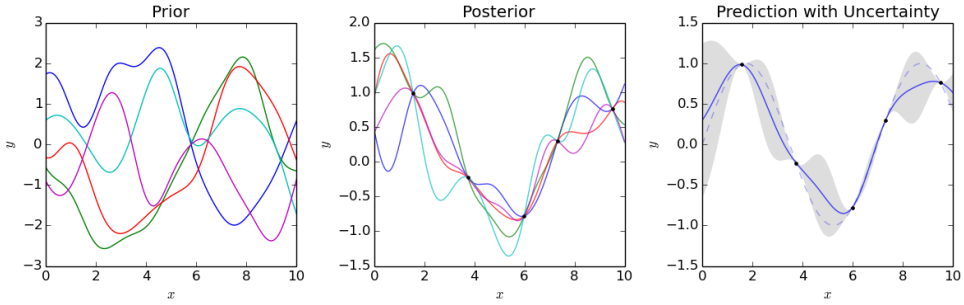



Figure 2.10: An example Gaussian process with a squared exponential kernel (Equation 2.16). Function draws from the (uniform) prior distribution are plotted in the left panel. Function draws from the posterior distribution conditioned on the data are plotted in the centre panel. The right panel shows the mean posterior prediction in blue, with shaded grey regions representing 1 standard deviation spread. Credit: Wikimedia user Cdipaulo96.

as

$$P(\boldsymbol{\theta} \mid \boldsymbol{X}) = \frac{P(\boldsymbol{X} \mid \boldsymbol{\theta})P(\boldsymbol{\theta})}{P(\boldsymbol{X})}. \tag{2.15}$$

Within an inference setting, the likelihood is the distribution of the observed data conditional on the model parameters, i.e. the probability of generating an observation $x_i \in \boldsymbol{X}$ given a model $\psi(\boldsymbol{\theta})$. The prior, $P(\boldsymbol{\theta})$, is the distribution of model parameters before any observed data. Careful thought must be given to ensure model priors are physically motivated and do not introduce large biases into the posterior distribution. Finally, the denominator, $P(\boldsymbol{X})$, is known as the marginal likelihood or the evidence. This is the distribution of the observed data marginalised over the parameter, and in general, the evidence is non-trivial to compute.

It is possible to introduce Bayes' Theorem into period finding methods, which allows the calculation of probabilistic errors on period estimates driven by prior knowledge of your observations. Such methods include Gaussian process regression which will be discussed in more detail or may be derived from other methods already mentioned, such as Bayesian periodograms (which will not be discussed in this thesis).

### 2.2.6 GP regression

**Gaussian process (GP) regression** is a somewhat different approach to time series modelling and periodicity detection. It is a non-parametric, Bayesian approach to regression. Instead of optimising over sets of model parameters that best fit the data, it optimises over sets of functions that best describe the data (Rasmussen & Williams 2006). Here non-parametric does not mean there are no parameters, but there are infinite possible parameters that are not defined; the number of parameters will change with the underlying model and data set size. We define a prior distribution, representing the expected outputs of a set of functions without any observed

data. For a Gaussian process, these functions are comprised of draws from multivariate normal (MVN) distributions. The underlying assumption here is that every data point is drawn from a multivariate Gaussian, and our time series consists of a set of draws from many MVNs. It is necessary to define *covariance kernels* to constrain the relationship between subsequent time points in our data, which defines how two samples from our MVNs at different times are related. One such example of a covariance kernel is the radial basis function (RBF) or squared-exponential kernel, defined as

$$\mathrm{cov}(x_\mathrm{i}, x_\mathrm{j}) = A \exp\left(-\frac{(x_\mathrm{i} - x_\mathrm{j})^2}{2l^2}\right) \tag{2.16}$$

where $x_\mathrm{i}$ and $x_\mathrm{j}$ are measurements taken at times $t_\mathrm{i}$ and $t_\mathrm{j}$ and $A$ and $l$ are the kernel hyperparameters. Kernel hyperparameter tuning forms a large basis of the modelling involved in GP regression. Selecting a relevant kernel that best fits the data is an important part of GP modelling. More complex covariance kernels such as Quasi-Periodic (QP) kernels can model complex periodic signals. One such definition of a QP kernel from Rasmussen & Williams (2006) is

$$k_{i,j} = A \exp\left[-\frac{(x_\mathrm{i} - x_\mathrm{j})^2}{2l^2} - \Gamma^2 \sin^2\left(\frac{\pi(x_\mathrm{i} - x_\mathrm{j})}{P}\right)\right] + \sigma^2 \delta_\mathrm{ij}. \tag{2.17}$$

Here, $P$ can be interpreted as a rotation period with $\Gamma$ controlling the amplitude of the $\sin^2$ term. $\Gamma$ parameterises how strongly correlated points lying one period apart are: large values of $\Gamma$ impose a more strict periodicity than small values, for which points separated by more or less than a period are still correlated. Additionally, a white noise term $\sigma$ captures any remaining jitter not well modelled by the GP. The paper by Angus et al. (2018) provides further insight into this QP kernel.

GP regression has proven to be an extremely effective method of accurately modelling complex stellar activity both photometrically (e.g., Evans et al. 2015; Vanderburg et al. 2015; Grunblatt et al. 2016; Aigrain et al. 2016; Littlefair et al. 2017; Gillen et al. 2020) and spectroscopically (e.g., Gibson et al. 2012; Rajpaul et al. 2015, 2016; Angus et al. 2018). Utilising complex kernels such as the QP kernel described above in addition to other kernel forms, it is possible to model not only the rotation or activity of a star but also correlated and uncorrelated noise arising from both astrophysical and instrumental sources.

A popular modelling framework with astronomy is `EXOPLANET` (Foreman-Mackey et al. 2021), which in turn uses `Celerite2` (Foreman-Mackey 2018) for model terms[c], `pymc3` (Salvatier et al. 2016) for probabilistic inference and `THEANO` (The Theano Development Team et al. 2016)

[c] As explained in Foreman-Mackey et al. (2017), Celerite optimises GP calculations for speed by restricting the set of available kernels to a specific class generated by a mixture of exponentials.

for efficient model evaluation. I will further explain the phrase 'probabilistic inference' in the context of GP regression and MCMC sampling in Section 2.3.4.

Stellar variability signals, particularly rotational signals, have been well modelled using a sum of simple harmonic oscillator (SHO) kernels that are efficient to calculate (Foreman-Mackey et al. 2017; Foreman-Mackey 2018). This kernel is used as part of the variability detection pipeline in Chapter 6; I will explain it in detail here. The kernel contains a single SHO term of the form

$$S(\omega) = \sqrt{\frac{2}{\pi}} \frac{S_0^2 \omega_0^4}{(\omega^2 - \omega_0^2)^2 + \omega_0^2 \omega^2 / Q^2}, \tag{2.18}$$

where $S_0$ is the power at $\omega = 0$, $Q$ is the quality factor, and $\omega_0$ is the undamped angular frequency of the system driven at an angular frequency $\omega$. Added to this SHO term is a `Celerite2 RotationTerm`[d] which is a mixture of two SHO terms used to model stellar rotation. The term contains two modes, one at the period and another at half the period, describing many astrophysical variability signals. The term is modelled precisely using the following parameters:

$$Q_1 = 1/2 + Q_0 + \delta Q \tag{2.19}$$

$$\omega_1 = \frac{4\pi Q_1}{P\sqrt{4Q_1^2 - 1}} \tag{2.20}$$

$$S_1 = \frac{\sigma^2}{(1+f)\omega_1 Q_1} \tag{2.21}$$

for the primary term and

$$Q_2 = 1/2 + Q_0 \tag{2.22}$$

$$\omega_2 = \frac{8\pi Q_1}{P\sqrt{4Q_1^2 - 1}} \tag{2.23}$$

$$S_2 = \frac{f\sigma^2}{(1+f)\omega_2 Q_2} \tag{2.24}$$

for the second term. $\sigma$ is the standard deviation of the process, $P$ is the primary period of variability, $Q_0$ is the quality factor for the secondary oscillation. $\delta Q$ is the difference between the quality factors of the first and second modes, which is constrained to ensure the primary mode always has higher quality. $f$ is the fractional amplitude of the secondary mode compared to the primary and should be $0 < f < 1$ to ensure the secondary mode has a smaller amplitude than the primary. Additionally, white noise is modelled by adding a diagonal log-jitter term to the kernel.

[d] `https://celerite2.readthedocs.io/en/latest/api/python/#celerite2.terms.RotationTerm`. Accessed: 22/01/2022.

## 2.3 Machine learning for automatic detection and classification

The phrase *Machine Learning* has been accredited to Arthur Samuel from his 1959 paper to describe a subset of artificial intelligence that utilises statistical methods to enable computers to *learn* through progressively improving performance on a task using data rather than explicit programming instructions. Since then, along with the exponential increase in computing power, machine learning methodology has been applied to many fields, including scientific research, health care, finance, sports betting, social media and more recently, the rise of automated systems such as self-driving cars and smart assistants.

Machine learning is becoming a much more widely used tool for solving big data problems within astrophysics. Many different methods have been applied successfully, such as Generative Adversarial Networks (GAN) for galaxy detection (Schawinski et al. 2017), Self Organising Maps (SOM) and Random Forests (RF) for exoplanet detection (Armstrong et al. 2016, 2017, 2018) and Neural Networks (NN) for gravitational wave detection and parameter estimation (George & Huerta 2018) and exoplanet detection and parameter estimation (Shallue & Vanderburg 2018).

Machine learning algorithms fall broadly into two categories: *supervised* and *unsupervised* learning[e]. There also exists a branch of machine learning based on probabilistic modelling, which includes Gaussian process Regression and Markov-based models.

### 2.3.1 Supervised learning

Supervised learning algorithms rely on user-inputted *Ground Truth* data with initial solutions (a training data set). This training set specifies a pre-defined output for each input. This output could be a discrete class (a classification problem) or a continuous value (a regression problem). The algorithm aims to learn a pattern or rule that maps the inputs to the outputs.

Many different algorithms exist, ranging from simple optimisation to complex neural networks that mimic the human brain's connections, all of which broadly use a large number of simple calculations to solve complex problems. A selection of well-defined and commonly used supervised algorithms will be outlined in this Section, and Section 2.3.2 will outline some unsupervised learning algorithms.

#### 2.3.1.1 Decision trees

A decision tree is an intuitive classification tool. The input is at the tree's root, with further 'branches' which can be evaluated and taken. As we move down the branches, we make further

[e]There are other types of algorithms, for example, *semi-supervised* learning, but these will remain outside of the scope of this thesis.

decisions until we reach a 'leaf' node at the end of the tree. This leaf represents an output or class. Decision trees are fairly simple to understand: they resemble human decision-making as an iterative process. They could be classified as 'white-box' learning, as we can see what decisions each node makes. Decision trees are relatively basic models, and as such, they are often prone to overfitting the training data and are not robust to changes in this training data. One can improve simple decision trees by using ensemble methods, combining simple algorithms to create a more robust ensemble method that forgoes some of the shortcomings of an isolated decision tree.

#### 2.3.1.2 Ensemble methods

To prevent models from overfitting the data, often a useful tactic is to combine model outputs. This can be done simply through averaging or by applying more complex methods such as bootstrapping. One well-known and often used ensemble method is a *random forest*, which improves the decision tree algorithm described previously. A random forest constructs many decision trees by randomly distributing where features are split in each tree. The output is determined as the average output of the forest of trees, and as such, should be robust to overfitting, unlike a single tree. As a corollary, the random forest can map the relative importance of features (i.e. decisions at each node) based on the collection of trees. RFs have been employed successfully by Armstrong et al. (2018) in exoplanet transit detection using NGTS data; by selecting a set of features from observed candidate transit events, a random forest was able to predict the likelihood that these candidate events were real planetary transits. Additionally, the use of an RF allowed for feature importance ranking, highlighting where the RF model gained the most information. In this case, a transit shape statistic, the signal-to-noise ratio and the transit depth were identified as key features. Jayasinghe et al. (2018) used a random forest approach to generate the ASAS-SN variable star catalogue. To classify variable stars, the authors used variability features such as period, amplitude and Fourier information in addition to stellar parameters.

#### 2.3.1.3 Neural networks and deep learning

Deep learning is a branch of machine learning based on algorithms that attempt to model high-level abstractions in data using multiple processing layers, with a complex structure composed of multiple non-linear transformations. Figure 2.11 shows a neural network composed of 3 layers: one input, one hidden and one output; in this example, 4 inputs provide one outcome.

Many network architectures exist and are often application dependent. Common to all neural networks is a graph of connected nodes, where each node applies a non-linear transformation to

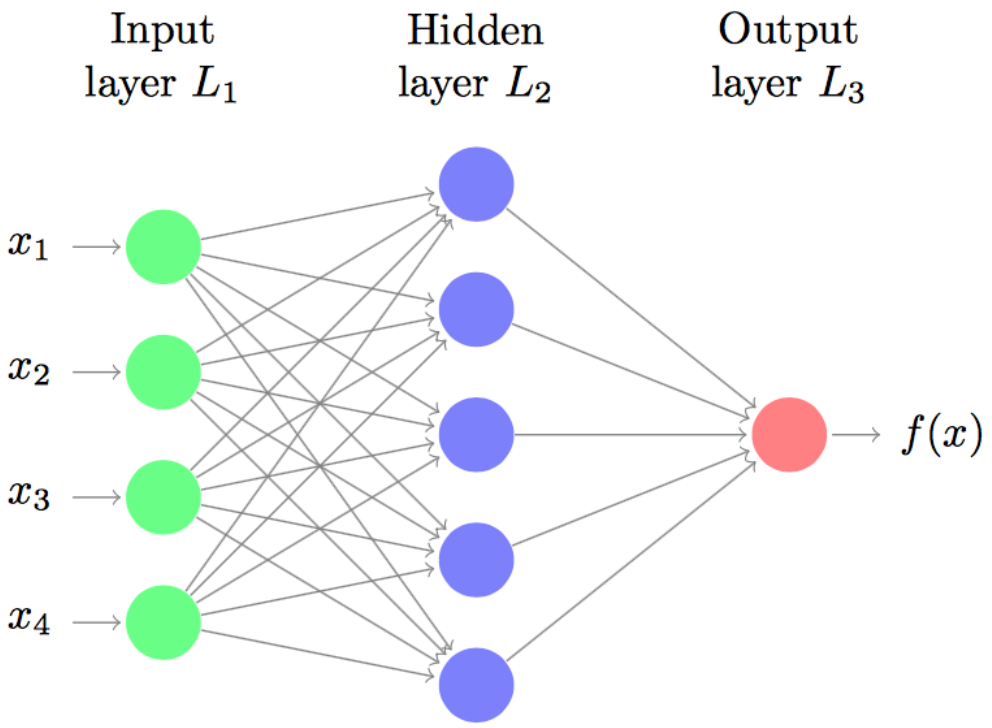



Figure 2.11: A cartoon of a neural network diagram with a single hidden layer. The hidden layer derives transformations of the inputs (non-linear transformations of linear combinations), which are then used to model the output. Credit: Efron & Hastie (2016).

the input, which is a linear combination of previous outputs. The non-linear transformation is known as an *activation function* and can take many forms, most commonly a sigmoid function, a hyperbolic tangent or a ReLU function of the form $\phi(z_i) = \max(0, z_i)$[f].

The training of a neural network is conducted by optimising the weight and/or bias parameters which combine a layer into inputs for the next layer. The general form for layer $k$ would be:

$$x_i^{(k+1)} \equiv y_i^{(k)} = \phi(\boldsymbol{w}_i \cdot \boldsymbol{x}^{(k)} + b_i) = \phi(z_i^{(k)}) \tag{2.25}$$

where $\boldsymbol{w}_i$ is a vector of weights and $b_i$ is a vector of biases to be optimised.

These parameters are optimised subject to the training data and a cost/loss function using a technique known as *Error Backpropagation* (Rumelhart et al. 1986). Backpropagation enables the calculation of the gradient of the cost function with respect to the weights and iteratively descends upon the optimum network weights. The exact architecture of a neural network will alter the input data and task to which a network is best suited. For example, **convolution neural networks (CNN)** use a series of convolution layers to retain local spatial structure

[f]There are many excellent introductions to the pros and cons of NN activation functions, such as https://www.v7labs.com/blog/neural-networks-activation-functions (accessed: 22/01/2022).

within image data by convolving nearby points and pooling the output. For time-series data, such as natural language processing tasks, a **recurrent neural network (RNN)** architecture may be more appropriate. Such networks contain a casual memory encoding through history vectors, such that the previous set of inputs alters the output of a given layer. Networks such as **autoencoder** networks can be used to reduce dimensionality. Encoding and subsequently decoding an input vector can often find a lower-dimensional representation of the input, which encodes all of the information necessary to decode the vector accurately. It is important to ensure your network architecture is optimised for the task when utilising deep learning methods. This architecture optimisation is an open topic of research for machine learning experts. Most applications choose the architecture that best fits their data set, tested against a few alternatives.

Neural networks have been applied successfully to photometric light curves, most prominently CNNs in the `astronet` series of papers for exoplanet discovery (Shallue & Vanderburg 2018; Ansdell et al. 2018; Dattilo et al. 2019; Osborn et al. 2020). The use of neural networks within this context affords two benefits: firstly, discovering new planets not previously seen by manual eyeballing such as Kepler-80g and Kepler-90i (Shallue & Vanderburg 2018) and secondly, reducing manual eyeballing time required to identify planet transits. With the addition of scientific domain knowledge, the ML model can perform considerably better (Ansdell et al. 2018), and this state-of-the-art NN model has been applied successfully to TESS data more recently (Osborn et al. 2020). A CNN-based approach has been applied to the transit identification stage of NGTS (Chaushev et al. 2019) and currently operates as a vetting stage of the NGTS transit detection pipeline.

### 2.3.2 Unsupervised learning

Unsupervised learning algorithms rely solely on the data; there is no training data set as with supervised learning algorithms. The algorithm must decide for itself what the outputs are. Examples of unsupervised machine learning applications are cluster detection, anomaly detection and expectation maximisation; some specific examples are discussed below.

#### 2.3.2.1 k-means clustering

k-means clustering partitions the data into $k$ clusters such that each point belongs to the cluster with the nearest mean (or centroid). It is an extremely simple algorithm to implement; however, it is non-trivial to solve computationally. It also relies on a fixed number of means as an input parameter. However, given considerable computing resources, it is possible to determine this number by calculating how well points lie within clusters for different values of $k$, such as

with a 'silhouette' score (Rousseeuw 1987) which compares how well points lie within their allocated cluster compared to nearby clusters.

#### 2.3.2.2 Density-based clustering

A problem with simple clustering algorithms like k-means is that loosely related observations may be clustered together based on the nearest centroid. The problem can be alleviated by considering the spatial distribution of observations and the density of observations; a cluster in data space can be thought of as a contiguous region of high density, separated from other clusters by a contiguous region of low density. The use of data density as an additional metric is the principle upon which the Density-Based Spatial Clustering of Applications with Noise algorithm (DBSCAN, Ester et al. 1996) is derived.

DBSCAN requires two parameters: $\epsilon$ and `MinPoints`. $\epsilon$ defines the radius of a hypersphere centred at each point in which the density should be considered, and `MinPoints` defines the minimum number of data points within that sphere for that data point to be considered a *core* point. Points are separated into three classes: core, border and noise. Border data points have fewer than `MinPoints` nearby, and noise data points have no nearby data points. A cluster is defined as a group of points that are 'density-connected'; two points are density-connected if there exists a core point that is density-reachable (connected via a series of core points) from both points. Every core point will be assigned to a new cluster unless they share the space with other core points, in which case they will be clustered together.

These simple rules and parameters generate excellent density-based clustering maps of complex data sets, such as those demonstrated in Figure 2.12. DBSCAN is extremely sensitive to the values of the parameters $\epsilon$ and `MinPoints`, and more sophisticated algorithms such as HDBSCAN (Hierarchical DBSCAN, Campello et al. 2013) seek to solve this problem by heuristically assessing the density distribution of the data points to ascertain optimal thresholds for cluster separation.

DBSCAN has been shown to effectively cluster real-world data, with applications in gene clustering (Edla & Jana 2012), anti-money laundering transaction detection (Yang et al. 2014) and astrophysics (Kounkel & Covey 2019; Kounkel et al. 2020; Cánovas et al. 2019; Hunt & Reffert 2021).

#### 2.3.2.3 Self-organising maps

A self-organising map (SOM)/ Kohonen Map (Kohonen 2001) is an unsupervised neural network. The algorithm maps high dimensional data into a low dimensional feature space. It calculates the shortest distance in this space from a point to a feature to classify or cluster the

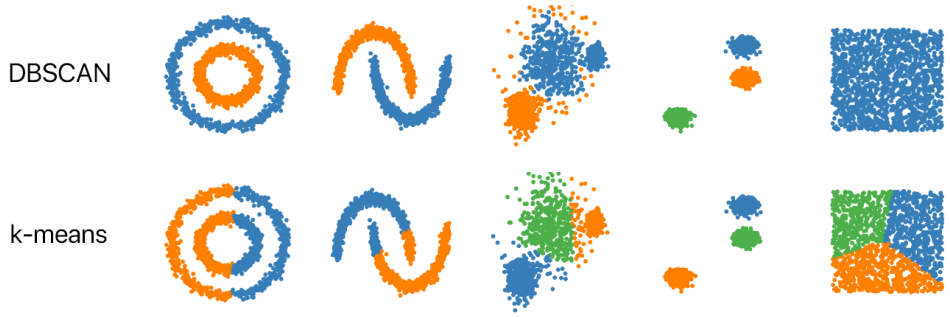



Figure 2.12: A comparison of the results of DBSCAN and k-means clustering algorithms on a variety of toy data sets. Credit: https://github.com/NSHipster/DBSCAN. Accessed: 22/01/2022.

data. It is often useful to consider this Kohonen layer as a 2d grid of pixels. Each new data point encountered is projected onto this layer and added to the closest pixel (for example, by calculating Euclidian distance). SOMs have been employed successfully in transit detection; Armstrong et al. (2016) demonstrated the use of SOMs in conjunction with RFs to classify variable stars.

### 2.3.3 Probabilistic models

Machine learning algorithms may be used to fit probabilistic models to data. These can be either supervised or unsupervised for classification or regression. A commonly used model within astrophysics is the Gaussian process model (Section 2.2.6), which can be used to create well-defined probabilistic predictions and interpolations of data sets. It is often impossible to directly infer values with such complex probabilistic models, so approximation methods must be used. One such way of approximating complex distributions is to sample the distribution randomly and build up an approximation to the distribution from these samples. Generating a set of independent draws from a distribution is known as Monte Carlo sampling. With small changes to the assumptions made on each draw, it is possible to improve the performance of Monte Carlo sampling.

### 2.3.4 Markov models

The Markov assumption is that a model state relies solely on the previous state of the model. The simplest Markov model is a *Markov Chain*, which models a random variable changing through time and uses the Markov property to stipulate that the distribution for this variable depends only on the distribution of the previous state. Applying the Markov assumption to

Monte Carlo sampling allows for efficient random sampling of high-dimension probability distributions; Markov Chain Monte Carlo (MCMC) samplers draw from a distribution where each draw depends solely on the value of the previous sample. Several different algorithms exist to calculate draws from a given distribution, the mathematical details of which are beyond the scope of this introduction. Broadly, these algorithms will generate samples of a calculable function proportional to the distribution of interest. The distribution of samples should more closely approximate the distribution of interest over time. In the case of the Metropolis-Hastings algorithm (Hastings 1970), this is done by accepting or rejecting each new sample based on dynamically calculated acceptance criteria. For a new sample, drawn randomly from a distribution about the previous sample, the probability of acceptance of the new sample is decided based on the posterior value at this new value. If the new proposal has a higher posterior value than the previous sample, the new sample is accepted. If the new proposal has a lower posterior value than the previous sample, the new sample is accepted with a probability equal to the ratio of the posterior value at the previous sample and the proposal. This algorithm is repeated until enough samples have been generated; the number of samples deemed 'enough' will depend on how fast the chain converges. For a well-formed, single-peaked distribution with good initial parameter estimates, MCMC can generate a good approximation rapidly. However, in the case of more complex distributions or poor initial estimates, the random nature of the MCMC sampling can mean the chain takes many steps or even fails to converge. Care should be taken when using MCMC samplers to ensure that the chains are well-formed and appear to converge sensibly. MCMC samplers are used frequently within computational astrophysics to optimise model fits such as Gaussian processes.[g]

[g] One popular package is `emcee`, an efficient `Python` based MCMC implementation widely used within exoplanet research (Foreman-Mackey et al. 2013) and used extensively in Chapter 6 of this thesis.

Chapter 3

# The Next Generation Transit Survey (NGTS)

As outlined in Wheatley et al. (2013), 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, particularly for stars that are sufficiently bright for radial velocity confirmation and mass determination. The purpose of this is to better populate the exoplanet mass–radius space with information on density and bulk composition with planets which are suitable candidates for atmospheric structure and composition follow-up. An important secondary scientific goal of NGTS is for efficient ground-based follow-up of candidates identified in space-based transit surveys. NGTS has a finer pixel scale than space telescopes such as TESS (Ricker et al. 2014) and the future PLATO mission (Rauer et al. 2014), which allows resolution of blended sources, detection of single transit events and ephemerides refinement of systems suitable for observation with future telescopes such as JWST (Gardner et al. 2006).

The science goals of NGTS have altered since its inception, even throughout my PhD. Working groups with different scientific goals now use NGTS data for more than just exoplanet detection and characterisation. These working groups include the traditional candidate selection and follow-up group, as well as groups targeting:

- bright star observations;
- faint star observations;
- open cluster observations;
- M dwarf observations;
- monotransit detection;

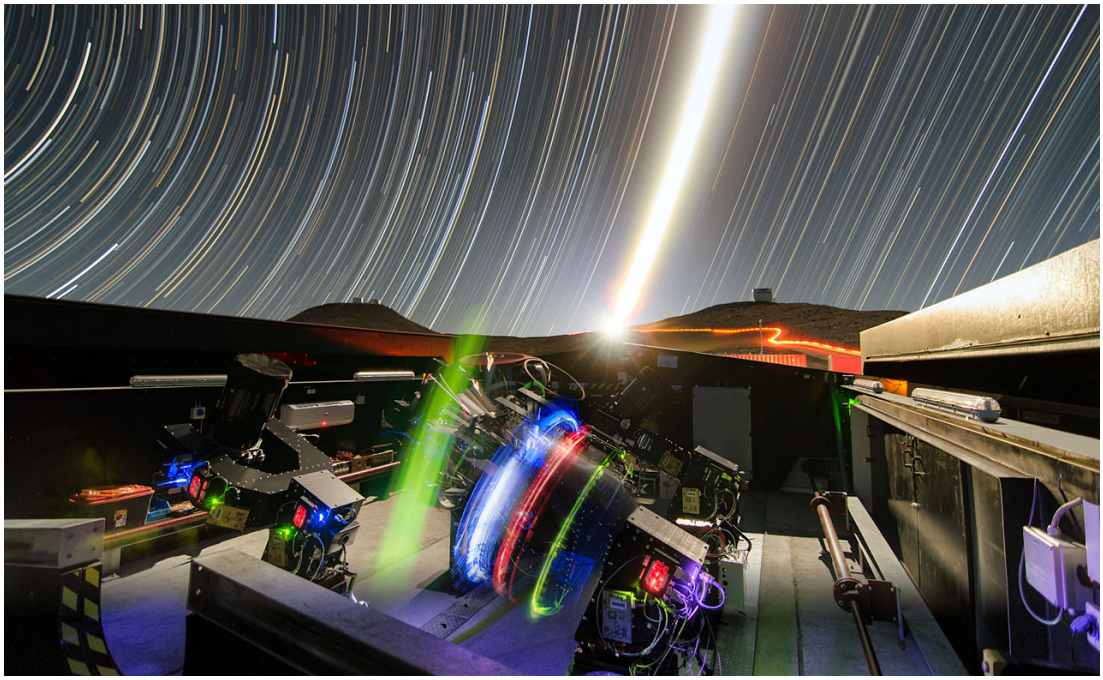


Figure 3.1: This nighttime long-exposure view shows the NGTS telescopes during testing. The brilliant Moon appears in the centre of the picture, and the VISTA (right) and VLT (left) domes can also be seen on the horizon. (Credit: ESO / G.Lambert).

- TESS collaboration and follow-up;
- photometric precision and
- citizen science projects.

These working groups have produced a plethora of different science in addition to the 18 and rising confirmed new planet detections: The detection of giant quasi-periodic flare signals on M-dwarf stars (Jackman et al. 2019a), confirmation of single transit candidates detected with TESS (Gill et al. 2020), eclipsing M-dwarf systems close to the hydrogen-burning limit (Acton et al. 2020), rotation within the Blanco 1 open cluster (Gillen et al. 2020), nanoflare signatures on flaring stars (Dillon et al. 2020) and long-period modulation of accreting white dwarf stars (Chote et al. 2021). The use of high-cadence, high-precision photometry from NGTS for astrophysical analysis outside of transiting planet detection is a driving force behind the scientific goals of this thesis.

## 3.1 Design

The design of NGTS was motivated by the consortium's primary science goal. Based on previous ground-based surveys, most transiting planets identified have around a 1 per cent transit depth and a few with significantly shallower transits. Follow-up observations can often

achieve considerably better than this, up to sub-mmag precision (for example, the <0.1 per cent depth transits of WASP-5 (Southworth et al. 2009) and WASP-52b (Kirk et al. 2016)) and secondary eclipses of hot Jupiters with 0.1 per cent depths, for example, OGLE-TR-56b (Sing & López-Morales 2009) and WASP-19b (Burton et al. 2012).

A transit depth of 0.1 per cent would correspond to detections of super-Earths around early M dwarfs. The target stars must be sufficiently bright for follow-up radial velocity confirmation and mass determination. This brightness threshold was set based on the visual magnitude thresholds of HARPS (Mayor et al. 2003) and ESPRESSO (Pepe et al. 2021) at around 13 and 15, respectively.

Yield simulations have shown that to detect a sample of tens of small planets, an instrument with an instantaneous field of view (FoV) of around 100 $\deg^2$ is required for a survey lasting a few years (Günther et al. 2017; Wheatley et al. 2013). It was decided that the large field of view should be built up from an array of individual telescopes, as a single telescope with such a large FoV would be subject to large atmospheric refraction effects. A further advantage of the telescope array is that efficient follow-up observations of multiple candidates can be carried out simultaneously. Alternatively, it is possible to maximise the collecting area and photometric precision by pointing all the telescopes at the same target. Bryant et al. (2020b) achieved a photometric precision of 152 ppm per 30 minutes for the bright (T-band magnitude 8.87) star WASP-166 using nine NGTS telescopes simultaneously, matching the precision of simultaneous TESS observations of the transit event.

The NGTS I filter has been designed with a bandpass of 520–890 nm, providing good sensitivity to late K and early M dwarfs. The red cutoff of the filter will minimise variations in the atmospheric extinction caused by strong water absorption bands beyond 900 nm, which are highly variable, even at Paranal, as shown by Noll et al. (2012). This cutoff ensures that the effective bandpass of NGTS is defined primarily by the instrument and not by the sky.

### 3.1.1 Data management

The NGTS data is stored in 4 MySQL databases: operations, data tracking, data reduction and candidate tracking. The operations database contains useful observation metadata such as current time, pointing, focus, autoguiding statistics and environmental data such as weather and the positions of the Sun and Moon. A subset of this data forms the FITS image headers.

The NGTS telescopes generate around 200 GB of image data per night. This data must be transferred to the University of Warwick, UK. The data is first compressed and then stored on 2 TB hard drives, and shipped to Warwick fortnightly. This data is stored on the NGTS cluster in Warwick and on a backup server in Cambridge. A database-driven tracking system jointly manages the servers in Paranal and Warwick to ensure no data is lost during this transfer.

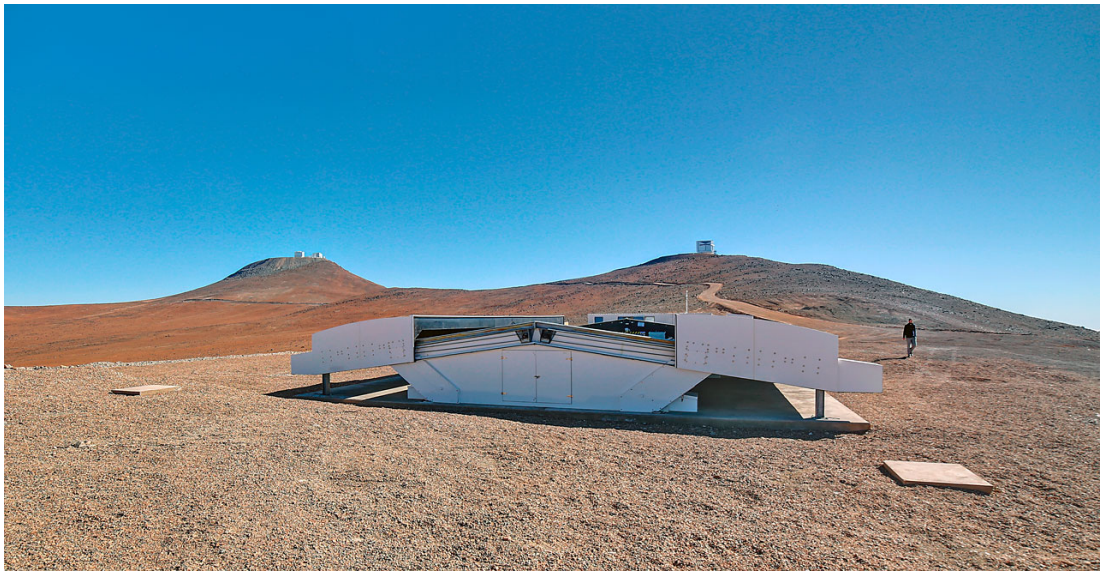


Figure 3.2: This image shows the NGTS enclosure during the day at the ESO Paranal Observatory in Northern Chile. The VISTA (right) and VLT (left) domes can also be seen on the horizon. (Credit: ESO/R. Wesson).

Figure 3.3: Most of the 20-centimetre telescopes that form the survey system are shown in this picture, which was taken during testing. (Credit: ESO/R. West).

The raw photometry and data reduction pipeline outputs (see section 3.3) are stored in the pipeline database in Warwick. The candidate database is used to store measured exoplanet candidate properties, external exoplanet catalogue data and summary statistics.

## 3.2 Operations

NGTS operates completely robotically aside from the human go/no-go decision each night. Once the enclosure roof has opened, flat-field images are taken, followed by an analysis of the optimum focus point for each telescope. Science data is taken when the sun is below $-15°$. Each telescope will operate in either survey or follow-up mode during the night.

**Survey Mode** The telescope will observe a sequence of survey fields. For baseline surveys, these fields are spaced such that one field rises above 30° as the previous field sets below 30°. Each telescope will typically observe two fields per night, resulting in around 500 h coverage per field spread over 250 nights.

**Follow-up Mode** The telescope will target a particular star, placed in the centre of the field to minimise differential atmospheric refraction effects.

For both modes, the default is to observe in focus, with a 10-second exposure time. A few special observation programmes exist beyond these two modes, particularly relevant for open clusters. I took the data used in Chapter 6 from cluster field observations, in which a single telescope was centred on a known open cluster.

### 3.2.1 Field selection

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

## 3.3 Data reduction and analysis

The NGTS data reduction pipeline is a custom-built, modular program run at the University of Warwick. A catalogue of target stars is generated, the night's science images are bias-corrected and flat-fielded, astrometric solutions are found, and photometric measurements are made. I will discuss each of these steps within the next sections. Once the data for a field have been reduced and photometric measurements made per science image, a light curve is assembled

per target star. This light curve is then detrended for red noise sources, and an exoplanet transit search is conducted. These detections are vetted, and the best candidates are selected for follow-up with further photometric and additional spectroscopic observations.

### 3.3.1 Catalogue generation

Source detection is done using the IMCORE module in CASUTOOLS (Irwin et al. 2004) to generate an object list that is cross-matched against other catalogues. NGTS generates its own input source catalogue, as explained in Section 5 of Wheatley et al. (2018). For pipeline runs before 2021 (including the data used in Chapter 5), the pipeline used this internally generated catalogue to generate light curves, with any catalogue cross-matching done post-pipeline. For more recent runs, including the data used in Chapter 6, this source catalogue is cross-matched against several external catalogues, including the Tess Input Catalogue (TICv8) and some consortium curated special target catalogues, including clusters, WDs and other astrophysical objects of interest. In both cases, cross-matching is done in position, colour and separation to limit spurious matches. Separation cross-matching allows flagging of potential unresolved binaries in NGTS apertures.

### 3.3.2 Astrometry

A full astrometric solution is required for each image to account for the stretching of the field resulting from atmospheric refraction and field rotation due to imperfect polar alignment. Individual images are solved for translation, rotation, skews and scales using the WCSFIT program from the CASUTOOLS software suite (Irwin et al. 2004). The 2-MASS catalogue is used for reference, with an initial estimated astrometric solution taken from astrometry.net (Lang et al. 2010).

### 3.3.3 Photometry

Photometric measurements are made using aperture photometry with the CASUTOOLS IMCORE_LIST program (Irwin et al. 2004). For each star in the input catalogue, a soft-edged circular aperture of radial 3 pixels (15 arcsec) is placed using the per-image astrometric solutions.

### 3.3.4 Light curve detrending

Light curve detrending is done with a custom implementation of the SysRem algorithm (Tamuz et al. 2005), based on the version used by the WASP project (Cameron et al. 2006). SysRem will remove signals common to multiple stars and is amplitude independent. A mean light curve is

calculated and used to correct first-order offsets seen in all the stars observed. SysRem does not completely remove systematic signals correlated with the Moon phase and sidereal time as these can have different shapes for different stars. Moon phase correlated signals generally arise from imperfect sky subtraction or low-level non-linearity of the detectors. Signals correlated with sidereal time can arise from airmass changes affecting sub-pixel movements of stars caused by differential atmospheric refraction. Alternatively, such systematics could arise from imperfect flat-fielding or sub-pixel sensitivity variations. 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. Significant periodic signals (which could arise from stellar variation, for example) are identified and removed from the light curves if they do not appear to have a transit shape by subtracting the calculated mean in the phase-folded light curve. This detrending step has been proven to increase transit detection efficiency by 10–30% (Wheatley et al. 2018); however, in this thesis, I elect to use the light curves before this detrending step to retain any significant stellar variability signals.

The NGTS pipeline provides flags per image and per timestamp per object light curve, which I use to pre-process light curves for variability analysis. These flags alert us to bad-quality data points due to pixel saturation, blooming spikes from nearby bright sources, cosmics and other crossing events (including weather and laser guide stars) and any sky background changes.

### 3.3.5 Transit detection

Detrended light curves are searched for transit-like signatures using a Box-Least-Squares (BLS) algorithm, a standard algorithm for basic transit detection. NGTS uses a custom code, ORION, based on the code used by Cameron et al. (2006) for the SuperWASP survey. ORION has several improvements to the Cameron et al. code, namely the fitting of boxes of multiple widths to allow detection of planets in inclined orbits. ORION can combine data from multiple cameras, fields and observing seasons; it also incorporates the Trend Filtering Algorithm from Kovács et al. (2005), which is used to correct systematics arising from common instrumental effects and data reduction anomalies.

Some transit detection pipeline steps are automated to reduce the dependence on manual inspection or 'eyeballing' of sources. A box-least-squares (BLS) periodogram is calculated for each light curve, and a transit model is fitted to significant peaks. Stellar parameters are taken from external catalogues such as Gaia DR2 (Gaia Collaboration et al. 2018c) and the Tess Input Catalogue (Stassun et al. 2019), and a spectral energy distribution (SED) model is fitted where cross-matching is unavailable. Following the work of Chaushev et al. (2019), a CNN model takes the stellar parameters, transit model fit and additional data on the transit and light curve and ranks candidates based on their likelihood of being a transiting exoplanet.

# Chapter 4

# The Generalised Autocorrelation Function (G-ACF)

This chapter is based on the paper *G-ACF: A generalised autocorrelation function for irregularly sampled time series* (Kreutzer et al. *submitted*[a]). I will outline the development of the algorithm and in particular the implementation and testing of the algorithm conducted in 2018. This includes the development of a `C++` and `Python` based implementation available via `GitHub`[b] and `PyPI`[c]. The paper's lead author, Lars Kreutzer, was a maths student working in the department and, with the input of Edward Gillen and Didier Queloz, created the initial mathematical definitions of the G-ACF. During the first year of my PhD, I worked with Ed, Didier and Lars to implement the G-ACF in `C++` and `Python` which involved parameter optimisation, functional form decisions and the formalising of the G-ACF into a scientific paper. My main contribution to this work is implementing and testing the functional forms and parameters used by the G-ACF on both synthetic data and real data from the Kepler mission. Additionally, I was heavily involved in the writing of the G-ACF paper. In particular, I conducted an extensive literature review which I have included as a part of the introduction and motivation for this Chapter. Where I describe work completed solely by me, I will use "I", and where the work was completed by others or as a joint effort I will use "we".

I will present the generalised autocorrelation function, G-ACF, an extended and generalised

[a] Submitted to MNRAS Feb 2022. Received a referee report March 2022 which was generally positive but recommended moderate revisions.

[b] https://github.com/joshbriegal/gacf. Accessed: 22/01/2022.

[c] https://pypi.org/project/gacf/. Accessed: 22/01/2022.

version of the standard autocorrelation function (ACF). G-ACF is a versatile definition that can robustly and efficiently extract periodicity and signal shape information from a time series, essentially independent of both the time sampling and underlying process. Calculating the autocorrelation of irregularly sampled time series becomes possible by generalising the lag of the autocorrelation function to a real parameter and introducing the notion of selection and weight functions. We showed that the G-ACF reduces to the standard ACF in the case of regularly sampled time series. We demonstrated the application of the G-ACF to astrophysical data by extracting rotation periods for KIC 5110407, which agree with periods obtained through other methods. The G-ACF has a wide range of potential applications and will be useful in quantitative science disciplines where irregularly sampled time series occur.

## 4.1 Background

The motivation behind developing a generalisation of the ACF lies in a broader context than just astrophysics. Time series are ubiquitous throughout the experimental sciences, and their analysis plays a central role in research. Fundamentally, they give insight into the temporal evolution of systems and their underlying processes. As we have seen already in this thesis, time series within astrophysics have been instrumental in our understanding of stellar and planetary systems: stellar light and radial velocity curves yield information about the temporal evolution of processes on the stellar surface, from the longitudinal inhomogeneity of starspot distributions and magnetic field mechanisms to the presence of orbiting bodies and material.

I have already introduced several periodicity detection techniques, focusing on either Fourier decomposition (for regularly sampled data) or fitting sinusoidal models (for irregularly sampled data). An example of the former is the Fast Fourier Transform (FFT; Cooley et al. 1969), and examples of the latter are the standard, modified and Bayesian Lomb–Scargle periodograms (Lomb 1976; Scargle 1982; Zechmeister & Kürster 2009; Mortier et al. 2015). While the Lomb–Scargle method can be used for arbitrary samplings, the accuracy of the estimated periods can be limited for quasi-periodic processes and evolving periodic signals due to the inherent assumption that the process is well-described by a pure sine wave of a fixed period. Similar issues affect methods based on phase folding and then minimising the variance or entropy of the data, such as Phase Dispersion Minimisation (PDM), as they also rely on strict periodicity and negligible phase evolution (e.g., Stellingwerf 1978; Graham et al. 2013a,b). More recently, flexible machine learning methods such as Gaussian processes, applicable to both regular and irregular time series, have been used to describe quasi-periodic variations in stellar light curves (e.g., Angus et al. 2018).

The above approaches share the same basic principle: they all fit a model to the data to

determine whether periodicity is present. The concept of autocorrelation, i.e. correlating the data with itself, is a distinct 'model-free' approach that uses only the time series data to extract periodicity (e.g., Shumway & Stoffer 2017). The autocorrelation function (ACF) is a powerful definition and a reliable method to obtain information from any regularly sampled time series. It can capture both strictly periodic and quasi-periodic processes. It has been widely used on space-based photometric data given the regular sampling available (e.g., McQuillan et al. 2013, 2014), as well as on solar data (Morris et al. 2019) for the same reason. However, the requirement of the ACF for regularly sampled data can be a limiting factor in its broader application, e.g., for ground-based photometric data.

Previous studies have attempted to address this problem by generalising the ACF to irregularly sampled data. Several of these methods create an approximately regularly sampled time series to apply the standard autocorrelation function, enhanced with rules on which terms to discard in the series. The method proposed in Lukatskaia (1975) assumes that the irregular sampling arises from missing data points in a regularly sampled series and further assumes that the statistical properties of the missing data are the same as the observed data. Therefore, it is possible to calculate a standard autocorrelation using only data points that fall onto this regular sampling, with the caveat that the time series must be much longer than the variability period of the signal of interest. The method from Andronov & Chinarova (2005) interpolates onto a regular sampling grid using a smoothing function. These methods can work well if the sampling is almost regular and only a subset of values are missing from a regularly sampled time series. As proposed in Edelson & Krolik (1988), The Discrete Autocorrelation Function relies on binning values in time intervals to account for missing overlaps. A similar method was also proposed in Mayo, Jr. et al. (1974) for laser velocimeter research.

As an extension of the binning proposed by Edelson & Krolik (1988) and drawing from the available kernel-based methods proposed by Hall et al. (1994), Stoica & Sandgren (2006) and Bjørnstad & Falck (2001) both use a kernel to weight the product of observations according to the difference between the observation interval and the desired lag bin centre. This technique is also known as 'fuzzy slotting'. The kernels proposed are smooth density functions that tend to zero as lag increases or decreases from the desired lag subject to a characteristic width parameter. Stoica & Sandgren (2006) propose a sinc function, demonstrating the efficacy of this weighting on examples from over-the-internet temperature data, pulsar time-of-arrival measurements and ice core $CO_2$ measurements. Bjørnstad & Falck (2001) use a Gaussian kernel in the context of estimating a spatial autocorrelation for sparse ecological population data. A comparison of correlation-analysis techniques for irregularly sampled time series (linear interpolation, Lomb–Scargle periodogram, correlation slotting and several kernel-based methods), in a geoscientific context, can be found in Rehfeld et al. (2011). These authors find

that while all methods investigated lead to consistent results for time series with a relatively constant sampling density, the kernel-based methods perform better for highly irregular time series.

Related to the problem of finding the ACF of irregularly sampled time series is the problem of finding the power spectral density (PSD); the Fourier transform of the power spectrum of a (stochastic) time series is equivalent to the ACF (see, e.g., Scargle 1989; Merrifield & McHardy 1994).

I will present a different generalisation of the standard autocorrelation function (ACF), named the generalised autocorrelation function, or G-ACF. The G-ACF is an extended and generalised version of the ACF, which applies to both regularly and irregularly sampled time series without making any assumptions about the time sampling or the statistical properties of the data. In particular, there are no assumptions about regularity in the time series sampling.

## 4.2 Basic definitions

I will outline the notation and wording used in Kreutzer et al. *submitted* to describe time series and autocorrelations.

### 4.2.1 Time series

A *time-series* $X_{\mathrm{I}}(t)$ can be defined to be a finite ordered set

$$X_{\mathrm{I}}(t) := \{(X_{\mathrm{i}}, t_{\mathrm{i}}) \in \mathbb{R} \times \mathbb{R}^{+} | i \in I \subset \mathbb{N}, \quad (t_{i+1} - t_{\mathrm{i}}) > 0 \; \forall i \in I\} \tag{4.1}$$

with $I \subset \mathbb{N}$ being a finite index set, which we can choose to be $I = \{0, 1, 2, \ldots, i_{\mathrm{max}}\}$. The set $T_{\mathrm{I}} := \{t_{\mathrm{i}} | i \in I\}$ is the set of *time labels* and the set $X_{\mathrm{I}} := \{X_{\mathrm{i}} | i \in I\}$ the set of *time series values*. In the case of a photometric light curve, the time labels would be the Julian-day time series and the time series values the flux.

It can be useful to think of a time series $X_{\mathrm{I}}(t)$ as a discrete sampling of a continuous process $X(t)$; the notation $X_{\mathrm{i}} = X(t_{\mathrm{i}})$ will be used. Hence, in this definition, I define a time series to be *regularly sampled* if there exists a *sampling constant* $\Delta t > 0$, such that $t_{\mathrm{k}} = t_0 + k \cdot \Delta t \quad \forall k \in I$, else we call the time series *irregularly sampled*.

### 4.2.2 Autocorrelation function (ACF)

We have already seen (in Section 2.2.3) one definition of an autocorrelation function, outlined in Equations 2.9 and 2.10. From these definitions, it is clear that the ACF can only be applied in this form to regularly sampled time series with a discrete set of a lag values $k \in \{0, \Delta t, \ldots, \Delta t \cdot i_{\mathrm{max}}\}$.

Previous efforts have been made to apply the ACF to irregularly sampled data, often employing methods that approximate a regularly sampled time series by imputing missing values to apply the standard ACF. These methods work well for 'near-regular' sampling, where only a few missing values from a regularly sampled time series must be imputed. The accuracy of such efforts will depend on the length of the gaps or irregularities in the time series sampling compared to the scale of structures in the signal. Suppose the time series has large temporal gaps compared to the scale of the underlying process. In that case, it will be very difficult to restore the missing information using interpolation or regression.

The G-ACF was developed to obtain information from arbitrary time series, regardless of their sampling. As the ACF is already applicable to any regularly sampled time series, this provided an excellent formulation to generalise.

## 4.3 The Algorithm

### 4.3.1 Definition

To generalise the ACF on to arbitrarily sampled time series, we introduced two functions: the *selection function* $\widehat{S}$ and the *weight function* $\widehat{W}$, as well as generalising the notion of the lag, $k$, to a *generalised lag*, $\widehat{k} \in [0\,, (\max(T_\mathrm{I}) - \min(T_\mathrm{I}))]$.

The generalised autocorrelation function can then be defined, for a time series of any sampling, to be the function $\widehat{\rho}(\widehat{k};\, \widehat{W}, \widehat{S})$ which, restricted to the generalised lag $\widehat{k}$, is a function of the form

$$\widehat{\rho}(\widehat{k}) : [0, (\max(T_\mathrm{I}) - \min(T_\mathrm{I}))] \rightarrow [-1, 1]. \tag{4.2}$$

A possible generalised definition is given by

$$\begin{aligned}\widehat{\rho}\left(\widehat{k}; \widehat{W}, \widehat{S}\right) :=& \\ \frac{1}{N} \sum_{\substack{i \in I \\ t_\mathrm{i}+\widehat{k} \leq \max(T_\mathrm{I})}} & \left[ (X(t_\mathrm{i}) - \langle X_\mathrm{I} \rangle) \times \left( X(\widehat{S}(t_\mathrm{i} + \widehat{k})) - \langle X_\mathrm{I} \rangle \right) \right. \\ & \left. \times\, \widehat{W}\left( \left| \widehat{S}\left(t_\mathrm{i} + \widehat{k}\right) - \left(t_\mathrm{i} + \widehat{k}\right) \right| \right) \right].\end{aligned} \tag{4.3}$$

Here $N := \sum_{i \in I} (X_\mathrm{i} - \langle X_\mathrm{I} \rangle)^2$ denotes the total sum of squares and $\langle X_\mathrm{I} \rangle$ is the mean of the time series values set. The general form of the G-ACF is very similar to that of the ACF (Equation 2.10). The G-ACF differs from the ACF by explicitly including the selection function in the second factor, the restriction on the sum, the generalised lag and an additional third factor given by the weight function. I will discuss in more detail the three new components: the generalised lag $\widehat{k}$ and the selection and weight functions.

### 4.3.2 The generalised lag $\widehat{k}$

As shown in Equation 4.3, the generalised lag $\widehat{k} \in [0\,, (\max(T_\mathrm{I}) - \min(T_\mathrm{I}))]$ can now take any value within an interval in time instead of solely integer values based on fixed regular sampling.

Even though the G-ACF is a well-defined function for any lag, it cannot contain meaningful information at a higher resolution than the time series itself. So it would be sensible to set the time-resolution of the generalised lag to values no smaller than the minimal difference between two neighbouring time labels $\delta\widehat{k} \geq \min(t_\mathrm{i} - t_{i+1})$ for $\{t_\mathrm{i}, t_{i+1}\} \in T_\mathrm{I}$. I use this default value when calculating the G-ACF of light curves throughout this thesis.

The condition $t_\mathrm{i} + \widehat{k} \leq \max(T_\mathrm{I})$ on the sum is the generalisation of the upper limit $i_\mathrm{max} - k$ of the sum in the ACF definition (Equation 2.10). The bound on the (generalised) lag enforces again that the maximum shifting of the process along itself is equal to the temporal length of the time series and thus when the first time label is matched up with the last time label.

### 4.3.3 The selection function $\widehat{S}$

The selection function is an integral part of the G-ACF definition (Equation 4.3): it deals with the irregular sampling issue at the core of the motivation behind this generalisation. A selection function $\widehat{S}$ is defined to be a function $\widehat{S} : \mathbb{R}^+ \to T_\mathrm{I}$ that projects an arbitrary point in time onto the set of available time labels, thus selecting a specific time label for each point in time. There are many sensible functions that one could choose to accomplish this; however, a natural selection function is the one that, for each point in time, selects the *closest allowed time label* (see Figure 4.1 for an illustration of this function). Suppose two time labels are equally close to the argument. In that case, one can employ the convention of always choosing the smaller or larger value or randomising the decision in any practical application of the G-ACF.

A possible alternative definition of the selection function would be to find the closest time label for the first shifted time label and then pair up all subsequent labels instead of finding the closest time label for each shifted label individually. While this definition reduces the computational complexity, on testing, it did not produce as accurate a reconstruction of the standard ACF as taking the closest time label for each particular time label when tested on synthetic time series.

### 4.3.4 The weight function $\widehat{W}$

We define a weight function $\widehat{W}$ to be a function $\widehat{W} : [0, \infty) \to [0, 1]$ with $\widehat{W}(0) \equiv 1$. An interpretation of the weight function is as a function that assigns time differences $\delta t \geq 0$ a weight within the interval $[0, 1]$. From Equation 4.3, we see that the weight function is used to assign a weight to the difference between the argument and the value of the selection function

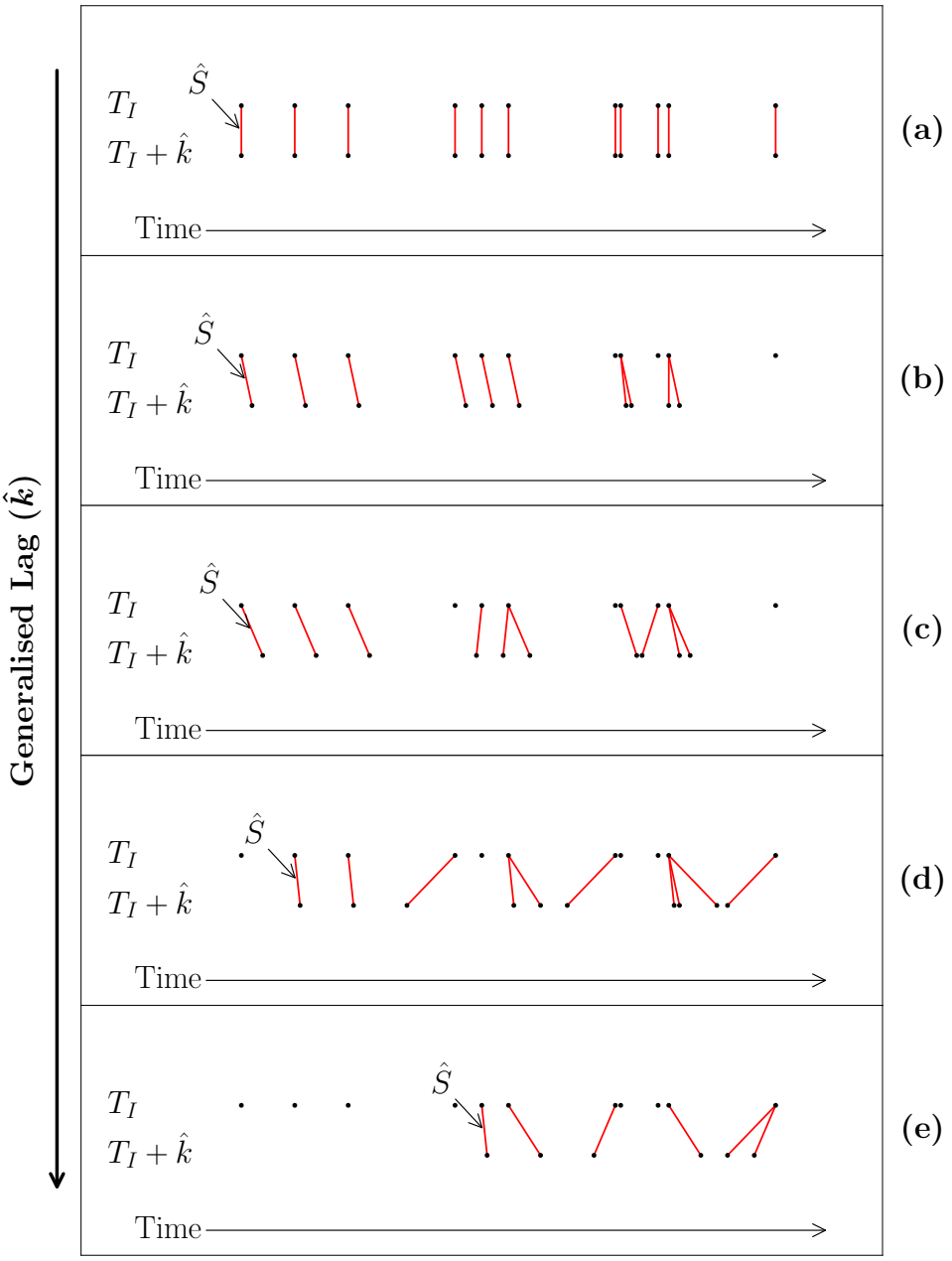



Figure 4.1: Each graphic (a) to (e) shows a set of time labels on the real axis and below them the same set of time labels shifted by a real generalised lag $\widehat{k}$. The red lines indicate how the selection function $\widehat{S}$ matches the shifted time labels to the original set of time labels above by choosing the closest time label from the set of time labels $T_{\mathrm{I}}$. The generalised lag increases from panel (a) to (e), corresponding to the lower labels 'shifting' to the right.

and can be considered to quantify the quality of the 'selection'. Every fixed point of the selection function $\widehat{S}\left(t_\mathrm{i}+\widehat{k}\right)=t_\mathrm{i}+\widehat{k}$, such as in the case of regular sampling, will therefore lead to a term in the G-ACF with weight equal to one because of the requirement that $\widehat{W}(0)\equiv 1$.

There are many choices for possible weight functions. However, the condition $\widehat{W}(0)\equiv 1$ must be observed since this is an important property in order to ensure the G-ACF is identical to the ACF for regular sampling. It would be natural for the weight function to be a monotonically decreasing function tending towards zero. This functional form reflects the interpretation that terms that involve time series values at similar points in time should be preferred. There are infinitely many such functions, including an exponential function or one half of a Gaussian distribution, both of which I tested on synthetic and real data. The effect of the exact shape of the weight function is not crucial to the overall shape of the ACF, provided it fits the above criteria. We propose a rational weight function such as

$$\widehat{W}(\delta t)=\frac{1}{1+\alpha\delta t},\quad \alpha>0,\ \delta t\geq 0 \tag{4.4}$$

where $\alpha$ is the characteristic scale parameter of the time series labels, e.g., one may choose $\alpha=1/\langle T_\mathrm{I}\rangle$. The $\delta t$ represents a generic time difference and should not be interpreted as a sampling constant. In testing, I also considered a half-Gaussian weight function of the form:

$$\widehat{W}(\delta t)=\exp\left(-\frac{\delta t^2}{2\alpha^2}\right),\quad \alpha>0,\ \delta t\geq 0. \tag{4.5}$$

The G-ACF will depend on the scale of the time labels since we are free to re-scale time labels arbitrarily, but the correlation between the different points in time of a process should not depend on the overall time scale. The scale parameter $\alpha$ cancels out any re-scaling of the time labels since it will re-scale inversely. Equation 4.4, the rational weight function, is a simple continuous weight function that fits the above criteria and is also efficient for explicit calculations. I use the rational weight function as the default weight function for the remainder of this work.

A different weight function we considered was a function which satisfies $\widehat{W}(0)=1$ but is zero in all other cases, thus discarding all terms that do not have matching shifted time labels and hence eliminating the selection function from the definition. However, in the case of irregularly sampled time series, this choice of weight function is likely to eliminate the majority of the terms contributing to the G-ACF for a given lag, including almost matching terms, which would not be considered at all. The method proposed in Lukatskaia (1975) assumes that the irregular sampling arises from missing data points in a regularly sampled series and further assumes that the statistical properties of the missing data are the same as the observed data. Therefore, it is possible to calculate a standard autocorrelation using only data points that fall onto this regular sampling, with the caveat that the time series must be much longer than the variability period

of the signal of interest. This method would be best suited for the case of an almost regular sampling where a small percentage of values are missing from an otherwise regularly sampled time series. In this case, most terms in the autocorrelation function will match when the lag corresponds to an integer multiple of the 'regular' sampling constant. Only a few terms without a matching time label would be discarded.

### 4.3.5 Reduction of the G-ACF to the ACF for regularly sampled time series

From the definition of the G-ACF (Equation 4.3), the selection function (Section 4.3.3), and the property $\widehat{W}(0) \equiv 1$ of the weight function, Lars Kreutzer was able to derive a consistency property of the G-ACF for the case of regularly sampled time series. The G-ACF reduces to the ACF for regularly sampled time series when restricting the generalised lag to multiples of the sampling constant. A full proof and detailed explanation of this property are given in Appendix A of Kreutzer et al. *submitted*. This reduction to the ACF is one of the core requirements of the generalisation and ensures that the G-ACF and the ACF are equivalent for regularly sampled time series. This requirement motivated some of the restrictions imposed on the selection and the weight functions.

Appendix B of Kreutzer et al. *submitted* demonstrates that the G-ACF predicts a perfect correlation for a zero time shift. This is a trivial result of the ACF, but it is necessary to demonstrate that if $\rho(0) = 1$ for the ACF, $\widehat{\rho}(0) = 1$ should also hold for the G-ACF.

## 4.4 Implementation and testing

### 4.4.1 Building the code base

The Python implementation used in this work is available open-source under the MIT license on GitHub.[d] It can be installed through `PyPI` using the command: `pip install gacf`. It was developed to allow testing of the G-ACF algorithm on both synthetic and real astrophysical data, with the requirement that it calculates the G-ACF of a time series both accurately and quickly.

The core algorithm, including the selection and weight functions, is written in `C++` as this provides a considerable speedup over the more commonly used `Python` through more precise memory management and lower overheads at runtime as it is a compiled, not interpreted language. In order to provide an easy to use package, the `C++` code has a `Python` wrapper, which is called similarly to the `astropy` Lomb–Scargle implementation (Robitaille et al. 2013).

[d] `www.github.com/joshbriegal/gacf`

The most computationally expensive aspect of the G-ACF is the selection function: for each lag time step, the closest point to each point in the lagged time series must be selected. Naïvely, this is at worst an $O(n^2)$ operation as each of the $n$ points in the time series will require an $O(n)$ lookup to be performed. As the lag increases, the computation time will decrease as fewer selection functions are evaluated per lag timestep. An improvement that was not implemented, but will result in fewer lookups, was to store the index of the previous closest time label and begin the search at this time label rather than searching the whole time series. The 'natural selection function' and the 'fast selection function' described in Section 4.3.3 were implemented. Although the fast selection function was considerably faster to evaluate, the G-ACF produced deviated from the standard ACF at increasing lags and began to deviate at small lags in the case of irregular sampling. I implemented the half-Gaussian and the rational weight functions (Equations 4.4 and 4.5); the default option is the rational weight function. The G-ACF can be calculated for a multi-dimensional array of values that share a common time series to improve efficiency further. This optimisation will speed up the calculation of such data arrays: the selection function needs to be evaluated on just one set of time labels rather than for each set of values. As the selection function is the most costly part of the algorithm to evaluate, we can achieve a speed-up of approximately $N$ times when considering $N$ light curves sampled on the same time series.

### 4.4.2 Simple examples

The simplest example is the process defined by a sine function. I generated a regularly sampled time series, a randomly sampled time series, and a structured but irregular time series. This last time series consisted of clusters of time labels with a fixed periodicity but larger gaps in between, representing a typical ground-based survey cadence for astronomical applications (Figure 4.2). It is important to note that all three time series possessed the same number of time labels ($|T_\mathrm{I}| = 250$) and differ only in the temporal distribution of the time labels.

As the time series differ only by their sampling, we can see how the distribution of time labels influences the G-ACF. From Figure 4.2, we see that the G-ACF of the regularly sampled time series (black) is identical to the ACF (by design), but the function is continuous due to the definition of the G-ACF. In Figure 4.2, I evaluate the continuous G-ACF function at a finite number of points and plot this as a line. The G-ACF of the random and cadence-like sampling are similar to the ACF but with small differences driven by data gaps. The differences depend on the exact position and size of gaps within the data. Increasing the sampling density will improve the accuracy of the G-ACF; however, if large gaps such as the cadence-based sampling gaps remain, there will be deviations from the regular ACF.

In Figure 4.3, I plot the sum of two sine functions, applying the three time-label sets

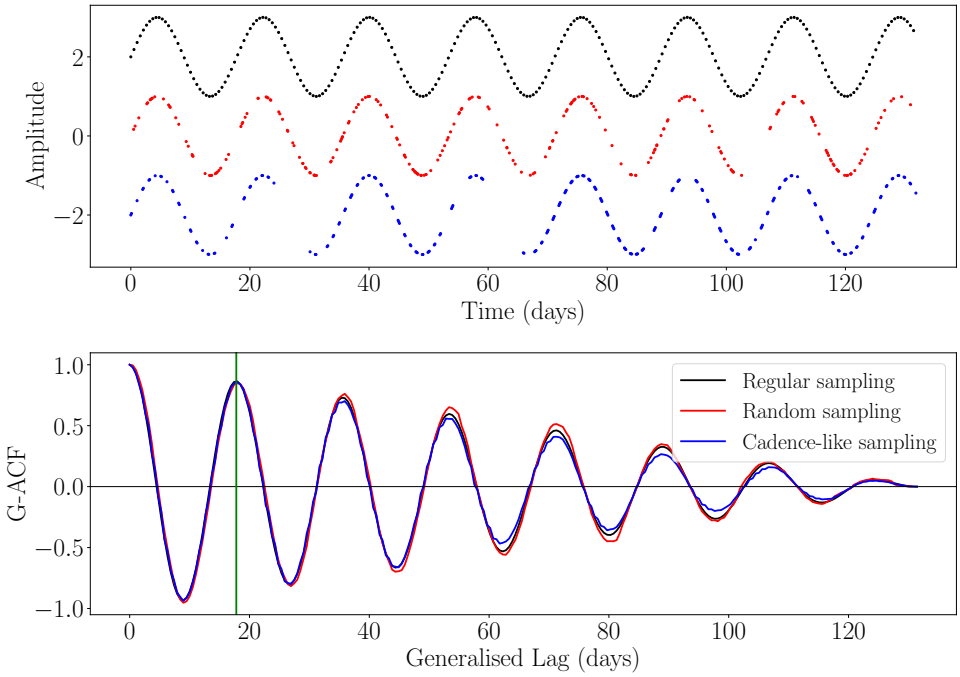



Figure 4.2: The top panel shows three time-series with an underlying sine process (17.8 day period), sampled regularly (black), randomly (red) and with a cadence-like sampling that possesses additional larger gaps (blue). All time-label sets have the same number of points (250). The bottom panel shows the generalised autocorrelation functions (G-ACF) of the above time series. A vertical green line is plotted at the period of the signal (17.8 days) in generalised lag.

described above and calculating their G-ACF. As for the single sine function example, the random (red) and cadence-like (blue) sampling cases display modest deviations from the regular ACF.

In the case of cadence-like sampling, with a (mostly) fixed periodicity of sampling gaps, alias signals can appear in the G-ACF due to the missing information in between well-sampled clusters of time labels. These aliases can be easily identified since their periodicity will be equal to the periodicity of the clusters of time labels. Their amplitude will be proportional to the relative size of the gaps between clusters. This effect will not be relevant for most applications unless one looks at the special case in which the structures of interest in the time series are of a comparable period to the structure of the sampling clusters. We suspect it may be possible to reduce this effect by generalising the normalisation to a lag and time label density-dependent function. However, full removal of these aliases will likely not be possible since gaps imply missing information that cannot be restored without additional information or assumptions. I note, however, that these effects are small if there are sufficient time labels per period of the

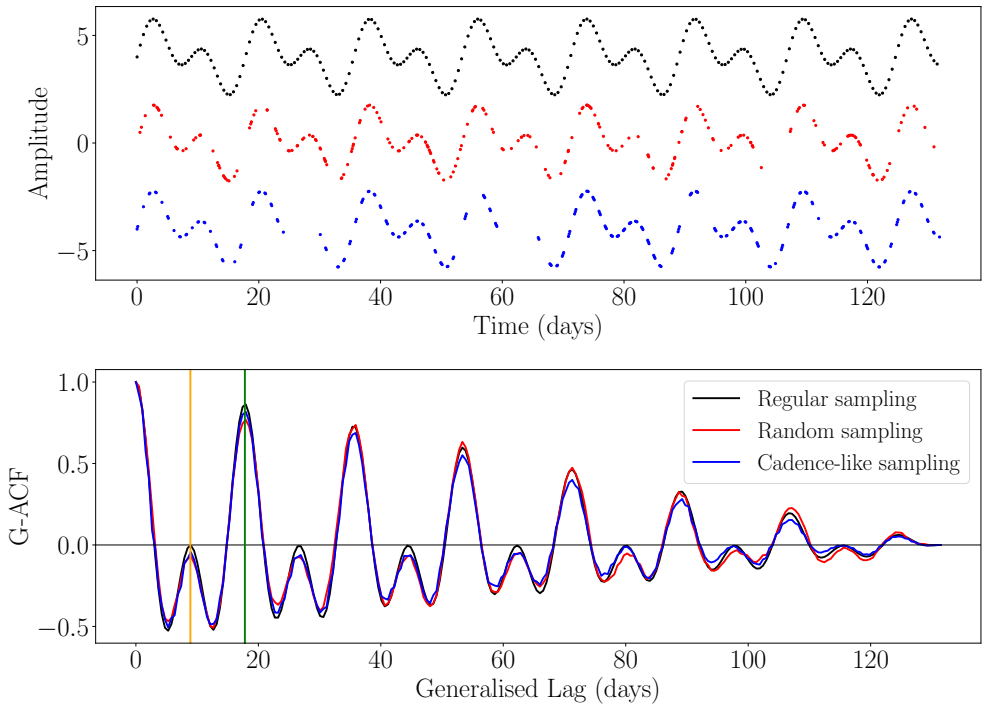



Figure 4.3: As Figure 4.2 but for a time series with an underlying process described by the sum of two sine functions with 8.9 and 17.8 day periods. Vertical orange and green lines are plotted at the periods of the signal (8.9 and 17.8 days, respectively) in generalised lag.

process.

The examples considered above both feature sampling such that the reconstruction of a signal is fairly accurate as the period of the signal is much greater than the sampling cadence. In cases where the signal is much less well sampled, the accuracy of any reconstruction (ACF or G-ACF) will be reduced. I chose the periods in Figures 4.2 and 4.3 as clear examples of the similarity between the G-ACF and the ACF. The periods are long enough to be sampled well and not close to multiples of 1 day, which drastically reduces the accuracy of the G-ACF on cadence-based sampled data.

In order to investigate the efficacy of the G-ACF on more realistic data sets, I generated a periodic signal with a large stochastic noise component. The periodic signal was again a sine function with a 17.8-day period. The stochastic component was drawn from a Gaussian process (GP) using a simple harmonic oscillator (SHO) kernel (with quality factor $Q = 1/3$ and characteristic timescale $\rho = 5$ days), as implemented in the `celerite2` Python package (Foreman-Mackey 2018). The amplitudes of the sinusoidal and stochastic components were comparable.

The same three temporal samplings were used as in the simpler cases, and the resulting time series and corresponding G-ACFs are shown in Figure 4.4. The G-ACF displays a prominent

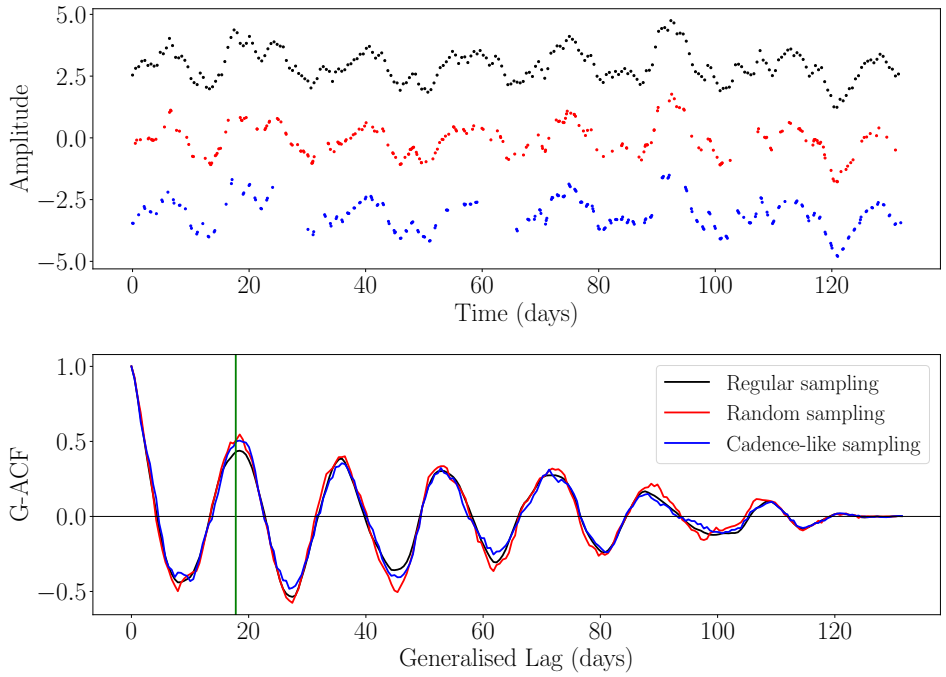



Figure 4.4: As Figure 4.2 but for a time series with an underlying process described by the sum of a comparable amplitude sine function and stochastic Gaussian process. A vertical green line is plotted at the period of the deterministic component of the signal (17.8 days) in generalised lag.

peak corresponding to the period of the sinusoidal component. However, the exact position of this peak will be moderately affected by the large noise component. The G-ACF can accurately recover a clear periodic signal in all three sampling cases despite the periodic and noise components having comparable amplitudes.

### 4.4.3 Application to real data: the Kepler light curve of KIC 5110407

The following section of work was conducted in collaboration with Ed Gillen. We tested the efficacy of the G-ACF on real time-series data. While the applications of the G-ACF are not restricted to astronomy, the standard ACF has been widely used to estimate the rotation periods of stars from time-series photometry. Therefore, as an illustrative example, we selected a spotted star observed by Kepler KIC 5110407 (e.g., Roettenbacher et al. 2013), and compare the period predictions of G-ACF to two other techniques for rotation period estimation: Gaussian process (GP) regression and Lomb–Scargle (LS) periodogram. Our approach to comparing these three models follows Gillen et al. (2020) and Section 3 of that paper contains full details, but I will give a brief overview below of the GP and LS models used here.

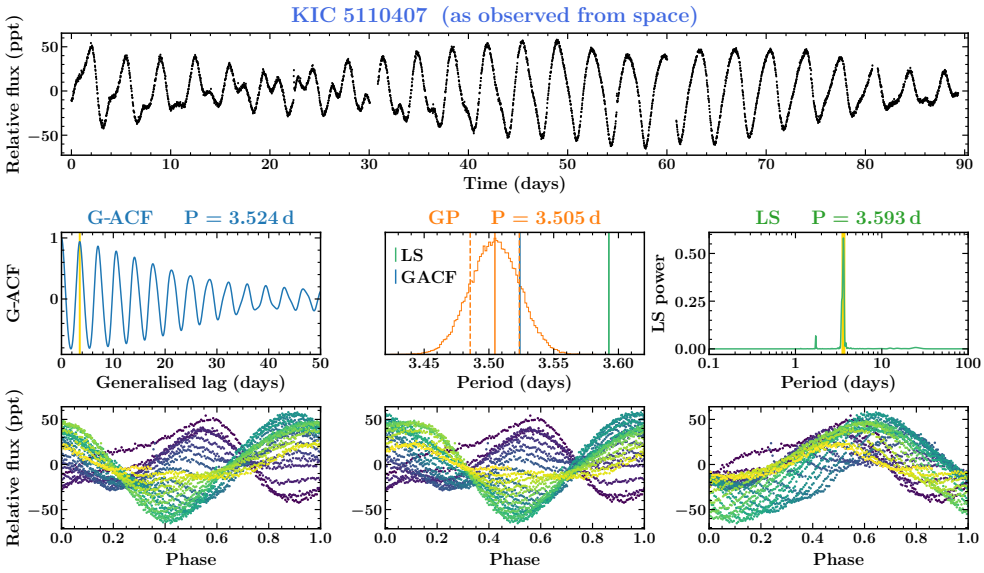



Figure 4.5: Rotation period estimates for the spotted star KIC 5110407 from G-ACF, Gaussian process (GP) regression, and Lomb–Scargle (LS) periodogram. *Top panel*: the system's quarter 7 Kepler light curve. *Middle left*: Generalised autocorrelation function (blue) with the identified period highlighted (yellow). *Middle centre*: GP posterior period distribution (orange) with the median and 1 $\sigma$ uncertainties highlighted (solid and dashed orange lines). For comparison, the G-ACF and LS periods are also shown (blue and green solid lines, respectively). *Middle right*: LS periodogram (green) with the identified period highlighted (yellow). *Bottom row*: The Kepler light curve phase-folded on the corresponding method's period (G-ACF, GP and LS; left-to-right) and coloured from the beginning (blue) to the end (yellow) of the observations. Credit: Ed Gillen (Kreutzer et al. *submitted*).

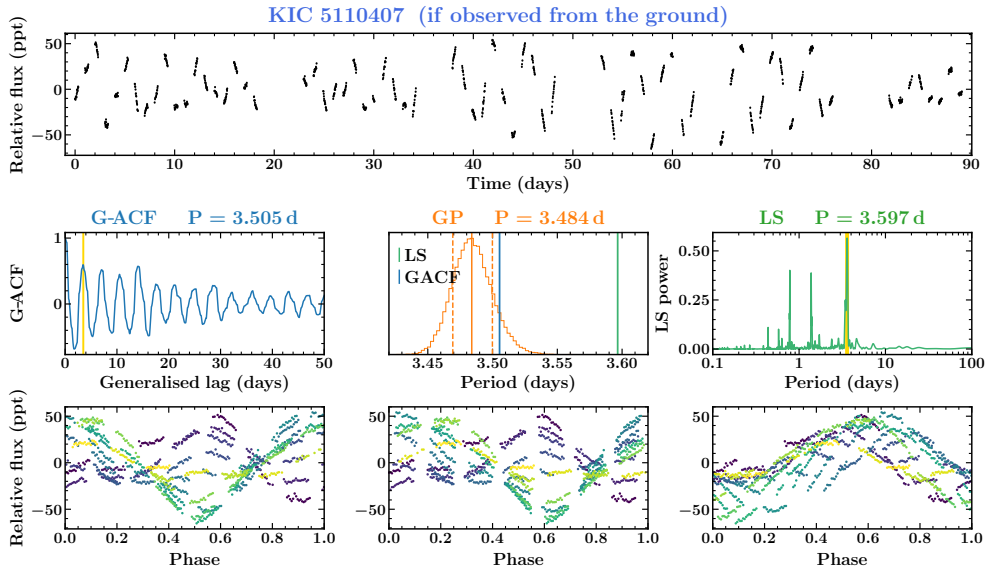



Figure 4.6: Same as Figure 4.5 but simulating KIC 5110407 being observed from the ground (i.e. observations during night time only with additional gaps from bad weather). Credit: Ed Gillen (Kreutzer et al. *submitted*).

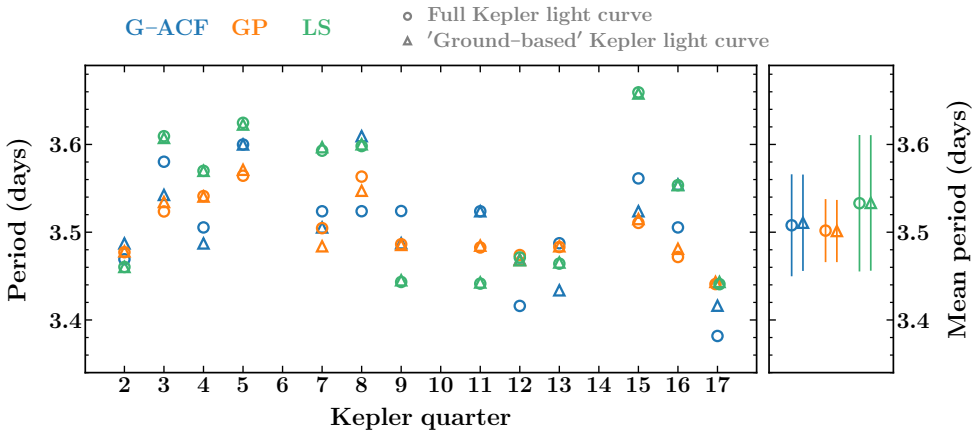



Figure 4.7: Rotation period estimates for KIC 5110407 from G-ACF (blue), GP (orange) and LS (green) for all quarters with Kepler data. Circles show the period estimates from the full Kepler light curve and triangles from the 'ground-based' version of the light curve. The right-hand panel shows the mean and standard deviation of the period estimates across all quarters. The mean G-ACF periods agree with both the GP and LS periods, as do the values from the full and 'ground-based' versions of the light curves. The scatter in the G-ACF periods across quarters is smaller than the scatter in LS periods but slightly larger than the scatter in GP periods. Credit: Ed Gillen (Kreutzer et al. *submitted*).

The GP model is based on the `celerite2` package (Foreman-Mackey et al. 2017; Foreman-Mackey 2018), as implemented through the `exoplanet` framework (Foreman-Mackey et al. 2021; Foreman-Mackey & Barentsen 2019), and uses the standard rotation kernel with an additional simple harmonic oscillator (SHO) kernel (with quality factor $Q = 1/3$) to capture any non-periodic structure in the light curves. The posterior parameter space was explored via gradient-based Markov-chain Monte Carlo (MCMC) using the No U-Turn Sampler (NUTS), as available through `exoplanet`, which in turn uses `PyMC3` and `theano` (Hoffman & Gelman 2014; Kumar et al. 2019; Salvatier et al. 2016; The Theano Development Team et al. 2016).

For each quarter, we ran five independent chains of 5,000 tuning steps followed by 10,000 sampling steps. It is worth noting that the GP model requires an initial period guess, in contrast to both G-ACF and LS, for which we give the average of the G-ACF and LS period estimates. The GP model is also sensitive to data not well captured by the chosen rotation kernel, such as stellar flares, which we account for by performing an initial maximum a posteriori fit, masking $3\,\sigma$ outliers, and refitting.

For the LS model, we use the version available through the `astropy` project (Robitaille et al. 2013; Price-Whelan et al. 2018). This implementation uses the formalism defined in Zechmeister & Kürster (2009), which introduced heteroscedastic weighting and zero-point estimation to the algorithm first proposed by Lomb (1976) and Scargle (1982). The LS and

G-ACF models were run on the data without further processing, such as flare masking. GP periods are estimated from the period posterior distributions, LS periods are estimated from the largest peak in the periodogram, and G-ACF periods are estimated by calculating a Fast Fourier Transform (FFT) of the first three peaks of the G-ACF and taking the largest peak in the periodogram. Restricting the lag time used in the period estimation reduces the effect of signal shape evolution on the autocorrelation function at long lag times. It correspondingly improves the accuracy of the period estimated. Using an FFT to calculate the periodicity of the G-ACF is possible as the G-ACF is a continuous function by definition. I note that another method of extracting periodicity from the G-ACF would be to calculate the position of the first peak in the G-ACF and then use the positions of subsequent peaks to refine this period estimate, such as the technique used in McQuillan et al. (2013).

Kepler observed KIC 5110407 for almost four years spanning 13 of the 17 quarters. Kepler quarters typically last ~90 days and have essentially continuous observations with a cadence of ~30 mins. The ACF has been successfully applied to such Kepler data (e.g., McQuillan et al. 2013, 2014) but, as noted, the ACF does not apply to non-continuous data that cannot be accurately interpolated onto a regularly spaced time series grid, i.e. time series with large data gaps, such as ground-based photometry. We, therefore, estimated the stellar rotation period of KIC 5110407 from two versions of its Kepler light curve: (i) the full Kepler light curve and (ii) the Kepler light curve as though it had been observed from the ground (i.e. with gaps during daytime and simulated 'bad weather' events[e]).

Figure 4.5 shows the results for the full Kepler light curve observed during quarter 7, and Figure 4.6 shows the results for the 'ground-based' version of the light curve. The Kepler data from this quarter shows moderate evolution throughout the light curve and displays both 'double-dip' patterns (e.g., at ~20 days) and sinusoidal modulation (e.g., at ~40–80 days). Therefore, the G-ACF and GP periods agree best for this quarter, whereas the LS period prediction is slightly larger. This is the case for both the full and 'ground-based' light curves. The better agreement between G-ACF and GP is because they are more flexible than LS (i.e. they do not assume a rigid sinusoidal model) and are more applicable to such evolving time series. The periods can be best compared in the middle centre panel of Figures 4.5 and 4.6 and by comparing the phase-folded light curves.

We performed the same analysis on each available quarter of Kepler data: Figure 4.7 compares the period predictions for G-ACF, GP and LS across quarters. Across quarters, and for both the full and 'ground-based' light curves, the G-ACF and GP periods agree best overall. The LS predictions agree well for some quarters, mainly those that show sinusoidal modulation,

[e] All quarters had the same relative times masked. Nighttime was considered to last 8 hours of each 24 hours, and bad weather was simulated between the following times: 18.5–22.5, 34.5–37.5, 48.5–52.5, 62.5–64.5 and 76.5–81.5 days (relative to the start of each quarter).

but less well for those that show evolving modulation patterns, resulting in a larger scatter and correspondingly larger uncertainties on the mean rotation period prediction than the G-ACF or GP. The mean periods and standard deviations across quarters are: G-ACF = 3.51 ± 0.06 and 3.51 ± 0.06 days for the full and 'ground-based' light curve, respectively; GP = 3.50 ± 0.04 and 3.50 ± 0.04 days; and LS = 3.53 ± 0.08 and 3.53 ± 0.08 days. Roettenbacher et al. (2013) estimate a rotation period for KIC 5110407 through light-curve inversion of 3.4693 days, which agrees to within $1\sigma$ for all three methods.

This comparison between the G-ACF and the GP and LS methods, for both continuous and irregularly sampled time series, illustrates the validity of the G-ACF for such applications. Furthermore, as the G-ACF is 'model-free' it can be applied to time series data of essentially any form without the need to adapt the kind of model chosen (in the case of GP) or assume a rigid sinusoidal model (in the case of LS). Additionally, the G-ACF is efficient to calculate: for example, calculating the G-ACF of the quarter 7 KIC 5110407 light curve took approximately 0.6 seconds with 4,117 data points on a single laptop core[f]. The GP regression was ∼14 times slower, taking ∼8.3 seconds for the maximum a posteriori fit (and ∼7 minutes for the MCMC). The LS periodogram was the fastest, taking approximately 0.01 seconds to run on the same laptop core, however the simple sinusoidal model employed may not well-model complex time series. The G-ACF is a powerful and efficient approach to extracting periodicity, quasi-periodicity and short-term self-similarity from time series data in general, and especially data for which the true functional form is unknown.

## 4.5 Application to NGTS Data

The first published application of the G-ACF to NGTS data was the rotation study of the Blanco 1 open cluster by Gillen et al. (2020), as discussed in Section 8.1. The remaining Chapters of work in this thesis discuss the wider application of the G-ACF to NGTS data in the form of a large-scale rotation study and additional open cluster studies.

The methods used in this study for assessing the accuracy of the G-ACF in recovering rotation periods are taken into account in the development of the `RoTo` package, which will be discussed in Chapter 6, in which multiple period extraction methods are made available for use on time series data in `Python`.

[f]The run time of the G-ACF is dependent on both the number of data points and the number of lag time steps.

## 4.6 Outlook

### 4.6.1 Accounting for measurement uncertainties

The definition of the G-ACF can be further extended to take into account the measurement uncertainties of the time series values. One approach would be to define an extended weight function that depends on the measurement uncertainties of both of the time series values in each product, such as

$$\widehat{W}(\delta t,\, \sigma_{X_\mathrm{i}}^n,\, \sigma_{X_\mathrm{j}}^n) = \frac{1}{1+\alpha\,\delta t} \cdot \frac{1}{1+\beta\,\sigma_{X_\mathrm{i}}^n} \cdot \frac{1}{1+\beta\,\sigma_{X_\mathrm{j}}^n} \tag{4.6}$$

where $\sigma_{X_\mathrm{i}}$ and $\sigma_{X_\mathrm{j}}$ represent the measurement uncertainties on the $i^{\mathrm{th}}$ and $j^{\mathrm{th}}$ time series values, respectively, and where $\delta t,\, \sigma_{X_\mathrm{i}},\, \sigma_{X_\mathrm{j}} \geq 0$ and $\alpha = 1/\langle T_\mathrm{I} \rangle$ as previously. We could define $\beta = 1/\langle X_\mathrm{I}^n \rangle$. The $n^{\mathrm{th}}$ power could take a value of either 1 or 2 to weight by the inverse of the measurement uncertainties or their variance, respectively.

This weight function does not satisfy the condition $\widehat{W}(\delta t = 0) = 1$ if the uncertainties are non-negligible, and thus the G-ACF with this extended weight function does not reduce to the ACF in the case of regular sampling. Instead, it reduces to a different generalisation of the ACF, which still weighs each product according to the measurement uncertainties. This result should be expected since the original definition of the ACF does not account for uncertainties in the time series. If the uncertainties of the time series values are equal, this method results in an overall re-scaling of the original G-ACF. Equation 4.6 does satisfy the condition $\widehat{W}(\delta t = 0,\, \sigma_{X_\mathrm{i}}^n = 0,\, \sigma_{X_\mathrm{j}}^n = 0) = 1$ and thus this generalised and extended ACF reduces to the ACF in the case of regular sampling if the uncertainty $\sigma_{X_\mathrm{i}}$ for each time series value is negligible and thus can be discarded. The viability of such an extension remains to be investigated.

## 4.7 Conclusions

The G-ACF, or generalised autocorrelation function, is a new and versatile definition that can reliably and efficiently extract, amongst others, periodicity and signal shape information from any time series, virtually independent of the time series sampling and independent of the underlying process. We show that the ACF can be generalised and applied to irregularly sampled time series by generalising the lag to a real variable and introducing selection and weight functions. We show that the G-ACF reduces to the ACF for regularly sampled time series and possesses the property of maximal correlation at zero lag.

I show that the G-ACF agrees well with the ACF for cases of aperiodic sampling, including the case of randomly sampled time labels and cadence-like sampling; however, there are slight deviations due to the data gaps and corresponding loss of information.

We compare the period predictions of G-ACF to those from GP regression and LS periodograms by extracting rotation periods for the spotted star, KIC 5110407. The G-ACF and GP periods typically agree best across the different Kepler quarters. LS periods are comparable in quarters with mainly sinusoidal modulation but more discrepant for quarters displaying more complex or evolving patterns. All three methods achieve consistent mean periods and uncertainties.

There are many potential applications for the G-ACF within astronomy and astrophysics and in other quantitative sciences where irregularly sampled time series occur, such as economics, climatology, geology, biology, and others.

I built and tested an implementation of the G-ACF algorithm in `C++` and `Python`, which has since been successfully applied to real astrophysical data beyond KIC 5110407, in the published work of Gillen et al. (2020), Briegal et al. (2022), and the remaining chapters of this work.

CHAPTER 5

# PERIODIC STELLAR VARIABILITY FROM ALMOST A MILLION NGTS LIGHT CURVES

The following chapter is based on the paper *Periodic stellar variability from almost a million NGTS light curves.* (Briegal et al. 2022), which was accepted for publication in MNRAS on 29th March 2022. I will outline the work completed for this study with a detailed explanation of the methods used and the results obtained. I will discuss how these results fit into a wider scientific context. This work draws on the stellar variability background outlined in Chapter 2 Section 2.1 and utilises the G-ACF method detailed in Chapter 4. I use data from NGTS as in Chapter 3. Almost the entirety of this work was completed by me; where other authors are responsible this will be made clear in the text.

I analyse 829,481 stars from the Next Generation Transit Survey (NGTS) to extract variability periods. I utilise a generalisation of the autocorrelation function (the G-ACF), which applies to irregularly sampled time series data. I extract variability periods for 16,880 stars from late-A through to mid-M spectral types and periods between ~0.1 and 130 days with no assumed variability model. I 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 me to identify distinct regions of variability in the Hertzsprung-Russell Diagram. I explore a sample of rotational main-sequence objects in period–colour space, in which we can observe a dearth of rotation periods between 15 and 25 days. This 'bi-modality' was previously only seen in space-based data from Kepler and K2 (McQuillan et al. 2014; Gordon et al. 2021). I

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. I 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.

As we have seen in Chapters 1 and 2, many of a star's physical properties can be inferred from its brightness variations over time. This variability can arise from many 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. 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 hypothesised that a star's rotation rate could be age dependent, obtaining the empirical relation between rotation period $P_{\rm rot}$ and age $t$: $P_{\rm rot} \propto t^{0.5}$. 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 magnetised stellar winds which depend on magnetic field topology and stellar mass (Booth et al. 2017). For young stars ($< 10$ Myr) angular momentum evolution may be dependent on magnetic coupling between the star and disc. Studies of pre-main-sequence stars in young clusters such as T-Tauri stars in the Taurus-Auriga molecular cloud (Hartmann & Stauffer 1989) or NGC 2264 (Sousa et al. 2016) show high levels of short period ($< 10$ day) 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 et al. 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 magnetised 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 (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. (2020) 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[a] 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

[a] Accessed on 09/11/2021

precision and excellent 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 (Ciardi et al. 2011; Basri et al. 2010; Affer et al. 2012; McQuillan et al. 2013; 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. (2013) and Davenport & Covey (2018) demonstrated a distinct bi-modal structure in the rotation periods of main-sequence stars with respect to colour. Gordon et al. (2021) followed up these studies with an analysis of data from the K2 mission, hypothesising the bi-modal 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 5.3.

NGTS routinely achieves milli-magnitude range photometric precision with 12-second sampling cadence and long observation baselines (typically 250 nights of data per target field). Such high-precision photometry lends itself well to ancillary stellar physics such as cluster rotation analysis (Gillen et al. 2020) or stellar-flare detection and characterisation (Jackman et al. 2019a). 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 particular, for this study, I employ a generalisation of the autocorrelation function (the G-ACF) which applies to this irregular sampling. I elected to use an autocorrelation function to extract variability as this has proven to be successful for extracting stellar variability by McQuillan et al. (2013, 2014) & Angus et al. (2018) and for NGTS data in Gillen et al. (2020). An Autocorrelation Function (ACF) also allows better detection of pseudo-periodic and phase-shifting variability often seen in young, active stars in comparison to a more rigid variability extraction technique such as Lomb–Scargle periodograms.

## 5.1 Methods

### 5.1.1 Data pre-processing

The NGTS pipeline as described in Chapter 3 provides flags per image and per timestamp per object light curve which I used 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. I removed any flagged data points from each light curve, and additionally checked if the majority of the light curve has been flagged ($> 80\%$ of data points). If this was the case, I removed the objects from processing entirely.

I clipped the flux data to remove any points lying further than 3 median-absolute-deviations

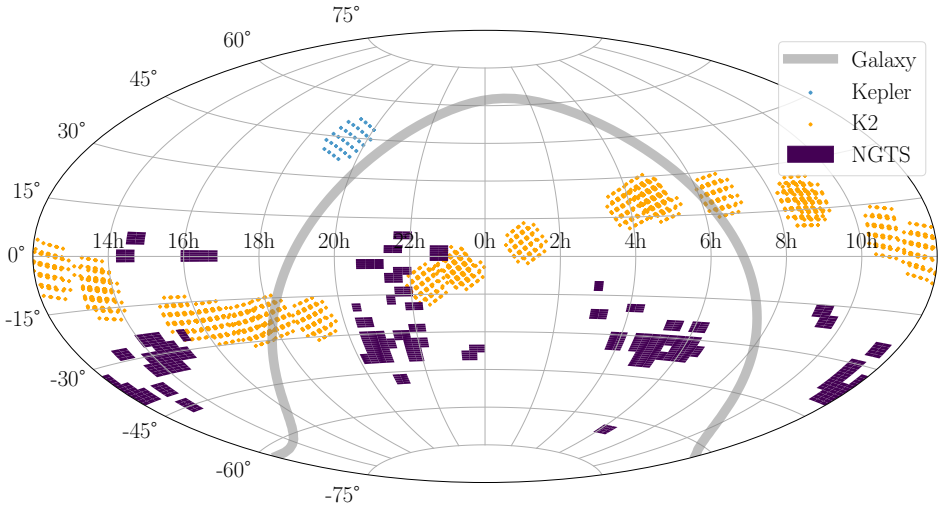



Figure 5.1: An ICRS plot of the position of the 94 NGTS fields used in this study (solid dark blue squares). The Kepler and K2 fields are included as blue and orange squares respectively, as well as the Galactic plane as a thick grey line.

(MAD) from the median to remove any outliers not caught by the NGTS pipeline flags. I note this cut may remove some variability signals such as long-period eclipsing binaries where the variability is a small fraction of the phase curve. Manual inspection of a single field confirmed that this was not the case, however, this cannot be guaranteed for all fields processed automatically. Finally, to speed up data processing, I binned each light curve into 20-minute time bins. This reduces the number of data points to process per light curve from 200,000 to roughly 10,000. The G-ACF computation time scales as $O(n^2 m)$ for $n$ data points with $m$ lag time steps, so reducing the number of timestamps in the light curve significantly improves processing time. This comes with a caveat that the pipeline will be unable to detect any periods below 40 minutes, however for this study that is focused on longer period variability this limit is not of concern.

I removed 6 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 open clusters. Removing these 6 fields reduced the number of light curves by 41,831, which 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 Figure 5.1. In this Figure, I 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 days) between September 2015 and November 2018. The shortest observational baseline for this data set was 84 days and the longest 272 days.

73 of the 94 fields had observational baselines over 200 days. I detect periodic variability in light curves spanning $8 < I_{\mathrm{NGTS}} < 16$ mag with 50% (90%) of detections being brighter than 13.5 (15.4) mag.

### 5.1.2 Use of the Cambridge HPC cluster

Despite the computation considerations explained in the previous section, there were a large number of light curves each with around 10,000 data points to process. The final manifestation of the period detection pipeline (as detailed in the following Sections) contains many processing steps per light curve. The light curves for all objects within a field are stored as FITS files (each field between 5 and 20 GB, totalling around 1.8 TB). The processing pipeline was written with this in mind, each field is processed separately and each light curve can be processed in parallel within the field. Processing of a single NGTS field took between 10 and 30 minutes depending on the number of objects in the field, which was processed on a single node with 32 CPUs allowing 32 objects to be processed simultaneously.

### 5.1.3 Period detection

The period detection pipeline is outlined in the flowchart in Figure 5.2. Further details of each step are given in the subsequent sections.

#### 5.1.3.1 G-ACF

The G-ACF algorithm has already been described in detail in Chapter 4. In this work, I used the `Python` package of the G-ACF algorithm as a part of the period finding pipeline. The G-ACF was calculated using the 'natural' selection function and the rational weight function as given in Equation 4.4. I calculated just the positive side of the ACF (i.e. positive lag values only), with a lag resolution of 20 minutes that corresponds to the minimum gap between time points as I have binned the data before processing.

#### 5.1.3.2 FFT

To extract a period from the G-ACF I elected to use a Fast Fourier Transform (FFT; Cooley & Tukey 1965). Extracting periods from an ACF can be done in several 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. I 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).

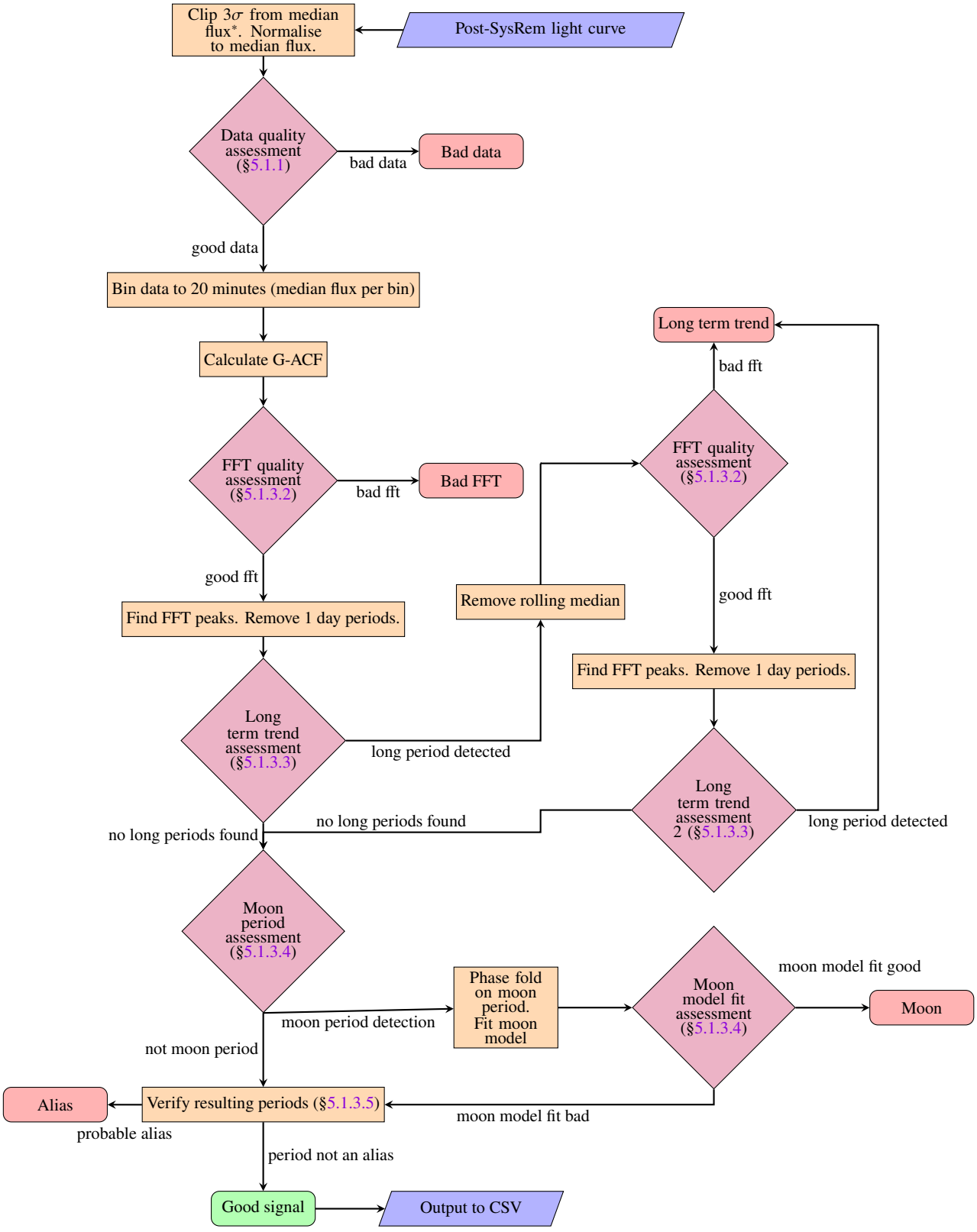



Figure 5.2: A schematic of the period detection pipeline, per NGTS light curve. $^*\sigma$ refers to 3 median absolute deviations (MAD) from the median.

This in general did not alter extracted periods enough to warrant the additional complexity for such an exploratory work. I also experimented with using fewer ACF peaks rather than the entire signal 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, I used the implementation in the `numpy.fft` package (Harris et al. 2020). I 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 I searched for peaks in the Fourier transform. I define a peak as the central point in a contiguous sequence of 5 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% of the highest peak in the periodogram to be included. Here an automated cut was made: any Fourier transforms with more than 6 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% had genuine periodic signals. Removing these objects greatly reduces the number of false positives extracted without removing many 'real' signals. 63% of processed objects were flagged as having no significant periodicity based on this FFT check.

#### 5.1.3.3 Long-term trend assessment

A time baseline of $\sim$ 250 days allows for robust extraction of periodic signals up to $\sim$ 125 days long. Signals longer than this may be present in the data, however, observing one or fewer complete variability cycles cannot definitively characterise 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 timescales. These objects may still contain interesting periodic variability at a shorter timescale, so by detecting and removing a long-term trend it is possible to more accurately calculate the period and amplitude of this variability.

If the most significant peak in the FFT (see Section 5.1.3.2) was at a period greater than half the length of the signal baseline it was flagged as a long-term trend. When this case occurred I computed a high-pass filter for the signal by calculating the median flux at each time step in a rolling window which is 10% of the time extent of the light curve. This will capture any long-term behaviour without removing any shorter period variability. I divided this median filter from the 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 I found noise or residuals of the median filter), the object was flagged as having a long-term trend and removed from processing.

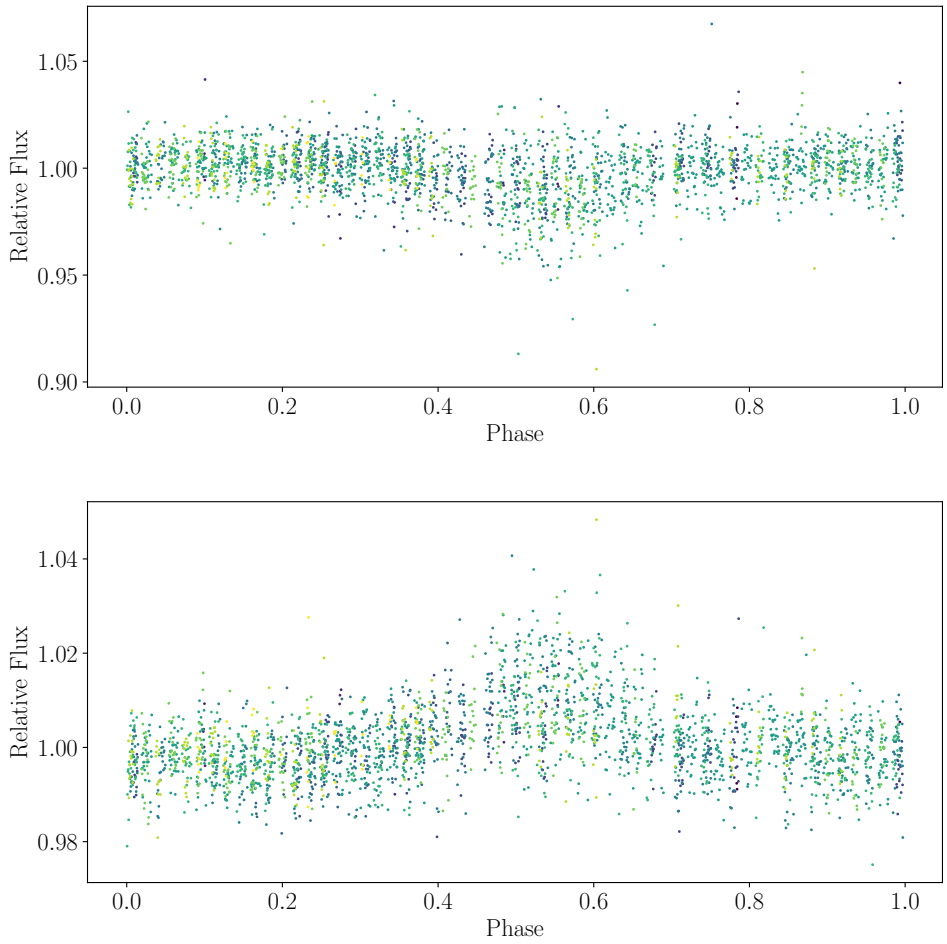



Figure 5.3: Two examples of typical Moon tainted signals. For each object, the light curve is phase folded on the expected Moon period and epoch. 0.0 & 1.0 phase are at new Moon, 0.5 phase is at full Moon. We see an example of an over-corrected signal with a typical decrease in flux at full Moon. An under-corrected signal demonstrates the opposite trend. Both signals exhibit an increase in scatter at full Moon, with an otherwise fairly flat light curve.

#### 5.1.3.4 Moon signal assessment

During the initial testing of the period extraction algorithm, I noted that a large number of periods between 27 and 30 days 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 Figure 5.3.

I fitted a model to these Moon correlated noise signals ('Moon signals') and flagged and

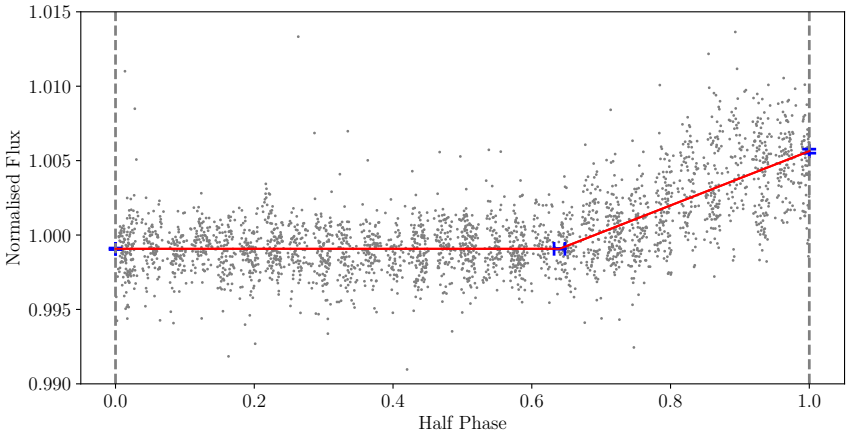



Figure 5.4: The three-parameter Moon model fit is used to assess if a signal is contaminated by the Moon. The flux data is phase folded on the period of the Moon and then again in half such that 0.0 in phase corresponds to new Moon and 1.0 in phase corresponds to full Moon.

removed any objects which fit the expected trend. To systematically detect Moon contaminated signals (for example as shown in Figure 5.3), I fitted a model to the flux data, phase folded on the expected Moon period for each NGTS field. The expected Moon period was 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 days, close to the 29.5-day synodic period of the Moon.

The model is a simple three-parameter piece-wise model described in Equation 5.1, where the parameter $x$ is the location in half phase $x \in [0, 1]$.

$$\begin{cases} \text{flux}_0 & 0 \leq x \leq \text{turnover} \\ mx + c & \text{turnover} < x \leq 1 \end{cases} \tag{5.1}$$

Where

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

I fitted for the 3 parameters $\text{flux}_0$, $\text{flux}_1$ and turnover. This model fit was assessed by checking the following criteria, with an example shown in Figure 5.4.

- Is the model turnover point at the expected point in phase? (between 0.2 and 0.8 in half-phase).

- Is there a flux RMS increase after the model turnover point?
- Is there a noticeable (i.e. $> 1\sigma$) change in flux from new to full Moon?
- Is there any missing data at full Moon?

If 3 or more of these criteria were met, the object was flagged as Moon contaminated and removed from the processing. The decision to remove these signals from processing rather than attempting to remove the Moon signal and re-process was made after re-running a single field with the Moon model subtracted, for which no new obvious periodic signals were found. This may in part have been due to the simplicity of this model, as the Moon signals visible in Figure 5.3 are not entirely modelled by the three-parameter solution. Given the scale of the data processing, I decided to remove the object from processing rather than attempting a more complex model fit which would require significantly more computation time for each object. One such model would be a Savitzky–Golay (SG) filter followed by a convolution that more accurately captures the Moon signal shape, as was applied to the Moon-affected light curves in Gillen et al. (2020).

#### 5.1.3.5 Alias checks

Using an FFT to extract periodicity from the G-ACF will be 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 using Equation 2.3. Note I 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 day. 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. Thus one can expect peaks in the FFT associated with 1-day systematic signals, as well as the true signal aliased with the 1-day sampling.

For each light curve, I first removed any periods arising from the 1-day sampling. I removed periods within 5% of 1 day, as well as within 5% of integer multiples of 1 day in period and integer multiples of 1 / day in frequency. I then assessed whether groups of periods were aliases of one another with respect to common sampling periods using basic graph theory. I constructed a graph of frequencies connected by the standard alias formula in Equation 2.3, using sampling periods of one day, 365.25636 days (one year), 27.32158 days (Lunar sidereal period) and 29.53049 days (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 2.3 given one of the sampling frequencies listed. Note that I considered aliases arising from both the synodic and sidereal Lunar period, however, given the 5% 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) I determined the frequency for which the phase folded light curve had the lowest spread in flux and took this to be the correct period. I calculated the $5^{\text{th}} - 95^{\text{th}}$ 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, I also checked the RMS of twice and half the periods, as in some cases I 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-day sampling. This same approach was taken by McQuillan et al. (2013), however, I elected to automate the process rather than by-eye confirmation of half- or double-period detections.

#### 5.1.3.6 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 are 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, I elected to find the longest period of continuous observation within these fields and remove any periods greater than half this time length. I define a period of continuous observation as a period in which there are no observation time gaps of greater than 20% of the entire field baseline. For our 250-night observation baseline, this equates to gaps of 50 days 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 the final sample, I elected to remove any signals with a relative amplitude $> 1.0$. This removed 58 signals, and manual inspection of all 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 the sample.

The 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.

#### 5.1.3.7 Cross-matching with Gaia

To assess this variability period sample within a meaningful scientific context, I elected to use Gaia Data Release 2 (DR2, Gaia Collaboration et al. 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 3.3.1 of this thesis.

As an extension of the Gaia DR2 data, the most recent Tess Input Catalogue (TICv8, Stassun et al. 2018) 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 (EDR3, Gaia Collaboration et al. 2021) contains improved precision on the astrometric fits to many objects from Gaia DR2, however as I use many derived parameters from external catalogues I elected to continue using the DR2 parameters throughout this study. The strengths of the EDR3 data will be demonstrated in Chapter 6.

#### 5.1.3.8 Extinction correction

In the final data products, I assess variability in the context of the colour–magnitude diagram which requires the calculation of absolute magnitudes. To be as accurate as possible, I combined Gaia G magnitudes ($G$) with distance estimates derived from Gaia parallax and accounted for extinction. I 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 ($A_{\mathrm{G}}$). 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. (2018). The final value for absolute magnitude was calculated using the formula:

$$M_{\mathrm{G}} = G - 5\log_{10}(\mathrm{distance}) + 5 - A_{\mathrm{G}}. \tag{5.2}$$

## 5.2 Results

Using the G-ACF period extraction pipeline, I derived variability periods for 16,880 stars observed with NGTS. A subset of these results is shown in Table 5.2, along with positions and cross-match data. The format of the results table is shown in Table 5.1.

Table 5.1: Variability periods, amplitudes 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 |

Table 5.2: A sample of variability periods, amplitudes and catalogue cross-match identifiers in the NGTS data set. Some catalogue cross-match columns have been excluded for publication clarity. The full table is available in a machine-readable format as supplementary material on the online journal and at CDS via anonymous ftp to `cdsarc.u-strasbg.fr` (130.79.128.5) or via `https://cdsarc.unistra.fr/viz-bin/cat/J/MNRAS`.

| 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 5.3: 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 5.1.3.6).

| 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 |

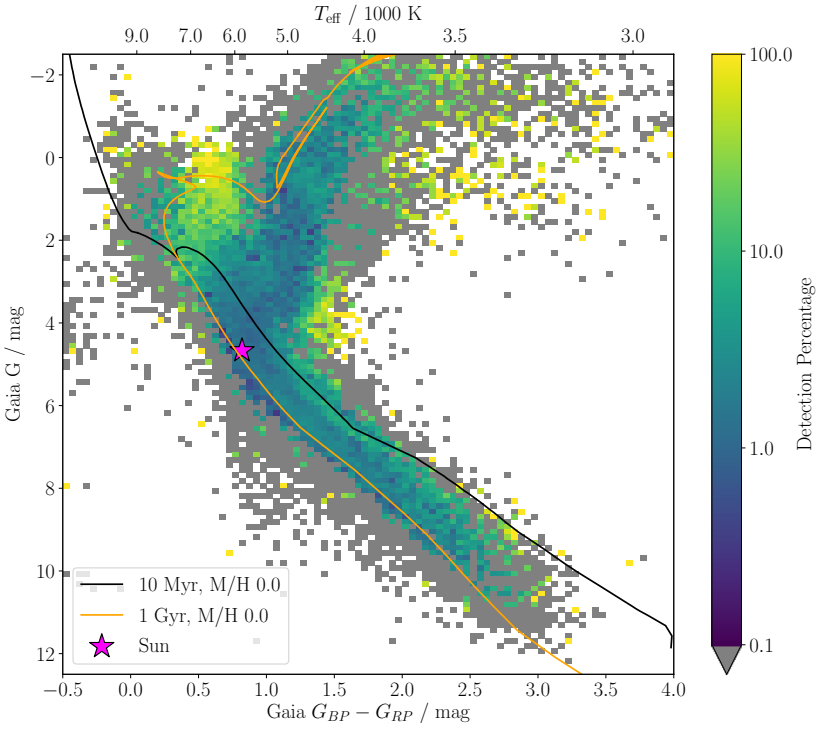



Figure 5.5: Binned colour–magnitude (HR) diagram of the NGTS variability sample. PARSEC v1.2 (Bressan et al. 2012) Solar metallicity isochrones of ages 10 Myr and 1 Gyr are included as solid black and orange lines respectively. The colour indicates the empirical detection percentage per bin. This is defined as the ratio of the number of detected periodic signals to all observed objects per bin. 0 detections within bins are coloured grey.

### 5.2.1 Periodicity in colour–magnitude space

Figure 5.2 shows the variability sample in colour–magnitude space. Table 5.3 details the breakdown of outputs from the pipeline. Once cross-matched with TICv8, I was 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%. The detection percentage varies in colour–magnitude space as shown in Figure 5.5, highlighting potential regions of increased variability or increased sensitivity of NGTS and the signal detection pipeline.

All conversions between $T_{\rm eff}$, $G_{BP} - G_{RP}$ 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)[b], interpolated using a univariate cubic spline. The isochrones in the HR diagrams are taken from PARSEC v1.2S (Bressan et al.

[b] 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

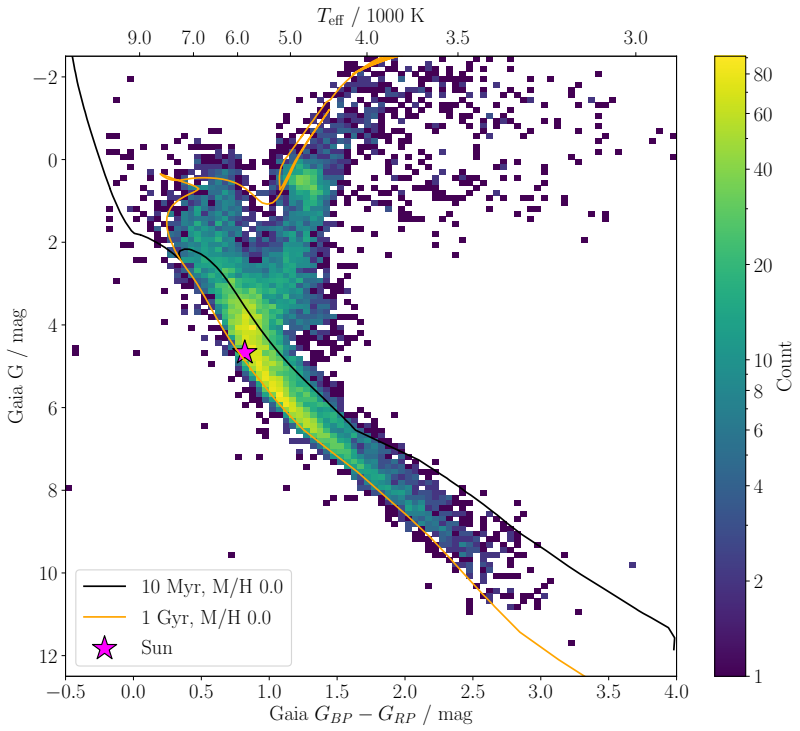



Figure 5.6: As Figure 5.5. The colour indicates the number of objects with detected variability within each colour–magnitude bin.

2012). I elected to use these isochrones as they have been proven to fit the Gaia DR2 main sequence well in Gaia Collaboration et al. (2018a). I produce isochrones using PARSEC v1.2S, selecting the Gaia DR2 passbands from Gaia Collaboration et al. (2018a)[c]. 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. I note, as shown in Gillen et al. (2020), that the PARSEC v1.2 models appear to be less reliable at pre-main-sequence ages, but should be sufficient for their indicative use in this study.

Figure 5.5 highlights regions of interest in terms of detection percentage. Additionally, Figure 5.6 shows the number of detections in each bin. Where detection percentage approaches 100% this is often indicative of a single variable object falling in this colour–magnitude bin. As in Gaia Collaboration et al. (2019), I 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 can also see a region of increased variability at the 'elbow'

[c] using the CMD 3.4 input form at http://stev.oapd.inaf.it/cgi-bin/cmd

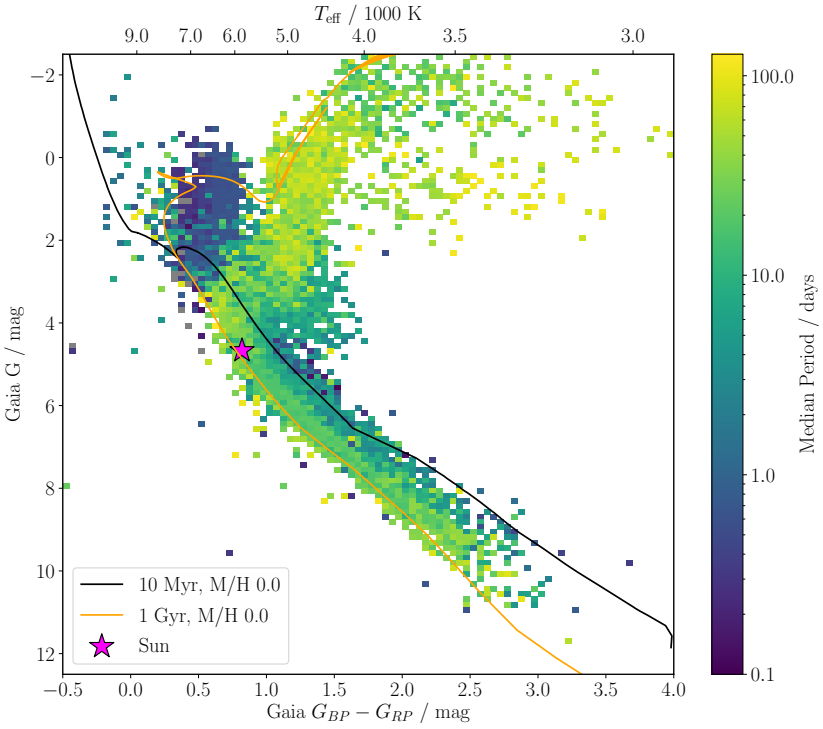



Figure 5.7: As Figure 5.5. The colour indicates the median variability period within each colour–magnitude bin.

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, or RS Canum Venaticorum variable binaries, such as those observed spectroscopically by Strassmeier et al. (1993).

In Figure 5.7 I plot the median period in each colour–magnitude bin. Of particular interest, we can 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 region will be attributed to pulsations; a comparison to Figure 2.7 indicates this region of the HR-diagram is occupied by pulsating variables such as RR-Lyrae, $\gamma$-Doradus and $\delta$-Scuti stars on the instability strip. There may also be 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 (Sikora et al. 2019; David-Uraz et al. 2019). These targets are prime candidates for future spectropolarimetric observations to detect and characterise the magnetic fields of these stars (e.g., Grunhut et al. 2017).

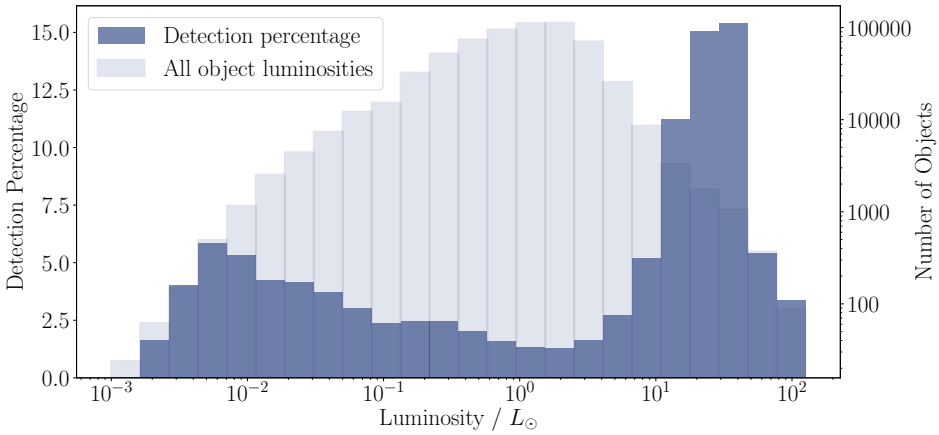



Figure 5.8: Histogram of the empirical detection percentage for all sources against luminosity, as well as the luminosity distribution for all observations.

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

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

1) One 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) One 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 pre-main-sequence (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. The median period observed for the bulk of main-sequence objects is 20 to 30 days, as expected.

I plot detection percentage vs luminosity in Figure 5.8. Luminosity values are taken from ticv8 (Stassun et al. 2018), calculated as:

$$\frac{L}{L_\odot} = \left(\frac{R}{R_\odot}\right)^2 \cdot \left(\frac{T_{\rm eff}}{5772}\right)^4 . \tag{5.3}$$

I 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). $T_{\rm eff}$ values come from spectroscopic catalogues where available, otherwise they are derived from the de-reddened $G_{BP} - G_{RP}$ colour.

As expected, I recover a much higher fraction of variable signals from more luminous stars, with up to 15% of the brightest objects in the sample having detectable variability signals. These objects will correspond to extremely bright giant stars, where one would expect large-amplitude variability arising from pulsations. The lowest number of variable objects coincides with the peak in the number of objects (at 1.5–2.5 $L_{\odot}$), where I detect variability in $< 2\%$ of objects. I also observe an increase in detection percentage for the faintest objects. Here one 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 (Figure 5.2), 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 $\sim 1.6$ in luminosity), this will not have a large effect on the plotted distribution.

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

### 5.2.2 Example variability signals

I show six examples of variability signals in Figure 5.9. A table of stellar parameters for each object is included for reference.

I selected the included objects to demonstrate a small selection of the variability I am 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 Figure 5.9 (1 to 6, top to bottom):

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 young stellar object (YSO). Objects above the main sequence with periods of 5 to 10 days 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-day variability.
6) A long period M-dwarf.

Within the observed G-ACF signals we see artefacts arising from 1-day sampling aliases. These aliases are particularly relevant for signals of period < 1 day, where it was necessary to perform the additional verification steps outlined in Section 5.1.3.5.

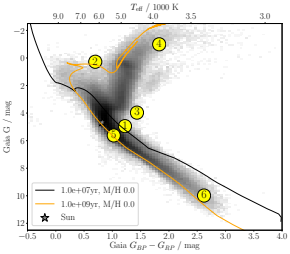
### 5.2.3 Cross-matching with previous catalogues

I cross-matched the NGTS variability periods with photometric variability catalogues in the literature. The ASAS-SN variability catalogue is a large catalogue of photometric variability. I took the latest available data, containing 687,695 variable stars from Jayasinghe et al. (2018) through to Jayasinghe et al. (2021)[d]. I cross-matched my catalogue with the ASAS-SN catalogue, matching on TICv8 ID and Gaia DR2 ID. I found 2,439 matches with periods in both catalogues. A period-period comparison is shown in the left panel of Figure 5.10. The majority (about 1,500 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-day sampling seen in both surveys. The NGTS period extraction pipeline will not return periods close to 1 day or multiples thereof to reduce the number of systematic false positive detections. We see many periods in the ASAS-SN catalogue falling on exact fractions of 1 day, 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-day sampling. Equation 2.3 can be used to calculate these connections and relations of the form

$$P_{\mathrm{ASASSN}} = \frac{1}{P_{\mathrm{sampling}} \pm \frac{1}{P_{\mathrm{NGTS}}}} \tag{5.4}$$

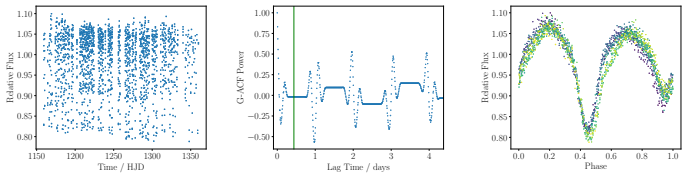
are shown in Figure 5.10. Three obvious sets of aliased periods exist that trace these relations, accounting for approximately 114 matches. We see two sets of related periods arising from 1-day sampling, with the same double phase folding for eclipsing binaries resulting in the set of periods approaching 2 days. There is also a small group of periods connected by aliases arising from 2-day sampling, however, the form of the relation is not shown in the Figure.

[d]The full catalogue is available at https://asas-sn.osu.edu/variables



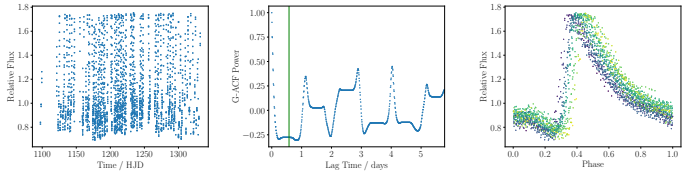

(1)

| Parameter | Value |
|---|---|
| NGTS ID | NG1615+0000.36958 |
| G-ACF Period | 0.438 days |
| Signal Amplitude | 23.6 % |
| Gaia $G$ Magnitude | 14.76 |
| Gaia Abs Magnitude | 4.95 |
| Gaia $G_{BP} - G_{RP}$ | 1.22 |
| $T_{eff}$ | 5098 K |
| Radius | 1.18 $R_{\odot}$ |



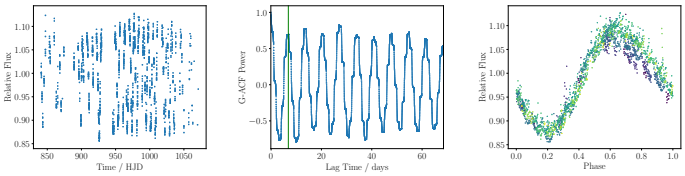

(2)

| Parameter | Value |
|---|---|
| NGTS ID | NG1442-3056.59475 |
| G-ACF Period | 0.579 days |
| Signal Amplitude | 91.4 % |
| Gaia $G$ Magnitude | 14.79 |
| Gaia Abs Magnitude | 0.28 |
| Gaia $G_{BP} - G_{RP}$ | 0.70 |
| $T_{eff}$ | 6540 K |
| Radius | 5.86 $R_{\odot}$ |



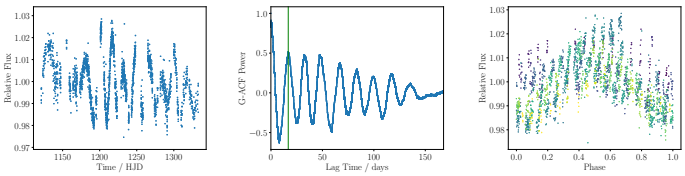

(3)

| Parameter | Value |
|---|---|
| NGTS ID | NG2028-2518.15457 |
| G-ACF Period | 6.85 days |
| Signal Amplitude | 21.4 % |
| Gaia $G$ Magnitude | 14.76 |
| Gaia Abs Magnitude | 3.97 |
| Gaia $G_{BP} - G_{RP}$ | 1.44 |
| $T_{eff}$ | 4512 K |
| Radius | 2.57 $R_{\odot}$ |



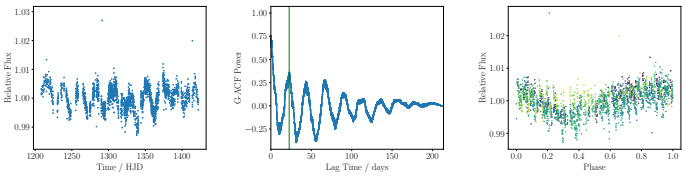

(4)

| Parameter | Value |
|---|---|
| NGTS ID | NG1444-2807.58383 |
| G-ACF Period | 16.8 days |
| Signal Amplitude | 3.48 % |
| Gaia $G$ Magnitude | 13.95 |
| Gaia Abs Magnitude | -0.99 |
| Gaia $G_{BP} - G_{RP}$ | 1.83 |
| $T_{eff}$ | 4072 K |
| Radius | 34.57 $R_{\odot}$ |




(5)

| Parameter | Value |
|---|---|
| NGTS ID | NG2158-3345.20872 |
| G-ACF Period | 22.6 days |
| Signal Amplitude | 1.2 % |
| Gaia $G$ Magnitude | 13.08 |
| Gaia Abs Magnitude | 5.62 |
| Gaia $G_{BP} - G_{RP}$ | 1.02 |
| $T_{eff}$ | 5257 K |
| Radius | 0.80 $R_{\odot}$ |




(6)

| Parameter | Value |
|---|---|
| NGTS ID | NG1146-3633.19225 |
| G-ACF Period | 85.3 days |
| Signal Amplitude | 15.5 % |
| Gaia $G$ Magnitude | 17.40 |
| Gaia Abs Magnitude | 9.98 |
| Gaia $G_{BP} - G_{RP}$ | 2.62 |
| $T_{eff}$ | 3496 K |
| Radius | 0.37 $R_{\odot}$ |

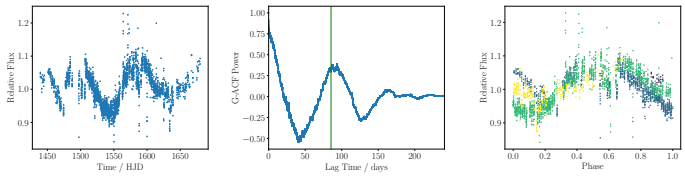



Figure 5.9: Example variable star signals across the HR diagram. From left to right: A table of stellar parameters. The NGTS light curve, binned to 20 minutes. The G-ACF of the light curve, with a green line indicating the extracted period. The light curve phase folded on the extracted period, each successive 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 to 6 correspond to the stars top to bottom. Solar metallicity PARSEC isochrones of ages 10 Myr and 1 Gyr are included as solid black and orange lines respectively.

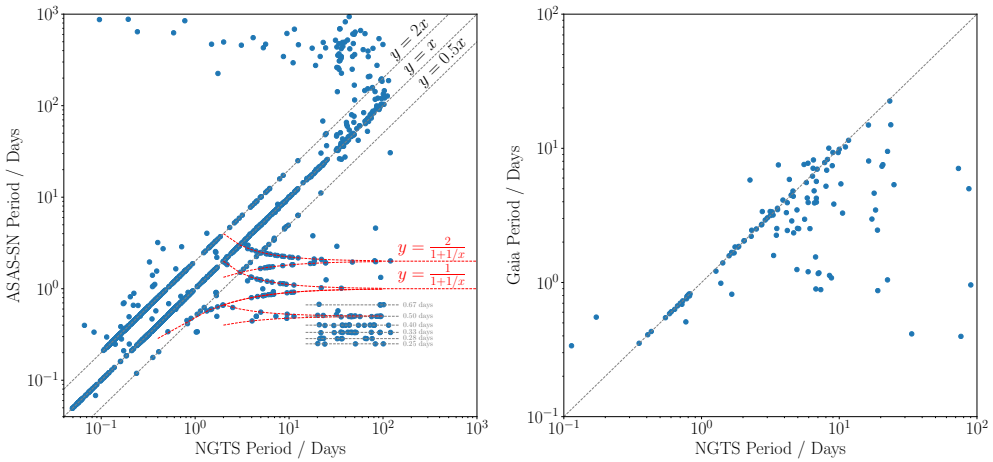



Figure 5.10: NGTS variability periods from this study compared with ASAS-SN periods (left) and Gaia (right). Lines of equal period from both surveys are plotted in light grey, and for the ASAS-SN comparison lines showing periods differing due to incorrect phase folding by a factor of two shorter or longer are also plotted in light grey. The red dashed lines and associated equations indicate relations between periods arising from 1-day 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.

I was 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-day rotation period for an object which not present in the MEarth catalogue (NG1444−2807.12982). For the two other objects (NG1214−3922.6732 and NG0458−3916.13434), NGTS detected a near 100-day 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 et al. 2018c), where photometric time-series data was processed to detect and classify variable sources (as described in Holl et al. 2018). Photometric time series from Gaia are sparsely sampled and not optimised to detect photometric variability, so may produce an incorrect period. I cross-matched 126 objects against the rotation period database provided by the Gaia Collaboration on VizieR[e], these period comparisons are shown in the right panel of Figure 5.10. For 60 of the 126 periods that differed, I 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 periods as in the cross-matching with ASAS-SN. This is a clear demonstration that the NGTS period recovery pipeline is well suited

[e] https://vizier.cds.unistra.fr/viz-bin/VizieR-3?-source=I/345/rm

to deal with aliases arising from 1-day sampling

Finally, I cross-matched the NGTS sample with the variability catalogue from Canto-Martins et al. (2020), which searched for rotation periods in 1000 TESS objects of interest. I 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 days and 5.4 days for NGTS and TESS respectively), although the phase fold on the NGTS data was cleaner using the NGTS period.

Although a large number of photometric variable stars are known in the Kepler field, I am unable to cross-match with these catalogues as NGTS does not observe this part of the sky. Additionally, I do not attempt to cross-match with small catalogues and papers reporting detections of individual variable objects. Two large variability catalogues I 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 (Masci et al. 2018; Tonry et al. 2018). Due to the small overlap in brightness and a large number of candidates in each catalogue, I elected not to perform a cross-match. Further cross-matching with smaller catalogues will be possible, as I provide the position in RA and Dec, as well as TICv8 and Gaia DR2 identifiers (where available) for all 16, 880 variable sources as a part of the paper.

### 5.2.4 Period ranges of interest

I break the results down into unevenly spaced intervals in variability period to assess how samples of similar variability periods are distributed in colour–magnitude space in Figure 5.11. This reveals more information than Figure 5.2 as it is possible to probe into the high-density main sequence. I have selected the period ranges empirically taking into account the sampling gaps at 14 and 28 days 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

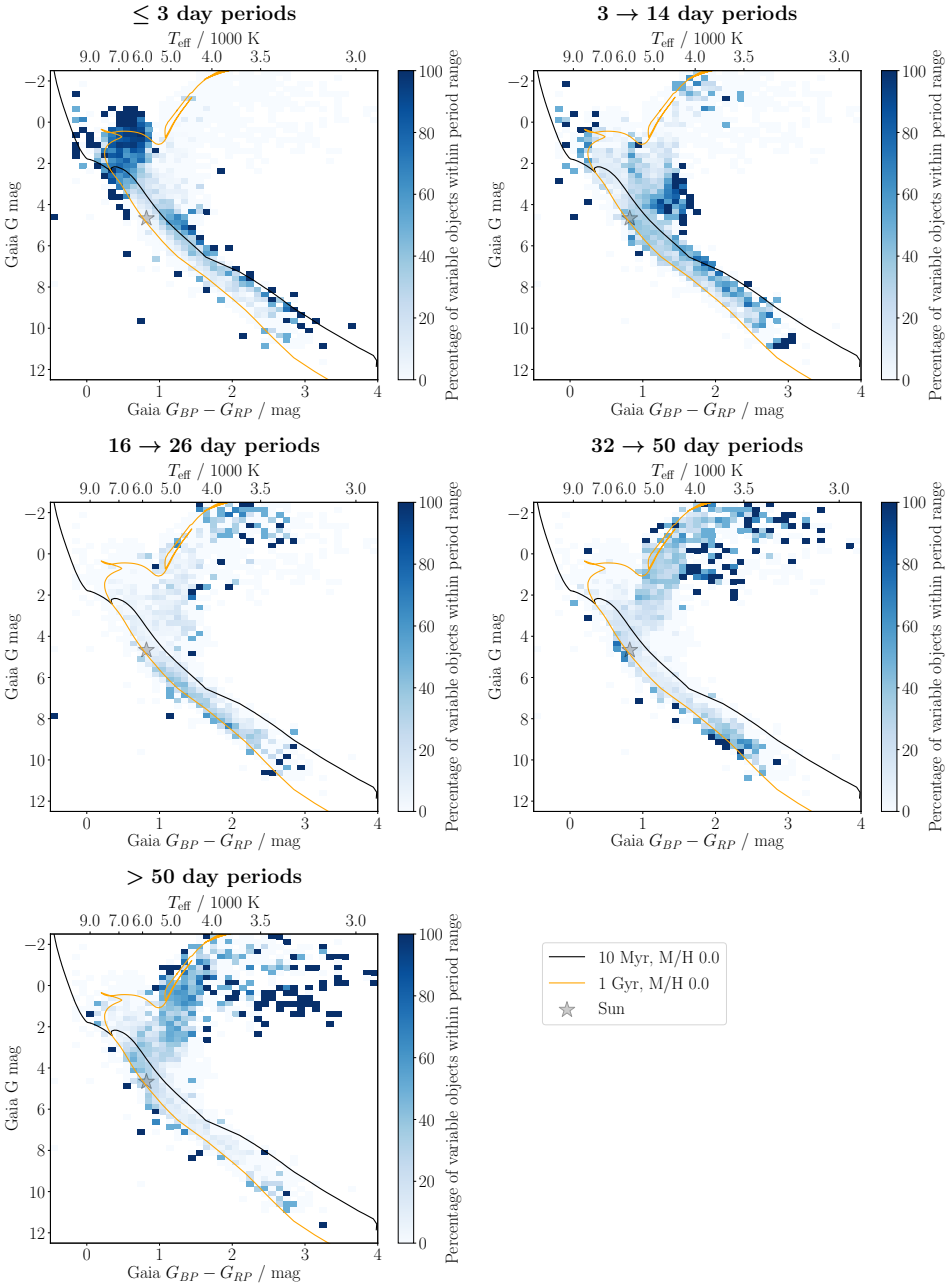



Figure 5.11: HR diagrams for the NGTS variability sample broken down into period ranges. Periods in the sample range from ∼ 0.1 to 130 days. The colour bar indicates the percentage of all variable objects across all period ranges that lie in this specific colour–magnitude–period bin. The sum of each bin across all 5 subplots will equal 100%. Solar metallicity PARSEC isochrones of ages 10 Myr and 1 Gyr are included as solid black and orange lines respectively.

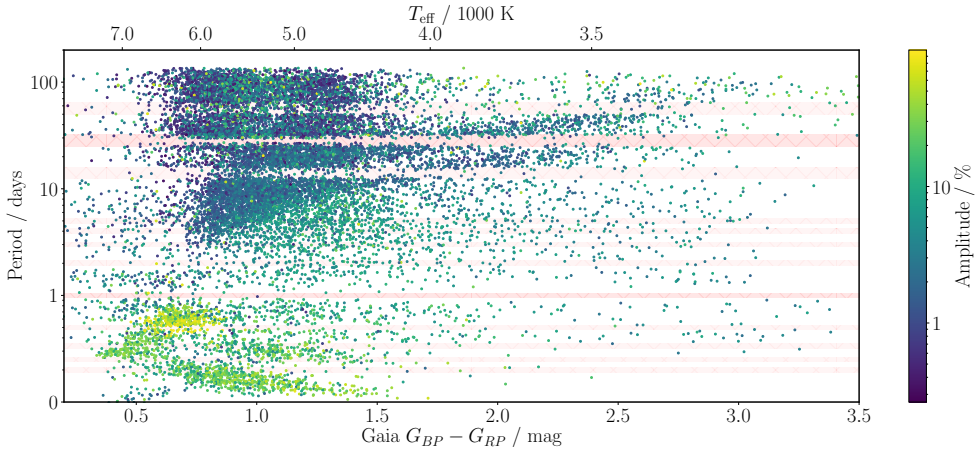



Figure 5.12: Effective temperature and Gaia $G_{BP} - G_{RP}$ colour against period for 16,880 stars. The colour indicates the $5^{\rm th} - 95^{\rm th}$ percentile spread of the signal in relative flux. To aid the eye, horizontal strips indicate regions of period space likely affected by systematics arising from the Moon or the 1-day sampling alias, with multiples of these periods more transparent.

pulsate with periods of less than 1 day. The peak density for less evolved stars is above the main sequence at this period range. Between 1 and 10 days, 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 et al. 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 of 3 to 14 days, 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 days, we see the peak density move towards the main sequence as well as a distinct lack of objects above the main sequence. At $>$ 30 days, 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 $32 - 50$ to $>$ 50 day periods we see the bulk of objects move further up the RGB and further down the main sequence towards cooler temperatures and redder colours.

### 5.2.5 Periodicity-colour comparison

I plot my variability periods against colour in Figure 5.12 and see several prominent features. Most striking is the high density of stars known in the literature as the 'I-Sequence' (Barnes 2003) or the 'Ridge' (Kovács 2015) spanning a period range from $4 - 40$ days 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 days) objects at $G_{BP} - G_{RP}$ of ~ 1.0. We would expect a higher density of detections at this colour due to the high-density main-sequence turnoff and red clump, as shown in Figure 5.2(b). Older main-sequence stars in this colour range may exhibit long period rotational modulation. Within this colour range lies the Cepheid instability strip, 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 day. This population corresponds to the top of the main sequence on an HR diagram.

We see two distinct groups of objects in a period range shorter than 1 day, trending to short periods with increasing colour index ($G_{BP} - G_{RP} \sim 0.75 \rightarrow\sim 1.5$). On further analysis, I confirmed that 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 Figure 5.9), 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 minimisation step described in Section 5.1.3.5. 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 can observe a period upper envelope of stars from $G_{BP} - G_{RP} > 1.5$ with the period increasing for the reddest stars. We see some objects with $G_{BP} - G_{RP} > 2.5$ having variability periods up to and exceeding 100 days. These objects are discussed in detail in Section 5.3.2.

### 5.2.6 Period bi-modality

Within the I-sequence envelope we can see a hint of a region lacking in periodic signals between ~ 3500 K and ~ 4500 K ($G_{BP} - G_{RP} \sim 2.5$ to 1.5) and ~ 15 to ~ 30 days. This gap has been the topic of extensive discussion in recent papers (such as McQuillan et al. (2013); Davenport & Covey (2018)), and although faint, I do observe this gap in this ground-based data set. This gap has previously been fitted using a gyrochrone, roughly following a ${T_{\rm eff}}^{1/2}$ relation (Davenport & Covey 2018), as well as an empirical model using a similar ${T_{\rm eff}}^{1/2}$ relation (Gordon et al. 2021).

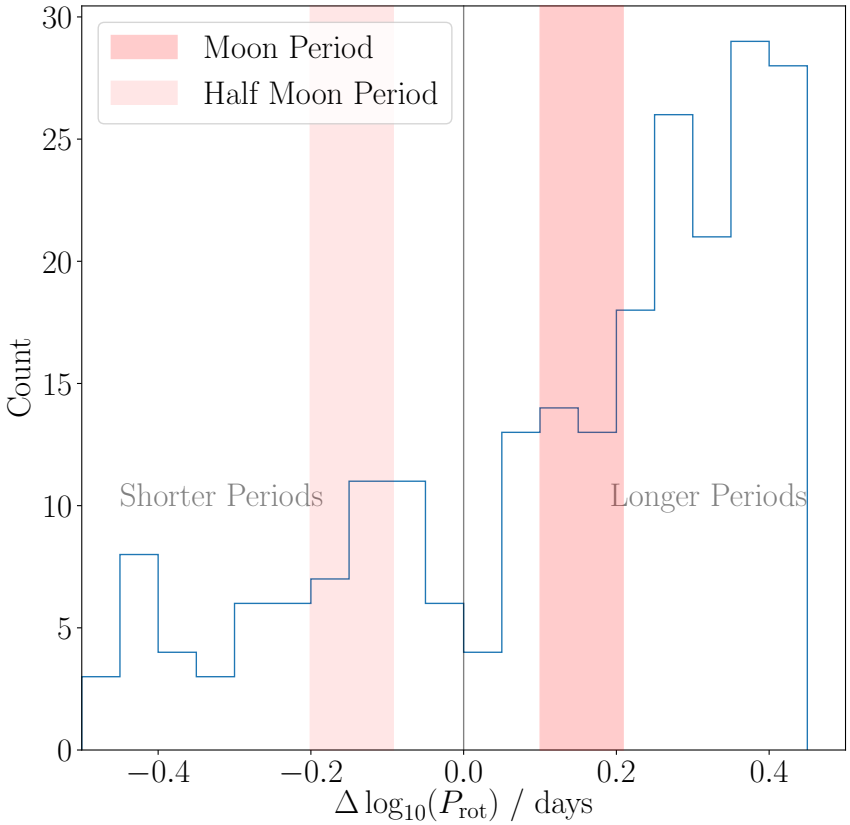



Figure 5.13: Distribution of the distance from a 600 Myr gyrochrone of the log periods for stars $1.4 < G_{BP} - G_{RP} < 2.2$. We see two peaks in the distribution, with a reduced number of rotation periods along the model gyrochrone (grey vertical line). The range of distances from the model to the Moon and half Moon period is included to demonstrate the lower density of objects does not arise from a gap due to the Moon.

To demonstrate the gap is present in my data, I conduct the same analysis as in Figure 3 of Davenport & Covey (2018). I subtract model periods calculated with a 600 Myr gyrochrone defined in Meibom et al. (2011) from our periods. I constrain the 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 Figure 5.13 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 Figure 5.14 I separate the sample into three sub-samples based on a bi-modality gap model and empirical short-period lower limit from Gordon et al. (2021). I calculate 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 Figure 5.2. I 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.

For this part of the analysis, I 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. I 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. I remove objects with the `'subgiant'` or `'RBG'` flags.

I define 3 sub-samples using several model constraints in period–colour space. I 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. I calculate the upper and lower edge of the gap using the model defined in Gordon et al. (2021), and select stars from the I-sequence envelope on either side of this branch. This model was only defined for $0.8 < G - G_{RP} < 1.05$, so I only use objects within this bound to define the sub-samples. The 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 precise details of the two models are given below and plotted in Figure 5.14a.

**Angus model**

I use the Praesepe-calibrated gyrochronology relation defined in Angus et al. (2019), as well as the parameters defined in Table 1 of this paper. The mathematical form of this fifth-order polynomial relationship is given in Equations 5.5 & 5.6 below for two different $G_{BP} - G_{RP}$ regimes:

$$\log_{10}(P_{\rm rot}) = c_{\rm A} \log_{10}(t) + \sum_{n=0}^{4} c_{\rm n} [\log_{10}(G_{BP} - G_{RP})]^n \tag{5.5}$$

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

$$\log_{10}(P_{\rm rot}) = c_{\rm A} \log_{10}(t) + \sum_{m=0}^{1} b_{\rm m} [\log_{10}(G_{BP} - G_{RP})]^m \tag{5.6}$$

for stars with $G_{BP} - G_{RP} > 2.7$. Here $P_{\rm rot}$ is the rotation period in days, and $t$ is age in years. I use the best-fit coefficients from Angus et al. (2019) in Table 5.4.

Table 5.4: A table of model coefficients used in Equations 5.5 & 5.6 as defined in Angus et al. (2019).

| Coefficient | Value |
|---|---|
| $c_A$ | $0.65 \pm 0.05$ |
| $c_0$ | $-4.7 \pm 0.5$ |
| $c_1$ | $0.72 \pm 0.05$ |
| $c_2$ | $-4.9 \pm 0.2$ |
| $c_3$ | $29 \pm 2$ |
| $c_4$ | $-38 \pm 4$ |
| $b_0$ | $0.9 \pm 0.5$ |
| $b_1$ | $-13.6 \pm 0.1$ |

Table 5.5: A table of model coefficients used in Equation 5.7 as defined in Gordon et al. (2021).

| | $A$ (days) | $B$ (days) | $x_0$ |
|---|---|---|---|
| upper edge | 68.2277 | −43.7301 | −0.0653 |
| lower edge | 34.0405 | -2.6183 | 0.3150 |

**Gordon model**

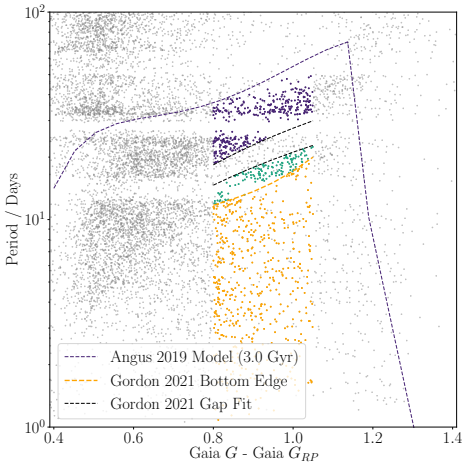
I use the K2 calibrated model from Gordon et al. (2021) to define the upper and lower edges of the bi-modality gap seen in the I-sequence envelope. The gap edges are fitted using a function of the form:

$$P = A(G - G_{RP} - x_0) + B(G - G_{RP} - x_0)^{1/2} \tag{5.7}$$

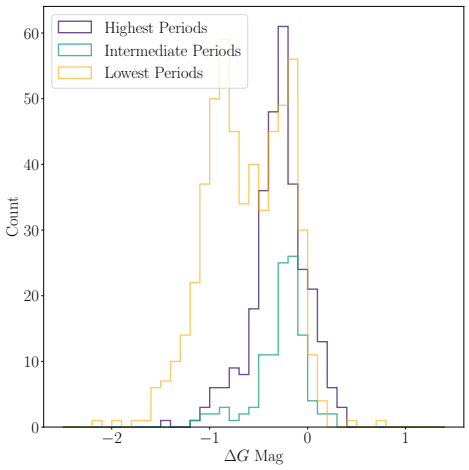
where $P$ is the rotation period in days. This equation is defined empirically for K2 stars with $0.8 < G - G_{RP} < 1.05$. We use the best fit coefficients defined by Gordon et al. (2021) in Table 5.5.

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. I instead define the lower edge by eye, taking the edge-detection fit line from the Gordon et al. (2021) paper.

The histograms in $\Delta G$ plotted in Figure 5.14b show two similar single-peaked distributions from the two longer period sub-samples and a distinct double-peak distribution for the shorter period sub-sample. I 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 a previous observation from the HR diagram: a group of very short period objects just above the main sequence, which could correspond to a sample heavily contaminated by binary sources. The two longer period




(a) Gaia $G_{BP} - G_{RP}$ 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 I do not plot the large model uncertainties defined for the Angus et al. (2019) model for stars outside the range $0.56 < G_{BP} - G_{RP} < 2.7$.




(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 5.14: 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.

sub-samples appear to have by-eye similar distributions of $\Delta G$, which leads me to believe the two branches are drawn from similar stellar populations in terms of colour, intrinsic brightness and multiplicity.

## 5.3 Discussion

### 5.3.1 Comparison to similar studies

The NGTS data set demonstrates that it is possible to use ground-based photometry to conduct stellar variability studies previously only done on this scale using space-based data. In contrast to, for example, the Kepler data set used by McQuillan et al. (2013) 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. I overlay data from the Kepler rotation study by

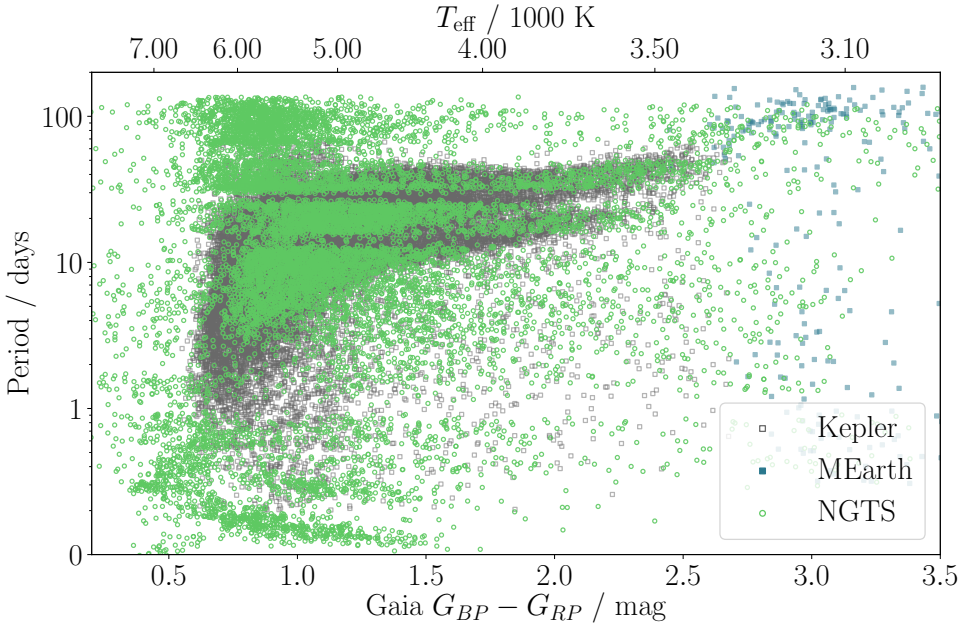



Figure 5.15: Effective Temperature vs Period data compared for this study (NGTS data, green circles), McQuillan et al. (2013) (Kepler data, grey squares) and Newton et al. (2018) (MEarth data, blue squares).

McQuillan et al. (2013) with my variability sample in Figure 5.15. In particular, we see a high density of objects at $G_{BP} - G_{RP} \sim 1.0$ with periods longer than roughly 40 days 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.

In addition to finding astrophysical signals of interest, I was also able to detect systematic periodicity within the entire data set down to amplitudes of 0.3%. As NGTS's primary scientific goal is to search for planets, in this context these very low amplitude systematic signals rarely matter, especially when follow-up observations and precise modelling are factored in. This study highlights the power of ground-based photometric surveys in terms of the size and precision of the data set. I have been able to extract a data set that 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 many increased sources of noise and aliasing. I note a lower recovery rate of periodic signals than in other studies. McQuillan et al. (2013) found variability in 25.6% of their $\sim 130,000$ objects, Gordon et al. (2021) found variability in almost 13% of their $69,000$

objects, and NGTS was able to find variability in about 2% of 829,481 objects. I note that 21% of all objects were flagged as having signals arising from Moon contamination, the largest source of systematic noise in this study.

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

### 5.3.2 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. I also find these extremely long periods ($>$ 100 days) in the M-dwarf population of this sample. Figure 5.12 shows an upwards trend in period in the mid-M dwarf sample at $< 3500K$. To provide a useful comparison to the MEarth rotation study, I also assessed this trend for just dwarf stars (as defined by `evolstate`). This sample contains751 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-day time series, which allows robust extraction of periods up to roughly 125 days in length. Newton et al. (2018) observed periods up to 140 days 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 days. The NGTS rotation periods qualitatively agree with the distribution of rotation periods seen in M dwarfs by Newton et al. (2018), however, these data reach the detectable period limit of these NGTS observations just shy of the $\sim$ 140 day 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. I 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 year continuously, with a one-year gap before another year of continuous observation. Most of the sky will only be observed for 28 days at a time, meaning a maximum of 14-day 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 a range of pointings.

### 5.3.3 Period bi-modality

I continue the ongoing discussion regarding the rotation period gap (McQuillan et al. 2013; Davenport & Covey 2018; Reinhold et al. 2019; Reinhold & Hekker 2020; Angus et al. 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, I align models from several previous works to a region of lower density in the NGTS data, as shown in Figure 5.13. To show this, I apply the same analysis as in Davenport (2017) and Davenport & Covey (2018), subtracting a 600 Myr gyrochrone taken from Meibom et al. (2011) which was assessed as the best fit for the observed gap in the Kepler data. A histogram of distance from this gyrochrone in $\log_{10}(P_{\rm rot})$ demonstrates a region of lower density close to this gyrochrone. The aliasing gaps created by the Moon and half-Moon period signals mean I am unable to probe the gap as it approaches these period ranges.

By utilising empirical models from previous studies on Kepler and K2 data, I separated my sample into three sub-samples: this is seen in Figure 5.14. 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. I note, similar to the Davenport & Covey (2018) study that I have not accounted 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 ($\sim$ 100 Myr) (Gillen et al. 2020), Praesepe ($\sim$ 800 Myr) (Rebull et al. 2016a, 2017), Ruprecht 147 ($\sim$ 3 Gyr) (Gruner & Barnes 2020) and M67 ($\sim$ 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 this 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. Before 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 re-couple. 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 $\sim$ 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 analyses 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. I do not observe such an obvious decrease in signal amplitude in the full NGTS sample, and when considering a smaller range of amplitudes more aligned with the K2 sample I 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% 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.

## 5.4 Conclusions

In this study I extracted robust variability periods for 16, 880 stars out of 829, 481 stars observed with the Next Generation Transit Survey (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 generalisation of the autocorrelation function to irregularly sampled data, I detected variability amplitudes down to levels of 0.3%. The contamination of signals by systematics demonstrates that using ground-based photometry requires further thought than using much cleaner space-based data 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-day periodic sampling intrinsic to ground-based observations. I demonstrated I can overcome these limitations and produce robust variability signals across the sample.

In comparison to previous large-scale stellar variability studies, I note that with NGTS we observe across the Southern sky (in comparison to Kepler's single pointing, as in McQuillan et al. (2013) and Davenport & Covey (2018)). We do not pre-select our targets as is the case for Kepler and K2, so I can observe variability across a more varied stellar sample. In particular, I extracted long-term variability periods for a population of cool dwarfs, similar to a population observed by Newton et al. (2018) using MEarth. This was 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-day) baseline allowed 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 the rotation period 'I-Sequence', I observed a gap between 15 and 25 days, first observed by McQuillan et al. (2013), 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) I demonstrated that the gap is present in this data set, and also showed 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.

I also concluded 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. I 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 of how stellar angular momentum evolves over the lifetime of a star. These conclusions lead to the work conducted in Chapter 6, in which open clusters are observed with NGTS and a search for variable signals is conducted. Additionally, I observed several 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 aid in confirming whether these stars are young, single stars such as T-Tauri objects, or multi-object 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.

The rotation period data produced in this Chapter will be made publicly available through the Vizier catalogue access tool and the MNRAS online journal upon publication in MNRAS.

# Chapter 6

# Periodic Stellar Variability in the Open Cluster NGC 6633

In this Chapter, I will outline rotational analysis completed on a subset of the NGTS data set targeting the young open cluster NGC 6633. This work draws upon the large-scale rotational study conducted in Chapter 5 but with a focus on open clusters for reasons explained in both Chapter 1 and the conclusions of Chapter 5. I will provide some background context for this work, describe the data taken with NGTS, discuss the methods used and the implementations and finally report the rotational results found.

The NGTS Open Clusters Working Group identified NGC 6633 as an open cluster of interest for observation during 2019 and 2020. NGC 6633 is a fairly young open cluster ($\sim$ 500 Myr, slightly younger than the Hyades and Praesepe), and at the time of observation, had limited previous photometric survey data available. As we have seen in Chapter 1, open clusters provide excellent stellar laboratories to understand stellar evolution. In particular, studying co-eval populations of stars allows us to understand the similarities and differences in stellar properties of stars of a given age, such as mass, temperature, surface gravity and angular momentum.

It is possible to place age estimates on open cluster populations more easily than single stars. Using complimentary ageing techniques can afford even greater precision, for example, combining isochrone ageing and gyrochronology such as in Angus et al. (2019) or the combination of lithium abundance and rotational age estimates in open clusters (Jeffries et al. 1997; Jeffries 1997). I have already touched on gyrochronology in Section 1.4.2, which fits empirical

relations to spectral type, rotation period and age. By measuring rotation periods of stars within open clusters, it is possible to fit calibrated gyrochronology relations to the colour–period diagram to estimate the age of the cluster (such as James et al. 2010; Delorme et al. 2011; Angus et al. 2015; Gillen et al. 2020; Gruner & Barnes 2020).

Isochrone fitting (Section 1.3.2) can also aid in understanding the age of an open cluster, however, with several caveats. It is possible to estimate the cluster's age by fitting an isochrone to the observed colour–magnitude diagram. The accuracy of this fit will depend on the spacing of isochrones of different ages and, in particular, along the main sequence, isochrones are extremely close together. Low mass stars with convective outer shells will remain on the main sequence considerably longer than high mass stars that will more rapidly deplete their hydrogen reserves and begin to evolve, adding further difficulty to isochrone ageing for main-sequence low-mass stars. At the main-sequence turnoff, isochrones are spread further apart; with sufficiently precise measurements for stars in an open cluster, it is possible to produce age estimates with errors of order 5–10% (Angus et al. 2019). Isochrone fitting works in complement to gyrochrone fitting; gyrochronology is best suited to main-sequence stars that exhibit spin-down, whereas isochrone fitting better suits off-main-sequence stars for which the rotational evolution is less well modelled.

It is also possible to age open clusters with fractional surface lithium abundance measurements. Lithium is the only metal produced in significant quantities in the big bang, and as such, stars are created with lithium fractions similar to primordial levels. Lithium is destroyed in the inner layers of stars through proton capture reactions when temperatures exceed 2.5 million Kelvin. Lithium is transported to these inner layers through internal mixing processes such as convection (Pinsonneault 1997). The presence (or absence) of photospheric lithium provides evidence as to whether enough time has passed for lithium to be transported and destroyed in the inner layers of a star. The Lithium Depletion Boundary (LDB) is the observational limit below which cores of low-mass stars do not reach high enough temperatures for lithium burning to occur; measurements of the LDB for low-mass stars within a cluster can yield accurate age measurements independent of other ageing techniques such as isochrones or rotation (Burke et al. 2004). Lower mass PMS stars ($\lesssim 0.5 M_{\odot}$) will rapidly burn lithium as they are fully convective during the pre-main-sequence, which enables rapid transport of surface lithium to the inner layers. However, lithium destruction will only occur once the stellar interior has reached a high enough temperature to enable Li-burning. The rate at which the stellar core temperature approaches the lithium destruction temperature is a strong function of mass. Hence, by measuring the luminosity dependence of lithium fractions in a cluster, it is possible to calculate an approximate age for open clusters (Burke et al. 2004; Jeffries & Oliveira 2005). Lithium depletion in the context of this work is used as a complementary source of literature ages for

open clusters; for more details on the technique, I refer the reader to a theoretical examination by Burke et al. (2004); Bildsten et al. (1997) or the Li-depletion ageing studies of open clusters such as Jeffries et al. (1997); Jeffries (2000); Jeffries & Oliveira (2005); Martín et al. (2018).

NGC 6633 is an open cluster in the constellation Ophiuchus thought to be of a similar or younger age to Praesepe and the Hyades at roughly 500 Myr, with lower metallicity (Harmer et al. 2001; Lyngå 1988; Strobel 1991). There has not been significant directed research towards NGC 6633: a study by Jeffries (1997) analysed a small sample of lithium abundances and spectroscopic rotation ($v \sin i$) rates for low-mass stars in the cluster, noting a similar rotation rate in cluster member stars to the Hyades but with a different lithium depletion pattern. Jeffries hypothesised that the increased lithium abundances, as well a greater spread in Li abundance values observed in NGC 6633 compared to the Hyades, could be as a result of lower metallicity and shallower convective zones of member stars rather than a significant age difference. They concluded that their spectroscopic rotational measurements were insufficiently sensitive to yield interesting results for this cluster. The relatively low metallicity of NGC 6633 was confirmed more precisely by Jeffries et al. (2002), spectroscopically estimating [Fe/H] = $-0.096 \pm 0.081$ for the cluster. This follow-up study also estimated the age of NGC 6633 as being marginally younger than Praesepe and the Hyades. A later study, by Harmer et al. (2001), analysed X-ray data from the cluster to determine whether the magnetic activity was similar to that of stars within the Hyades and Praesepe; however, their work was fairly inconclusive due to weak X-ray signals resulting in a high luminosity threshold for detection.

The previous inconclusive studies may be due to the large distance and interstellar extinction of NGC 6633. The distance to the NGC 6633 has previously been reported as $\sim$ 348 pc by Schmidt (1976), $\sim$ 312 pc in Lyngå (1988) and more recently as $394.3 \pm 2.4$ pc by Pang et al. (2021) using the latest Gaia EDR3 data. Pang et al. (2021) find significant extinction of about 0.558 mag in the visual using Gaia's latest data release (EDR3): this extremely precise astrometry and photometry allows insight into this distant cluster not previously possible.

NGTS has previously conducted similar targeted observations of open clusters, and Gillen et al. (2020) (described in Section 8.1) has demonstrated the utility of NGTS photometry in the context of open cluster variability studies. This study draws on the methods used by Gillen et al. (2020), particularly the use of a combination of period extraction techniques to confirm stellar variability periods from photometric light curve data. I extend the variability analysis techniques developed in earlier chapters of this thesis, particularly the G-ACF as a variability extraction tool in conjunction with previously well-established variability detection methods.

This Chapter is laid out as follows: Section 6.1 will explain the data used in this study: NGTS photometry and Gaia astrometry, and the literature membership lists used for the cluster. Section 6.2 will detail the methods used for extracting rotation periods from NGTS light curves,

Table 6.1: A comparison of approximate open cluster properties for four open clusters discussed in this study. References: [1]Rebull et al. (2016a). [2]Melis et al. (2014). [3]Soderblom et al. (2009). [4]Jeffries et al. (2002). [5]Pang et al. (2021). [6]Douglas et al. (2016). [7]van Leeuwen (2009). [8]Cummings et al. (2017).

| Cluster | Age | Distance | # of Members | [Fe/H] |
|---|---|---|---|---|
| The Pleiades | 125 ± 8 Myr[1] | 136.2 ± 1.2 pc[2] | >1000[1] | +0.03 ± 0.02[3] |
| NGC 6633 | 400 – 600 Myr[4,5] | 394 ± 2.4 pc[5] | 300[4] | −0.096 ± 0.08[4] |
| Praesepe | 670 ± 67 Myr[6] | 181.5 ± 6.0 pc[7] | 743[6] | +0.156 ± 0.004[8] |
| The Hyades | 727 ± 75 Myr[6] | 46.5 ± 0.5 pc[7] | 786[6] | +0.146 ± 0.004[8] |

as well as the development of the open-source rotational period finding tool `RoTo`. I will outline the results of this study in Section 6.3, exploring both the rotational modulation of individual objects and assessing the period–colour slow-rotator sequence of the cluster. I compare the slow-rotator sequence of NGC 6633 to clusters of a similar age (Praesepe and the Hyades) and metallicity (the Pleiades) and discuss the gyrochronological results of this study. Table 6.1 outlines some basic properties for the four open clusters discussed in this Chapter to aid the reader with comparisons.

## 6.1 Data

### 6.1.1 Literature membership lists and clustering

Cluster membership was determined from previous catalogues and clustering analysis of the Gaia EDR3 astrometric data. I will briefly explain the catalogues used in this study in this Section; Section 6.1.2 will detail how I combined these catalogues to assess cluster membership for candidate NGC 6633 members.

Two large cluster survey catalogues pre-dating Gaia DR2 were included in the global membership list: the cluster survey from Kharchenko et al. (2013) and the catalogue from Dias et al. (2014). Kharchenko et al. (2013) used stellar data from the PPMXL all-sky catalogue (Roeser et al. 2010) and 2MASS (Skrutskie et al. 2006) to determine kinematic and photometric membership probabilities for stars in a cluster region. Dias et al. (2014) presented a catalogue of mean proper motions and membership probabilities using data from the UCAC4 catalogue (Zacharias et al. 2013). Both of these catalogues used astrometric measurements pre-Gaia, and as such, contain large errors, with Dias et al. (2014) claiming positional errors of 15 to 100 mas and proper motion errors from 4 to $\gtrsim$ 10 mas/yr.

Since Gaia's second data release (Gaia Collaboration et al. 2018c), precise stellar positional and kinematic parameters have allowed cluster membership lists to be refined. The Zari et al. (2018) catalogue used this data to construct precision three-dimensional maps of stellar density

within 500 pc to assess the distribution of young star-forming regions. Two catalogues from Cantat-Gaudin et al. (2018, 2019) used Gaia astrometry to compile a list of 1,229 clusters using an unsupervised membership assignment algorithm. Over-densities of stars within astrometric space ($\mu_{\alpha^*}$, $\mu_\delta$, $\varpi$) were found using an iterative k-means clustering approach to produce membership probabilities for each star. The slightly more recent catalogue from Cantat-Gaudin et al. (2020) uses a neural-network-based approach on the Gaia DR2 astrometry to assign membership probabilities and predict cluster parameters such as age, distance modulus, extinction, and sometimes, metallicity. Kounkel & Covey (2019) applied an unsupervised machine learning algorithm to Gaia DR2's 5-dimensional dataset (3d position and 2d velocity) to identify clusters, associations and co-moving groups. They used HDBSCAN to find clusters of varying densities within the large DR2 dataset and identified 1,901 individual groups consisting of a total of 288,370 stars. These groups were not initially linked to previously known clusters but referred to as 'Theia'. Once linked with Cantat-Gaudin et al. (2018), 198 known open clusters were found, including NGC 6633, which, based on other membership catalogues, is contained within Theia 924. Finally, as part of the work of Gaia Collaboration et al. (2018b), 32 open clusters were analysed, and membership lists were generated based on an agreement of astrometric solutions of candidate cluster members to the assumed astrometric motion of the cluster centre (taken from previous catalogues such as Kharchenko et al. 2013), including radial velocities where available. This membership test was conducted iteratively; I refer the reader to Appendix A of Gaia Collaboration et al. (2018b) for full details of the process.

The release of Gaia Early DR3 (EDR3) (Section 6.1.3, Gaia Collaboration et al. 2021) provided even more precise astrometric parameters. Pang et al. (2021) produced a membership catalogue of 13 open clusters using this data release, including NGC 6633. The authors used an unsupervised machine learning method on the 5-dimensional astrometric solution from Gaia EDR3 (`StarGO`, from Yuan et al. 2018) based on self-organising maps. `StarGO` was originally developed to cluster stars kinematically using DR2 data to determine the galactic origins of halo stars in the Milky Way.

Additionally, I conducted a machine-learning-based clustering analysis using the Gaia EDR3 astrometric parameters. I used the unsupervised clustering algorithm DBSCAN (Ester et al. 1996) to cluster objects based on distance and proper motion from Gaia EDR3. Although the hierarchical version of DBSCAN (HDBSCAN) has previously been shown to be an effective tool for clustering DR2 astrometric data (Kounkel & Covey 2019; Kounkel et al. 2020; Cánovas et al. 2019), since NGTS observed just a single field centred around NGC 6633, I elected to use the standard DBSCAN algorithm to identify candidate cluster members and outliers from a single cluster. Furthermore, as the NGTS field is centred around NGC 6633, I do not consider the positions in RA and Dec when clustering.

I ran DBSCAN in three dimensions: distance (derived from EDR3 parallax, see Section 6.2.1 for details), proper motion in right ascension ($\mu_{\alpha^*}$) and declination ($\mu_\delta$). I first make broad cuts on the data to remove far-outliers in distance and proper motion: I do not attempt to cluster any objects further than 900 pc or with $\mu_{\alpha^*} > 3$ mas y$^{-1}$ mas y$^{-1}$ or $\mu_\delta > 6$ mas y$^{-1}$. These cuts were determined empirically based on the observed distributions of distances and proper motions from EDR3 for the sample and broadly agree with the limits of these values from previous catalogues (9 candidate members were outside this range). I normalised the three dimensions using a linear min-max scaling and ran DBSCAN using $\epsilon = 0.1$, with a minimum number of samples of 10.[a] Varying the value of $\epsilon$ between 0.05 and 0.15 altered the number of identified candidate cluster members by ±77, highlighting the strong dependence of DBSCAN on $\epsilon$. I decided the value of 0.1 based on visualising the three dimensions: this gave a tight grouping of candidate cluster members, with clear outliers in each of the 2d plots of the 3d space (as shown in Figure 6.1). Varying the minimum number of samples did not affect the clustering above a value of 5, for which DBSCAN always returned a single cluster and outliers. I elected to use 10 to ensure this property.

Further work could be conducted into optimising the parameters of DBSCAN used, such as cross-validation against other catalogues. In this study, the groupings from DBSCAN are used to ensure no highly-probable new members with updated EDR3 parameters have been missed by previous catalogues and that any candidate members flagged by previous catalogues do not appear spurious. It is immediately noticeable from Figure 6.1 that several candidate members from previous catalogues appear to have large calculated distances, which are not flagged as cluster members by DBSCAN. I do not remove these objects at this stage, but further considerations are taken for these objects when assessing the rotational properties of the cluster.

### 6.1.2 NGTS Observations and membership

NGTS observed a $\sim 9$ deg$^2$ region around NGC 6633 between 21/03/2019 and 24/07/2019, taking 134,597 images centred on right ascension 18:27 and declination +06.36 deg. Due to the COVID-19 pandemic, these images were not processed until August 2021. The standard NGTS photometric pipeline was run on the field, resulting in 11,335 candidate light curves. Membership data from all the catalogues listed in Section 6.1.1 provided 1,340 candidate cluster members, of which NGTS was able to produce photometric light curves for 1,042. I searched for periodic signals within these 1,042 candidate cluster members. 342 candidate members with NGTS light curves were cross-matched with catalogues using DR2 and EDR3 data. I found 235 matches from Gaia Collaboration et al. (2018b), 41 from Zari et al. (2018), 184 from Cantat-Gaudin et al. (2018), 319 from Kounkel & Covey (2019), 117 from Cantat-Gaudin et al.

[a] See Section 2.3.2.2 for details on DBSCAN parameters

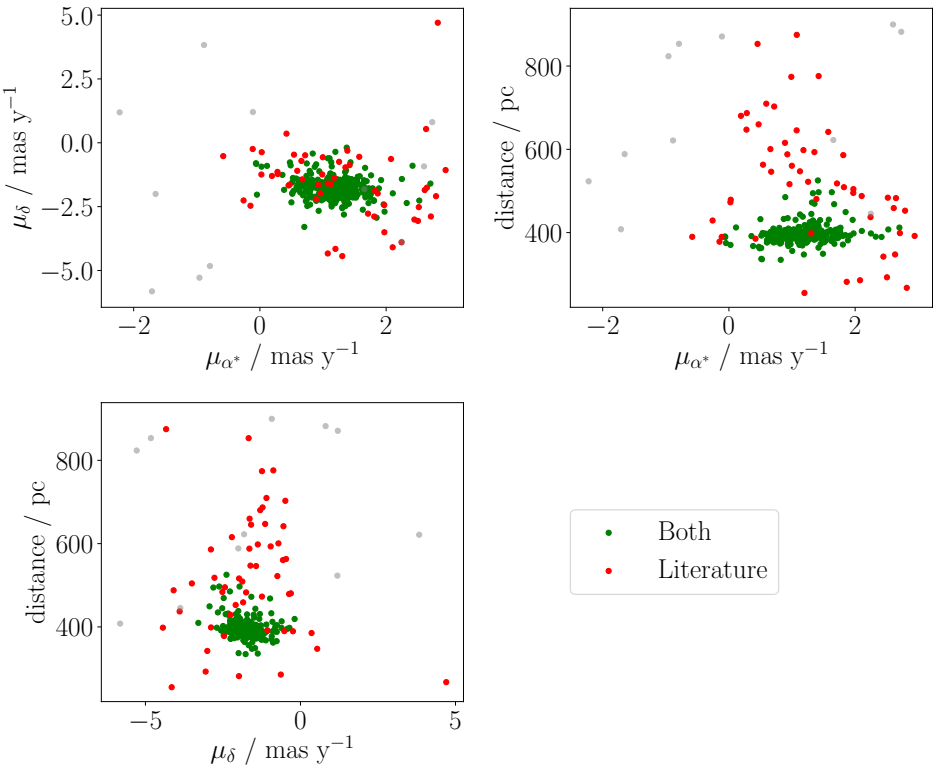



Figure 6.1: Cluster membership for a subset of 342 candidate NGC 6633 cluster members is shown in three dimensions (distance, proper motion in right ascension ($\mu_{\alpha^*}$) and in declination ($\mu_\delta$), coloured by the source of the cluster membership prediction. Red points have been identified as candidate cluster members in previous literature catalogues (52 points). Green points have been identified as candidate cluster members by both DBSCAN and found in previous literature catalogues (281 points). Gray points indicate objects identified as NGC 6633 members by literature catalogues prior to Gaia DR2 which are not contained in later catalogues or predicted to be cluster members by DBSCAN.

(2020), and 45 from Pang et al. (2021). Objects were often contained in multiple lists, so the catalogue cross-match numbers do not sum to 342. The 342 objects had an observed Gaia $G$ magnitude ranging from 17.9 to 9.8 mag.

Before searching for periodic signals, the data were binned into 20-minute bins to allow faster computation. This binning will prevent any signals shorter than around 40 minutes from being detected, but based on the expected rotation periods for main-sequence cluster members; I do not expect this to limit our detections (for example Gruner & Barnes 2020).

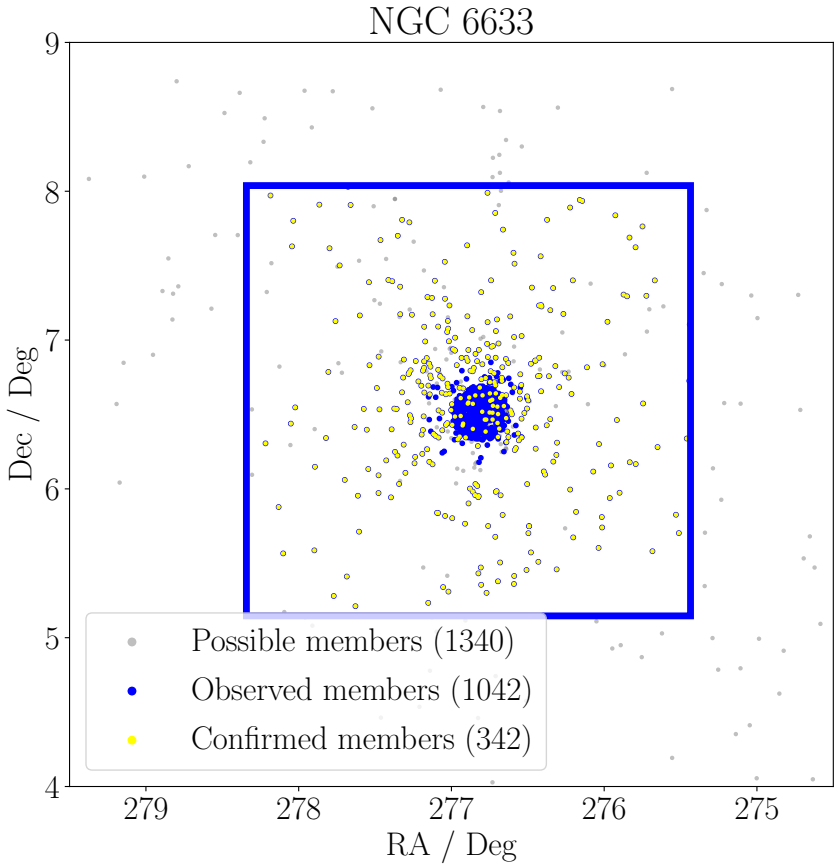



Figure 6.2: The positions of candidate members of the open cluster NGC 6633. Grey points indicate candidate members from all cluster membership lists available. The blue rectangle indicates the NGTS field of view around the cluster. Blue points are stars with available NGTS photometry, and yellow points are candidate cluster members from cluster membership lists that use Gaia DR2 or EDR3 data, which have been observed by NGTS. The yellow points are taken as the cluster sample in this work.

### 6.1.3 Gaia

The third Gaia data release is split into two instalments, with the early data release (EDR3, Gaia Collaboration et al. 2021) at the end of 2020 and the full Gaia DR3 planned for the first half of 2022[b]. EDR3 contains five-parameter astrometric solutions for around 1.468 billion sources, boasting improved precision over DR2 with an overall reduction in systematic noise. Due to its increased precision, the EDR3 catalogue can resolve previously unresolved binary star systems or faint background objects which appeared as one source in DR2. When cross-matching with this catalogue, it is essential to ensure the EDR3 source matches with the DR2 source. The Gaia catalogue enables this to be done easily by ensuring the DR2 and EDR3 `source_id` match for objects for which the stellar parameters are not vastly different between the two catalogues. In cases where the single DR2 source became multiple EDR3 sources with a poor match on all EDR3 sources, I removed this object from the catalogue. Figure 6.2 shows the position of candidate NGC 6633 members in RA and Dec, as well as the extent of the NGTS observation field of view. The 342 candidate members taken from Gaia DR2 and EDR3 confirmed sources are highlighted.

Despite the increased precision of the Gaia EDR3 photometry, Riello et al. (2021) suggests setting limits on a calculated quality metric (the $BP$ and $RP$ flux excess factor, $C$) when using Gaia EDR3 data. This is defined as a simple ratio between the total flux in $G_{\rm BP}$ and $G_{\rm RP}$, and the $G$-band flux: $C = (I_{\rm BP} + I_{\rm RP})/I_{\rm G}$. Where an object has considerably more flux in the $G_{\rm BP}$ and $G_{\rm RP}$ bands than the $G$ band, this is indicative of problems in the $G_{\rm BP}$ and/or $G_{\rm RP}$ photometry. Due to the design of the Gaia telescope, the $BP$ and $RP$ flux is measured using a wider photometric aperture than the $G$ flux, which is much more susceptible to contamination from nearby sources or an unusually bright sky background than the $G$ flux (Gaia Collaboration et al. 2018a).

I followed the formulation outlined in Sections 6 and 9.4 of Riello et al. (2021) to flag potentially problematic sources. Firstly, I calculated a colour-corrected $BP$ and $RP$ flux excess value ($C^*$) using the polynomial relation and Table 2 of coefficients defined in Riello et al. (2021) for each candidate NGC 6633 member to remove the dependence of colour on the flux excess. The $C^*$ values for all 342 candidate stars are plotted against Gaia $G$ magnitude in Figure 6.3. Secondly, I considered a magnitude dependent threshold in $C^*$ defined in Section 9.4 of that work:

$$\sigma_{C^*}(G) = c_0 + c_1 G^m, \tag{6.1}$$

with $c_0 = 0.0059898$, $c_1 = 8.817481 \times 10^{12}$, and $m = 7.618399$. This relation was considered to represent the 1 $\sigma$ scatter for a sample of well-behaved isolated stellar sources with good

[b] https://www.cosmos.esa.int/web/gaia/release. Accessed: 22/01/2022.

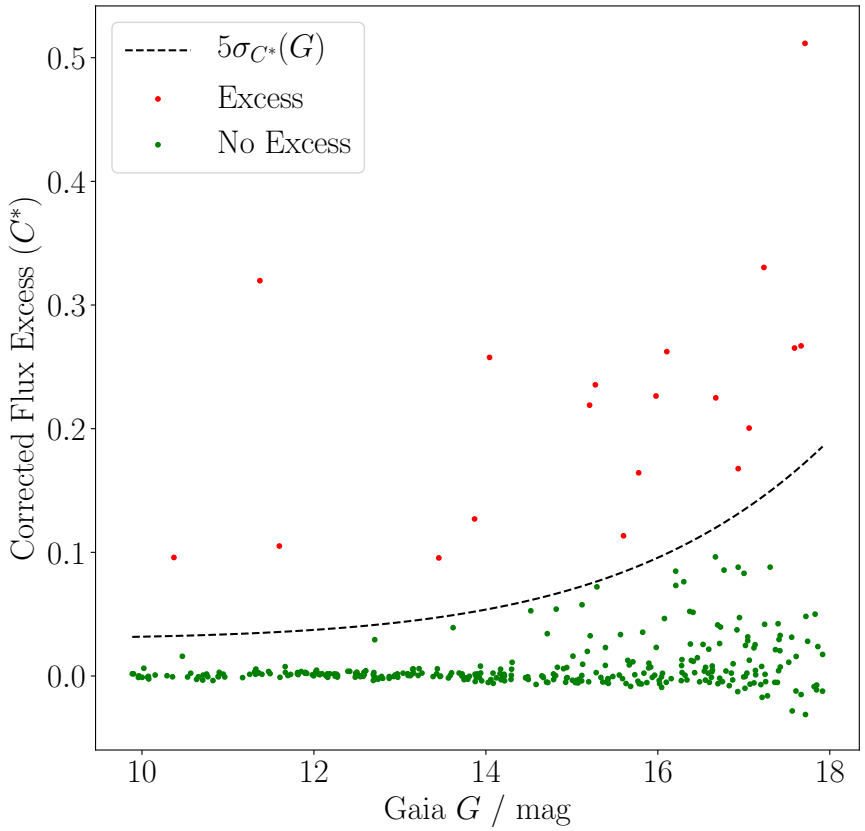



Figure 6.3: Gaia corrected colour excess ($C^*$) vs $G$ magnitude for 342 candidate members of NGC 6633. The dotted line indicates a 5 $\sigma$ colour dependant limit on flux excess, 19 objects above this threshold (red points) are flagged as having potentially bad Gaia photometry, the remaining points are shown in green.

quality Gaia photometry (Riello et al. 2021). I used their conservative limit of 5 $\sigma$ from their good photometry sample, which flagged 19 of 342 objects as having potentially spurious Gaia photometry. The 5 $\sigma$ threshold is plotted in Figure 6.3. The 342 candidate members of NGC 6633 all lay well below the $C^*$ cutoff of 5.0 in the Gaia catalogue; the highest value of $C^*$ in my sample was $\sim 0.5$. I did not remove objects which were flagged as having potentially bad Gaia photometry, but I consider this flag when assessing cluster properties.

## 6.2 Methods

### 6.2.1 Distance calculations

Distances to each star are required to calculate absolute magnitudes and interstellar extinction. As the distance to NGC 6633 is approximately around 400 pc, a simple inverse parallax will not be appropriate as a distance estimate. Furthermore, catalogues such as Bailer-Jones et al. (2018) and Bailer-Jones et al. (2021) use a galactic stellar density prior with assumptions that are statistically sensible for the entire galactic stellar population but may not hold for a cluster population at a specific pointing (Meingast et al. 2021). I elected to use a simple exponentially decreasing density prior and the EDR3 parallaxes for each star. This density prior has been demonstrated to work well for data sets when we have very little information other than parallax and want to make minimal assumptions (Bailer-Jones 2015). To generate the posterior distribution, the authors first assume that the parallax $\varpi$ is normally distributed with an unknown mean $1/r$ and known standard deviation $\sigma_\varpi$. Secondly, they assume that the volume density of stars is exponentially decreasing, i.e. $P(V) \sim \exp(-r/L)$ for some characteristic length scale $L$. I used the formalism outlined in Section 7 of Bailer-Jones (2015) with the following functional form for the distance posterior $P$:

$$P(r \mid \varpi, \sigma_\varpi) = \begin{cases} \dfrac{r^2 e^{-r/L}}{\sigma_\varpi} \exp\left[-\dfrac{1}{2\sigma_\varpi^2}\left(\varpi - \dfrac{1}{r}\right)^2\right] & \text{if } r > 0 \\ 0 & \text{otherwise} \end{cases}$$

for distance $r$, where $\varpi$ is the parallax, $\sigma_\varpi$ is the error on the parallax, and $L > 0$ is a characteristic length scale, here taken to be 1000 pc as in Bailer-Jones (2015). For distances $r \ll L$ this corresponds to a constant space density of stars; the approximate distance to NGC 6633 is 400 pc, so most of the stars fitted will fall into this regime. I found the modal distance value for this posterior given a Gaia EDR3 parallax and error value by setting $dP/dr = 0$ and numerically solving for the roots. The error on each distance measurement was calculated as ± half the FWHM spread in the distance posterior.

I calculated an estimated median distance to NGC 6633, considering only the 281 candidate members confirmed by previous catalogues and the DBSCAN clustering. This removes potential distance outliers as seen in Figure 6.1. The calculated median distance and 16$^{\text{th}}$ to 84$^{\text{th}}$ percentile spread was $(394 \pm 13)$ pc, which agrees with the EDR3 estimate from Pang et al. (2021) (394.2 ± 2.4 pc, derived from 300 candidate members).

### 6.2.2 Extinction correction

It has previously been noted by Pang et al. (2021) that NGC 6633 has a significant reddening coefficient ($E(B-V) = 0.18$ mag) which would correspond to an expected extinction of 0.558 mag in the V-band using $R_V = 3.1$ (Cardelli et al. 1989). I calculated differential extinction values for each star within the cluster, using the precise positions and distances calculated from Gaia EDR3. I referenced a 3-dimensional dust map of the Galaxy to calculate the line-of-sight interstellar extinction for each object. I then converted this extinction into the three Gaia EDR3 bandpasses: $G$, $G_{BP}$ and $G_{RP}$. Although Gaia provides extinction values within DR2, these extinction values are often unreliable as they only use data from the Gaia bandpasses, train on synthetic photometry, and use inverse parallax as a distance proxy without an informed prior[c].

I used the 3d dustmaps from Green et al. (2019), hereafter referred to as 'Bayestar19'. The latest iteration of this dustmap uses parallaxes from Gaia in conjunction with stellar photometry from Pan-STARRS and 2MASS to produce detailed 3d dustmaps of the Galaxy with reported reddening uncertainties approximately 30% smaller than those reported in the Gaia DR2 catalogue. Bayestar19 can be queried using the `Python` package `dustmaps`[d] for any right ascension, declination and distance north of a declination of -30°.

The units of Bayestar19 extinction differ slightly from standard units such as $E(B-V)$ colour excess. Green et al. (2019) provides empirical conversions into standard reddening units; I used the conversion $E(B-V) = 0.884\times$ (Bayestar19). To calculate the effect of extinction, I converted into the V band using $A_V = R_V \times E(B-V)$ with $R_V = 3.1$ (Cardelli et al. 1989) and then into the three Gaia EDR3 passbands using the following relations:

$$A_G = 0.87 A_V, \tag{6.2}$$

$$A_{BP} = 1.10 A_V, \tag{6.3}$$

$$A_{RP} = 0.636 A_V. \tag{6.4}$$

The coefficients were taken from the SVO Filter Profile Service[e] (Rodrigo & Solano 2020), which provides the ratio between the V-band extinction and the extinction in a large number of survey filters, which in turn are calculated using the extinction law from Fitzpatrick (1999).

For the 281 members of NGC 6633 confirmed by both literature sources and DBSCAN, the median extinction and 16$^{th}$ to 84$^{th}$ percentile spread in the three EDR3 bands were calculated as $A_G = (0.41 \pm 0.08)$ mag, $A_{BP} = (0.45 \pm 0.09)$ mag and $A_{RP} = (0.26 \pm 0.05)$ mag. The EDR3 G-band median extinction value is slightly lower than the value estimated by Pang et al. (2021)

[c] A full explanation of the Gaia DR2 catalogue extinction value calculations is given here: https://gea.esac.esa.int/archive/documentation/GDR2/Data_analysis/chap_cu8par/sec_cu8par_process/ssec_cu8par_process_priamextinction.html. Accessed: 22/01/2022.

[d] http://argonaut.skymaps.info/usage. Accessed: 22/01/2022.

[e] https://svo.cab.inta-csic.es/. Accessed: 22/01/2022.

(converted to the G band using the relation in Equation 6.2), however, agrees within 1 standard deviation. I calculated a fairly wide spread of differential extinction values for candidate cluster members of $\sim 0.2 < A_{\rm G} < 1.0$.

### 6.2.3 $B - V$ to $G_{BP} - G_{RP}$ conversion

In Chapter 5, I convert between passbands using the relations defined in the 'Modern Mean Dwarf Stellar Colour and Effective Temperature Sequence' (Pecaut & Mamajek 2013). It is possible to use more accurate passband conversions within open clusters with a tightly defined colour–magnitude main sequence. The Pecaut & Mamajek (2013) relations are defined only on a relatively sparse grid in colour and not specifically calibrated to the Gaia passbands, so interpolation would be necessary. Gruner & Barnes (2020) define empirical, formulaic colour transformations between $G_{BP} - G_{RP}$ and $B - V$. These transforms are fitted to photometric data for the Hyades, the Pleiades, Ruprecht 147 and a selection of red stars taken from Pecaut & Mamajek (2013), which have colour information in both $B - V$ and Gaia DR2 $G_{BP} - G_{RP}$. I refer the reader to Appendix A of Gruner & Barnes (2020) for full details of the transforms, for which both forward and inverse transformations are given. For the cluster data used in this study, I convert from $G_{BP} - G_{RP}$ to $B - V$ and vice-versa using the Gruner & Barnes (2020) transforms.

### 6.2.4 Identifying single stars

I identified potential photometric binary and higher-order systems by assessing their position on the colour–magnitude diagram (CMD). I fitted a single-star cluster sequence and flagged stars lying above this trend as potential multiple-star systems. The CMD I used compared Gaia $M_{\rm G}$ (absolute magnitude) versus $G_{BP} - G_{RP}$ colour, using extinction corrected EDR3 data. This will identify any near-equal-mass binary systems; I note that lower mass-ratio binary systems may not be flagged using this method, and using multiple CMDs in different bands would help to identify lower mass-ratio binary systems with different spectral types (e.g., as in Gillen et al. 2020). I elected to fit a cluster sequence over using an isochrone fit as this provided a tighter fit to the entirety of the cluster sequence in comparison to the PARSEC isochrones, which struggled to fit the exact shape of the cluster sequence, particularly at roughly $0.6 < G_{BP} - G_{RP} < 1.4$. Figure 6.4 shows both the line of best fit for the cluster main-sequence and a PARSEC isochrone of similar age and metallicity to NGC 6633.

I use a fifth-order spline fit in $M_{\rm G}$ versus $G_{BP} - G_{RP}$ for all DR2/EDR3 identified cluster members as this was by eye determined to be the best order spline fit across the entire $G_{BP} - G_{RP}$ range. The spline is fitted to the entire data set, and an iterative removal and re-fitting procedure

is implemented. I calculate residuals of the spline fit for all points and remove any data points which lie greater than three standard deviations from the spline fit. I then re-fit the same fifth-order spline to the remaining points. This procedure is repeated until only one point is removed from the fit. The natural stopping criterion would be when zero points are removed; in practice, the final removal steps just removed points one by one, so a stopping criterion of one point was deemed sensible. The final spline fit was fitted to $M_{\rm G}$ and $G_{BP} - G_{RP}$ values for 14 stars of the 342 initially used for NGC 6633.

A binary cutoff was defined as 0.375 mag above this line: an equal-mass binary system will lie 0.75 mag above the main sequence, so 0.375 mag halves this distance. I consider any points above this line candidate binary or higher-order systems. The spline fit appears to turn off the main sequence at $G_{BP} - G_{RP} \sim 2.2$, and based on the PARSEC isochrone, the main-sequence turnoff for this cluster occurs around $G_{BP} - G_{RP} \sim 0.2$. There are very few objects in the sample outside of this range $0.2 < G_{BP} - G_{RP} < 2.2$; 1 object was flagged as a potential binary below $G_{BP} - G_{RP} \sim 0.2$ and 11 were flagged above $G_{BP} - G_{RP} \sim 2.2$. These objects were flagged as potential binaries in my sample, and extra care was taken when analysing these stars. In practice, this made no difference to the rotational analysis, as none of these objects had detected periodic variability. This cut-off flagged 78 objects as potential binary or higher-order systems and 65 objects within the colour range $0.2 < G_{BP} - G_{RP} < 2.2$. I note that I do not include the errors on the calculated absolute magnitudes when flagging potential binary systems, although I expect this to affect a small number of stars with imprecise distance estimates.

### 6.2.5 The `RoTo` package

I use three methods to determine rotation periods from the NGTS light curves. We have already been introduced to these three methods throughout this work: The G-ACF (Kreutzer et al. *submitted*), a Lomb–Scargle Periodogram (Lomb 1976; Scargle 1982) and a Gaussian process model. To streamline the process of period retrieval and confirmation, I developed a software package, `RoTo`, to allow myself and other users to quickly and repeatably detect periodic variability from photometric light curves. The package provides tools for period determination with user-defined parameters for all models and plotting tools to confirm periodic variability signals. As well as the three methods used to analyse this NGTS data set, `RoTo` will provide additional period finding algorithms such as Fourier Transforms and Phase Dispersion Minimisation (not currently implemented, Stellingwerf 1978), which provides greater freedom to the user in finding and using a suitable method for their data set.

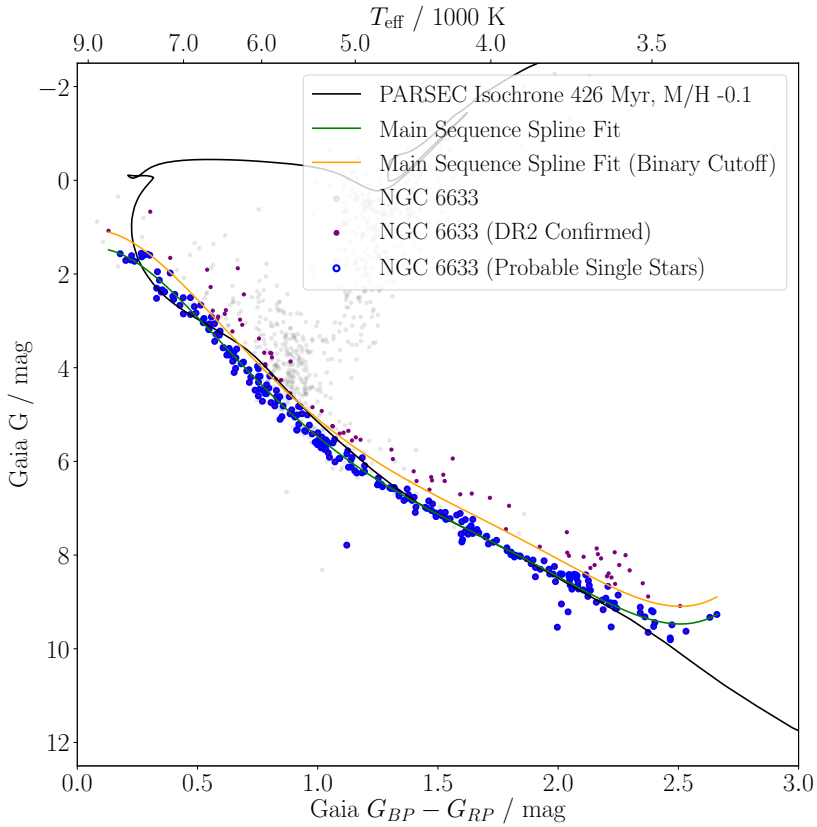



Figure 6.4: Gaia EDR3 $M_{\rm G}$ (absolute magnitude) plotted against Gaia $G_{BP} - G_{RP}$ for NGC 6633 objects. Light grey points indicate all possible cluster members. Purple points indicate cluster members confirmed using DR2 or EDR3 data. Objects not flagged as possible binaries based on the main sequence fit are circled in blue. The black line is an isochrone generated using PARSEC v1.2 in EDR3 passbands of age 426 Myr and [M/H] = −0.1. A fifth-order iterative spline fit is plotted in green, with a main sequence binary cutoff plotted in orange, which lies 0.375 mag above the green line.

#### 6.2.5.1 Lomb–Scargle periodogram

I use the `astropy` Lomb–Scargle implementation (Robitaille et al. 2013) to calculate a Lomb–Scargle periodogram for each light curve. To provide an error estimate on this value and confirm the stability of any variability signal within the light curve, I implement a 'sliding Lomb–Scargle periodogram'. A window of 5 times the initial period estimate (i.e. highest peak in the Lomb Scargle periodogram of the entire light curve) is stepped along the light curve. A Lomb–Scargle periodogram is calculated for each windowed data set; the largest peak for each periodogram is selected and used as the variability period within this window. The variability periods across all windows are aggregated, and a representative period is selected. `RoTo` provides flexibility on the window size, number of windows and aggregation function. The default window size is five times the initial period estimate, with a maximum of 100 windows allowed before this window is increased in size. By default, the representative period is selected as the median across windows, with an error given as the $16^{\text{th}}$ to $84^{\text{th}}$ percentile spread (i.e. 1 $\sigma$). Users may also use mean and standard deviation or the modal period and a percentile spread.

#### 6.2.5.2 G-ACF

I use the same parameters for the G-ACF as in Chapter 5. I calculate the positive lag values only, using a lag resolution of the minimum time difference between data points. I use a natural selection function (Section 4.3.3), and rational weight function (Equation 4.4) with the scaling parameter $\alpha$ taken as the median time value of the time series.

To extract a period from the G-ACF, `RoTo` provides two methods: an FFT period or a G-ACF peak-based period. An FFT (Cooley & Tukey 1965) of the G-ACF is calculated, and the largest FFT peak is taken as the period. This does not provide an error estimate on the peak. The latter, peak based method, is similar to the method employed by McQuillan et al. (2013). The G-ACF is convolved with a 1-D Gaussian Kernel of FWHM 18 lag points over a window of 56 lag points. These values are taken directly from McQuillan et al. (2013), which provides a good compromise between noise reduction and ensuring a strong ACF signal without prior knowledge of the period. This smooths the G-ACF signal, and the smoothed G-ACF is then assessed for peaks. If there is just one peak in the smoothed G-ACF, the period and error are taken as the centre point and FWHM of the single smoothed peak. Where multiple peaks are found (up to a maximum of 10), the period is taken as the median of the gap between peaks. The period uncertainty is calculated using the scatter of these values, calculated as

$$\sigma_{\text{P}} = \frac{1.483 \times \text{MAD}}{\sqrt{N-1}} \tag{6.5}$$

where $N$ is the number of peaks and MAD is the median absolute deviation from the median

period value. The MAD is used as a proxy for standard deviation but is more robust to outliers than the standard deviation of a Gaussian.

In practice, the FFT often provided more robust period estimates than the peak finding method, and so the FFT method was adopted as the default for `RoTo` at this stage. Further investigation of the peak finding method should be conducted to assess its shortcomings on this dataset.

#### 6.2.5.3 Gaussian process regression

I adapt the Gaussian process fitting for period extraction as used in Gillen et al. (2020), which utilised a quasiperiodic kernel composed of a sum of SHO terms. The details of the kernel used as a part of `RoTo` (which is slightly different from the kernel used in Gillen et al. 2020) were given in Section 2.10. The default parameters for the SHO kernel terms are fixed, taken from the `EXOPLANET` documentation[f], however, can be specified as user-defined parameters within `RoTo`. The sampler is initialised with a period estimate taken from a Lomb Scargle periodogram, and a *maximum a posteriori* (MAP) fit is performed. Should the MAP fit produce a good solution, $3\sigma$ outliers from this model are removed from the data and, the model is re-fitted. This masking will remove non-rotational variability phenomena such as flares and deep eclipses. A full model fit is performed using the 'No U-Turn Sampler' (Hoffman & Gelman 2014) MCMC sampling. I used seven independent Markov chains for this work, each running on a separate CPU core with 500 tuning steps followed by 2000 production steps. This typically took around 20 minutes to run for NGTS light curves using the Cambridge HPC cluster.

The GP model is the most computationally expensive period determination method within `RoTo` and the total run-time of each light curve with `RoTo` was determined almost entirely by the speed of the GP modelling. When running on the HPC cluster, it was necessary to implement a timeout for the GP MAP fit and MCMC posterior sampling of 40 minutes. In cases where the GP model was difficult to fit (for example, light curves with rapidly evolving signals or data with poor error estimates), I re-ran the processing with a longer timeout. For a few light curves, `RoTo` was still unable to compute a GP model solution within a feasible time, and in these cases, just the G-ACF and LS periods were considered.

#### 6.2.5.4 Combining period estimates

In addition to the period detection methods described above, `RoTo` provides methods for outputting a 'best' period for a photometric light curve. `RoTo` will run a user-specified set of period detection methods using user-specified parameters and will return a 'best' period using

[f] https://gallery.exoplanet.codes/tutorials/stellar-variability/. Accessed: 22/01/2022.

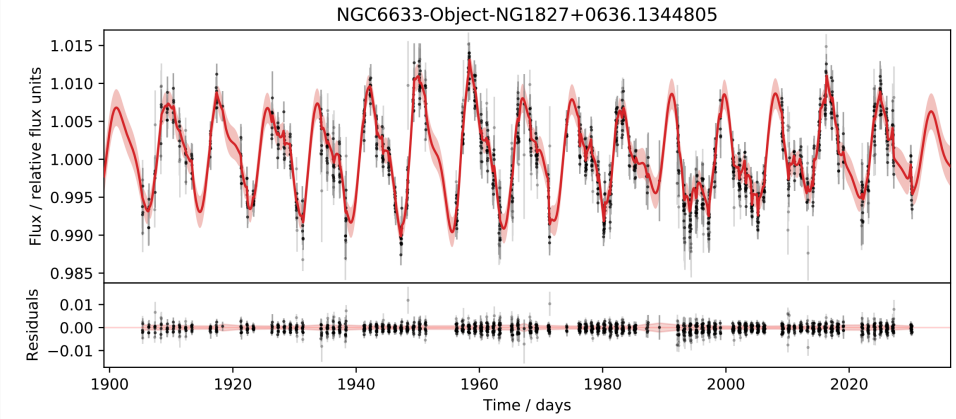



Figure 6.5: Example RoTo data plot. The light curve is plotted as black points with error bars (top). A GP model fit is overlaid in red, with 1 $\sigma$ uncertainty intervals in light red. The residuals of the GP model fit are plotted in black (bottom), with the uncertainty of the GP model overlaid in light red.

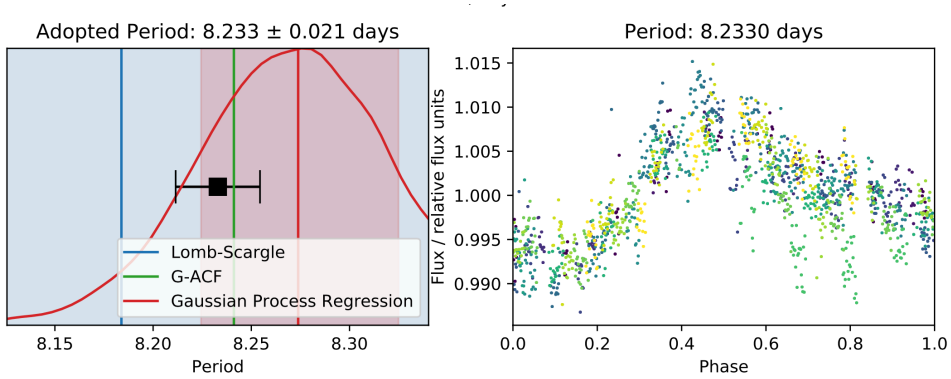



Figure 6.6: Example RoTo combined period estimate plot. Outputs from three period estimation methods are plotted (left). Vertical blue and green lines show the period estimate from an LS periodogram and a G-ACF, respectively, with error bars plotted as the same colour shading. The GP posterior is shown as a red line, with the mean (vertical red line) and 1 $\sigma$ uncertainty (light red shading). The combined period and uncertainty are plotted as a black point with an error bar. The right-hand plot shows the light curve phase folded on this combined period.

a user-specified aggregation method. The aggregation method can return the period estimate from one method if one performs best; otherwise, I calculate a median and MAD or a mean and standard deviation. In practice, this combined rotation period estimate may be less useful than the individual period estimates from all methods.

#### 6.2.5.5 Plotting tools

RoTo can generate plots that show the estimated periods from each method, as well as diagnostic plots and phase folds for each method, to enable validation of the estimated periods. I split an

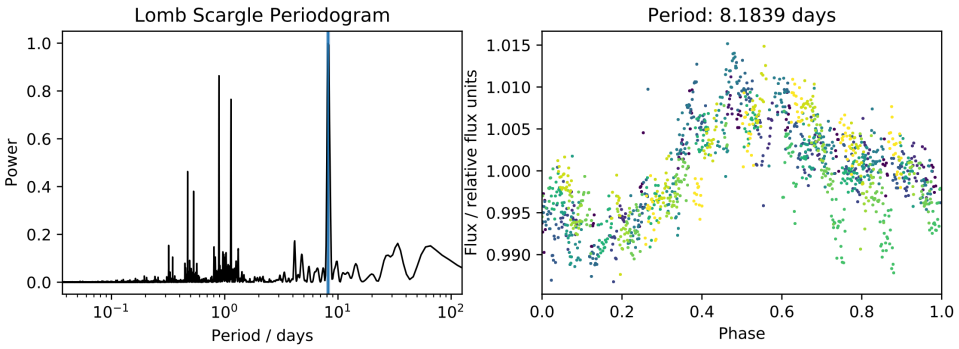



Figure 6.7: Example `RoTo` method plot. The Lomb–Scargle periodogram of the entire light curve is shown (right). The estimated period is plotted as a blue line with errors plotted as a light blue region. In this example, the errors on the estimated period are extremely small, and hence may not be visible. The left plot shows the light curve phase folded on this LS estimated period.

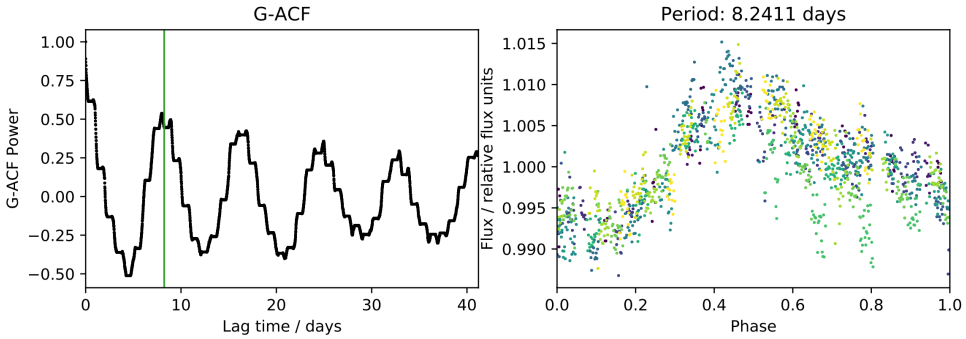



Figure 6.8: Example `RoTo` method plot. The G-ACF of the light curve is plotted (right). The estimated period is shown as a green line with errors plotted as a light green region. In this example there are no error estimates on the period, and hence not visible. The left plot shows the light curve phase folded on this G-ACF estimated period.

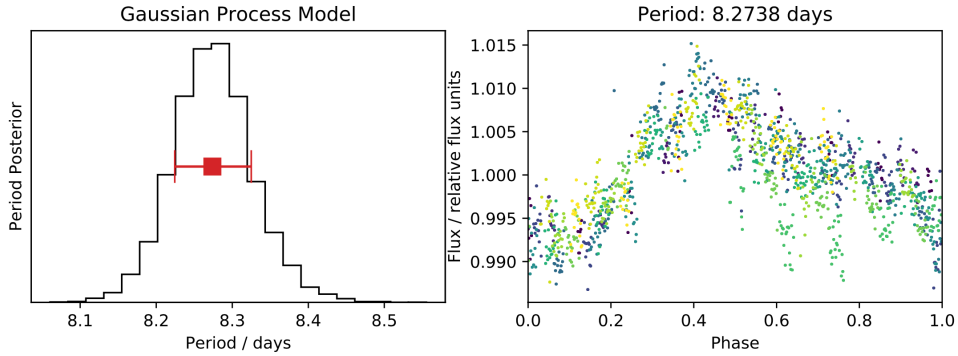



Figure 6.9: Example `RoTo` method plot. The GP model period posterior is plotted as a black histogram, with the estimated period and 1 $\sigma$ uncertainty shown as a red point with error bars (right). The left plot shows the light curve phase folded on this GP estimated period.

example of a `RoTo` generated `.pdf` into five plots (Figures 6.5 – 6.9), and explain each plot in turn. Examples of the full `.pdf` are shown for NGC 6633 in Appendix A.

At the top of a `RoTo` output `.pdf` (Figure 6.5), the entire light curve is plotted as black scatter points, with errors on each point. Overlaid is the MAP GP model fit in red, with 1 $\sigma$ uncertainty intervals in light red. Below this, the residuals of this model fit are plotted in black. 1 $\sigma$ uncertainty intervals for the GP model are plotted in light red.

The second row of the `.pdf` output is shown in Figure 6.6. The left plot shows the results of the individual `RoTo` methods overlaid with error bars. In the case of the GP model (red), the period posterior distribution is plotted, along with the mean (vertical red line) and 1 $\sigma$ uncertainty (light red). For the G-ACF (green) and LS (blue), the estimated period is shown as a vertical line, with uncertainty as light shading of the same colour. In this example, the uncertainty on the LS period extends beyond the x-axis. The combined period and uncertainty are shown as a black point with error bars. The right plot shows the light curve phase folded on this combined period.

Figures 6.7, 6.8 and 6.9 include the remainder of the `RoTo` output. The left-hand plots show details of the method, which can be a periodogram, an ACF or a period posterior distribution. The right-hand plot in each example will show the light curve phase folded on the period found by that method.

`RoTo` will dynamically generate the plot based on the methods specified by the user and the data available for each method. For example, if `RoTo` has produced a MAP fit with no MCMC solution, the package will plot a MAP model, but the plot will show no period posterior distribution. Additionally, `RoTo` can plot detailed diagnostic plots for the GP, including a corner plot for all parameter distributions and associated trace plots for the MCMC chains.

### 6.2.6 Rotational analysis pipeline

For each of the NGTS objects identified as a candidate cluster member by any of the catalogues mentioned in Section 6.1.1, I ran `RoTo` to determine any periodic variability signals present in the light curve. I calculated a period from a G-ACF, an LS periodogram and a GP regression model, using the parameters and settings defined in Sections 6.2.5.1, 6.2.5.2 and 6.2.5.3. I generated an array of jobs to run on the Cambridge HPC system, each using seven cores to evaluate seven MCMC samplers simultaneously. Each run produced a `.csv` file with the estimated rotation period and errors for each method successfully run, plus the combined period estimate. A `.pdf` of the outputs, plus any diagnostic plots for the MCMC fit where applicable, were also generated for each run. I combined the `.csv` files, and I joined the rotation period data with the cross-match data from Gaia.

The validation of the rotation periods was done manually: each object was assessed by eye using the generated `.pdf` file showing phase folded data, model fits and periodograms. Identifying many systematically incorrect period estimates from 1-day aliasing and Moon correlated background noise was straightforward. I elected not to run the period validation pipeline developed in Chapter 5 due to having fewer objects and the increased complexity in the period extraction stages. Where the signal was not systematic, I assessed if the phase fold of the light curve appeared to be a valid variability signal. Running multiple period extraction methods also allowed cross-validation of the estimated periods; if multiple methods agree on the period and it is not an obvious systematic, it is likely a real signal in the data. This also had the advantage of highlighting spurious detections if just one method detects a signal and other methods detect common systematics.

Gillen et al. (2020) applied additional data cleaning to the NGTS light curves by using a Savitzky–Golay (SG) filter followed by a convolution to remove any longer-term trends, namely signals arising from the Moon. I did not apply this step in this analysis. However, manual inspection of the `RoTo` data products showed a large number of variability signals that could be attributed to Moon correlated signals. It is unclear how many more variable objects would have been detected due to this processing step; in Chapter 5, I found approximately 68% of periodic variability signal detections (21% of all light curves analysed) from NGTS light curves were dominated by Moon correlated noise; however no attempt at correction of these signals was made using the simple three-parameter model.

## 6.3 Results and discussion

I generated variability periods, extinctions and distances for 58 NGC 6633 objects in total, of which 11 were flagged as possible photometric binary systems based on CMD position. Table 6.2 outlines the format of the provided data tables; the full tables and object plots are given in Appendix A. Of the 342 candidate cluster members analysed, 214 stars lie within the colour range $0.47 < G_{BP} - G_{RP} < 1.49$ for which periodic signals were detected. The detection efficiency for this sample of stars within NGC 6633 was 27%.

### 6.3.1 Individual light curves

All 58 manually vetted `RoTo` outputs are shown in Appendix A and follow the same format described in Section 6.2.5.5.

Of the 58 objects assessed to be variable, 11 were flagged as potential binary objects based on CMD position. Manual inspection of these objects highlights examples of clear eclipsing binary signals, for example, `NG1827+0636.1025611` for which the G-ACF finds a 3.66 day

Table 6.2: Table format for the final data product of this study.

| Column | Format | Units | Label | Description |
|---|---|---|---|---|
| 1 | A19 | — | NGTS_ID | NGTS source designation |
| 2 | F9.5 | deg | RA | Gaia EDR3 Source right ascension (J2000) |
| 3 | F7.5 | deg | DEC | Gaia EDR3 Source declination (J2000) |
| 4 | F8.5 | mag | G_MAG | Gaia EDR3 G-band magnitude |
| 5 | A8 | — | METHOD | Period extraction method |
| 6 | F8.5 | days | PROT | Extracted variability period |
| 7 | F7.5 | days | PROT_ERR_N | Extracted variability period negative error |
| 8 | F7.5 | days | PROT_ERR_P | Extracted variability period positive error |
| 9 | F8.5 | days | PROT_LS | LS Extracted variability period |
| 10 | F7.5 | days | PROT_LS_ERR_N | LS Extracted variability period negative error |
| 11 | F7.5 | days | PROT_LS_ERR_P | LS Extracted variability period positive error |
| 12 | F8.5 | days | PROT_GACF | GACF Extracted variability period |
| 13 | F7.5 | days | PROT_GACF_ERR_N | GACF Extracted variability period negative error |
| 14 | F7.5 | days | PROT_GACF_ERR_P | GACF Extracted variability period positive error |
| 15 | F8.5 | days | PROT_GP | GP Extracted variability period |
| 16 | F7.5 | days | PROT_GP_ERR_N | GP Extracted variability period negative error |
| 17 | F7.5 | days | PROT_GP_ERR_P | GP Extracted variability period positive error |
| 18 | F7.5 | — | AMPLITUDE | 5–95 percentile relative flux |
| 19 | F9.5 | parsec | DISTANCE | Estimated distance to source |
| 20 | F7.5 | mag | A_0 | Estimated V-band Extinction |
| 21 | I1 | — | BINARY | Possible Photometric Binary Flag |
| 22 | I19 | — | GAIA_DR2_ID | Cross-matched Gaia DR2 identifier |
| 23 | I19 | — | GAIA_DR3_ID | Cross-matched Gaia EDR3 identifier |
| 24 | I10 | — | TIC_ID | Cross-matched Tess Input Catalogue (v8) identifier |
| 25 | A16 | — | TWOMASS_ID | Cross-matched 2MASS identifier |
| 26 | A19 | — | WISE_ID | Cross-matched WISE identifier |
| 27 | A10 | — | UCAC4_ID | Cross-matched UCAC4 identifier |

period which appears to be a rotation signal, but additionally, we see clear flux drops which are indicative of eclipses in an EA (Algol-type) system. Some light curves did not show flux drops indicative of an eclipse but displayed signals with multiple rotation period detections within the target aperture. Object `NG1827+0636.1233779` displays clear photometric variability; however, the three period-recovery methods find different periods within the data. In the LS periodogram, we can see a large peak at 1.42 days and approximately 3.5 days, which the G-ACF picked up. This could indicate the presence of a two-star system, in which one has a rotation period of $\sim$ 1.4 days and the other a period of $\sim$ 3.5 days. I note, however, that these signals may be aliases of one another with respect to 1-day sampling. There is a third large peak in the LS periodogram at $\sim$ 0.6 days, and it is possible to relate 0.6, 1.4 and 3.5 days as sampling aliases with respect to 1-day sampling. Other potential binary objects, such as `NG1827+0636.1401169` and `NG1827+0636.1452024` display single variability periods of 3.88 and 4.14 days, respectively, with no indication of eclipses or multiple rotation periods. The signals are fairly sinusoidal, which could indicate that the rotation periods of the stars are synchronised to the orbital period via tidal interactions (Pan 1997). Follow-up of such objects spectroscopically would aid in revealing multiple stars, if present.

For objects where RoTo returned periods that disagreed between methods, manual inspection provided the 'best' variability period. Object NG1827+0636.1010770 was one such example, where the G-ACF and the GP model found a long-term trend within the data of period $\sim$ 120 days, whereas, within the LS periodogram, there was a clear peak at $\sim$ 6.2 days, which when phase folded appears to be a real periodic signal. In this case, the LS period was deemed the correct one and used as the reported rotation period of this object. Object NG1827+0636.1041927 had a well-bounded period posterior sampled by the GP centred at 7.28 days, which did not appear to agree with the LS or G-ACF rotation periods within error. In this case, combining the rotation period estimates would result in a less accurate rotation period for the object than found with the GP, so the GP and associated error was adopted as the correct period. For some binary objects, as described above, and such as object NG1827+0636.1025611, the G-ACF was able to extract a rotation-like periodic signal of $\sim$ 3.66 days, but the GP and LS methods return poor period estimates due to the in-transit flux drops of the eclipsing system. In this case, the G-ACF was reported as the correct signal with a zero error estimate.

### 6.3.2 Global variability

I plot the rotation periods against extinction corrected Gaia $G_{BP} - G_{RP}$ colour to ascertain if there is a tight cluster rotation sequence as expected. Figure 6.10 shows a well-defined sequence, broadly increasing in rotation period for redder stars. To aid the eye, I plot two model gyrochrones from Angus et al. (2019) as defined in Section 5.2.6 at 400 and 600 Myr, hereafter referred to as the 'Angus Model'. I plot two gyrochrone models from Spada & Lanzafame (2020) at 400 and 600 Myr (dotted lines). The literature age of NGC 6633 is roughly 500 Myr, and the rotation periods found for the cluster lie broadly between the 400 and 600 Myr gyrochrones for both models, particular around Solar colour where the models are well-calibrated. The Angus models do not well-describe the rotation periods of stars hotter than around 6250K; these stars will have thin convective layers and weak magnetic dynamos and hence will not converge onto the Skumanich braking law Angus et al. (2019). This behaviour is noticeable as the model fails to accurately model the sharp dip in rotation period observed for the bluest stars in the NGC 6633 sample.

The majority of objects flagged as binary systems are below the well-defined slow-rotator sequence of NGC 6633 in period, which is indicative of potentially tidally locked close-in binary systems (Gillen et al. 2020). I highlight two stars of $G_{BP} - G_{RP}$ colour $\sim$ 1.8 and period $\sim$ 4 days, which have not been flagged as binary systems based on their position on the CMD. Neither of these systems appeared to be a photometric binary upon inspection of the light curves, implying that these systems are either multiple star systems containing very low mass companions or single stars whose angular momentum loss has been reduced during

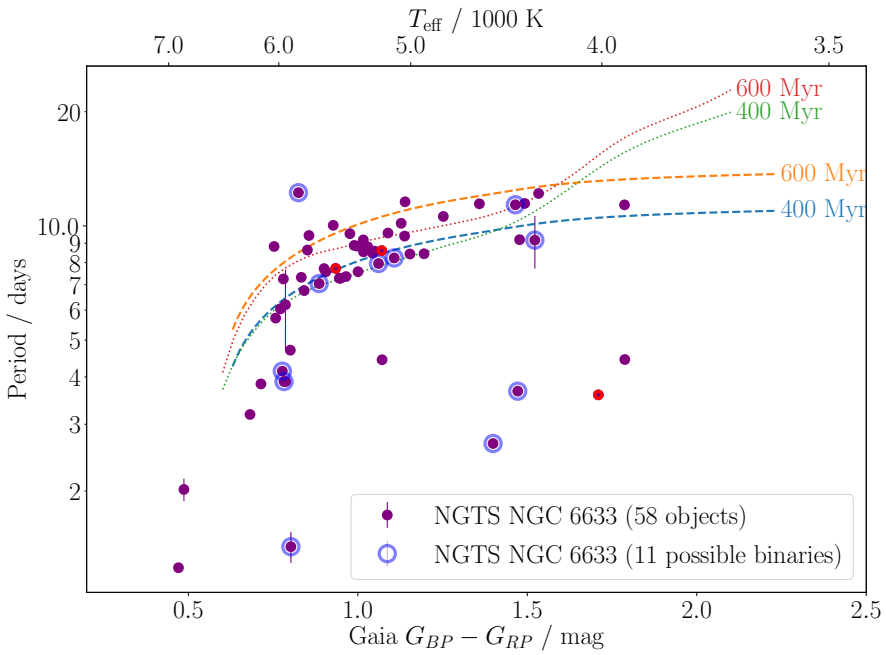



Figure 6.10: Period vs Gaia $G_{BP} - G_{RP}$ colour for objects in NGC 6633. Two gyrochrones using models from Angus et al. (2019) are plotted for 400 and 600 Myr (labelled). Two gyrochrone models from Spada & Lanzafame (2020) are plotted for 400 and 600 Myr (dotted lines). 58 stars in total had rotation periods, of which 11 were flagged as possible photometric binary systems based on CMD position. Three objects (in red) had significant $BP$ and $RP$ excess flux.

evolution. One of these objects, `NG1827+0636.1794649` has been flagged as having significant $BP$ and $RP$ flux excess; this may indicate the colour of the star is incorrect, however, given the magnitude of the colour excess it is unlikely that the true colour would place this object onto the slow-rotator sequence of NGC 6633. Two other objects with significant $BP$ and $RP$ flux excess have been flagged at $G_{BP} - G_{RP} \sim 1.0$. I looked at the nearest neighbour objects within the Gaia EDR3 catalogue and it appears that the closest sources to these two objects were of similar $G_{BP} - G_{RP}$ colours, implying that the contamination of these sources giving rise to the flux excess does not appear to significantly affect their $G_{BP} - G_{RP}$ colours.

One object, `NG1827+0636.1439675`, has a rotation period of $\sim 4.43$ days, and $G_{BP} - G_{RP}$ of $\sim 1.0$. It has not been flagged as a binary system, and manual inspection of the light curve revealed a clear single-peaked periodicity. The calculated distance to this object was $518 \pm 8$ pc, which is well beyond the expected cluster distance of $\sim 400$ pc. This object was not flagged as a cluster member by DBSCAN and additionally was only confirmed as a cluster member by one literature source (Kounkel & Covey 2019). Based on the poor fit to the slow-rotator sequence and the poor agreement to the cluster astrometry, I remove this object from further

processing because it is likely not a member of NGC 6633.

As is visible on the CMD, NGTS observed objects up to $(G_{\rm BP} - G_{\rm RP}) \lesssim 2.0$, and so I am unable to trace the period–colour relation beyond this point. Previous studies of the Hyades and Praesepe demonstrate this relationship holds up to $(G_{\rm BP} - G_{\rm RP}) \lesssim 3.0$ (Douglas et al. 2019), where rotation periods of approximately 20 days are seen in the reddest stars. The large distance and high interstellar reddening of NGC 6633 will increase the telescope power required to observe the cluster, particularly for the already faint redder objects within the cluster. This potentially means that detailed observations of redder stars within NGC 6633 will fall below the precision threshold of NGTS.

### 6.3.3 Spada and Lanzafame (SL20) model

As discussed briefly in Section 5.3, Spada & Lanzafame (2020) developed a set of stellar evolution codes which incorporate the competing effects of wind-braking and interior angular momentum coupling to the rotational evolution of solar-like stars. These models were motivated by observations of the slow-rotator sequence in the open-clusters Praesepe and the ~ 1Gyr NGC 6811 cluster, which was previously not well modelled by stellar evolution codes or empirical gyrochronology models. The authors claim the model well captures the mass dependence of the slow-rotator sequence in the range 0.4–1.3 $M_{\odot}$ for stars between 700 Myr and 1 Gyr. However, the model was also shown to reproduce the slow-rotator sequence of the younger Pleiades cluster satisfactorily. The model provides a more physically motivated gyrochronology relation by considering the internal transport of angular momentum in a two-zone stellar interior. This model is therefore not valid for fully convective stars due to the two-zone nature of the model.

An analytic form of the gyrochronology relations is not feasible; however, the authors provide a grid of gyrochrones in $B - V$ colour for ages ranging from 0.1 to 4.57 Gyr. To allow fitting of the models, I interpolate this provided model grid in two dimensions. Firstly the $B-V$ colour is converted into a $G_{BP} - G_{RP}$ colour using the equations from Gruner & Barnes (2020) (as described in Section 6.2.3), this provides a valid range of $0.57 < G_{BP} - G_{RP} < 2.1$ For this work, the interpolation will not allow model periods to be calculated for colours outside of the colour range defined by the provided model grids. The interpolation is conducted using `scipy.interp2d` (Virtanen et al. 2020), in which I interpolate in both $G_{BP} - G_{RP}$ and age using a 2-dimensional cubic spline. This interpolation was deemed appropriate: for the range of interest for NGC 6633 (400–600 Myr), three model gyrochrones are provided (at 400, 500 and 600 Myr), which are of similar shape and close together in period. I will refer to this model as the 'SL20 Model'.

### 6.3.4 Comparison to other clusters

To contextualise the rotation sequence for NGC 6633, I plot the period–colour diagram for NGC 6633 alongside data from 3 other well-studied clusters: Figure 6.11 shows the period–colour relationship for NGC 6633 (this work), as well as Praesepe (Douglas et al. 2019), the Hyades (Douglas et al. 2019) and the Pleiades (Rebull et al. 2016a) clusters. I cross-match these three catalogues with Gaia EDR3 parameters to avoid any possible colour-related inconsistencies. I de-redden the $G_{BP} - G_{RP}$ colours, using average values for each of the clusters converted into Gaia $G_{\rm BP}$ and $G_{\rm RP}$ bandpasses as in Section 6.2.2. The $V$ band extinction values for each of the clusters are as follows: Praesepe $A_{\rm V} = 0.035$, the Hyades $A_{\rm V} = 0$ (Douglas et al. 2019) and the Pleiades $A_{\rm V} = 0.12$ (Rebull et al. 2016a). Although the extinction values for Praesepe and the Pleiades are small compared to NGC 6633, this aids in ensuring a fair comparison is being made of the slow-rotator sequences. Figure 6.11 also shows two gyrochronology period–colour relations: a 480 Myr Angus model as a dashed line and a 575 Myr SL20 model as a dotted line. The model ages for these two gyrochrones best fit the NGC 6633 cluster sequence; Section 6.3.6.1 will detail how these best-fit models are calculated.

Qualitatively, it appears that the slow-rotator sequence of NGC 6633 agrees with those of the Hyades and Praesepe. This confirms previous conclusions that NGC 6633 is approximately the same age as these two clusters. The slow-rotator sequence appears to be fractionally lower in period than these two clusters but above the younger Pleiades cluster. Again, this would be expected given previous conclusions that NGC 6633 is fractionally younger than the Hyades and Praesepe.

It is immediately noticeable that the NGTS NGC 6633 data does not extend into as red colours as the three comparison cluster data sets. There are two objects at $G_{BP} - G_{RP} \gtrsim 1.5$ that have not been identified as possible binary systems, which appear to drop below the slow-rotator sequence of the cluster. For the younger Pleiades cluster, the slow-rotator sequence breaks down for objects redder than this, however for the older Hyades and Praesepe clusters, we do not see this rotational slow-sequence turnoff until $G_{BP} - G_{RP} \gtrsim 2.5$. There are just two objects observed in NGC 6633 at these colours, so it is difficult to confirm a trend, but these objects may indicate that NGC 6633 is younger than the Hyades and Praesepe due to this turnoff. This trend was coined the fast-rotator or C-sequence in Barnes (2003) and bifurcates off the slow-rotator I-sequence at redder colours for older clusters, as seen in Figure 6.11. I note that we do see shorter period stars of these older clusters from Douglas et al. (2019) which do not lie on the C-sequence, and so these two NGC 6633 objects may be outliers.

Studying redder stars within clusters of similar ages to NGC 6633 is a vital step in helping to establish gyrochronology relations for M-dwarf cluster stars, which are currently ill-defined.

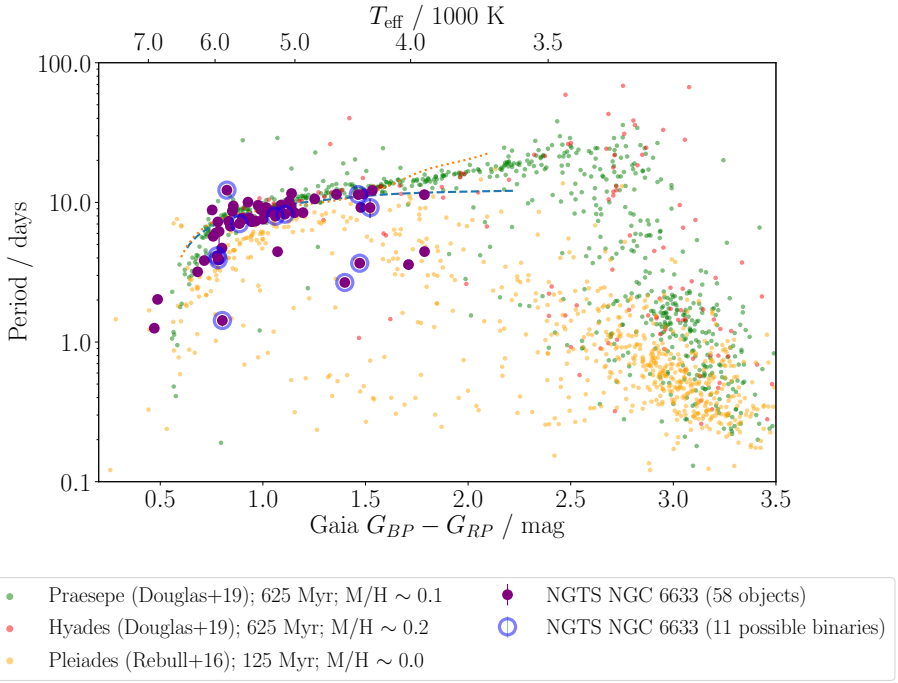



Figure 6.11: Period vs Gaia $G_{BP} - G_{RP}$ colour for objects in NGC 6633 (purple), Praesepe (green, Douglas et al. 2019), the Hyades (red, Douglas et al. 2019) and the Pleiades (yellow, Rebull et al. 2016a) clusters. The NGTS data for NGC 6633 are plotted in purple as in Figure 6.10, with possible binary objects highlighted with blue circles. A 480 Myr Angus et al. (2019) gyrochrone (dashed) and a 575 Myr Spada & Lanzafame (2020) gyrochrone (dotted) are plotted as best fitting to the NGC 6633 data.

The recent work by Popinchalk et al. (2021) combines data from clusters ranging from 10 Myr (Upper Sco) to 750 Myr (the Hyades) and older field stars to ascertain if age-rotation relations exist with M-dwarf populations. They observe the 'elbow' of the slow-rotator sequence of clusters moves to redder colours with increased age (the elbow of Praesepe can be seen at $G_{BP} - G_{RP} \sim 2.75$ in Figure 6.11). However, the authors note a distinct lack of M-dwarf rotation periods in clusters aged between 200 and 700 Myr: exactly the age of NGC 6633.

The metallicity of NGC 6633 is closer to that of the relatively metal-poor Pleiades than the near-solar metallicity of Praesepe and the Hyades. Gyrochronology models have only been calibrated for near-solar metallicity stars, and the full effect of metallicity on these relations is not well understood (Metcalfe & Egeland 2019).

### 6.3.5 Extinction

Jeffries (1997) claim their measurements rule out any differential reddening within the cluster above 0.04 mag through the observed scatter of equivalent line widths (EWs). This contradicts the large differential reddening calculated using 3d dustmaps and EDR3 positions and distances

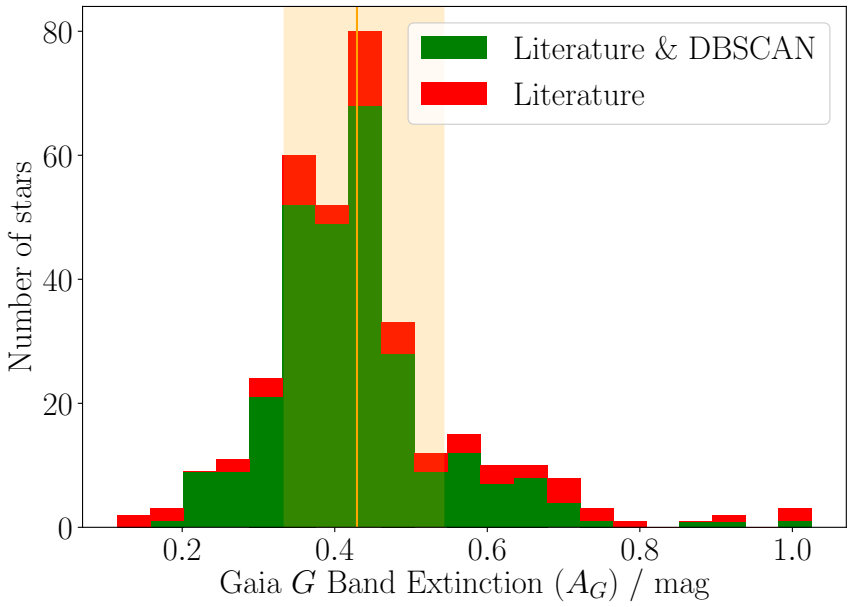



Figure 6.12: A stacked histogram of the dustmap calculated Gaia $G$ band extinction ($A_{\rm G}$) for 342 NGC 6633 cluster members. Objects identified in the literature and by the DBSCAN clustering are shown in green, while objects outside the DBSCAN cluster identified in the literature are plotted in red. The orange vertical line and shaded region represent the median and $16^{\rm th}$–$84^{\rm th}$ percentile spread of $0.43 \pm 0.10$ mag.

in this study. Figure 6.12 shows a histogram of the Gaia $G$ band extinction ($A_{\rm G}$) for NGC 6633 calculated in this study, clearly demonstrating a large scatter in $A_{\rm G}$, which corresponds to a range in $E(B-V)$ from $\sim 0.05$ to $\sim 0.35$. I calculated reddening values for 342 cluster members confirmed with Gaia DR2 or later astrometric parameters compared to the 23 stars used in Jeffries (1997). Furthermore, it is highly probable that with astrometry from Gaia DR2 and sophisticated dustmap models, the reddening calculated in this work is more accurate than those derived through EW calculations.

### 6.3.6 Gyrochronology

From Figure 6.10, it is clear that the rotation period for stars within NGC 6633 follows a colour–period relationship as expected for an open cluster. No previous studies of NGC 6633 have derived gyrochronological ages for the cluster due to poor photometry, which means a good rotational sample is not available. Therefore, confirming the previously derived Li-abundance and isochrone age of NGC 6333 through gyrochronology is an important result. The two plotted gyrochrones from both the Angus and SL20 models at 400 and 600 Myr are extremely close

together; the spread of rotation periods within NGC 6633 is of a similar order of magnitude to the difference between these two gyrochrones. It is, therefore, possible to state NGC 6633 appears to be 400–600 Myr old, which agrees with the LDB measurement of ∼ 600 Myr (Jeffries et al. 2002) and the isochrone measurement of 426 Myr (Pang et al. 2021).

#### 6.3.6.1 Gyrochronology model fitting

To generate a probabilistic age estimate for NGC 6633, I elected to fit two gyrochronology models (from Angus et al. (2019) and Spada & Lanzafame (2020)) to the slow-rotator sequence for the cluster. These models should be suitable as age estimators as they are both fitted on the similarly aged Praesepe cluster within a colour range similar to the population observed with NGTS.

I fitted the models by sampling a posterior distribution with an MCMC sampler, `emcee` implemented in `Python` (Foreman-Mackey et al. 2013). Assuming the errors on the period estimates are Gaussian, the log-likelihood can be defined as

$$\ln p(P_{\rm rot} \mid \boldsymbol{\theta}, f) = -\frac{1}{2}\sum_{\rm n}\left[\frac{(P_{{\rm rot},n} - \psi(\boldsymbol{\theta}))^2}{s_{\rm n}^2} + \ln(s_{\rm n}^2)\right], \tag{6.6}$$

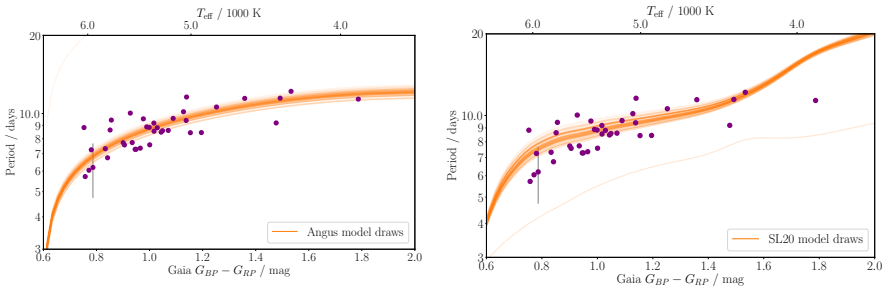
where

$$s_{\rm n}^2 = \sigma_{\rm n}^2 + f^2(\psi(\boldsymbol{\theta}))^2 \tag{6.7}$$

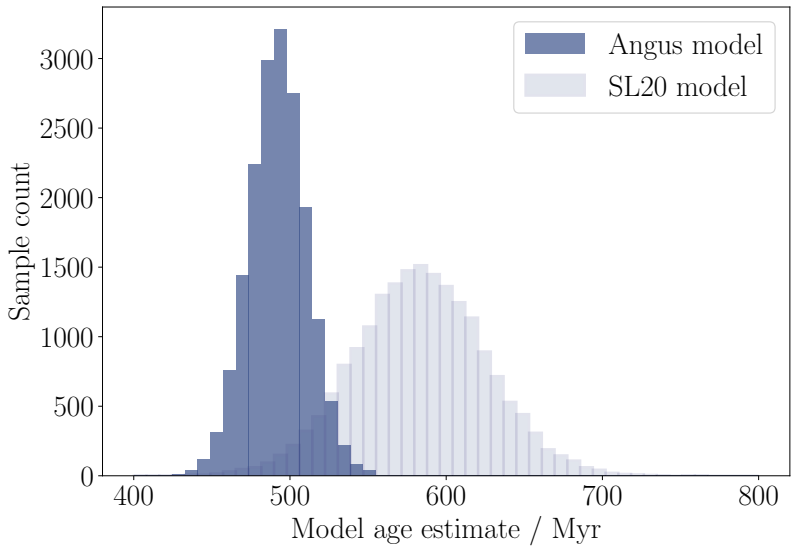
for a model $\psi$ with parameters $\boldsymbol{\theta}$ which estimates a rotation period $P_{\rm rot}$. $\sigma_{\rm n}$ is the error on a measured rotation period $P_{\rm rot}$. This likelihood function is a Gaussian, but the standard deviation is underestimated by some fraction $f$. The introduction of the parameter $f$ affords some flexibility in the model fit by assuming that the period's error values are often underestimated. The simplest way to include this into the model fit is as a constant fractional underestimation of the errors. This parameter $f$ is marginalised over in the quoted age estimates for the model fit.

I used the same uniform priors for fitting both models: $\ln p(f) \sim U[-10.0, 1.0]$ and $p(\text{age}) \sim U[100, 800]$ Myr for the Angus model and $p(\text{age}) \sim U[0.1, 0.8]$ Gyr for the SL20 model. The SL20 model is defined in Gyr, so I fitted in Gyr. The Angus model is defined in Myr, and so I re-scaled using a base-10 logarithm to a similar scale to $f$ in order to fit this model.

The MCMC sampler was run for 10,000 steps, with initial solutions of 500 Myr and $f = 0.5$. Of the 58 objects with detected periodic variability detected, I first removed the 11 possible photometric binaries. I removed object `NG1827+0636.1439675` as a possible false member. I removed the two objects with $G_{BP} - G_{RP} \sim 1.8$ and period ∼ 3 days, as these are not on the slow-rotator sequence for NGC 6633. Finally, I removed the two bluest objects of the sample with rotation periods of approximately 1–2 days ($G_{BP} - G_{RP} \sim 0.5$) as these fall outside the




(a) Period plotted against Gaia $G_{BP} - G_{RP}$ for the 38 NGC 6633 objects, which best follow the cluster slow-rotator sequence (purple points with black error bars in period). Model draws from the MCMC sampler of the fitted Angus model (left), and the SL20 model (right) are shown in orange.




(b) A comparison of posterior samples from the MCMC model fits to the Angus and SL20 models for NGC 6633. The model posteriors have been transformed into Myr for comparison; the Angus model was fitted in log years and the SL20 model in Gyr.

Figure 6.13: Two models were fitted to the slow-rotator sequence of NGC 6633, the Angus model (Angus et al. 2019) and the SL20 model (Spada & Lanzafame 2020). The model fits in colour–period and posterior samples in age are shown in (a) and (b), respectively.

colour range for which the models appear well defined. This left a sample of 38 stars that appear to follow the slow-rotator sequence of the cluster to fit the gyrochronology relations. Several model draws from the sampler are plotted in Figure 6.13a for the Angus model (left) and the SL20 model (right), and the sample distribution of the posterior for both model fits is shown in Figure 6.13b. The best fit age for the Angus model to NGC 6633 was $496^{+20}_{-46}$ Myr and for SL20 $587^{+43}_{-49}$ Myr.

Both estimates agree with previous studies, which give an age slightly younger than Praesepe

and the Hyades, and the two model-fit age estimates are less than 2-sigma discrepant. Despite small errors on the MCMC best-fit age of both models, there are likely much larger errors on the models themselves, which are harder to quantify. I converted models defined in $B-V$ into $G_{BP}-G_{RP}$ (Section 6.2.3), which will introduce a small error into these results, though this is difficult to quantify. Interestingly, the SL20 model gives a significantly older age estimate for the cluster than the Angus model. This model was derived for the slow-rotator sequences of clusters older than NGC 6333, which may explain the older age estimate. However, the model was shown to fit well to the much younger Pleiades rotation sequence (Spada & Lanzafame 2020). Fitting a gyrochronological model may not yield an accurate age estimate for the cluster; Bouma et al. (2021) labels this 'an exercise in tautology', as gyrochronology models are empirically calibrated to a small number of clusters (in the case of the Angus models, just Praesepe). This leads me to the conclusion that although the age estimates are derived with errors for each of the models, the implicit error on the models themselves will outweigh any small calculated errors from an MCMC fit. Taking a mean and min/max spread of the two age estimates would give $542^{+88}_{-92}$ Myr as an age estimate for NGC 6633 from these two model fits, which well agrees with the by-eye estimate of 400–600 Myr, as well as quantitatively with the Li-abundance and main-sequence isochrone turnoff estimates.

#### 6.3.6.2 Simultaneous gyrochronology and isochrone model fitting

Following the formulation from Angus et al. (2019) and using the `stardate` `Python` package from the same work, I also conducted simultaneous gyro- and iso-chronological fits to each of the rotationally variable stars in NGC 6633 not flagged as potential photometric binaries. Isochrone fitting and gyrochronology provide complimentary ageing methods for a wide range of spectral types. Gyrochronology is well calibrated for FGK dwarfs, whereas isochrone fitting works better for hotter, evolved objects. `stardate` (Angus et al. 2019)[g] combines isochrone fitting with gyrochronology in order to increase the precision of stellar age estimates. The gyrochronology model has already been discussed in Section 5.2.6, empirically calibrated to Praesepe. The isochrone fitting is done through the `Python` package `isochrones`(Morton 2015)[h] which is a simple interface to interact with stellar evolution models from the MESA Isochrones and Stellar Tracks (Paxton et al. 2010)[i].

For each star, I supplied `stardate` with extinction corrected relative $G$, $G_{\rm BP}$ and $G_{\rm RP}$ magnitudes and parallax from EDR3 as well as the calculated extinction values in the $V$ band (Section 6.2.2) and derived $T_{\rm eff}$ and $\log g$ from TICv8 (Stassun et al. 2019). The errors on

[g] `https://github.com/RuthAngus/stardate`. Accessed: 22/01/2022.
[h] `https://github.com/timothydmorton/isochrones`. Accessed: 22/01/2022.
[i] `http://waps.cfa.harvard.edu/MIST/`. Accessed: 22/01/2022.

the magnitudes were taken to be the expected Gaia EDR3 photometric magnitude errors: $\sigma_{\mathrm{G}} = 0.001$ mag, $\sigma_{G_{\mathrm{BP}}} = 0.006$ mag and $\sigma_{G_{\mathrm{RP}}} = 0.012$ mag[j]. I assumed the metallicity of the objects to be [Fe/H] = $-0.096 \pm 0.0081$, as in Jeffries et al. (2002).

For the rotation period, I used the adopted period from `RoTo`, and initially, the error on the period estimate from `RoTo`. Following a trial run of `stardate` with 10,000 MCMC steps per star, the age estimates appear to be non-physically large. One object, `NG1827+0636.1378217`, had an estimated age of $2211^{+4092}_{-1572}$ Myr. This could be attributed to a zero error on the period as the best estimated period came from the FFT of the G-ACF. Other large errors on age estimates were for stars with small errors on their period estimates. Angus et al. (2019) assume a constant underestimation of period errors, as Aigrain et al. (2015) and Angus et al. (2018) suggest that often measured rotation period errors are smaller than the true error on the rotation period, which can also arise from a highly non-Gaussian noise distribution. They assume a constant measurement uncertainty of 5% on both their simulated data set and Kepler data of the $\sim 2.5$ Gyr open cluster NGC 6819. The measurement uncertainties on `RoTo` periods were roughly an order of magnitude below this. To provide more realistic period uncertainties, I elected to adopt the same approach as Angus et al. (2019) by adding a 5% period uncertainty to all period estimates prior to running `stardate`.

I used the standard priors from `stardate` as described in the Appendix of Angus et al. (2019), noting that the only significant difference to the rest of this Chapter's work is the distance prior for which the package uses a distance-squared prior ($P(D) \propto D^2$) in comparison to the exponentially decreasing distance prior as defined in Section 6.2.1. I used non-default initial estimates of log(age) = 8.7 (500 Myr), [Fe/H] = $-0.1$, distance = 400 pc and $A_{\mathrm{V}} = 0.558$, taken from my previous calculations or other studies. I found the optimum parameters using an MCMC sampler; I ran 50 walkers, each with 100,000 steps per star, which took around 40 minutes to run on a single laptop core. These are the recommended settings from Angus et al. (2019), and manual inspection of the MCMC chains and derived posteriors of several objects showed well-converged solutions; however, the walkers in age were fairly noisy.

The outputted ages are plotted against $G_{BP} - G_{RP}$ in Figure 6.14. Most noticeably, there are extremely large error bars for several points that extend beyond the y-axis. Error bars were particularly large for objects with small errors on the estimated period, even with the inflated period error estimates as described above. Angus et al. (2019) emphasises the importance of good estimates on the errors for stellar parameters and period to well-constrain stellar ages from `stardate`. The period estimates from `RoTo` contain reasonable period estimates where available, but for cases with just two methods of estimation or just a G-ACF period estimate, the error on the estimated period may be underestimated, which causes a much less precise estimate of

[j] Taken from `https://www.cosmos.esa.int/web/gaia/earlydr3`. Accessed: 22/01/2022.

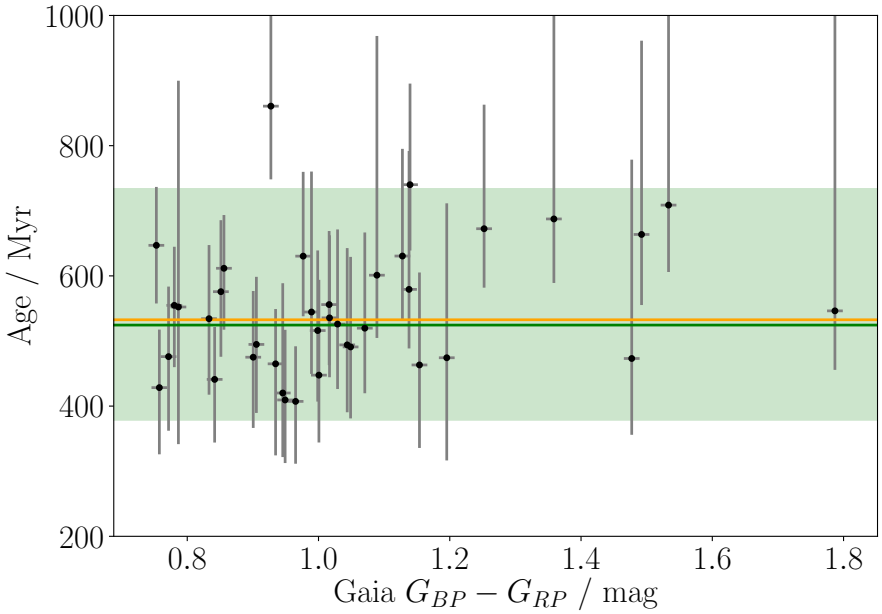



Figure 6.14: `stardate` derived ages and error bars for 38 NGC 6633 stars plotted against Gaia $G_{BP} - G_{RP}$ colour. The orange line shows the error-weighted mean age (532 Myr), and the green line and shaded region show the combined sampled posterior median and 16$^{\text{th}}$–84$^{\text{th}}$ percentile range ($524^{+209}_{-145}$ Myr).

the stellar age. The combination of the period estimates from `RoTo` is not physically motivated; taking a mean or median of the period estimates will not reflect uncertainty in the period arising from astrophysical noise such as latitudinal movement of spots. A full understanding of how period uncertainties affect gyrochronology-derived ages should be explored in subsequent work, beyond the scope of this thesis. Additionally, the relative importance of physical variations in rotation period versus measurement error must be better understood. However, it is clear from this study that with a poor understanding of the period uncertainty, gyrochronological fitting of individual stars will result in poor age estimates.

There is also a slight trend with age; we see bluer stars have, on average, lower ages than redder stars. This is not surprising as `stardate` uses the Angus model to fit for gyrochronological ages. We see from the model's shape (e.g., in Figure 6.10) that the slope appears shallower than the data may suggest, leading to younger age estimates for bluer stars and vice versa.

I combined the age estimates for the cluster stars in two ways: firstly, I calculated an error-weighted mean and error-weighted standard deviation of the sample of 30 `stardate` derived ages. This gave an estimated age for NGC 6633 of 532 ± 531 Myr (plotted as an orange line in Figure 6.14). This result is largely skewed by the large error estimate on the upper age limit of

these stars, so I discarded this estimate. Secondly, I combine posterior samples from each star to create an overall distribution of age samples from the cluster. This combined age posterior distribution gives an age estimate for NGC 6633 of $524^{+209}_{-145}$ Myr when taking the 16$^{\text{th}}$, 50$^{\text{th}}$ and 16$^{\text{th}}$ percentiles (plotted as a green line and shaded region in Figure 6.14). Although the errors are large, this age estimate agrees with previous literature values.

The age estimates of young clusters from isochrone turnoff ageing can be unreliable (Barnes 2003) as, by the age of NGC 6633, most of the cluster members will be on the main sequence. Isochrone fitting will not yield accurate age estimates for populations of main-sequence dwarf stars as low mass stars will spend a significant amount of time on the main sequence, leading to very small changes in the best-fit isochrone. So, including isochrone information for each star may not provide an age estimate as precise as gyrochronological ageing. I have not considered the relative accuracy of each star's isochrone and gyrochronology age estimates; however, this would be possible by running separate isochrone and gyrochronology fits per star (as in Figure 7 of Angus et al. 2019). As I compare each star individually, I cannot use the data from all stars observed to generate an age prediction using based on the slow-rotator sequence fit, potentially reducing the accuracy of any age estimates over the gyrochronological estimates.

## 6.4 Conclusions

I have developed a general-purpose variability detection pipeline and associated `Python` package, `RoTo`, which is freely available online. `RoTo` combines several variability extraction methods to produce robust period estimates from variable light curve data. Currently, I have implemented Lomb–Scargle periodograms, the generalised autocorrelation function (G-ACF) and Gaussian process (GP) regression. The package also provides simple plotting tools to allow quick and easy validation of the results.

I applied `RoTo` to NGTS observations of the young open cluster NGC 6633 to find rotational signals. NGTS light curves for 1,042 stars were analysed, of which 342 were from cluster members confirmed using Gaia DR2 and EDR3 astrometry. I computed accurate distances and interstellar extinctions for candidate NGC 6633 cluster members using Gaia EDR3 stellar parameters and 3d galactic dustmap models. For NGC 6633, the calculated median extinction and 16$^{\text{th}}$ to 84$^{\text{th}}$ percentile spread in the three EDR3 bands were $A_{\text{G}} = (0.41 \pm 0.08)$ mag, $A_{\text{BP}} = (0.45 \pm 0.09)$ mag and $A_{\text{RP}} = (0.26 \pm 0.05)$, which agrees within 1 standard deviation of previous studies.

I produced validated rotation periods for 58 candidate NGC 6633 cluster members, of which 11 are flagged as potential near-equal-mass binaries based on their position on the CMD. I plotted a period–colour diagram for the cluster which shows a reasonably tight slow-rotator

sequence that broadly follows empirical gyrochronological relations. The slow-rotator sequence of NGC 6633 was shown to roughly agree with the slow sequences of the similarly aged Praesepe and Hyades open clusters. However, it may indicate that NGC 6633 is fractionally younger. Based on a by-eye gyrochrone fit, I conclude that the rotational age of NGC 6633 is 400–600 Myr, which agrees with previous age estimates from Li-abundance data and isochronal fitting. I fit two model gyrochrones from Angus et al. (2019) and Spada & Lanzafame (2020) and run a simultaneous isochrone and gyrochrone fit on each of the cluster members to produce probabilistic age estimates, which give a cluster age of $479^{+18}_{-20}$ Myr, $567^{+52}_{-50}$ Myr and $524^{+209}_{-145}$ Myr respectively. These age estimates agree quantitatively with previous results from Li-abundance ageing and main-sequence isochrone turnoff fitting and with the statement that NGC 6633 is slightly younger than the Hyades and Praesepe clusters. I note that these age estimates do not account for the relatively low metallicity of NGC 6633 compared to the Hyades and Praesepe. However, the cluster slow-rotator sequence broadly fits into the collective knowledge of the rotational evolution of stars in open clusters.

CHAPTER 7

# OVERALL CONCLUSIONS AND FUTURE WORK

I present a summary of the three major projects undertaken during my PhD, including important results and findings. I consider these results within the context of this PhD and suggest further improvements and next steps for each of the three studies.

## 7.1 Development of the G-ACF

This project aimed to implement and test a generalisation of the autocorrelation function (ACF), which applies to irregularly sampled data. This algorithm was developed in collaboration with Lars Kreutzer, Edward Gillen and Didier Queloz to allow fast estimation of the ACF of ground-based photometric data. The ACF has been previously shown to be a robust method of extracting non-sinusoidal periodicity from astrophysical data. My contribution to this project was in the implementation and testing of the algorithm, which was implemented in `C++`, with a `Python` wrapper for ease of use. I experimented with the functional form of the weight function and the parameters of the G-ACF, using three simulated data sets: two simple sine-waves and a more complex stochastic process with a periodic component. I also demonstrated the similarity of the G-ACF to a standard ACF for these three data sets when sampled regularly, randomly and with a cadence-like sampling structure similar to ground-based data.

In collaboration with Ed Gillen, I applied the G-ACF to real astrophysical data: a photometric light curve from the Kepler mission of the previously studied spotted star KIC 5110407. We demonstrated that the G-ACF can extract rotation period information with comparable accuracy to a much more complex Gaussian process for this phase-shifting signal, but with a

much shorter computation time and fewer model assumptions.

The G-ACF applies to any time series domain data where the sampling is not regular, and an ACF would yield interesting information within the time series. Despite testing the G-ACF on synthetic and real astrophysical data, I did not apply the G-ACF to any other problems. The sampling of these data sets can differ from astronomical data, and the scales and forms of noise within these data sets may also pose additional problems. It may transpire that the optimal set of parameters and the form of the weight function differ within different contexts. Additionally, I only tested a few simple, functional forms for the weight function. There are strict criteria on the form of the weight function, and a simple function brings fewer assumptions to the model. It would, however, be interesting to include the effects of any errors on the data into the weight function, down-weighting points with large errors. This has not been implemented yet into the G-ACF; however, the codebase and testing framework would make it simple to implement and test.

It would also be interesting to compare the results of the G-ACF with other generalisations of the ACF, particularly the 'slotting' based approaches and the discrete correlation function. A few algorithmic improvements could speed up the compute time of the G-ACF, which are not yet implemented; however, in general, it should be faster than any approach that requires aggregation or binning.

Finally, further assessments of the mathematics of the G-ACF could shed light on errors on G-ACF-derived periods, which in turn can be used to assess how wide a range of circumstances the algorithm can be effectively used. This includes assessing how the G-ACF is affected by sampling and formally defining sampling-related aliases that plague the G-ACF of cadence-like sampled data.

The motivation for this project was within time-domain astrophysics, particularly photometric light curve analysis, and I was able to successfully demonstrate that for typical data sets within this field, the G-ACF can accurately reproduce the ACF of data with irregular sampling. The G-ACF is available for use in `Python` and `C++` for users to begin to experiment with the G-ACF within wider contexts.

## 7.2 Periodic stellar variability from almost a million NGTS light curves

I applied the G-ACF algorithm to the entirety of the NGTS photometric light curve data set to extract periodic variability. This involved the development of a large data processing pipeline, which was able to process almost one million photometric light curves rapidly with consideration of the noise sources and aliasing present within this data. This work represents

the largest ground-based systematic variability study using such precise and high-cadence photometry but highlights the advantages and challenges of such a data set. In particular, systematic variability arising from imprecise background correction and aliased signals from the 1-day sampling cadence of ground-based telescopes caused large numbers of systematic noise detections. Despite these challenges, I was able to extract periodic variability signals from 16,880 photometric light curves down to amplitudes of 0.3% in relative flux.

The large data set of NGTS spans a range of positions and types of stars, and I was able to produce a sample of field star variability periods on stars from late-A through to mid-M spectral types, with rotation periods from $\sim$ 0.1 to 130 days. I compared this sample with several previous variability studies, demonstrating the power of the period extraction pipeline and the precision of the NGTS photometry compared to ASAS-SN, Gaia, MEarth and TESS. The data set contains a large number of main-sequence stars, similar to the rotational data sets from Kepler and K2, but also contains many redder objects and giants, which highlight interesting and diverse variability. I explored how this variability was distributed within the HR diagram to highlight interesting populations of objects such as binary stars, PMS objects and giant pulsating variables. I also assessed the variability sample in period–colour space, where I observed an absence of detected variability (most likely rotation of main-sequence stars) between 15 and 25 days. This gap was previously observed within Kepler and K2 data; this study confirms the presence of this gap from the ground. The presence of a rotation period gap aids the development of rotational evolution models by providing empirical evidence for such models as Lanzafame & Spada (2015) and Spada & Lanzafame (2020), which include a 'broken spin-down' evolution of stellar rotation.

One line of investigation not conducted during this work was into the prevalence of half- and twice-period aliases. Such aliases could arise as a result of the period detection pipeline, or from astrophysical phenomena such as a spotted star with active regions on opposing hemispheres. Despite this work demonstrating the clear presence of a dearth of detected stellar rotation signals, the two distinct branches may in related by these half- and twice-period aliases. A similar gap was noticed in flux–colour diagrams when measuring the chromospheric Ca–II H and K emissions in field stars (The 'Vaughan–Preston Gap', Vaughan & Preston 1980), however recent works by Zhao et al. (2015) and Boro Saikia et al. (2018) have called into question the existence of the Vaughan–Preston Gap based on recent large scale analyses. If the observed rotation period gap could be explained by such $P/2$ aliases, this would negate the need for complex stellar evolution models such as Spada & Lanzafame (2020).

Studying large populations of field stars is useful for analysing broad trends within stars grouped into regions within the CMD or period–colour space, however as field stars will range in age, this is not a useful exercise for calibration of empirical ageing methods such

as gyrochronology. This requires targeted observations of fixed-age populations, such as open clusters, which have been considered in the final chapter of this thesis. However, this does not mean that additional follow-up work is not possible from field studies. Populations of interesting objects highlighted within the study are suitable for follow-up; for example, objects with short variability periods above the main sequence in the CMD may be interesting candidates for spectroscopic follow-up as PMS stars. These early-age stars provide insights into the beginnings of the lives of stars and may shape stellar evolution models, which are often poorly defined at these early ages.

Since the publication of this work, NGTS has observed an additional 158 fields (as of Feb 2022), and the period detection pipeline could easily be re-run on these. Further modifications and improvements to the NGTS light curve processing pipeline are constantly being made, such as a finer grid for the background correction and a more comprehensive systematic removal step. Running the periodicity detection software on these re-processed light curves may yield fewer false-positive detections and bring to light signals which were previously below the noise.

Beyond this re-processing, it would be a fairly straightforward task to run the outputs of the period detection pipeline through a machine-learning-based classifier to characterise the variability detected. This has been done for other variability studies to great success, and by using either a neural-network-based approach on the light curve data or a clustering approach using derived rotation periods and stellar parameters, it would be possible to generate clusters of similar periodic signals in terms of period, amplitude and signal shape as well as by spectral type. Further to this, this work may be able to be linked back to the original scientific goals of NGTS, exoplanet discovery. By assessing the types of variable signals typical of intrinsic stellar variability and binary systems, it may be possible to filter out false-positive detections from the exoplanet search pipeline. Furthermore, understanding the typical variability signals associated with a certain spectral type may aid in understanding the likelihood of (potentially habitable) planets existing around such stars.

## 7.3 Periodic stellar variability in the open cluster NGC 6633

NGTS observed the ~ 500 Myr old open cluster NGC 6633 during 2019 and 2020. This study aimed to assess the rotational variability of stars within this cluster using NGTS photometry. As highlighted in the rotational study of field stars, rotational analysis of groups of coeval populations such as open clusters provides much more insight into the rotational evolution of stars.

To facilitate this, I developed a general-purpose periodicity extraction software package, `RoTo`. `RoTo` combines multiple periodicity detection methods, currently a Lomb–Scargle peri-

odogram, the G-ACF and a Gaussian process (GP) regression method, to provide precise period information from a time series. Using multiple period detection aids in confirming whether a detected period is a real signal or a false positive of that specific method.

I conducted an in-depth study of the open cluster NGC 6633, which has previously not been studied in great detail; in particular, this cluster lacks targeted photometric data, of which rotational variability is one data product. I produced robust membership lists for the cluster, drawing from literature periods using a variety of clustering algorithms on Gaia DR2 data and implemented a clustering algorithm using DBSCAN and the latest Gaia EDR3 parameters to confirm potential members.

I confirmed the distance of the cluster using Gaia EDR3 stellar parameters and derived differential extinction values for cluster members using EDR3 parameters and a 3d-dustmap model. These values are particularly important for this cluster, as it is fairly distant, which will cause significant reddening, and hence may bias any results which rely on uncorrected magnitude values or colour.

I assessed how the cluster members are positioned in a CMD and period–colour space. I identified a clear main sequence for the cluster in colour–magnitude space and highlighted potential binary systems that lie above this. Within period–colour space, there is a clear slow-rotator sequence for the cluster, which qualitatively agrees with previous age estimates, lying fractionally below the slow-rotator sequences of the similarly but slightly older Praesepe and Hyades clusters. I produced probabilistic age estimates for the cluster by fitting two gyrochronology models (from Angus et al. 2019 and Spada & Lanzafame 2020) and conducting a simultaneous gyrochronology and isochrone fit. These methods produced estimated cluster ages of $479^{+18}_{-20}$ Myr, $567^{+52}_{-50}$ Myr and $524^{+209}_{-145}$ Myr respectively. These age estimates all agree quantitatively with previous age estimates for NGC 6633, although the different models produced slightly different age estimates. This highlights the errors associated with empirical gyrochronology models, and once again, the importance of targeted observations of similar age populations to calibrate stellar evolution models.

Improvements could be made to the rotational analysis pipeline, such as the inclusion of the data pre-processing steps conducted in previous NGTS studies to remove systematic signals including Moon correlated signals.

In several objects, a GP model was not fitted to the light curve. By re-running the analysis pipeline with longer timeouts, it may be able to fit a GP model to such objects. Alternatively, an in-depth analysis of these objects may provide clues to further refinements to the GP model which result in a better fit, such as alterations to the default kernel hyperparameters.

Within the NGC 6633 sample, several objects lie well below the slow-rotator sequence. These objects provide interesting candidates for follow-up observations: spectroscopic, astro-

metric or photometric. Confirmation of a binary rotational sequence in the cluster (such as in Gillen et al. 2020) may illuminate how stars evolving in these systems differ from their single-star counterparts. Alternatively, a similar half-period aliasing effect as hypothesised in the further work to Chapter 5 may affect the rotation period of these stars and should be taken into consideration.

There is also the open question of how metallicity affects the angular momentum evolution of stars. Previous studies are mainly theoretical, providing insights into how metallicity may affect the internal structure and hence chemical and angular momentum transport within stars. Continuing to observe and produce rotational information for clusters of differing ages and metallicities should enable a picture to be painted of how metallicity affects the rotational distribution of stars of a specific age.

Further development of the `RoTo` package, such as the implementation of additional period detection methods, would be straightforward. The package is written in a modular format, meaning that new period detection methods are straightforward to implement. The first new method to be implemented will be phase-dispersion minimisation, which has been proven to work well for photometric light curve data. Further refinements to the GP model could be made, as the model itself is currently a 'one-size-fits-all' stellar rotation model. For example, the package should allow the user more freedom to specify model terms and parameters. Further testing should be conducted on how to best combine periods outputted by `RoTo`. In this study, manual inspection of the outputs was deemed the most appropriate method for assessing the 'best' period, which may not always be feasible. Furthermore, it was unclear how best to combine the period estimates and their errors to produce a sensible period range. This was further highlighted in the simultaneous gyro- and isochrone fit, in which underestimated errors on the period led to unphysically large age estimate errors. Not only should the combination of period estimates be further considered, but also the errors within each period estimate. This leads back to further work for the G-ACF project, including flux errors in the G-ACF calculation, and gaining a better understanding of how sampling affects the G-ACF may also allow a better understanding of the order of the error in the G-ACF period estimate.

Finally, NGTS has observed a number of other potentially interesting clusters as part of the open clusters working group. One such cluster is the $\sim$ 50–60 Myr old Trumpler 10 cluster. This cluster has very little previous data, in particular photometric, and so conducting a similar study to that of NGC 6633 would be both straightforward and also informative. The rotation of extremely young stars and in particular star-disc interaction for such stars is an active topic of scientific interest.

## 7.4 Summary

In summary, I have developed and tested algorithms and pipelines for extracting periodic signals from irregularly sampled data. This work has been conducted within the astrophysical context of detecting and characterising the periodic variability of stars observed photometrically.

I aided in developing the generalised autocorrelation function (G-ACF), a generalisation of the ACF to irregularly sampled data. My contributions to this project were implementing and testing the algorithm, including computational considerations and the optimal functional forms and parameters of the weight and selection functions central to the G-ACF. I demonstrated that the implementation of the G-ACF accurately reproduces the expected ACF for simple synthetic data, as well as more complex stochastically driven examples and real Kepler photometry.

I then applied the G-ACF to the entirety of the NGTS photometric light curve data set. This project's scientific focus was to search for and characterise periodic variability observed with NGTS, as well as demonstrate the usefulness of the G-ACF within astrophysics. In Briegal et al. (2022), I presented 16, 880 variability periods from 829, 481 objects observed with NGTS between 2015 and 2018, which span late-A through to mid-M spectral types and with periods between $\sim 0.1$ and 130 days. I explored how these variable objects are distributed in colour–magnitude and colour–period space and demonstrated we could observe a distinct bi-modality in colour–period, previously only observed within space-based data.

Finally, based on the conclusions of the NGTS periodic variability study, I applied period finding algorithms to the $\sim 500$ Myr open cluster NGC 6633 observed by NGTS from 2019 to 2021. I conducted an in-depth search for periodic variability within NGC 6633, using a periodicity detection package, `RoTo`, which I developed to include three commonly used period extraction methods: a Lomb–Scargle periodogram, an ACF (using my G-ACF implementation), and a Gaussian-Process regression method. Using the latest Gaia EDR3 astrometry, I derived accurate distances and differential extinction values for cluster members. I assessed the colour–period distribution for NGC 6633, comparing it with other cluster rotational main-sequences and gyrochronology models, to give a probabilistic age estimate for NGC 6633 that agrees with age estimates from complementary methods.

This work is firmly rooted in big data principles: I applied complex algorithms to large data sets to ascertain statistical summaries of the data set in terms of rotation period and other stellar parameters. I made extensive use of the Cambridge HPC facilities and spent time implementing, testing, and optimising algorithms and periodicity detection pipelines to streamline discoveries. I have contributed a large amount of open-source code from this project, including two publicly available `Python` packages: `GACF` and `RoTo`, which astronomers are already using within UK universities. I have also demonstrated that large ground-based photometric datasets can give

statistical results comparable with those from space-based data. The caveat is that this requires comprehensive analysis pipelines that account for the complex noise sources within this data. I have also highlighted the importance of ancillary science focuses within telescope consortia. This project started within an exoplanet detection context but has produced many interesting empirical results within stellar evolution and gyrochronology. This view is synonymous with the recent introduction of the NGTS working groups, which target results beyond exoplanet detection and characterisation with NGTS.

# Chapter 8

# Publications and other work

Over the course of my PhD, I have been involved with the NGTS consortium. This includes responsibilities for manual eyeballing of fields to select potential planet candidates, attendance at quarterly meetings and participation in collaborative research. I will briefly discuss my contributions to a number of papers I have been authored on over the course of my PhD, including a significant contribution to the first NGTS clusters working group publication on rotation in the open cluster Blanco 1. A full list of publications I have been authored on is given in Table 8.1.

As a part of the CDT in Data Intensive Sciences, I spent six months (from August 2019 to January 2020) working full-time at the AI cyber-security firm Spherical Defence. During this work placement, I worked on many aspects of the production codebase for their anomaly detection pipeline; this included a large-scale refactor of their data delivery pipeline. This placement allowed me to further develop my skills working with production code, including aspects of version control and DevOps, which I have applied to the open source code produced throughout this PhD.

## 8.1 NGTS clusters survey – I. Rotation in the young benchmark open cluster Blanco 1

This study, led by Edward Gillen with me as second author, determined rotation periods for 127 stars in the $\sim$ 115-Myr-old open cluster Blanco 1 using NGTS data. We determined rotation periods using three methods: GP regression, the Generalised Autocorrelation Function

| Description | Reference | Journal and Status | Contribution |
|---|---|---|---|
| Stellar variability from almost a million NGTS light curves | Briegal et al. (2022) | MNRAS, accepted | See Chapter 5 |
| G-ACF: A generalised autocorrelation function for irregularly sampled time series | Kreutzer et al. *submitted* | MNRAS, under review | See Chapter 4 |
| NGTS clusters survey – I. Rotation in the young benchmark open cluster Blanco 1 | Gillen et al. (2020) | MNRAS, 2020 | See Section 8.1 |
| Stellar flares detected with the Next Generation Transit Survey | Jackman et al. (2021) | MNRAS, 2021 | G-ACF analysis & proof-reading |
| NGTS clusters survey – III. A low-mass eclipsing binary in the Blanco 1 open cluster spanning the fully convective boundary | Smith et al. (2021b) | MNRAS, 2021 | G-ACF analysis & proof reading |
| NGTS-14Ab: a Neptune-sized transiting planet in the desert | Smith et al. (2021a) | A&A, 2021 | Eyeballing |
| NGTS 15b, 16b, 17b, and 18b: four hot Jupiters from the Next-Generation Transit Survey | Tilbrook et al. (2021) | MNRAS, 2021 | Eyeballing |
| A Transiting Warm Giant Planet around the Young Active Star TOI-201 | Hobson et al. (2021) | AJ, 2021 | Eyeballing |
| NGTS-12b: A sub-Saturn mass transiting exoplanet in a 7.53 day orbit | Bryant et al. (2020a) | MNRAS, 2020 | Eyeballing |
| NGTS-8b and NGTS-9b: two non-inflated hot Jupiters | Costes et al. (2020) | MNRAS, 2020 | G-ACF analysis & eyeballing |
| NGTS-10b: the shortest period hot Jupiter yet discovered | McCormac et al. (2020) | MNRAS, 2020 | Eyeballing |
| Classifying exoplanet candidates with convolutional neural networks: application to the Next Generation Transit Survey | Chaushev et al. (2019) | MNRAS, 2019 | Proof-reading |
| NGTS-7Ab: an ultrashort-period brown dwarf transiting a tidally locked and active M dwarf | Jackman et al. (2019b) | MNRAS, 2019 | Eyeballing |
| NGTS-6b: an ultrashort period hot-Jupiter orbiting an old K dwarf | Vines et al. (2019) | MNRAS, 2019 | G-ACF analysis & eyeballing |
| NGTS-4b: A sub-Neptune transiting in the desert | West et al. (2019) | MNRAS, 2019 | Eyeballing |
| NGTS-2b: an inflated hot-Jupiter transiting a bright F-dwarf | Raynard et al. (2018) | MNRAS, 2018 | Eyeballing |

Table 8.1: A list of authored publications during my PhD, as well a brief summary of my contribution to the work.

(G-ACF, Chapter 4) and LS periodograms; the period determination using the G-ACF was conducted by me.

As we saw in Chapters 1 and 2, open clusters are an excellent target for observations and science programmes seeking information on the rotational evolutions of stars. Blanco 1 is a similar age to the Pleiades cluster, and pairs of similarly aged clusters provide a means to determine the rotation period distribution at a given age from two independent samples of stars in different cluster environments. This was the motivation behind the NGTS observations of Blanco 1, as the Pleiades has a rich history of previous rotational studies (Hartman et al. 2010; Rebull et al. 2016a,b; Covey et al. 2016; Stauffer et al. 2016) in comparison to the relative dearth of rotational detections within Blanco 1: just 33 photometric rotation periods were reported by Cargile et al. (2014).

NGTS observed Blanco 1 using a single camera over a 195-night long baseline from May to November 2017, taking a total of 201,773 images at 13-second cadence on 134 nights. Light curves were extracted for confirmed cluster members; the membership list of the cluster was taken from Gaia Collaboration et al. (2018b) in which groups of objects clustered astrometrically are selected and assessed for tight main-sequence distributions in colour–magnitude. NGTS was able to generate photometric light curves for 429 of the 489 members of Blanco 1 from Gaia Collaboration et al. (2018b).

This work presented the first use of the G-ACF on a large astrophysical data set and compared the outputs from the G-ACF with two other periodicity detection methods. In order to extract a rotation period from the computed G-ACF, I elected to use a two-stage fast Fourier transform (FFT): the FFT of the entire G-ACF was calculated to find an approximate first period, and this period is then refined using a smaller section of the G-ACF up to 5 times the initial period estimate in lag. This refinement reduces signal degradation caused by the evolution of the rotational signals of solar-type stars: at large lag values, we will be shifting the time series against another part of the time series in which the signal shapes are no longer matching, Figure 8.1 show two such examples. This Figure shows NGTS light curves and detected periods for two stars within Blanco 1; in both stars, the G-ACF and the GP fit have excellent agreement with the LS lying outside the bulk of the GP posterior. This highlights the importance of using non-sinusoidal models when evaluating rotation within stars, and in particular young stars such as the stars within Blanco 1. The rotation periods of all 127 stars were manually assessed for validity, and the by-eye best method was selected in each case. The GP method produced the most reliable periods across the entire sample, however in the case of four short period stars the other methods were used. We note that the GP is the only method to produce an error on the outputted periods as it is able to sample a posterior distribution. The multiplicity of each star was also considered, as the presence of companion stars will affect the rotational evolution

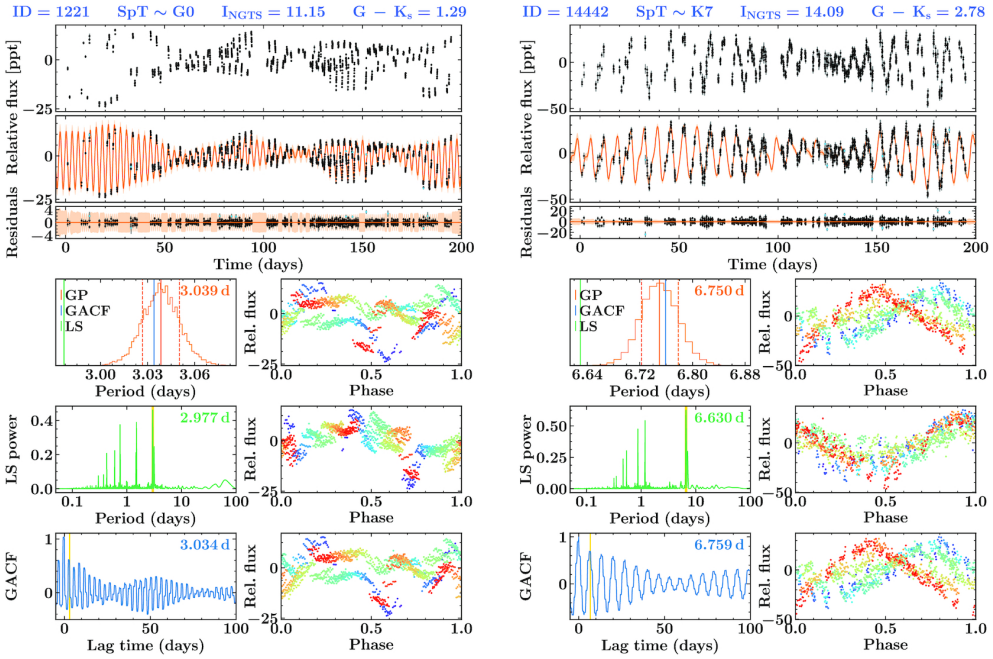



Figure 8.1: NGTS light curves and detected periods from two Blanco 1 stars. The top three plots show the NGTS light curve, with a GP model overlayed in the middle plot and the residuals of this model fit in the bottom plot. The bottom six plots represent the periods extracted from the GP, LS and G-ACF methods, as well as phase folded light curves on each of the periods on the right. The GP period shows the posterior period distribution, as well as lines showing the period detections from the G-ACF and LS. The LS periodogram is shown in the middle-left plot and the G-ACF of the light curve in the bottom left plot, with lines in both plots indicating the position of the detected period. (Credit: Gillen et al. 2020).

of the system. This was done using two complementary approaches: firstly by fitting the single-star cluster sequence in colour–magnitude space and identifying stars above this trend and secondly through cross-matching with literature RV surveys.

The colour–period diagram of the rotation periods detected within Blanco 1 shows a strong sequence between mid-F and mid-K stars in colour, with photometric multiples in general sitting below this sequence (Figure 8.2). It was noted that the rotation sequence appears to break down for the redder objects, with a much broader distribution of rotation periods in the M-dwarf sample of the population. Conversely, the M stars appeared to show much more stable rotation periods than the FK stars, which exhibited significant phase evolution within the observation spans. This could be attributed to different magnetic field morphologies between the two populations and hints at a possible relation between magnetic field topology and the convergence onto the well-defined rotation sequence for a given age.

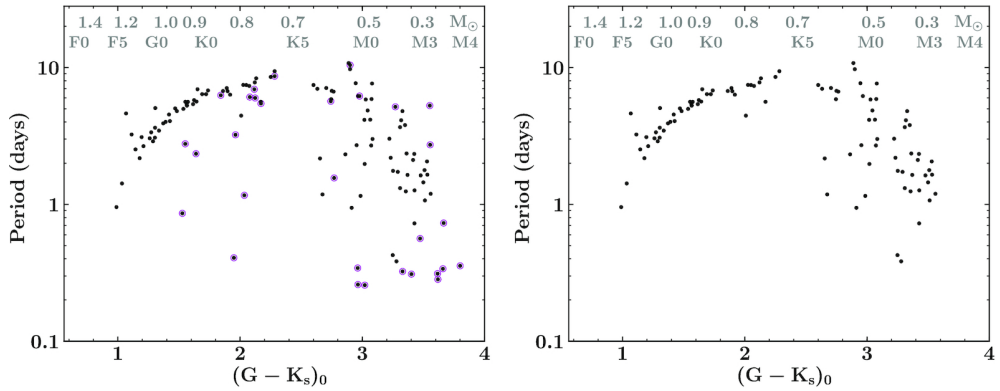



Figure 8.2: Rotation period versus colour for stars in Blanco 1. The left plot shows all objects, including those identified as multiple star systems. The right plot shows only the apparent single-star systems, highlighting a clear rotation period dependence on colour. (Credit: Gillen et al. 2020).

## 8.2 NGTS planet discoveries

Additionally, I have been authored on a number of NGTS planet discovery papers during my PhD. My contributions to these papers have been in the form of confirming the rotation period of the host star (or lack thereof) using the G-ACF stellar variability pipeline from Chapters 4 and 5 and in manual candidate eyeballing. As most of the NGTS planet detections have been planets orbiting main-sequence Solar-type stars, generally the rotational signals have been weak and without significant phase evolution.

# APPENDIX A

# NGC 6633 RoTo OBJECTS

Tables A.1 through to A.5 give the estimated variability periods and derived stellar parameters for the 58 objects analysed from the open cluster NGC 6633 in Chapter 6. I include the `RoTo` generated plots for these objects. The format of these plots is explained below, taken from Section 6.2.5.5.

`RoTo` can generate plots that show the estimated periods from each method, as well as diagnostic plots and phase folds for each method, to enable validation of the estimated periods. I split an example of a `RoTo` generated `.pdf` into five plots (Figures A.1 – A.5), and explain each

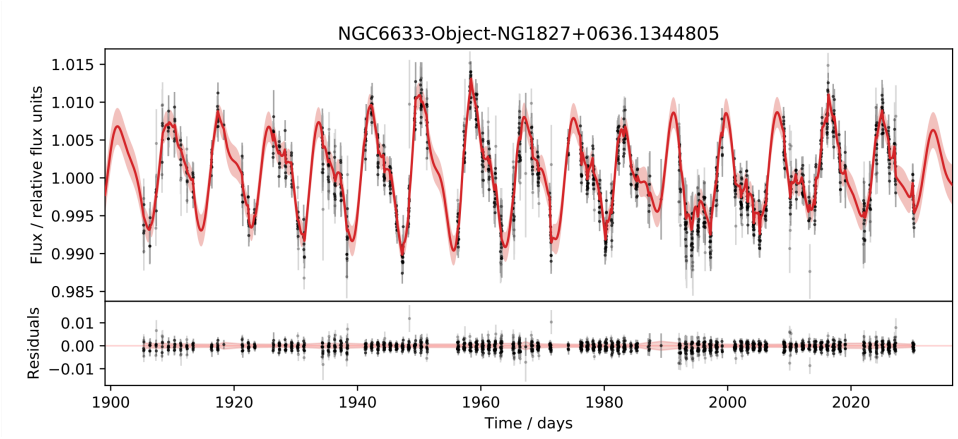



Figure A.1: Example `RoTo` data plot. The light curve is plotted as black points with error bars (top). A GP model fit is overlaid in red, with 1 $\sigma$ uncertainty intervals in light red. The residuals of the GP model fit are plotted in black (bottom), with the uncertainty of the GP model overlaid in light red.

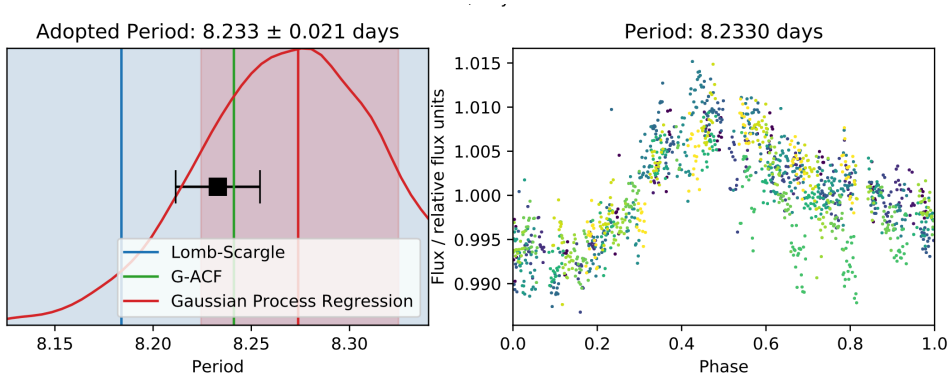



Figure A.2: Example RoTo combined period estimate plot. Outputs from three period estimation methods are plotted (left). Vertical blue and green lines show the period estimate from an LS periodogram and a G-ACF, respectively, with error bars plotted as the same colour shading. The GP posterior is shown as a red line, with the mean (vertical red line) and 1 $\sigma$ uncertainty (light red shading). The combined period and uncertainty are plotted as a black point with an error bar. The right-hand plot shows the light curve phase folded on this combined period.

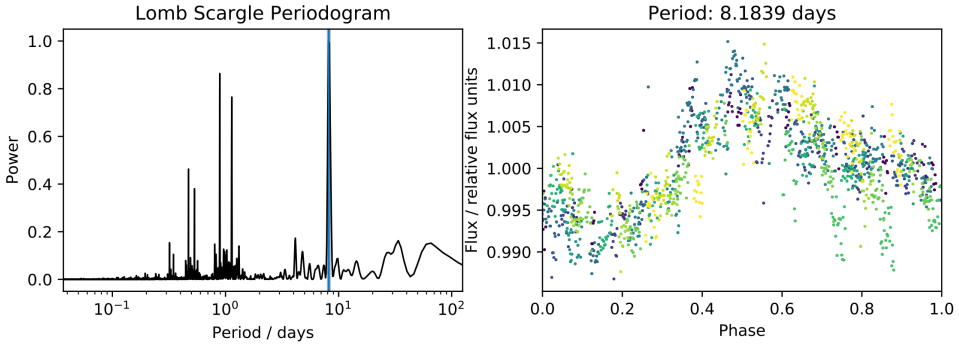



Figure A.3: Example RoTo method plot. The Lomb–Scargle periodogram of the entire light curve is shown (right). The estimated period is plotted as a blue line with errors plotted as a light blue region. In this example, the errors on the estimated period are extremely small, and hence may not be visible. The left plot shows the light curve phase folded on this LS estimated period.

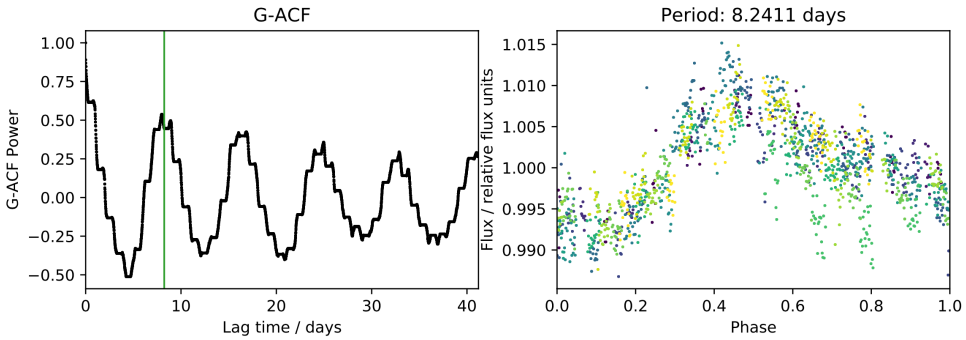



Figure A.4: Example `RoTo` method plot. The G-ACF of the light curve is plotted (right). The estimated period is shown as a green line with errors plotted as a light green region. In this example there are no error estimates on the period, and hence not visible. The left plot shows the light curve phase folded on this G-ACF estimated period.

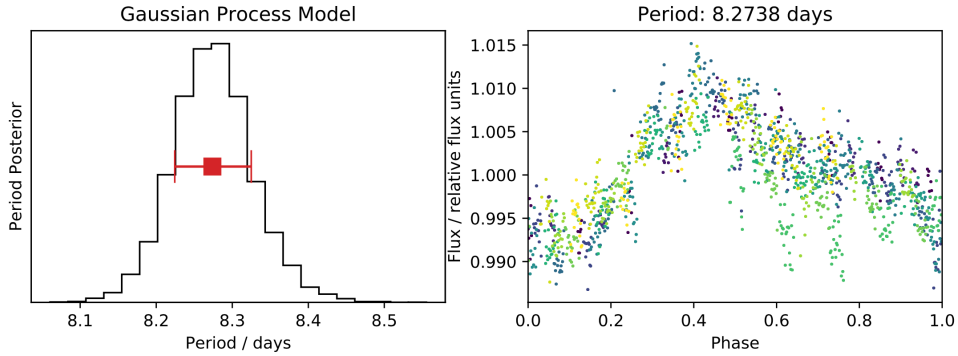



Figure A.5: Example `RoTo` method plot. The GP model period posterior is plotted as a black histogram, with the estimated period and 1 $\sigma$ uncertainty shown as a red point with error bars (right). The left plot shows the light curve phase folded on this GP estimated period.

plot in turn.

At the top of a `RoTo` output `.pdf` (Figure A.1), the entire light curve is plotted as black scatter points, with errors on each point. Overlaid is the MAP GP model fit in red, with 1 $\sigma$ uncertainty intervals in light red. Below this, the residuals of this model fit are plotted in black. 1 $\sigma$ uncertainty intervals for the GP model are plotted in light red.

The second row of the `.pdf` output is shown in Figure A.2. The left plot shows the results of the individual `RoTo` methods overlaid with error bars. In the case of the GP model (red), the period posterior distribution is plotted, along with the mean (vertical red line) and 1 $\sigma$ uncertainty (light red). For the G-ACF (green) and LS (blue), the estimated period is shown as a vertical line, with uncertainty as light shading of the same colour. In this example, the uncertainty on the LS period extends beyond the x-axis. The combined period and uncertainty are shown as a black point with error bars. The right plot shows the light curve phase folded

on this combined period.

Figures A.3, A.4 and A.5 include the remainder of the RoTo output. The left-hand plots show details of the method, which can be a periodogram, an ACF or a period posterior distribution. The right-hand plot in each example will show the light curve phase folded on the period found by that method.

Table A.1: Adopted rotation periods and method used for the 58 objects with detected periodic variability in NGC 6633. Objects are sorted by period, with potential binary objects separated at the end of the table. The page number of each object's corresponding RoTo plot is given.

| NGTS ID | $P_{\rm rot}$ | $\Delta P^{-}_{\rm rot}$ | $\Delta P^{+}_{\rm rot}$ | Method | Figure Page Number |
|---|---|---|---|---|---|
| NG1827+0636_1321241 | 1.25610 | 0.00547 | 0.00527 | GP | 184 |
| NG1827+0636_1451334 | 2.01984 | 0.13767 | 0.13767 | Combined | 185 |
| NG1827+0636_1100060 | 3.18430 | 0.05449 | 0.05449 | Combined | 186 |
| NG1827+0636_1794649 | 3.58670 | 0.02526 | 0.02526 | Combined | 187 |
| NG1827+0636_1097244 | 3.83260 | 0.02541 | 0.02541 | Combined | 188 |
| NG1827+0636_1008149 | 3.88881 | 0.01768 | 0.01768 | Combined | 189 |
| NG1827+0636_1439675 | 4.43841 | 0.01877 | 0.01877 | Combined | 190 |
| NG1827+0636_1422346 | 4.44441 | 0.00153 | 0.00153 | Combined | 191 |
| NG1827+0636_1319323 | 4.70421 | 0.00059 | 0.00059 | Combined | 192 |
| NG1827+0636_1439611 | 5.71558 | 0.02011 | 0.02011 | Combined | 193 |
| NG1827+0636_1452712 | 6.04603 | 0.05811 | 0.05811 | Combined | 194 |
| NG1827+0636_1010770 | 6.19723 | 1.46872 | 1.46872 | LS | 195 |
| NG1827+0636_1329847 | 6.75315 | 0.01451 | 0.01451 | Combined | 196 |
| NG1827+0636_1328554 | 7.24699 | 0.01334 | 0.01334 | Combined | 197 |
| NG1827+0636_1327844 | 7.27716 | 0.04115 | 0.04115 | Combined | 198 |
| NG1827+0636_1041927 | 7.27979 | 0.09911 | 0.10627 | GP | 199 |
| NG1827+0636_1300498 | 7.31989 | 0.14121 | 0.14121 | Combined | 200 |
| NG1827+0636_1106701 | 7.35425 | 0.01780 | 0.01780 | Combined | 201 |
| NG1827+0636_1027580 | 7.56903 | 0.01087 | 0.01087 | Combined | 202 |
| NG1827+0636_1769107 | 7.57452 | 0.01025 | 0.01025 | Combined | 203 |
| NG1827+0636_558862 | 7.71833 | 0.09950 | 0.09950 | Combined | 204 |
| NG1827+0636_958489 | 7.72781 | 0.01946 | 0.01946 | Combined | 205 |
| NG1827+0636_849720 | 8.43288 | 0.16477 | 0.16477 | Combined | 206 |
| NG1827+0636_2096690 | 8.44428 | 0.06988 | 0.06988 | Combined | 207 |
| NG1827+0636_1447254 | 8.47575 | 0.00337 | 0.00337 | Combined | 208 |
| NG1827+0636_1321895 | 8.54611 | 0.00367 | 0.00367 | Combined | 209 |
| NG1827+0636_1435074 | 8.58229 | 0.00993 | 0.00993 | Combined | 210 |
| NG1827+0636_885537 | 8.60776 | 0.00086 | 0.00086 | Combined | 211 |
| NG1827+0636_1429043 | 8.63872 | 0.07396 | 0.07396 | Combined | 212 |
| NG1827+0636_1376142 | 8.81288 | 0.04790 | 0.04790 | Combined | 213 |
| NG1827+0636_868397 | 8.82793 | 0.01365 | 0.01365 | Combined | 214 |
| NG1827+0636_641157 | 8.84373 | 0.05506 | 0.05506 | Combined | 215 |
| NG1827+0636_1378217 | 8.88301 | 0.00000 | 0.00000 | GP | 216 |
| NG1827+0636_869793 | 9.18723 | 0.00441 | 0.00441 | Combined | 217 |
| NG1827+0636_870224 | 9.19814 | 0.01212 | 0.01212 | Combined | 218 |
| NG1827+0636_1258396 | 9.40107 | 0.01105 | 0.01105 | Combined | 219 |
| NG1827+0636_539015 | 9.43472 | 0.03092 | 0.03092 | Combined | 220 |
| NG1827+0636_1764944 | 9.54031 | 0.03821 | 0.03821 | Combined | 221 |
| NG1827+0636_1394490 | 9.57805 | 0.08676 | 0.08676 | Combined | 222 |
| NG1827+0636_1422940 | 10.03600 | 0.10067 | 0.10067 | Combined | 223 |
| NG1827+0636_1453345 | 10.16039 | 0.04369 | 0.04369 | Combined | 224 |
| NG1827+0636_842826 | 10.59263 | 0.00625 | 0.00625 | Combined | 225 |
| NG1827+0636_1191801 | 11.36309 | 0.02239 | 0.02239 | Combined | 226 |
| NG1827+0636_1032588 | 11.43760 | 0.01836 | 0.01836 | Combined | 227 |
| NG1827+0636_1346046 | 11.45982 | 0.09839 | 0.09839 | Combined | 228 |
| NG1827+0636_639925 | 11.57708 | 0.02392 | 0.02392 | Combined | 229 |
| NG1827+0636_1427699 | 12.17651 | 0.06126 | 0.06126 | Combined | 230 |
| NG1827+0636_1233779 | 1.42685 | 0.13324 | 0.13324 | LS | 231 |
| NG1827+0636_1291615 | 2.66976 | 0.00000 | 0.00000 | G-ACF | 232 |
| NG1827+0636_1025611 | 3.66796 | 0.00000 | 0.00000 | G-ACF | 233 |
| NG1827+0636_1401169 | 3.88896 | 0.00629 | 0.00629 | Combined | 234 |
| NG1827+0636_1452024 | 4.14111 | 0.01876 | 0.01876 | Combined | 235 |
| NG1827+0636_1371389 | 7.04782 | 0.10702 | 0.10702 | Combined | 236 |
| NG1827+0636_1321860 | 7.95933 | 0.02446 | 0.02446 | Combined | 237 |
| NG1827+0636_1344805 | 8.23295 | 0.02147 | 0.02147 | Combined | 238 |
| NG1827+0636_1119676 | 9.17334 | 1.45090 | 1.45090 | Combined | 239 |
| NG1827+0636_907345 | 11.36973 | 0.04146 | 0.04146 | Combined | 240 |
| NG1827+0636_1317309 | 12.23621 | 0.00391 | 0.00391 | Combined | 241 |

Table A.2: Estimated variability periods for 58 objects in NGC 6633 from Lomb–Scargle, G-ACF and a Gaussian Process model estimate. The table is continued in Table A.3.

| NGTS ID | $P_{\rm rot}$(LS) | $\Delta P^{-}_{\rm rot}$(LS) | $\Delta P^{+}_{\rm rot}$(LS) | $P_{\rm rot}$(GACF) | $\Delta P^{-}_{\rm rot}$(GACF) | $\Delta P^{+}_{\rm rot}$(GACF) | $P_{\rm rot}$(GP) | $\Delta P^{-}_{\rm rot}$(GP) | $\Delta P^{+}_{\rm rot}$(GP) |
|---|---|---|---|---|---|---|---|---|---|
| NG1827+0636_1321241 | 1.25687 | 2.39501 | 2.39501 | 4.75819 | 0 | 0 | 1.25610 | 0.00547 | 0.00527 |
| NG1827+0636_1451334 | 1.74118 | 3.13413 | 3.13413 | 2.32364 | 0 | 0 | 1.99471 | 1.49629 | 2.65419 |
| NG1827+0636_1100060 | 3.05229 | 1.94325 | 1.94325 | 3.26732 | 0 | 0 | 3.23330 | 0.02376 | 0.02375 |
| NG1827+0636_1794649 | 3.64853 | 5.84466 | 5.84466 | 3.55359 | 0 | 0 | 3.55798 | 0.00938 | 0.00915 |
| NG1827+0636_1097244 | 3.86854 | 2.44568 | 2.44568 | 3.79666 | 0 | 0 | | | |
| NG1827+0636_1008149 | 3.91580 | 1.68223 | 1.68223 | 3.84598 | 0 | 0 | 3.90466 | 0.05660 | 0.05600 |
| NG1827+0636_1439675 | 4.46496 | 0.33848 | 0.33848 | 4.41187 | 0 | 0 | | | |
| NG1827+0636_1422346 | 4.44657 | 4.32100 | 4.32100 | 4.44225 | 0 | 0 | | | |
| NG1827+0636_1319323 | 4.70338 | 0.36841 | 0.36841 | 4.70504 | 0 | 0 | | | |
| NG1827+0636_1439611 | 5.74402 | 0.46440 | 0.46440 | 5.68713 | 0 | 0 | | | |
| NG1827+0636_1452712 | 5.90711 | 5.77667 | 5.77667 | 6.14234 | 0 | 0 | 6.08864 | 0.04824 | 0.04790 |
| NG1827+0636_1010770 | 6.19723 | 1.46872 | 1.46872 | 123.76466 | 0 | 0 | 129.99061 | 22.74655 | 53.92254 |
| NG1827+0636_1329847 | 6.77367 | 0.57694 | 0.57694 | 6.73263 | 0 | 0 | | | |
| NG1827+0636_1328554 | 7.22813 | 0.40286 | 0.40286 | 7.26586 | 0 | 0 | | | |
| NG1827+0636_1327844 | 7.33535 | 0.37706 | 0.37706 | 7.21897 | 0 | 0 | | | |
| NG1827+0636_1041927 | 7.43146 | 1.05312 | 1.05312 | 7.98961 | 0 | 0 | 7.27979 | 0.09911 | 0.10627 |
| NG1827+0636_1300498 | 7.12019 | 7.08664 | 7.08664 | 7.51960 | 0 | 0 | | | |
| NG1827+0636_1106701 | 7.37942 | 0.34794 | 0.34794 | 7.32907 | 0 | 0 | | | |
| NG1827+0636_1027580 | 7.58440 | 0.93082 | 0.93082 | 7.55366 | 0 | 0 | | | |
| NG1827+0636_1769107 | 7.56001 | 0.19989 | 0.19989 | 7.58902 | 0 | 0 | | | |
| NG1827+0636_558862 | 7.85904 | 7.41169 | 7.41169 | 7.57762 | 0 | 0 | | | |
| NG1827+0636_958489 | 7.75533 | 6.50077 | 6.50077 | 7.70028 | 0 | 0 | | | |
| NG1827+0636_849720 | 8.66590 | 0.71198 | 0.71198 | 8.19985 | 0 | 0 | | | |
| NG1827+0636_2096690 | 8.54310 | 0.33157 | 0.33157 | 8.34545 | 0 | 0 | | | |
| NG1827+0636_1447254 | 8.48052 | 8.14058 | 8.14058 | 8.47099 | 0 | 0 | | | |
| NG1827+0636_1321895 | 8.54092 | 0.20670 | 0.20670 | 8.55130 | 0 | 0 | | | |
| NG1827+0636_1435074 | 8.59633 | 0.23388 | 0.23388 | 8.56825 | 0 | 0 | | | |
| NG1827+0636_885537 | 8.60898 | 4.92148 | 4.92148 | 8.60654 | 0 | 0 | | | |
| NG1827+0636_1429043 | 8.53413 | 1.24647 | 1.24647 | 8.74331 | 0 | 0 | | | |

Table A.3: (Continued from Table A.2.)

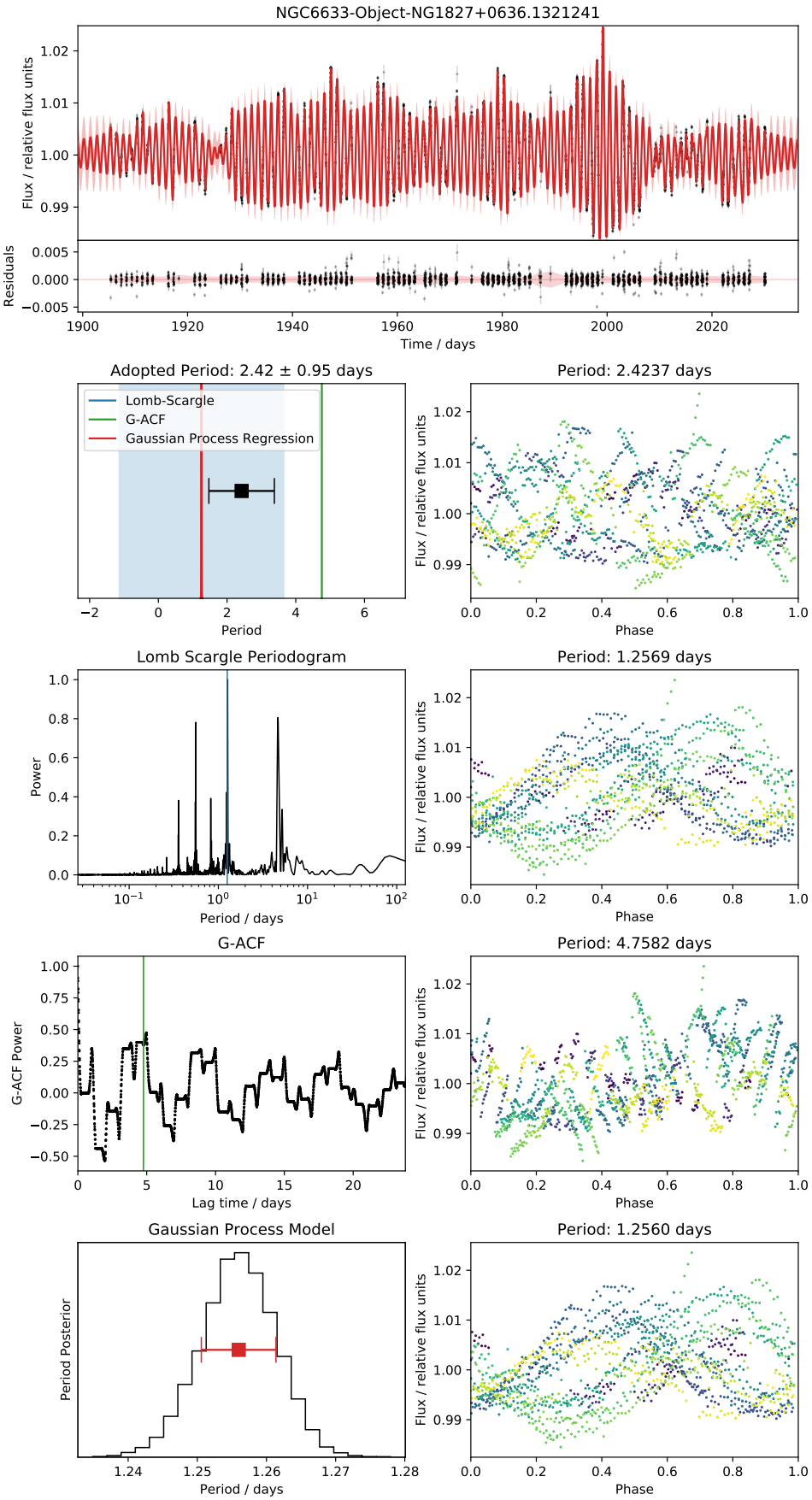
| NGTS ID | $P_{\rm rot}$(LS) | $\Delta P^{-}_{\rm rot}$(LS) | $\Delta P^{+}_{\rm rot}$(LS) | $P_{\rm rot}$(GACF) | $\Delta P^{-}_{\rm rot}$(GACF) | $\Delta P^{+}_{\rm rot}$(GACF) | $P_{\rm rot}$(GP) | $\Delta P^{-}_{\rm rot}$(GP) | $\Delta P^{+}_{\rm rot}$(GP) |
|---|---|---|---|---|---|---|---|---|---|
| NG1827+0636_868397 | 8.80863 | 0.39164 | 0.39164 | 8.84723 | 0 | 0 | | | |
| NG1827+0636_641157 | 8.76587 | 1.09832 | 1.09832 | 8.92160 | 0 | 0 | | | |
| NG1827+0636_1378217 | 8.65059 | 7.76979 | 7.76979 | 8.97982 | 0 | 0 | 8.88301 | 0.00000 | 0.00000 |
| NG1827+0636_869793 | 9.19346 | 0.57445 | 0.57445 | 9.18099 | 0 | 0 | | | |
| NG1827+0636_870224 | 9.21529 | 0.56225 | 0.56225 | 9.18099 | 0 | 0 | | | |
| NG1827+0636_1258396 | 9.38544 | 0.39442 | 0.39442 | 9.41669 | 0 | 0 | | | |
| NG1827+0636_539015 | 9.39099 | 2.33039 | 2.33039 | 9.47845 | 0 | 0 | | | |
| NG1827+0636_1764944 | 9.59434 | 0.95097 | 0.95097 | 9.48627 | 0 | 0 | | | |
| NG1827+0636_1394490 | 9.50376 | 5.40119 | 5.40119 | 9.78762 | 0 | 0 | 9.44277 | 0.09727 | 0.10986 |
| NG1827+0636_1422940 | 9.79105 | 3.05091 | 3.05091 | 10.18317 | 0 | 0 | 10.13377 | 0.17834 | 0.19019 |
| NG1827+0636_1453345 | 10.22217 | 0.12857 | 0.12857 | 10.09860 | 0 | 0 | | | |
| NG1827+0636_842826 | 10.58380 | 0.41146 | 0.41146 | 10.60147 | 0 | 0 | | | |
| NG1827+0636_1191801 | 11.33143 | 0.36395 | 0.36395 | 11.39476 | 0 | 0 | | | |
| NG1827+0636_1032588 | 11.46357 | 0.47101 | 0.47101 | 11.41164 | 0 | 0 | | | |
| NG1827+0636_1346046 | 11.22286 | 5.71205 | 5.71205 | 11.54030 | 0 | 0 | 11.61631 | 0.15190 | 0.16374 |
| NG1827+0636_639925 | 11.61091 | 0.34615 | 0.34615 | 11.54325 | 0 | 0 | | | |
| NG1827+0636_1427699 | 12.26314 | 0.53292 | 0.53292 | 12.08988 | 0 | 0 | | | |
| NG1827+0636_1233779 | 1.42685 | 0.13324 | 0.13324 | 3.45437 | 0 | 0 | 2.83745 | 0.00593 | 0.00567 |
| NG1827+0636_1291615 | 2.61056 | 1.07780 | 1.07780 | 2.66976 | 0 | 0 | 0.50142 | 0.00000 | 0.00000 |
| NG1827+0636_1025611 | 1.86713 | 3.02680 | 3.02680 | 3.66796 | 0 | 0 | | | |
| NG1827+0636_1401169 | 3.88007 | 1.47594 | 1.47594 | 3.89785 | 0 | 0 | | | |
| NG1827+0636_1452024 | 4.11457 | 1.48303 | 1.48303 | 4.16764 | 0 | 0 | | | |
| NG1827+0636_1371389 | 6.89647 | 3.81839 | 3.81839 | 7.19917 | 0 | 0 | | | |
| NG1827+0636_1321860 | 7.99392 | 0.65009 | 0.65009 | 7.92475 | 0 | 0 | | | |
| NG1827+0636_1344805 | 8.18387 | 0.18470 | 0.18470 | 8.24113 | 0 | 0 | 8.27385 | 0.04947 | 0.05113 |
| NG1827+0636_1119676 | 10.80983 | 10.52993 | 10.52993 | 11.08720 | 0 | 0 | 5.62299 | 0.00000 | 0.00000 |
| NG1827+0636_907345 | 11.42836 | 0.45170 | 0.45170 | 11.31109 | 0 | 0 | | | |
| NG1827+0636_1317309 | 12.24174 | 0.80835 | 0.80835 | 12.23067 | 0 | 0 | | | |

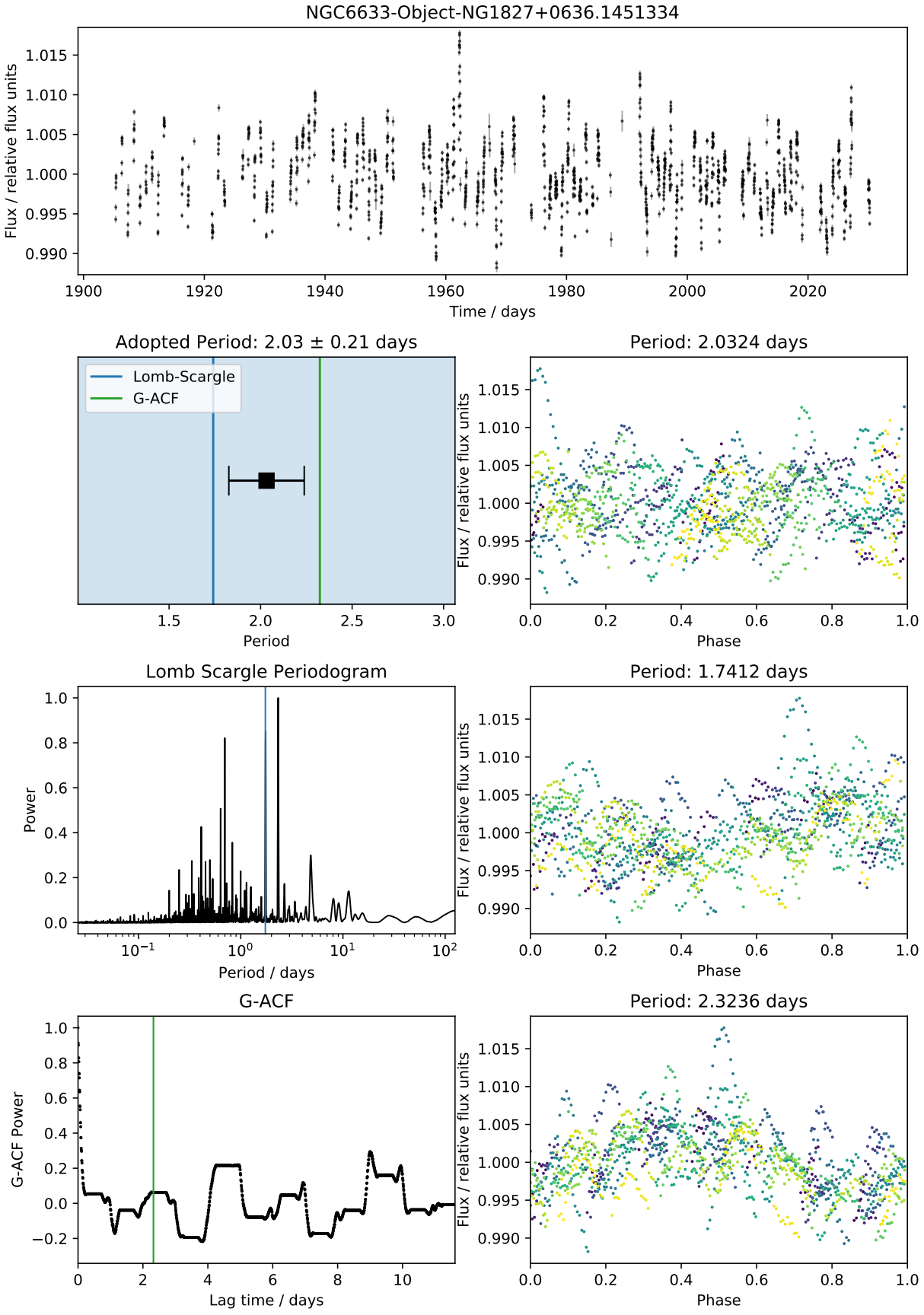
Table A.4: Stellar parameters and cross-match identifiers for 58 objects in NGC 6633. A subset of the table columns have been printed for publication clarity. The table is continued in Table A.5.

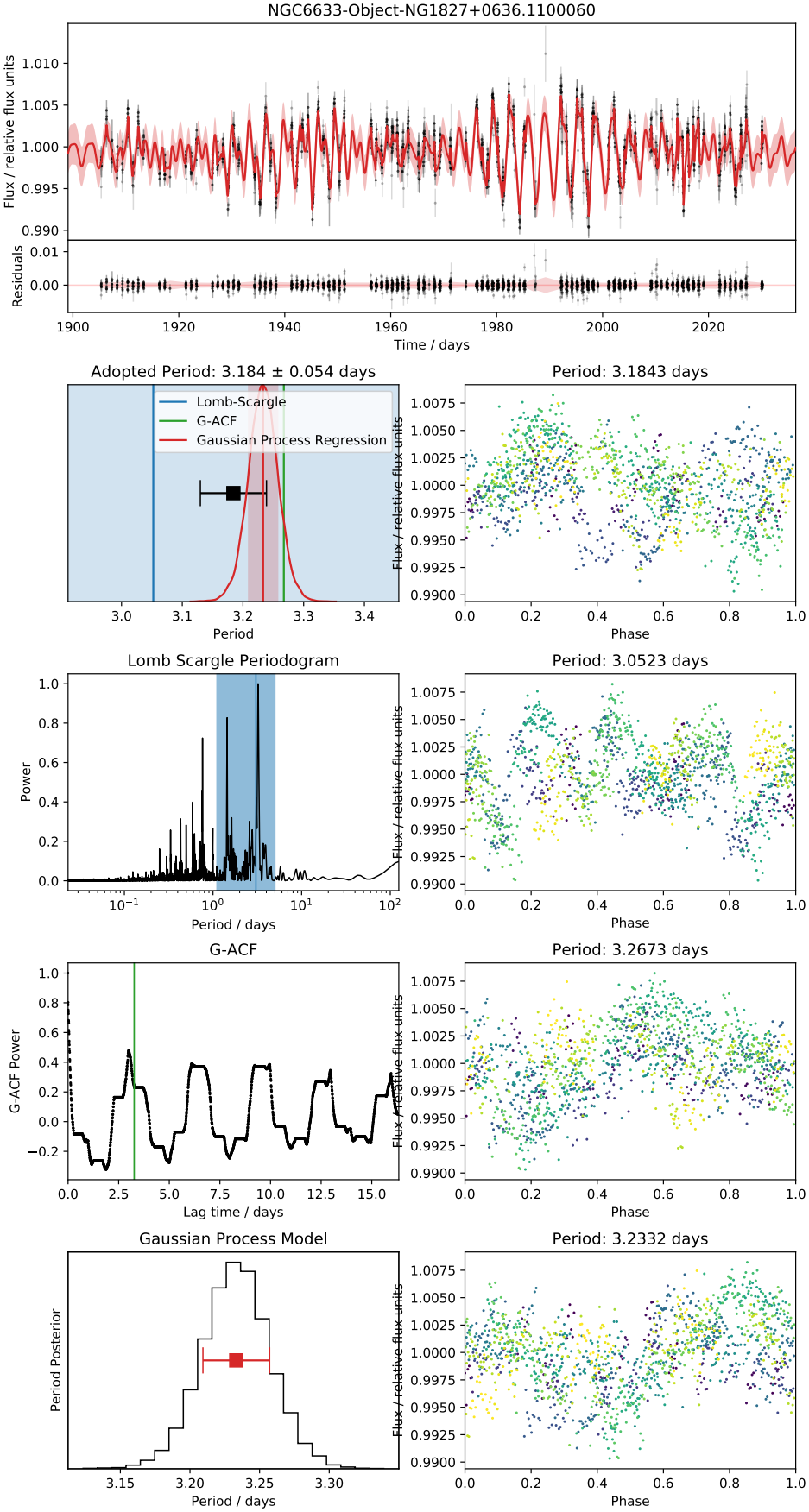
| NGTS ID | RA | Dec | G Mag | $A_0$ | Distance | Binary | Amplitude | Gaia EDR3 ID | TICv8 ID |
|---|---|---|---|---|---|---|---|---|---|
| NG1827+0636_1321241 | 276.83235 | 6.85826 | 11.32500 | 0.56808 | 391.96335 | 0 | 0.02219 | 4477266156402806784 | 319341055 |
| NG1827+0636_1451334 | 276.73081 | 6.46595 | 10.94568 | 0.35625 | 387.23465 | 0 | 0.01420 | 4477215544507181568 | 414958614 |
| NG1827+0636_1100060 | 277.47710 | 6.09102 | 12.60373 | 0.73384 | 407.12855 | 0 | 0.01149 | 4284992736985442304 | 320631423 |
| NG1827+0636_1794649 | 276.66243 | 7.74214 | 16.10443 | 0.35625 | 404.07005 | 0 | 0.03796 | 4477642120658537984 | 1677163143 |
| NG1827+0636_1097244 | 277.18675 | 6.39579 | 12.52017 | 0.57548 | 393.70056 | 0 | 0.01572 | 4477158399947061760 | 320152834 |
| NG1827+0636_1008149 | 278.13127 | 5.87808 | 13.46705 | 0.79472 | 480.63865 | 0 | 0.01800 | 4284935905976928256 | 371346403 |
| NG1827+0636_1439675 | 276.45135 | 6.86242 | 14.94039 | 0.41106 | 518.05328 | 0 | 0.01429 | 4477436271461228032 | 25747566 |
| NG1827+0636_1422346 | 276.60439 | 6.50090 | 16.31666 | 0.52068 | 390.25558 | 0 | 0.05904 | 4477219392798051328 | 26026946 |
| NG1827+0636_1319323 | 277.01167 | 6.64042 | 12.58592 | 0.52068 | 390.23586 | 0 | 0.01381 | 4477247773942721536 | 319768780 |
| NG1827+0636_1439611 | 276.47980 | 6.84072 | 12.75196 | 0.42140 | 384.09316 | 0 | 0.00909 | 4477433247804199424 | 25747465 |
| NG1827+0636_1452712 | 276.69507 | 6.50287 | 12.80034 | 0.40098 | 393.91653 | 0 | 0.01275 | 4477218838735244800 | 414958465 |
| NG1827+0636_1010770 | 277.17824 | 7.08946 | 12.94473 | 0.49327 | 390.63119 | 0 | 0.01247 | 4477370811849637888 | 320145514 |
| NG1827+0636_1329847 | 276.93383 | 6.88011 | 13.08959 | 0.56982 | 393.40048 | 0 | 0.01151 | 4477267599512021504 | 319708851 |
| NG1827+0636_1328554 | 276.84236 | 6.90705 | 12.87522 | 0.46587 | 383.82754 | 0 | 0.01908 | 4477271001126161408 | 319341296 |
| NG1827+0636_1327844 | 276.97754 | 6.92862 | 13.72151 | 0.52068 | 384.61132 | 0 | 0.01563 | 4477274089194001408 | 319708610 |
| NG1827+0636_1041927 | 277.57377 | 6.53746 | 13.87438 | 0.63029 | 391.97390 | 0 | 0.01874 | 4477186819764957696 | 320634537 |
| NG1827+0636_1300498 | 276.54043 | 7.02682 | 13.03638 | 0.43846 | 388.93988 | 0 | 0.01494 | 4477463450015170560 | 26024570 |
| NG1827+0636_1106701 | 277.36945 | 6.24886 | 13.90143 | 0.74595 | 403.19491 | 0 | 0.02279 | 4284998887379010048 | 320159454 |
| NG1827+0636_1027580 | 277.19727 | 6.69302 | 13.38575 | 0.40619 | 394.05667 | 0 | 0.01630 | 4477202311692687360 | 320149744 |
| NG1827+0636_1769107 | 277.10842 | 7.39866 | 13.72388 | 0.43846 | 389.44539 | 0 | 0.02196 | 4477570514961452032 | 319774610 |
| NG1827+0636_558862 | 276.47947 | 5.57290 | 13.42785 | 0.60289 | 378.13901 | 0 | 0.01257 | 4476727941145239040 | 170038531 |
| NG1827+0636_958489 | 276.07148 | 6.11632 | 15.60070 | 0.65770 | 1020.99258 | 0 | 0.03545 | 4476818169823058944 | 25547760 |
| NG1827+0636_849720 | 276.82120 | 5.95099 | 14.66889 | 0.49327 | 405.41464 | 0 | 0.02408 | 4476767699673917952 | 415076095 |
| NG1827+0636_2096690 | 277.33510 | 7.17360 | 14.29588 | 0.38366 | 374.81009 | 0 | 0.02246 | 4477372534154684416 | 320169391 |
| NG1827+0636_1447254 | 276.74850 | 6.64141 | 14.16443 | 0.43846 | 391.79277 | 0 | 0.01868 | 4477238428093129216 | 414957895 |
| NG1827+0636_1321895 | 276.85330 | 6.81253 | 13.77254 | 0.39864 | 387.85112 | 0 | 0.02045 | 4477265778445672960 | 319340834 |
| NG1827+0636_1435074 | 276.89305 | 6.57332 | 14.30510 | 0.52068 | 387.33913 | 0 | 0.01794 | 4477223550326185984 | 167684274 |
| NG1827+0636_885537 | 276.70398 | 5.98499 | 14.04281 | 0.54808 | 404.59781 | 0 | 0.01616 | 4476768627386930176 | 167654172 |
| NG1827+0636_1429043 | 276.93956 | 6.60999 | 13.47312 | 0.53916 | 390.71403 | 0 | 0.01400 | 4477224100081995776 | 319710165 |

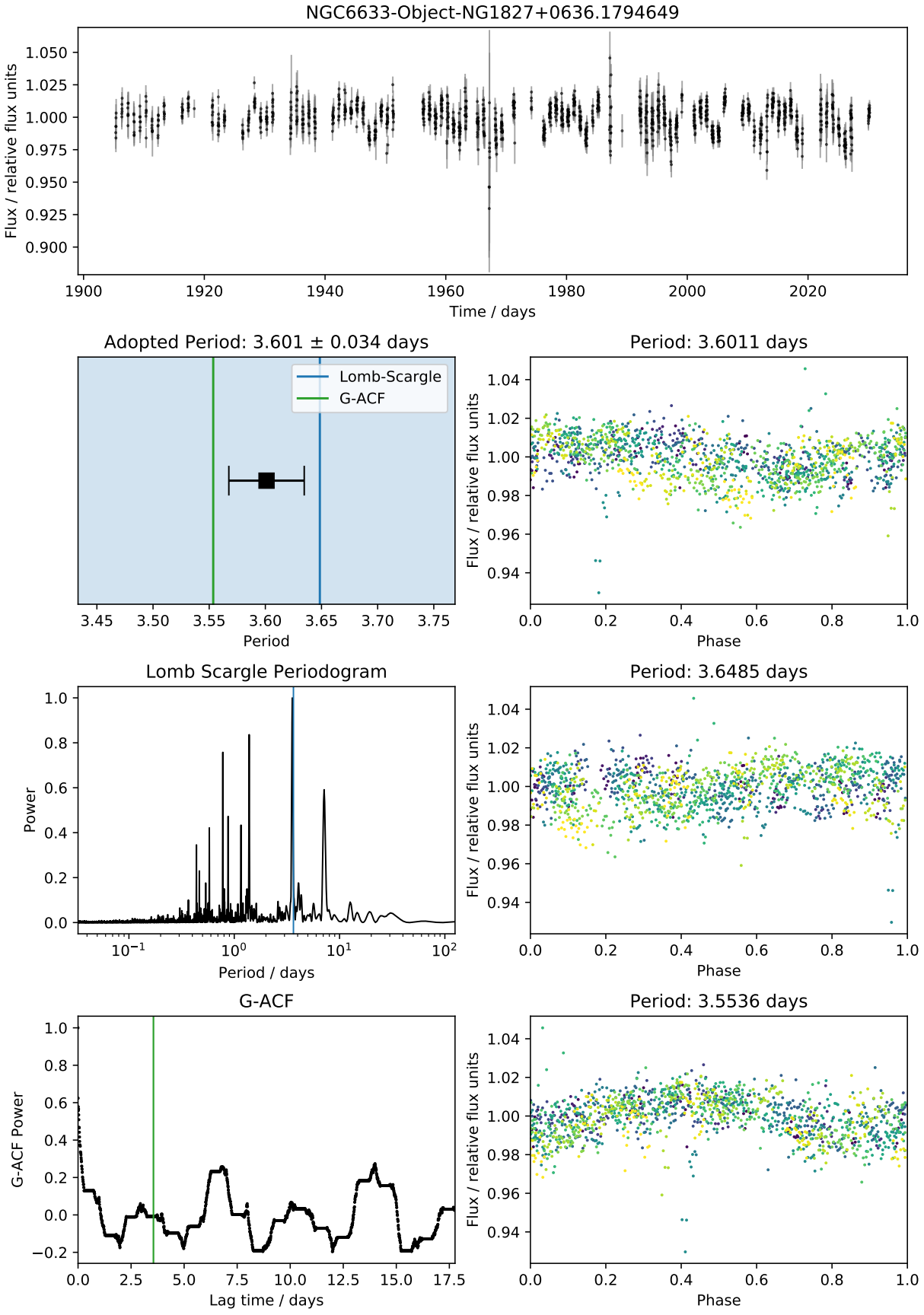
Table A.5: (Continued from Table A.4.)

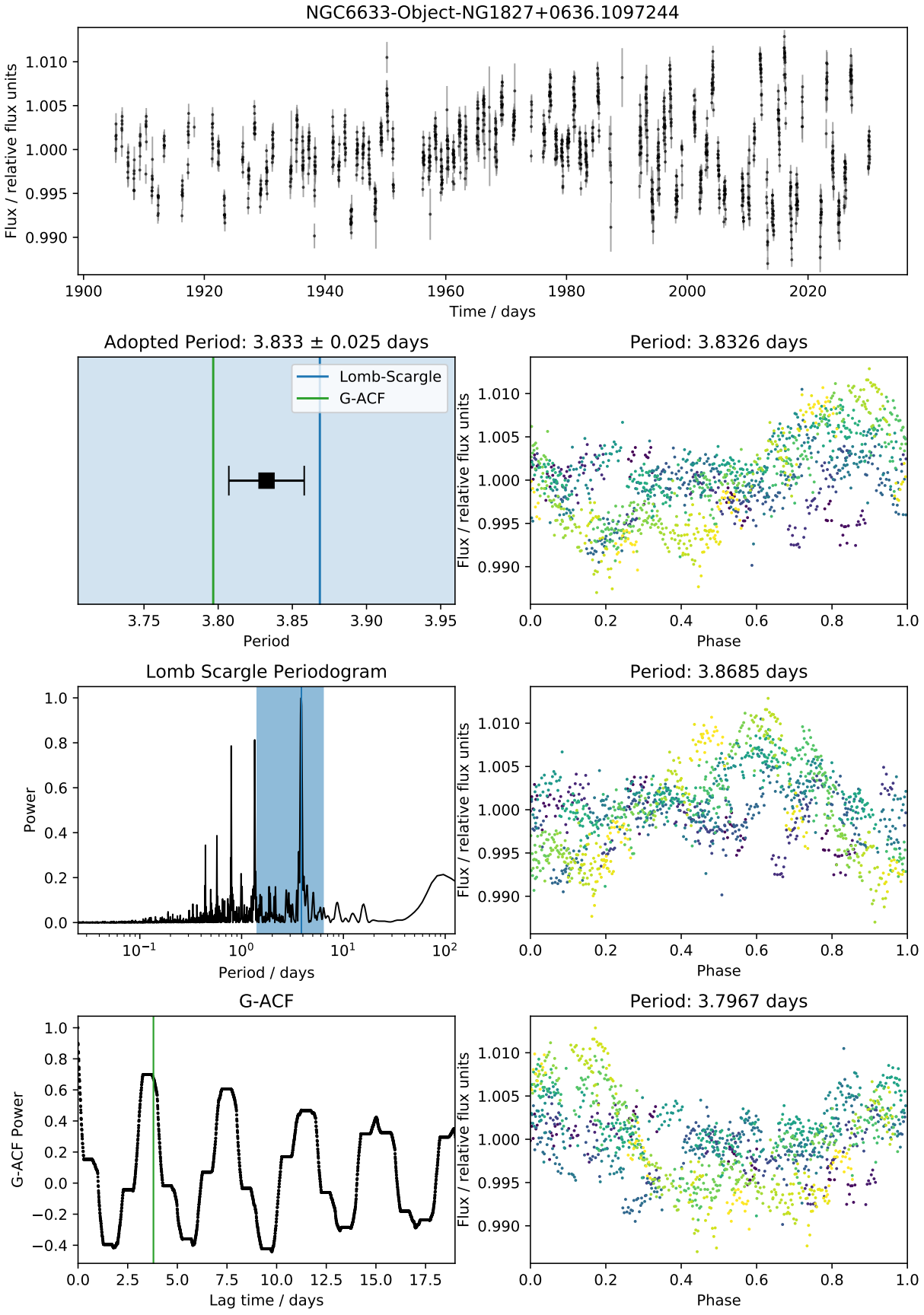
| NGTS ID | RA | Dec | G Mag | $A_0$ | Distance | Binary | Amplitude | Gaia EDR3 ID | TICv8 ID |
|---|---|---|---|---|---|---|---|---|---|
| NG1827+0636_868397 | 276.49618 | 5.70217 | 13.14814 | 0.62988 | 397.75949 | 0 | 0.02000 | 4476734233288121856 | 170039786 |
| NG1827+0636_641157 | 277.62850 | 5.21160 | 14.17656 | 0.74275 | 378.08550 | 0 | 0.01878 | 4284326707816352768 | 404856397 |
| NG1827+0636_1378217 | 276.87885 | 6.00200 | 14.05823 | 0.54808 | 396.61832 | 0 | 0.01577 | 4476769486380313600 | 415380280 |
| NG1827+0636_869793 | 276.49196 | 5.74903 | 14.17095 | 0.59506 | 391.61327 | 0 | 0.01737 | 4476735435878983168 | 170040275 |
| NG1827+0636_870224 | 276.49186 | 5.74986 | 15.39047 | 0.59507 | 391.62027 | 0 | 0.02205 | 4476735435867877376 | 1676517940 |
| NG1827+0636_1258396 | 276.33414 | 7.34938 | 14.21571 | 0.34362 | 387.68541 | 0 | 0.02339 | 4477534643395003392 | 25661732 |
| NG1827+0636_539015 | 277.05452 | 5.34018 | 14.46555 | 0.87886 | 598.07846 | 0 | 0.02247 | 4284506134368316416 | 415383151 |
| NG1827+0636_1764944 | 277.12697 | 7.28314 | 13.77982 | 0.49327 | 386.77430 | 0 | 0.01154 | 4477565494129081344 | 319774064 |
| NG1827+0636_1394490 | 276.55874 | 6.32616 | 14.29897 | 0.43128 | 395.11513 | 0 | 0.01751 | 4476840267443301376 | 26027722 |
| NG1827+0636_1422940 | 276.50520 | 6.53974 | 13.15397 | 0.38366 | 395.42900 | 0 | 0.01318 | 4477230692842056704 | 25746131 |
| NG1827+0636_1453345 | 276.73117 | 6.51187 | 14.48681 | 0.46587 | 390.20313 | 0 | 0.03153 | 4477218907451595776 | 414958433 |
| NG1827+0636_842826 | 276.01531 | 5.81078 | 13.74564 | 0.16442 | 281.78407 | 0 | 0.01959 | 4476792507405858304 | 169359509 |
| NG1827+0636_1191801 | 275.83430 | 7.68878 | 16.04301 | 0.43846 | 365.79650 | 0 | 0.03947 | 4478320274506277376 | 319003107 |
| NG1827+0636_1032588 | 277.27183 | 6.49468 | 15.18019 | 0.52401 | 378.47693 | 0 | 0.03468 | 4477194447626711040 | 320162057 |
| NG1827+0636_1346046 | 276.21393 | 6.63765 | 15.21300 | 0.21923 | 370.49096 | 0 | 0.01427 | 4477051408031741952 | 414817093 |
| NG1827+0636_639925 | 277.77033 | 5.28017 | 15.41872 | 1.06202 | 472.81685 | 0 | 0.03938 | 4284329044278504448 | 370256692 |
| NG1827+0636_1427699 | 276.81451 | 6.68094 | 15.47711 | 0.34308 | 387.24461 | 0 | 0.02677 | 4477228360689766400 | 319340171 |
| NG1827+0636_1233779 | 276.70201 | 7.13871 | 12.06443 | 0.43846 | 397.76036 | 1 | 0.02279 | 4477466022687624192 | 414955846 |
| NG1827+0636_1291615 | 276.97616 | 7.01612 | 14.52250 | 0.43846 | 385.26114 | 1 | 0.03502 | 4477277941793605632 | 1676882623 |
| NG1827+0636_1025611 | 277.37405 | 6.53937 | 14.90309 | 0.63029 | 401.86635 | 1 | 0.04763 | 4477184723821187584 | 320521170 |
| NG1827+0636_1401169 | 276.62705 | 6.34842 | 12.27502 | 0.41106 | 396.66303 | 1 | 0.01445 | 4477212761368341504 | 26027621 |
| NG1827+0636_1452024 | 276.72816 | 6.55827 | 12.05453 | 0.38366 | 405.54492 | 1 | 0.01193 | 4477225676323174912 | 414958238 |
| NG1827+0636_1371389 | 277.09436 | 5.83923 | 13.13597 | 0.68510 | 396.63868 | 1 | 0.01351 | 4284606842764014080 | 415380975 |
| NG1827+0636_1321860 | 276.87528 | 6.79426 | 13.60107 | 0.41092 | 397.99392 | 1 | 0.02061 | 4477264953811943936 | 319340743 |
| NG1827+0636_1344805 | 276.41318 | 6.65606 | 13.61859 | 0.41075 | 375.59065 | 1 | 0.01887 | 4477423180400704512 | 25746616 |
| NG1827+0636_1119676 | 277.82113 | 6.34166 | 14.81098 | 0.60289 | 398.25100 | 1 | 0.01909 | 4285014486700919296 | 405008485 |
| NG1827+0636_907345 | 276.37611 | 6.42734 | 15.01244 | 0.38366 | 412.07283 | 1 | 0.02400 | 4476854213190251520 | 25742487 |
| NG1827+0636_1317309 | 277.02126 | 6.68363 | 12.70012 | 0.54808 | 394.02860 | 1 | 0.01164 | 4477248289338945536 | 319769244 |

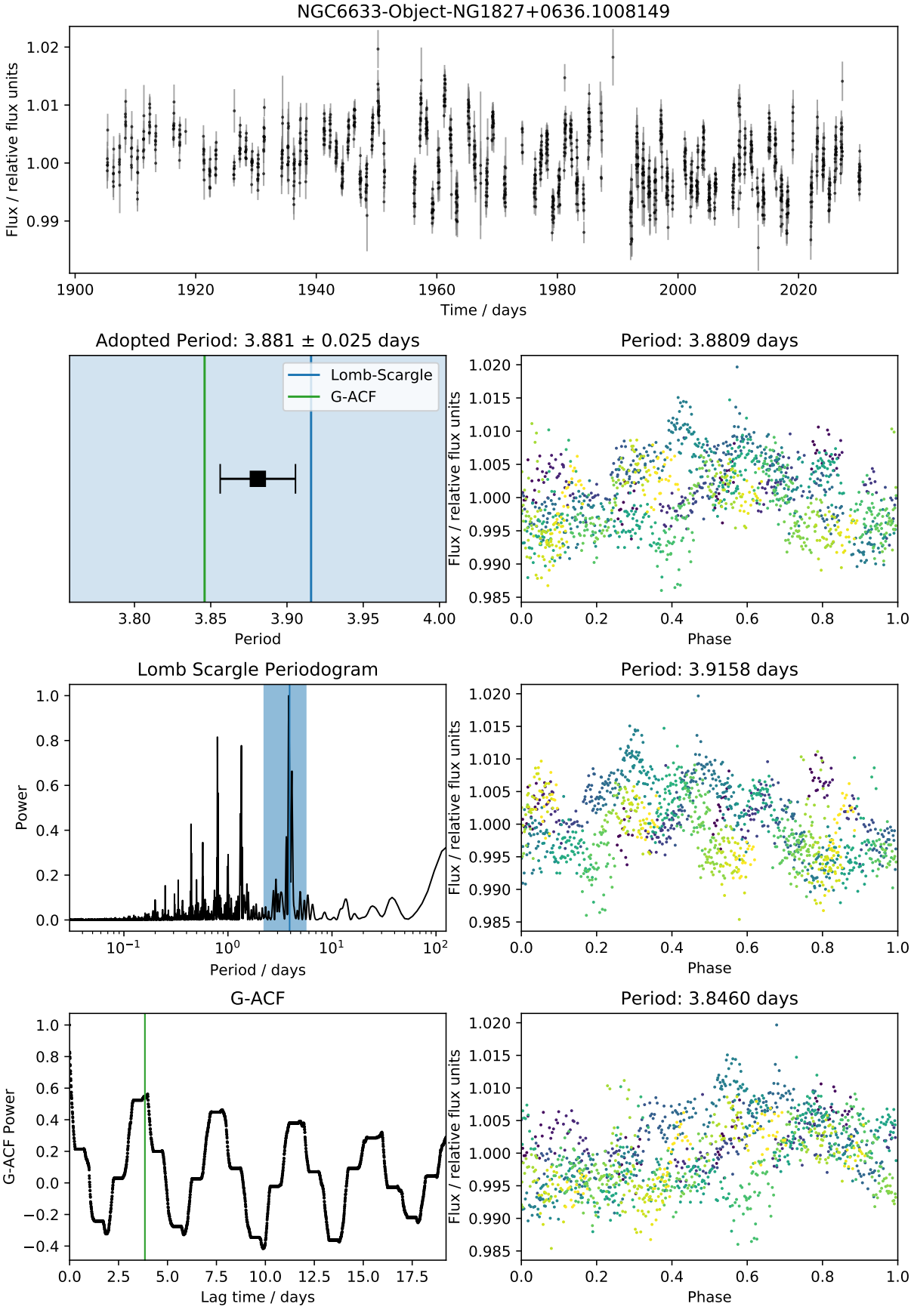

NGC6633-Object-NG1827+0636.1321241
Flux / relative flux units
Residuals
Time / days
Adopted Period: 2.42 ± 0.95 days
Lomb-Scargle
G-ACF
Gaussian Process Regression
Period
Period: 2.4237 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 1.2569 days
G-ACF
G-ACF Power
Lag time / days
Period: 4.7582 days
Gaussian Process Model
Period Posterior
Period / days
Period: 1.2560 days

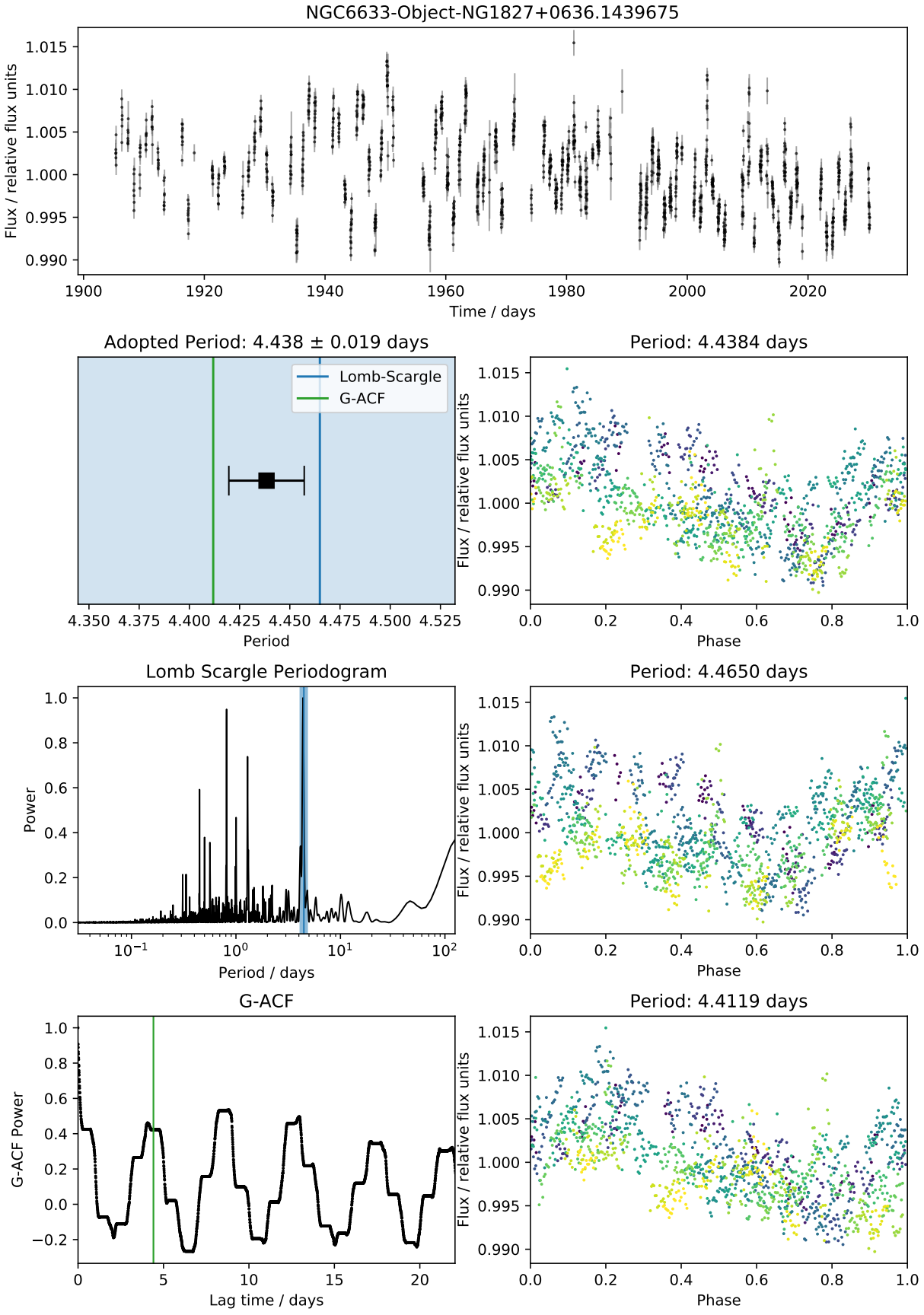

NGC6633-Object-NG1827+0636.1451334
Flux / relative flux units
Time / days
Adopted Period: 2.03 ± 0.21 days
Lomb-Scargle
G-ACF
Period
Period: 2.0324 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 1.7412 days
Phase
G-ACF
G-ACF Power
Lag time / days
Period: 2.3236 days
Phase

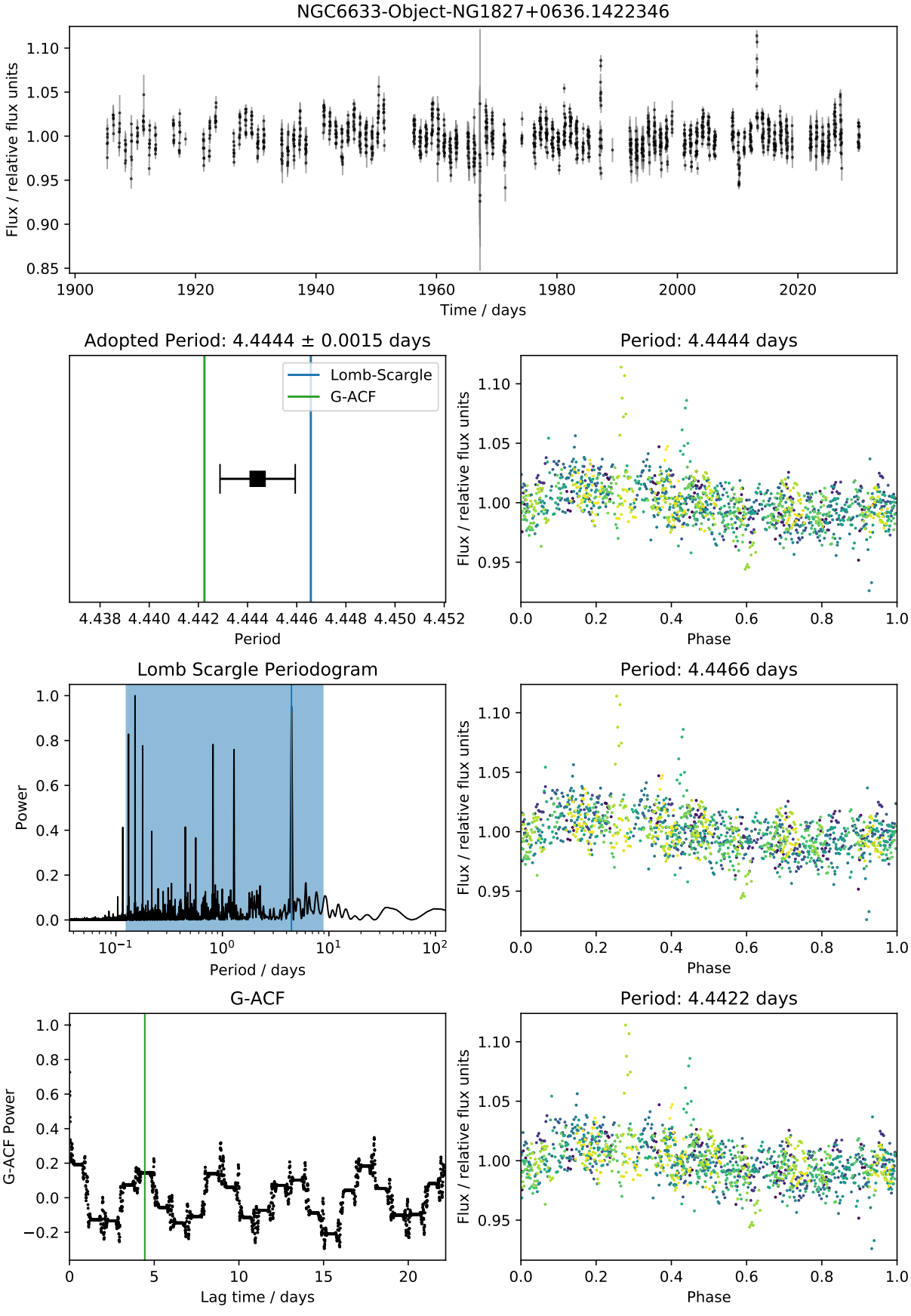

NGC6633-Object-NG1827+0636.1100060
Flux / relative flux units
Residuals
Time / days
Adopted Period: 3.184 ± 0.054 days
Lomb-Scargle
G-ACF
Gaussian Process Regression
Period
Period: 3.1843 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 3.0523 days
G-ACF
G-ACF Power
Lag time / days
Period: 3.2673 days
Gaussian Process Model
Period Posterior
Period / days
Period: 3.2332 days

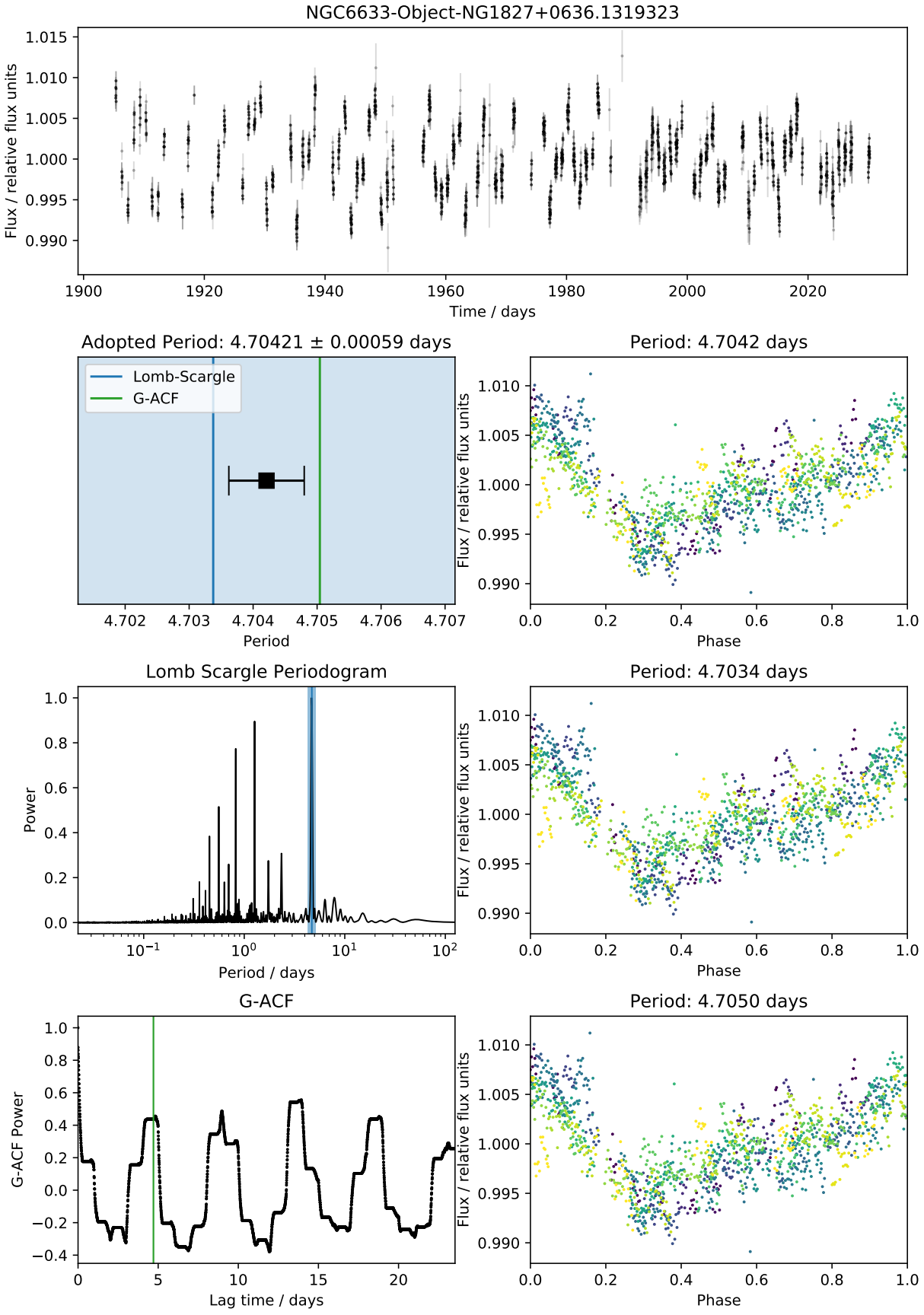

NGC6633-Object-NG1827+0636.1794649
Flux / relative flux units
Time / days
Adopted Period: 3.601 ± 0.034 days
Lomb-Scargle
G-ACF
Period
Period: 3.6011 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 3.6485 days
G-ACF
G-ACF Power
Lag time / days
Period: 3.5536 days

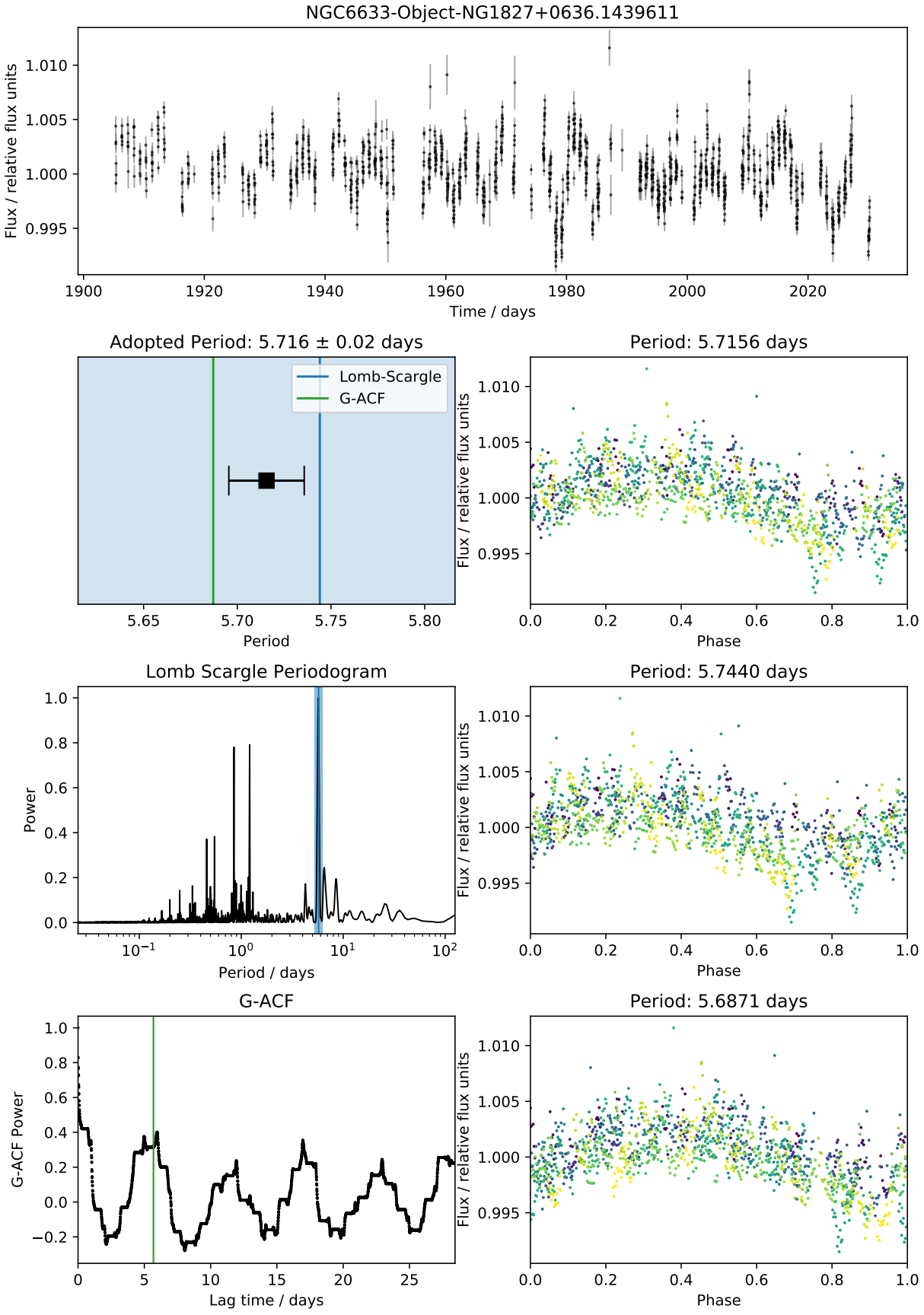

NGC6633-Object-NG1827+0636.1097244
Flux / relative flux units
Time / days
Adopted Period: 3.833 ± 0.025 days
Lomb-Scargle
G-ACF
Period
Period: 3.8326 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 3.8685 days
Phase
G-ACF
G-ACF Power
Lag time / days
Period: 3.7967 days
Phase

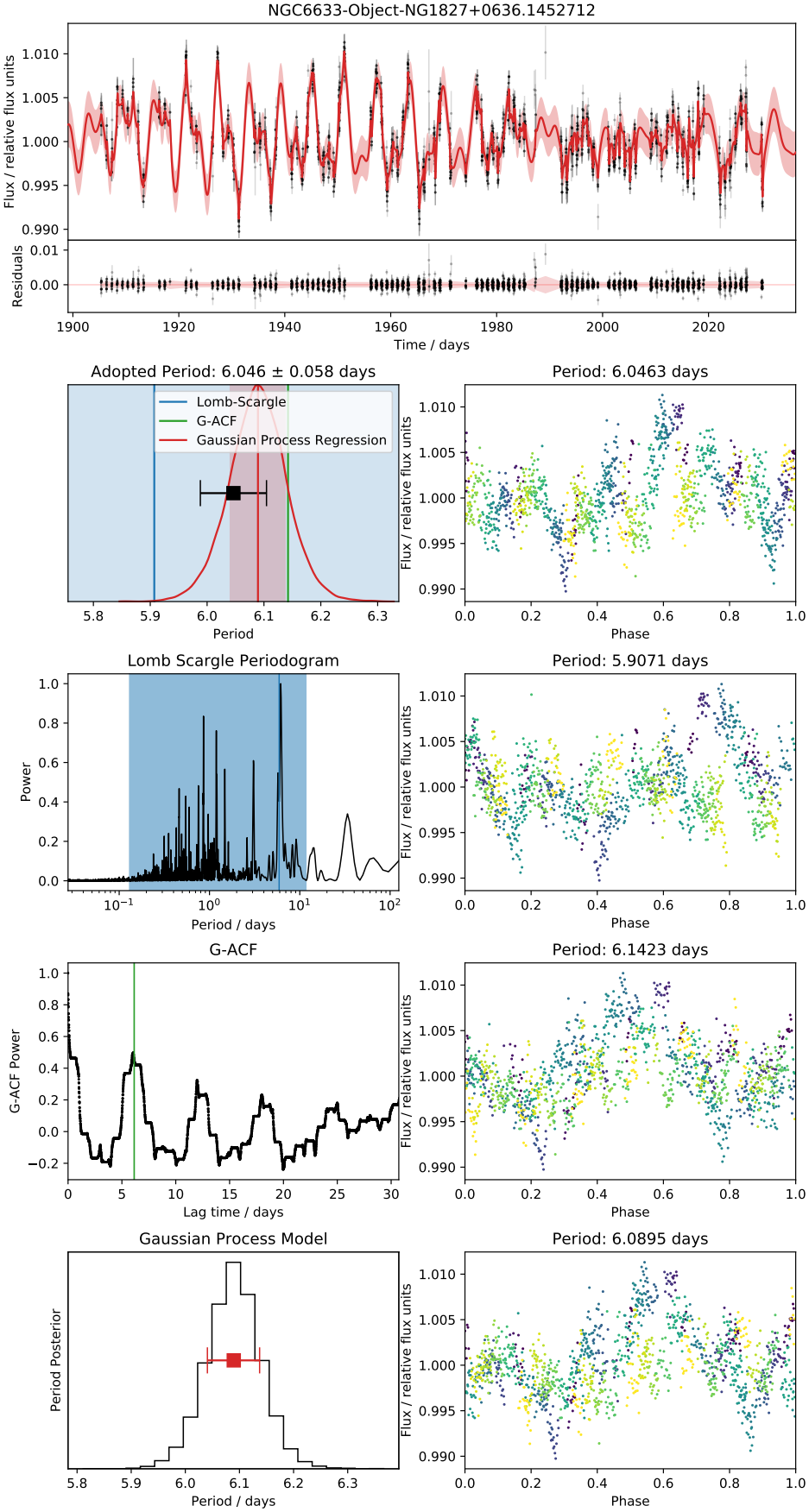

NGC6633-Object-NG1827+0636.1008149
Flux / relative flux units
Time / days
Adopted Period: 3.881 ± 0.025 days
Lomb-Scargle
G-ACF
Period
Period: 3.8809 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 3.9158 days
G-ACF
G-ACF Power
Lag time / days
Period: 3.8460 days

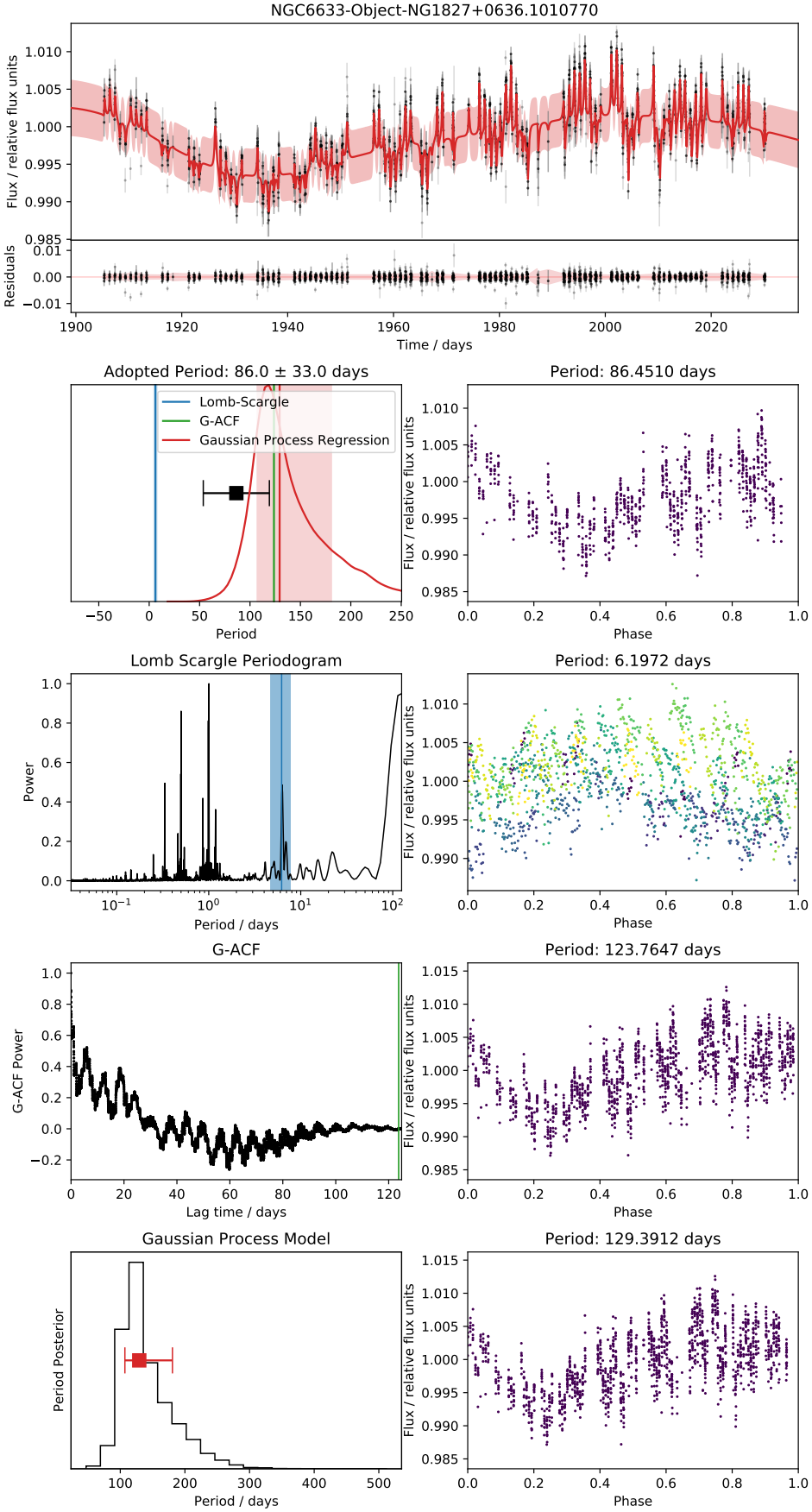

NGC6633-Object-NG1827+0636.1439675
Flux / relative flux units
Time / days
Adopted Period: 4.438 ± 0.019 days
Lomb-Scargle
G-ACF
Period
Period: 4.4384 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 4.4650 days
G-ACF
G-ACF Power
Lag time / days
Period: 4.4119 days

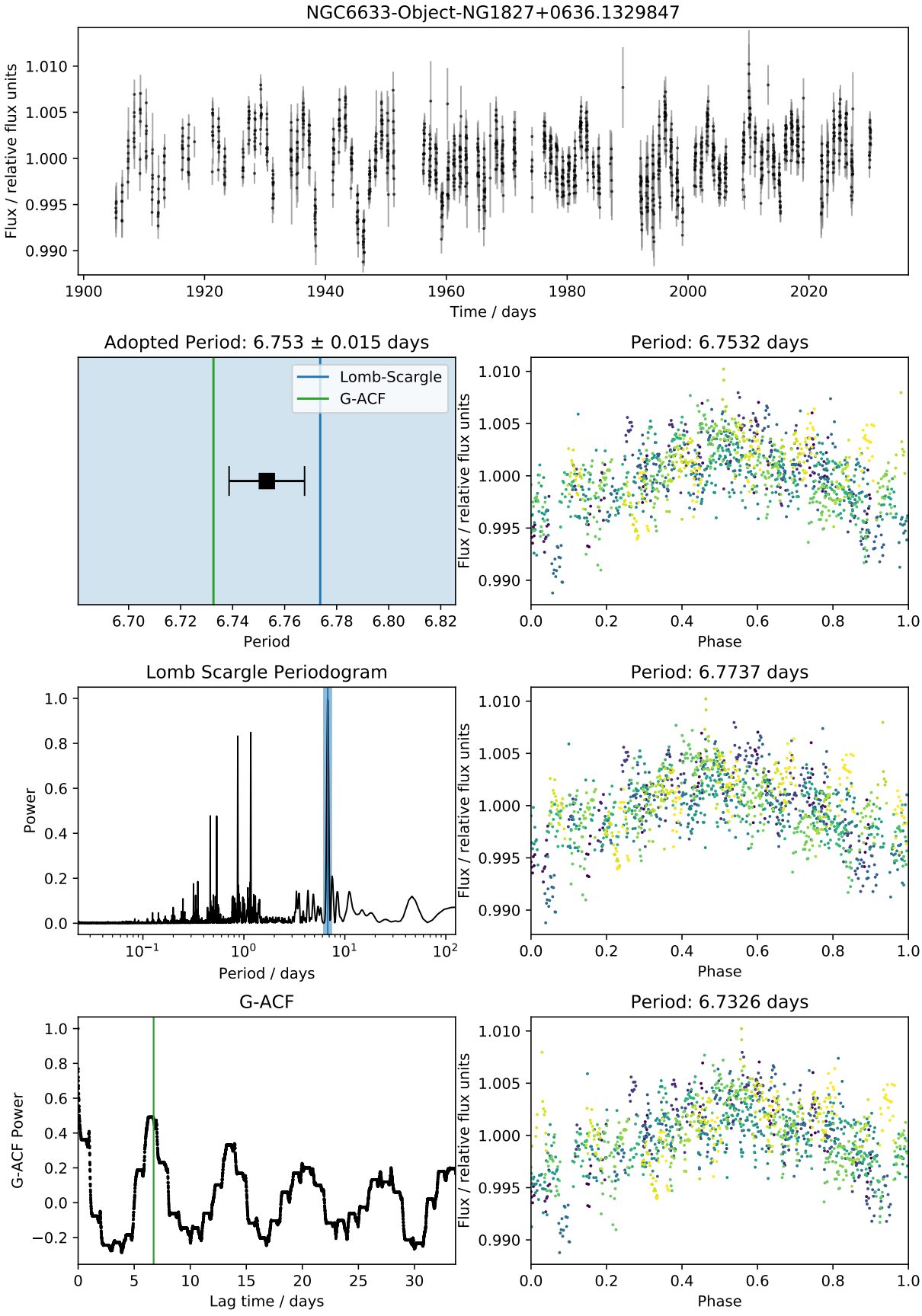

NGC6633-Object-NG1827+0636.1422346
Flux / relative flux units
Time / days
Adopted Period: 4.4444 ± 0.0015 days
Lomb-Scargle
G-ACF
Period
Period: 4.4444 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 4.4466 days
G-ACF
G-ACF Power
Lag time / days
Period: 4.4422 days

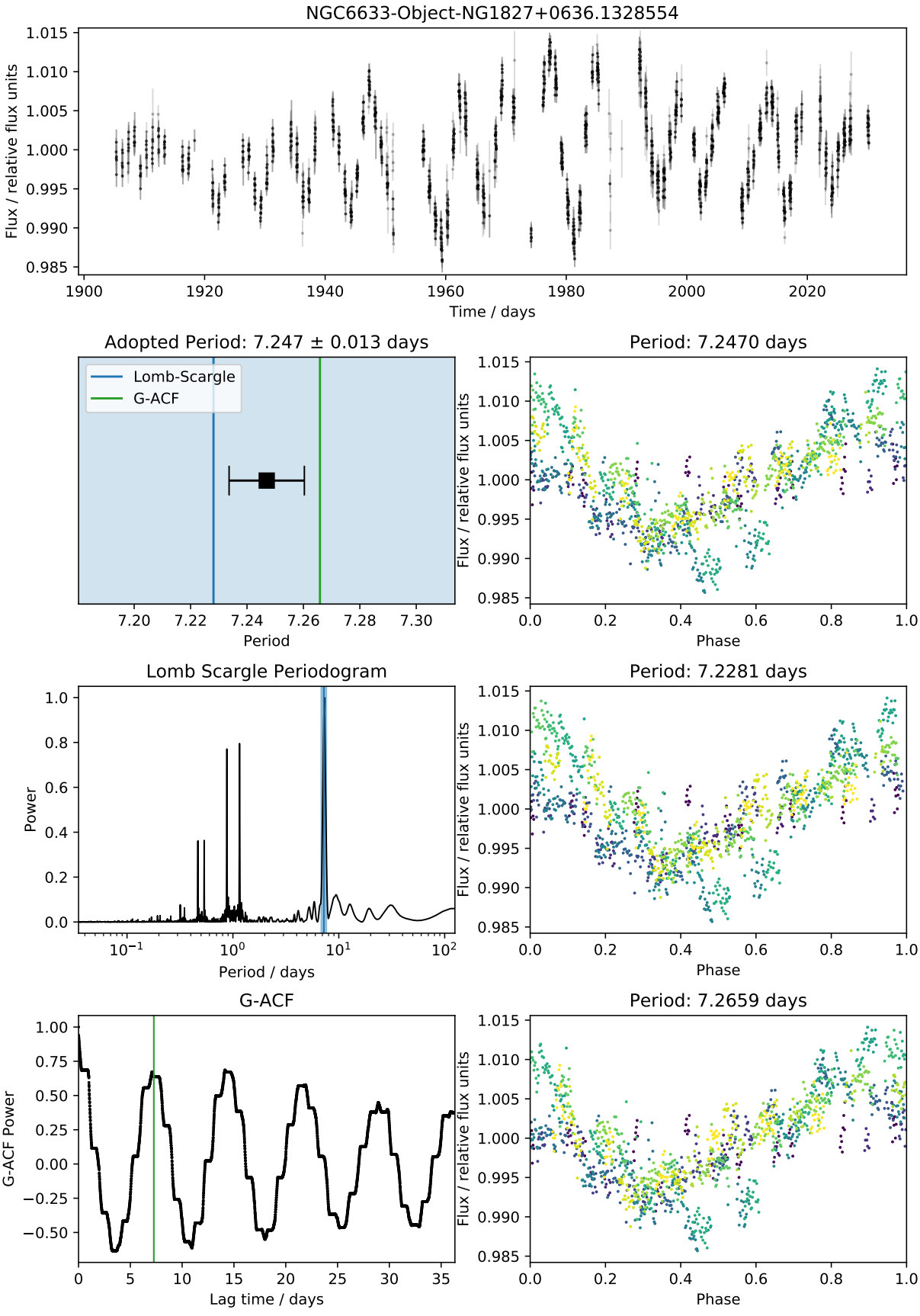

NGC6633-Object-NG1827+0636.1319323
Flux / relative flux units
Time / days
Adopted Period: 4.70421 ± 0.00059 days
Lomb-Scargle
G-ACF
Period
Period: 4.7042 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 4.7034 days
Phase
G-ACF
G-ACF Power
Lag time / days
Period: 4.7050 days
Phase

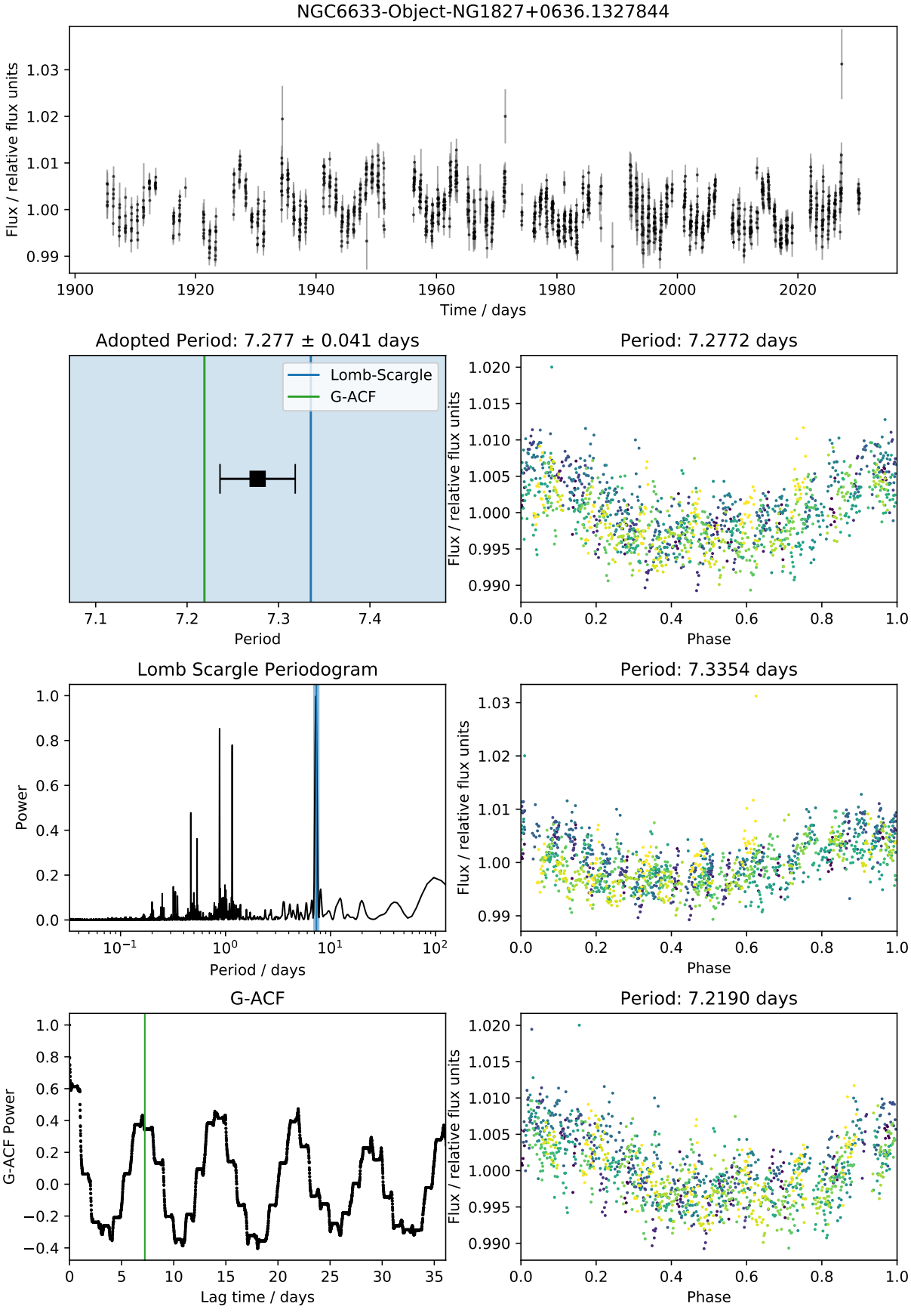

NGC6633-Object-NG1827+0636.1439611
Flux / relative flux units
Time / days
Adopted Period: 5.716 ± 0.02 days
Lomb-Scargle
G-ACF
Period
Period: 5.7156 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 5.7440 days
G-ACF
G-ACF Power
Lag time / days
Period: 5.6871 days

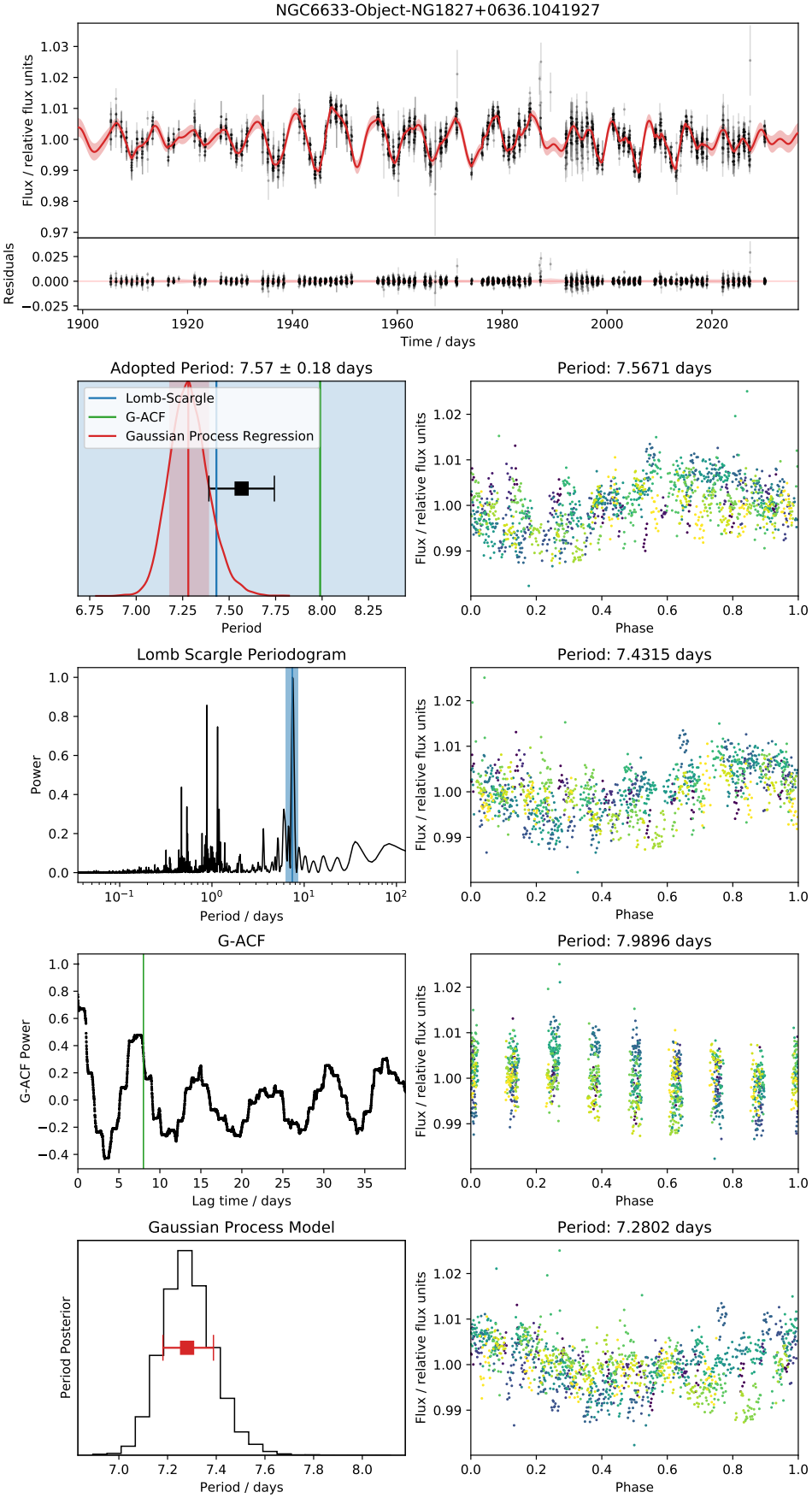

NGC6633-Object-NG1827+0636.1452712
Flux / relative flux units
Residuals
Time / days
Adopted Period: 6.046 ± 0.058 days
Lomb-Scargle
G-ACF
Gaussian Process Regression
Period
Period: 6.0463 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 5.9071 days
G-ACF
G-ACF Power
Lag time / days
Period: 6.1423 days
Gaussian Process Model
Period Posterior
Period / days
Period: 6.0895 days

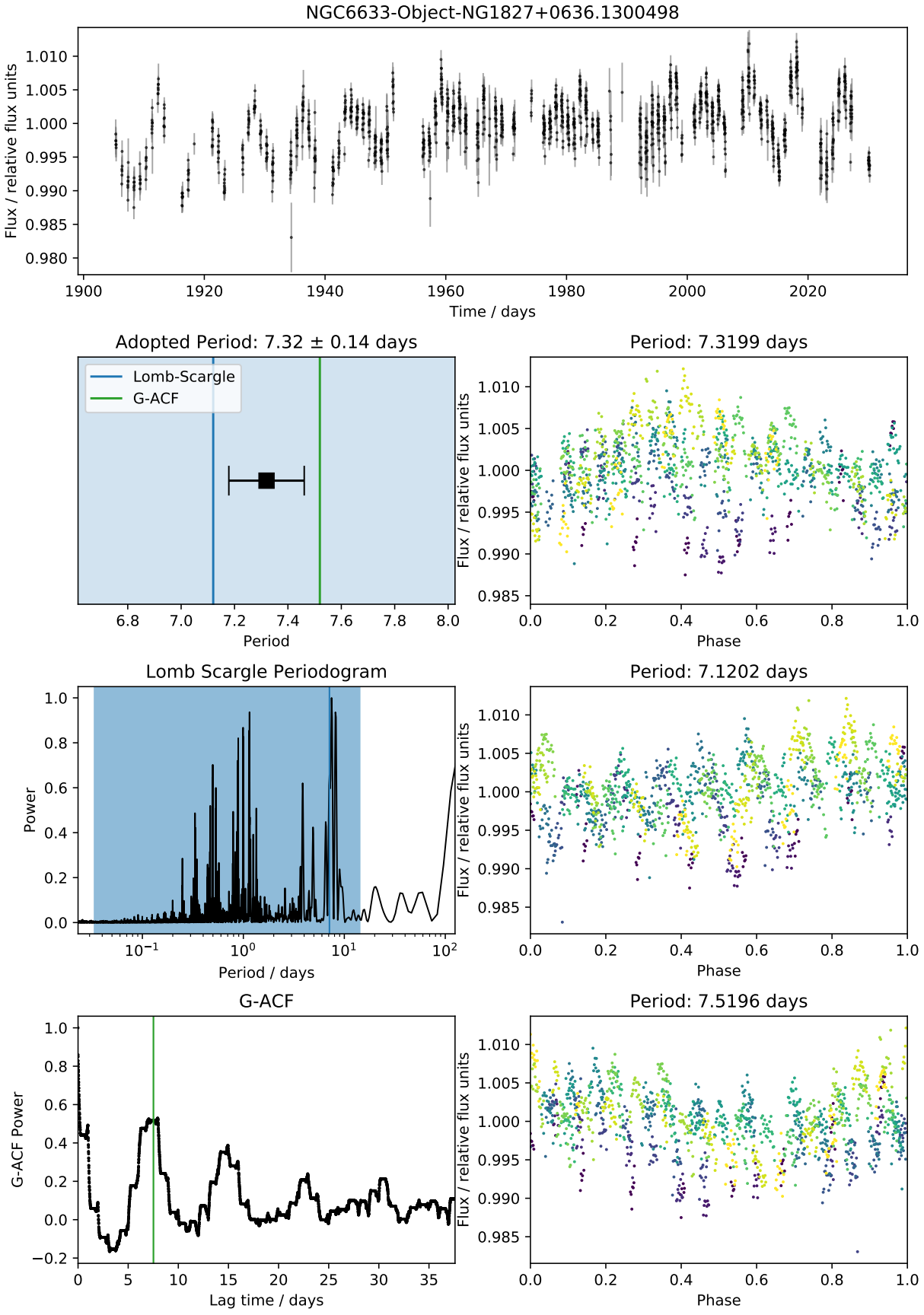

NGC6633-Object-NG1827+0636.1010770
Flux / relative flux units
Residuals
Time / days
Adopted Period: 86.0 ± 33.0 days
Lomb-Scargle
G-ACF
Gaussian Process Regression
Period
Period: 86.4510 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 6.1972 days
G-ACF
G-ACF Power
Lag time / days
Period: 123.7647 days
Gaussian Process Model
Period Posterior
Period / days
Period: 129.3912 days

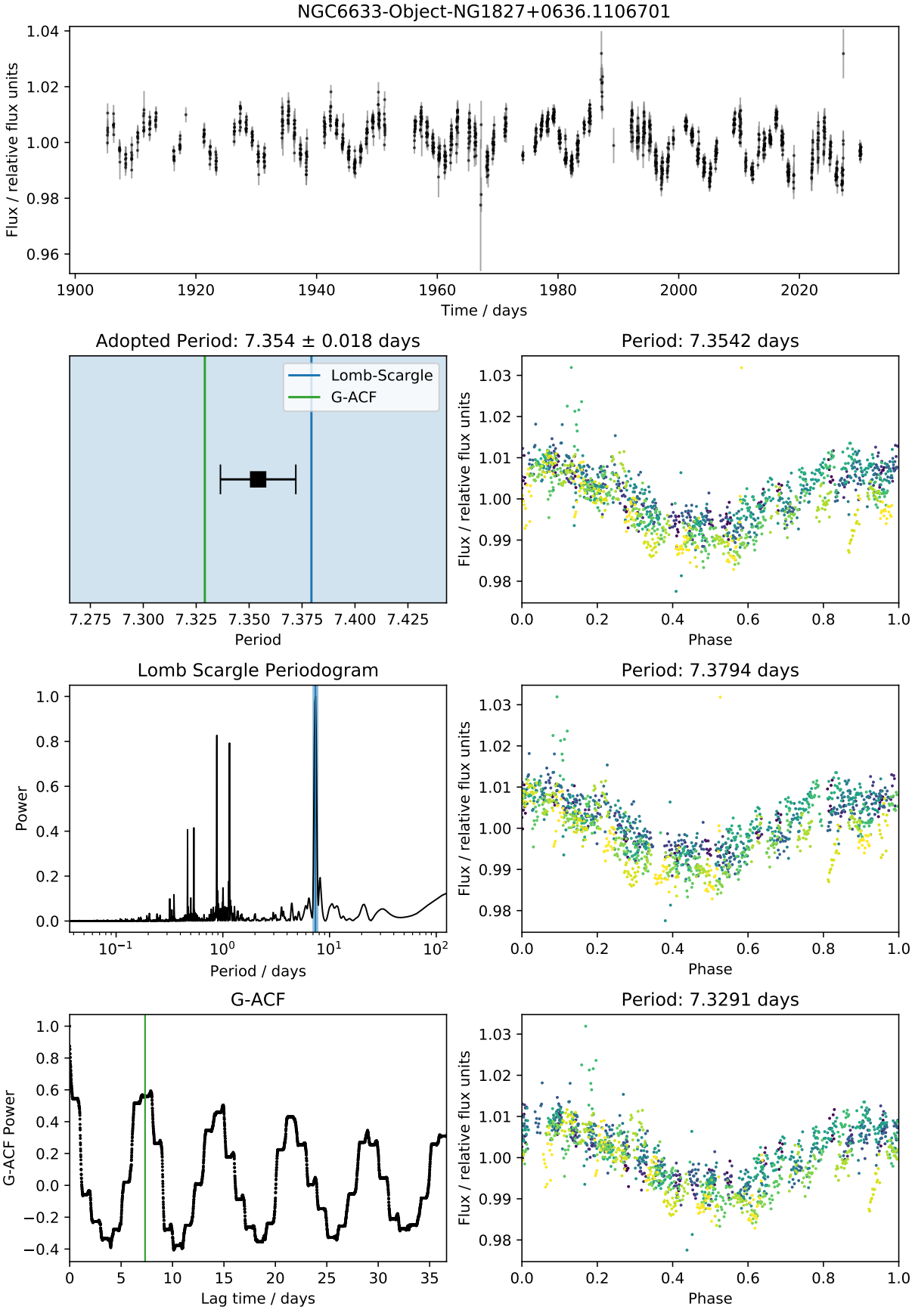

NGC6633-Object-NG1827+0636.1329847
Flux / relative flux units
Time / days
Adopted Period: 6.753 ± 0.015 days
Lomb-Scargle
G-ACF
Period
Period: 6.7532 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 6.7737 days
Phase
G-ACF
G-ACF Power
Lag time / days
Period: 6.7326 days
Phase

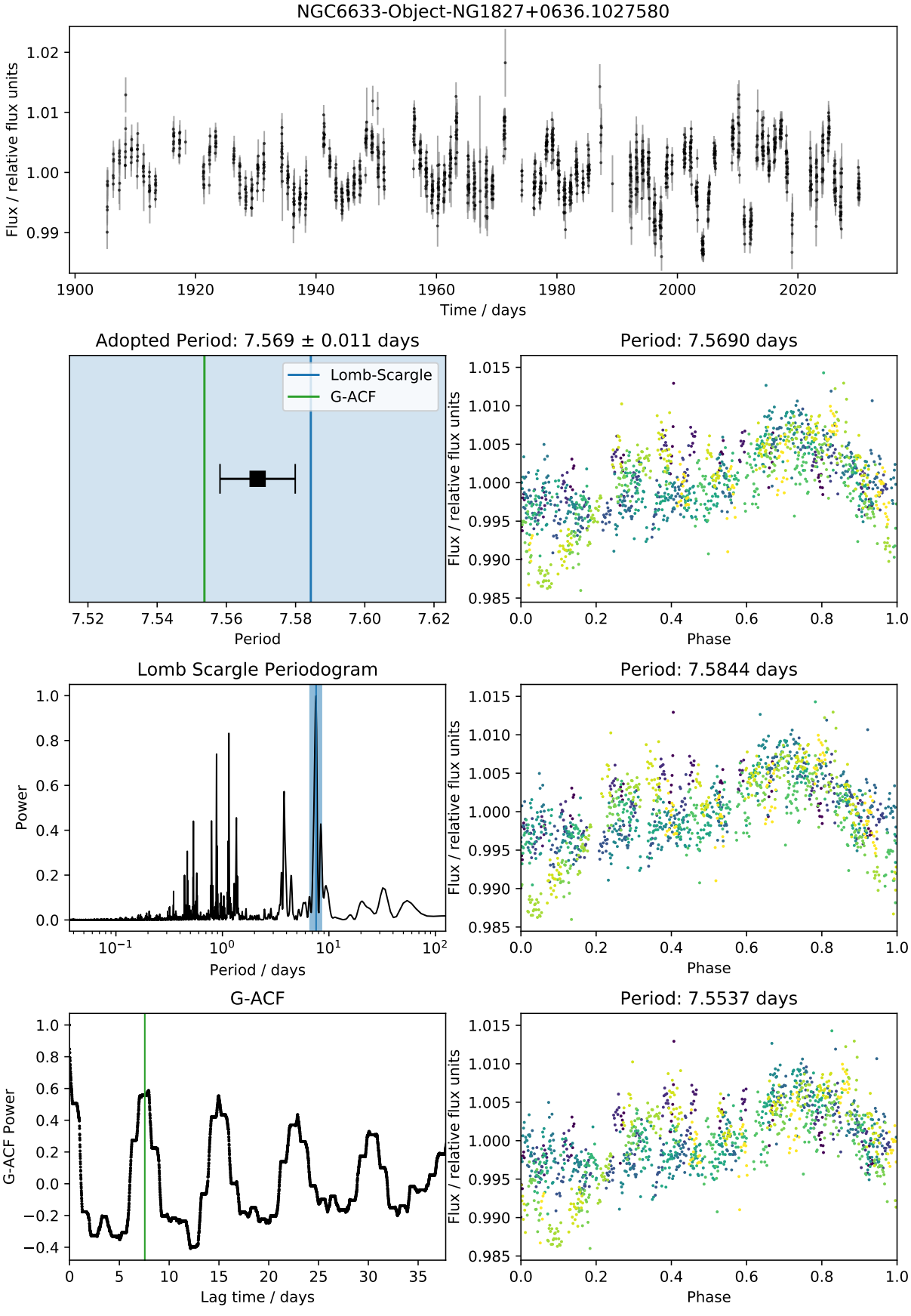

NGC6633-Object-NG1827+0636.1328554
Flux / relative flux units
Time / days
Adopted Period: 7.247 ± 0.013 days
Lomb-Scargle
G-ACF
Period
Period: 7.2470 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 7.2281 days
G-ACF
G-ACF Power
Lag time / days
Period: 7.2659 days

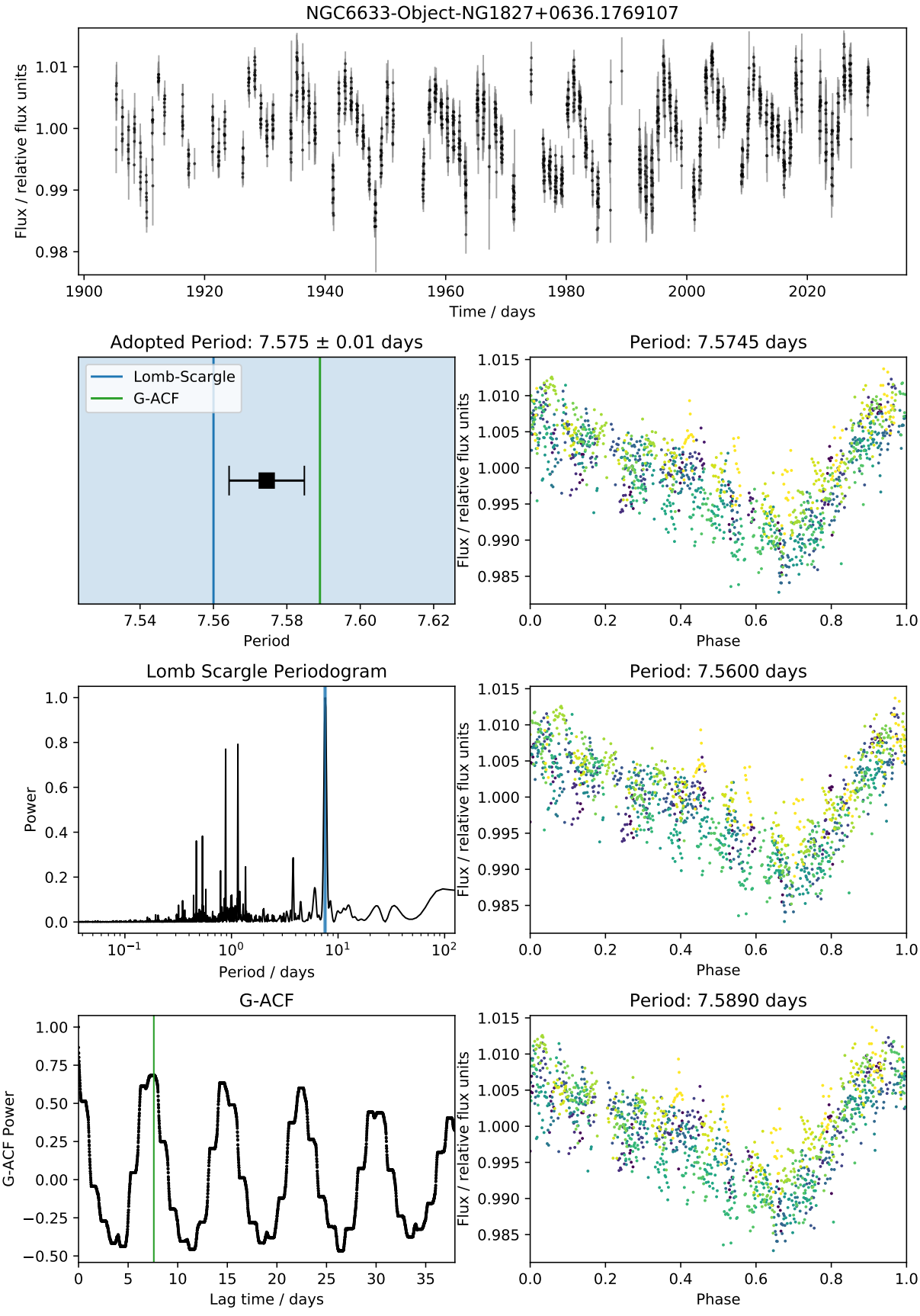

NGC6633-Object-NG1827+0636.1327844
Flux / relative flux units
Time / days
Adopted Period: 7.277 ± 0.041 days
Lomb-Scargle
G-ACF
Period
Period: 7.2772 days
Flux / relative flux units
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 7.3354 days
Flux / relative flux units
Phase
G-ACF
G-ACF Power
Lag time / days
Period: 7.2190 days
Flux / relative flux units
Phase

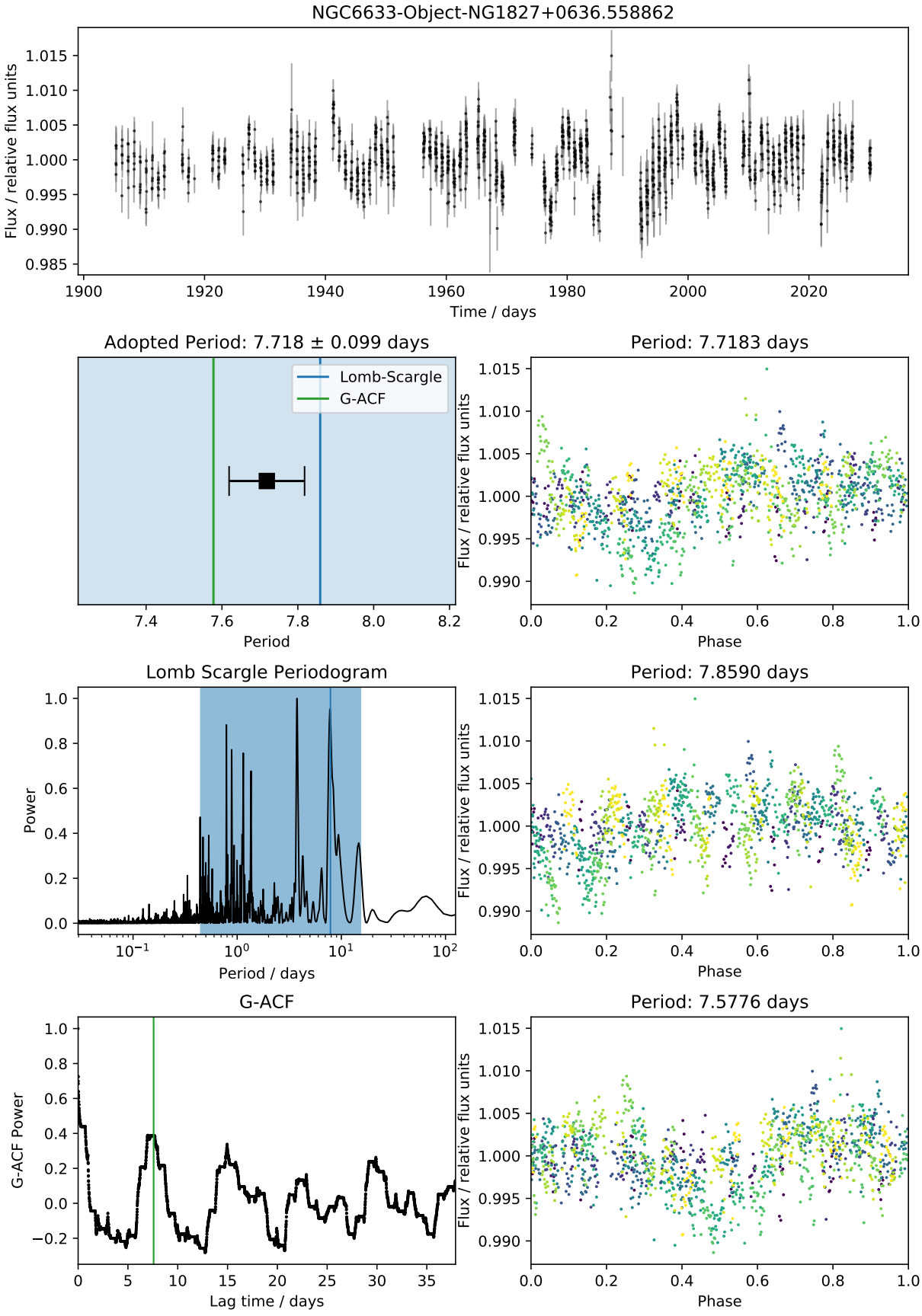

NGC6633-Object-NG1827+0636.1041927
Flux / relative flux units
Residuals
Time / days
Adopted Period: 7.57 ± 0.18 days
Lomb-Scargle
G-ACF
Gaussian Process Regression
Period
Period: 7.5671 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 7.4315 days
G-ACF
G-ACF Power
Lag time / days
Period: 7.9896 days
Gaussian Process Model
Period Posterior
Period / days
Period: 7.2802 days

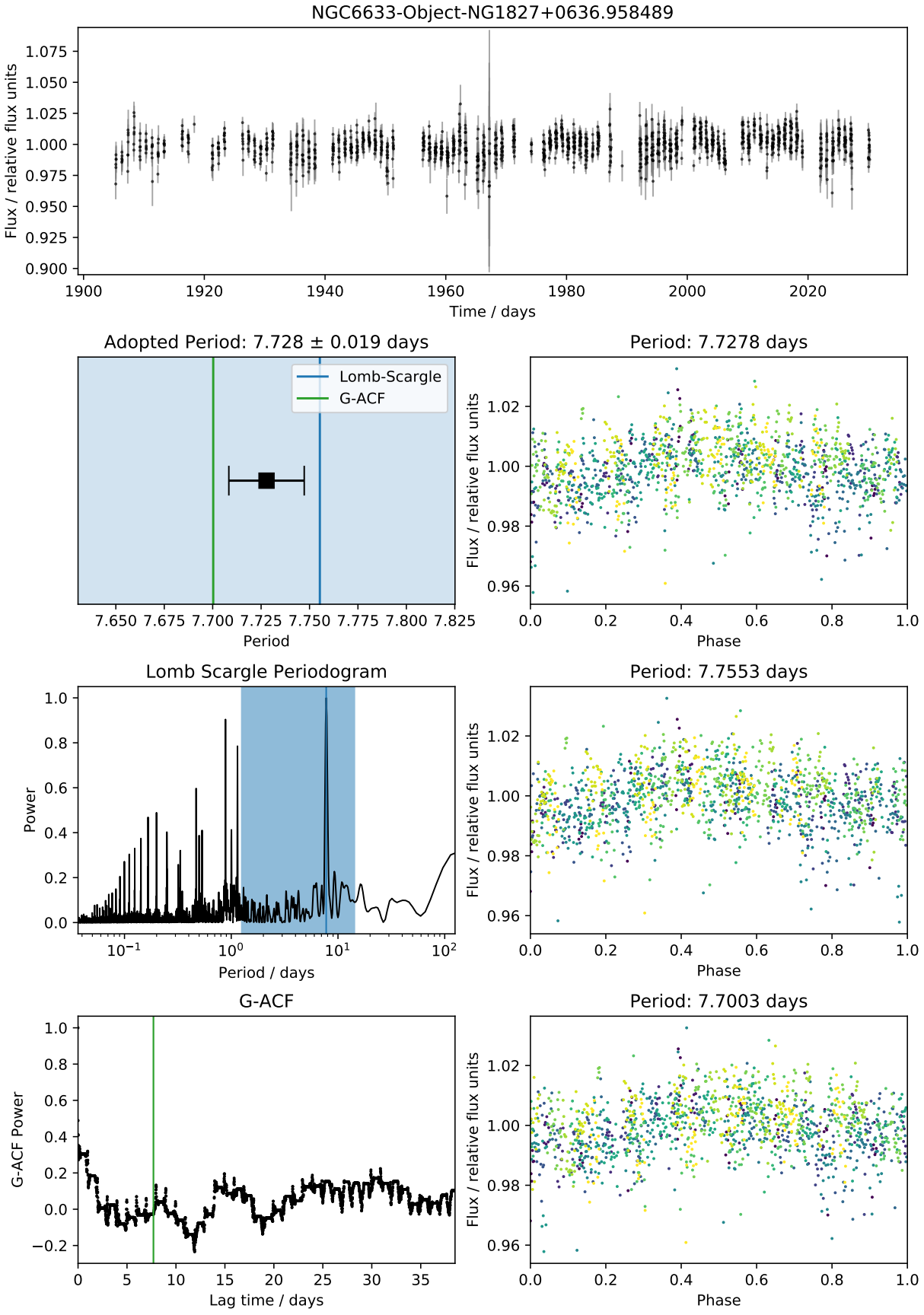

NGC6633-Object-NG1827+0636.1300498
Flux / relative flux units
Time / days
Adopted Period: 7.32 ± 0.14 days
Lomb-Scargle
G-ACF
Period
Period: 7.3199 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 7.1202 days
G-ACF
G-ACF Power
Lag time / days
Period: 7.5196 days

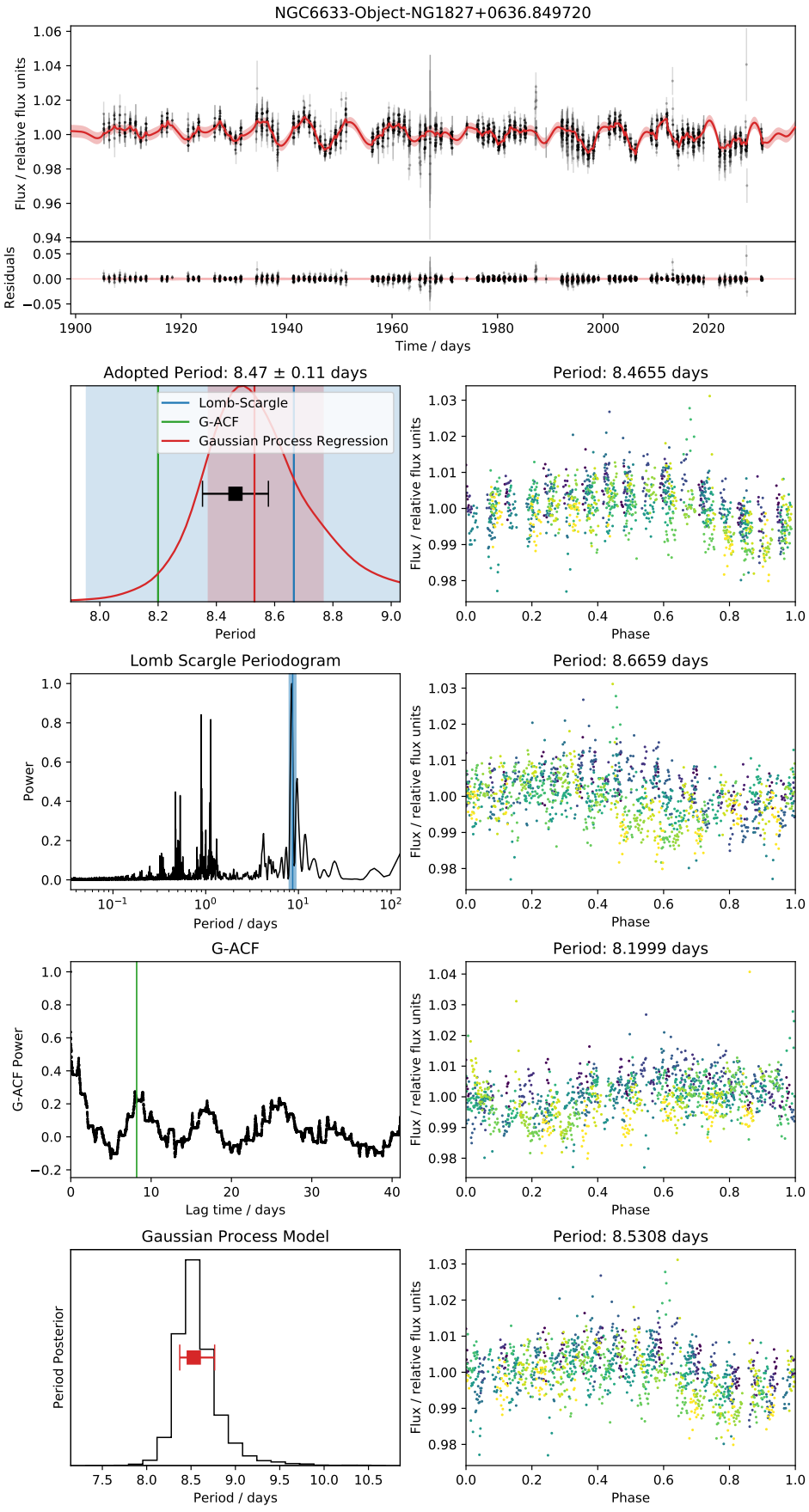

NGC6633-Object-NG1827+0636.1106701
Flux / relative flux units
Time / days
Adopted Period: 7.354 ± 0.018 days
Lomb-Scargle
G-ACF
Period
Period: 7.3542 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 7.3794 days
Phase
G-ACF
G-ACF Power
Lag time / days
Period: 7.3291 days
Phase

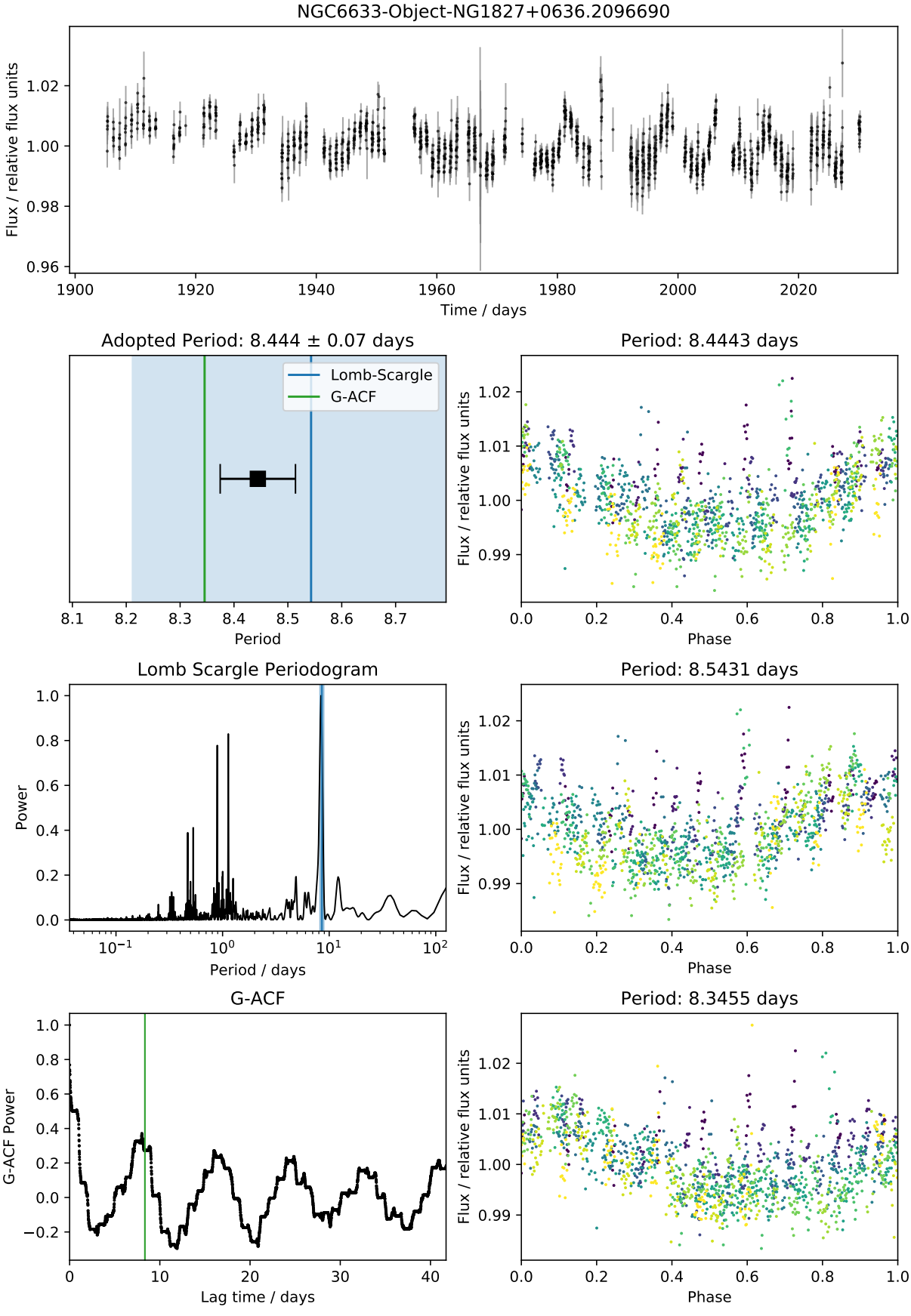

NGC6633-Object-NG1827+0636.1027580
Flux / relative flux units
Time / days
Adopted Period: 7.569 ± 0.011 days
Lomb-Scargle
G-ACF
Period
Period: 7.5690 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 7.5844 days
Phase
G-ACF
G-ACF Power
Lag time / days
Period: 7.5537 days
Phase

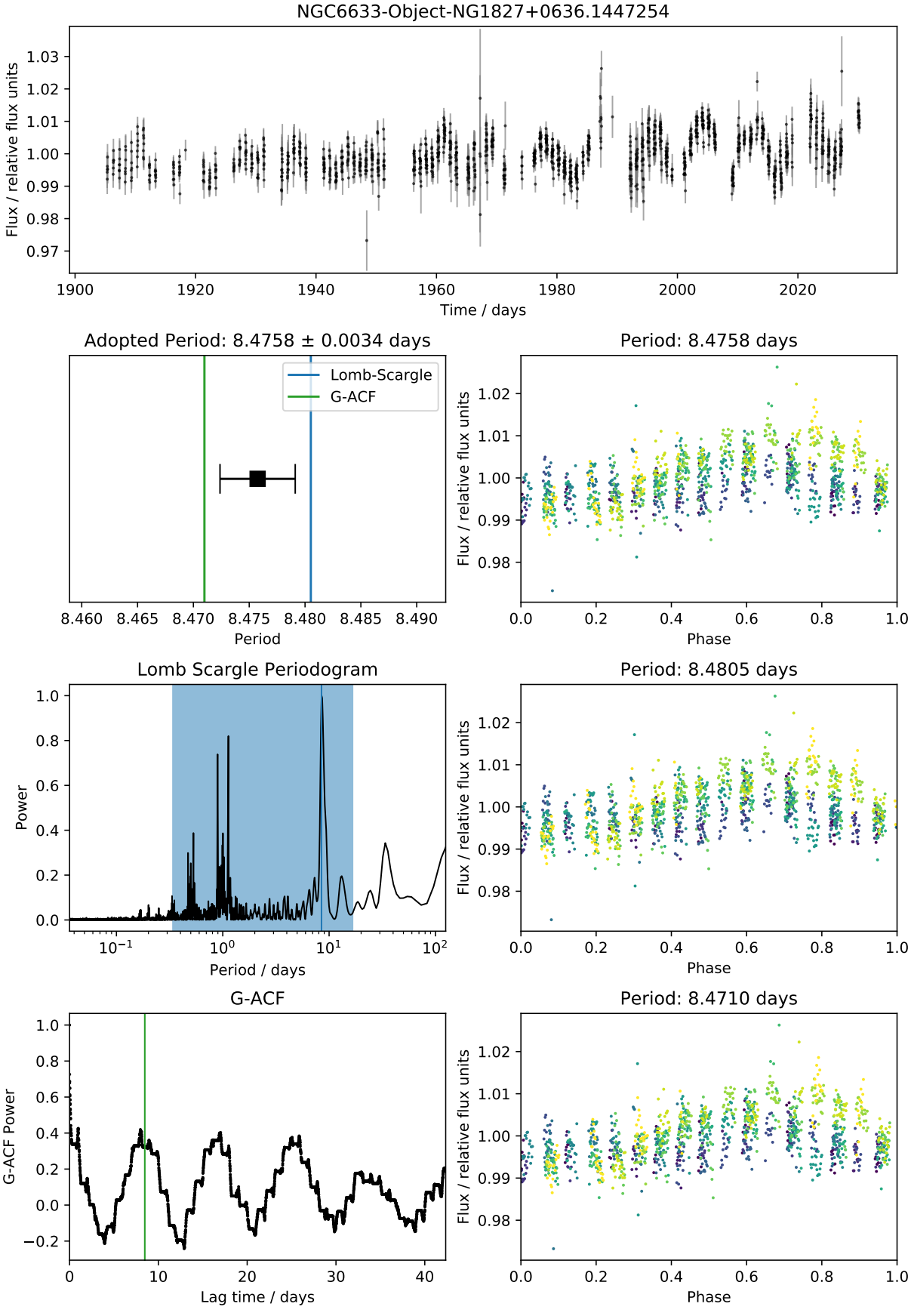

NGC6633-Object-NG1827+0636.1769107
Flux / relative flux units
Time / days
Adopted Period: 7.575 ± 0.01 days
Lomb-Scargle
G-ACF
Period
Period: 7.5745 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 7.5600 days
G-ACF
G-ACF Power
Lag time / days
Period: 7.5890 days

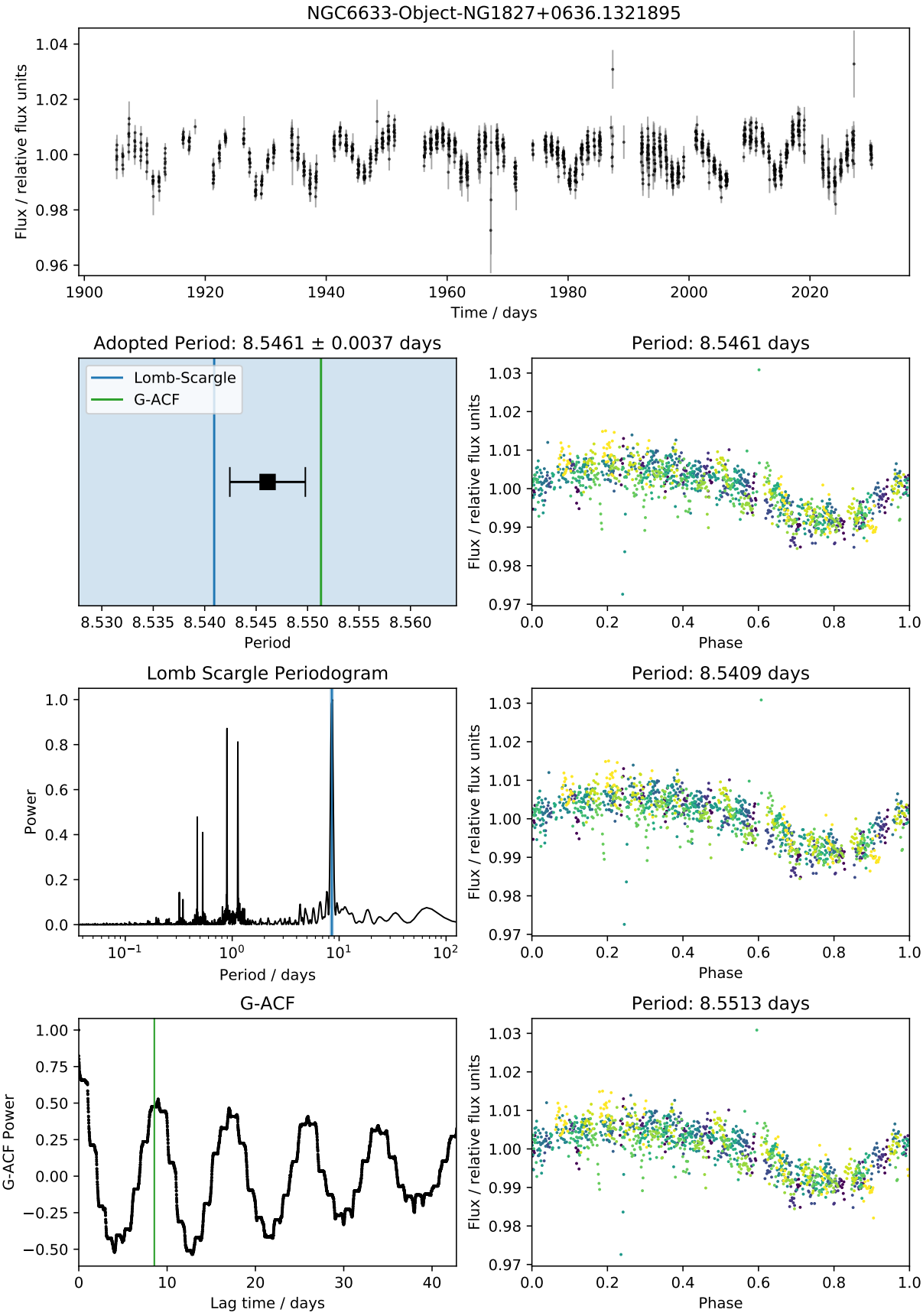

NGC6633-Object-NG1827+0636.558862
Flux / relative flux units
Time / days
Adopted Period: 7.718 ± 0.099 days
Lomb-Scargle
G-ACF
Period
Period: 7.7183 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 7.8590 days
G-ACF
G-ACF Power
Lag time / days
Period: 7.5776 days

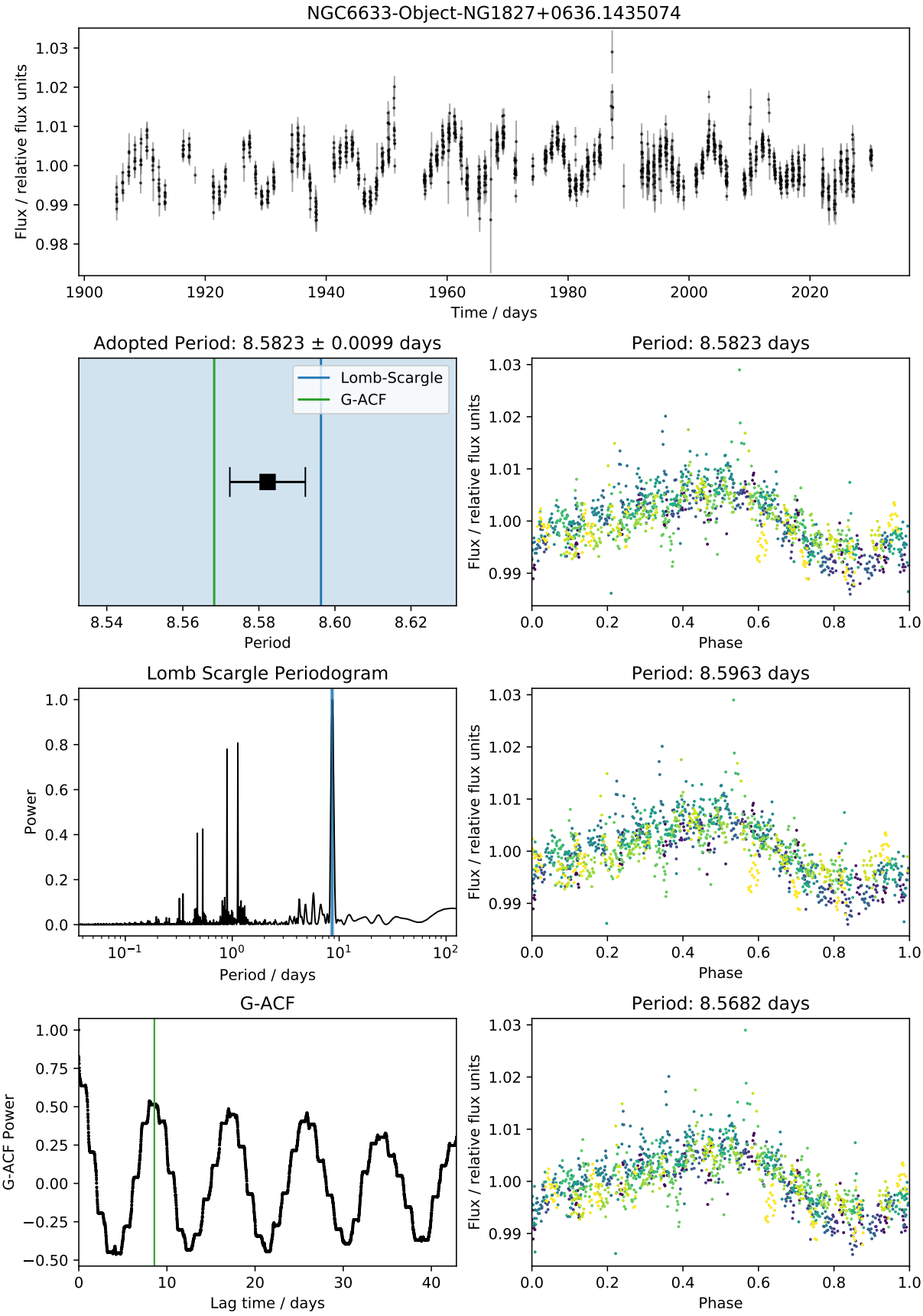

NGC6633-Object-NG1827+0636.958489
Flux / relative flux units
Time / days
Adopted Period: 7.728 ± 0.019 days
Lomb-Scargle
G-ACF
Period
Period: 7.7278 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 7.7553 days
G-ACF
G-ACF Power
Lag time / days
Period: 7.7003 days

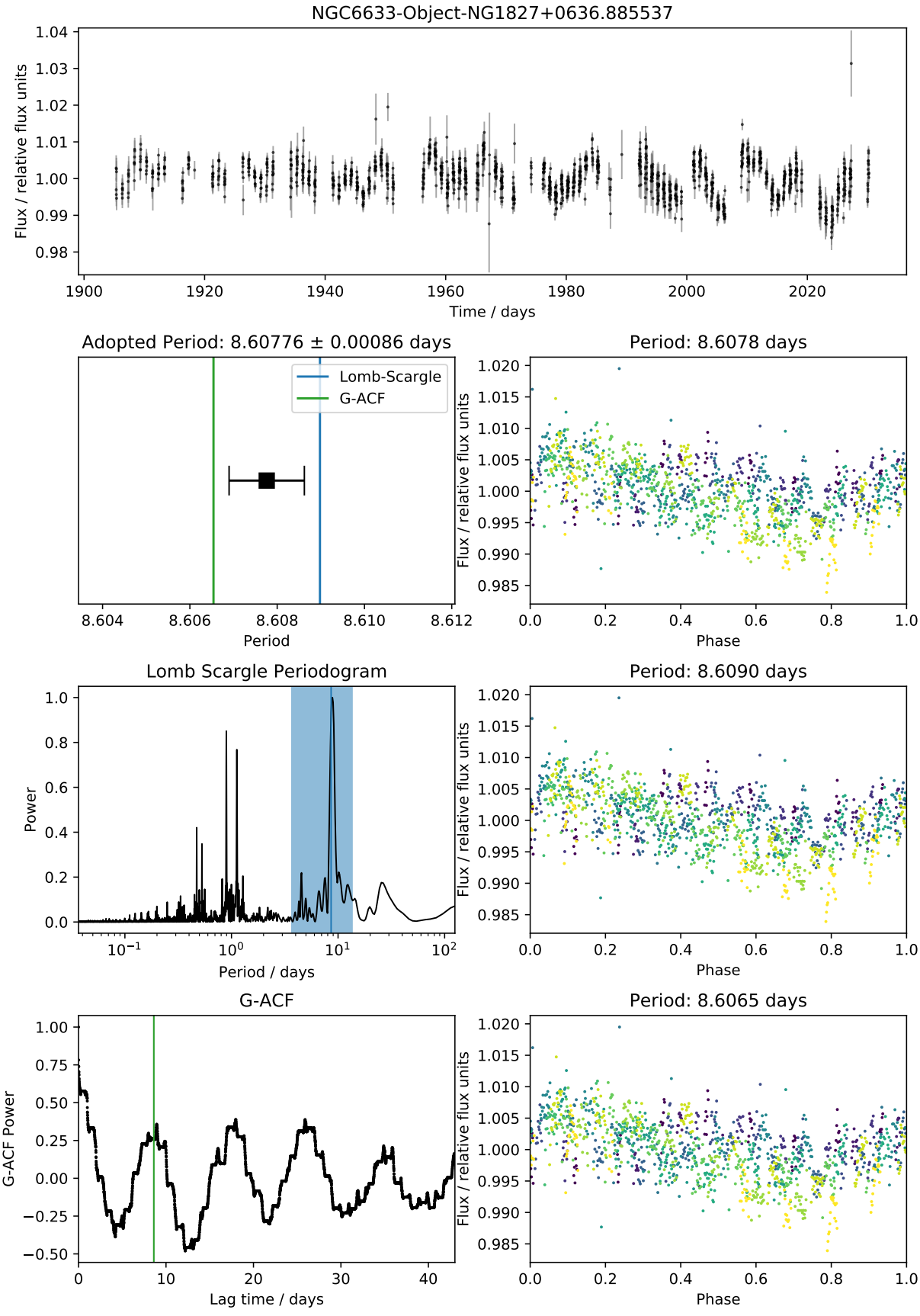

NGC6633-Object-NG1827+0636.849720
Flux / relative flux units
Residuals
Time / days
Adopted Period: 8.47 ± 0.11 days
Lomb-Scargle
G-ACF
Gaussian Process Regression
Period
Period: 8.4655 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 8.6659 days
G-ACF
G-ACF Power
Lag time / days
Period: 8.1999 days
Gaussian Process Model
Period Posterior
Period / days
Period: 8.5308 days

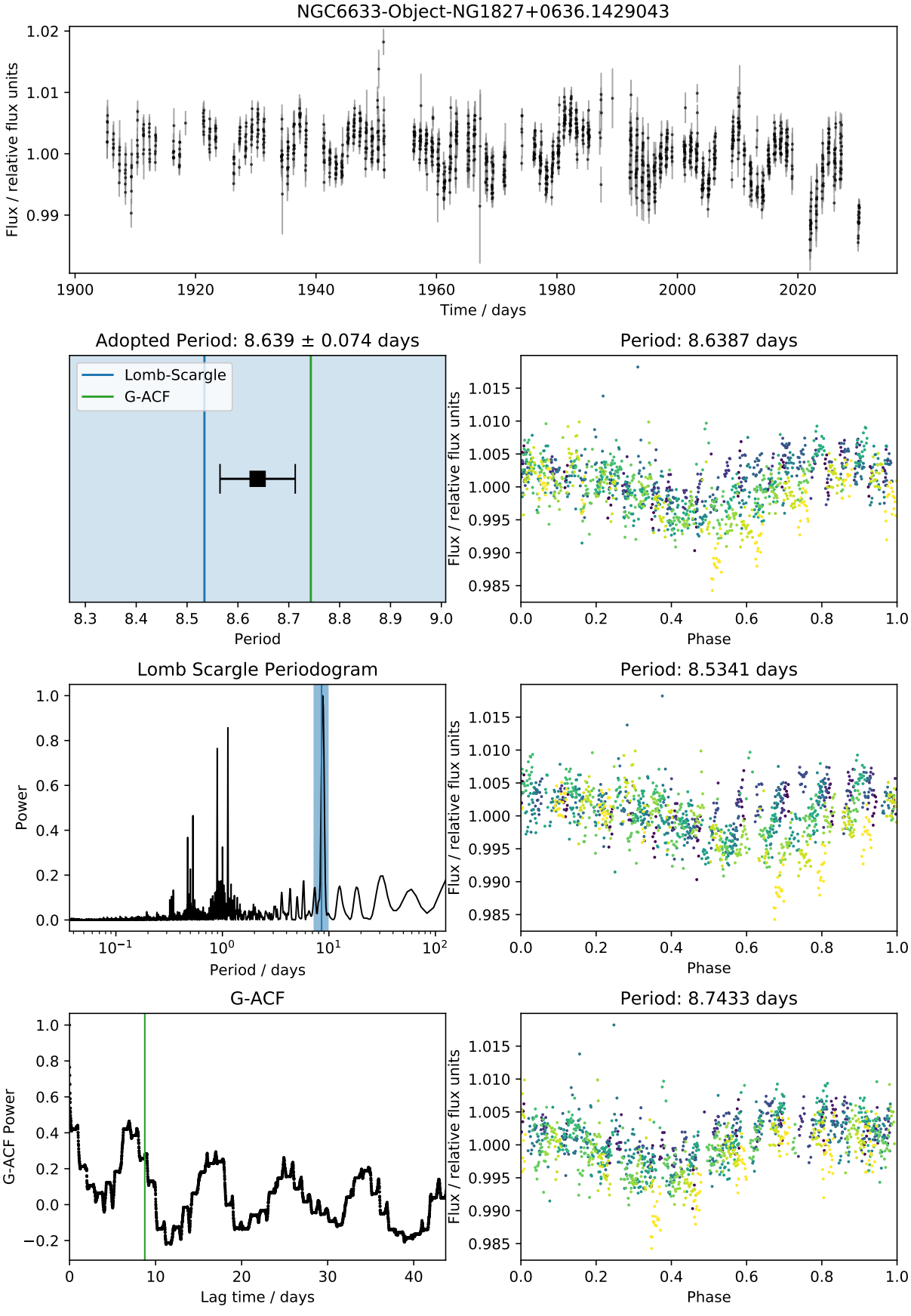

NGC6633-Object-NG1827+0636.2096690
Flux / relative flux units
Time / days
Adopted Period: 8.444 ± 0.07 days
Lomb-Scargle
G-ACF
Period
Period: 8.4443 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 8.5431 days
G-ACF
G-ACF Power
Lag time / days
Period: 8.3455 days

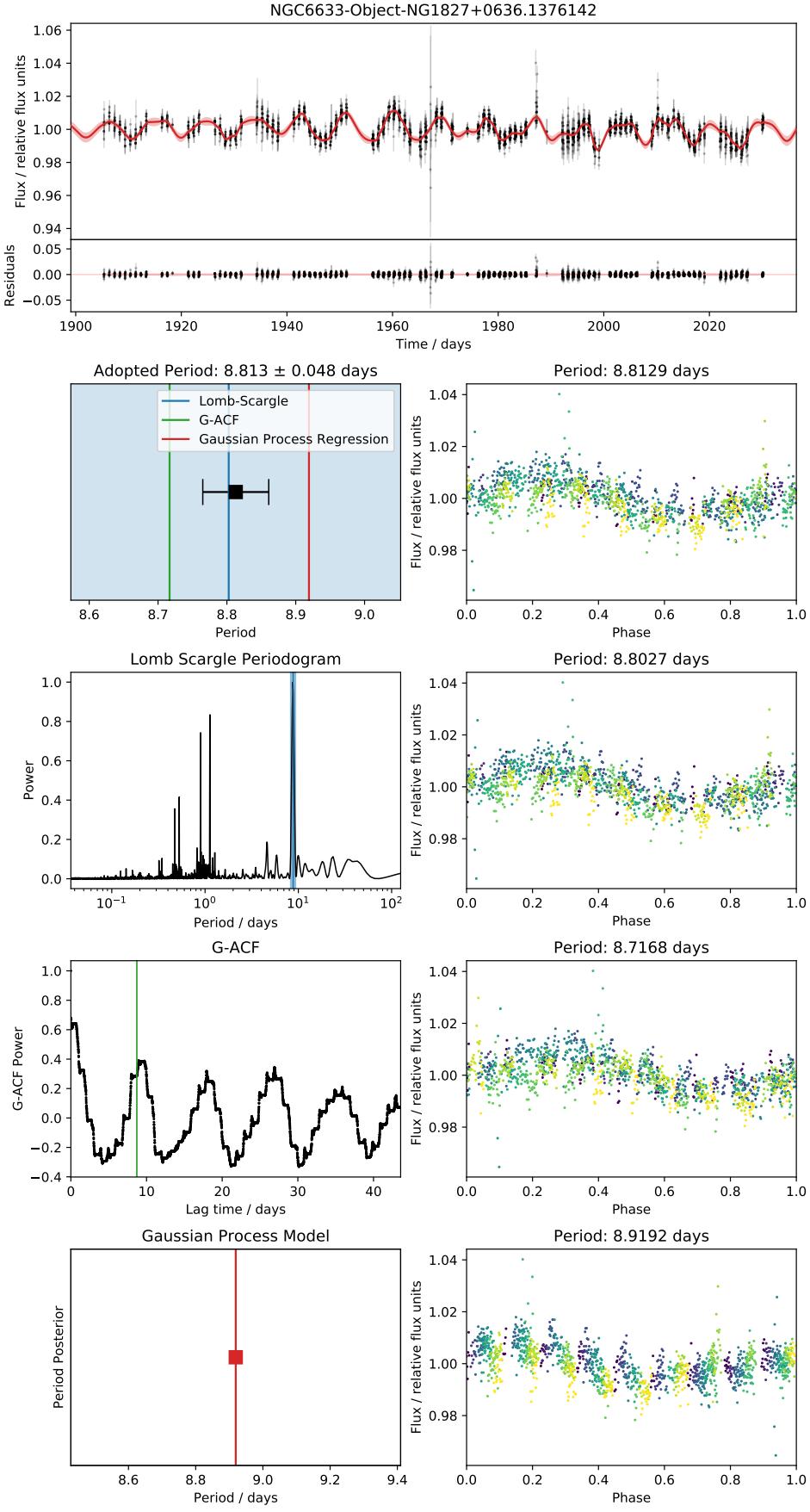

NGC6633-Object-NG1827+0636.1447254
Flux / relative flux units
Time / days
Adopted Period: 8.4758 ± 0.0034 days
Lomb-Scargle
G-ACF
Period
Period: 8.4758 days
Flux / relative flux units
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 8.4805 days
Flux / relative flux units
Phase
G-ACF
G-ACF Power
Lag time / days
Period: 8.4710 days
Flux / relative flux units
Phase

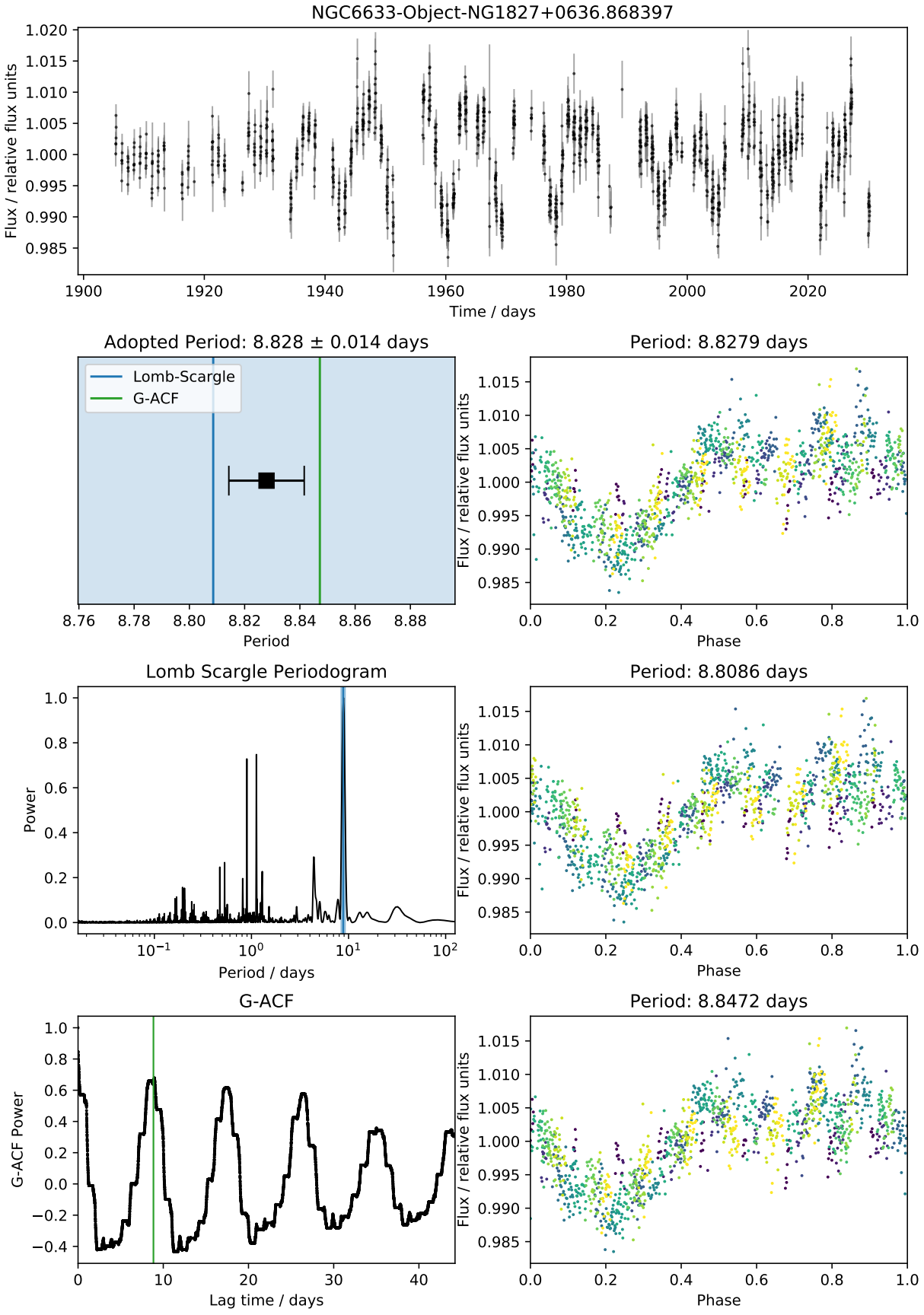

NGC6633-Object-NG1827+0636.1321895
Flux / relative flux units
Time / days
Adopted Period: 8.5461 ± 0.0037 days
Lomb-Scargle
G-ACF
Period
Period: 8.5461 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 8.5409 days
Phase
G-ACF
G-ACF Power
Lag time / days
Period: 8.5513 days
Phase

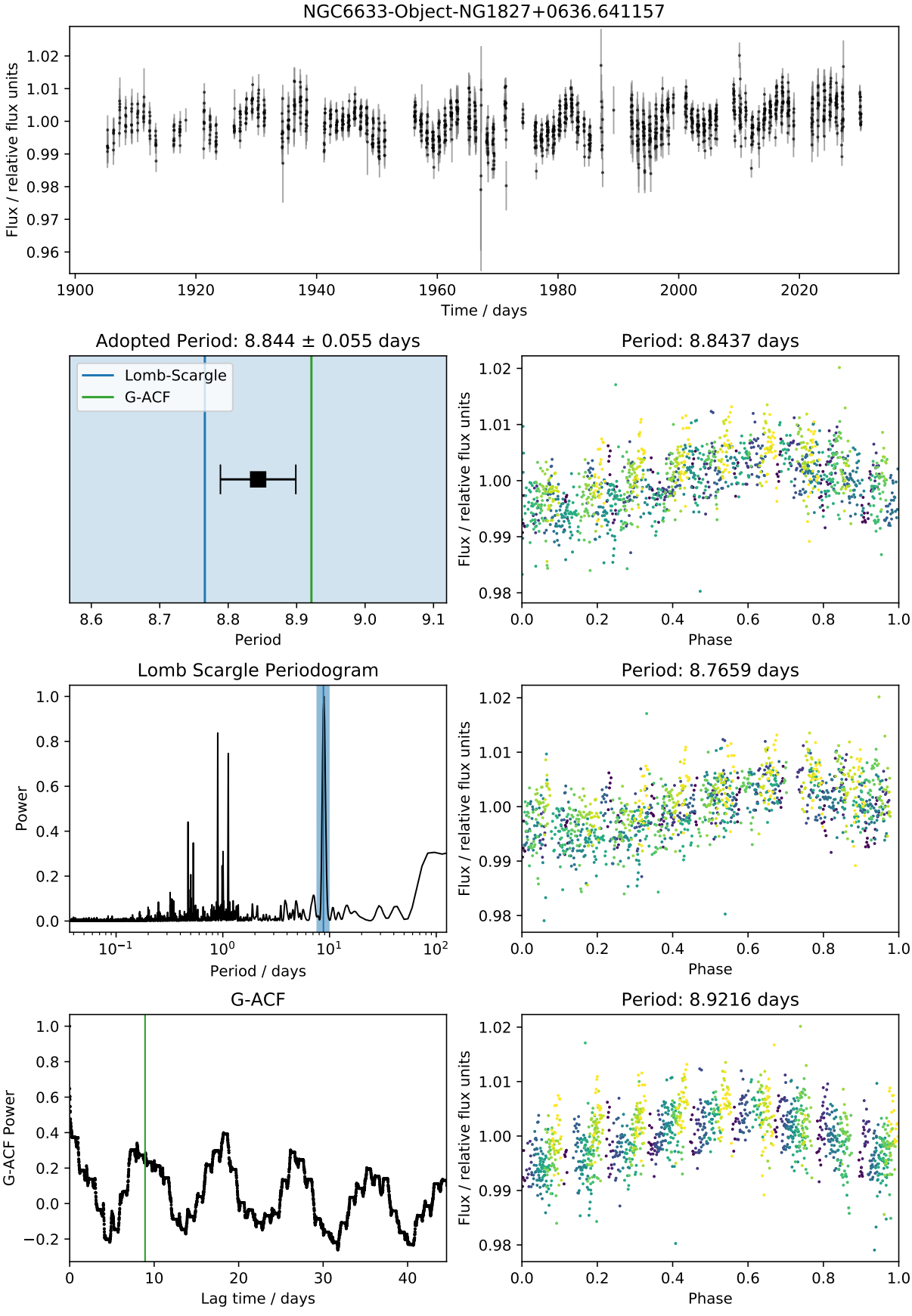

NGC6633-Object-NG1827+0636.1435074
Flux / relative flux units
Time / days
Adopted Period: 8.5823 ± 0.0099 days
Lomb-Scargle
G-ACF
Period
Period: 8.5823 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 8.5963 days
Phase
G-ACF
G-ACF Power
Lag time / days
Period: 8.5682 days
Phase

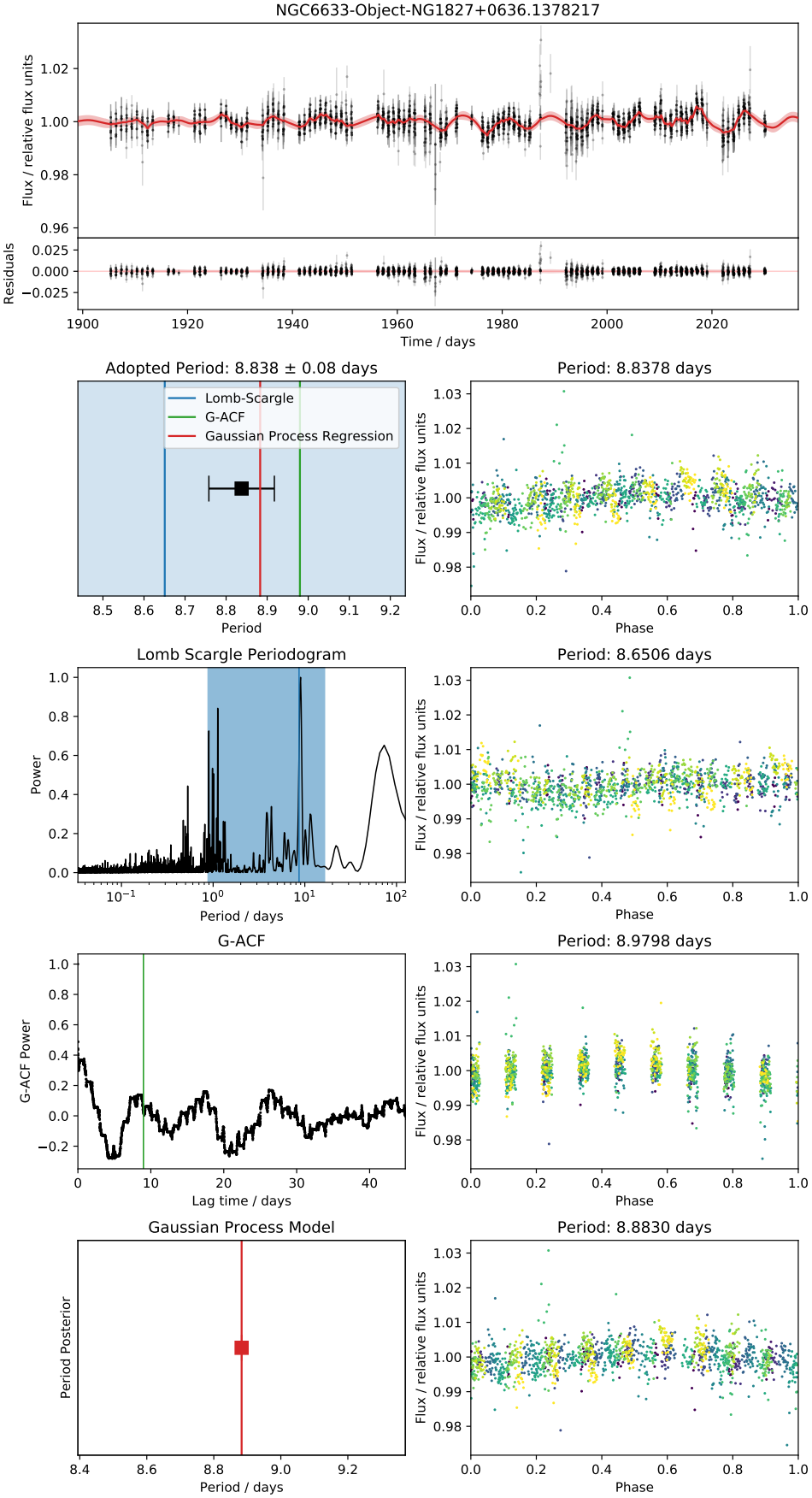

NGC6633-Object-NG1827+0636.885537
Flux / relative flux units
Time / days
Adopted Period: 8.60776 ± 0.00086 days
Lomb-Scargle
G-ACF
Period
Period: 8.6078 days
Flux / relative flux units
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 8.6090 days
Flux / relative flux units
Phase
G-ACF
G-ACF Power
Lag time / days
Period: 8.6065 days
Flux / relative flux units
Phase

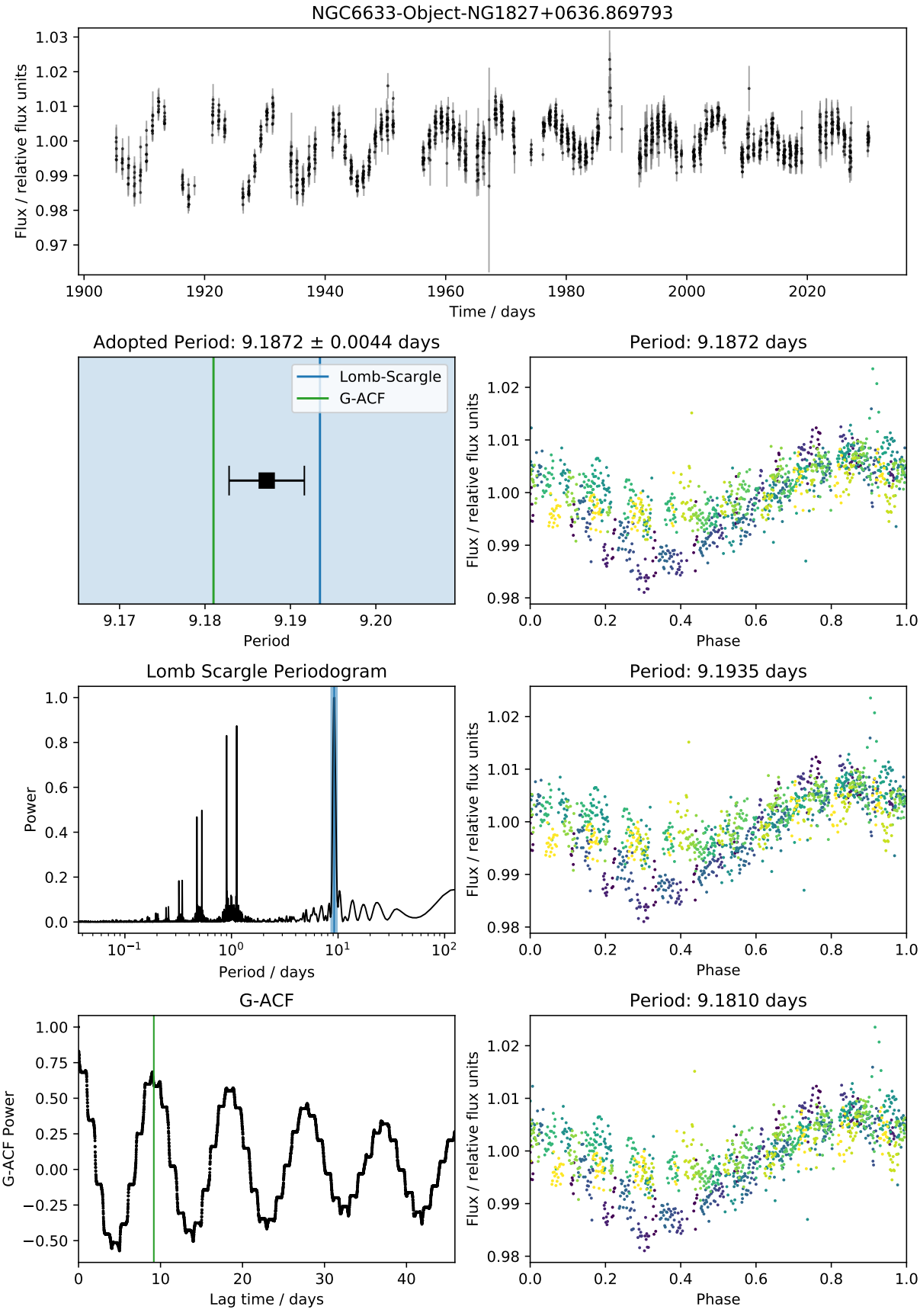

NGC6633-Object-NG1827+0636.1429043
Flux / relative flux units
Time / days
Adopted Period: 8.639 ± 0.074 days
Lomb-Scargle
G-ACF
Period
Period: 8.6387 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 8.5341 days
Phase
G-ACF
G-ACF Power
Lag time / days
Period: 8.7433 days
Phase

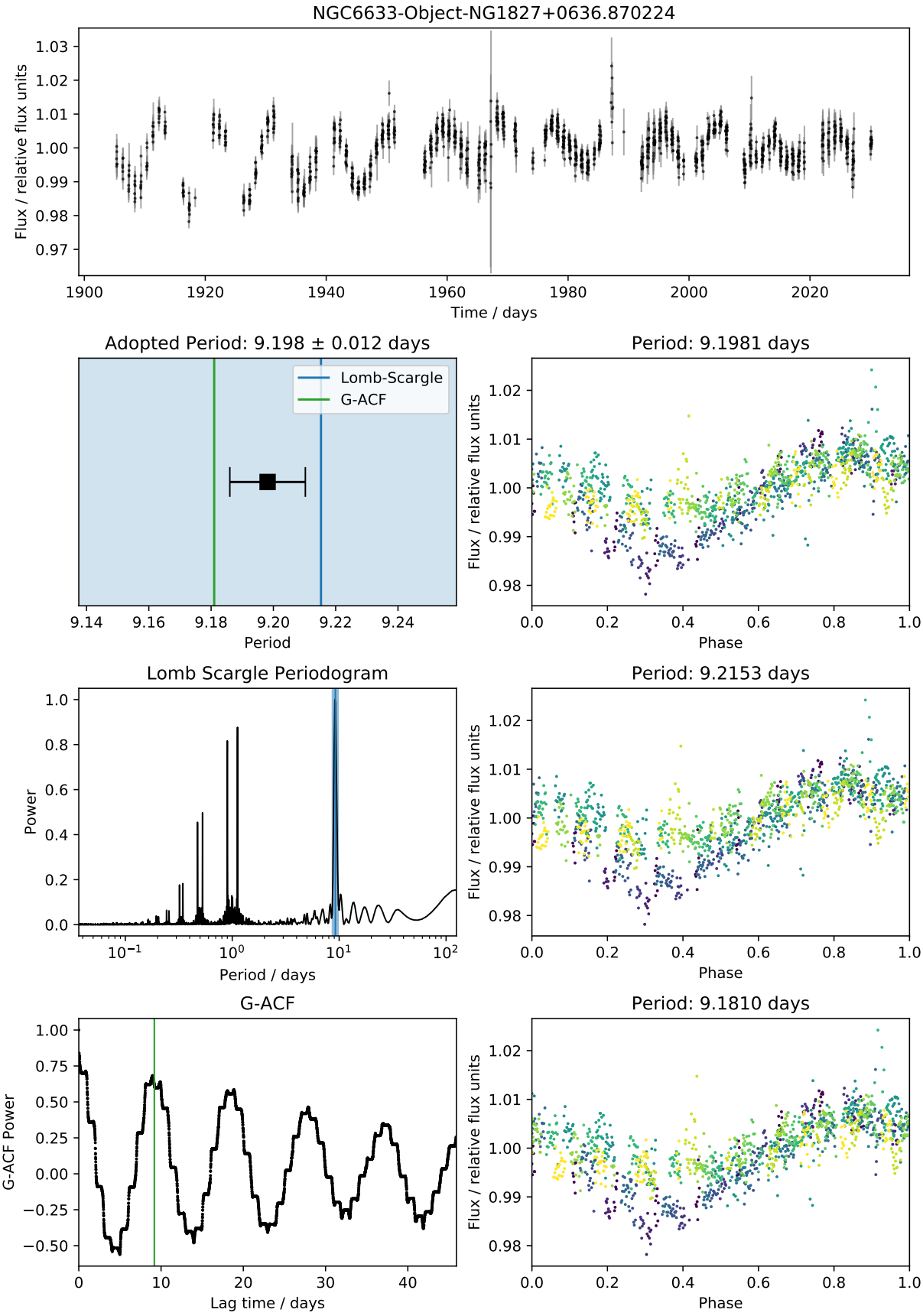

NGC6633-Object-NG1827+0636.1376142
Flux / relative flux units
Residuals
Time / days
Adopted Period: 8.813 ± 0.048 days
Lomb-Scargle
G-ACF
Gaussian Process Regression
Period
Period: 8.8129 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 8.8027 days
G-ACF
G-ACF Power
Lag time / days
Period: 8.7168 days
Gaussian Process Model
Period Posterior
Period / days
Period: 8.9192 days

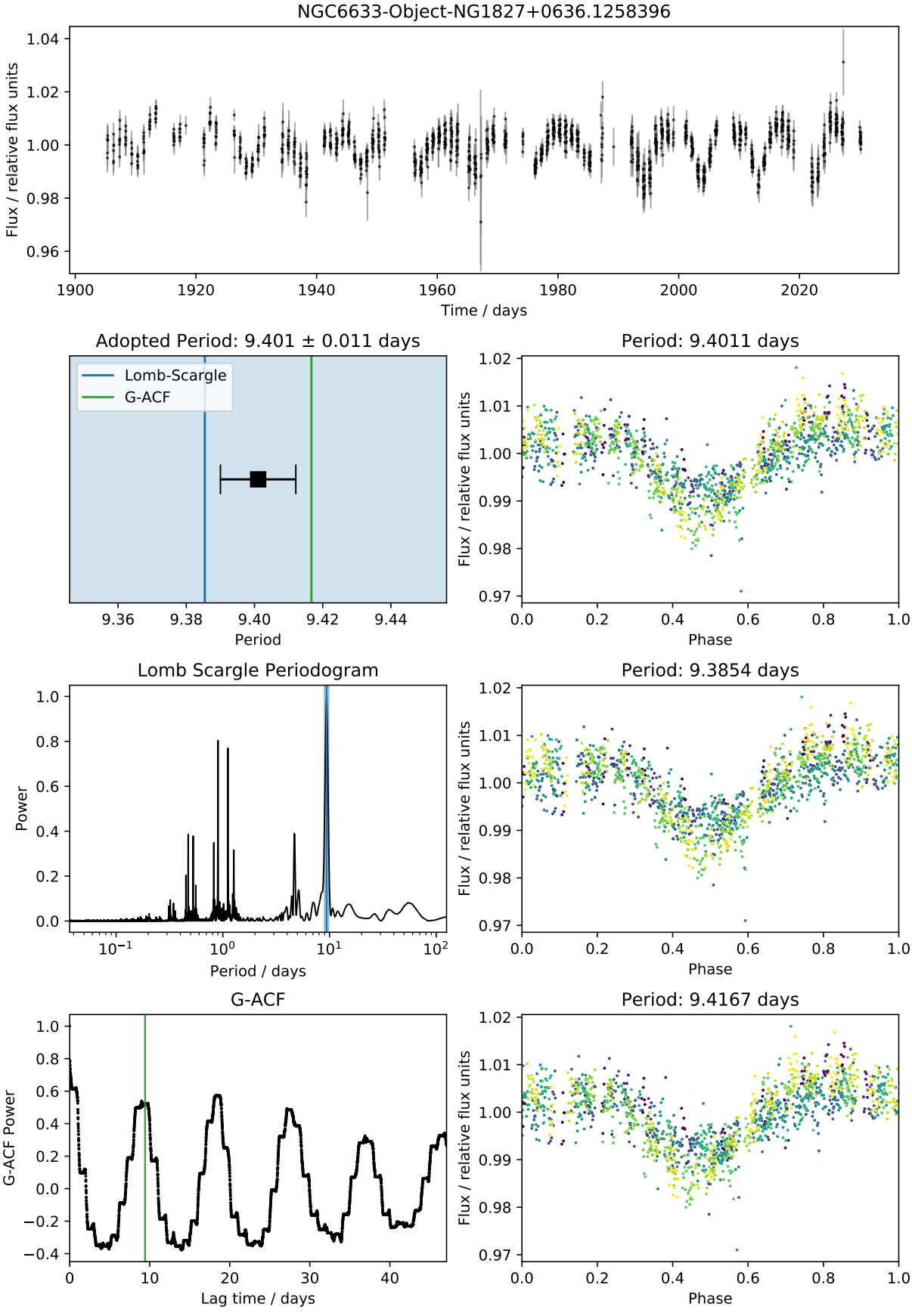

NGC6633-Object-NG1827+0636.868397
Flux / relative flux units
Time / days
Adopted Period: 8.828 ± 0.014 days
Lomb-Scargle
G-ACF
Period
Period: 8.8279 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 8.8086 days
Phase
G-ACF
G-ACF Power
Lag time / days
Period: 8.8472 days
Phase

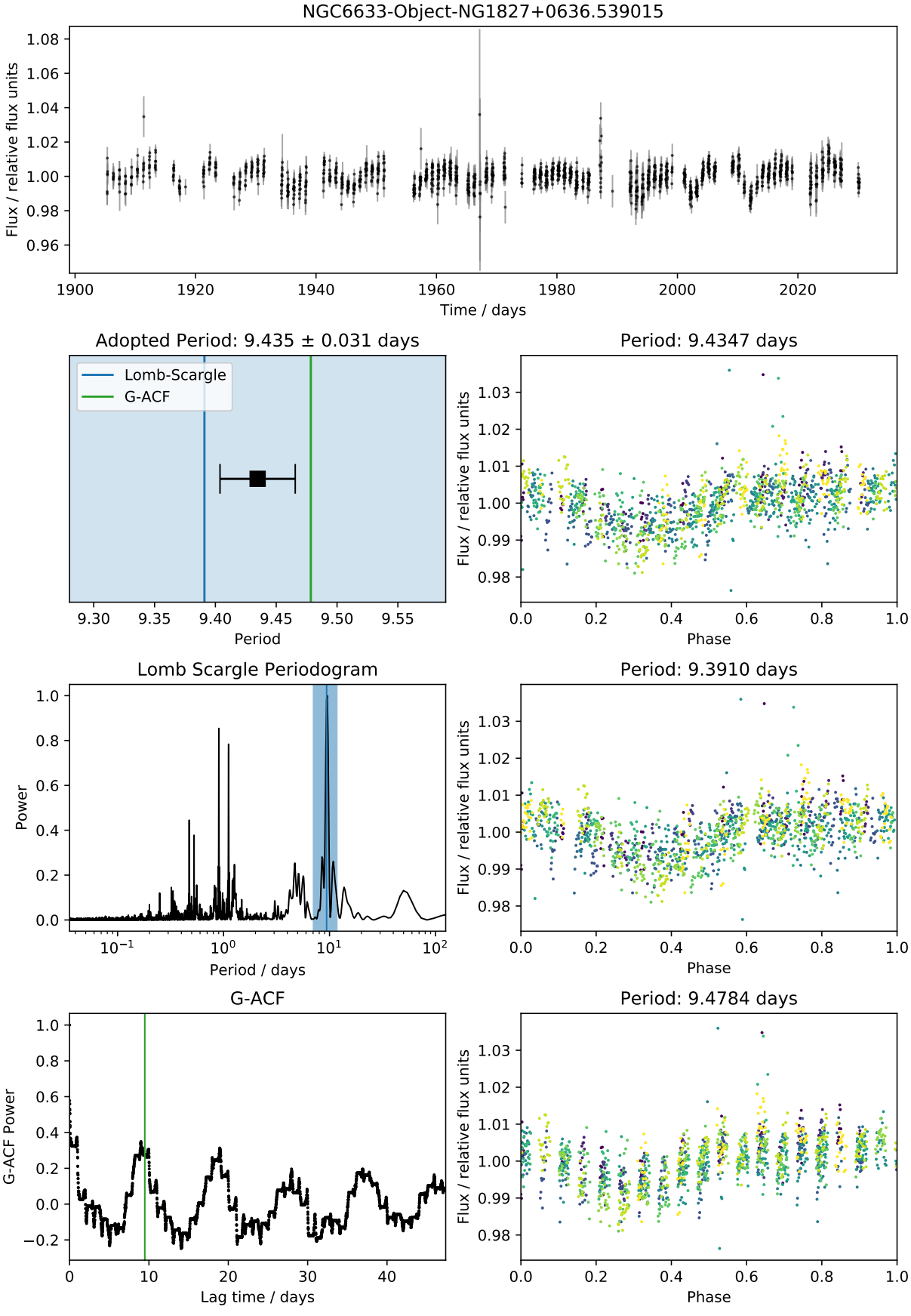

NGC6633-Object-NG1827+0636.641157
Flux / relative flux units
Time / days
Adopted Period: 8.844 ± 0.055 days
Lomb-Scargle
G-ACF
Period
Period: 8.8437 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 8.7659 days
Phase
G-ACF
G-ACF Power
Lag time / days
Period: 8.9216 days
Phase

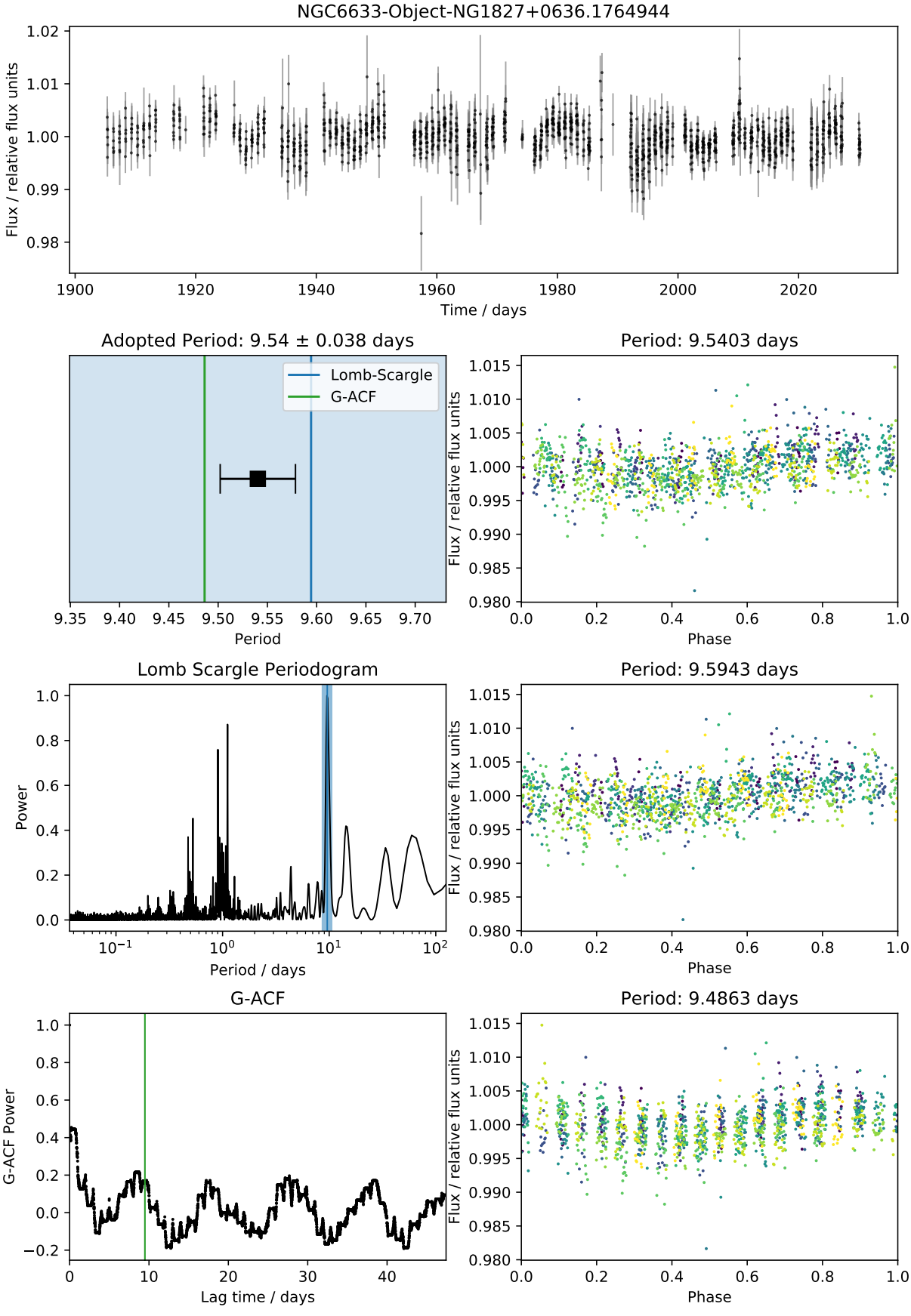

NGC6633-Object-NG1827+0636.1378217
Flux / relative flux units
Residuals
Time / days
Adopted Period: 8.838 ± 0.08 days
Lomb-Scargle
G-ACF
Gaussian Process Regression
Period
Period: 8.8378 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 8.6506 days
G-ACF
G-ACF Power
Lag time / days
Period: 8.9798 days
Gaussian Process Model
Period Posterior
Period / days
Period: 8.8830 days

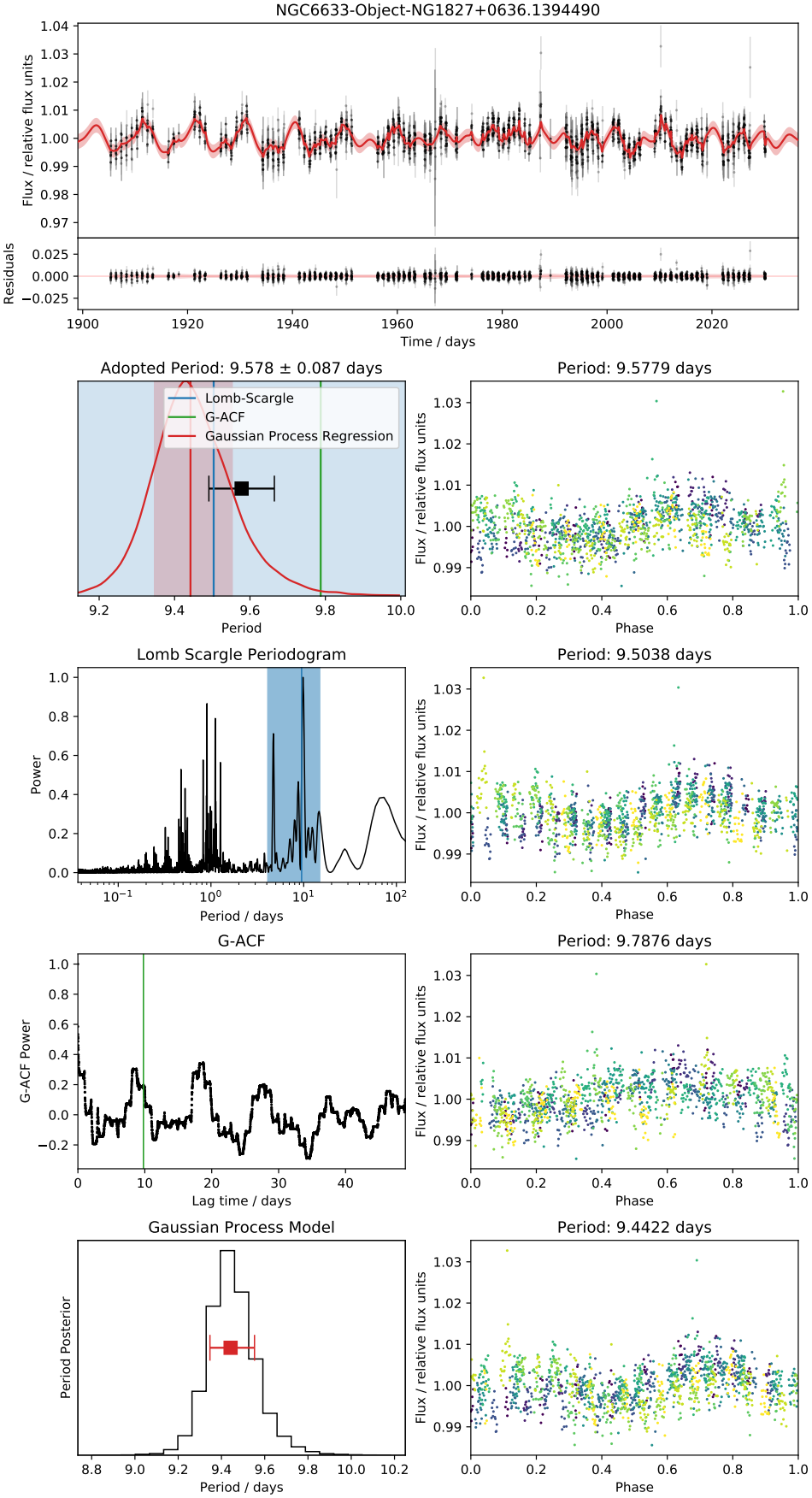

NGC6633-Object-NG1827+0636.869793
Flux / relative flux units
Time / days
Adopted Period: 9.1872 ± 0.0044 days
Lomb-Scargle
G-ACF
Period
Period: 9.1872 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 9.1935 days
Phase
G-ACF
G-ACF Power
Lag time / days
Period: 9.1810 days
Phase

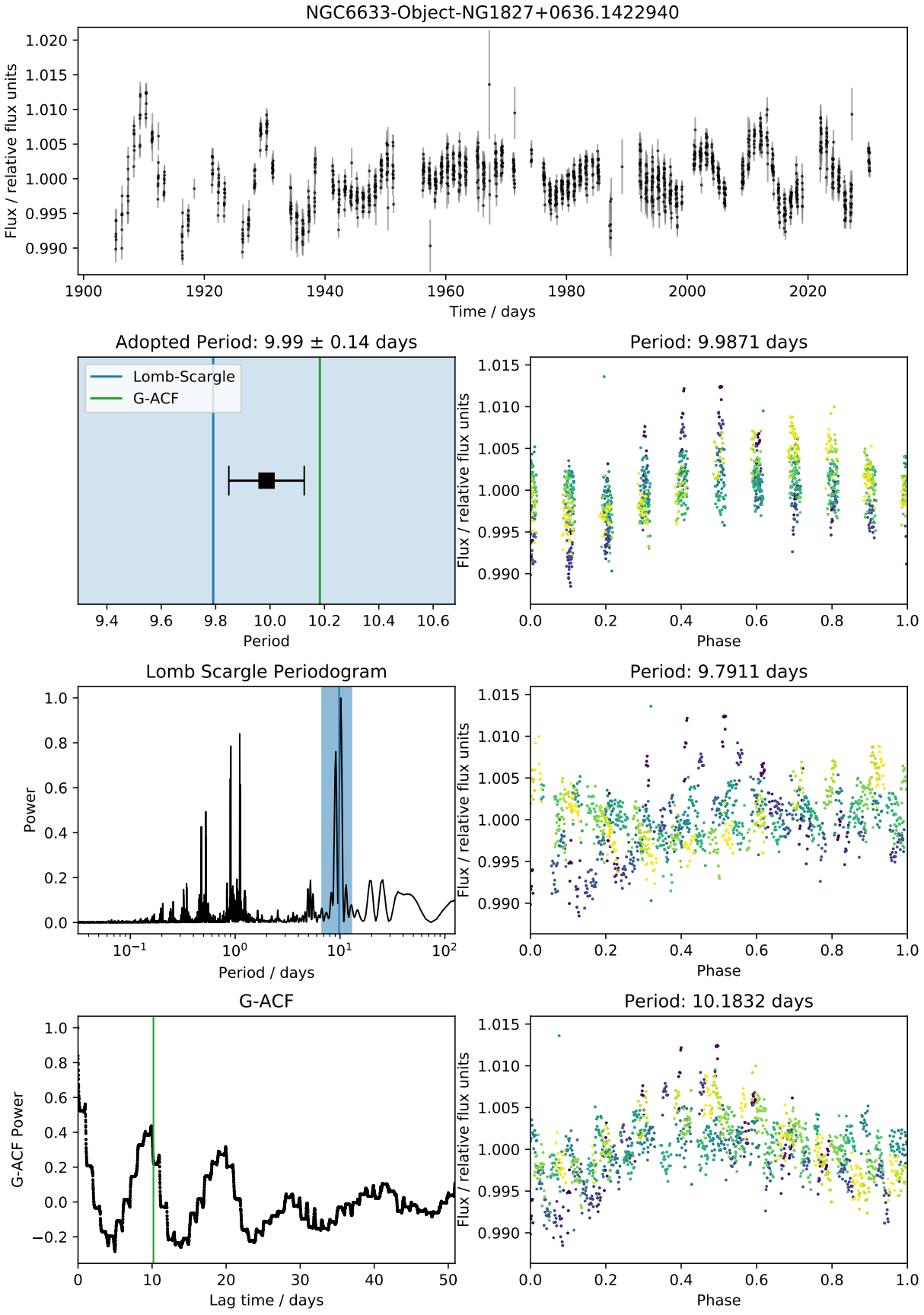

NGC6633-Object-NG1827+0636.870224
Flux / relative flux units
Time / days
Adopted Period: 9.198 ± 0.012 days
Lomb-Scargle
G-ACF
Period
Period: 9.1981 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 9.2153 days
G-ACF
G-ACF Power
Lag time / days
Period: 9.1810 days

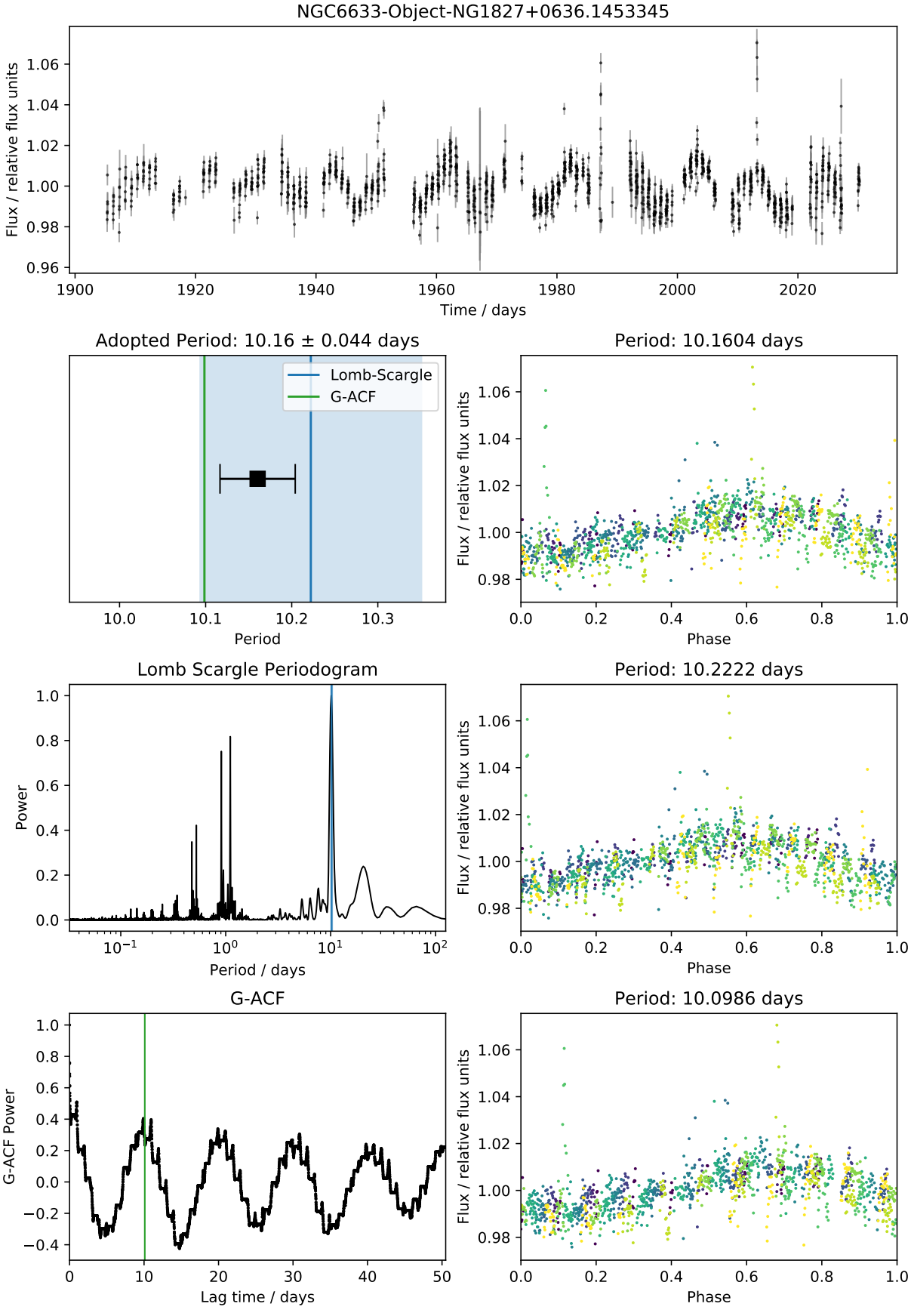

NGC6633-Object-NG1827+0636.1258396
Flux / relative flux units
Time / days
Adopted Period: 9.401 ± 0.011 days
Lomb-Scargle
G-ACF
Period
Period: 9.4011 days
Flux / relative flux units
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 9.3854 days
Flux / relative flux units
Phase
G-ACF
G-ACF Power
Lag time / days
Period: 9.4167 days
Flux / relative flux units
Phase

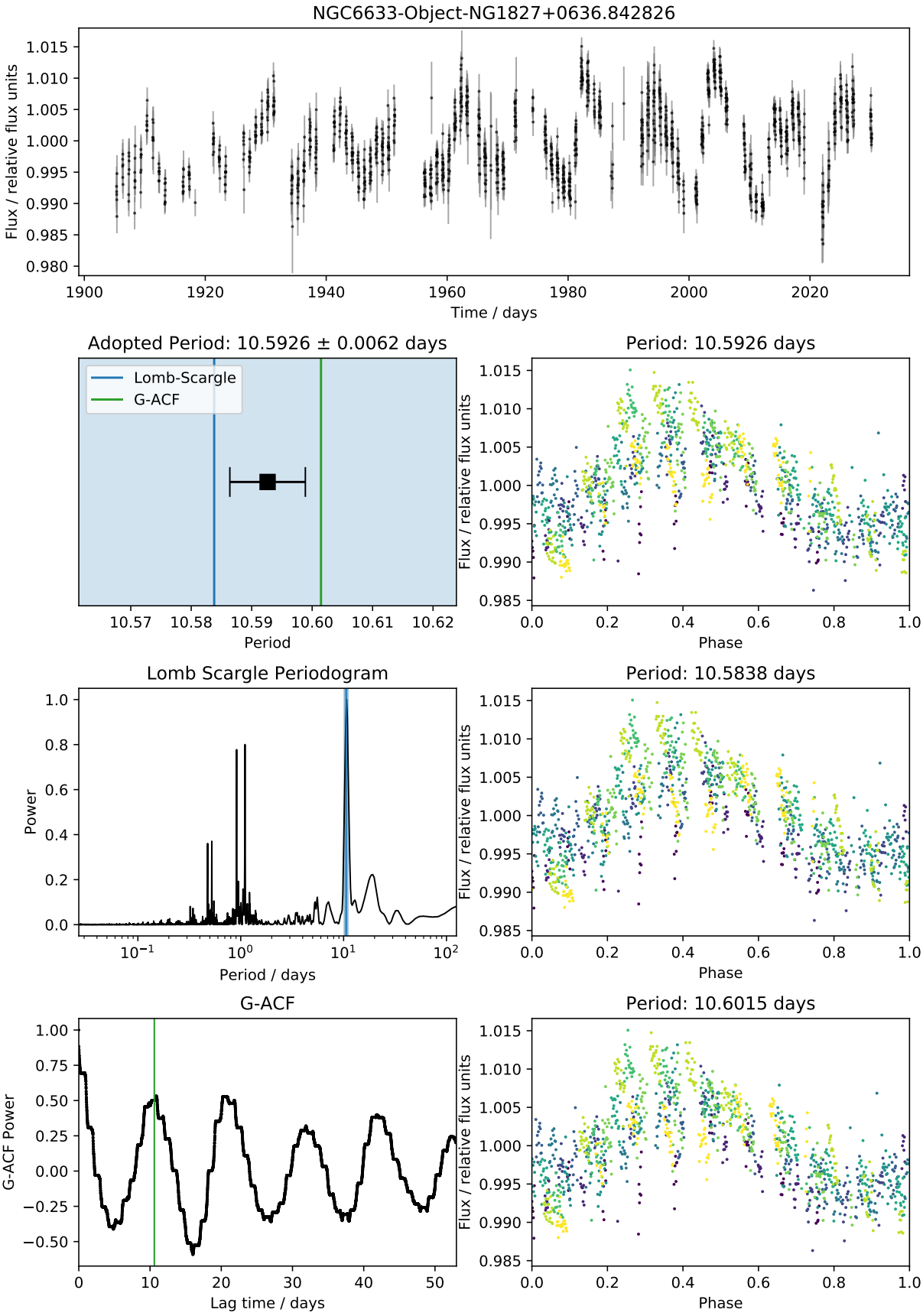

NGC6633-Object-NG1827+0636.539015
Flux / relative flux units
Time / days
Adopted Period: 9.435 ± 0.031 days
Lomb-Scargle
G-ACF
Period
Period: 9.4347 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 9.3910 days
G-ACF
G-ACF Power
Lag time / days
Period: 9.4784 days

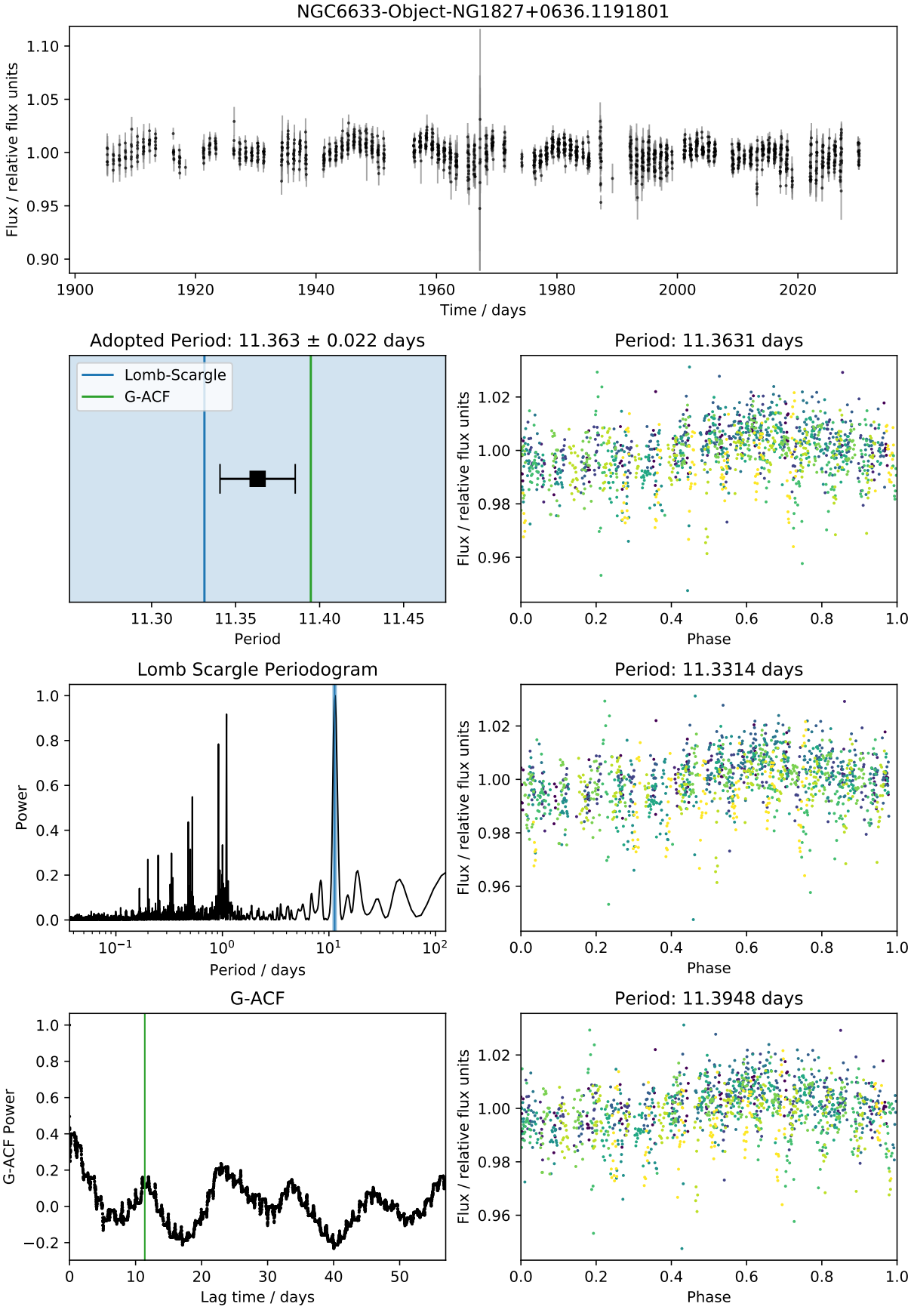

NGC6633-Object-NG1827+0636.1764944
Flux / relative flux units
Time / days
Adopted Period: 9.54 ± 0.038 days
Lomb-Scargle
G-ACF
Period
Period: 9.5403 days
Flux / relative flux units
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 9.5943 days
Flux / relative flux units
Phase
G-ACF
G-ACF Power
Lag time / days
Period: 9.4863 days
Flux / relative flux units
Phase

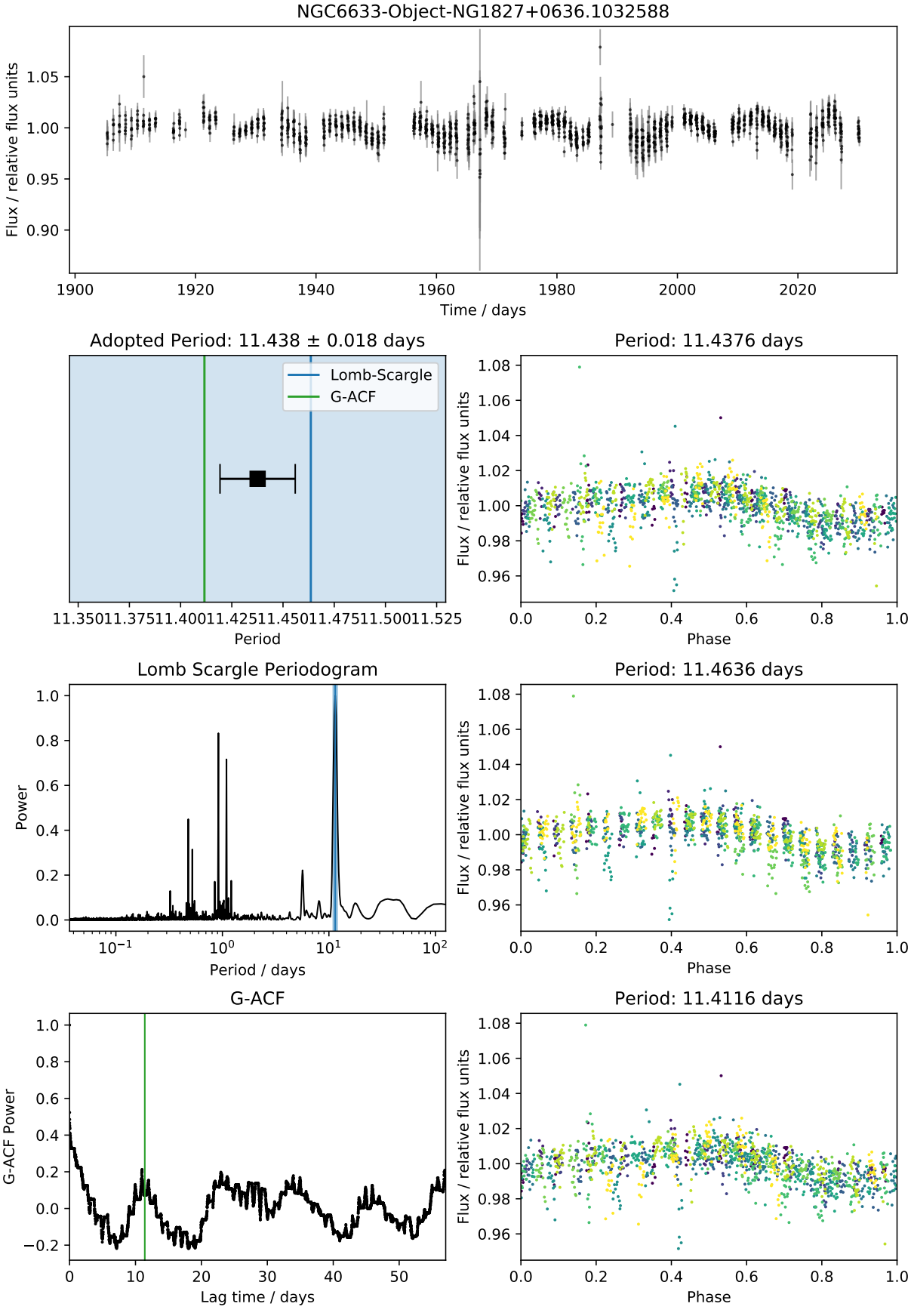

NGC6633-Object-NG1827+0636.1394490
Flux / relative flux units
Residuals
Time / days
Adopted Period: 9.578 ± 0.087 days
Lomb-Scargle
G-ACF
Gaussian Process Regression
Period
Period: 9.5779 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 9.5038 days
G-ACF
G-ACF Power
Lag time / days
Period: 9.7876 days
Gaussian Process Model
Period Posterior
Period / days
Period: 9.4422 days

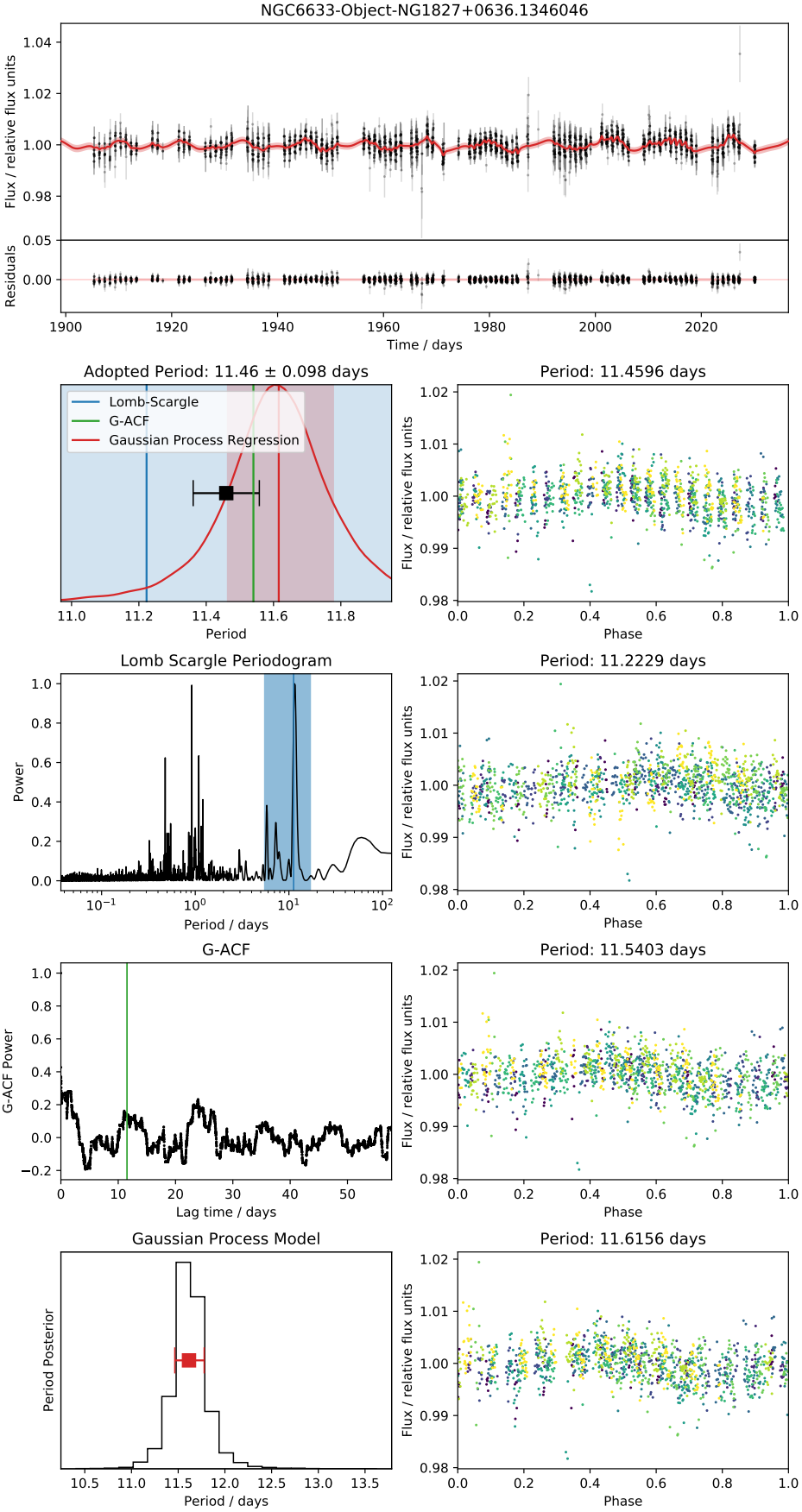

NGC6633-Object-NG1827+0636.1422940
Flux / relative flux units
Time / days
Adopted Period: 9.99 ± 0.14 days
Lomb-Scargle
G-ACF
Period
Period: 9.9871 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 9.7911 days
G-ACF
G-ACF Power
Lag time / days
Period: 10.1832 days

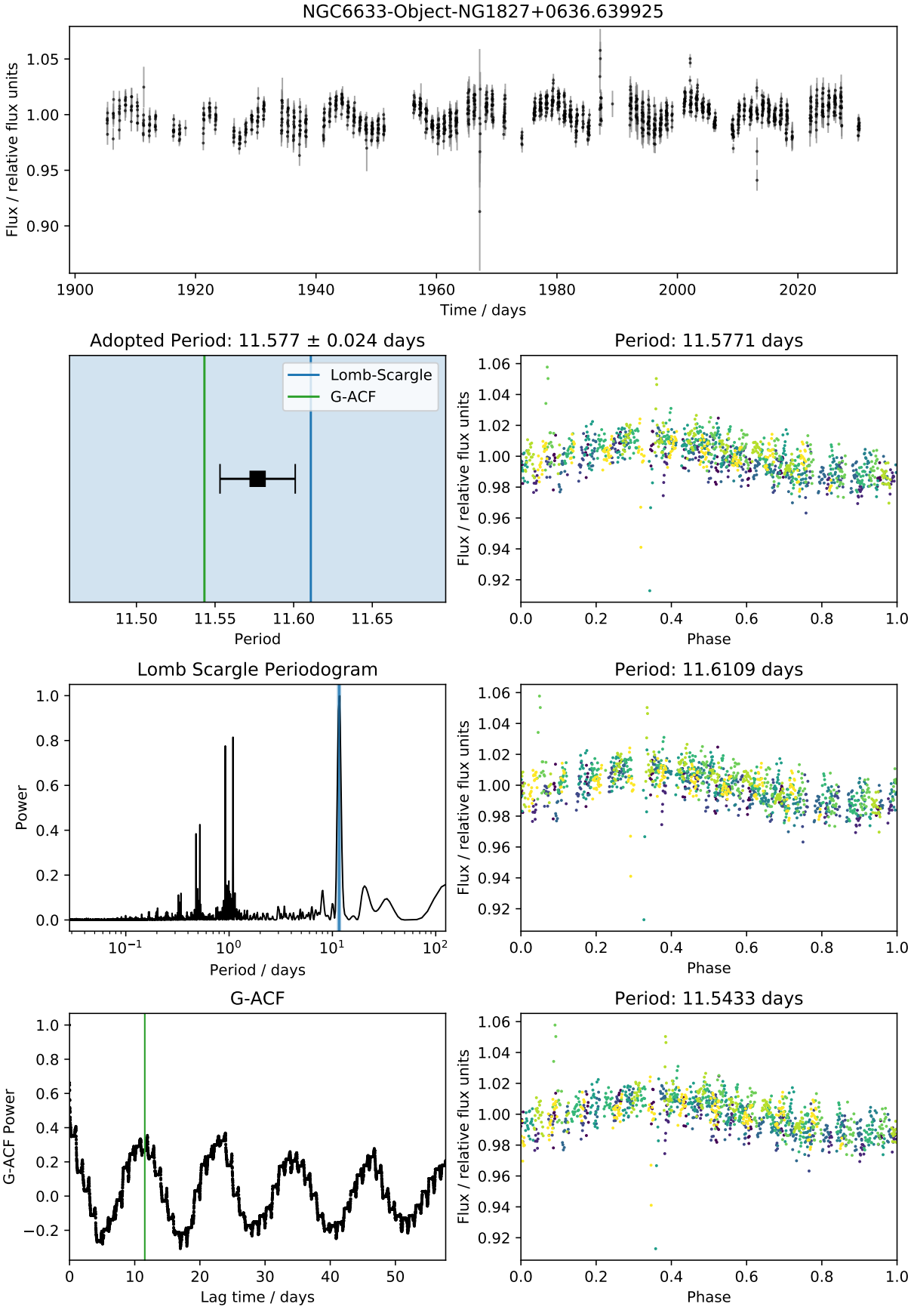

NGC6633-Object-NG1827+0636.1453345
Flux / relative flux units
Time / days
Adopted Period: 10.16 ± 0.044 days
Lomb-Scargle
G-ACF
Period
Period: 10.1604 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 10.2222 days
G-ACF
G-ACF Power
Lag time / days
Period: 10.0986 days

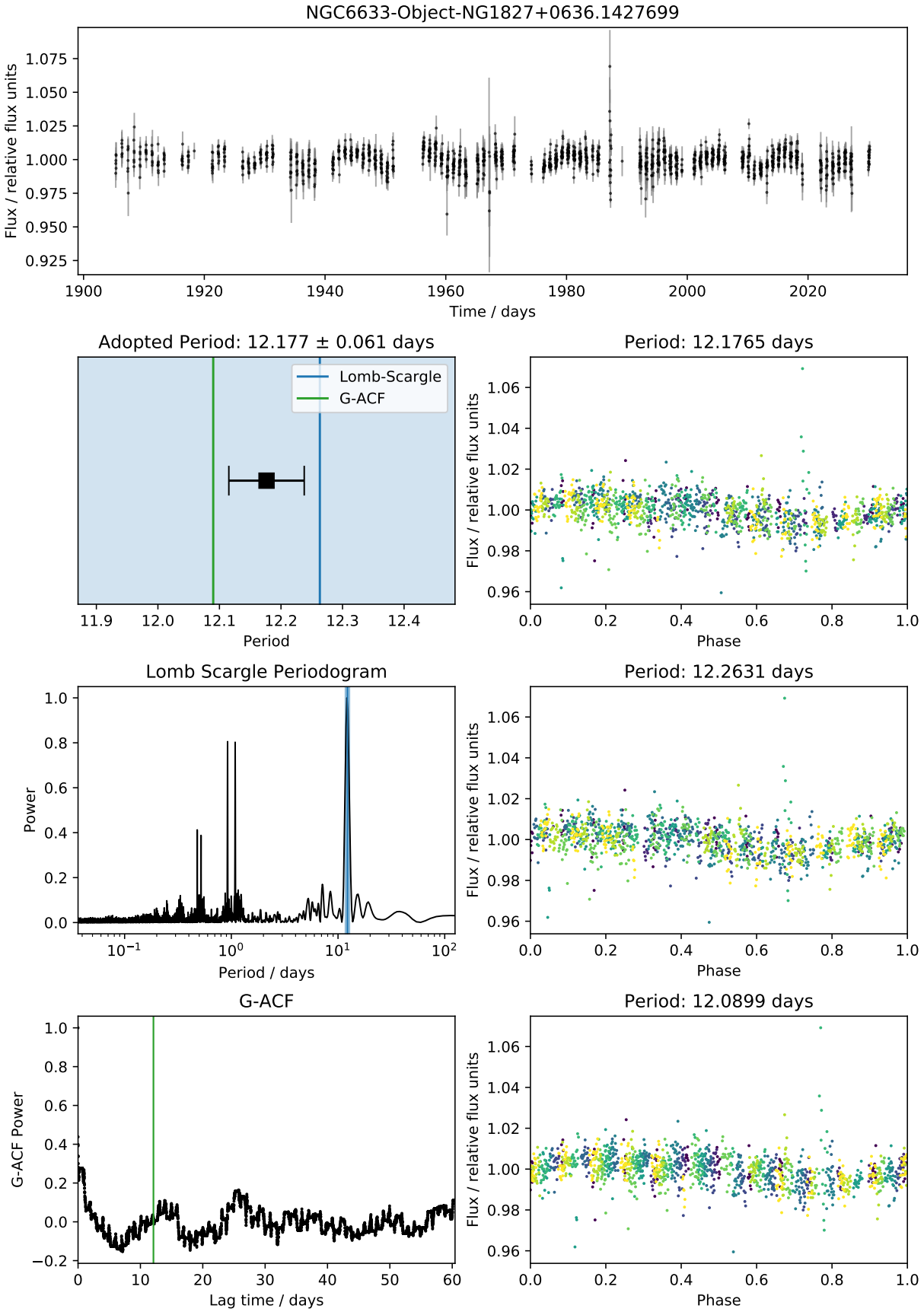

NGC6633-Object-NG1827+0636.842826
Flux / relative flux units
Time / days
Adopted Period: 10.5926 ± 0.0062 days
Lomb-Scargle
G-ACF
Period
Period: 10.5926 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 10.5838 days
G-ACF
G-ACF Power
Lag time / days
Period: 10.6015 days

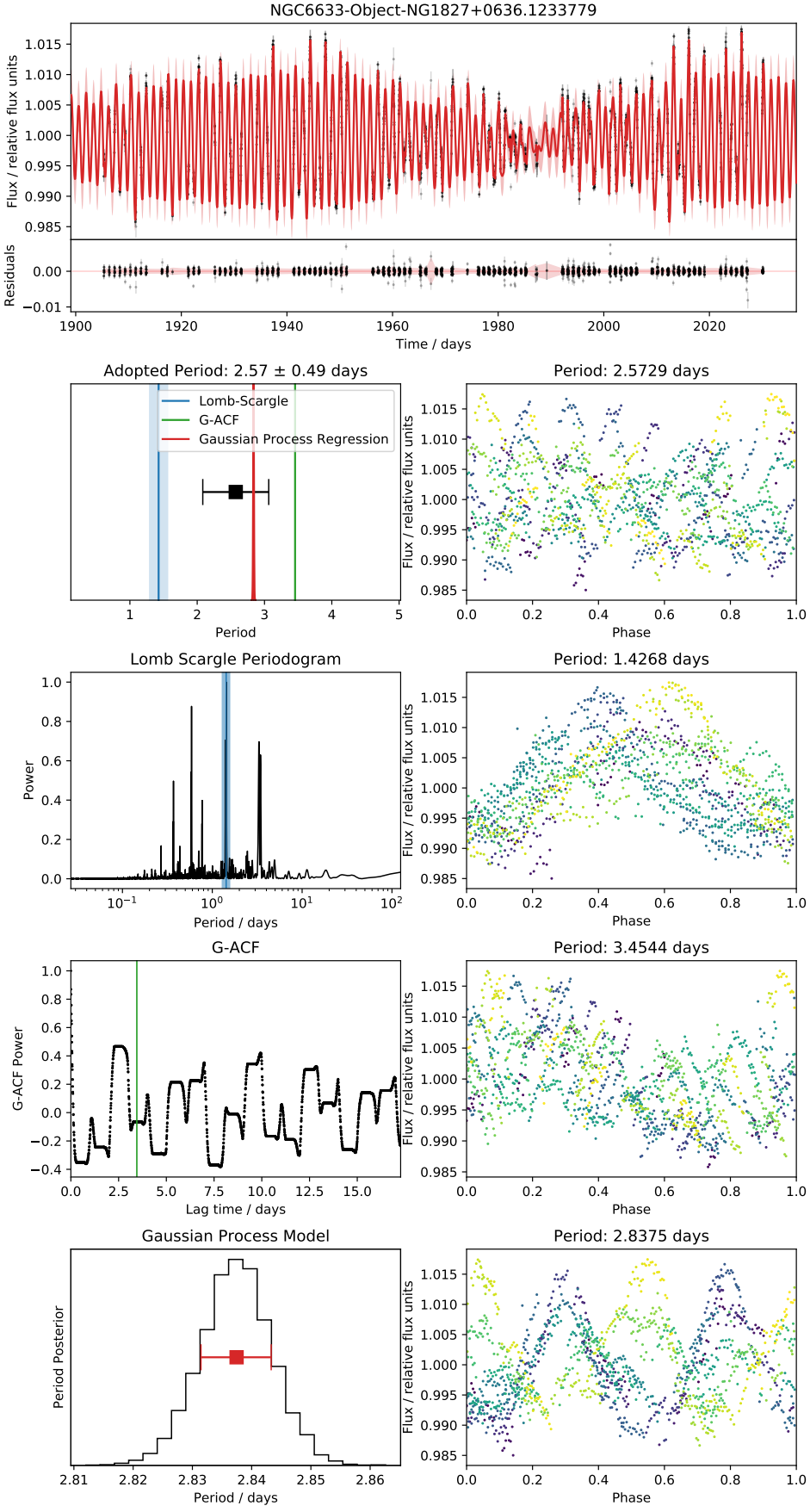

NGC6633-Object-NG1827+0636.1191801
Flux / relative flux units
Time / days
Adopted Period: 11.363 ± 0.022 days
Lomb-Scargle
G-ACF
Period
Period: 11.3631 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 11.3314 days
G-ACF
G-ACF Power
Lag time / days
Period: 11.3948 days

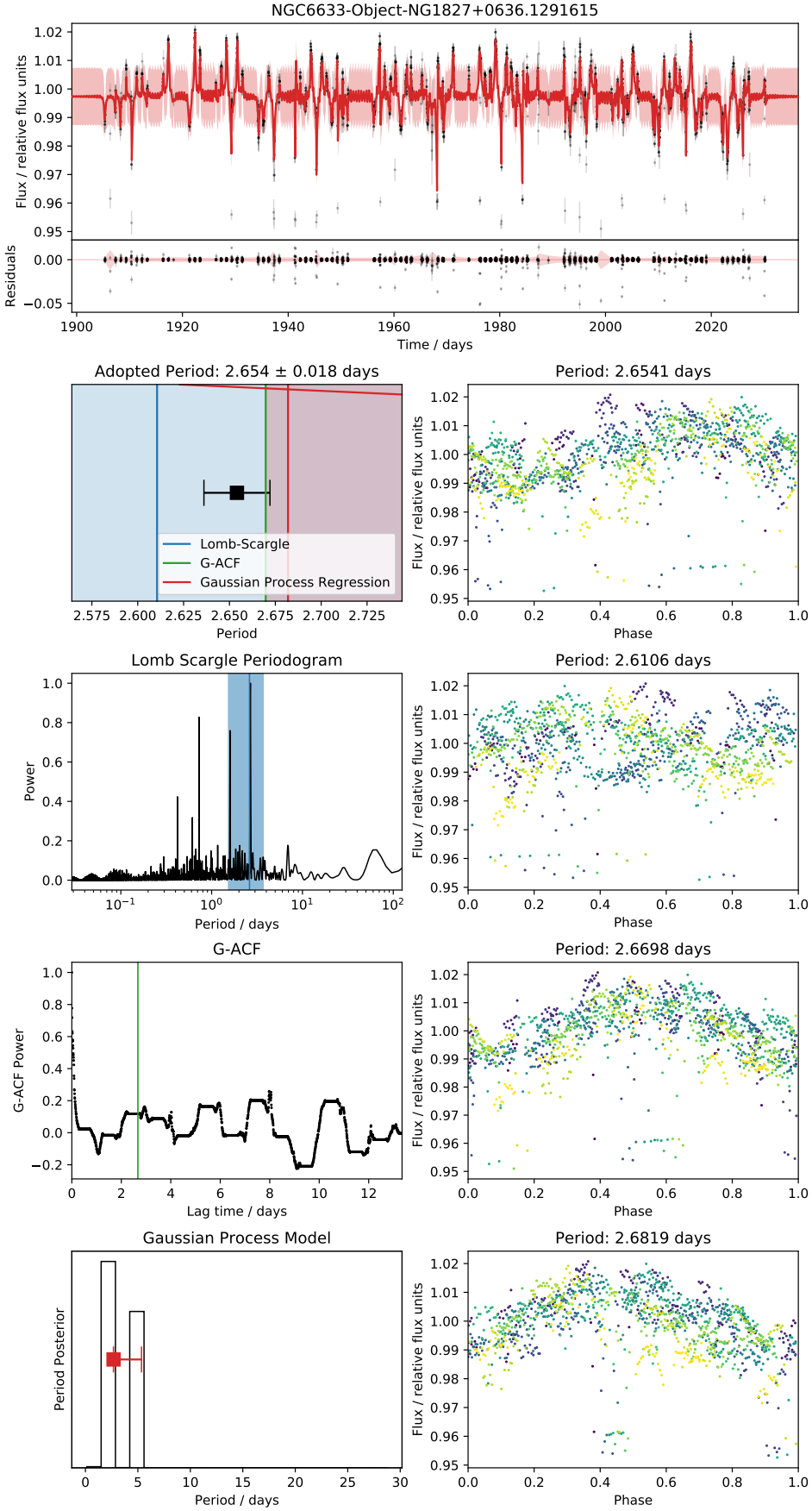

NGC6633-Object-NG1827+0636.1032588
Flux / relative flux units
Time / days
Adopted Period: 11.438 ± 0.018 days
Lomb-Scargle
G-ACF
Period
Period: 11.4376 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 11.4636 days
Phase
G-ACF
G-ACF Power
Lag time / days
Period: 11.4116 days
Phase

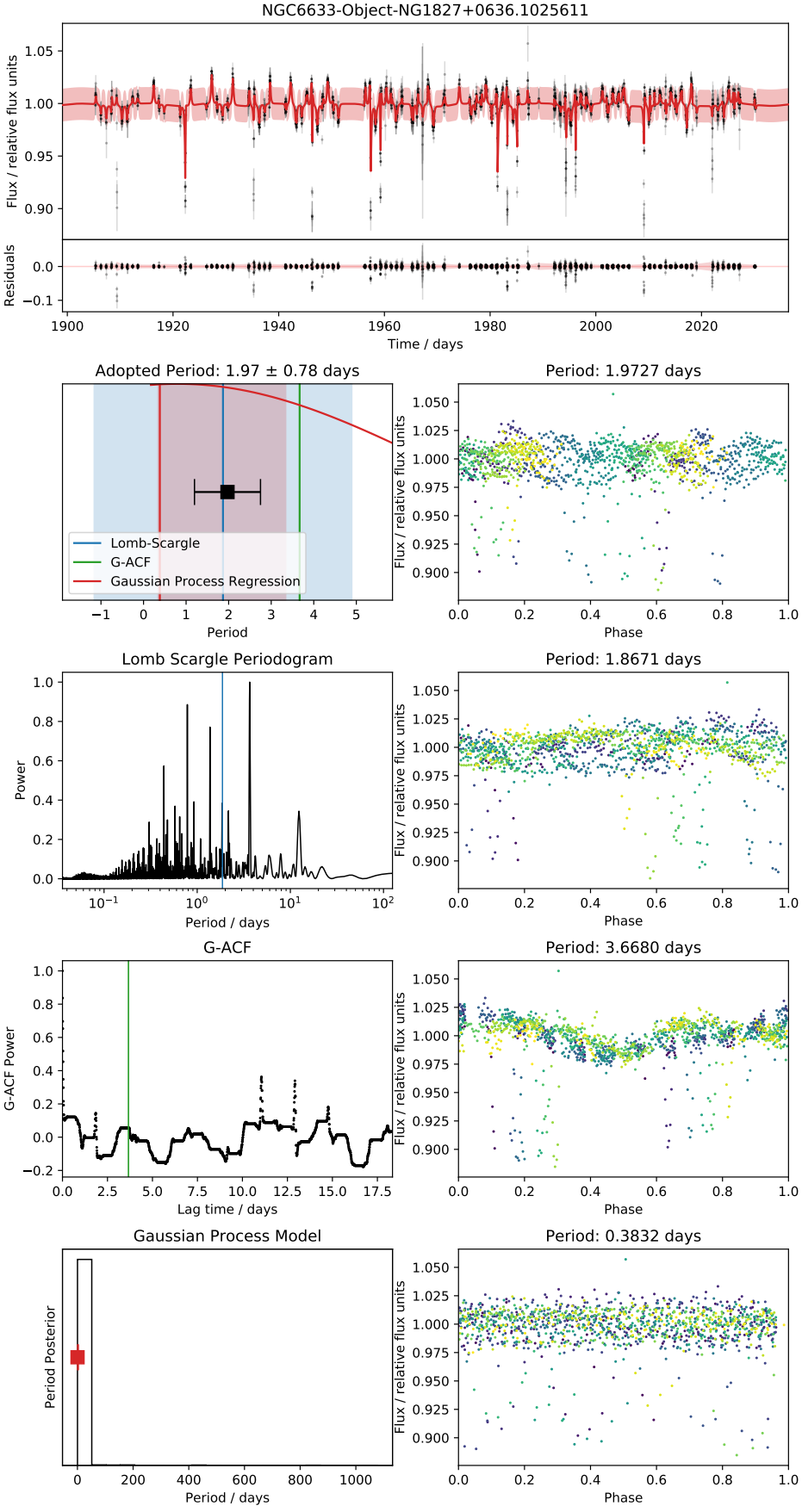

NGC6633-Object-NG1827+0636.1346046
Flux / relative flux units
Residuals
Time / days
Adopted Period: 11.46 ± 0.098 days
Lomb-Scargle
G-ACF
Gaussian Process Regression
Period
Period: 11.4596 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 11.2229 days
G-ACF
G-ACF Power
Lag time / days
Period: 11.5403 days
Gaussian Process Model
Period Posterior
Period / days
Period: 11.6156 days

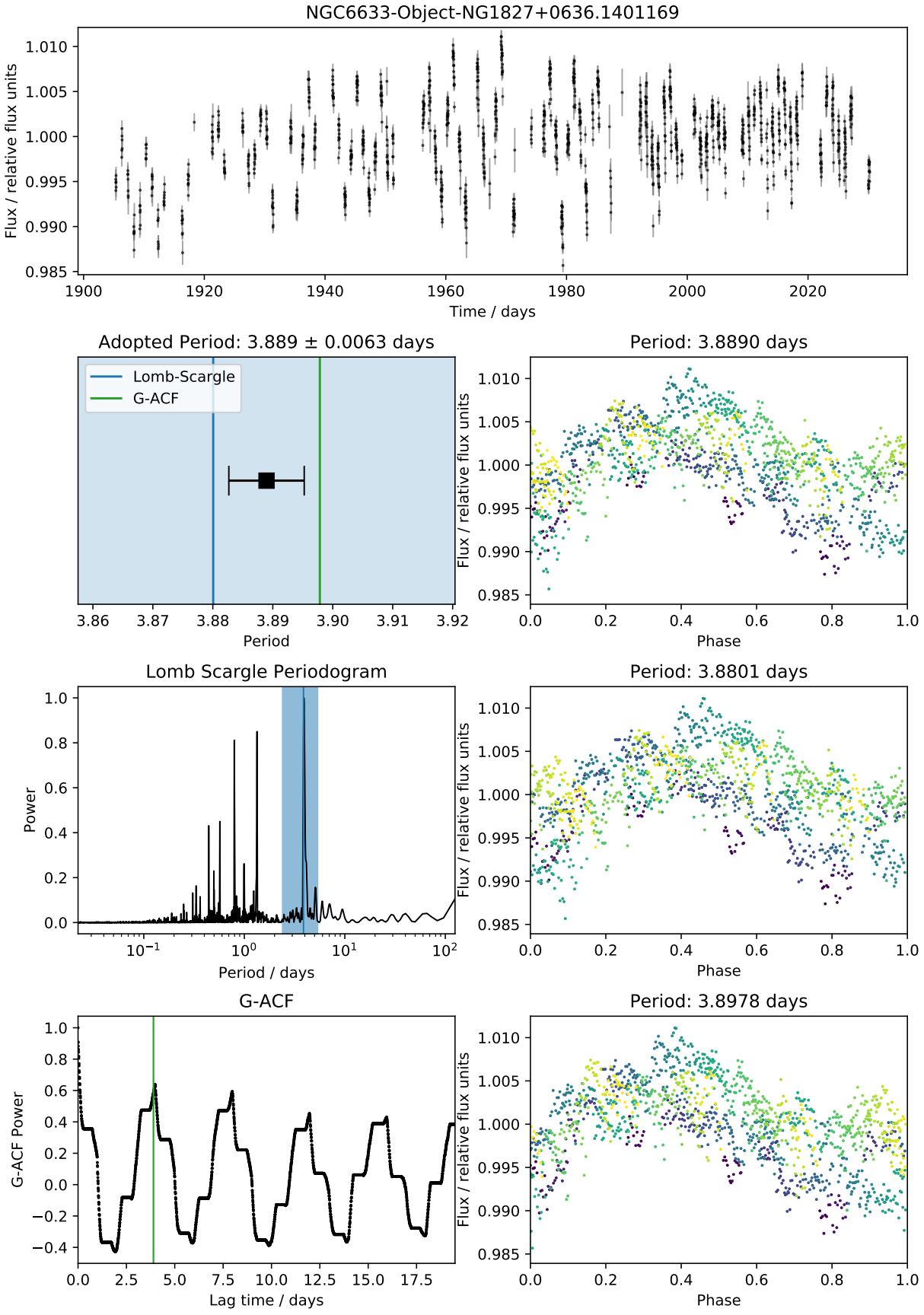

NGC6633-Object-NG1827+0636.639925
Flux / relative flux units
Time / days
Adopted Period: 11.577 ± 0.024 days
Lomb-Scargle
G-ACF
Period
Period: 11.5771 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 11.6109 days
G-ACF
G-ACF Power
Lag time / days
Period: 11.5433 days

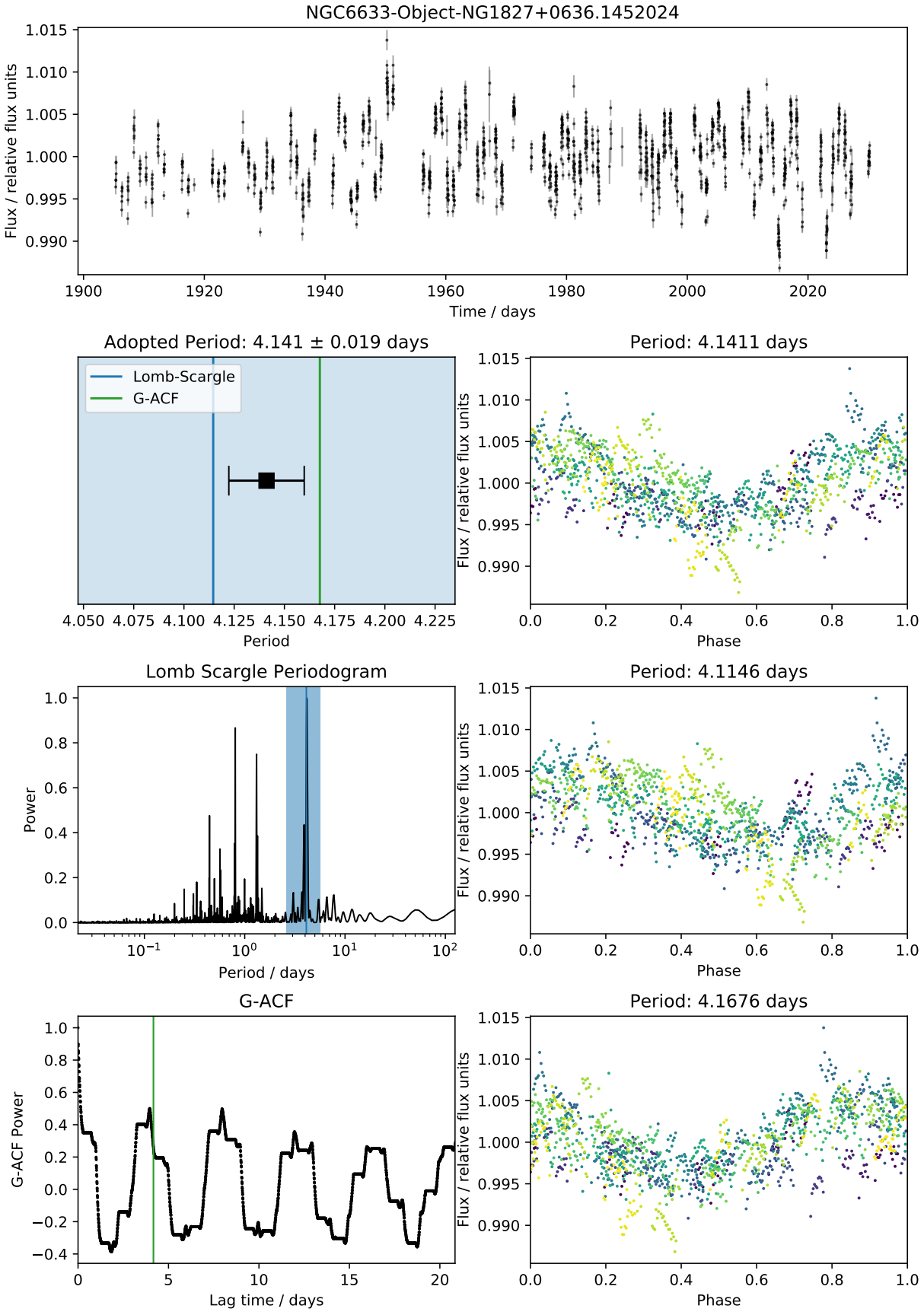

NGC6633-Object-NG1827+0636.1427699
Flux / relative flux units
Time / days
Adopted Period: 12.177 ± 0.061 days
Lomb-Scargle
G-ACF
Period
Period: 12.1765 days
Flux / relative flux units
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 12.2631 days
Flux / relative flux units
Phase
G-ACF
G-ACF Power
Lag time / days
Period: 12.0899 days
Flux / relative flux units
Phase

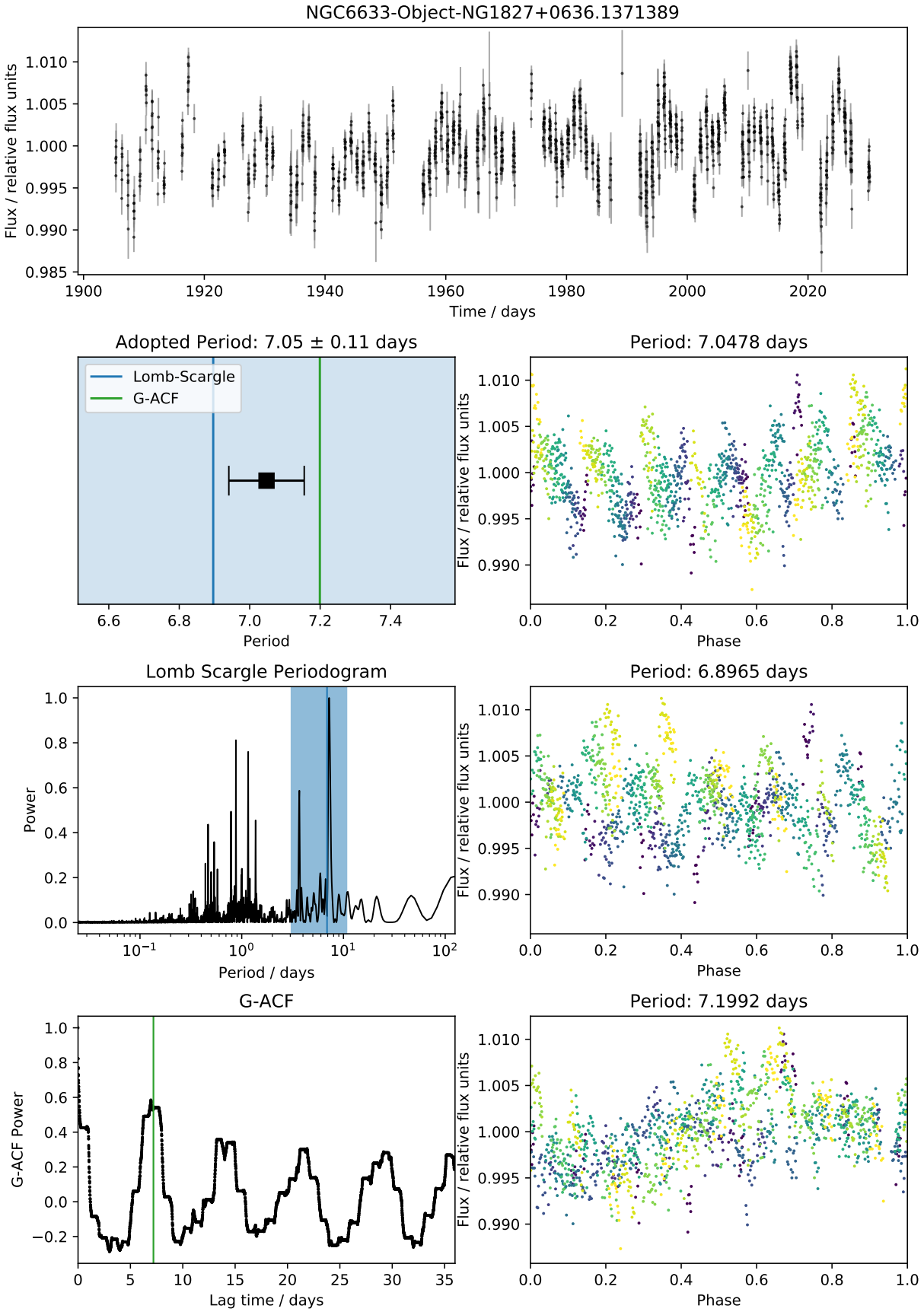

NGC6633-Object-NG1827+0636.1233779
Flux / relative flux units
Residuals
Time / days
Adopted Period: 2.57 ± 0.49 days
Lomb-Scargle
G-ACF
Gaussian Process Regression
Period
Period: 2.5729 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 1.4268 days
G-ACF
G-ACF Power
Lag time / days
Period: 3.4544 days
Gaussian Process Model
Period Posterior
Period / days
Period: 2.8375 days

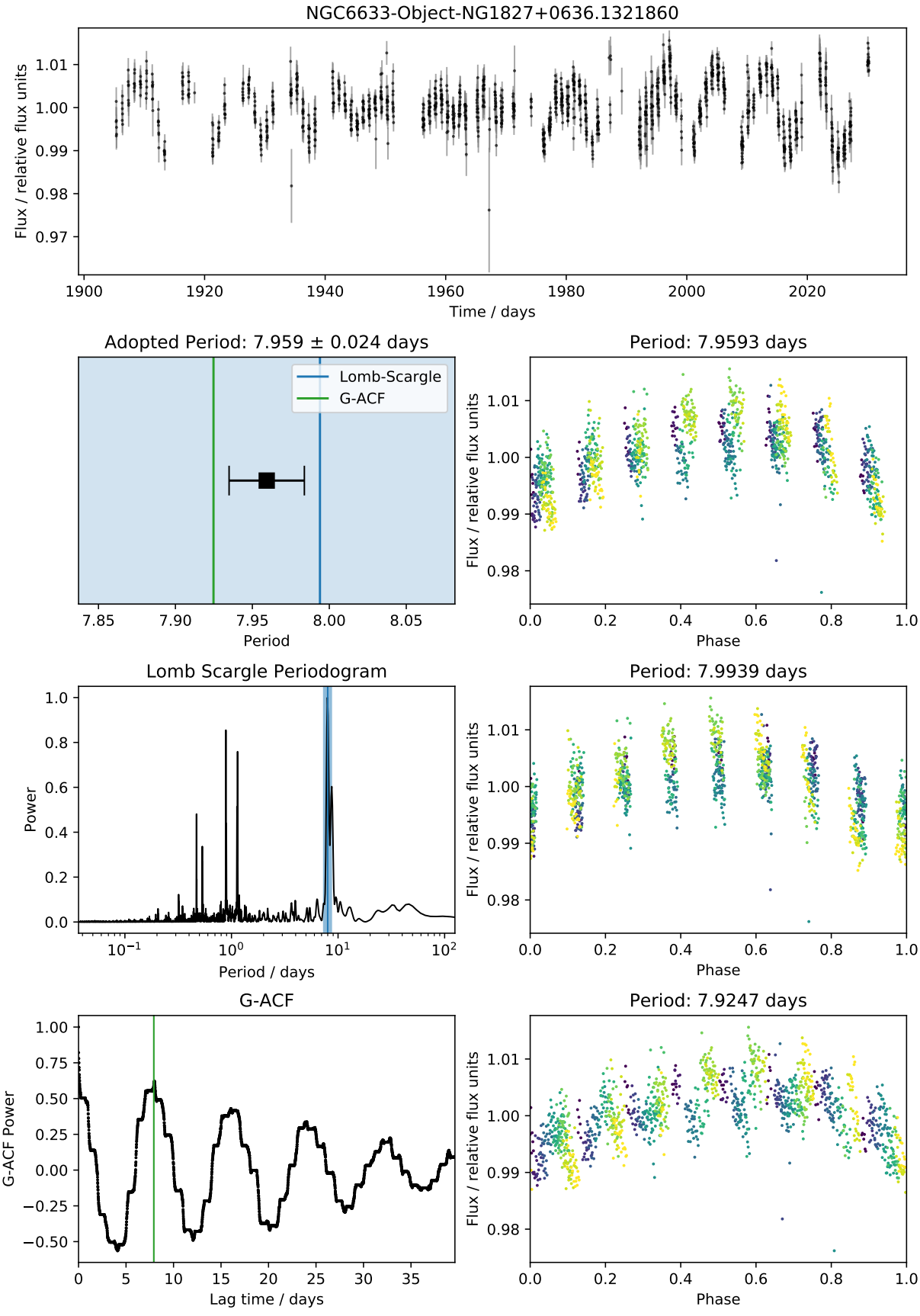

NGC6633-Object-NG1827+0636.1291615
Flux / relative flux units
Residuals
Time / days
Adopted Period: 2.654 ± 0.018 days
Lomb-Scargle
G-ACF
Gaussian Process Regression
Period
Period: 2.6541 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 2.6106 days
G-ACF
G-ACF Power
Lag time / days
Period: 2.6698 days
Gaussian Process Model
Period Posterior
Period / days
Period: 2.6819 days

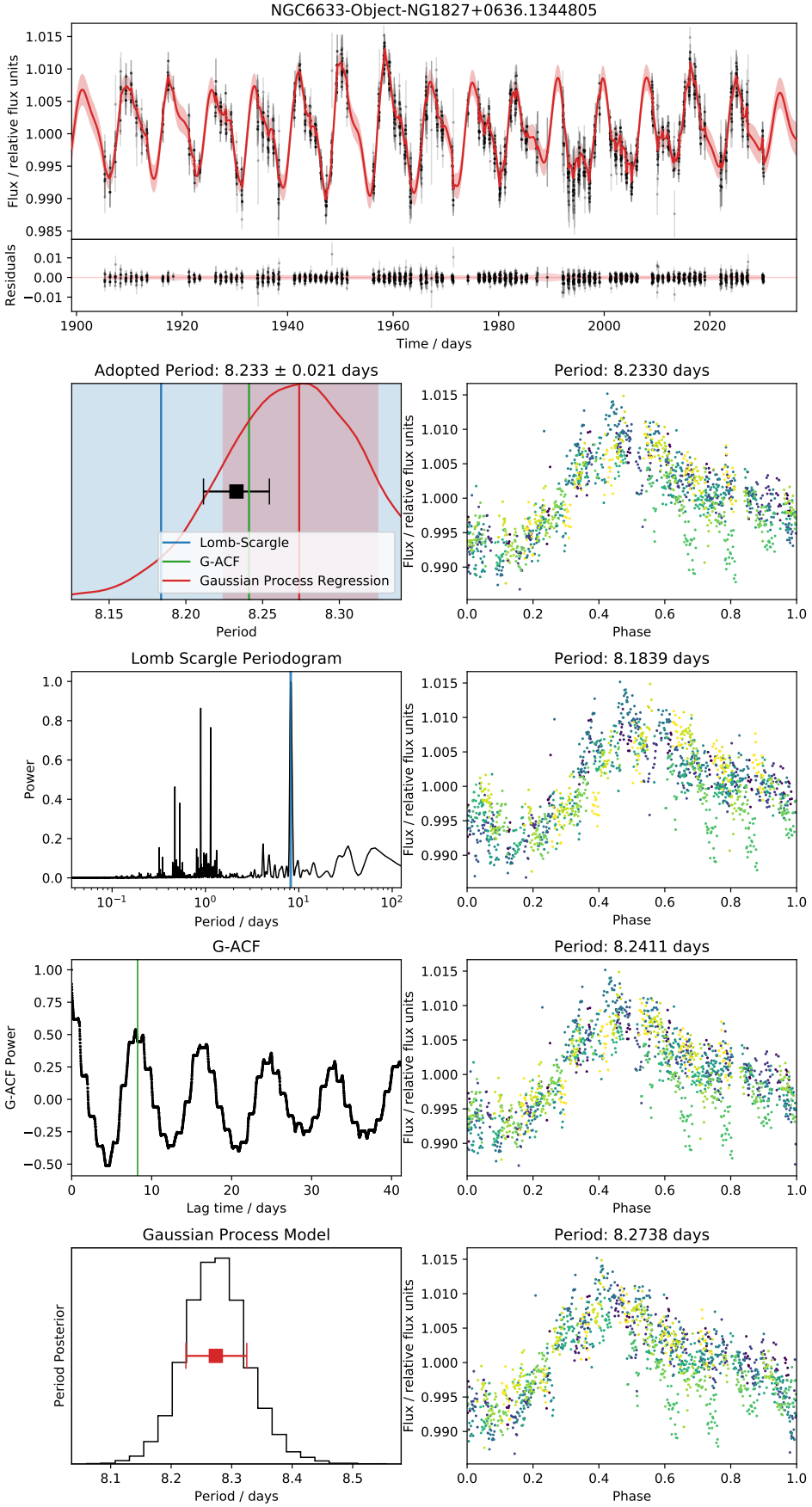

NGC6633-Object-NG1827+0636.1025611
Flux / relative flux units
Residuals
Time / days
Adopted Period: 1.97 ± 0.78 days
Lomb-Scargle
G-ACF
Gaussian Process Regression
Period
Period: 1.9727 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 1.8671 days
G-ACF
G-ACF Power
Lag time / days
Period: 3.6680 days
Gaussian Process Model
Period Posterior
Period / days
Period: 0.3832 days

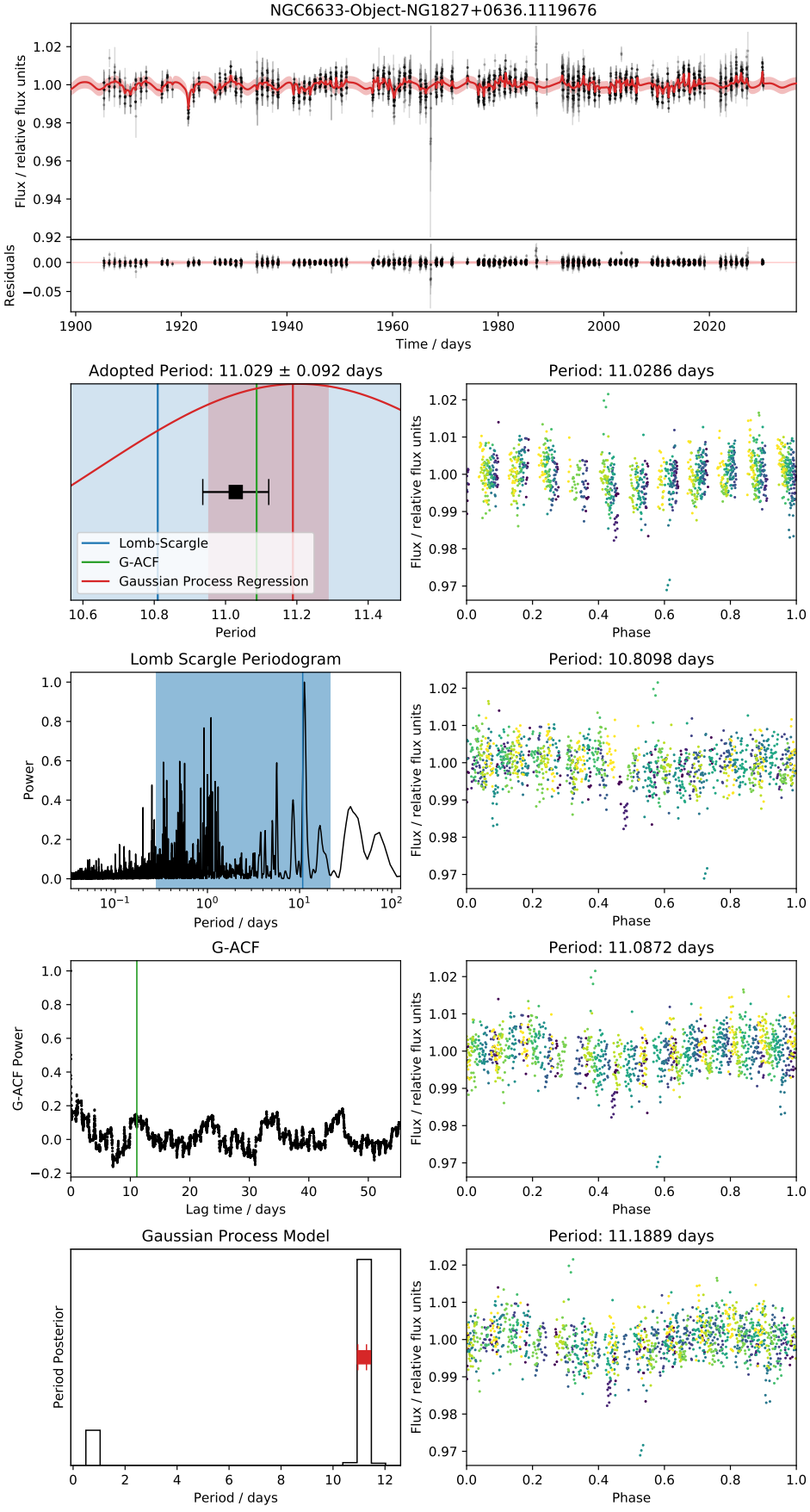

NGC6633-Object-NG1827+0636.1401169
Flux / relative flux units
Time / days
Adopted Period: 3.889 ± 0.0063 days
Lomb-Scargle
G-ACF
Period
Period: 3.8890 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 3.8801 days
Phase
G-ACF
G-ACF Power
Lag time / days
Period: 3.8978 days
Phase

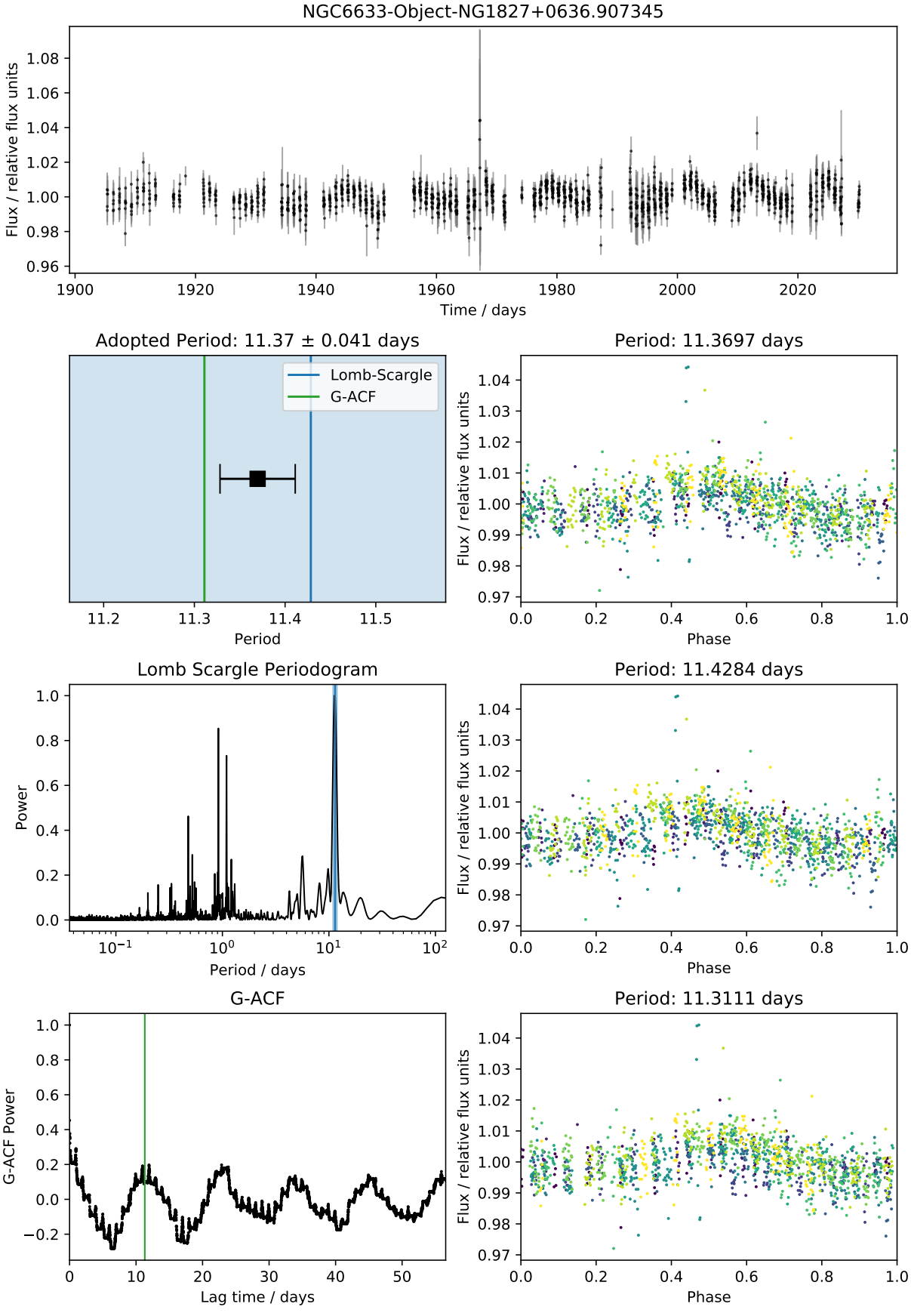

NGC6633-Object-NG1827+0636.1452024
Flux / relative flux units
Time / days
Adopted Period: 4.141 ± 0.019 days
Lomb-Scargle
G-ACF
Period
Period: 4.1411 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 4.1146 days
Phase
G-ACF
G-ACF Power
Lag time / days
Period: 4.1676 days
Phase

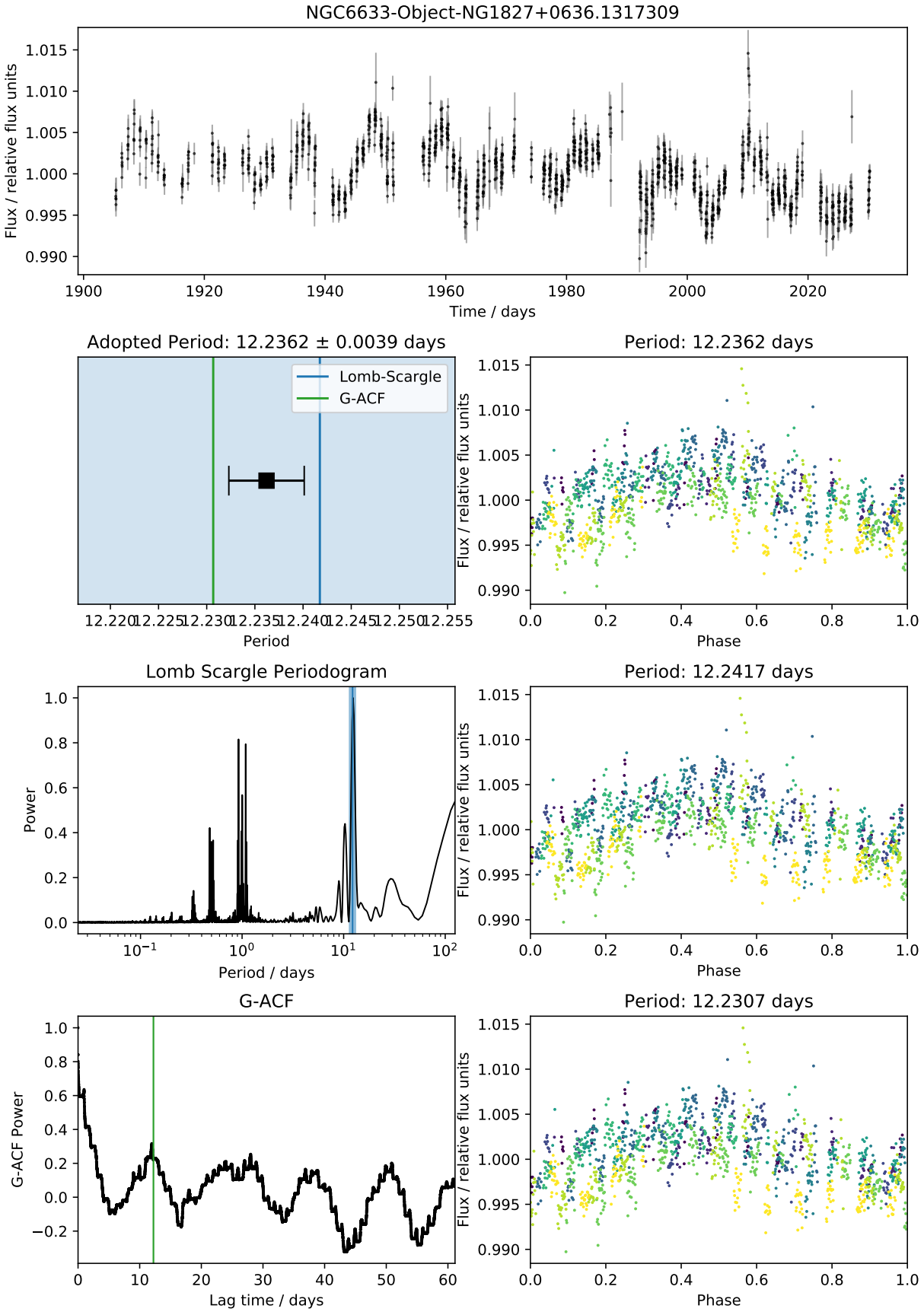

NGC6633-Object-NG1827+0636.1371389
Flux / relative flux units
Time / days
Adopted Period: 7.05 ± 0.11 days
Lomb-Scargle
G-ACF
Period
Period: 7.0478 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 6.8965 days
G-ACF
G-ACF Power
Lag time / days
Period: 7.1992 days


NGC6633-Object-NG1827+0636.1321860
Flux / relative flux units
Time / days
Adopted Period: 7.959 ± 0.024 days
Lomb-Scargle
G-ACF
Period
Period: 7.9593 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 7.9939 days
G-ACF
G-ACF Power
Lag time / days
Period: 7.9247 days


NGC6633-Object-NG1827+0636.1344805
Flux / relative flux units
Residuals
Time / days
Adopted Period: 8.233 ± 0.021 days
Lomb-Scargle
G-ACF
Gaussian Process Regression
Period
Period: 8.2330 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 8.1839 days
G-ACF
G-ACF Power
Lag time / days
Period: 8.2411 days
Gaussian Process Model
Period Posterior
Period / days
Period: 8.2738 days


NGC6633-Object-NG1827+0636.1119676
Flux / relative flux units
Residuals
Time / days
Adopted Period: 11.029 ± 0.092 days
Lomb-Scargle
G-ACF
Gaussian Process Regression
Period
Period: 11.0286 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 10.8098 days
G-ACF
G-ACF Power
Lag time / days
Period: 11.0872 days
Gaussian Process Model
Period Posterior
Period / days
Period: 11.1889 days


NGC6633-Object-NG1827+0636.907345
Flux / relative flux units
Time / days
Adopted Period: 11.37 ± 0.041 days
Lomb-Scargle
G-ACF
Period
Period: 11.3697 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 11.4284 days
Phase
G-ACF
G-ACF Power
Lag time / days
Period: 11.3111 days
Phase


NGC6633-Object-NG1827+0636.1317309
Flux / relative flux units
Time / days
Adopted Period: 12.2362 ± 0.0039 days
Lomb-Scargle
G-ACF
Period
Period: 12.2362 days
Phase
Lomb Scargle Periodogram
Power
Period / days
Period: 12.2417 days
Phase
G-ACF
G-ACF Power
Lag time / days
Period: 12.2307 days
Phase

# References


Acton, J. S., Goad, M. R., Casewell, S. L., et al. 2020, MNRAS, 498, 3115

Affer, L., Micela, G., Favata, F., & Flaccomio, E. 2012, MNRAS, 424, 11

Aigrain, S., Hodgkin, S., Irwin, J., et al. 2007, MNRAS, 375, 29

Aigrain, S., Hodgkin, S. T., Irwin, M. J., Lewis, J. R., & Roberts, S. J. 2015, MNRAS, 447, 2880

Aigrain, S., Parviainen, H., & Pope, B. J. S. 2016, MNRAS, 459, 2408

Alonso, R., Brown, T. M., Charbonneau, D., et al. 2007, PASP, 366, 13

Alsubai, K. A., Parley, N. R., Bramich, D. M., et al. 2013, Acta Astron., 63, 465

Andronov, I. L., & Chinarova, L. L. 2005, in 14th European Workshop on White Dwarfs, ed. D. Koester & S. Moehler, Vol. 334 (Kiel: Astronomical Society of the Pacific), 659

Angus, R., Aigrain, S., Foreman-Mackey, D., & McQuillan, A. 2015, MNRAS, 450, 1787

Angus, R., Morton, T., Aigrain, S., Foreman-Mackey, D., & Rajpaul, V. 2018, MNRAS, 474, 2094

Angus, R., Morton, T. D., Foreman-Mackey, D., et al. 2019, AJ, 158, 173

Angus, R., Beane, A., Price-Whelan, A. M., et al. 2020, AJ, 160, 90

Ansdell, M., Ioannou, Y., Osborn, H. P., et al. 2018, ApJ, 869, L7

Armstrong, D. J., Pollacco, D., & Santerne, A. 2017, Astrophysics Source Code Library, 2017ascl.soft03010A

Armstrong, D. J., Kirk, J., Lam, K. W. F., et al. 2016, MNRAS, 456, 2260

Armstrong, D. J., Günther, M. N., McCormac, J., et al. 2018, MNRAS, 478, 4225

Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, ARA&A, 47, 481

Attridge, J. M., & Herbst, W. 1992, ApJ, 398, L61

Auvergne, M., Bodin, P., Boisnard, L., et al. 2009, A&A, 506, 411

Bailer-Jones, C. A. L. 2015, PASP, 127, 994

Bailer-Jones, C. A. L., Rybizki, J., Fouesneau, M., Demleitner, M., & Andrae, R. 2021, AJ, 161, 147

Bailer-Jones, C. A. L., Rybizki, J., Fouesneau, M., Mantelet, G., & Andrae, R. 2018, AJ, 156, 58

Bakos, G., Lázár, J., Papp, I., Sári, P., & Green, E. 2002, PASP, 114, 974
Bakos, G., Noyes, R., Kovács, G., et al. 2004, PASP, 116, 266
Ball, N. M., & Brunner, R. J. 2010, International Journal of Modern Physics D, 19, 1049
Baraffe, I., Chabrier, G., Allard, F., & Hauschildt, P. H. 1998, A&A, 337, 403
Baraffe, I., Homeier, D., Allard, F., & Chabrier, G. 2015, A&A, 577, A42
Baranne, A., Queloz, D., Mayor, M., et al. 1996, A&AS, 119, 373
Barnes, J. R., Jeffers, S. V., Haswell, C. A., et al. 2017, MNRAS, 471, 811
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., Walkowicz, L. M., Batalha, N., et al. 2010, AJ, 141, 20
Benedict, G. F., McArthur, B. E., Feast, M. W., et al. 2007, AJ, 133, 1810
Berger, T. A., Huber, D., Gaidos, E., & van Saders, J. L. 2018, ApJ, 866, 99
Bildsten, L., Brown, E. F., Matzner, C. D., & Ushomirsky, G. 1997, ApJ, 482, 442
Bjørnstad, O. N., & Falck, W. 2001, Environmental and Ecological Statistics, 8, 53
Blažko, S. 1907, Astronomische Nachrichten, 175, 325
Booth, R. S., Poppenhaeger, K., Watson, C. A., Aguirre, V. S., & Wolk, S. J. 2017, MNRAS, 471, 1012
Borissova, J., Rejkuba, M., Minniti, D., Catelan, M., & Ivanov, V. D. 2009, A&A, 502, 505
Boro Saikia, S., Marvin, C. J., Jeffers, S. V., et al. 2018, A&A, 616, A108
Borucki, W. J., Koch, D., Basri, G., et al. 2010, Science, 327, 977
Bouma, L. G., Curtis, J. L., Hartman, J. D., Winn, J. N., & Bakos, G. Á. 2021, AJ, 162, 197
Bouvier, J., Forestini, M., & Allain, S. 1997, A&A, 326, 1023
Bressan, A., Marigo, P., Girardi, L., et al. 2012, MNRAS, 427, 127
Briegal, J. T., Gillen, E., Queloz, D., et al. 2022, MNRAS, 513, 420
Brown, A. G. A., Vallenari, A., Prusti, T., et al. 2016, A&A, 595, A2
Brown, A. G. A., Vallenari, A., Prusti, T., et al. 2018, A&A, 616, A1
Bryant, E. M., Bayliss, D., Nielsen, L. D., et al. 2020a, MNRAS, 499, 3139
Bryant, E. M., Bayliss, D., McCormac, J., et al. 2020b, MNRAS, 494, 5872
Burdanov, A., Delrez, L., Gillon, M., & Jehin, E. 2018, in Handbook of Exoplanets, ed. H. J. Deeg & J. A. Belmonte (Cham: Springer International Publishing), 1007
Burke, C. J., Pinsonneault, M. H., & Sills, A. 2004, ApJ, 604, 272
Burton, J. R., Watson, C. A., Littlefair, S. P., et al. 2012, ApJS, 201, 36
Butters, O. W., West, R. G., Anderson, D. R., et al. 2010, A&A, 520, L10
Caffau, E., Ludwig, H. G., Steffen, M., Freytag, B., & Bonifacio, P. 2010, Sol. Phys., 268, 255
Cameron, A. C., Pollacco, D., Street, R. A., et al. 2006, MNRAS, 373, 799

Cameron, A. C., Davidson, V. A., Hebb, L., et al. 2009, MNRAS, 400, 451
Campello, R. J. G. B., Moulavi, D., & Sander, J. 2013, in Advances in Knowledge Discovery and Data Mining, ed. J. Pei, V. S. Tseng, L. Cao, H. Motoda, & G. Xu (Berlin, Heidelberg: Springer Berlin Heidelberg), 160
Cánovas, H., Cantero, C., Cieza, L., et al. 2019, A&A, 626, A80
Cantat-Gaudin, T., Jordi, C., Vallenari, A., et al. 2018, A&A, 618, A93
Cantat-Gaudin, T., Jordi, C., Wright, N. J., et al. 2019, A&A, 626, A17
Cantat-Gaudin, T., Anders, F., Castro-Ginard, A., et al. 2020, A&A, 640, A1
Canto-Martins, B. L., Gomes, R. L., Messias, Y. S., et al. 2020, ApJS, 250, 20
Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, ApJ, 345, 245
Cargile, P. A., James, D. J., Pepper, J., et al. 2014, ApJ, 782, 29
Carlberg, J. K., Majewski, S. R., Patterson, R. J., et al. 2011, ApJ, 732, 39
Ceillier, T., Tayar, J., Mathur, S., et al. 2017, A&A, 605, A111
Chaboyer, B., Demarque, P., & Pinsonneault, M. H. 1995, ApJ, 441, 865
Chaplin, W. J., Lund, M. N., Handberg, R., et al. 2015, PASP, 127, 1038
Charbonneau, D., Brown, T. M., Latham, D. W., & Mayor, M. 1999, ApJ, 529, L45
Charbonnel, C., Decressin, T., Amard, L., Palacios, A., & Talon, S. 2013, A&A, 554, A40
Chaushev, A., Raynard, L., Goad, M. R., et al. 2019, MNRAS, 488, 5232
Chen, X., Wang, S., Deng, L., et al. 2020, ApJS, 249, 18
Choi, P. I., Herbst, W., Choi, P. I., & Herbst, W. 1996, AJ, 111, 283
Chote, P., Gänsicke, B. T., McCormac, J., et al. 2021, MNRAS, 502, 581
Ciardi, D. R., von Braun, K., Bryden, G., et al. 2011, AJ, 141, 108
Collier Cameron, A., & Campbell, C. G. 1993, A&A, 274, 309
Cooley, J. W., Lewis, P. A. W., & Welch, P. D. 1969, IEEE Transactions on Education, 12, 27
Cooley, J. W., & Tukey, J. W. 1965, Mathematics of Computation, 19, 297
Costes, J. C., Watson, C. A., Belardi, C., et al. 2020, MNRAS, 491, 2834
Covey, K. R., Agüeros, M. A., Law, N. M., et al. 2016, ApJ, 822, 81
Cox, J. P., & Giuli, R. T. 1968, Principles of stellar structure, 2nd edn. (Cambridge Scientific Publishers)
Cummings, J. D., Deliyannis, C. P., Maderak, R. M., & Steinhauer, A. 2017, AJ, 153, 128
Cunha, M. S., Antoci, V., Holdsworth, D. L., et al. 2019, MNRAS, 487, 3523
Damiani, C., & Lanza, A. F. 2015, A&A, 574, A39
Dattilo, A., Vanderburg, A., Shallue, C. J., et al. 2019, AJ, 157, 169
Davenport, J. R. A. 2017, ApJ, 835, 16
Davenport, J. R. A., & Covey, K. R. 2018, ApJ, 868, 151
David-Uraz, A., Neiner, C., Sikora, J., et al. 2019, MNRAS, 487, 304

Debosscher, J., Sarro, L. M., Aerts, C., et al. 2007, A&A, 475, 1159

Delorme, P., Cameron, A. C., Hebb, L., et al. 2011, MNRAS, 413, 2218

Demarque, P., Woo, J.-H., Kim, Y.-C., & Yi, S. K. 2004, ApJS, 155, 667

Di Mauro, M. P. 2016, in Frontier Research in Astrophysics II (FRAPWS2016), 29

Dias, W. S., Monteiro, H., Caetano, T. C., et al. 2014, A&A, 564, A79

Dillon, C. J., Jess, D. B., Mathioudakis, M., et al. 2020, ApJ, 904, 109

Dotter, A., Chaboyer, B., Jevremović, D., et al. 2008, ApJS, 178, 89

Douglas, S. T., Agüeros, M. A., Covey, K. R., et al. 2016, ApJ, 822, 47

Douglas, S. T., Curtis, J. L., Agüeros, M. A., et al. 2019, ApJ, 879, 100

Dravins, D. 1994, The Messenger, 78, 9

Dubath, P., Rimoldini, L., Süveges, M., et al. 2011, MNRAS, 414, 2602

Dumusque, X., Borsa, F., Damasso, M., et al. 2017, A&A, 598, A133

Edelson, R. A., & Krolik, J. H. 1988, ApJ, 333, 646

Edla, D. R., & Jana, P. K. 2012, Procedia Technology, 6, 485

Efron, B., & Hastie, T. 2016, Computer Age Statistical Inference, Institute of Mathematical Statistics Monographs (Cambridge University Press)

Eggenberger, P. 2013, EPJ Web of Conferences, 43, 1005

Eggenberger, P., Maeder, A., & Meynet, G. 2005, A&A, 440, L9

Ekström, S., Georgy, C., Eggenberger, P., et al. 2012, A&A, 537, A146

ESA. 1997, European Space Agency, (Special Publication) ESA SP, Volume 1

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. 1996, in KDD'96: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (AAAI Press), 226

Evans, T. M., Aigrain, S., Gibson, N., et al. 2015, MNRAS, 451, 680

Eyer, L., & Blake, C. 2005, MNRAS, 358, 30

Eyer, L., & Mowlavi, N. 2008, in Journal of Physics Conference Series, Vol. 118, 12010

Eyer, L., Rimoldini, L., Audard, M., et al. 2019, A&A, 623, A110

Feiden, G. A. 2016, A&A, 593, A99

Feissel, M., & Mignard, F. 1998, A&A, 331, L33

Fekel, F. C., & Balachandran, S. 1993, ApJ, 403, 708

Fitzpatrick, E. 1999, PASP, 111, 63

Foreman-Mackey, D. 2018, Research Notes of the American Astronomical Society, 2, 31

Foreman-Mackey, D., Agol, E., Ambikasaran, S., & Angus, R. 2017, AJ, 154, 220

Foreman-Mackey, D., & Barentsen, G. 2019, dfm/exoplanet: exoplanet v0.1.3, doi.org/10.5281/zenodo.2536576

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306

Foreman-Mackey, D., Savel, A., Luger, R., et al. 2021, exoplanet-dev/exoplanet: v0.5.1,

doi.org/10.5281/zenodo.1998447
Fritzewski, D. J., Barnes, S. A., James, D. J., & Strassmeier, K. G. 2020, A&A, 641, A51
Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, A&A, 595, A1
Gaia Collaboration, Evans, D. W., Riello, M., et al. 2018a, A&A, 616, A4
Gaia Collaboration, Babusiaux, C., van Leeuwen, F., et al. 2018b, A&A, 616, A10
Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018c, A&A, 616, A1
Gaia Collaboration, Eyer, L., Rimoldini, L., et al. 2019, A&A, 623, A110
Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2021, A&A, 649, A1
Gallet, F., Bolmont, E., Bouvier, J., Mathis, S., & Charbonnel, C. 2018, A&A, 619, A80
Gallet, F., & Delorme, P. 2019, A&A, 626, A120
Gardner, J. P., Mather, J. C., Clampin, M., et al. 2006, Space Science Reviews, 123, 485
Gary, B. 2007, Exoplanet Observing for Amateurs (Reductionist Publications)
George, D., & Huerta, E. A. 2018, Physics Letters B, 778, 64
Gibson, N. P., Aigrain, S., Roberts, S., et al. 2012, MNRAS, 419, 2683
Gill, S., Bayliss, D., Cooke, B. F., et al. 2020, MNRAS, 491, 1548
Gillen, E., Briegal, J. T., Hodgkin, S. T., et al. 2020, MNRAS, 492, 1008
Gillon, M., Jehin, E., Magain, P., et al. 2011, EPJ Web of Conferences, 11, 06002
Goodricke, J. 1786, Philosophical Transactions of the Royal Society of London, 76, 48
Gordon, T. A., Davenport, J. R. A., Angus, R., et al. 2021, ApJ, 913, 70
Gossage, S., Conroy, C., Dotter, A., et al. 2018, ApJ, 863, 67
Graham, M. J., Drake, A. J., Djorgovski, S. G., Mahabal, A. A., & Donalek, C. 2013a, MNRAS, 434, 2629
Graham, M. J., Drake, A. J., Djorgovski, S. G., et al. 2013b, MNRAS, 434, 3423
Green, G. M., Schlafly, E., Zucker, C., Speagle, J. S., & Finkbeiner, D. 2019, ApJ, 887, 93
Grunblatt, S. K., Huber, D., Gaidos, E. J., et al. 2016, AJ, 152, 185
Gruner, D., & Barnes, S. A. 2020, A&A, 644, A16
Grunhut, J. H., Wade, G. A., Neiner, C., et al. 2017, MNRAS, 465, 2432
Günther, M. N., Queloz, D., Demory, B.-O., & Bouchy, F. 2017, MNRAS, 465, 3379
Hall, D. S. 1976, International Astronomical Union Colloquium, 29, 287
Hall, P., Fisher, N. I., & Hoffmann, B. 1994, The Annals of Statistics, 22, 2115
Harmer, S., Jeffries, R. D., Totten, E. J., & Pye, J. P. 2001, MNRAS, 324, 473
Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Nature, 585, 357
Hartig, E., Cash, J., Hinkle, K. H., et al. 2014, AJ, 148, 123
Hartman, J. D., Bakos, G. A., Kovacs, G., & Noyes, R. W. 2010, MNRAS, 408, 475
Hartman, J. D., Gaudi, B. S., Pinsonneault, M. H., et al. 2009, ApJ, 691, 342
Hartmann, L., & Stauffer, J. R. 1989, AJ, 97, 873

Hastings, W. K. 1970, Biometrika, 57, 97
Hayashi, C. 1961, PASJ, 13, 450
Haywood, R. D., Cameron, A. C., Queloz, D., et al. 2014, MNRAS, 443, 2517
Heinze, A. N., Tonry, J. L., Denneau, L., et al. 2018, AJ, 156, 241
Hennebelle, P., Fromang, S., & Mathis, S. 2013, in EAS Publications Series, Vol. 62, EAS Publications Series
Henning, T. K., Dullemond, C. P., Klessen, R. S., & Beuther, H. 2014, Protostars and Planets VI (Tucson: University of Arizona Press)
Herbst, W., Bailer-Jones, C. A. L., & Mundt, R. 2001, ApJ, 554, L197
Hertzsprung, E. 1913, Astronomische Nachrichten, 196, 201
Hobson, M. J., Brahm, R., Jordán, A., et al. 2021, AJ, 161, 235
Hodgkin, S. T., Irwin, J. M., Aigrain, S., et al. 2006, Astronomische Nachrichten, 327, 9
Hoffman, M. D., & Gelman, A. 2014, Journal of Machine Learning Research, 15, 1593
Høg, E., Fabricius, C., Makarov, V. V., et al. 2000, A&A, 355, L27
Holl, B., Audard, M., Nienartowicz, K., et al. 2018, A&A, 618, A30
Hopkins, M. M. 1916, Journal of the Royal Astronomical Society of Canada, 10, 498
Howell, S. B., Sobeck, C., Haas, M., et al. 2014, PASP, 126, 398
Huber, D., Zinn, J., Bojsen-Hansen, M., et al. 2017, ApJ, 844, 102
Hunt, E. L., & Reffert, S. 2021, A&A, 646, A104
Irwin, J., Aigrain, S., Bouvier, J., et al. 2009, MNRAS, 392, 1456
Irwin, J., Hodgkin, S., Aigrain, S., et al. 2007, MNRAS, 377, 741
Irwin, J. M., Berta-Thompson, Z. K., Charbonneau, D., et al. 2014, in 18th Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun, 767
Irwin, M. J., Lewis, J., Hodgkin, S., et al. 2004, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 5493, Optimizing Scientific Return for Astronomy through Information Technologies, ed. P. J. Quinn & A. Bridger, 411
Jackman, J. A. G., Wheatley, P. J., Pugh, C. E., et al. 2019a, MNRAS, 482, 5553
Jackman, J. A. G., Wheatley, P. J., Bayliss, D., et al. 2019b, MNRAS, 489, 5146
Jackman, J. A. G., Wheatley, P. J., Acton, J. S., et al. 2021, MNRAS, 504, 3246
James, D. J., Barnes, S. A., Meibom, S., et al. 2010, A&A, 515, A100
Jayasinghe, T., Kochanek, C. S., Stanek, K. Z., et al. 2018, MNRAS, 477, 3145
Jayasinghe, T., Stanek, K. Z., Kochanek, C. S., et al. 2019, MNRAS, 485, 961
Jayasinghe, T., Stanek, K. Z., Kochanek, C. S., et al. 2020, MNRAS, 491, 13
Jayasinghe, T., Kochanek, C. S., Stanek, K. Z., et al. 2021, MNRAS, 503, 200
Jeffers, S. V., Min, M., Waters, L. B. F. M., et al. 2012, A&A, 539, A56
Jeffries, R. D. 1997, MNRAS, 292, 177

Jeffries, R. D. 2000, in Astronomical Society of the Pacific Conference Series, Vol. 198, Stellar Clusters and Associations: Convection, Rotation, and Dynamos, ed. R. Pallavicini, G. Micela, & S. Sciortino, 245

Jeffries, R. D., & Oliveira, J. M. 2005, MNRAS, 358, 13

Jeffries, R. D., Thurston, M. R., & Pye, J. P. 1997, MNRAS, 287, 350

Jeffries, R. D., Totten, E. J., Harmer, S., & Deliyannis, C. P. 2002, MNRAS, 336, 1109

Jurcsik, J., Sódor, A., Szeidl, B., et al. 2009, MNRAS, 400, 1006

Kawaler, S. D. 1988, ApJ, 333, 236

Kharchenko, N. V., Piskunov, A. E., Schilbach, E., Röser, S., & Scholz, R.-D. 2013, A&A, 558, A53

Kilic, M., Munn, J. A., Harris, H. C., et al. 2017, ApJ, 837, 162

Kirk, J., Wheatley, P. J., Louden, T., et al. 2016, MNRAS, 463, 2922

Klioner, S. A. 2003, AJ, 125, 1580

Kohonen, T. 2001, Self-Organizing Maps, Springer Series in Information Sciences (Springer-Verlag Berlin Heidelberg)

Kolenberg, K. 2008, in Journal of Physics Conference Series, Vol. 118, 012060

Kopal, Z. 1955, Annales d'Astrophysique, 18, 379

Kounkel, M., & Covey, K. 2019, AJ, 158, 122

Kounkel, M., Covey, K., & Stassun, K. G. 2020, AJ, 160, 279

Kovács, G. 2015, A&A, 581, A2

Kovács, G., Bakos, G., & Noyes, R. W. 2005, MNRAS, 356, 557

Kovtyukh, V. V., Luck, R. E., Chekhonadskikh, F. A., & Belik, S. I. 2012, MNRAS, 426, 398

Krisciunas, K. 1993, in American Astronomical Society Meeting Abstracts, Vol. 183

Kumar, R., Carroll, C., Hartikainen, A., & Martin, O. A. 2019, The Journal of Open Source Software, doi.org/10.21105/joss.01143

Lada, C. J., & Lada, E. A. 2003, ARA&A, 41, 57

Lagarde, N., Decressin, T., Charbonnel, C., et al. 2012, A&A, 543, A108

Landolt, A. U. 2007, in Astronomical Society of the Pacific Conference Series, Vol. 364, The Future of Photometric, Spectrophotometric and Polarimetric Standardization, ed. C. Sterken, 27

Lang, D., Hogg, D. W., Mierle, K., Blanton, M., & Roweis, S. 2010, AJ, 139, 1782

Lanzafame, A. C., & Spada, F. 2015, A&A, 584, A30

Lindegren, L., Lammers, U., Hobbs, D., et al. 2012, A&A, 538, A78

Littlefair, S. P., Burningham, B., & Helling, C. 2017, MNRAS, 466, 4250

Lomb, N. R. 1976, Astrophysics and Space Science, 39, 447

Lukatskaia, F. I. 1975, in Variable Stars and Stellar Evolution; Proceedings of the Symposium,

Vol. 67, Moscow, 179
Lyngå, G. 1988, in European Southern Observatory Conference and Workshop Proceedings, Vol. 28, 379
Mamajek, E. E., & Hillenbrand, L. A. 2008, ApJ, 687, 1264
Mandal, S., Chatterjee, S., & Banerjee, D. 2017, ApJ, 835, 158
Martín, E. L., Lodieu, N., Pavlenko, Y., & Béjar, V. J. S. 2018, ApJ, 856, 40
Masci, F. J., Laher, R. R., Rusholme, B., et al. 2018, PASP, 131, 018003
Massarotti, A., Latham, D. W., Stefanik, R. P., & Fogel, J. 2008, AJ, 135, 209
Mathieu, R. D. 1994, ARA&A, 32, 465
Mayo, Jr., W. T., Shay, M. T., & Riter, S. 1974, Defense Technical Information Center, accession number: AD0784891
Mayor, M., & Queloz, D. 1995, Nature, 378, 355
Mayor, M., Pepe, F., Queloz, D., et al. 2003, The Messenger, 114, 20
McCormac, J., Gillen, E., Jackman, J. A. G., et al. 2020, MNRAS, 493, 126
McCullough, P. R., Stys, J. E., Valenti, J. A., et al. 2005, PASP, 117, 783
McQuillan, A., Aigrain, S., & Mazeh, T. 2013, MNRAS, 432, 1203
McQuillan, A., Mazeh, T., & Aigrain, S. 2014, ApJS, 211, 24
Meibom, S., Mathieu, R. D., & Stassun, K. G. 2009, ApJ, 695, 679
Meibom, S., Barnes, S. A., Latham, D. W., et al. 2011, ApJ, 733, L9
Meingast, S., Alves, J., & Rottensteiner, A. 2021, A&A, 645, A84
Melis, C., Reid, M. J., Mioduszewski, A. J., Stauffer, J. R., & Bower, G. C. 2014, Science, 345, 1029
Merrifield, M. R., & McHardy, I. M. 1994, MNRAS, 271, 899
Metcalfe, T. S., & Egeland, R. 2019, ApJ, 871, 39
Meunier, N., Desort, M., & Lagrange, A. M. 2010, A&A, 512, A39
Michell, J. 1767, Philosophical Transactions of the Royal Society of London, 57, 234
Morris, B. M., Davenport, J. R. A., Giles, H. A. C., et al. 2019, MNRAS, 484, 3244
Mortier, A., Faria, J. P., Correia, C. M., Santerne, A., & Santos, N. C. 2015, A&A, 573, A101
Morton, T. D. 2015, Astrophysics Source Code Library, ascl:1503.010
Moskalik, P., & Dziembowski, W. A. 2005, A&A, 434, 1077
Muirhead, P. S., Dressing, C. D., Mann, A. W., et al. 2018, AJ, 155, 180
National Oceanic and Atmospheric Administration. 2021, Solar Cycle Progression | NOAA / NWS Space Weather Prediction Center
Newton, E. R., Mondrik, N., Irwin, J., Winters, J. G., & Charbonneau, D. 2018, AJ, 156, 217
NExSci. 2021, NASA Exoplanet Archive
Nguyen, C. T., Costa, G., Girardi, L., et al. 2022, arXiv e-prints, arXiv:2207.08642

Noll, S., Kausch, W., Barden, M., et al. 2012, A&A, 543, A92

Oláh, K., Kővári, Z., Petrovay, K., et al. 2016, A&A, 590, A133

Osborn, H. P., Ansdell, M., Ioannou, Y., et al. 2020, A&A, 633, A53

Palate, M., Koenigsberger, G., Rauw, G., Harrington, D., & Moreno, E. 2013, A&A, 556, A49

Palla, F., & Stahler, S. W. 1999, ApJ, 525, 772

Pan, K. K. 1997, A&A, 321, 202

Pang, X., Li, Y., Yu, Z., et al. 2021, ApJ, 912, 162

Parks, J. R., White, R. J., Schaefer, G. H., Monnier, J. D., & Henry, G. W. 2011, in Astronomical Society of the Pacific Conference Series, Vol. 448, 16th Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun, ed. C. Johns-Krull, M. K. Browning, & A. A. West, 1217

Paxton, B., Bildsten, L., Dotter, A., et al. 2010, ApJS, 192, 3

Pecaut, M. J., & Mamajek, E. E. 2013, ApJS, 208, 9

Pepe, F., Cristiani, S., Rebolo, R., et al. 2021, A&A, 645, A96

Pepper, J., Kuhn, R. B., Siverd, R., James, D., & Stassun, K. 2012, PASP, 124, 230

Pepper, J., Pogge, R. W., DePoy, D. L., et al. 2007, PASP, 119, 923

Percy, J. R. 2007, Understanding variable stars (Cambridge University Press)

Perryman, M. A. C., Lindegren, L., Kovalevsky, J., et al. 1997, A&A, 500, 501

Pinsonneault, M. 1997, ARA&A, 35, 557

Pollacco, D. L., Skillen, I., Cameron, A. C., et al. 2006, PASP, 118, 1407

Pont, F., Zucker, S., & Queloz, D. 2006, MNRAS, 373, 231

Popinchalk, M., Faherty, J. K., Kiman, R., et al. 2021, ApJ, 916, 77

Price-Whelan, A. M., Sipőcz, B. M., Günther, H. M., et al. 2018, AJ, 156, 123

Pustylnik, I. 1998, Astronomical and Astrophysical Transactions, 15, 357

Queloz, D., Mayor, M., Weber, L., et al. 2000, A&A, 354, 99

Queloz, D., Henry, G. W., Sivan, J. P., et al. 2001, A&A, 379, 279

Raghavan, D., McAlister, H. A., Henry, T. J., et al. 2010, ApJS, 190, 1

Rahman, M. 2011, Applications of Fourier Transforms to Generalized Functions (WIT Press)

Rajpaul, V., Aigrain, S., Osborne, M. A., Reece, S., & Roberts, S. 2015, MNRAS, 452, 2269

Rajpaul, V., Aigrain, S., & Roberts, S. 2016, MNRAS: Letters, 456, L6

Ranjan, S., Wordsworth, R., & Sasselov, D. D. 2017, ApJ, 843, 110

Rasmussen, C. E., & Williams, C. K. I. 2006, Gaussian Processes for Machine Learning, Vol. 3 (The MIT Press)

Rauer, H., Catala, C., Aerts, C., et al. 2014, Experimental Astronomy, 38, 249

Raynard, L., Goad, M. R., Gillen, E., et al. 2018, MNRAS, 481, 4960

Rebull, L. M., Stauffer, J. R., Hillenbrand, L. A., et al. 2017, ApJ, 839, 92

Rebull, L. M., Stauffer, J. R., Bouvier, J., et al. 2016a, AJ, 152, 113

Rebull, L. M., Stauffer, J. R., Bouvier, J., et al. 2016b, AJ, 152, 114

Rehfeld, K., Marwan, N., Heitzig, J., & Kurths, J. 2011, Nonlinear Processes in Geophysics, 18, 389

Reiners, A., & Mohanty, S. 2012, ApJ, 746, 43

Reinhold, T., Bell, K. J., Kuszlewicz, J., Hekker, S., & Shapiro, A. I. 2019, A&A, 621, A21

Reinhold, T., & Gizon, L. 2015, A&A, 583, A65

Reinhold, T., & Hekker, S. 2020, A&A, 635, A43

Richards, J. W., Starr, D. L., Butler, N. R., et al. 2011, ApJ, 733, 10

Ricker, G. R., Winn, J. N., Vanderspek, R., et al. 2014, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9143, Space Telescopes and Instrumentation 2014: Optical, Infrared, and Millimeter Wave, ed. J. O. J. M., M. Clampin, G. G. Fazio, & H. A. MacEwen, 914320

Riello, M., Angeli, F. D., Evans, D. W., et al. 2021, A&A, 649, A3

Rimmer, P. B., Xu, J., Thompson, S. J., et al. 2018, Science Advances, 4, eaar3302

Robitaille, T. P., Tollerud, E. J., Greenfield, P., et al. 2013, A&A, 558, A33

Rodrigo, C., & Solano, E. 2020, in Contributions to the XIV.0 Scientific Meeting (virtual) of the Spanish Astronomical Society (Spanish Astronomical Society), 182

Roelofs, G. H. A., Groot, P. J., Nelemans, G., Marsh, T. R., & Steeghs, D. 2006, MNRAS, 371, 1231

Roeser, S., Demleitner, M., & Schilbach, E. 2010, AJ, 139, 2440

Roettenbacher, R. M., Monnier, J. D., Harmon, R. O., Barclay, T., & Still, M. 2013, ApJ, 767, 60

Rousseeuw, P. J. 1987, Journal of Computational and Applied Mathematics, 20, 53

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. 1986, Nature, 323, 533

Sahlmann, J., Lazorenko, P. F., Ségransan, D., et al. 2013, A&A, 556, A133

Saio, H. 1993, Astrophysics and Space Science, 210, 61

Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. 2016, PeerJ Computer Science, 2, e55

Samuel, A. L. 1959, IBM Journal of Research and Development, 3, 210

Sandage, A. R. 1953, AJ, 58, 61

Scaife, A. M. M. 2020, Philosophical Transactions of the Royal Society A, 378, 20190060

Scargle, J. D. 1982, ApJ, 263, 835

Scargle, J. D. 1989, ApJ, 343, 874

Schawinski, K., Zhang, C., Zhang, H., Fowler, L., & Santhanam, G. K. 2017, MNRAS: Letters, 467, L110

Schmidt, E. G. 1976, PASP, 88, 63

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

Shallue, C. J., & Vanderburg, A. 2018, AJ, 155, 94

Shannon, C. 1984, Proceedings of the IEEE, 72, 1192

Shapiro, A. I., Solanki, S. K., Krivova, N. A., Yeo, K. L., & Schmutz, W. K. 2016, A&A, 589, A46

Shappee, B. J., Prieto, J. L., Grupe, D., et al. 2014, ApJ, 788, 48

Shumway, R. H., & Stoffer, D. S. 2017, Springer Texts in Statistics, Vol. 1, Time Series Analysis and Its Applications, 4th edn. (Cham: Springer International Publishing)

Siddiqui, H., Els, S. G., Guerra, R., et al. 2014, in Observatory Operations: Strategies, Processes, and Systems V, ed. A. B. Peck, C. R. Benn, & R. L. Seaman, Vol. 9149, International Society for Optics and Photonics (SPIE), 851

Sikora, J., David-Uraz, A., Chowdhury, S., et al. 2019, MNRAS, 487, 4695

Sing, D. K., & López-Morales, M. 2009, A&A, 493

Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163

Skumanich, A. 1972, ApJ, 171, 565

Slettebak, A. 1982, ApJS, 50, 55

Smith, A. M. S., Acton, J. S., Anderson, D. R., et al. 2021a, A&A, 646, A183

Smith, G. D., Gillen, E., Queloz, D., et al. 2021b, MNRAS, 507, 5991

Smith, H. A. 2003, RR Lyrae stars, 1st edn. (Cambridge University Press)

Soderblom, D. R., Duncan, D. K., & Johnson, D. R. H. 1991, ApJ, 375, 722

Soderblom, D. R., Laskar, T., Valenti, J. A., Stauffer, J. R., & Rebull, L. M. 2009, AJ, 138, 1292

Somers, G., Cao, L., & Pinsonneault, M. H. 2020, ApJ, 891, 29

Soszyński, I., Udalski, A., Szymański, M. K., et al. 2014, Acta Astron., 64, 177

Soszynski, I., Udalski, A., Szymanski, M. K., et al. 2015, Acta Astron., 65, 297

Soszyński, I., Udalski, A., Szymański, M. K., et al. 2018, Acta Astron., 68, 89

Sousa, A. P., Alencar, S. H. P., Bouvier, J., et al. 2016, A&A, 586, A47

Southworth, J., Hinse, T. C., Jørgensen, U. G., et al. 2009, MNRAS, 396, 1023

Spada, F., Gellert, M., Arlt, R., & Deheuvels, S. 2016, A&A, 589, A23

Spada, F., & Lanzafame, A. C. 2020, A&A, 636, A76

Stahler, S. W., & Palla, F. F. 2004, The formation of stars (Wiley-VCH)

Stassun, K. G., Oelkers, R. J., Pepper, J., et al. 2018, AJ, 156, 102

Stassun, K. G., Oelkers, R. J., Paegert, M., et al. 2019, AJ, 158, 138

Stauffer, J., Rebull, L., Bouvier, J., et al. 2016, AJ, 152, 115

Stauffer, J. R., & Hartmann, L. E. E. W. 1986, PASP, 98, 1233

Stauffer, J. R., Hartmann, L. W., & Latham, D. W. 1987, ApJ, 320, L51

Stellingwerf, R. F. 1978, ApJ, 224, 953

Stoica, P., & Sandgren, N. 2006, Digital Signal Processing, 16, 712

Strassmeier, K. G. 1999, A&A, 347, 225

Strassmeier, K. G. 2009, A&A Rev., 17, 251

Strassmeier, K. G., Hall, D. S., Fekel, F. C., & Scheck, M. 1993, A&AS, 100, 173

Strobel, A. 1991, A&A, 247, 35

Tamuz, O., Mazeh, T., & Zucker, S. 2005, MNRAS, 356, 1466

Terndrup, D. M., Stauffer, J. R., Pinsonneault, M. H., et al. 2000, AJ, 119, 1303

The Theano Development Team, Al-Rfou, R., Alain, G., et al. 2016, arXiv e-prints, arXiv:1605.02688

Tilbrook, R. H., Burleigh, M. R., Costes, J. C., et al. 2021, MNRAS, 504, 6018

Tilley, M. A., Segura, A., Meadows, V., Hawley, S., & Davenport, J. 2019, Astrobiology, 19, 64

Tonry, J. L., Denneau, L., Flewelling, H., et al. 2018, ApJ, 867, 105

Torres, G., Vaz, L. P. R., Lacy, C. H. S., & Claret, A. 2014, AJ, 147, 36

van Leeuwen, F. 2009, A&A, 497, 209

van Maanen, A. 1945, ApJ, 102, 26

Vanderburg, A., Montet, B. T., Johnson, J. A., et al. 2015, ApJ, 800, 59

VanderPlas, J. T. 2018, ApJS, 236, 16

Vaughan, A. H., & Preston, G. W. 1980, PASP, 92, 385

Vines, J. I., Jenkins, J. S., Acton, J. S., et al. 2019, MNRAS, 489, 4125

Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, Nature Methods, 17, 261

Vogt, S. S., & Penrod, G. D. 1983, PASP, 95, 565

Weber, E. J., & Davis, Leverett, J. 1967, ApJ, 148, 217

West, R. G., Gillen, E., Bayliss, D., et al. 2019, MNRAS, 486, 5094

Wheatley, P. J., Pollacco, D. L., Queloz, D., et al. 2013, in European Physical Journal Web of Conferences, Vol. 47, European Physical Journal Web of Conferences, 13002

Wheatley, P. J., West, R. G., Goad, M. R., et al. 2018, MNRAS, 475, 4476

Winn, J. N. 2010, in Exoplanets, ed. S. Seager (Tucson, Arizona, USA: University of Arizona Press), 55

Yang, Y., Lian, B., Li, L., Chen, C., & Li, P. 2014, in 2014 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery, 60

Yi, S., Demarque, P., Kim, Y.-C., et al. 2001, ApJS, 136, 417

Yu, J., Huber, D., Bedding, T. R., et al. 2018, ApJS, 236, 42

Yuan, Z., Chang, J., Banerjee, P., et al. 2018, ApJ, 863, 26

Zacharias, N., Finch, C. T., Girard, T. M., et al. 2013, AJ, 145, 44

Zari, E., Hashemi, H., Brown, A. G. A., Jardine, K., & de Zeeuw, P. T. 2018, A&A, 620, A172

Zechmeister, M., & Kürster, M. 2009, A&A, 496, 577

Zhao, J.-K., Oswalt, T. D., Chen, Y.-Q., et al. 2015, Research in Astronomy and Astrophysics, 15, 1282

Zwart, S. F. P., McMillan, S. L. W., & Gieles, M. 2010, ARA&A, 48, 431