The Detection and Characterization of Terrestrial Exoplanets and Exoplanetary Satellites Via Their Transits

Tyler Gordon

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2023

Reading Committee:
Eric Agol, Chair
Daniel Foreman-Mackey
James Davenport

Program Authorized to Offer Degree:
Astronomy

University of Washington

Abstract

The Detection and Characterization of Terrestrial Exoplanets and Exoplanetary Satellites Via Their Transits

Tyler Gordon

Chair of the Supervisory Committee:
Professor Eric Agol
Astronomy

In recent decades the observation of planetary transits has emerged as the most successful method of exoplanet detection and characterization. Transits give us access to regions of parameter space that are inaccessible to RV and direct imaging – notably low-mass planets in the habitable zones of dwarf stars are best observed via their transits. Followup observations of planetary transits can, in principle, constrain molecular signatures in rocky exoplanet atmospheres via transit transmission spectroscopy, measure the densities of planets via transit timing effects, and even potentially reveal the presence of exomoon companions.

Despite the success of the transit method, challenges remain with regards to objects with shallow and/or infrequent transits. In this dissertation I detail some of those challenges and describe the progress that I have made towards mitigating them. Along the way I will describe a number of projects on subjects ranging from the characterization of stellar variability to modeling planetary interiors, all of which relate back to the basic theme of characterizing small, rocky transiting exoplanets.

In Chapter 2 I review Gaussian process methods for modeling stellar variability, and demonstrate the utility of these methods by applying a one-dimensional GP model to measure stellar rotation periods in the K2 sample. In Chapter 3 I will extend this one-dimensional GP model to two dimensions in order to capture the wavelength-dependence of starspots

and granulation. I then apply this model to simulated transit observations in order to demonstrate improved inference over the one-dimensional case. In Chapter 4 I summarize some additional applications of the Gaussian process framework including Fisher Information analysis, outlier detection and modeling, and non-stationary Gaussian processes. In Chapter 5 I review the subject of transiting exomoons and detail the development of an exomoon transit model that is intended to be combined with the stellar variability model from Chapter 2 in order to enable the detection of transiting exomoons. Finally, in chapter 6 I describe the prospects for exomoon detection in the near future with JWST.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

This work represents the culmination of seven years of graduate study in the astronomy department at the University of Washington. I would not have been capable of completing these studies without the support of countless individuals. My advisor, Eric Agol, was instrumental in guiding my work, providing feedback at every stage, and helping to secure the funding and other resources that allowed me to devote my time to these projects. I would also like to thank Daniel Foreman-Mackey for his mentorship and advising on various aspects of this work, particularly those involving Gaussian Processes. I'm grateful to James Davenport as well for advising me on the work that makes up chapter 2, and providing mentorship throughout my graduate career. Finally, I would like to thank Drew Gorman-Lewis and Sarah Stroup for serving on my dissertation committee.

The encouragement, friendship, and sometimes commiseration of my fellow graduate students was at least as essential as any advising relationship, and without the consistent support of that community I doubt that I would have been able to complete my studies. While I'm not up to the task of naming every individual who's words or actions bouyed me over the last seven years, I'd like to mention Meredith Durbin, Nicole Sanchez, and Joachim Moeyens, with whom I joined the department in 2016. It was a joy to learn alongside you all.

I'm also grateful to my parents, Shelley and Dana Gordon, for their endless support of my academic endeavours. I've also relied heavily on the support of many friends outside of the field who helped to ground me and to serve as a constant reminder that there is more to life than astronomy.

Finally, I'm grateful to the administrative staff in the Department of Astronomy for the essential work that they perform, without which none of my academic work would be possible. I also thank the custodial professionals and facilities staff that keep the Physics and Astronomy Building safe, clean, and accessible.

I would like to acknowledge that the work that makes up this dissertation was completed predominantly on the lands of the Coast Salish people, and specifically of the Duwamish Tribe. I am grateful to the land itself, as well as to the people who have stewarded this land since time immemorial. I acknowledge the history of land theft and forced removal which is ongoing today, and which is foundational to the existence of the University of Washington and the city of Seattle. Portions of this work were also completed on the lands of the Gabrielino/Tongva peoples in what is now known as Los Angeles, California, and the lands of the Shoshone and Bannock peoples in what is now known as Boise, Idaho.

DEDICATION

To my grandmother, Virginia Reynolds, who always believed in my abilities, and to my grandfather, Donald Reynolds, who instilled in our family a great love for education.

Chapter 1

INTRODUCTION

1.1 A Short History of Exoplanet Observation

The scientific study of planets beyond our solar system has only been possible for the past several decades. However interest in, and speculation about, the uniqueness of the Earth and life on it began much earlier. In order to understand the motivation for this thesis it is important to situate the work within the context of historical inquiry on these subjects. While this section is necessarily brief, it should nevertheless capture the essence of the historical development of these ideas.

1.1.1 Historical Inquiry into the Uniqueness of the Earth

Speculation about the uniqueness of the Earth and its biosphere has been ongoing since at least the classical period of Ancient Greece, with philosophers staking competing views about the possibility of life in other parts of the cosmos (Crowe & Dowd 2013). In the centuries prior to the systematization of astronomy as a science the influence of each position in this debate ebbed and flowed. With the onset of the Copernican revolution, though not by the work of Copernicus alone, the view that Earth-like worlds with life of their own may exist in abundance took hold (Danielson 2013) and remained the popular view through the beginning of the space age.

With the advent of robotic space exploration in the 1950s it became possible to investigate the habitability of nearby solar system bodies. Until this point it had seemed possible, maybe even likely, that life was so abundant in the universe that even the nearest planets might support ecosystems. The fantastical observations of canals on Mars by Perceval Lowell (Lowell 1897) and others demonstrate the ease with which an educated person of the time could accept the idea of extraterrestrial life, just as science fiction novels from the decades leading up to the space race demonstrate the popularity of the notion in the public sphere (Asimov 1978; Bradbury 2017). Early robotic landings on Venus quickly demonstrated that the planet was certainly not teaming with biology (Avduevskij et al. 1971), and both telescopic and robotic exploration of Mars by Mariner 9 and the Viking orbiter quickly

demonstrated that any biology present on the planet would likely be microscopic (Carr et al. 1976; Sagan & Lederberg 1976).

While searches for signs of biology in the Martian subsurface as well as in the oceans of Jupiter and Saturn's icy moons are ongoing, by the late 1970s all signs pointed to a likely dearth of life in the Solar System. The dawning realization that we were alone in our immediate neighborhood coincided with a burgeoning interest in planets beyond the Solar System, with studies on exoplanets beginning to appear consistently in the literature (see figure 1.1.1). By the 1980s astronomers were developing dedicated instruments and surveys to detect exoplanets (Campbell et al. 1988; Gatewood & Stein 1986; McMillan et al. 1982; Cochran & Young 1985).

1.1.2 Early Claims of Exoplanet Detection and the First Confirmed Exoplanets

Early attempts to find exoplanets focused on measuring the motion of the host star in order to detect gravitational perturbations induced by the unseen planet. The motion of the star towards or away from the observer can be measured from the Doppler shift of a star using a dedicated radial velocity spectrograph, while the motion of a star in the plane of the sky can be measured by direct observation of the position of the star over time in a process known as astrometry. This latter method was the first to be used to claim an exoplanet detection when, in 1855, W. S. Jacob presented evidence for a substellar companion of 70 Ophiucus B based on a systematic error between the observed motion of the star and the best-fitting elliptical orbit (Jacob 1855). Further claimed detections of substellar companions in the 70 Ophiucus and 61 Cygni systems were made in subsequent decades, but none have been confirmed (See 1896; Berman 1932; Reuyl & Holmberg 1943; Strand 1943).

The first claimed radial velocity detection of an exoplanet came in 1963 when Peter van de Kamp published observations of the motion of Barnard's Star which he claimed indicated the presence of a 1.6 Jupiter-mass companion (van de Kamp 1963). While the presence of a companion was ruled out by Gatewood & Eichhorn in 1973 (Gatewood & Eichhorn 1973) and attributed to a systematic anomaly in the telescope (Hershey 1973), van de Kamp continued

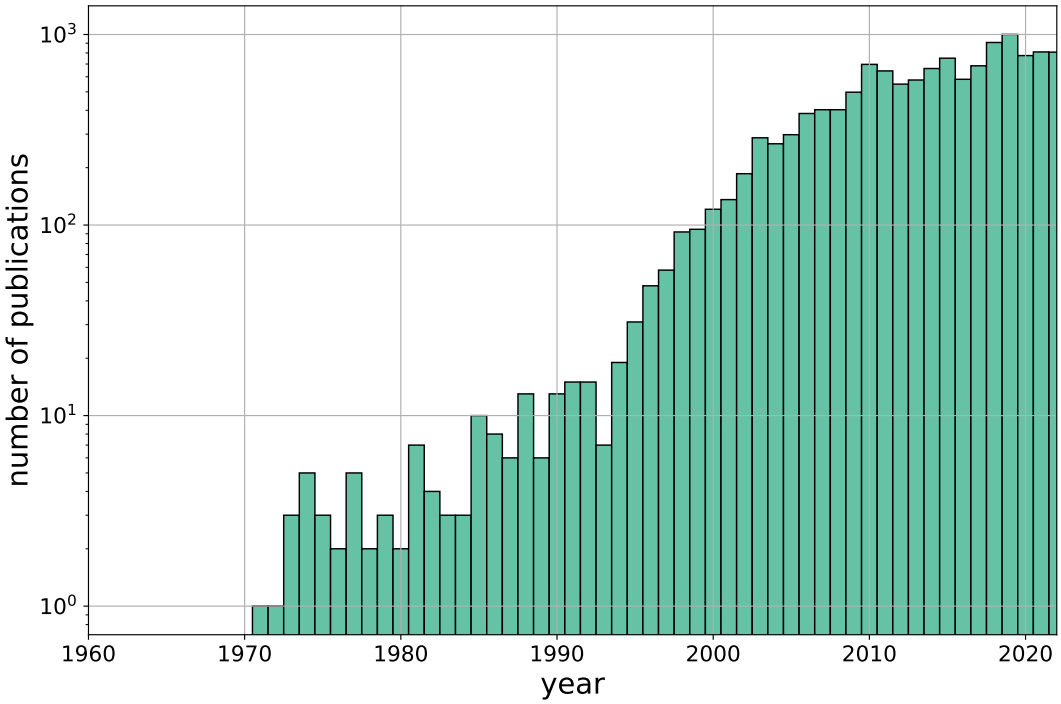

Figure 1.1 Number of publications including "exoplanet" or "extrasolar planet" in the title for the years 1960 through 2022. This data was obtained through NASA's Astrophysics Data System Bibliographic Services. It should be noted that many recent studies on specific exoplanets do not contain the word "exoplanet" but rather refer directly to the exoplanet in question. It is therefore likely that this count underestimates the number of exoplanet studies published in recent years.

to publish evidence in support of his claim throughout the remainder of his career (van de Kamp 1969a,b, 1975, 1982).

The first confirmed exoplanet detections were not made by either of these methods, nor were these planets observed around main-sequence stars. Instead they came in the form of "pulsar planets" – planetary-mass companions orbiting the remnants of massive stars (Wolszczan & Frail 1992). These planets were detected via their influence on the regular periodicity of the pulsar's emissions. These pulsar planets received little attention compared to that garnered by the discovery of the first exoplanet orbiting a main-sequence star, 51 Pegasi b, which was announced by Michel Mayor and Didier Queloz in 1995 (Mayor & Queloz 1995).

1.2 The Transit Method: A Window to Worlds Like Our Own

Despite the excitement surrounding the discovery of 51 Pegasi b, the planet is not much more habitable than the pulsar planets discovered earlier in the decade. 51 Pegasi b is a member of the class of planets we now refer to as "hot Jupiters". These are gas giants that orbit extremely close to their host stars. They have neither liquid water nor rocky surfaces, and their atmospheric temperatures can be thousands of degrees Celsius. This brings up the question of why the announcement of 51 Pegasi b was met with so much more fanfare than that of the pulsar planets. Lisa Messeri suggests an answer to this question in her book "Placing Outer Space: An Earthly Ethnography of Other Worlds"(Messeri 2016). The problem with pulsar planets is that they are so unlike the planets that we recognize as "worlds" endowed with a sense of place that they cannot be thought of as a step towards finding a second *Earth-like* world. Despite the uninhabitable nature of 51 Pegasi b, a planet around a sun-like star opens the door to an imagined universe of millions or even billions of such planets, many of which could indeed be "worlds" like our own – a notion as motivating for scientists as it is exciting for the scientifically interested public. Ever since these first discoveries the trend of exoplanetary research has been to focus increasingly on the Earth-

likeness[1] of the planets that we study, and the transit method is uniquely positioned to be our window to these worlds.

1.2.1 Outline of the Transit Method

Before exploring the history and current state of transit observations in depth, it will be helpful to briefly outline the technique. The transit method relies on the observation that, for a planet orbiting its star in a plane edge-on to the observer, an eclipse of the star by the planet will be observed once in each orbit. During this eclipse the flux from the star observed at the Earth will decrease by the amount of light blocked by the planet. The light blocked by the planet is proportional to the planet's projected area, and the light emitted by the star is proportional to the star's projected area. Therefore the flux blocked by the planet as a fraction of the star's total flux is proportional to the ratio of those areas:

$$F_{\text{blocked}} \propto \frac{A_{\text{planet}}}{A_{\text{star}}} = \left(\frac{R_{\text{planet}}}{R_{\text{star}}}\right)^2 . \tag{1.1}$$

This means that the size of the planet relative to the star can be directly observed, and if the radius of the star is known then the radius of the planet can be computed.

Information about the planet's orbital motion can also be obtained through the transit method by examining the duration and periodicity of the transit. The time between successive transits gives us the orbital period of the planet, and the duration of the transit

[1] In the remainder of this work we will use the word "Earth-like" to refer to planets which are similar to the Earth in their size and composition, namely being large enough to retain an atmosphere but small enough to not accrete a thick gaseous envelope like Jupiter and Saturn, and being composed of a combination of iron and silicates. By "Earth-like" we do not mean to imply that the planet maintains liquid surface water or possesses an atmosphere, or that it orbits at the right distance from its star to maintain a temperate climate. We will use the phrase "habitable-zone planets", or indicate that the planet is "orbiting in the habitable zone of its star" to indicate a planet that receives the right amount of instellation from its host star to maintain a temperate climate with the potential for liquid water on its surface. However it is important to note that even an "Earth-like, habitable-zone planet" does not necessarily have a temperate climate and oceans of liquid water as does the Earth. These traits depend on the present day composition of the planet and its atmosphere, which in turn depend on its formation conditions, the subsequent history of the planet, its star, and the corresponding exoplanetary system, and the availability of the requisite elements and molecules in its local environment – all details which cannot be known with any certainty given the present state of the art in exoplanet observation.

constrains the planet's velocity and the path that it takes across the stellar disk. This path is parameterized by the impact parameter, which refers to the distance of closest approach between the planet and the center of the stellar disk as illustrated in figure 1.2.1.

One complicating factor in transit modeling is that the stellar disk is not uniform in brightness. Because light emitted near the edges of the stellar disk must traverse a longer path through the star's atmosphere than light emitted near the center, the edges of the star appear dimmer than the center of the disk. This effect is known as limb-darkening, and it causes the transit curve to be shallower near the edges and deeper in the center than indicated by the planet's radius, as seen in figure 1.2.1.

Measuring Atmospheres with Transmission Spectroscopy

A straightforward extension of the transit method is to observe the lightcurve in multiple bands rather than in a single filter. While a planet with no atmosphere will block the same fractional amount of starlight at all wavelengths, a planet with an atmosphere exhibits wavelength-dependent absorption as different molecules absorb light at different wavelengths. The effect of this absorption is to make the transit depth larger, and thus increase the apparent radius of the planet, in bands where more absorption occurs Brown et al. (2001). The observation of this wavelength-dependent transit depth is known as transmission spectroscopy. This data allows for the extraction of information about the composition of the exoplanet's atmosphere through a technique known as atmospheric retrieval.

For dim stars or shallow transits it may be necessary to combine photons from across the spectrum into a single band to obtain the necessary signal-to-noise to detect and measure the transit, making transmission spectroscopy challenging for these systems. It is therefore limited to bright stars and to systems with large planet/star radius ratios such as for gas giants orbiting G and K dwarfs or rocky planets around bright M dwarfs.

A second challenge for transmission spectroscopy is the presence of clouds and hazes in exoplanet atmospheres. Because cloud and haze particles are opaque to light across the spectrum, they tend to cause transmission spectra to appear featureless even when otherwise

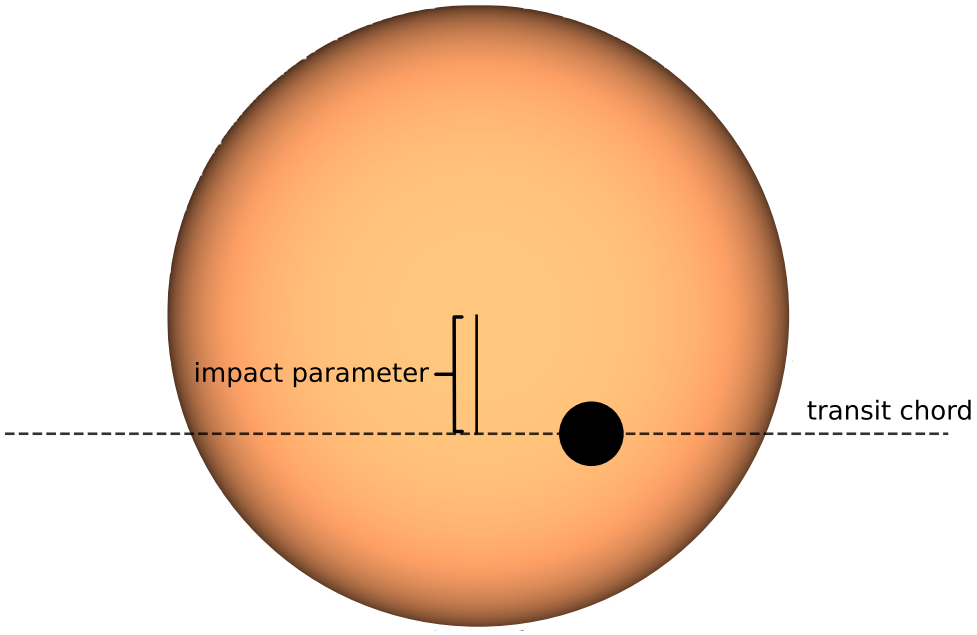

Figure 1.2 Illustration of a transit showing the transit chord, or the path along which the planet transits the star, and the impact parameter. The impact parameter is the distance, in units of the stellar radius, between the star's center and the planet at the point of its closest passage.

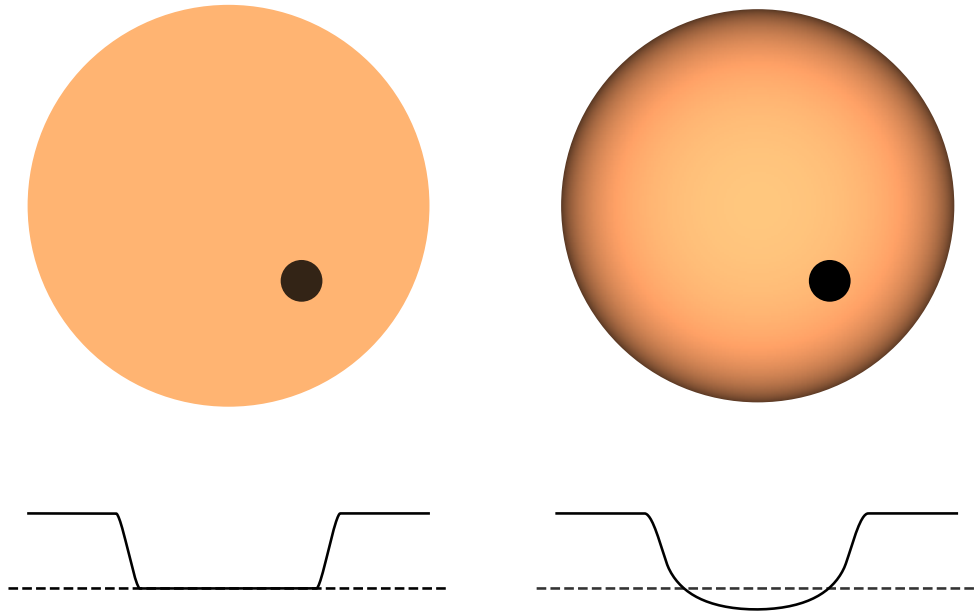

Figure 1.3 Demonstration of the effects of limb-darkening on a transit lightcurve. **Left:** A transit with no limb-darkening shows sharp transitions at the beginning and end of ingress and egress. **Right:** Limb-darkening results in a smooth transition from the out-of-transit to in-transit flux, and a transit profile that is shallower at the edges and deeper in the center. The dashed line shows the in-transit flux for the no-limb-darkening case, which is $F = 1 - (r_p/r_*)^2$ for a planet with radius r_p transiting a star with radius r_*.

detectable molecules exist in the atmosphere. At present it is not known how ubiquitous clouds and hazes may be, and determining whether a flat spectrum is due to the presence of clouds or the absence of a detectable atmosphere is not possible from the transmission spectrum alone (see Lustig-Yaeger et al. 2023).

Measuring Density with Transit Timing Variations

Exoplanet transits only occur with perfect periodicity when no other gravitationally interacting bodies are present in the system. When other exoplanets are present, perturbations to the Keplerian orbit of the transiting planet alter the timing of the transits. For pairs of planets near resonance this effect can be large enough to be measurable, in which case the deviations in the time of the transits can be used to constrain the masses of the planets. This effect is known as transit timing variations (Agol et al. 2005)

Coupled with the known radius of the planet from the transit depth, this technique allows for constraints to be placed on the planet's density; this is otherwise only possible with radial velocity followup, which is not feasible for all planetary systems.

Measuring Surface Temperature with Secondary Eclipses and Phase Curves

When the planet is sufficiently bright, either due to a high temperature or a highly-reflective surface, a secondary eclipse may be measured. This secondary eclipse occurs when the planet passes behind the star and the total flux from the system drops from

$$F_{\text{total}} = F_{\text{star}} + F_{\text{planet}} \tag{1.2}$$

to

$$F_{\text{total}} = F_{\text{star}}. \tag{1.3}$$

In the infrared, the amount of missing flux can then be converted to a brightness temperature in order to obtain the surface temperature of the planet (see, for instance Greene et al. 2023).

By observing the star between the primary and secondary eclipses it may be possible to measure a phase curve. The phase curve measures the gradual brightening of the system as the planet moves away from its transit, during which it presents only its dim night-side to the observer, towards its eclipse, during which (just prior to the eclipse itself) it presents its bright day-side to the observer. By measuring the phase curve we can determine the temperature contrast between night-side and day-side (Winn 2010). Because atmospheres tend to redistribute heat from the day-side to the night-side of a planet, the temperature contrast can help to determine whether an atmosphere is present, even when the atmosphere may be too thin or cloudy to be measured with transmission spectroscopy.

1.2.2 The History of Transit Observations

The potential for detecting an exoplanet using the effect of its eclipse on the brightness of the host star was first suggested by observations of eclipsing binary systems, in which a star passes in front of its gravitationally bound companion. This method was first suggested by Otto Struve in 1952 (Struve 1952). A program for detecting exoplanets based on the color change of their stars as the planet moved from the reddened stellar limb across the bluer center of the star was proposed by Frank Rosenblatt in 1971 (Rosenblatt 1971), but that study made over-optimistic predictions for the ease of a detection with the technology available at the time. In 1984 William Borucki (Borucki & Summers 1984) demonstrated that sufficient photometric precision had already been achieved to detect transiting giant planets from the ground, and suggested that even terrestrial planets could be detected by a space-based telescope. One of the primary challenges to detecting a transit, besides the small amplitude of the expected transit signal, is the low probability of a transit occuring for a given system. This is because the system must be aligned nearly perfectly edge-on to the Earth in order for the planet to pass directly in front of the star. In Borucki's observational program, as in Rosenblatt's, these odds would be overcome simply by observing a huge number of stars simultaneously.

Throughout the late 1980s and 1990s sporadic attempts were made to detect exoplanets

by their transits. Most notable was the campaign to monitor the eclipsing binary system CM Draconis for signs of a transiting substellar body (Schneider & Doyle 1995; Deeg et al. 1998). CM Draconis was chosen based on the principle that any planets orbiting the system would likely be aligned with the orbital plane of the binary stars, therefore nearly guaranteeing the correct geometry for a transit. The presence of two stars would mean that two transits could potentially occur, and the small size of both stars in the system would mean that the transit depth, which increases for smaller star/planet radius ratios, would be greater than for more massive stars. This attempt at selecting a system for which the minuscule odds of a detectable transit would be maximized was unsuccessful, and to date there are no known exoplanets in the CM Draconis system.

The first exoplanetary transit was finally observed in 1999 by David Charbonneau and Timothy Brown (Charbonneau et al. 2000) using the STARE instrument, which was a small, wide-field telescope designed to deliver simultaneous photometry for 40,000 stars. The target of this first detection, HD 209458, was not just a random member of that 40,000 star sample, but was intentionally chosen as the known host of a planet detected previously using the radial velocity technique (Brown & Charbonneau 2000).

In 2001, after several rounds of rejections and on the heels of the first ground-based transit detections, William Borucki's transit missiong proposal, eventually dubbed Kepler, was selected by NASA with the goal of detecting transiting exoplanets from space (Impey 2013). By it's 16th month of operations over 2,300 exoplanet candidates had been detected (Batalha et al. 2013), far eclipsing the number of planet candidates detected by the radial velocity method.

In addition to space-based missions such as Kepler, CoRoT, K2, CHEOPS, and TESS, a host of ground-based transiting exoplanet surveys have turned up hundreds of candidates and confirmed planets. These include the TRAPPIST-1 system of seven rocky planets, four of which orbit in or near the habitable zones of the host star (Gillon et al. 2017), and several of which have already been extensively observed with JWST.

1.2.3 Comparing the Transit Method to Other Exoplanet Observation Techniques

Different methods of observing exoplanets offer us access to different properties of the planet, and some are better suited to specific types of planets. Of the methods commonly in use today, the transit and radial velocity techniques are best suited for the observation of terrestrial worlds in the habitable zones of their stars (see figure 1.2.3). However, the wealth of additional observations accessible through transit observation allows for a deeper characterization of the planet and its system than is possible using radial velocities alone. As outlined above, transit observations can yield information about atmospheric composition through transmission spectroscopy, planetary density and bulk composition through transit timing variability, and surface temperature, rotation state, and atmospheric circulation through phase curves. While some of these techniques have yet to be productive for Earth-like worlds, continued advancements in space-based observatories and data analysis techniques promise to make transit observations an indispensable tool for in-depth characterization of habitable worlds.

In order to support this assertion about the utility of the transit method for characterizing terrestrial exoplanets we will now outline several of the most commonly used methods for observing exoplanets and identifying the advantages as disadvantages of each. The methods we will highlight, in addition to the transit technique, are: astrometry, radial velocities, direct imaging, and microlensing.

Astrometry is the detection of an exoplanet via the gravitational perturbation to the motion of its host star in the plane of the sky. For a star and planet orbiting a mutual center of mass, the semi-major amplitudes of the star's orbit is given by

$$a_{\text{star}} = \frac{m_{\text{planet}}}{m_{\text{star}}} a_{\text{planet}}. \tag{1.4}$$

From this expression we can see that the magnitude of the star's motion will be greater for a more massive planet and for one orbiting with a larger semi-major axis. As a result, this technique is most sensitive to massive planets at large distances from their host stars. For

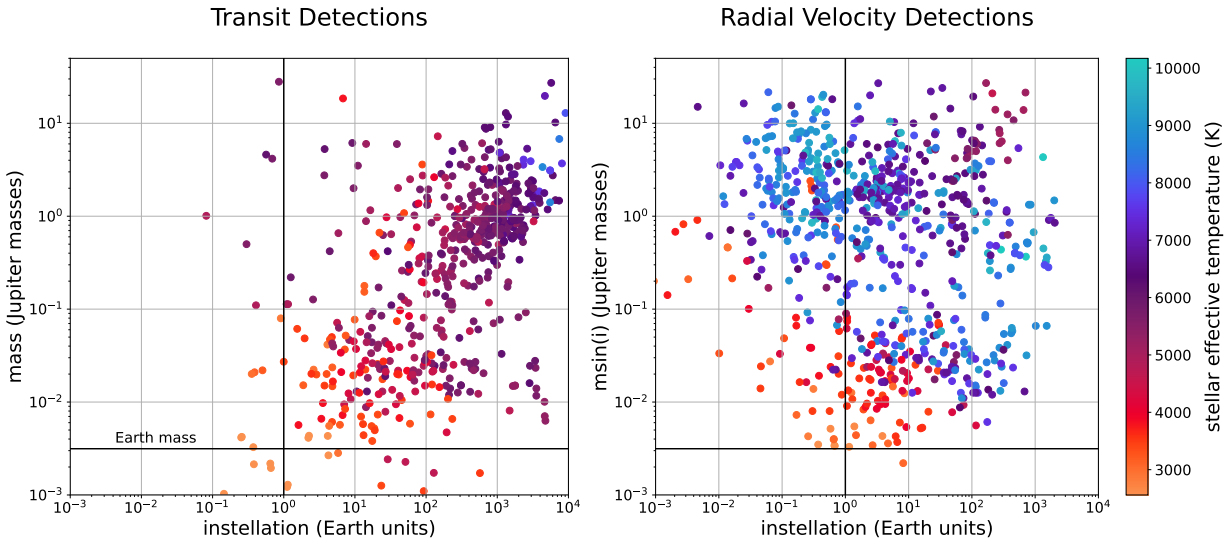

Figure 1.4 **Left:** Mass and relative instellation for exoplanets detected via the transit method. **Right:** Same as left panel but for exoplanets detected via the radial method, with the y-axis representing $msin(i)$, the minimum possible mass of the planet assuming an edge-on orbit. The colors of the points indicate the effective temperature of the host star. For reference the Sun's effective temperature is about 5780 K. Horizontal and vertical black lines represent the mass and insolation of the Earth. While the radial velocity method can, when pushed to its limit, detect Earth-mass planets around the coolest stars, the transit method is the only method that has demonstrated the ability to detect planets with masses below that of the Earth within the habitable zones of their stars..

instance, predicted exoplanet yields for the Gaia mission are expected to peak at semi-major axes of 3-4 AU around F, G, and K stars (Casertano et al. 2008; Perryman et al. 2014). Because Earth-like worlds are necessarily small, and habitable-zone planets orbit relatively close to their stars, astrometric detections are biased away from these types of worlds.

Unlike the transit technique, astrometric detections only yield information about the mass and orbit of the planet. Unless the planet also transits, there is no way to obtain information about the planets radius (and therefore density), or atmospheric temperature and composition. Furthermore, because of the bias towards large orbital separations for astrometric planets, the geometric probability of a transit is low for these systems.

Radial velocities are used to detect exoplanets by the gravitational acceleration of the star towards and away from the observer, induced by the presence of the planet. This is similar to astrometry, except that the component of motion observed in the radial velocity technique is at a right angle to the motions observed with the astrometry technique, and the velocity of the star's motion is the quantity directly observed rather than the position over time. For a planet with $m_{\rm planet}$ orbiting a star with mass $m_{\rm star}$ with a period P, semi-major axis a, and inclination i, the semiamplitude of the radial velocity variations is given by

$$K = \frac{m_{\rm planet}}{m_{\rm star} + m_{\rm planet}} \frac{2\pi a \sin i}{P\sqrt{1-e^2}}. \tag{1.5}$$

Applying Kepler's second law we can express the radial velocity semiamplitude as

$$K = \frac{m_{\rm planet} \sin i}{\sqrt{m_{\rm star} + m_{\rm planet}}} \sqrt{\frac{G}{(1-e^2)a}} \tag{1.6}$$

As a result of the linear scaling of the radial velocity amplitude with semi-major axis and planet mass, the technique is most successful for massive planets at large orbital separations, and less-massive planets at smaller orbital separations. This latter group encompasses the Earth-like, habitable-zone planet regime, making this the only technique besides transits

currently suited for both detection and followup characterization of these worlds. However, figure 1.2.3 demonstrates an important qualifier to this statement: RV detections of terrestrial, habitable-zone planets have so far only been accomplished for M dwarfs. Accomplishing this for sun-like stars will require new techniques for mitigating the effects of stellar variability on RV timeseries.

Another disadvantage of the RV method is that, like for the astrometric technique, radial velocities only tell us the minimum mass and planetary orbit, and contain no information about the density or composition of the planet.

Direct imaging is the observation of an exoplanet for which the light emitted and reflected by the planet can be resolved separately from its host star. For this method to be effective the planet must be bright (therefore large and often hot), and widely separated from its host star. The method is therefore biased to young, hot, gas giants with very wide orbits. While Earth-like worlds in the habitable zone of their stars may be accessible to this method in the future, the technology to make such observations is not yet in practical use. Another disadvantage of direct imaging is that the masses and radii of planets detected in this way cannot be directly inferred without information from other exoplanet observing techniques.

Microlensing can be used to detect exoplanets in systems that coincidentally pass between the Earth and a bright background star. During this chance alignment the gravitational fields of the foreground system magnifies the light of the background star in a manner that allows the extraction of information about the mass and positions of the planet and its host star. As it relates to Earth-like, habitable-zone planets this method has two major disadvantages. The first is that not only does it yield no information about the density or composition of the planet, but it also does a poor job of constraining the planet's orbit since only the instantaneous planet-star separation is measured. The second is that because the method relies on the coincidental alignment between a foreground and background star, followup is impossible; the information obtained about the system from a single lensing event is all of

the information that will ever be known.

This leave the transit technique as the only method that can constrain the radius and bulk composition of a habitable-zone, Earth-like exoplanet. Transit observations can also frequently be combined with radial velocity followup to obtain the mass and therefore density of the planet. In some cases the mass can even be determined without resorting to radial velocity followup by observing transit timing variations due to gravitational interactions with other planets in the system. While the transit technique does suffer from observational biases and limitations much like the other methods, which will be outlined presently and which form a large part of the motivation for this work, at present it is the most effective method of observing worlds similar to our own.

1.2.4 Challenges to Observing Transiting Exo-Earths

Despite the success of the transit method, a number of barriers remain to the full realization of the method's potential. In this section we survey some of the limitations of the method and summarize the directions in which progress is being made towards overcoming them.

Instrumental and Atmospheric Effects for Ground-based Telescopes

The observation of exoplanet transits from the ground is complicated by the effects of the passage of the starlight through Earth's atmosphere. These effects include scintillation, which is the variability in the intensity of the starlight due to turbulent motion in the upper atmosphere (Osborn et al. 2011, 2015) and variable extinction due to absorption by water vapor and other species in the atmosphere (Mann et al. 2011).

In recent years progress has been made in mitigating scintillation noise by defocusing the star on the detector, or by inserting engineered diffusers into the optical path to spread the star's light across the detector (Stefansson et al. 2017). By diffusing the star's light across the detector variations in the sensitivity of individual pixels are decreased, and individual pixels take longer to saturate, allowing for longer integration times in order to average out

the scintillation. Along with the use of narrow-band filters targeted towards windows in the water vapor absorption spectrum (Garcia-Mejia et al. 2020) and advances in telescope guidance and data reduction techniques, the precision of ground-based transit observations has begun to approach that of space-based observations from Spitzer and TESS for bright stars.

Stellar Variability

All transit observations, whether taken from the ground or from space, must contend with the presence of stellar variability. This variability occurs across a wide range of timescales from days or weeks for starspot-induced variability down to minutes or hours for variability arising from granulation patterns on the stellar surface and seconds to minutes for stellar pulsations and oscillations. Starspot variability occurs on the timescale of the stellar rotation period (and can in fact be used to measure the rotation period as discussed in Chapter 2 of this work). Because the transit duration is typically much shorter than the rotation period of the star this type of variability is easily dealt with through a number of methods, with the simple application of a moving average filter with a window larger than the expected transit duration often being sufficient.

Variations in the granulation pattern on the surface of the star are reflected in the disk-integrated stellar flux. These variations occur on timescales similar to the transit duration and can have a large impact on the measurement of transit parameters (Pereira et al. 2019; Barros et al. 2020). Many different methods have been used to model granulation variability in single-band transit observations by Kepler, K2, and TESS. In Chapter 3 of this work we discuss the use of Gaussian process (GP) variability models for single-band lightcurves and present an extension of these models meant to be applied to multiband lightcurves. While the effects of granulation variability cannot be fully mitigated, careful treatment of this effect can substantially improve measurements of important transit parameters.

Stellar oscillations occur in stars with convective envelopes, being excited by convective motions within the star (Bedding 2014). These oscillations occur on the shortest timescales

and often do not present a significant source of correlated noise on the timescale of transit duration. However they may be impactful for small planets around bright, sun-like stars (Sarkar et al. 2018). Since these are the most compelling targets for observation by upcoming transit missions as well as JWST, understanding these sources of variability and mitigating their effects is also important.

Starspots and The Transit Light Source Effect

Even for a perfectly static stellar disk, inhomogeneities across the disk can confound the interpretation of transit signals. For instance when a planet transits a starspot the resulting lightcurve will deviate from the expected shape of an ideal transit. If not accounted for this effect can confound the interpretation of the transit signal and result in incorrect measurements of the transit time and depth. Even unnoculted starspots present an issue. A star with a signficant starspot covering fraction will emit less flux than a star of the same mass and radius without starspots. As a result a planet transiting the star without crossing a spot will have a larger transit depth than it would if no spots were present. If this effect is not recognized or accounted for the transit signal will be interpreted as that of a planet with a larger radius than the true radius of the planet.

In the case that the planet transits a spot, the transit model can be modified to account for the spot in order to better measure the transit parameters. However, in the case that the starspots do not lay in the path of the transiting planet, there is no way to account for their effects without independent knowledge of the spot covering fraction of the star. This points to the importance of understanding starspot covering fractions and temperatures as a function of stellar type. For systems of special interest, such as TRAPPIST-1, understanding the influence of starspots on transit observations will be of crucial importance.

Beyond their effect on white-light transits, starspots can also have a deleterious effect on the measurement of transmission spectra. Transmission spectroscopy is the technique of measuring the absorption spectrum of an exoplanet's atmosphere by the apparent radius of the planet as a function of wavelength. At wavelengths where the atmosphere is relatively

opaque due to absorption by abundant chemical constituents, the planet will exhibit a deeper transit compared to wavelengths where the atmosphere is relatively transparent. In order to correctly interpret this wavelength-dependent transit depth however, the spectrum emitted by the portion of the star being eclipsed must be taken into account. For a star without surface inhomogeneities the spectrum of the transit chord (the portion of the star eclipsed by the planet over the course of the transit) is identical to the disk-integrated spectrum of the star when the planet is out of transit. However, when the disk is inhomegeneous, the difference in spectrum between the occulted region of the star and the unocculted region will imprint a deviation from the atmosphere's absorption spectrum onto the transmission spectrum. This is known as the "transit light source effect" (Rackham et al. 2018). This effect is most problematic for late spectral types (e.g. m dwarfs) observed at short wavelengths, and could potentially be a limiting factor in the detection of biosignatures around m dwarf planets such as those in the TRAPPIST-1 system (Rackham et al. 2019).

1.2.5 State of the Art: TESS, JWST, and near-future Transit Missions

As of this writing there are two space-based telescopes in operation dedicated to observing transiting exoplanets: TESS (Ricker et al. 2015) and CHEOPS (Broeg et al. 2013). In addition two transit missions are currently in development by the European space agency. These are PLATO (Rauer et al. 2014), a mission designed to both detect and characterize transiting exoplanets, and ARIEL, a mission designed to characterize exoplanets via transits, secondary eclipses, and phase curves (Tinetti et al. 2018).

In addition to these dedicated missions, the large space-based observatory JWST is well-equipped to observe transiting exoplanets and has already been used to characterize the atmospheres of a number of planets at much higher levels of detail than was previously possible (Ahrer et al. 2023; Lustig-Yaeger et al. 2023; Radica et al. 2023). Development is already underway for large space-based telescopes to succeed JWST. In 2021 the National Academy of Sciences released the long-awaited decadal survey which recommended the development of a large IR/O/UV mission to be launched sometime in the mid 2040s. It is

likely that the design of this mission will be in part informed by the goals of the exoplanet community, especially with regards to the characterization of, and search for biosignatures on, Earth-like, habitable-zone worlds (National Academies of Sciences & Medicine 2021).

1.3 New Frontiers in Exoplanetary Transits

In this section we summarize several emerging areas of research in transiting exoplanet science and relate them to the work presented in this thesis.

1.3.1 Characterizing Complete Planetary Systems

In the early days of transiting exoplanet science planets were, due to a dearth of information, treated in isolation. As observational techniques have improved and decades of monitoring have revealed thousands of transiting exoplanets, many in multi-planet systems, it has become possible to understand many of these worlds in context.

The word "context" here refers broadly to the characteristics of the system in which the exoplanet orbits. This includes additional planets in the system; key system parameters like its age; the metallicity, rotation state, and activity levels of the host star; and the presence of exocomets, exomoons, and exorings in the vicinity of the planet. This context is important because it gives us clues to the planets formation history, any subsequent migration it might have undergone, the dynamical forces that it is and has been subject to, and its irradiation history.

As the exoplanet community seeks to identify habitable worlds, having access to this context will become even more essential. Instantaneous habitability at this moment in time is not sufficient to establish the potential for life to exist on a planet; the planet must remain habitable over evolutionary timescales (Kasting et al. 1993), as well as never having been subject to conditions which would result in desiccation or permanent atmospheric loss (Luger & Barnes 2015). Obtaining that knowledge about the history of a world is only possible if the planetary system it resides in is also well-understood.

Another reason to pursue the complete characterization of the planetary system is that

exoplanet observations often admit multiple interpretations which can only be discriminated between by appealing to our understanding of the planet's history and environment (Rogers & Seager 2010; Dorn et al. 2015). For instance, even with constraints on the density of a planet the composition cannot be determined with certainty due to degeneracies between different structures that yield the same bulk density. A planet with a bulk density similar to the Earth may have a dense, iron-rich core and a large envelope of low-density volatile elements, or it may have no atmosphere at all and be composed of silicates with an intermediate density. These degeneracies can be partially broken by incorporating information about the abundance of heavy elements in the star (Dorn et al. 2015, 2017; Adibekyan et al. 2021), with the expectation that the stellar abundances should relate to the abundances in the planet. Certain possible planetary structures may also be eliminated by arguments pertaining to the likelihood of a planet maintaining a large volatile envelope given its age, radiation environment, and migration history (for example: Piaulet et al. 2023).

1.3.2 Exomoons

Exomoons are the constituents of the exoplanetary system that exist in closest physical proximity to the planet itself. As such they are uniquely situated to act as probes of a planet's physical properties and dynamical history.

The moons of the Solar System formed in a variety of ways. The larger moons of Jupiter and Saturn are thought to have co-accreted from the same disk as the giant planets themselves, while some of their smaller moons show evidence of having formed elsewhere before being captured into their current orbits. Earth's moon is a particularly unique case in our Solar System as it appears to have formed from much the same material as the Earth based on isotopic data (Wiechert et al. 2001; Young et al. 2016) but with a depletion of certain volatile elements which suggest an energetic early history (Canup et al. 2021). This has led to the development of a variety of impact scenarios for the Moon's formation.

Given the abundance and variety of Solar System moons, we expect that moons should also be found in exoplanetary systems. For now, however, they elude our detection. Perhaps

this should come as no surprise since the moons in our own Solar System are universally too small to be detectable either by their transits or by their gravitational influence on the planets they orbit.

The driving motivation for continued attempts at exomoon detection is that exomoons have the potential to unlock new avenues of investigation into the formation and evolution of planetary systems and to add a new dimension to our understanding of planetary habitability. Exomoons are expected to be easily stripped from their planet's orbits in the presence of planet-planet scattering events and during certain planetary migration events (Kane 2017; Ramírez & Sucerquia 2018; Trani et al. 2020). Their presence in a system would therefore constrain possible dynamical histories. Much like exoplanets, their size and composition also contains information about where in the system and under what conditions they formed.

Much work has been done on the topic of the moon's impact on Earth's habitability (Laskar et al. 1993; Ward & Brownlee 2000; Lissauer et al. 2012). In recent years this work has been extended to exoplanets in order to understand the importance of exomoons for establishing and maintaining the habitability of their host planets (Piro 2018). The habitability of exomoons as worlds in and of themselves has also been considered. A particularly exciting notion is that tidal heating in exomoon interiors may provide a pathway for exomoons to remain warm enough to harbor liquid water far beyond the planetary habitable zone (Tjoa et al. 2020; Scharf 2006; Forgan & Dobos 2016).

In Chapters 5 and 6 we discuss the state of current attempts at exomoon detection, outline the prospects for exomoon detection with JWST, and describe the development of a transit model for exomoon/exoplanet transits and its application to the problem of exomoon detection.

1.3.3 Constraining Planetary Composition

When information about the mass of a planet is available in addition to the radius measurements inferred from the planet's transit it becomes possible to investigate the structure and composition of the planet. Mass measurements can either be made by radial velocity fol-

lowup of the system, or by the analysis of transit timing variations in systems with multiple planets orbiting in near resonances.

Placing constraints on the structure and composition of exoplanets is enormously important for understanding how and where in their systems they formed, the history of the delivery and removal of volatiles to their surfaces, and their present-day habitability. The processes by which exoplanets, including habitable-zone Earth-like exoplanets, obtain, retain, or lose volatile compounds throughout their evolution is an open question in planetary science in part because of the difficulty in constraining their volatile contents to begin with. Along with the importance of volatiles for habitability this is a primary motivator for developing methods of constraining planetary composition from transit and/or radial velocity observations.

Current techniques for constraining planetary compositions suffer from fundamental degeneracies in the allowed compositions for a given planetary density. Even so, it is sometimes possible to place upper or lower limits on the volatile envelope or core mass fraction of a planet (Agol et al. 2021; Piaulet et al. 2023; Suissa & Kipping 2018). These constraints may be improved by the inclusion of additional information about stellar elemental abundances in the form of Bayesian priors (Dorn et al. 2015).

1.4 Structure of Work

In this introduction we have outlined the current state of the field of exoplanetary transit science and identified areas in which the field is poised to make progress in the next decade. The bulk of this thesis focuses on two of these areas: The impact of stellar variability on transiting exoplanet science and exomoon transit modeling and detection.

In Chapter 2 we review the concept of a Gaussian process variability model and demonstrate the application of such a model to a large sample of light curves in order to infer their rotation periods. The implications of this work for our understanding of stellar evolution on the main sequence are investigated and directions of future research in this area are discussed.

In Chapter 3 we describe an extension to the model used in Chapter 2 which enables modeling of multiwavelength lightcurves. We show that this model has the potential to substantially improve transit measurements when correlated noise is present on the transit timescales.

A persistent problem for Gaussian process models is their poor performance in the presence of outliers. In Chapter 4 we describe a procedure for detecting outliers and conducting inference with Gaussian process models in their presence.

In Chapter 5 we shift directions to focus on the problem of exomoon detection. While the stellar variability models described in the first three chapters will be essential to measuring shallow exomoon transits, another necessary ingredient is a transit model that handles cases of overlap between the planet and moon. Typical transit models are not equipped to model these overlapping transits, and of the few models that do, most are not compatible with advanced inference methods such as Hamiltonian Monte Carlo, which require access to the gradients of the lightcurve. We develop a model for the mutual transit of an exoplanet and an exomoon companion that is exact and differentiable. We demonstrate that our model is general enough to be applied to the mutual transit of two planets in a multiplanetary system and use the model to infer the parameters of a putative two-planet transit in the Kepler-51 system.

Finally, in Chapter 6, we pull together the threads of the first five chapters to examine the prospects for exoplanet detection with JWST. We include the effects of correlated stellar variability in the form of stellar flicker, and assume that we can accurately model this variability using the Gaussian process framework described in Chapter 3. We find that moderately-sized exomoons are likely detectable around the brightest exoplanet host stars with JWST and we suggest an observing strategy for making a detection. We also discuss the selection of targets for a JWST exomoon search.

Chapter 2

GAUSSIAN PROCESS MODELS OF STELLAR VARIABILITY

Portions of this chapter were originally published in collaboration with James Davenport, Ruth Angus, Daniel Foreman-Mackey, Eric Agol, Kevin Covey, Marcel Agüeros, and David Kipping in the May 2021 issue of The Astrophysical Journal (Tyler A. Gordon et al 2021 ApJ 913 70) and are reproduced here with permission of the American Astronomical Society.

2.1 Summary

We introduce the concept of a Gaussian process and describe its use as a model for stochastic stellar variability. A GP model specifically designed to model rotationally-modulated starspot variability is developed and applied to the lightcurves of 284,834 unique K2 targets. The primary motivation for this chapter is to establish the existence of the basic features of the rotation period-color diagram for Kepler stars in the K2 sample. In particular we recover the gap in the rotation period-color diagram first reported by McQuillan et al. (2013). While the gap was tentatively detected in Reinhold & Hekker (2020), this chapter represents the first robust detection of the gap in K2 data for field stars. This is significant because K2 observed along many lines of sight at wide angular separation, in contrast to Kepler's single line of sight. Together with recent results for rotation in open clusters, we interpret this gap as evidence for a departure from the $t^{-1/2}$ Skumanich spin down law, rather than an indication of a bimodal star formation history. We provide maximum likelihood estimates and uncertainties for all parameters of the quasi-periodic light curve model for each of the 284,834 stars in our sample.

2.2 Introduction

2.2.1 Gaussian Processes as a Model for Stellar Variability

While more general definitions of Gaussian processes may be formulated, for our purposes a Gaussian processes is an ordered collection of random variables along one or more axes. In the case of a stellar lightcurve the random variables model a series of observations of the star's flux taken at discrete times. The Gaussian aspect of a Gaussian process describes the

relationship between random variables; we model N' observations with an N'-dimensional Gaussian distribution. The covariance of the multi-dimensional Gaussian is described by a kernel function, which gives the covariance between any pair of observations as a function of their separation in time or space. The kernel function then defines the covariance matrix K. For a kernel $k(x_i, x_j)$, we have

$$K_{i,j} = k(x_i, x_j) + \delta_{i,j}\sigma_i^2, \tag{2.1}$$

where $\delta_{i,j}$ is the Kronecker delta function and σ_i is the white noise component for the i^{th} datapoint. In addition to the kernel function, a GP is characterized by its mean function, $\mu(t)$, which describes the deterministic component of the process. In the absence of any deterministic phenomenon like transits or flares, the mean function is generally taken to be constant. The GP likelihood function, $\mathcal{L}$, describes the likelihood that a set of observations, $\boldsymbol{y}$ taken at times $\boldsymbol{t} = \{t_i\}$, is drawn from the GP. It is written as

$$\ln\ \mathcal{L} = -\frac{1}{2}(\boldsymbol{y} - \boldsymbol{\mu})^{\mathrm{T}} K^{-1}(\boldsymbol{y} - \boldsymbol{\mu}) - \frac{1}{2}\ln\ \det(K) - \frac{N'}{2}\ln(2\pi) \tag{2.2}$$

where $\boldsymbol{\mu}$ is a mean vector with entries given by $\mu_i = \mu(t_i)$.

Another more intuitive way to think of a Gaussian process is as a statistical distribution from which random functions with a given covariance structure may be drawn. A useful analogy here is to a Gaussian *distribution*. Just as a random individual point may be drawn from a distribution, a collection of points representing a set of observations may be drawn from a process. Given an individual observation, the likelihood that the observation was drawn from a Gaussian distribution with a specific mean, μ and standard deviation, σ can be computed using the Gaussian likelihood function

$$\ln \mathcal{L} = -\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 - \frac{1}{2}\ln\left(2\pi\sigma^2\right). \tag{2.3}$$

This is the one-dimensional analog to the likelihood function for the GP given above.

Gaussian processes have been used extensively to model variability in light curves (see Dawson et al. 2014, Barclay et al. 2015, and Chakrabarty & Sengupta 2019 for examples from studies of transiting exoplanets and MacLeod et al. (2010) for an example in which a GP is used to model AGN variability).

The kernel function of the GP can vary widely and choosing the optimal kernel for a given application is often non-trivial. The parameters of the kernel function, which control the form of the covariance, are referred to as "hyperparameters" in the Gaussian process literature in order to distinguish them from the direct "parameters" of the GP and its likelihood function which are the covariance matrix and mean.

Modeling observations with a GP generally takes the form of Gaussian process regression, which is the process of finding optimal parameters for the kernel and mean functions given the data. This can be accomplished by maximizing the likelihood function. For a more detailed primer on Gaussian processes in astronomy, see Foreman-Mackey et al. (2017), or, for a more complete resource on GPs across fields we refer the reader to Rasmussen & Williams (2006).

In this chapter we use GP regression to find the optimal hyperparameters of a kernel function chosen to describe rotational variability in a stellar light curve given a set of K2 observations. We then sample the likelihood using an MCMC algorithm to obtain uncertainties for the hyperparameters. Our choice of kernel function and the details of the procedure used to optimize the hyperparameters and sample the likelihood function is described in section 2.3.2.

2.2.2 The Rotation of Main-Sequence Stars as Seen by Kepler

Stellar rotation is a key physical property for understanding individual stars as well as stellar populations. Rotation drives the stellar dynamo which produces surface magnetic fields. These magnetic fields in turn give rise to stellar activity (e.g. starspots and flares). The rotation period ($P_{\rm rot}$) of a star is tied to its age through magnetic braking, which slows the star's rotation over time (Durney 1972; Skumanich 1972). Age is a fundamental stellar parameter, but is difficult to determine from the position of a star on a color-magnitude

diagram, especially for stars on the main-sequence. Any information we can extract about the age of a star from its rotation period is therefore very valuable. This is the subject of gyrochronology, which seeks to measure stellar ages by observing the star's rate of rotation (Barnes 2003).

The Kepler mission (Borucki et al. 2010) revolutionized the study of stellar rotation by producing high-precision light curves for hundreds of thousands of stars, from which over 34,000 rotation periods have been inferred (Nielsen et al. 2013; McQuillan et al. 2014). The distribution of rotation periods measured by McQuillan et al. (2013) showed an unexpected bimodality in the field M dwarfs, which was found to extend to the K dwarfs by McQuillan et al. (2014). This bimodality was recovered in the G dwarfs by Davenport & Covey (2018), who used Gaia astrometry to limit their analysis to main-sequence stars with well-determined Gaia photometric solutions, removing contamination by subgiants.

Several explanations have been put forward to explain this bimodal period distribution. Davenport & Covey (2018) and McQuillan et al. (2013, 2014) suggest that the bimodality may be the result of a bimodal star formation history, with a recent burst of star formation accounting for the fast-rotating branch of the bimodality and an older population of stars forming the slow-rotating branch. Reinhold et al. (2019) propose that the gap between modalities may represent a minimum in detectability of rotation periods due to the transition from spot-dominated to faculae-dominated stellar activity.

A third possibility, to be discussed in more detail in the Section 3.6, is that the gap results directly from the spin evolution of G, K and M dwarfs. An epoch of stalled spindown followed by a period of rapid angular momentum loss before the resumption of Skumanich spindown may be able to explain such a feature. To date, rotation periods from open clusters have provided the most compelling evidence that modified spin-evolution is indeed the cause of the gap. These fixed-age populations have shown that rotation periods for low-mass stars break from the expected Skumanich spindown model, such as the cluster of rotation periods at $P_{\rm rot} \approx 10$ days found in the 1 Gyr old cluster NGC 6811 by Meibom et al. (2011). Similar deviations from the traditional Skumanich spindown profile have been seen in e.g. Praesepe

at 650 Myr (Douglas et al. 2017)and NGC 752 at 1.3 Gyr (Agüeros et al. 2018). Curtis et al. (2019) note in their analysis of NGC 6811 that the stall in spin-down appears to be mass and age dependent. Further, Curtis et al. (2020) show compelling evidence that this deviation from Skumanich spin-down indeed corresponds to the gap in field rotation periods, with individual cluster sequences "crossing" the rotation period gap.

This scenario may be explained in terms of time-variable and mass-dependent rotational coupling between the core and envelope of the star (Spada & Lanzafame 2020). Magnetic braking slows the rotation of the convective envelopes of stars. However, if the core and envelope are only weakly coupled, the stellar core may continue to spin rapidly even as the envelope slows down. A decoupled core and envelope with reduced angular momentum exchange is expected for young stars (Endal & Sofia 1981; MacGregor & Brenner 1991; Bouvier 2008; Denissenkov et al. 2010; Gallet & Bouvier 2013; Lanzafame & Spada 2015; Somers & Pinsonneault 2016). After a period of time, which appears to depend on stellar mass, the core and the envelope begin exchanging angular momentum. When angular momentum transport is happening efficiently, the core's angular momentum transferred to the envelope would offset magnetic braking, allowing the envelope to maintain a constant rotation period. After the core and envelope have coupled and the star rotates as a solid body, the star would resume spinning down. This process would result in a departure from the Skumanich spin-down law (Skumanich 1972), which prescribes a smooth spindown over time following the relation

$$P_{\rm rot} \propto t^{-1/2}, \tag{2.4}$$

where $P_{\rm rot}$ is the stellar rotation period as a function of the age of a star, t. This scenario may explain the convergence of cluster sequences below the gap. Explaining the under-density within the gap requires us to posit an additional stage in stellar spin evolution consisting of a period of accelerated spindown immediately after the epoch of stalled spindown and before the resumption of Skumanich spindown (Curtis et al. 2020). This accelerated spindown is not predicted by the coupling scenario presented here, and its explanation will likely require

further theoretical work.

The Kepler data alone gives us a limited ability to explore these various hypotheses due to its single pointing, which admits the possibility that the bimodality is unique to the Kepler field. In contrast, K2 observed the sky in 18 separate campaigns, each having a different line of sight (save for a few overlapping campaigns). As van Saders et al. (2019) note, if the Kepler line of sight happened to point directly through a late burst of star formation, thereby accounting for the fast rotating branch of the bimodal period distribution, we would not expect this feature to be visible in all 18 K2 campaigns.

In this work we measure and report probabilistic constraints on periodic signals for 284,834 K2 stars from all 18 campaigns and analyze a subset of 8,943 highly accurate rotation periods. For those stars appearing in multiple campaigns we run our analysis separately for each light curve. We use a modification of the Gaussian Process regression method described in Angus et al. (2018) to measure periodic signals. We find that the bimodality is visible in all K2 campaigns, lending support to the idea that the feature is related to stellar physics rather than being a product of the star formation history within the Kepler field.

2.3 Measuring Rotation Periods

We begin by describing the model that we use to infer probabilistic rotation periods from the EVEREST light curves (Luger et al. 2018). Stellar magnetic activity induces starspots and faculae on the star's surface. As the star rotates, these features are carried into and out of view, introducing periodicity into the light curve. If the starspots and faculae were static over time, we would observe a perfect periodicity with the star returning to the same luminosity once every period. However, starspots and faculae are not static, but rather evolve over time, emerging, changing shape, and disappearing as the star's rotation brings them into and out of view. As a result, the light curves do not display perfectly periodic variations, but rather a quasi-periodic variability with the shape and amplitude of the variability changing from period to period. This means that inferring rotation periods using straightforward sinusoidal variability models does not give good results. Instead, non-inference based methods such as

autocorrelation functions (McQuillan et al. 2013), or Lomb-Scargle periodograms (Reinhold et al. 2013) can be employed. An alternative to these non-inference methods is to use a stochastic variability model such as a Gaussian process (Angus et al. 2018). In this work we use the autocorrelation function to derive a multi-modal "prior" over the period. We then use a Markov chain Monte Carlo (MCMC) method to estimate prosteriors for the parameters of the quasi-periodic Gaussian Process (GP) variability model defined in section 2.3.2. The use of quotation marks around the word "prior" references the fact that this is not technically a Bayesian prior, because it does not strictly depend on our prior beliefs about the period distribution. We explain this further and describe the effect that this has on our analysis in the next section.

2.3.1 Autocorrelation Function Analysis

We use long cadence K2 EVEREST light curves from Luger et al. (2018) as the starting point for our analysis. We use the co-trending basis vector corrected flux (keyword FCOR in the EVEREST FITS files) which removes systematic trends from the raw light curves. To further remove long-term trends we subtract a third order polynomial from each light curve before computing the autocorrelation function (ACF). This has the effect of flattening the overall decay of the ACF at short time lags, improving our ability to detect the rotation period from the ACF peak. We note that the third order polynomial may overfit and remove some rotation signals for the slowest rotators in our sample. This primarily affects rotation periods longer than about 25 days, which does not interfere with our detection or analysis of the period gap in section 2.4. In our initial experiments we found that a second order polynomial did not sufficiently flatten the decay of the ACF and higher order polynomials were too likely to overfit the rotation signal. We remove outliers from our light curves by masking all flux observations greater than 3-sigma from a running median with a kernel width of 5 K2 long cadences.

After pre-processing the light curve as described above, we compute the ACF for each light curve using the implementation provided in `exoplanet`, which wraps the `astropy` ACF

function (Astropy Collaboration et al. 2018; Foreman-Mackey et al. 2019). We smooth the ACF with a Gaussian filter with a kernel width of 0.5 days. We then use the smoothed ACF to construct a unique multi-modal period "prior", which we find aids in convergence during the MCMC step. We place "prior" in quotes because, strictly speaking, a prior should only reflect our prior beliefs about the period distribution rather than depending on the data itself. The distribution we derive here is not technically a prior although we use it as such. Since we are building our "prior" from the same data that we use to fit the GP, we risk underestimating the uncertainty on the period. Since our analysis focuses on point estimates of the period rather than the full posterior, we elected to accept this risk in exchange for the benefit of recovering more rotation periods.

Our prior is a Gaussian mixture with $3N$ components, where N is given by

$$N = \begin{cases} N_{\text{peaks}}(\text{p} > 0.01) & N_{\text{peaks}}(\text{p} > 0.01) < 10 \\ 10 & N_{\text{peaks}}(\text{p} > 0.01) \geq 10 \end{cases} \tag{2.5}$$

where $N_{\text{peaks}}(p)$ is an integer corresponding to the number of peaks in the ACF with topographical prominence greater than p. The topographical prominence is computed with respect to the adjacent ACF minima, by `scipy`'s `signal` library (Virtanen et al. 2020). We take each of these peaks to represent a candidate period, recognizing that for a well-defined periodic signal there will be multiple peaks corresponding to the same period.

The factor of 3 in $3N$ arises from our inclusion of candidate periods at $\tau_i/2$ and $2\tau_i$ where τ_i is the lag of the i^{th} peak, so that for each peak we have 3 candidate periods. The weight of the component of the Gaussian mixture prior corresponding each peak is given by

$$w_i = h_i \sqrt{p_i} \tag{2.6}$$

where h_i is the height of the peak, and is the same for the candidate periods at $\tau_i/2$ and $2\tau_i$ as for the candidate period at τ_i itself. The standard deviation of each Gaussian component,

σ_i, is given by the width of the peak at half of the peak height. This means that the width of the Gaussian component of the prior is wider by a factor of approximately 2.35 than the standard deviation of the Gaussian equivalent to the ACF peak. The standard deviation associated with the candidate period at τ_i is also used for the candidate periods at $2\tau_i$ and $\tau_i/2$. In the case that no peaks are detected in the ACF, we adopt a uniform prior over the range $P = (0, \Delta T/2)$ where ΔT is the total duration of the light curve.

The choices we made in computing w_i and σ_i are motivated by the logic that a higher peak should be given more weight in the mixture than a lower peak, and that the component in the mixture should have a smaller standard deviation if the peak in the ACF is sharper, reflecting the smaller uncertainty on the corresponding period. We also aim to construct a prior that is informative, but that doesn't prohibit the MCMC from exploring periods not identified as period candidates by our algorithm. As can be seen in Figures 2.1, the period priors we construct tend to have wide regions of high probability and therefore limited influence over our point estimates of the rotation period.

2.3.2 Gaussian Processes and MCMC Analysis

To construct a GP stellar rotation model we take the mean function of the GP to be constant and allow the kernel function to model the correlated variability introduced into the star's light curve by spots and faculae as they rotate in and out of view and evolve over time. Our GP model has three terms: two quasi-periodic terms to capture the rotationally induced variability and one that is aperiodic to capture any leftover variability originating from other astrophysical sources or instrumental effects. The power spectrum of each term is given by

$$S_i(\omega) = \sqrt{\frac{2}{\pi}} \frac{S_i \omega_i^4}{(\omega^2 - \omega_i^2)^2 + 2\omega^2 \omega_i^2}. \tag{2.7}$$

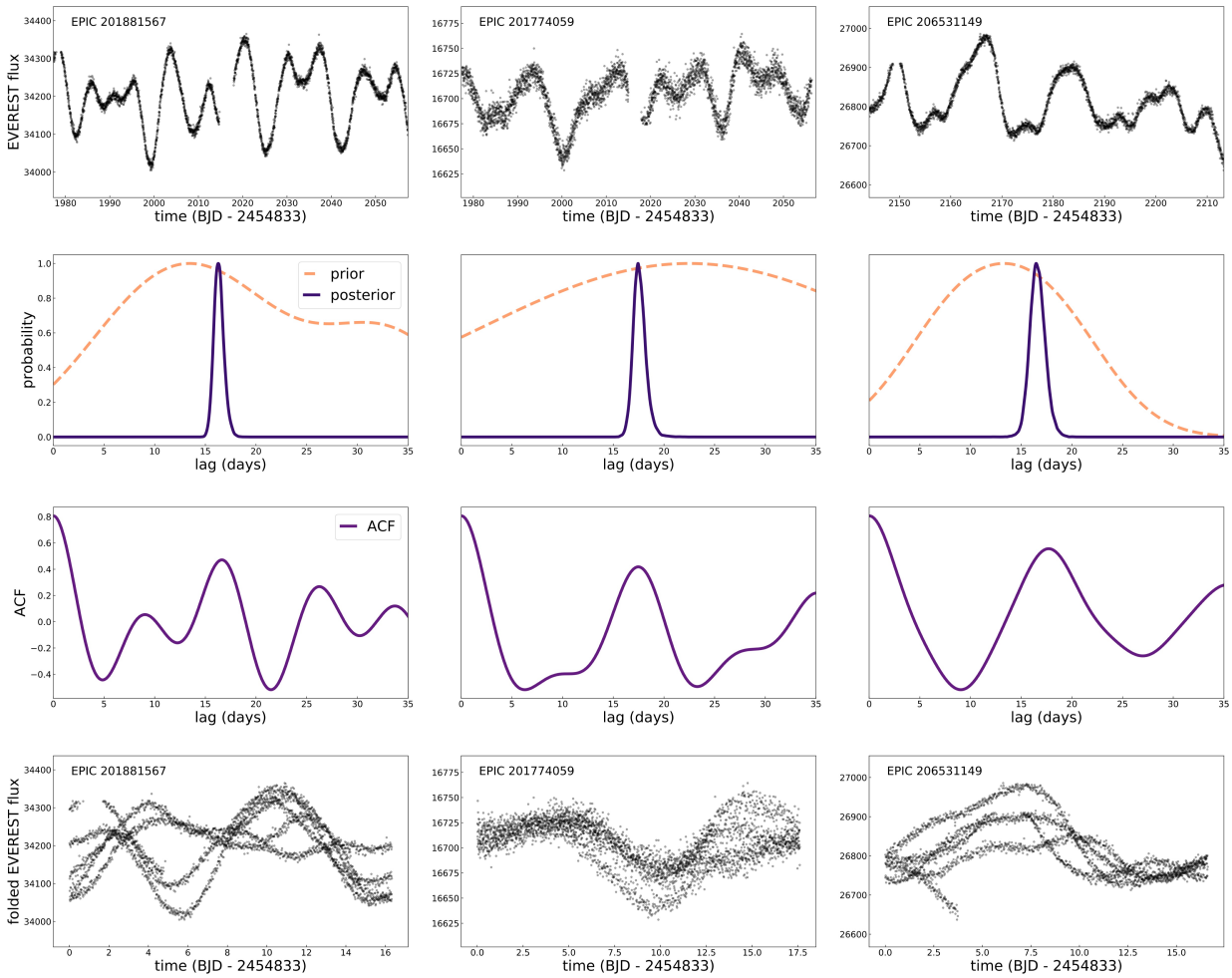

Figure 2.1 Sample output from our period detection procedure for three K2 stars with well-determined rotation periods. The top row of panels show the cotrending basis vector-detrended EVEREST flux. The second row shows the period prior and MCMC-estimated posterior. The third row shows the smoothed autocorrelation function, and the bottom row shows the light curve folded on the mean of the posterior for the period.

For visibility of the prior, both the prior and posterior are normalized such that their maximum probability is 1.

For the periodic terms we follow Foreman-Mackey et al. (2017) in setting

$$\begin{aligned} Q_1 &= 1/2 + Q + \Delta Q \\ Q_2 &= 1/2 + Q \\ \omega_1 &= \frac{4\pi Q_1}{P\sqrt{4Q_1^2 - 1}} \\ \omega_2 &= \frac{8\pi Q_1}{P\sqrt{4Q_1^2 - 1}} \\ S_1 &= \frac{\sigma^2}{(1+f)\omega_1 Q_1} \\ S_2 &= \frac{f\sigma^2}{(1+f)\omega_2 Q_2} \end{aligned} \tag{2.8}$$

where Q is the quality factor, ΔQ specifies the offset in the quality factor between the two oscillators, P is the period of the oscillator, σ^2 is the variance of the oscillation, and f specifies the fractional contribution of the oscillator at the half period $P/2$ compared to the oscillator at the full period. For the aperiodic term we set:

$$Q_3 = 1/\sqrt{2} \tag{2.9}$$

while ω_3 and S_3 are free parameters. Setting $Q_3 = 1/\sqrt{2}$ means that the third oscillator is critically damped and will not display periodic oscillations. For this term the power spectrum simplifies to

$$S_3(\omega) = \sqrt{\frac{2}{\pi}} \frac{S_3}{(\omega/\omega_3)^4 + 1}. \tag{2.10}$$

The full variability model including both the quasi-periodic and aperiodic terms has the power spectrum

$$S(\omega) = \sum_{i}^{3} S_i(\omega). \tag{2.11}$$

To compute the GP model we use the `celerite` GP method (Foreman-Mackey et al. 2017) as implemented in `exoplanet` (Foreman-Mackey et al. 2019). We maximize the GP likelihood with respect to the EVEREST co-trending basis vector detrended flux for the parameters $\{P, Q, \Delta Q, A, f, S_3, \omega_3\}$, where $A = \sigma^2$ is the variance of the periodic component of the GP. We then compute the maximum likelihood solution, initialized at the period obtained from our autocorrelation function analysis, to use as a starting point for the MCMC analysis. We use uninformative priors for all GP hyperparameters except the period, for which we use the multi-modal Gaussian mixture prior described previously. We use the NUTS sampler provided by PyMC3 **?** to run 1000 tuning samples followed by 500 production samples on each of 28 cores for a total of 28,000 tuning and 14,000 production samples. We have found that a relatively large number of tuning samples is helpful for achieving convergence when using a multi-modal period prior, in order to allow the sampler to fully explore the multi-modal likelihood space.

In Figure 2.2 we show the variation in the binned mean of the period P, maximum quality factor $Q_{\max} = \max(Q_1, Q_2)$, the logarithm of the ratio between the variances of the periodic and aperiodic components of the model, and the logarithm of the fractional uncertainty in rotation period.

2.3.3 Selecting Main Sequence Stars

We begin by making selections based on the quality of the Gaia DR2 photometric solutions (Gaia Collaboration et al. 2018). We require that the following conditions be met:

- $\sigma(G)/G < 0.01$
- $\sigma(G_{\rm RP})/G_{\rm RP} < 0.01$.

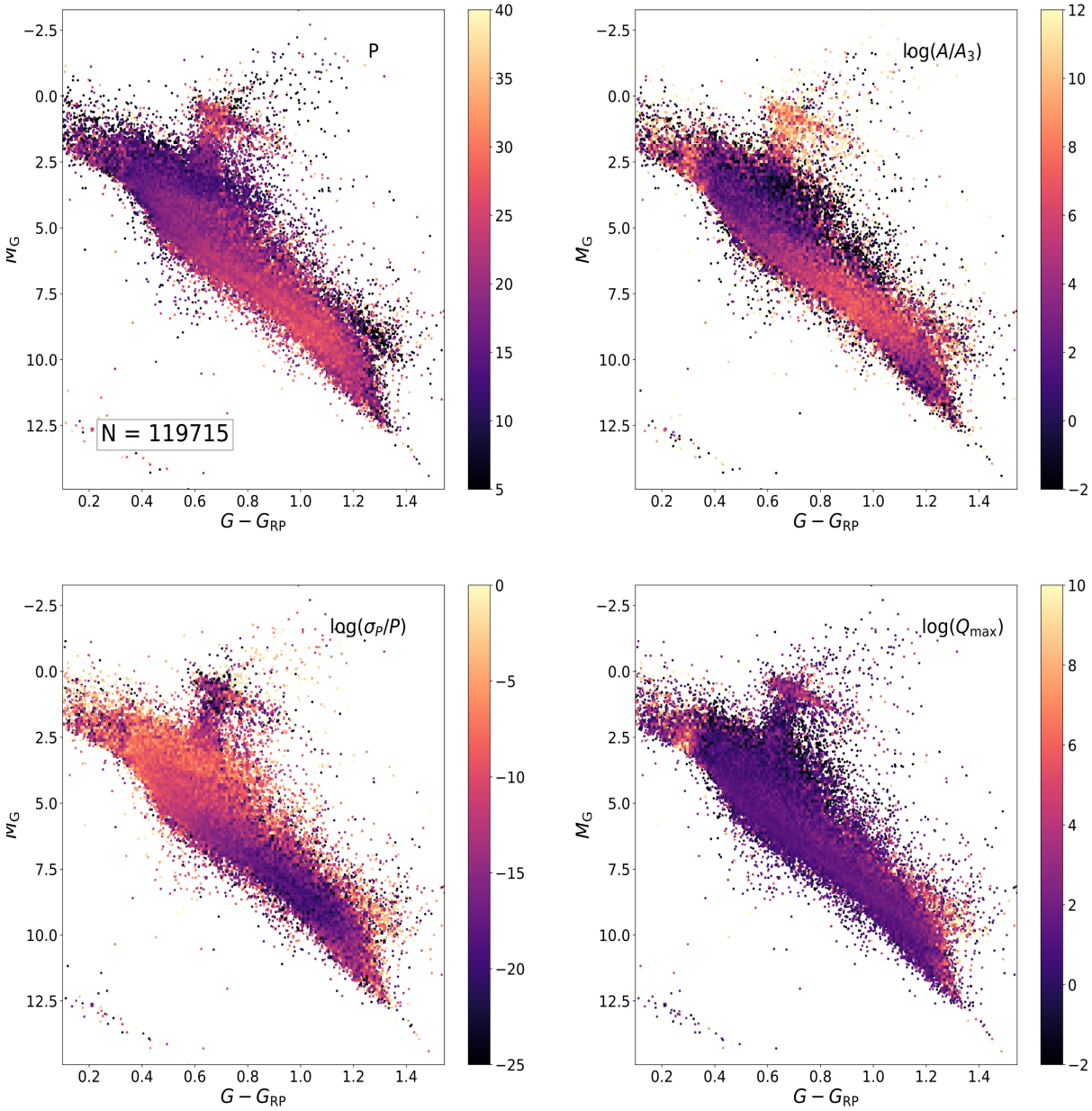

Figure 2.2 Selected hyperparameters plotted over the Gaia color-magnitude diagram. For each plot the color in a bin indicates the mean of the quantity given in the upper right-hand corner of the plot and the scattered points are colored by that quantity in regions where the density of stars is low. **Upper left:** Rotation period. **Upper right:** Log of the ratio between the periodic variance, A, and the variance of the aperiodic component, A_3. **Lower left:** Log of the fractional uncertainty for the inferred period. **Lower right:** Log of the mean quality factor, $Q_{\max}$, with larger $Q_{\max}$ indicating stronger periodicity.

where G and $G_{\rm RP}$ refer to the passbands used in Gaia DR1 and DR2.

In order to reduce contamination from giants, subgiants and unresolved binary stars, we require that stars in our sample be on or near the main sequence, as defined by a MIST isochrone (Choi et al. 2016; Dotter 2016; Paxton et al. 2011, 2013, 2015) with an age of 200 Myr and a metallicity of [Fe/H] = +0.25 to identify the nominal main sequence, and we select stars within 0.3 mag below and 0.9 mag above the isochrone, as shown in Figure 2.3. This wide slice of magnitude space allows us to encompass different ages and metallicities while reducing contamination from the giant and subgiant branches. The cost of selecting such a wide slice is that we likely incorporate a significant number of unresolved binaries into our final sample, which will add some amount of contamination. We find this acceptable since we don't expect this contamination to have a systematic influence on the overall shape of the period-color diagram.

It should be emphasized that we made no attempt to choose an isochrone that represents the actual main-sequence for stars in our sample, which would be infeasible because the K2 sample contains stars with a wide range of ages and metallicities. The choice of [Fe/H] = +0.5 was made on the basis that this isochrone does an adequate job of matching the trend of the main-sequence. We have found that the exact choice of age and metallicity does not have a significant impact on our results, so long as the isochrone and the width of the box in M_G selects a sufficient number of stars for our edge finding algorithm to perform well. Our final sample consists of 8,943 stars near the main sequence which pass the quality cuts described in the next section.

2.3.4 Vetting Rotation Periods

We select a final sample of well-measured rotators from the main-sequence sample based on MCMC convergence and period measurement precision. For inclusion of a star in our final sample, we require the following conditions be met:

- $P/\sigma_P > 15$

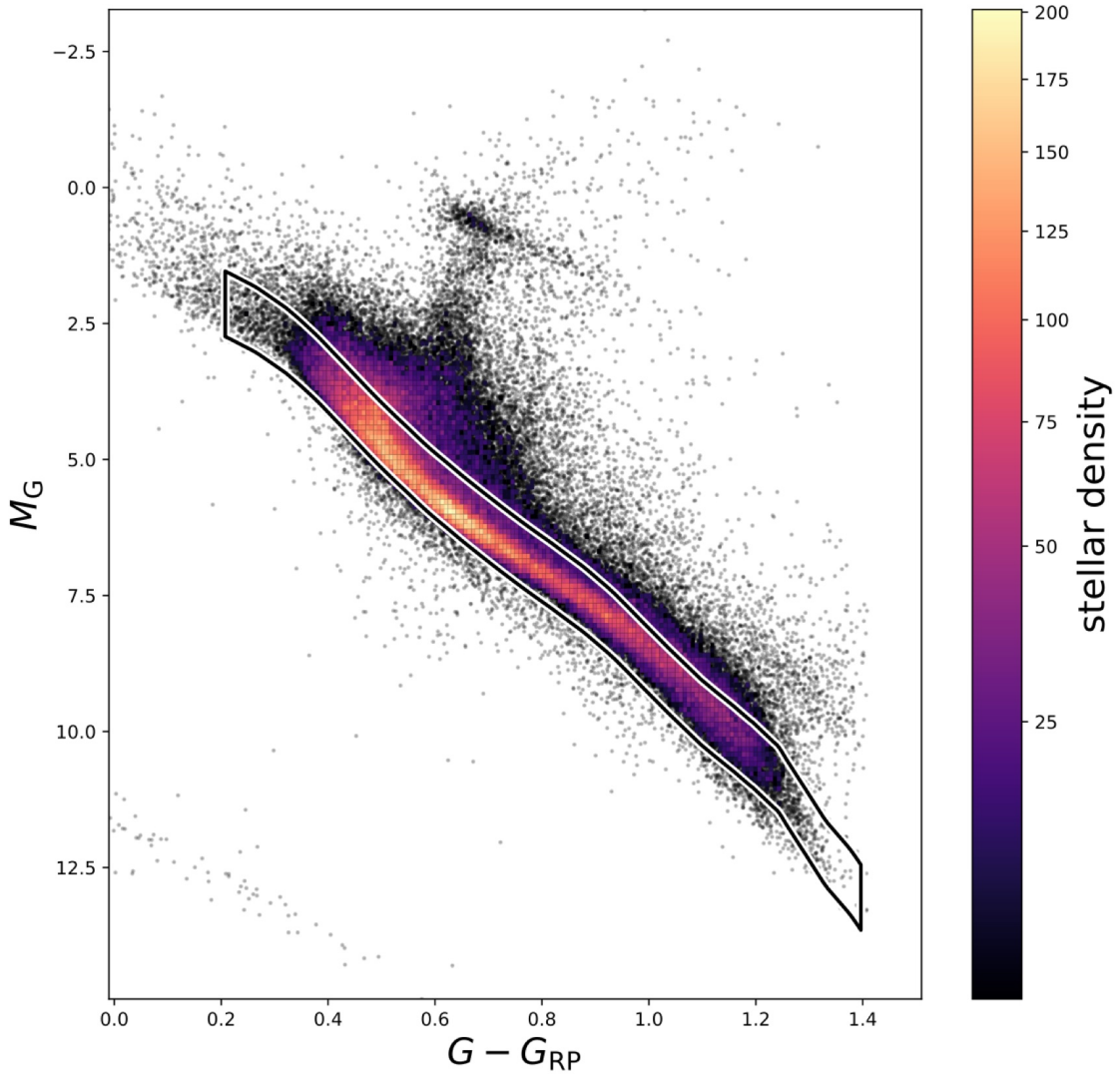

Figure 2.3 K2 stars on the Gaia color-magnitude diagram. The boxed region shows the area selected as the main sequence in order to exclude, e.g., evolved stars and unresolved binaries from our final sample. The main sequence is defined by a MIST isochrone with an age of 10^9 years and [Fe/H] = +0.5. We identify 123,079 stars belonging to the main sequence, of which 8,943 meet the requirements for our final sample.

- $\log_{10}(A/A_3) > -3$
- $0.9 < \hat{R}_P < 1.1$

where P is the measured period, σ_P is the error on the period derived from MCMC, A is the variance of the periodic GP component (the amplitude is $\sigma = \sqrt{A}$, and $A_3 = S_3\omega_3 Q_3$ is the amplitude of the aperiodic GP component. $\hat{R}_P$ is the Gelman-Rubin statistic (Gelman & Rubin 1992) which compares the variance of samples for an individual parameter (in this case the period, P) within a chain to the variance between chains. For chains that have converged to the same solution, these values will be approximately the same and their ratio, $\hat{R}$, will be close to 1. The cutoff on $\log_{10}(A/A_3)$ is meant to exclude stars for which the periodic component is very small compared to the non-periodic variability on the basis that these stars are more likely to be showing periodicity due to contamination or systematics, rather than rotation. We find that when we do not include this cutoff, a pileup of stars at a period of around two days is observed. This pileup, as shown in Figure 2.4, spans a range of stellar masses but is extremely localized in period-space and therefore appears to be artificial, though its origin is not known. Figure 2.5 shows our final sample including the main-sequence cuts described in Section 2.3.3 in blue compared to the full main-sequence sample in the P/σ_P vs. $\log_{10}(A/A_3)$ plane.

We have chosen these specific conditions with the goal of being conservative about selecting only the highest quality period measurements in order to highlight structure in the color-period diagram. As a result, many periodic signals are excluded from our final sample because they don't meet the condition $P/\sigma_P > 15$. A less selective criteria could be used to obtain a much larger sample size for applications that don't require such high precision, such as an analysis of the rotation period-metallicity relation reported by Amard et al. (2020). The exclusion of signals on the basis of P/σ_P affects the completeness of our sample most at long periods. In Figure 2.6 this means that the upper envelope of the rotation period-color distribution is not as well-defined as it might appear in the plot, especially towards the fainter K and M dwarfs towards the right-hand side of the plot.

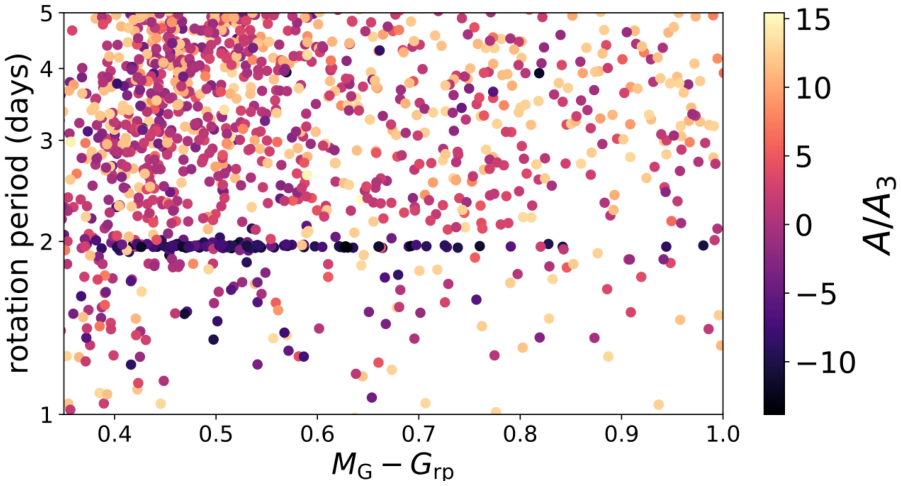

Figure 2.4 A segment of the period-color diagram showing the presumed artificial pileup at a period of two days. Points are colored by A/A_3, the amplitude of the periodic component of the GP relative to the aperiodic component. Stars in the pileup are notable for having a very small value for this ratio relative to the rest of the stars in the sample, allowing us to effectively remove this feature by imposing a cutoff in A/A_3 for our final sample.

2.4 Features in period-color space

Figure 2.6 shows the distribution of rotation periods in period-color space for our final sample. There are several prominent features in this space, among them the afformentioned gap, the upper edge of the envelope of rotation periods, the lower edge of the same envelope, and the overdensity of m dwarfs with short rotation periods in the lower right-hand corner. In this section we focus on the gap and the overdensity of fast-rotating m dwarfs. The upper edge of the envelope is not well-measured in our sample as it is largely determined by the exclusion of rotation periods longer than 32 days, as well as by our cutoff in the precision of the rotation period measurement, which preferentially excludes slow rotators.

2.4.1 The Period Gap

The gap in rotation periods extends from $\sim$ 15 days at $G - G_{\rm RP} = 0.75$ to $\sim$ 25 days at $G - G_{\rm RP} = 1.1$. This gap has been extensively studied in the Kepler data, first by McQuillan

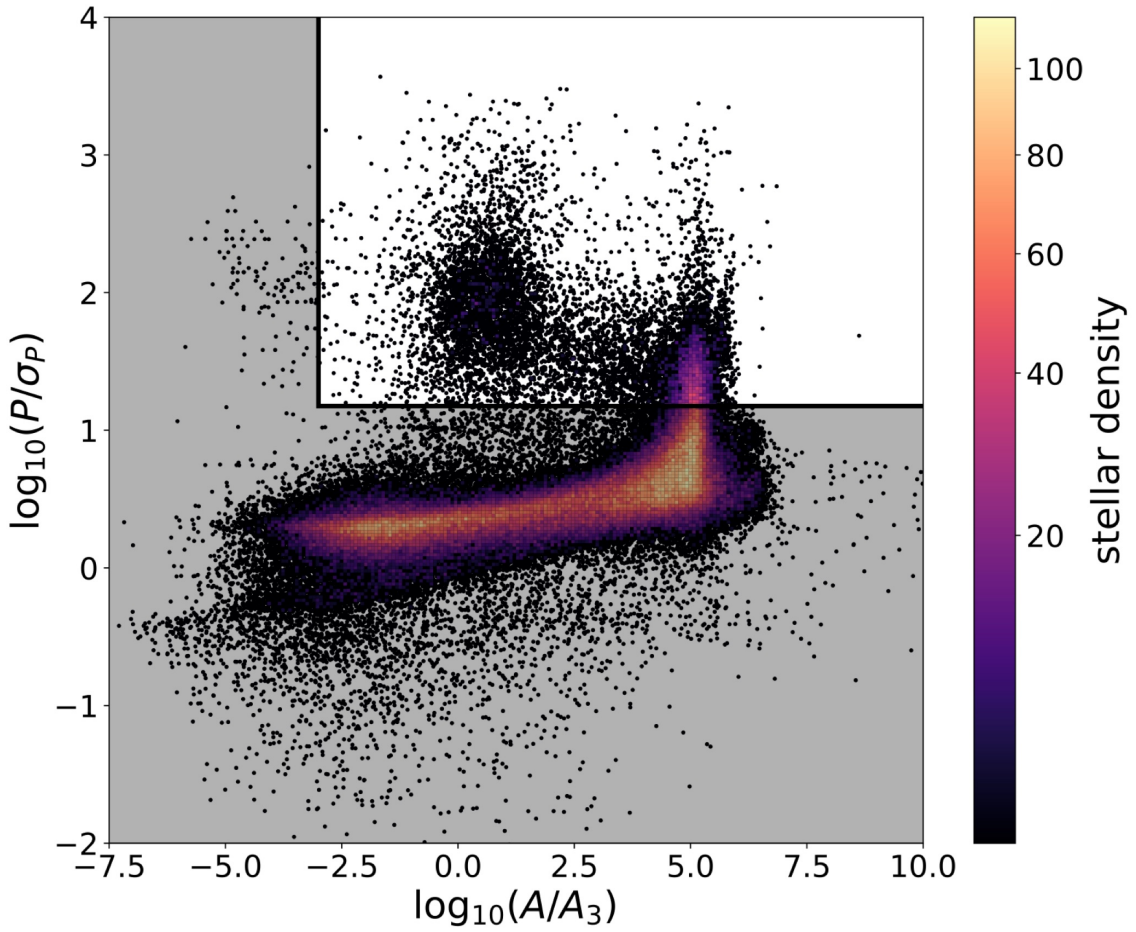

Figure 2.5 Main-sequence stars plotted in the P/σ_P vs $\log(A/A_3)$ plane. The full main-sequence sample is shown with the greyed out area representing the region excluded by the cuts detailed in section 2.3.4. Our final sample consists of the stars in the highlighted region.

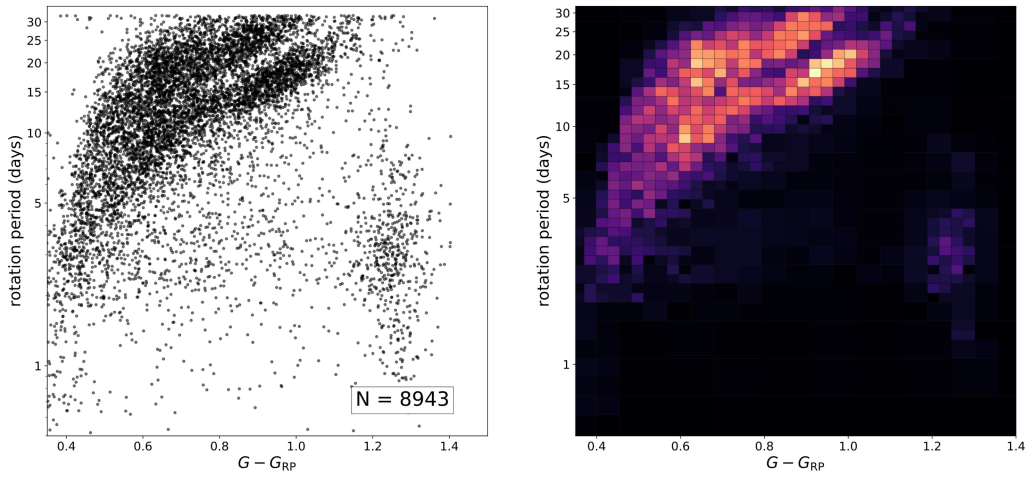

Figure 2.6 Inferred rotation periods for the 8,943 main sequence K2 stars, plotted against Gaia $G - G_{\rm RP}$ color. **Left:** scatter plot showing the measured periods and **Right:** The same rotation periods presented as a two-dimensional histogram, in order to highlight the variations in the density of stars across period-color space.

et al. (2013) for M dwarfs in the Kepler sample. In McQuillan et al. (2014) the same feature was found in Kepler K dwarfs. Davenport (2017) identified the gap for the G dwarfs, and Davenport & Covey (2018) showed that the gap is present in stellar populations out to 525 pc, beyond which the feature can no longer be recovered due to the difficulty in recovering rotation periods at large distances. Besides the tentative detection by Reinhold & Hekker (2020), our work represents the first robust measurement of this feature outside of the Kepler data. In contrast to the Kepler data, the gap detected here appears wider and has more sharply defined boundaries than is seen in the Kepler sample.

The robustness of our detection of this gap allows us to constrain new details of the feature. We use an edge detection procedure based on the Canny edge detection algorithm (Canny 1986) to find the edges of the period gap. We then fit a parametric model to the gap edges as a function of color. We measure the locations of the gap edge in each campaign individually in order to verify the presence of the gap in each campaign. We confirm that

the gap properties do not appear to depend on the direction of the K2 pointing.

We begin by applying our edge detection algorithm to the rotation period-color distribution for all campaigns combined. Our edge detection algorithm is based on the Canny edge detection algorithm, which operates on a two-dimensional array. The Canny algorithm first applies a Sobel operator to the image, which produces an approximation to the gradient. It then identifies local maxima and minima of the approximated gradient which correspond to edges (points where the intensity of the image is changing most quickly) in the original image. Our modification replaces the first step of applying the Sobel operator with the computation of a kernel gradient estimator which allows us to apply the algorithm directly to the distribution of stars in rotation period-color space, which is not possible to do directly with the Canny algorithm as it requires a two-dimensional image grid rather than a set of points in the plane. The kernel gradient estimator is defined to be the gradient of the kernel density estimator. The kernel gradient estimator is defined

$$\nabla \hat{f}(x,y) = \frac{1}{nh} \sum_{i}^{n} \nabla K_i(xh^{-1}, yh^{-1}). \tag{2.12}$$

which is the gradient of the more widely known kernel density estimator. K_i is the kernel function taken here to be a Gaussian centered at the coordinates of the i^{th} data point, n is the number of data points used to make the estimate, h is the width of the kernel which we set to 0.04, and (x, y) are the coordinate at which the kernel estimate is computed. We then apply the second part of the Canny algorithm as implemented in `scikit-image` (van der Walt et al. 2014) to identify local maxima and minima of the kernel gradient estimate. We take these local extrema to be the edges of the distribution. Figure 2.7 shows the output of this algorithm applied to our sample.

We parameterize the gap edges using a function of the form

$$\begin{aligned} P_{\text{upper}} &= A(G - G_{\text{RP}} - x_0) \\ &+ B(G - G_{\text{RP}} - x_0)^{1/2} \end{aligned} \tag{2.13}$$

Table 2.1. Best fit parameters from Equation 2.13.

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

Table 2.2. Measured gap edges and widths.

$G_{\rm BP} - G_{\rm RP}$ (mag)	$P_{\rm lower}$ (days)	$P_{\rm upper}$ (days)	gap width (days)
0.80	15.20	17.97	2.771
0.85	15.92	19.90	3.98
0.90	17.97	22.26	4.29
0.95	19.54	24.89	5.35
1.00	20.66	28.61	7.95

using the edges identified in the slice of color-space given by

$$0.8 < G - G_{\rm RP} < 1.05. \tag{2.14}$$

which corresponds to the stellar mass range:

$$0.57 M_\odot < M < 0.76 M_\odot. \tag{2.15}$$

Equation 2.13 is taken from the gyrochronology model of Barnes (2003). Our decision to fit the gap edges with this equation is motivated by the observation that the gap edges appear to have a similar trend to the gyrochrones from that work, but this choice is not meant to imply that the gap edges occur at constant age. The best-fit parameters are given in Table 2.4.1, and Figure 2.7 shows the best-fit models in period-color space.

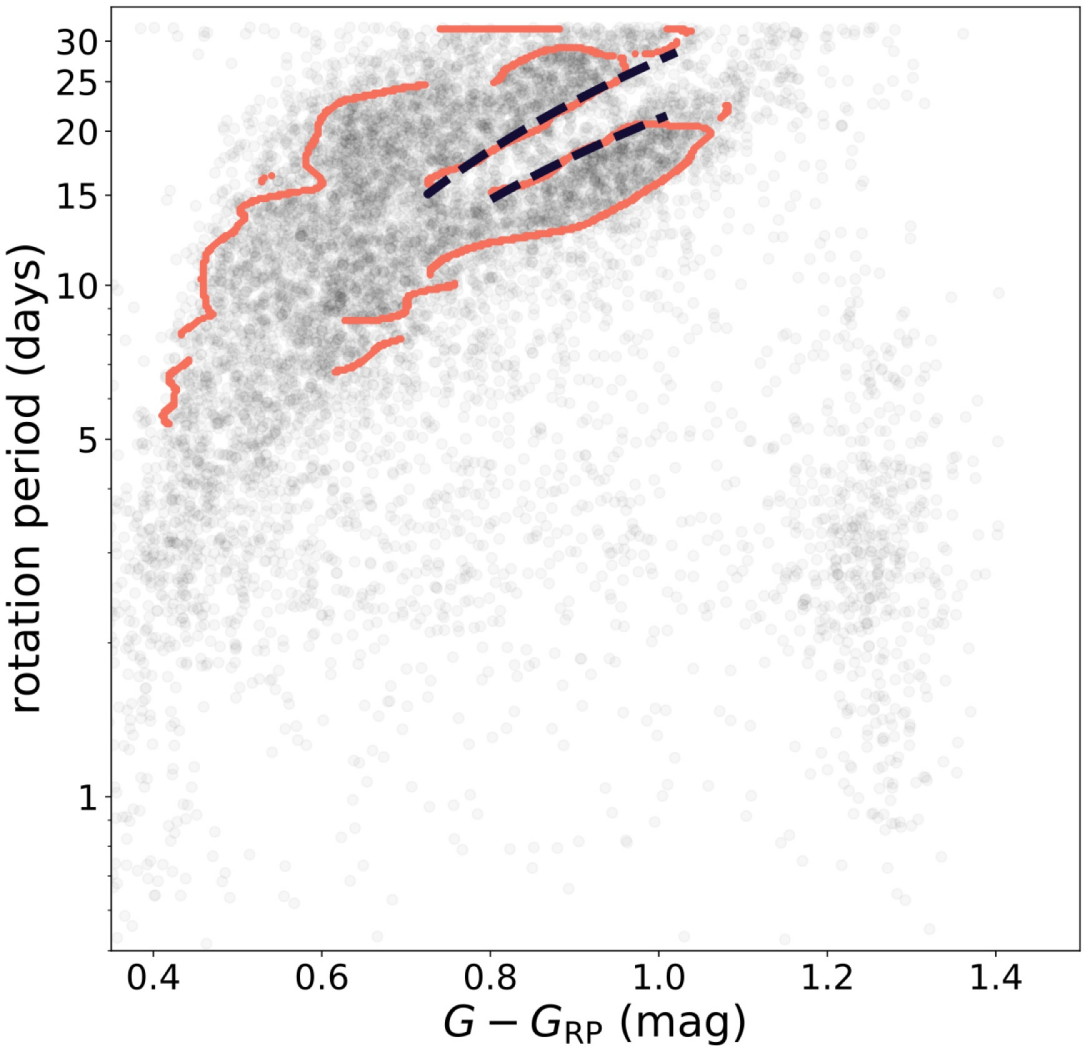

Figure 2.7 Detected edges of the rotation-period diagram using a modified version of the Canny edge detection algorithm. Best-fit models to the gap edges are shown with blue dashed lines. The model used to fit the edges is given in Equation 2.13 and the best-fit parameters are in Table 2.4.1.

This two-dimensional fit to the gap edges was conducted for all campaigns simultaneously. In order to investigate the variation between campaigns which observe different regions of the galaxy, we have estimated the locations of the gap edges in each Campaign in terms of their offset from the best-fit model for the full sample. Because the sample sizes are small for some campaigns we collapse the problem to one dimensions rather than considering the full two-dimensional period-color diagram. For each edge of the gap (upper and lower) in each Campaign, we first subtract off the gap trend and then sum over color in the range of colors for which the best-fit edge model is valid:

$$0.8 < G - G_{\rm RP} < 1.05. \tag{2.16}$$

This gives us the one-dimensional period distribution in terms of the offset from the edge locations defined in table 2.4.1. Treating each gap edge separately, we apply the Gaussian kernel derivative estimator, which is the one-dimensional anologue of Equation 2.12, to the period distribution. We then identify the local maximum (in the case of the upper edge) or minimum (in the case of the lower edge) nearest to the zero-offset point and take this to be an estimate of the location of the edge relative the edges in Figure 2.7. Figure 2.8 illustrates this procedure for Campaign 8.

Figure 2.10 shows the lower edge locations plotted against the upper edge locations for 16 of the 18 campaigns. We have excluded Campaigns 0, 2, and 11 which have too few stars to make an accurate determination of the edge locations. The error bars are determined by bootstrap resampling from the full sample. We do not observe an obvious correlation between the upper and lower gap edges, which would indicate a shifting of the gap towards longer or shorter periods for some campaigns. There are a few outlier campaigns, with Campaign 18 being the most significant. Campaigns 4, 8, 10, 13, and 15 also deviate noticeably from at least one of the measured gap edges for the all-campaign sample. In Figure 2.11 we show the period-color diagrams for these outliers. We note that Campaigns 5 and 18 observed the Praesepe cluster which imprints a visible sequence of stars corresponding

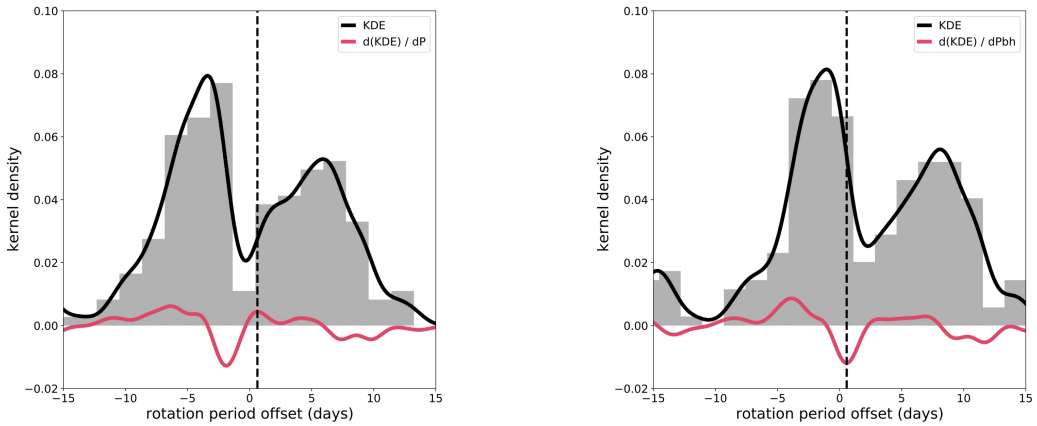

Figure 2.8 Histograms showing the distribution of rotation periods for stars in Campaign 8, in the color range $0.8 < G - G_{\rm RP} < 1.05$. The kernel density estimate is shown in black, and its derivative is shown in red. The locations determined for the gap edges are shown by the dashed vertical line. The left panel shows the upper gap edge and the periods are given as the difference between the observed rotation period and the trend of the upper gap edge. The right panel shows the same for the lower gap edge.

to a 600-700 Myr gyrochrone onto the diagram. This does not appear to influence the gap measurement for Campaign 5, but for Campaign 18 the Praesepe cluster is likely responsible for the displacement of the lower gap edge from the expected position. For the rest of the outliers we don't see any obvious evidence of a systematic displacement in the gap edges from the locations derived from the all-campaign data. In general we don't find the outliers to be significant, and we conclude that the rotation period gap shows no dependence on the direction of the K2 pointing.

To further illustrate this point, Figure 2.12 shows the locations of the K2 footprints for these outlying campaigns on the sky relative to the rest of the campaigns. We note that the outlier campaigns are in general widely separated on the sky, appear both above and below the galactic plane, and show no evidence of clumping. We therefore state our conclusion that the rotation period gap appears to be an isotropic feature of the stellar populations in the nearby Milky Way.

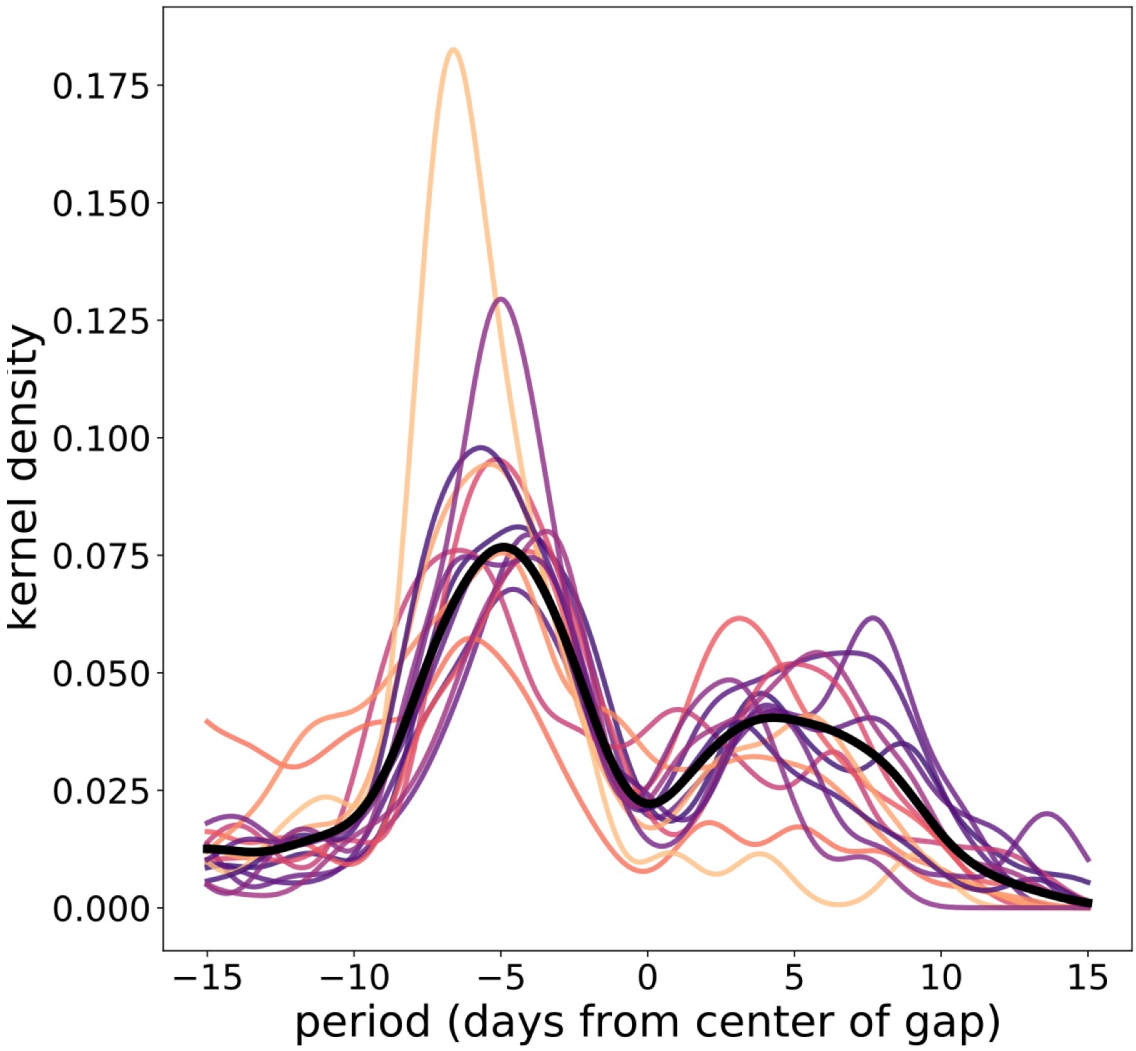

Figure 2.9 Kernel density estimates for the 16 campaigns with $N > 200$, for stars with $G - G_{\mathrm{RP}}$ in the range defined in equation 2.16. The thick black curve is the kernel density estimate for all 16 campaigns combined.

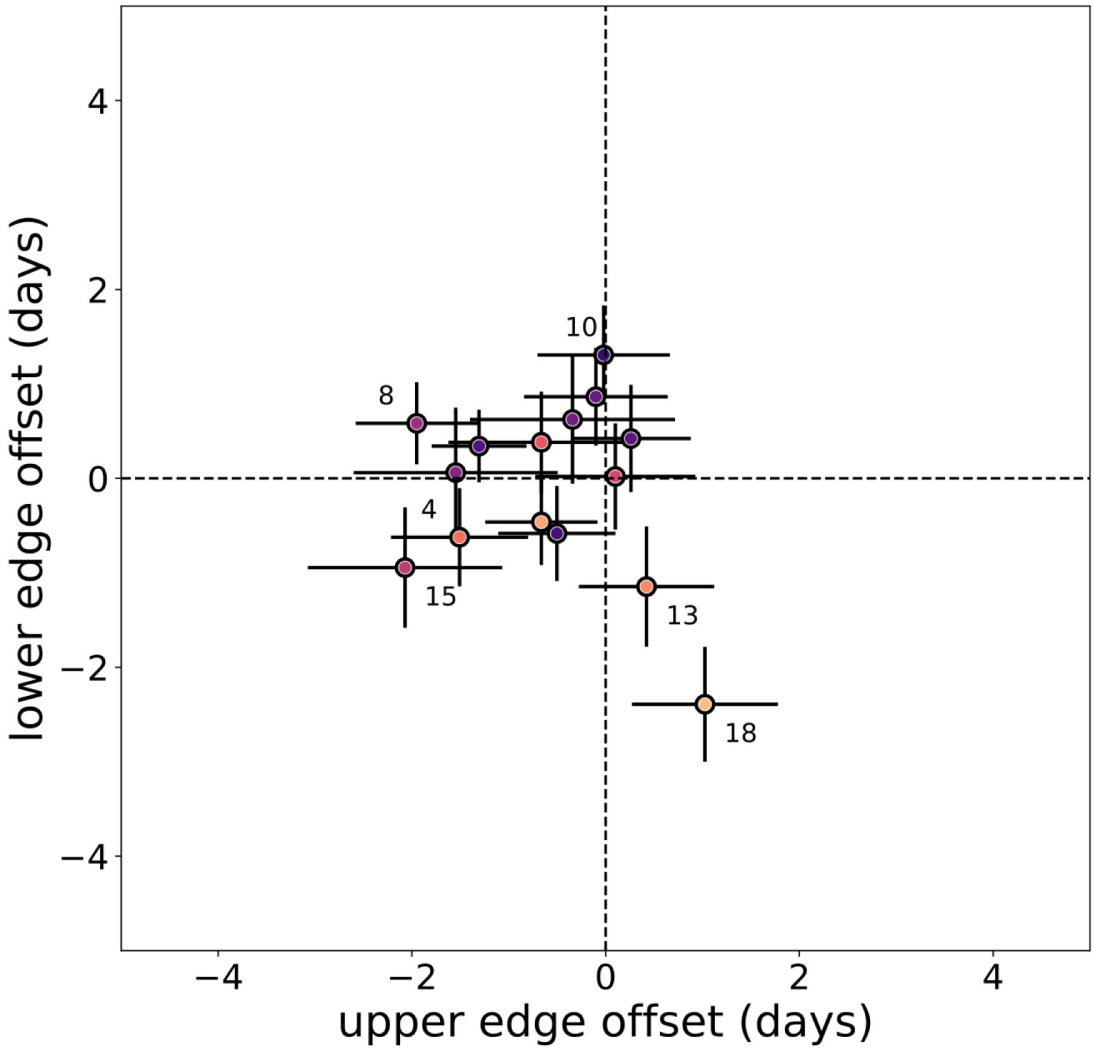

Figure 2.10 Location of the lower edge of the gap plotted against the location of the upper edge. Both gap edge locations are given as the displacement in days from the best-fit gap edges in Figure 2.7 and are determined from the slope of the one-dimensional kernel density estimates in figure 2.9. The four most notable outliers are labeled and their period-color diagrams are shown in figure 2.11.

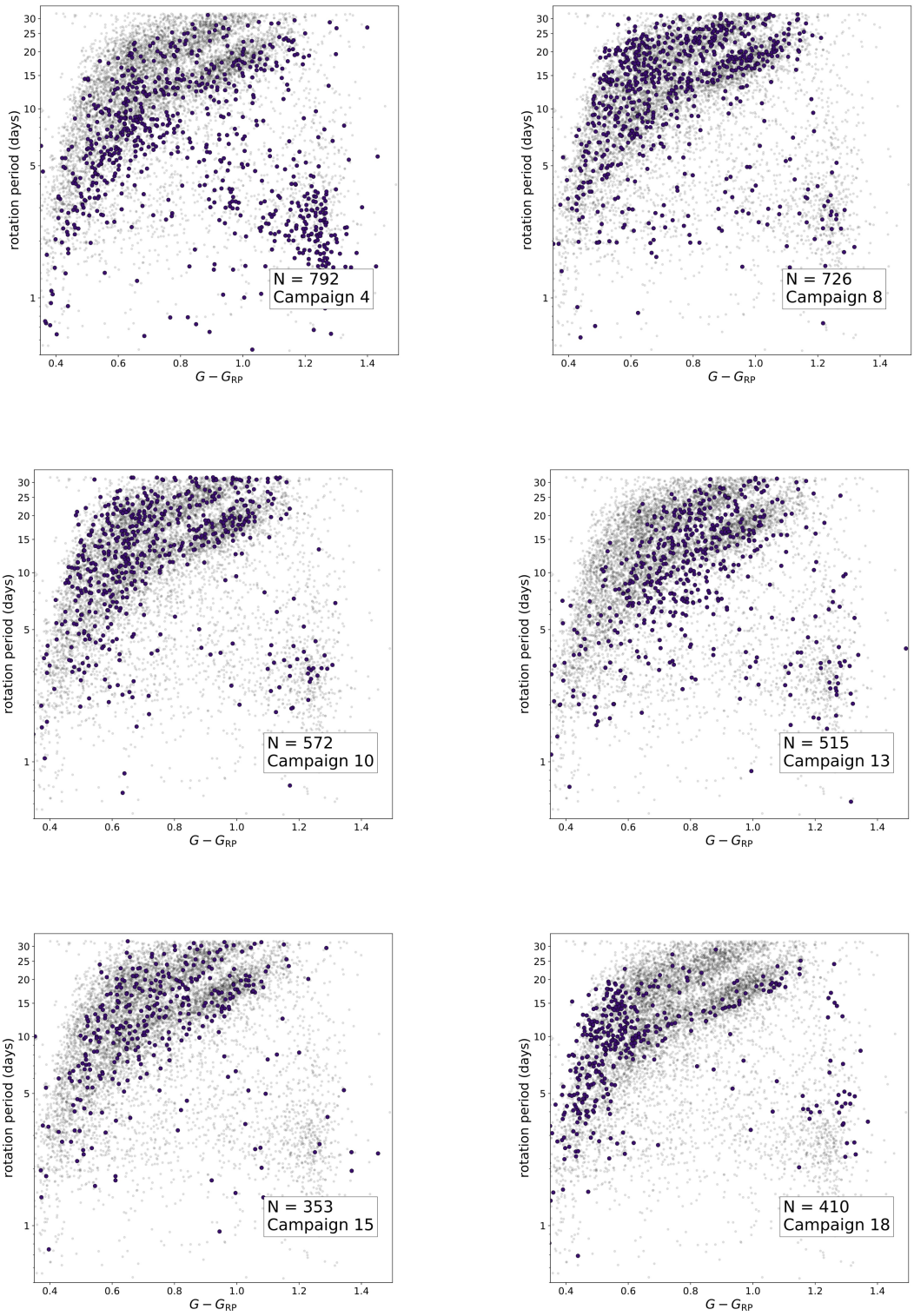

Figure 2.11 Period-color plots for the outliers identified in Figure 2.10. The dark purple points show the stars from the individual campaign, while the gray points are the full sample of 8,943 stars. The locations of the K2 footprints on the sky for these campaigns is shown in Figure 2.12. The Praesepe sequence can be seen in Campaign 18, but it only appears

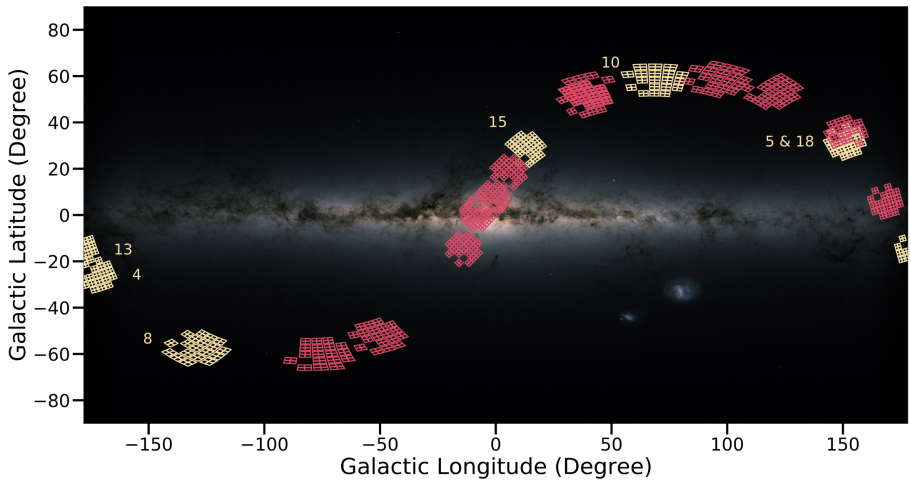

Figure 2.12 Positions of the campaigns shown in Figure 2.11 with Milky Way as seen by Gaia DR2 for reference. There is no obvious correlation between the direction of the K2 pointing and the change in shape or position of the gap for the outlier campaigns. Background image credit: ESA/Gaia/DPAC.

2.4.2 Other features in the rotation period-color diagram

In addition to the prominent period gap, a major feature of the period-color distribution seen here is the over-density in the bottom right corner, which represents a population of fast rotating M-dwarfs. This population has been studied in young open clusters (Rebull et al. 2016, 2018) as well as in the MEarth sample (Newton et al. 2016). The break between the slow rotating M dwarfs for which the rotation period increases with decreasing mass, and the fast-rotating sequence for which the rotation period decreases with decreasing mass occurs at approximately the mass at which M dwarfs become fully convective and corresponds to a change in the morphology of the surface magnetic field from a more complex towards a more simple configuration (Morin et al. 2010; Garraffo et al. 2018).

We find that the light curves of these fast-rotating M dwarfs show rotational modulation that is more periodic than those of the other stars in our sample. For our GP rotation model the degree to which a light curves shows periodic variations is measured by the parameters

Q_1 and Q_2 in equation 6.3. Larger values of Q_1 and Q_2 mean that the power spectrum of the variability is more sharply peaked about ω_1 and ω_2. In this analysis we consider the maximum of (Q_1, Q_2) which we call $Q_{\rm max}$. Figure 2.14 demonstrates how this parameter effects the appearance of the light curve. We interpret a larger $Q_{\rm max}$ value to indicate that features on the star's surface are stable over a longer period of time, giving rise to variations that are coherent across many periods.

The right panel of Figure 2.13 shows how $Q_{\rm max}$ varies across the period-color diagram. From this figure we see that stars with large $Q_{\rm max}$, indicating stronger periodicity, cluster in the fast-rotating M dwarfs. These stars are, however, not limited to this cluster and occur in lower densities across the full range of $G - G_{\rm RP}$ at short rotation periods. The left panel of Figure 2.13 shows our sample in period-$Q_{\rm max}$ space with stars colored by their $G - G_{\rm RP}$ color. In this space we observe a distinct population of strongly periodic rotators with a negative correlation between rotation period and $Q_{\rm max}$.

2.5 Discussion

There has been much interest in, and discussion of, the origin of the rotation period gap, with several promising possible explanations having been put forward since its discovery (Angus et al. 2020; Davenport & Covey 2018; Reinhold et al. 2013, 2019). We now consider these potential explanations in light of our new measurements, as well as taking into account recent work on the 2.7 Gyr cluster Ruprecht 147 by Curtis et al. (2020) and Gruner & Barnes (2020), which crosses the rotation period gap.

McQuillan et al. (2014) and Davenport & Covey (2018) propose that the gap may be an artifact of a recent (< 500 Myr) burst of star formation in either the Solar neighborhood or in the direction of the Kepler field, which would have produced a population of young, fast-rotating stars that make up lower branch of the observed bimodality. The single pointing of the Kepler mission admitted the possibility that this feature confined to that field. Our sample has the benefit of K2's multiple pointings, which has allowed us to demonstrate that the bimodality is present in all directions and is therefore not unique to the stellar popula-

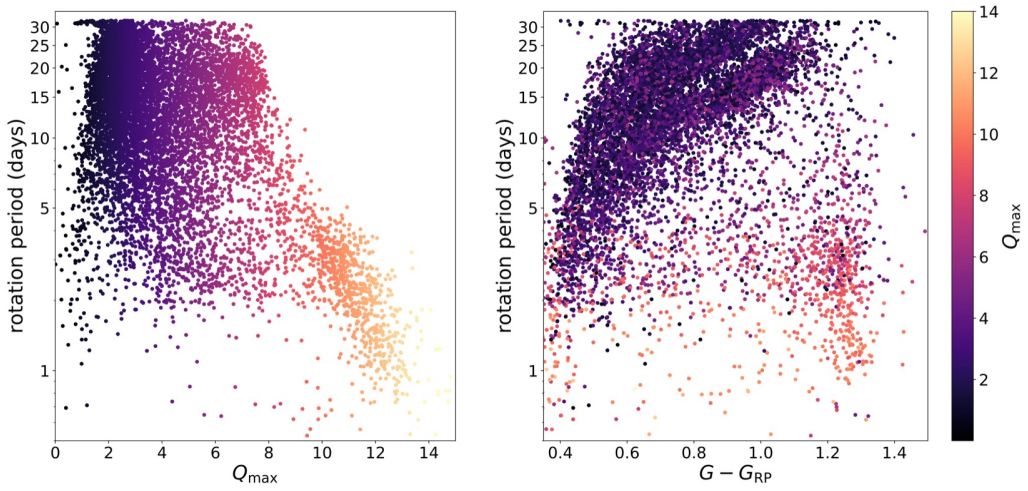

Figure 2.13 **Left:** Sample in period-$Q_{\max}$ space, showing that sinusoidal rotators (high-Q stars) cluster separately from the main population and preferentially occur at short rotation periods. **right:** Sample in period-color space with stars colored by the maximum quality factor $Q_{\max}$. While stars with higher $Q_{\max}$ values cluster in the fast-rotating M dwarfs, they also occur across all colors, and hence across all stellar masses in our sample.

tion observed by Kepler. The possibility remains that the bimodal star formation history suggested by McQuillan et al. (2014) and Davenport & Covey (2018) might be isotropic. However, the position and shape of the gap revealed by our sample makes this explanation untenable, as the trend of the gap shows a sharper slope than the sequences associated with constant age populations from Praesepe and NGC 6811 (e.g. Curtis et al. 2019). It is interesting to note that the sequence of stars associated with the 2.7 Gyr cluster Ruprecht 147 appears to cross the gap around $G - G_{\rm RP} \sim 0.7$. While caution should be exercised due to the fact that this sequence has a relatively smaller number of stars than for the younger clusters and shows a large intrinsic scatter, this apparent crossing of the gap lends further evidence against the hypothesis that the gap represents a feature at constant age. Indeed, as Curtis et al. (2020) note in their analysis of periods for Ruprecht 147 and other clusters, this gap-crossing seems to occur at a roughly fixed Rossby number, rather than at a single fixed age, which agrees with our assessment.

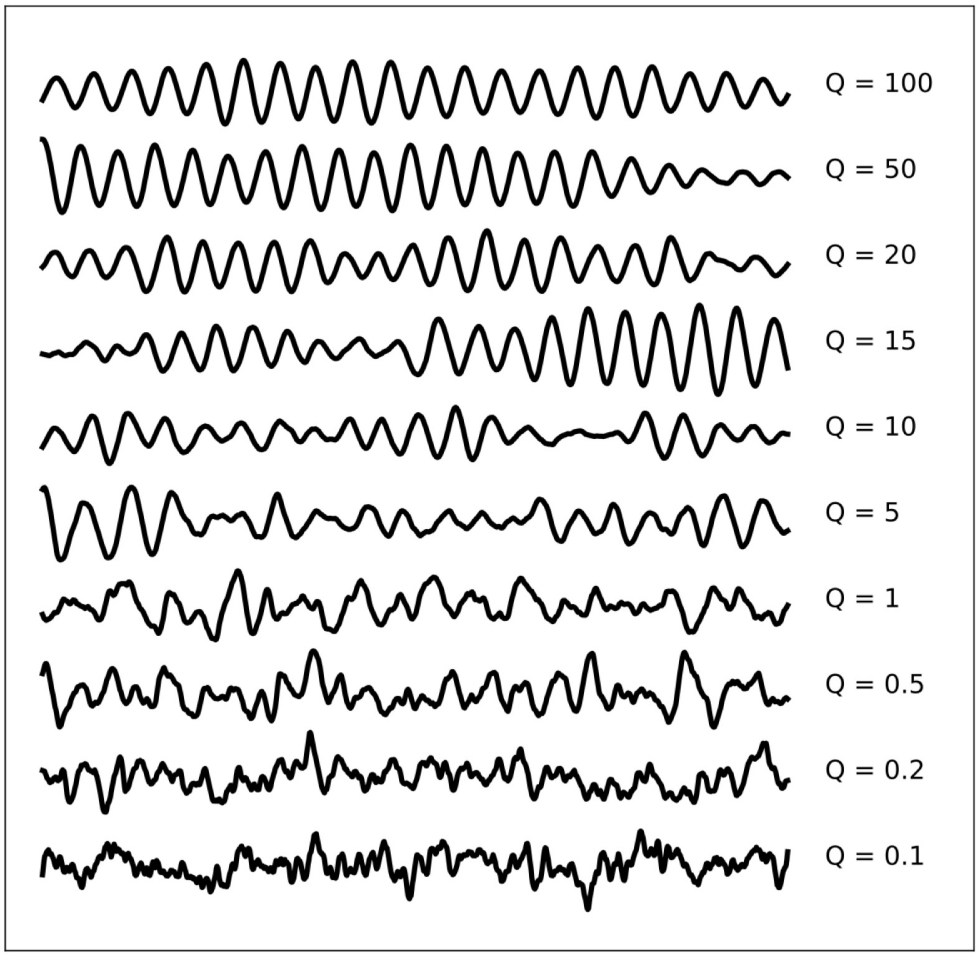

Figure 2.14 Samples drawn from our GP model showing the effect of increasing the quality factor, Q, on a light curve. All light curves have the same period and amplitude. We have set $\Delta Q = 0$ for these simulations so that $Q = Q_{\max}$. A higher Q value means that the light curve shows stronger periodicity. In terms of stellar rotation, this likely indicates that surface features are stable for a longer period of time when Q is large.

Reinhold et al. (2013) and Reinhold et al. (2019) suggest that the gap is an artifact of the transition from spot-dominated to faculae-dominated photospheres as stars age. In this explanation the gap would result from a minimum in the detectability of rotation periods for stars at the point in this transition where neither spot nor facula-induced variability are able to dominate the light curves of these stars. Our measurements of the gap do not rule out this explanation. To do this would require more work on the evolution of stellar activity over a range of ages and spectral types.

Our preferred explanation is that the gap emerges from a period of accelerated spindown immediately after the stalled spindown noted by Curtis et al. (2020). This explanation was first put forth by McQuillan et al. (2013), but the hypothesis was dismissed in favor of the "two populations" hypothesis preferred by McQuillan et al. (2014) and Davenport & Covey (2018).

In this scenario, a young star with its envelope initially decoupled from its core would experience magnetic braking, reducing the spin of the envelope while the decoupled core would be allowed to continue its faster rotation. At a later time the core and envelope would begin to exchange angular momentum. At this point the transfer of angular momentum from the core to the envelope would slow or even halt spindown by offsetting magnetic braking at the surface, resulting in an overdensity of stars just below the period gap. The underdensity making up the gap itself could then be explained by a period of increased spindown once this coupling is complete and before the star resumes ordinary Skumanich spindown. This could be due to a temporary increase in magnetic activity. Lanzafame & Spada (2015) and Spada & Lanzafame (2020) have developed a spindown model featuring a mass-dependent core-envelope coupling timescale which reproduces the stalling behavior and has been applied to observations of open clusters. Curtis et al. (2020) have found the stall in spindown corresponds to a track of roughly constant Rossby number. Angus et al. (2020) suggests that this mechanism may explain the period gap as a break between a "young" regime in which rotation periods increase with decreasing mass from an "old" regime in which rotation periods are nearly constant or even decreasing with decreasing mass, with

the gap representing a period of relatively fast spin evolution during the transition between these regimes. In Figure 2.15 we plot rotation measurements for several important clusters over the distribution of K2 field stars to show the correspondence between the location of the gap in the field stars and the apparent stalling of spindown in clusters. We use the open clusters Praesepe, which has an age of 600-700 Myrs, NGC 6811 which has an age of approximately 1 Gyr (Agüeros et al. 2018; Curtis et al. 2019), and Ruprecht 147, which is older than both at 2.7 Gyrs. The sequences for Praesepe and NGC 6811 sit on top of each other for low mass stars, but have diverged for stars more massive than about 0.9 $M_{\odot}$, suggesting that lower mass stars have stalled in their angular momentum loss while higher mass stars have continued to spin down. By the time we reach the age of Ruprecht 147, spindown has resumed for stars down to about $0.7M_{\odot}$. The location of the gap in the rotation measurements for the K2 field stars coincides with the point at which the clusters transition from stalled spindown at low masses to resumed Skumanich spindown at higher masses. This supports the notion that the gap represents a discontinuity between these two regimes of spindown.

There is still much work to be done to determine whether core-envelope coupling and decoupling fully explains these observations. On the observational side it will be important to continue to benchmark clusters of ages between 1 Gyr and solar age, as clusters in this age range may cross the gap (similar to Ruprecht 147). On the theoretical side, models of rotational evolution that can explain the period of rapid spindown after the epoch of stalled spindown and in doing so reproduce the shape and trend of the gap will be important for testing this explanation. Another promising avenue of investigation may be kinematic dating of field stars (Angus et al. 2020). Proper motion measurements from Gaia may provide us with the ability to estimate ages of stellar populations by the vertical component of their motion with respect to the galactic disc, since stars become excited in this direction by dynamical interactions over time. This would allow for independent calibration of gyrochronological relations, which may shed light on how stars evolve across the gap.

Finally, while the population of stars within the gap is small, it appears to be nonzero,

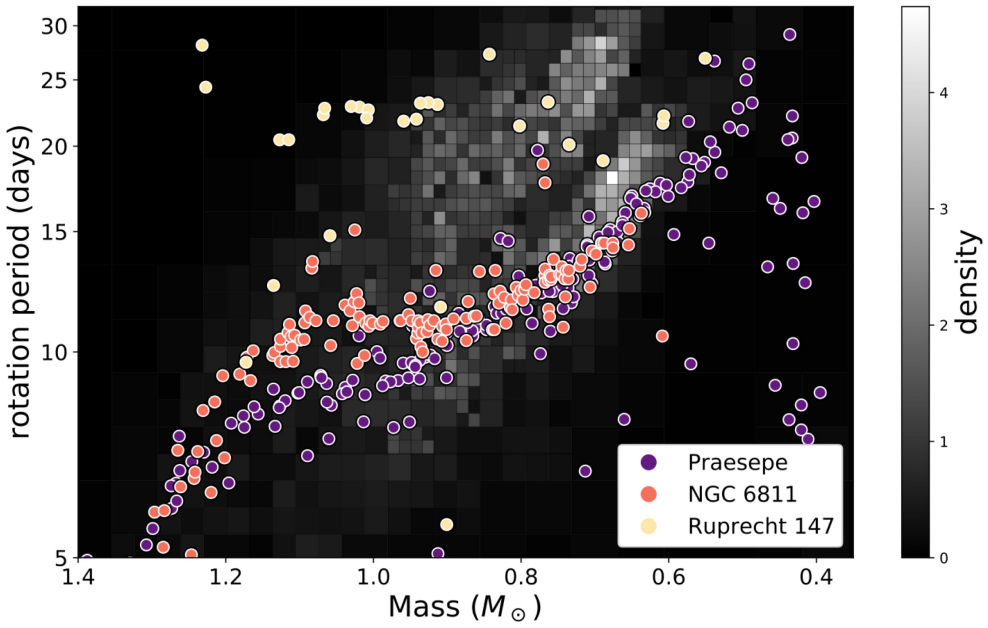

Figure 2.15 Clusters Praesepe, NGC 6811, and Ruprecht 147 superimposed on the distribution of field stars. For Praesepe, we use our own rotation measurements with membership in the cluster taken from Douglas et al. (2019). Rotation periods for NGC 6811 are from Curtis et al. (2019), and for Ruprecht 147 from Curtis et al. (2020).

which opens up the possibility of targeted studies of stars that are currently crossing the gap. Detailed observations of individual stars in the gap or near the lower boundary of the gap may reveal interesting aspects of their activity and the processes that shepherd them across this span of the color-period diagram.

2.6 Conclusions

We have measured precise rotation periods for 8,943 main-sequence K2 stars by Gaussian process regression. We perform MCMC simulations on each light curve to obtain estimates of the GP hyperparameters and their uncertainties. We detect and measure the gap in the rotation period distribution and show that this feature appears in all K2 campaigns and is thus unlikely to result from a peculiarity of the stellar populations observed by Kepler. We review several explanations for the gap and argue that the most likely is that the gap results from stalled spindown on the fast-rotating sequence for low-mass stars, followed by rapid evolution across the gap to the slow-rotating sequence. This evolution may be governed by time-variable core-envelope coupling, which controls the rate of transfer of angular momentum from the core to the surface of the star.

In the future, TESS observations will provide a large sample of light curves for field stars. We expect that a similar distribution of rotation periods will be observed for this sample. One key observation that TESS may enable is whether or not the gap extends to stars more massive than $\sim 0.8 M_{\odot}$. If the gap represents the space between two separate stellar populations at different ages, then it should extend to higher mass stars, but if the gap emerges from the physics of core-envelope coupling then we may expect to observe a mass-dependence for the phenomenon.

Finally, one dimension that has been left out of this work is that of metallicity. Amard et al. (2020) reports a metallicity dependence to stellar rotation in the Kepler which may be detectable in our K2 sample as well. As the inner structure and the evolution of a star is known to be dependent on its chemical composition, this dependence of rotation period on metallicity may help to illuminate the relationship between interior structure and spindown.

We leave the task of exploring this relationship to future work.

By making the full results of our MCMC simulations available to the community we hope to make it possible for other researchers to make different choices about which periods to include and exclude. Machine learning techniques such as convolutional neural networks or random forests may also be useful for identifying rotation signals in EVEREST light curves. Combining these techniques with our period measurements may result in a larger final sample without sacrificing quality. Our sample may also be of use as a training set for machine learning algorithms seeking to identify stellar rotation signals.

In this chapter we have demonstrated the application of a Gaussian process variability model to a large sample of variable white-light lightcurves. In the next chapter we develop a two-dimensional GP model suitable for modeling multiband observations. While the machinery employed to optimize the GP hyperparameters and sample the likelihood function will be much the same as in this chapter, the method differs in that the goal of the work is to produce superior estimates of the mean hyperparameters rather than the kernel hyperparameters. Despite that small difference, the lessons learned about GP modeling in this chapter will serve us well moving forward.

Chapter 3

IMPROVING TRANSIT MEASUREMENTS WITH WAVELENGTH-DEPENDENT STELLAR VARIABILITY MODELS

Portions of this chapter were originally published in collaboration with Eric Agol and Daniel Foreman-Mackey in the November 2020 issue of The Astronomical Journal (Tyler A. Gordon et al 2020 AJ 160 240) and are reproduced here with permission of the American Astronomical Society.

3.1 Summary

Gaussian processes (GPs) are commonly used as a model of stochastic variability in astrophysical timeseries. In particular, GPs are frequently employed to account for correlated stellar variability in planetary transit light curves. The efficient application of GPs to light curves containing thousands to tens of thousands of datapoints has been made possible by recent advances in GP methods, including the `celerite` method. Here we present an extension of the `celerite` method to two input dimensions, where, typically, the second dimension is small. This method scales linearly with the total number of datapoints when the noise in each large dimension is proportional to the same `celerite` kernel and only the amplitude of the correlated noise varies in the second dimension. We demonstrate the application of this method to the problem of measuring precise transit parameters from multiwavelength light curves and show that it has the potential to improve transit parameters measurements by orders of magnitude. Applications of this method include transmission spectroscopy and exomoon detection, as well a broader set of astronomical problems. This work has particular relevance to JWST transit observations because JWST is capable of observing spectral timeseries at high precision. In this regime wavelength-dependent stellar variability may impact transit measurements.

3.2 Introduction

All exoplanet transit observations must contend with the presence of noise. Light curves can display both uncorrelated, or white, noise and correlated noise. While white noise often results from the the statistics of photon counting, and may only be ameliorated by collecting more photons, correlated noise can arise from a variety of sources. These can be broadly

divided into two categories: astrophysical noise, which results from physical processes at the source of the observed photons such as stellar granulation and oscillations (Pereira et al. 2019; Barros et al. 2020; Sulis et al. 2020; Morris et al. 2020), and instrumental noise, which results from imperfections in detectors, errors in spacecraft pointing, or other processes taking place at the location of the observer rather than at the source.

Our ability to detect transits and infer their parameters depends on how well we can model both white and correlated noise. While white noise is straightforward to model as a Gaussian distributed random variable,[1] correlated noise can be more challenging to account for. Additionally, as more powerful telescopes yield more precise observations, photon-counting noise will decrease while astrophysical correlated noise (which does not depend on photon counts) will not. In fact, correlated noise will become more dominant as decreasing white noise amplitudes reveal previously undetectable variability.

A number of methods have been used to model, filter, or otherwise account for correlated noise in astrophysics, dating back to work by Rybicki & Press (1992) and Rybicki & Press (1995). Among these techniques are wavelet filtering (Carter et al. 2008) and Kalman filtering (Kelly et al. 2014). A comprehensive study of various detrending methods is given in Hippke et al. (2019). These include various sliding filter methods (such as a sliding mean or median), sums of sines or cosines (Kipping et al. 2013; Mazeh & Faigler 2010), and others.

Our work focuses on the Gaussian process method of modeling correlated noise. In this chapter we introduce an extension to the popular `celerite` code which can be used to model correlated noise in two dimensions. We use this extension to simulate multiwavelength stellar variability in transit observations. We show that by accurately modeling correlation across wavelengths we can improve measurements of transit parameters by orders of magnitude in some common limits.

While this chapter focuses on multiwavelength transit observations with a small number of bands, our method also naturally extends to transit spectroscopy as the number of bands

[1] Which is the limit of a Poisson distribution at high photon count rates.

becomes large. In this chapter we consider a trapezoidal transit model that has no wavelength dependence, but a wavelength-dependent transit model can easily be incorporated. For transit spectroscopy, in practice, the transit depth and limb-darkening parameters should be allowed to vary between bands.

We further assume that the data has been preprocessed to remove instrumental systematics. Long-term trends in the observations may either be removed during preprocessing or incorporated into the mean of the GP model. If removed during preprocessing any uncertainties introduced should be carefully accounted for so that they can be incorporated into the GP. In the case studies presented in section 3.5 we assume a zero mean, indicating that the observations have been normalized to zero in each band. We assume there are either no long-term trends present or that they have been removed in pre-processing without introducing any meaningful additional uncertainty.

The final assumption we make is that the correlated component of the variability is stationary, by which we mean that the parameters describing the correlated noise do not change with time. This is a fundamental limitation that is inherited from the 1D `celerite` method. Our method does, however, allow for a heteroscedastic white noise component. This means that each datapoint is allowed to have a unique measured uncertainty that varies from point to point in both time and wavelength.

3.2.1 Gaussian Processes as a Noise Model for Transit Lightcurves

A typical and much simplified procedure for measuring exoplanet transit parameters using a Gaussian process noise model (as applied in Dawson et al. 2014; Barclay et al. 2015; Chakrabarty & Sengupta 2019, among others) can be summarized as follows:

1. Choose a suitable kernel function to describe the correlated noise.

2. Choose a transit model to use as the GP mean function.

3. Optionally, maximize the GP likelihood with respect to the mean and kernel parameters

or use another method to obtain a starting point for initializing MCMC chains.

4. Use a Monte Carlo method to sample the posterior (defined by the GP likelihood and priors for each parameter) in order to estimate uncertainties for the transit and kernel parameters.

Choosing a suitable kernel function can be a complicated task. The choice of kernel might be motivated by prior knowledge of the characteristics of the data, or might be "learned" from the data, such as with spectral mixing kernels (Wilson & Adams 2013). In section 3.3 we discuss and justify our choice of a kernel for modeling stellar variability which has the added benefit of being easily expressible in the `celerite` framework. For a more complete discussion of model selection in the general case we refer the reader to chapter 5 of Rasmussen & Williams (2006).

When searching for a previously undetected transit, the results of step 3 will suggest the most likely parameters of the transit. In a Bayesian framework the posterior estimates obtained from the MCMC analysis can then be used to estimate the evidence for a transit with respect to a flat mean.

In the case of a monochromatic light curve this procedure is effective at identifying transits when the depth or duration of the transit differs significantly from the amplitude and characteristic timescales of the noise. For instance, a transit that is much deeper than the noise amplitude is poorly described by the GP noise model and thus the likelihood will be sharply peaked at the location of the correct transit parameters. Similarly, a transit that occurs on a much shorter timescale than the characteristic timescale of the variability will be poorly described by the GP and hence easily detectable via the likelihood. Figures 3.5 and 3.6 illustrate these instances.

A problem occurs when the transit depth and duration are comparable to the noise amplitude and timescale. In this case the GP covariance alone is able to fit the transit without the need for a mean model. The result is that the GP likelihood is not sharply peaked about the location of the correct transit parameters and the transit is thus difficult

or impossible to detect. Gathering more photons with a larger telescope does not fix the problem as the correlated noise does not decrease with higher photon count rates as white noise does. One simply obtains a better measurement of the correlated noise, but the transit remains masked by the variability.

One solution to this problem is to gather light in multiple wavebands. With a multiband light curve we can leverage the difference in the spectral dependence of the transit as compared to the correlated variability to disentangle the transit from the noise, and thus detect shallower transits across a broader range in duration than is possible with monochromatic observations. This approach depends upon the assumption that the correlated noise has the same time dependence for each component of the power spectrum, but varies in amplitude with wavelength. If, on the other hand, the correlated noise is achromatic, multiple wave bands will not improve upon the monochromatic case. For the remainder of this chapter we will assume that there is in fact a wavelength dependence to the correlated noise which shares a common time dependence. While in the general case a time delay may be included in these models, the common time dependence is a requirement of the fast GP methods derived here. We exploit the information contained in this time dependence to demonstrate an improvement on the inference of transiting planet parameters. We also ignore instrumental/systematic variations, and assume that the white noise is dominated by Poisson photon counting uncertainty.

In the next section we describe our wavelength-dependent stellar variability model (§3.3). We then review the one-dimensional version of `celerite` before describing our extension to two dimensions (§3.4). Next, we conduct an Information analysis to derive approximate, semi-analytic upper bounds on the precision that can be achieved when inferring transit parameters from multiband light curves with different noise properties, and compare the results of our Information analysis to a full MCMC treatment for select noise parameters (§3.5). In the discussion (§3.6) we outline additional applications of our method including exomoon detection and transmission spectroscopy. We conclude with a discussion of the limitations of and potential improvements to our method (§3.7).

3.3 Multiwavelength noise model

Here we describe our model for noise that is correlated across both time and wavelength. We start with a description of the time-dependence of the noise.

3.3.1 Time-correlated variability model

Foreman-Mackey et al. (2017) describe how `celerite` can be used as a physically-motivated model for stellar variability. The following discussion is closely based on the discussion in that paper.

We follow Anderson & Jefferies (1990) in modeling stellar oscillations as the result of stochastic excitations that are damped by convection and turbulent viscosity in the star. This process is described by the differential equation

$$\frac{1}{\omega_0^2}\frac{\mathrm{d}^2}{\mathrm{d}t^2}y(t) + \frac{1}{\omega_0 Q}\frac{\mathrm{d}}{\mathrm{d}t}y(t) + y(t) = \epsilon(t) \tag{3.1}$$

where ω_0 is the characteristic frequency of the oscillator, Q is the quality factor of the oscillator, $\epsilon(t)$ is a stochastic driving force, and $y(t)$ is the amplitude of the oscillations. If $\epsilon(t)$ is Gaussian distributed then the solution to Equation 3.1 is a Gaussian process with the power spectral density

$$S(\omega) = \sqrt{\frac{2}{\pi}}\frac{S_0\omega_0^4}{(\omega^2-\omega_0^2)^2 + \omega_0^2\omega^2/Q^2}. \tag{3.2}$$

Figure 3.1 shows this power spectrum for several values of Q. For our modeling we set $Q = 1/\sqrt{2}$, in which case the power spectral density simplifies to

$$S(\omega) = \sqrt{\frac{2}{\pi}}\frac{S_0}{(\omega/\omega_0)^4 + 1}. \tag{3.3}$$

This power spectrum has been used to describe granulation-driven stellar variability (Kallinger et al. 2014). We employ this kernel in our work both because stellar granulation is a significant source of noise on transit timescales and to simplify our discussion of relevent noise

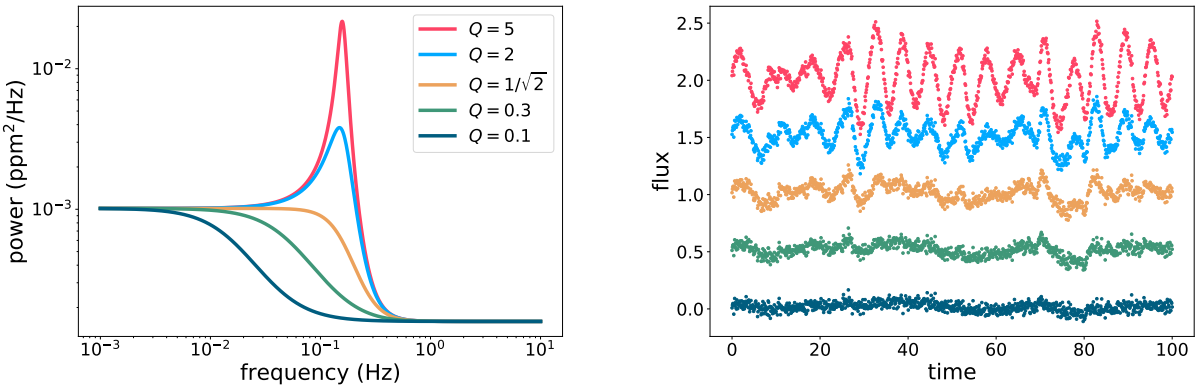

Figure 3.1 **Left:** Power spectrum of the SHO kernel for several values of the quality factor Q. For $Q < 1/\sqrt{2}$ the system is overdamped. For $Q > 1/\sqrt{2}$ the system is underdamped and the GP shows oscillations at the characteristic frequency. For our simulations we set $Q = 1/\sqrt{2}$, in which case the system is critically damped. **Right:** Noise realizations for each power spectrum on the left. Note the decreasing coherency of the oscillations as we move from high to low values of Q. The decreasing noise amplitudes from top to bottom are a result of the fact that the GPs with larger Q values have more total power at constant S_0.

timescales in section 3.5.1 which would be complicated by the presence of oscillations at the characteristic frequency. Choosing different values for Q will not affect the qualitative aspects of our results.

The corresponding kernel function to equation 3.3 is

$$k(\tau) = S_0 \omega_0 e^{-\omega_0 \tau/\sqrt{2}} \cos\left(\frac{\omega_0 \tau}{\sqrt{2}} - \frac{\pi}{4}\right), \tag{3.4}$$

where $\tau = |t_i - t_j|$.

3.3.2 Wavelength dependence of variability

We are now interested in constructing a simple model for the wavelength dependence of stellar variability based upon our time-dependent correlated variability model. To begin, we consider a two-component photosphere where each component has a unique spectrum and covering fraction. The star's variability is then a result of variations in the covering fraction

of these components, and the covering fractions vary according to the stochastic process described in Section 3.3.1.

We label the two components "hot" and "cold." Their spectra are given by $S_{\rm h}(\lambda)$ and $S_{\rm c}(\lambda)$ and their covering fractions are given by $x_{\rm h}$ and $x_{\rm c} = 1 - x_{\rm h}$. In the absence of limb-darkening the flux observed in a band B_1 is given by

$$F_{B_1} = \frac{\pi R_*^2}{d^2} \int (x_{\rm c} S_{\rm c}(\lambda) + x_{\rm h} S_{\rm h}(\lambda)) \mathcal{R}_{B_1}(\lambda) {\rm d}\lambda \tag{3.5}$$

where $\mathcal{R}_{B_1}(\lambda)$ is the response curve for the filter and the integral is taken over all wavelengths, d is the distance from the observer to the star, and R_* is the stellar radius. Substituting $x_{\rm h} = 1 - x_{\rm c}$ allows us to rewrite this expression as

$$\begin{aligned} F_{B_1} &= \frac{\pi R_*^2}{d^2} \left(\int S_{\rm h}(\lambda) \mathcal{R}_{B_1}(\lambda) {\rm d}\lambda \right) \\ &- \frac{\pi R_*^2}{d^2} x_{\rm c} \left(\int (S_{\rm h}(\lambda) - S_{\rm c}(\lambda)) \mathcal{R}_{B_1}(\lambda) {\rm d}\lambda \right). \end{aligned} \tag{3.6}$$

The first term of Equation 3.6 is the total flux for a photosphere completely covered by the hot component, and the second term is a correction dependent on the contrast between the hot and cold components. For simplicity, we define:

$$F_{\rm B_1,hot} = \frac{\pi R_*^2}{d^2} \int S_{\rm h}(\lambda) \mathcal{R}_{B_1}(\lambda) {\rm d}\lambda \tag{3.7}$$

and

$$\alpha_1 = \frac{\pi R_*^2}{d^2} \sigma_c \int (S_{\rm h}(\lambda) - S_{\rm c}(\lambda)) \mathcal{R}_{B_1}(\lambda) {\rm d}\lambda, \tag{3.8}$$

where $\sigma_c^2 = {\rm var}(x_c)$. With these definitions we have

$$F_{B_1} = F_{\rm B_1,hot} - \frac{x_{\rm c}}{\sigma_c} \alpha_1. \tag{3.9}$$

We can do the same for a second hypothetical band B_2, giving us

$$F_{B_2} = F_{B_2,\mathrm{hot}} - \frac{x_\mathrm{c}}{\sigma_c}\alpha_2. \tag{3.10}$$

Since the only time-dependent quantity in Equations 3.9 and 3.10 is the covering fraction of the cold component x_c, we see that the flux in each band will vary coherently with the same power spectral density and the amplitude of the variability will be set by the contrast between the hot and cold components of the photosphere in each band.

The covariance between two bands can now be computed:

$$\begin{aligned} \mathrm{cov}(F_{B_1}, F_{B_2}) &= \sigma_c^{-2}\mathrm{cov}(x_c\alpha_1, x_c\alpha_2) \\ &= \alpha_1\alpha_2\mathrm{corr}(x_c, x_c). \end{aligned} \tag{3.11}$$

Now we let x_c be a function of time, $x_c(t)$, and assert that it is drawn from a one-dimensional Gaussian process evaluated at times t_i for $i = 1, \ldots, N$ (i.e. a correlated time-series) with a kernel which can be described with the `celerite` formalism. Then the full covariance matrix for the time and wavelength dimensions is given by the block matrix

$$K = \begin{bmatrix} \Sigma_1 + T_{1,1}R & T_{1,2}R & \ldots & T_{1,N}R \\ T_{2,1}R & \ddots & & \\ \vdots & & & \\ T_{N,1}R & & & \Sigma_N + T_{N,N}R \end{bmatrix}, \tag{3.12}$$

where

$$\Sigma_i = \begin{pmatrix} \sigma_{i,1}^2 & 0 \\ 0 & \sigma_{i,2}^2 \end{pmatrix} \tag{3.13}$$

is a diagonal matrix containing the white noise components for each band at time i; $T_{i,j} = \mathrm{corr}(x_c(t_i), x_c(t_j))$ is the time covariance matrix for the process described in Section 3.3.1

normalized by the variance of x_c; and R is the covariance matrix across bands, defined:

$$R = \begin{pmatrix} \alpha_1^2 & \alpha_1\alpha_2 \\ \alpha_2\alpha_1 & \alpha_2^2 \end{pmatrix}. \tag{3.14}$$

For M bands $B_1, B_2, \ldots B_M$ with amplitudes given by $\alpha_1, \alpha_2, \ldots \alpha_M$, R becomes

$$\begin{aligned} R &= \begin{pmatrix} \alpha_1^2 & \alpha_1\alpha_2 & \ldots & \alpha_1\alpha_M \\ \alpha_2\alpha_1 & \ddots & & \alpha_2\alpha_M \\ \vdots & & & \vdots \\ \alpha_M\alpha_1 & \alpha_M\alpha_2 & \ldots & \alpha_M^2 \end{pmatrix} \\ &= \boldsymbol{\alpha}\boldsymbol{\alpha}^T \end{aligned} \tag{3.15}$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_M)^T$. The covariance matrix can now be written

$$K = \Sigma + T \otimes R, \tag{3.16}$$

where Σ is the block matrix

$$\Sigma = \begin{bmatrix} \Sigma_1 & \ldots & 0 \\ 0 & \ddots & \\ \vdots & & \Sigma_N \end{bmatrix}, \tag{3.17}$$

and where $\otimes$ denotes the Kronecker product. The Kronecker product is defined for two matrices A and B with dimensions $N \times M$ and $P \times Q$, respectively, as the $NP \times MQ$ block matrix

$$A \otimes B = \begin{bmatrix} a_{1,1}\boldsymbol{B} & a_{2,1}\boldsymbol{B} & \ldots & a_{1,N}\boldsymbol{B} \\ a_{2,1}\boldsymbol{B} & a_{2,2}\boldsymbol{B} & & \\ \vdots & & \ddots & \\ a_{N,1}\boldsymbol{B} & & & a_{N,N}\boldsymbol{B} \end{bmatrix}. \tag{3.18}$$

An important consideration for constructing new GP covariance matrices is that a valid

covariance matrix must be positive-definite for all inputs (Rasmussen & Williams 2006). For a detailed discussion of the postive-definiteness of the 1D `celerite` kernel we refer the reader to Appendix A of Foreman-Mackey et al. (2017). Assuming that the covariance matrix T is positive-definite, the positive-definiteness of the full covariance $K = T \otimes R + \Sigma$ can be ascertained by considering the eigenvalues of the Kronecker product plus diagonal covariance matrix, which are uniformly positive if and only if the matrix is positive-definite. For now we state the conclusion that K is positive-definite if R is positive-definite, or if R is positive-semidefinite and Σ has all nonnegative entries along its diagonal. A proof is given in Appendix A.5. In the case that R is the outer product $\boldsymbol{\alpha}\boldsymbol{\alpha}^T$, R is positive-semidefinite and positive-definiteness is thus ensured as long as a nonzero white noise component is given for each datapoint.

In practice, providing too small of a white noise component when R is positive-semidefinite may result in numerical instabilities. For this reason we recommend that care be taken when applying this method to extremely high-precision data. For arbitrary definitions of R positive definiteness should be ensured on a case-by-case basis. In general it is sufficient to show that R is positive-definite, or that R is positive-semidefinite with a nonzero white noise amplitude provided for each datapoint.

When the number of bands, M, is small, this covariance matrix can be used to model multiband observations. We can also allow M to become arbitrarily large, in which case the resultant covariance matrix can be used to model spectral observations. Here each entry in $\boldsymbol{\alpha}$ would represent the amplitude of the correlated variability in one wavelength bin of the spectrum. The linear scaling of our method with respect to both the time and wavelength dimension makes it feasible to model high spectral resolution timeseries this way. We include additional discussion on the subject of modeling transmission spectra in section 3.6.

To validate this model of multiwavelength stellar variability, we compare with observed Solar variability in Figure 3.2. This figure shows a timeseries from the SOHO VIRGO three-channel sun photometer Frohlich et al. (1995). The SPM monitors the Sun's variability in three visible light wave bands at one-minute cadence, and each of these bands exhibits a

power spectrum which has the same shape, but with amplitude which increase from red (862 nm) to blue (402 nm) as shown in Sulis et al. (2020). Alongside the SOHO SPM data we show a Gaussian Process drawn from our two-dimensional celerite algorithm in which the amplitudes in each band have been scaled to match the SOHO SPM multiband data. The qualitative agreement between the observed and simulated data is remarkable, and indicates that our model contains the necessary properties to capture high-precision multiwavelength stellar variability.

The algorithm used to simulate multiwavelength stellar variability and to compute the likelihood model is described in §3.4.2. Our implementation of the multiband GP, which is based on the `celerite` GP method, achieves $\mathcal{O}(NMJ^2)$ scaling where N is the size of T corresponding to the length of the vector x_c and M is the number of bands and corresponds to the size of the vector $\boldsymbol{\alpha}$. Appendix A.1 introduces a more general form of the two-dimensional GP which scales as $\mathcal{O}(NJ^2M^3)$ for arbitrary covariance in the second dimension. The remaining component of the likelihood function is the mean model, which for this chapter we take to be a transit model, described next.

Similarly, Loper et al. (2020) recently derived a multivariate generalization of `celerite` with linear scaling for a class of covariance functions called Latent Exponentially Generated (LEG) kernels. These LEG kernel functions are presented for multivariate *outputs* instead of multivariate *inputs* as described here, but it should be possible to express the kernels described here as members of the LEG family. However, for our restricted application, the computational cost and scaling of our method is better, since LEG GPs will scale as $\mathcal{O}(NJ^2M^3)$, in the notation above.

3.3.3 Transit model

To simplify and sharpen our simulated light curves, we use a trapezoidal transit model (Carter et al. 2008); this is the mean model whose parameters we wish to infer. For all our simulations the out-of-transit flux is normalized to unity in order to reduce the number of parameters to be inferred, though we note that this would represent an additional free

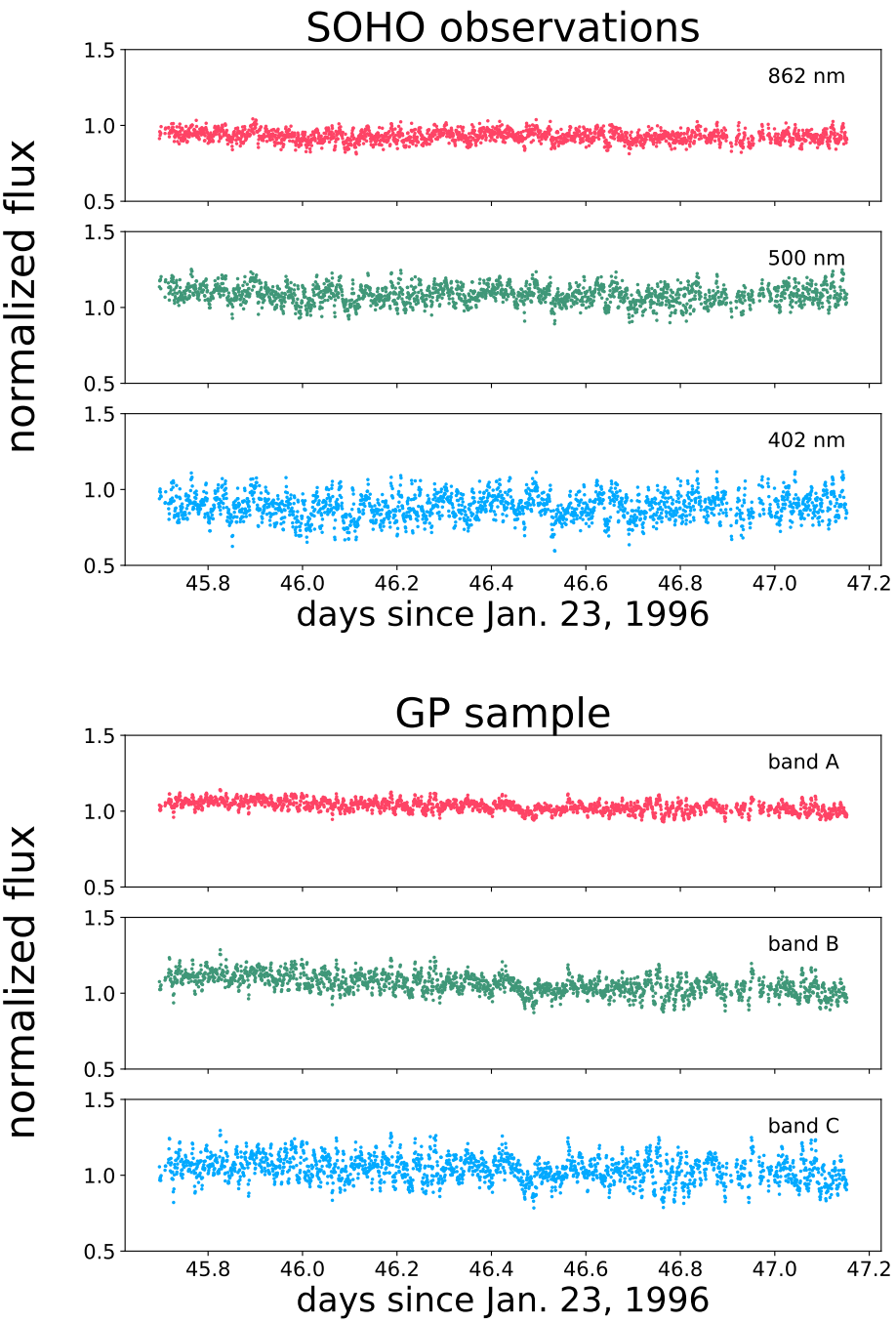

Figure 3.2 **Top:** SOHO three-channel sunphotometer timeseries of the Sun. **Bottom:** A three-band light curve simulated from a GP with a kernel consisting of three Kronecker-product terms (see equation 3.57), each term having the covariance described by equation 3.15. The GP hyperparameters were obtained by optimizing the GP likelihood with respect to the data in the top panel.

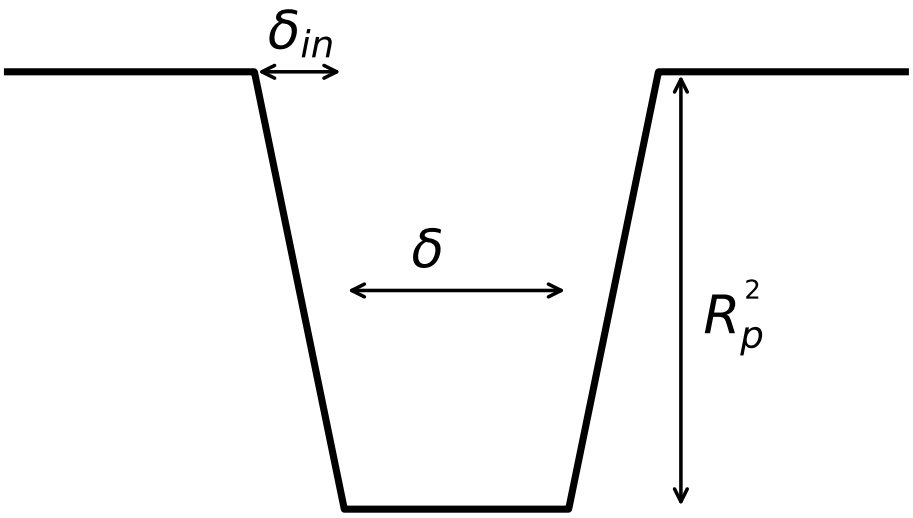

Figure 3.3 Schematic of the trapezoidal transit model. The center of transit t_0 is the midpoint of the transit.

parameter when modeling real observations. A schematic of this transit model is shown in Figure 3.3. For the purposes of this chapter, we ignore limb-darkening (which can have a wavelength dependence), and we ignore the slight curvature which occurs during ingress and egress. We also assume that the radius of the transiting planet is constant with respect to wavelength. This requirement can be relaxed to accommodate transmission spectroscopy, which we discuss in section 3.6.

The model is described by the function $\mu_{\rm trap}(t, \boldsymbol{\theta})$ with $\boldsymbol{\theta} = (R_p, t_0, \delta, \delta_{\rm in})$ where R_p is the planet's radius in units of the star's radius, t_0 is the time at center of transit, δ is the transit duration, and $\delta_{\rm in}$ is the ingress/egress duration. Note that we set the normalization of this model to one under the assumption that the out-of-transit data will be sufficiently lengthy to constrain the unocculted stellar flux.

With the noise and mean models specified, we next describe our simulated data.

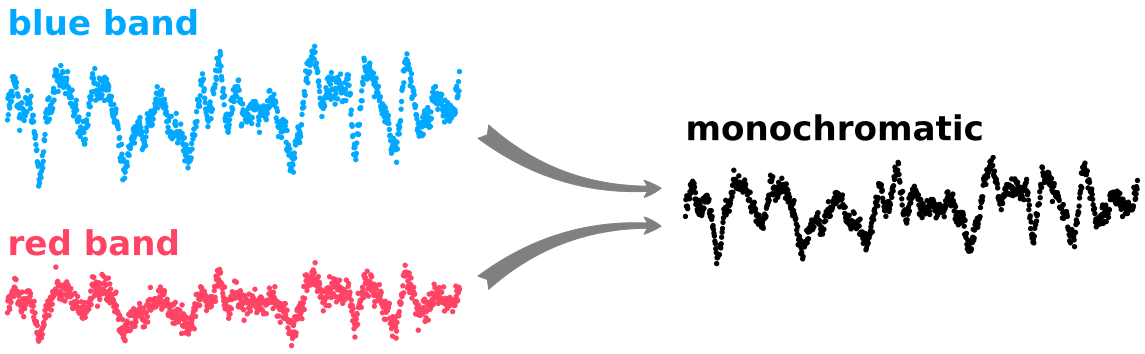

Figure 3.4 Two bands from a multiband simulation are combined to simulate a monochromatic light curve with the same noise realization. Note that the white noise amplitude is smaller in the monochromatic light curve than for either individual band, while the amplitude of the correlated noise is the photon-weighted mean of the amplitude in the two bands. Here the blue band has a correlated noise amplitude twice that of the red band.

3.3.4 Simulations

We simulate a suite of multiband light curves and construct a parallel set of monochromatic light curves by summing the flux between the bands of our multiband light curves. Figure 3.4 shows schematically how we produce a monochromatic light curve from the simulated multiband light curve. We compute the Information matrix (see 3.3.5) and run MCMC analysis on each light curve using our multiband GP model. The Information matrix tells us the theoretical lower limit for the uncertainty of each parameter, while the MCMC analysis gives us an estimate of the uncertainty on the parameters.

We split our simulations into three noise regimes based on the ratio between the characteristic variability timescale, and the ingress/egress and total duration of the transit. The characteristic variability timescale is given by $2\pi\omega_0^{-1}$ where ω_0 is the characteristic frequency of the variability appearing in equation 3.2. We define the three regimes as follows:

- Regime I: $1/f_0 > \delta$
- Regime II: $\delta_{\rm in} < 1/f_0 < \delta$

- Regime III: $1/f_0 < \delta_{\rm in}$

where $f_0 = \omega_0/(2\pi)$ is the characteristic frequency of the variability. Figure 3.5 contains representative light curves from each regime, chosen where the white and correlated noise amplitudes are comparable. In regime I the transit signal is distinguishable from the noise by its duration — all of the power in the correlated variability is on longer timescales than the transit duration. In regime II the characteristic timescale of the noise is smaller than the transit duration, but longer than the ingress/egress timescale. The transit still stands out from the noise because the transition into and out of transit is sharper than is characteristic for the SHO variability. In regime III the variability timescale is shorter than all of the relevant transit durations. We can see from Figures 3.1 and 3.6 that the SHO power spectrum allocates equal power to all oscillations on timescales longer than the characteristic timescale. The transit durations are thus swamped by correlated noise. As a result in the monochromatic case it is difficult to differentiate between the transit signal and noise, both by eye and with the GP. Fortunately the multiband GP is able to make use of additional information in the correlation between bands to disentangle the transit signal from the variability.

Amongst all of our simulations we hold constant the total noise, $\bar{\alpha}^2 + \sigma^2$ where $\bar{\alpha}^2$ is the weighted variance of the correlated noise over all bands and σ^2 is the variance of the white noise summed over all bands. We then vary the ratio between the noise amplitudes in order to analyze the simulations as a function of $\bar{\alpha}/\sigma$. For the multiband simulations, $\bar{\alpha}$ is the weighted mean of the amplitudes of variability in the individual bands, given by α_i. For all of our simulations, unless otherwise specified, we use a two-band model with $\alpha_2 = 2\alpha_1$ to represent the multiband case.

We hold the transit duration and ingress/egress duration constant so that the value of ω_0 changes to determine which noise regime we fall under.

Into all of our simulations we inject a transit signal with a fractional depth of 1% of the star's flux. We use a transit duration of 12 hours in the middle of a 10 day baseline. The ingress/egress duration is set to 1.2 hours.

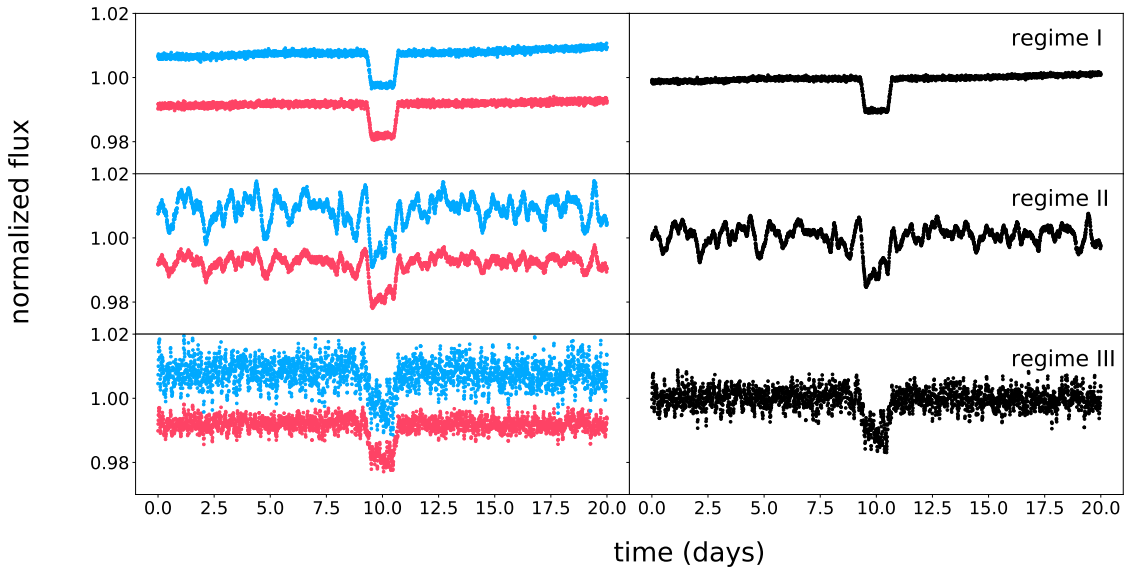

Figure 3.5 Representative light curves for the three noise regimes. The left panels show the two bands separately and the right panels show the monochromatic light curve resulting from the summation of the two bands. **Top:** In regime I the variability timescale is much longer than the transit duration. **Middle:** In regime II the variability timescale is between the transit duration and ingress/egress duration. **Bottom:** In regime II the variability timescale is shorter than the ingress/egress duration. Figure 3.6 shows power spectra corresponding to each of these regimes (but not to the light curves pictured here).

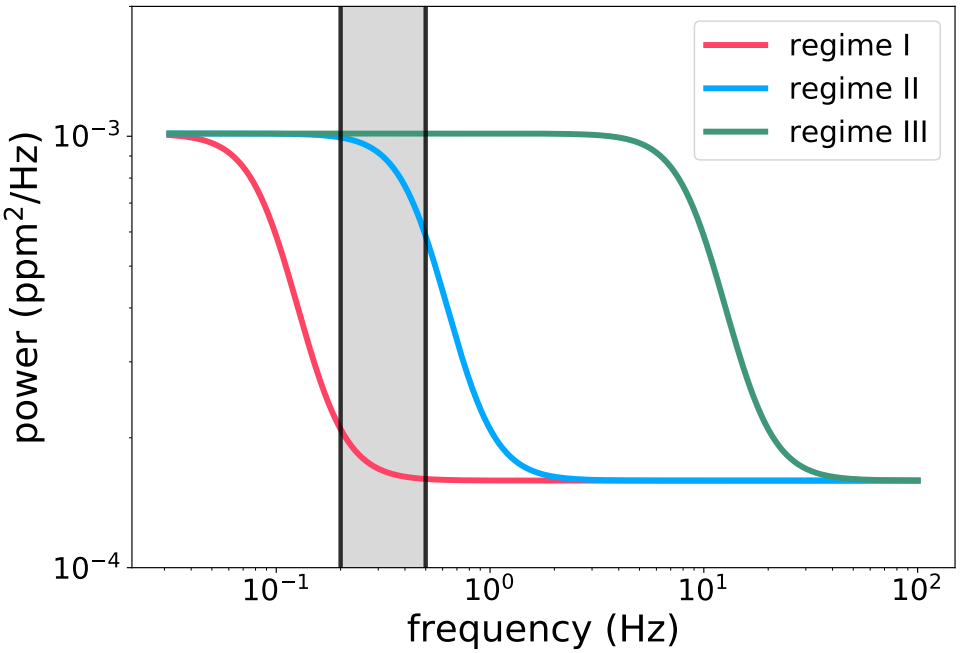

Figure 3.6 Power spectral densities for the three regimes. The shaded region spans from the inverse transit duration on the left to the inverse ingress/egress duration on the right. Note that the densities plotted here are only meant to be illustrative, and do not correspond to the power spectra of the light curves in figure 3.5

3.3.5 Information matrix analysis

The Information matrix encodes the amount of information about a signal that can be determined from observations taken in the presence of noise with a given covariance. For a model made up of a mean function $\boldsymbol{\mu_\theta}$ with N_θ parameters $\theta_1, \theta_2, \ldots \theta_{N_\theta}$ obscured by noise drawn from a multivariate Gaussian with covariance K, the Information matrix is the $N_\theta \times N_\theta$ matrix with entries given by

$$[\mathcal{I}_{\boldsymbol{\theta}}]_{i,j} = \left(\frac{d\boldsymbol{\mu}}{d\theta_i}\right)^{\mathrm{T}} K^{-1} \left(\frac{d\boldsymbol{\mu}}{d\theta_j}\right). \tag{3.19}$$

The covariance between parameters of the mean are then approximated by

$$\left[\mathcal{I}_{\boldsymbol{\theta}}^{-1}\right]_{i,j} \approx \mathrm{cov}(\theta_i, \theta_j). \tag{3.20}$$

This approximation represents a lower limit on the covariance that can be estimated in practice via methods such as MCMC simulation. It is valid in the limit that the posterior probability is a multi-dimensional Gaussian distribution near the maximum likelihood solution. This corresponds to the limit in which a signal may be approximated as linear with respect to its parameters, known as the linear signal approximation (LSA). Vallisneri (2008) shows that in order for LSA to apply we must be in the high signal-to-noise limit. Accordingly, the following analysis should be taken to apply only to a transit with a depth much larger than both the correlated and white noise components of the noise. While the approximation may continue to be accurate for smaller signal-to-noise, a full quantification of the uncertainty in the low SNR limit should rely on sampling the posterior directly via MCMC analysis.

We compute the Information matrix for the transit parameters assuming that the hyperparameters of the GP are known exactly. In practice the GP hyperparameters will be unknown, and should be fit simultaneously with the transit parameters. Our results thus represent a scenario in which there are sufficient out-of-transit observations to determine the

covariance of the noise to arbitrary precision.

We adopt a semi-analytic approach to computing the Information matrix by using exact derivatives of the trapezoidal transit model and using `celerite` to compute products of the inverted covariance matrix with the transit model's derivatives. This approach is necessary because the covariance matrix for our GP model cannot be inverted analytically except in special cases.

3.3.6 Analytical Estimates For Parameter Uncertainties

The Information matrix approach can yield analytic results for the depth uncertainty in the limit that limb-darkening is ignored, the ingress/egress duration is short, $\delta_{in} \approx 0$, and all other parameters are assumed to have no uncertainty. In particular we make the approximation that the out-of-transit flux is measured to high precision from extensive monitoring. In this limit the transit model has a derivative of

$$\frac{\partial \boldsymbol{\mu}_{\rm trap}}{\partial R_p^2} = \begin{cases} 0 & {\rm out-of-transit} \\ -1 & {\rm in-transit} \end{cases}, \tag{3.21}$$

where R_p^2 is the depth of the transit. If we assume that the transit duration matches exactly a single observation cadence then the covariance matrix may be written in the two-band case as:

$$K = \begin{pmatrix} \sigma_1^2 + \alpha_1^2 & \alpha_1 \alpha_2 \\ \alpha_1 \alpha_2 & \sigma_2^2 + \alpha_2^2 \end{pmatrix}, \tag{3.22}$$

where $\sigma_{1,2}$ are the white noise components on the timescale of the transit and $\alpha_{1,2}$ are the correlated noise amplitudes on the timescale of the transit in the two bands.

For this covariance matrix, the Information matrix gives an uncertainty on the depth of

the transit, $\sigma_{R_p^2}$, of

$$\sigma^2_{R_p^2,poly} = \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right)^{-1} \left(\frac{1 + \left(\frac{\alpha_1}{\sigma_1}\right)^2 + \left(\frac{\alpha_2}{\sigma_2}\right)^2}{1 + \frac{(\alpha_1 - \alpha_2)^2}{\sigma_1^2 + \sigma_2^2}}\right). \tag{3.23}$$

Note that the prefactor equals the noise in the limit of no correlated noise component ($\alpha_1 = \alpha_2 = 0$).

In the monochromatic case we can compute the uncertainty assuming that the noise is Poisson, in which case the mean amplitude of correlated noise is given by

$$\sigma^2_{R_p^2,mono} = \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right)^{-1} + \bar{\alpha}^2, \tag{3.24}$$

where we have assumed the noise to be Poisson and $\bar{\alpha}$ is defined to be the weighted mean amplitude of the correlated noise in both bands, given by

$$\bar{\alpha} = \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right)^{-1} \left(\frac{\alpha_1}{\sigma_1^2} + \frac{\alpha_2}{\sigma_2^2}\right). \tag{3.25}$$

The relations for the polychromatic and monochromatic cases are plotted in Figure 3.7 in the case $\alpha_2 = 2\alpha_1$ and $\sigma_1 = \sigma_2 = \sigma$ in which the sum of the white noise and correlated noise is held fixed. Compare with Figure 3.12 to see the similarity of this analytic approximation with the Information matrix results for the full trapezoidal model.

We can generalize these expressions for the depth uncertainty in the monochromatic and two-band case to an arbitrary number of bands when the white noise is identical for each band (i.e. $\sigma_i = \sigma$ for M bands indexed by i). In this case the uncertainties are given by

$$\frac{\sigma^2_{R_p^2,M,poly}}{\sigma^2} = \frac{1 + \sigma^{-2}\sum_{i=1}^{M}\alpha_i^2}{M\left(1 + \sigma^{-2}\sum_{i=1}^{M}\alpha_i^2\right) - \left(\sigma^{-1}\sum_{i=1}^{M}\alpha_i\right)^2}, \tag{3.26}$$

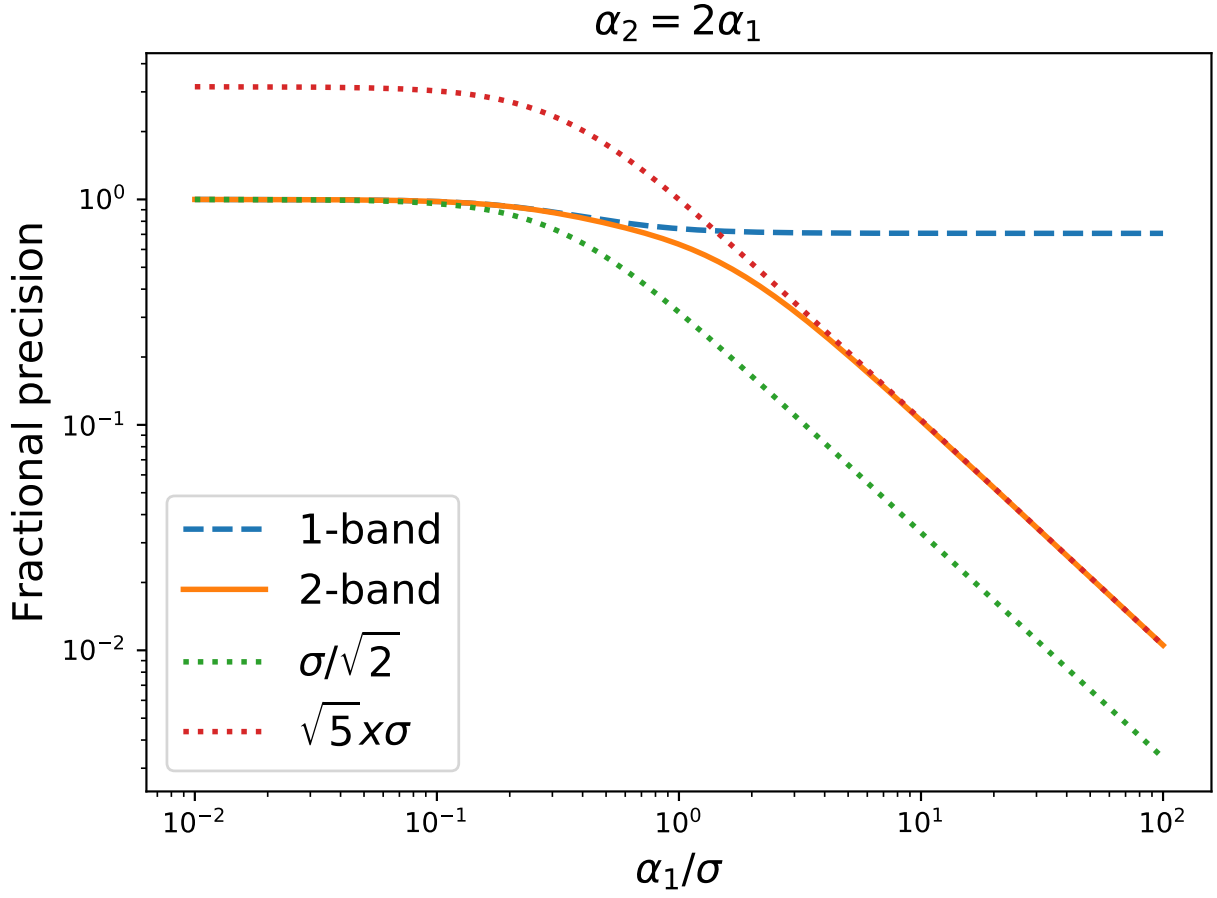

Figure 3.7 Analytic approximation for the fractional uncertainty on depth for two bands versus the ratio of the correlated noise to white noise in first band, α_1/σ, in the limit of a constant amplitude of the sum of correlated and white noise (so that the white noise declines as the correlated noise increases). The ratio of the correlated noise in the two bands is two, i.e. $\alpha_2 = 2\alpha_1$. Plotted are the single-band case (blue dashed), two-band case (orange solid), and the white noise in each band, σ, times $1/\sqrt{2}$ and $\sqrt{5}$ (dotted). The fractional precision is normalized to the case $\alpha_1 = 0$.

for the M-band case, and

$$\frac{\sigma^2_{R_p^2,M,mono}}{\sigma^2} = \frac{1}{M} + \left(\frac{1}{M\sigma}\sum_{i=1}^{M}\alpha_i\right)^2. \tag{3.27}$$

for the corresponding monochromatic case. Similar expressions may likely be found for the other transit parameters as well as for non-uniform noise in M bands, which we leave to future work.

While the uncertainties predicted by these equations differ from those found by a full Information matrix analysis of the trapezoidal transit, we find that they correctly predict the relationship between the monochromatic and multiband uncertainties in the limits $\alpha \gg \sigma$ and $\alpha \ll \sigma$ not only for the depth, but for the other parameters of the trapezoidal transit as well.

This is illustrated by Figure 3.12 which shows the Information uncertainties for each parameter of the trapezoidal transit model in the presence of correlated noise. We use a two band noise model with $\alpha_2 = 2{*}\alpha_1$. When the white noise dominates over the correlated noise ($\sigma \gg \alpha_{1,2}$), the Information uncertainties for the model with correlated noise is identical to a white noise-only model with the same white noise component, as we expect given that the correlated noise component is insignificant in this limit. We can use equation 3.26 to predict the Information matrix for the two band model in the limit that the correlated noise component dominates over the white noise component ($\sigma \ll \alpha_{1,2}$). Taking this limit equation 3.26 becomes

$$\lim_{\sigma \ll \alpha_{1,2}} \frac{\sigma^2_{R_p^2,M,\mathrm{poly}}}{\sigma^2} = \frac{\sum_{i=1}^{M}\alpha_i^2}{M\sum_{i=1}^{M}\alpha_i^2 - \left(\sum_{i=1}^{M}\alpha_i\right)^2}. \tag{3.28}$$

Setting $\alpha_1 = 1$ and $\alpha_2 = 2$, we find $\sigma_{R_p^2,M,\mathrm{poly}}/\sigma = \sqrt{10}$ which explains the scaling of the Information uncertainty at large α/σ in Figure 3.12.

We also examine the Information uncertainties as a function of number of bands. We consider a photon spectrum for which the variability increases from a value of $\alpha_{\min}$ to $\alpha_{\max}$. We assume that the photon spectrum variability is split into M bands with an equal photon

count rate in each band to give equivalent Poisson noise across all bands. In addition, we assume that α varies linearly with the photon count rate across all bands, so that the ith band has a correlated noise amplitude of $\alpha_i = \alpha_{\min} + (\alpha_{\max} - \alpha_{\min})(i - 1/2)/M$. For example, in the case of two bands with $\alpha_{\max}/\alpha_{\min} = 5$, we have $\alpha_2 = 2\alpha_1$, as in Figure 3.7. Figure 3.8 shows the uncertainty for the planet-star radius ratio as a function of the ratio between the minimum and maximum variability, $\alpha_{\max}/\alpha_{\min}$, for several values of M. The minimum achievable uncertainty as M approaches infinity and $\alpha_{\max} \gg \alpha_{\min}$, which can be arrived at by taking the appropriate limits of Equation 3.26 and transforming the sums into integrals as M approaches infinity. In these limits the minimum achievable uncertainty is twice the uncertainty for the white noise-only case, which is represented by the dashed line in Figure 3.8.

The same calculation may be performed for alternative spectra. For a blackbody spectrum we arrive at a limit of 2.2 times the white noise-only case when the number of bands and the contrast ratio is large. For arbitrary spectra the integrals can be computed numerically to yield the minimum achievable uncertainties for realistic stellar spectra and spot models.

The Information matrix and analytic approaches describe approximations to the parameter uncertainties. We next summarize our MCMC analysis to check and validate these approximations.

3.3.7 MCMC analysis

We use the `exoplanet` package Foreman-Mackey et al. (2019) which interfaces with PyMC3 to conduct our MCMC simulations. Each simulation is initialized with the true parameters. During MCMC we hold the GP hyperparameters constant as we did for the Information matrix analysis, and vary only the parameters of the trapezoidal transit model. We use PyMC3's implementation of No U-Turn sampling Hoffman & Gelman (2014), which requires the derivatives of the log likelihood to carry out the Hamiltonian markov chain integration. The NUTS sampler is initialized by tuning each simulation for 2000 steps. Subsequently, the simulation is run another 2000 steps to sample the posterior. This procedure results

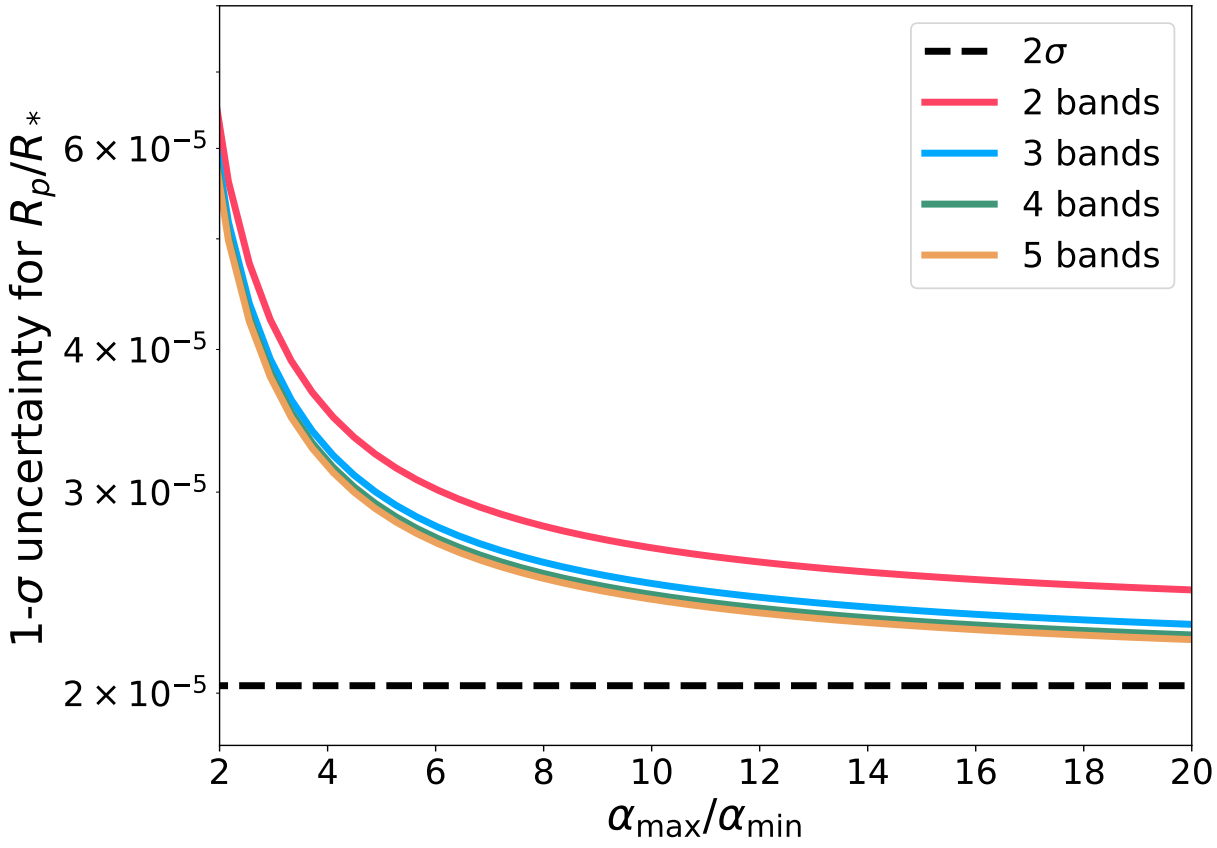

Figure 3.8 Information uncertainty curves for the planet/star radius ratio as a function of contrast ratio for a spectrum that increases linearly with photon flux from $\alpha_{\min}$ to $\alpha_{\max}$. We plot the Information uncertainty for different values of M, the number of bands into which the spectrum is binned for modeling. The dashed line is the minimum uncertainty achievable as the contrast ratio becomes infinite which, for the two-band case, is equal to 2σ where σ is the Information uncertainty in the absence of correlated noise.

in about 10^3 effective samples for each parameter of the model for each simulation as the autocorrelation length of the chains is extremely short (one of the advantages of using the NUTS sampler).

The final ingredient needed for our Information matrix and MCMC simulations involves our novel 2D version of `celerite`, which we describe next.

3.4 Implementation of the multiwavelength variability model

We implement our multiwavelength variability model as an extension of the `celerite` GP method to two dimensions. The `celerite` algorithm Foreman-Mackey et al. (2017) is a method for computing Gaussian processes in one dimension that scales as $\mathcal{O}(NJ^2)$ where N is the number of datapoints being modeled and J is the number of terms used to represent the covariance matrix. While one-dimensional GPs are suitable for a wide range of applications, there are many problems for which we need to model covariance between datapoints in two or more dimensions. Here we describe a method for computing a two dimensional GP when the covariance in the second dimension can be written as the outer product of a vector with itself. This covariance matrix is relevant to the common task of modeling time-variable spectra, as in our multiband transit model application. Our method is scalable, with computational time increasing linearly with the number of datapoints. In this section we introduce the method and revisit the `celerite` algorithm for Cholesky decomposition of the covariance matrix as it applies to a two-dimensional dataset. In appendices A.2, A.3, and A.4 we discuss the algorithms for computing the likelihood, predicting or extrapolating from the GP, and sampling from the GP.

For problems where the covariance cannot be modeled as an outer product we offer a more general extension of `celerite` where the covariance matrix for the second dimension can be arbitrary. We discuss our implementation of the arbitrary covariance method in Appendix A.1.

3.4.1 The one-dimensional `celerite` *method*

Until the past few decades, the adoption of GP methods was limited by computational expense. As a reminder, the log-likelihood function for a GP model for a series of N flux measurements, $\boldsymbol{y} = (y_1, y_2, \ldots y_N)$, so that the total number of datapoints $N' = N$, taken at times $\boldsymbol{t} = (t_1, t_2, \ldots t_N)$ is given by

$$\ln\ \mathcal{L} = -\frac{1}{2}(\boldsymbol{y} - \boldsymbol{\mu})^{\mathrm{T}} K^{-1} (\boldsymbol{y} - \boldsymbol{\mu}) - \frac{1}{2} \ln\ \det(K) - \frac{N}{2} \ln(2\pi) \tag{3.29}$$

where $\boldsymbol{\mu} = (\mu(t_1), \mu(t_2), \ldots \mu(t_N))$ and K is the covariance matrix of the GP. This equation involves the inverse and determinant of the $N \times N$ matrix K. In general, computing the inverse and determinant of an $N \times N$ matrix requires $\mathcal{O}(N^3)$ operations. Thus computing the likelihood for a GP by directly inverting K becomes prohibitively expensive for datasets larger than about 10^4 observations (Deisenroth & Ng 2015). This is especially true for applications that require repeated calls to the likelihood function as is the case for minimization and MCMC.

Because of this, much work has been done to reduce the complexity of GP computations. This can be accomplished primarily in two nonexclusive ways. The first is by employing inexact methods in which the full GP covariance matrix is approximated by a matrix for which the relevant matrix operations (primarily inversion and computation of the determinant) can be computed more efficiently than $\mathcal{O}(N^3)$ (Rasmussen & Williams 2006). Members of this class of methods include the HODLR factorization method of Ambikasaran et al. (2015) which achieves $\mathcal{O}(n \log^2 n)$ scaling as well as various sparse GP methods (Almosallam et al. 2016; Csató & Opper 2002; Snelson & Ghahramani 2006).

The second is by restricting the user to covariance matrices of a specific form. These methods are often known as structure-exploiting methods since they take advantage of the properties of specially structure matrices (e.g. low-rank matrices, Toeplitz matrices, Kronecker-product matrices) to speed up Gaussian process operations (Wilson & Nickisch 2015; Zhang et al. 2005; Nickson et al. 2015).

The `celerite` algorithm is a fast, one-dimensional GP method which exploits the properties of semiseparable plus diagonal matrices to accelerate GP computations, achieving $\mathcal{O}(NJ^2)$ scaling where J is the number of `celerite` terms that make up the kernel function and N is the number of datapoints. For commonly-used kernel models the number of terms will be very small compared to N.

`celerite` works by representing the GP covariance matrix as the sum of a diagonal matrix and J semi-separable matrices. The Cholesky factorization of J semi-separable matrices plus a diagonal matrix can be computed in $\mathcal{O}(NJ^2)$ rather than the $\mathcal{O}(N^3/3)$ required for an ordinary matrix. Once the Cholesky factors are in hand, the inverse and determinant of the covariance matrix can be computed in $\mathcal{O}(NJ)$ and $\mathcal{O}(N)$ respectively. Here we briefly describe the `celerite` algorithm, referring the reader to Foreman-Mackey et al. (2017) for a more detailed exposition of the method.

Consider a one-dimensional Gaussian process evaluated at the coordinates

$$\boldsymbol{x} = \begin{pmatrix} t_1 & \cdots & t_N \end{pmatrix} \tag{3.30}$$

The `celerite` kernel is given by

$$\begin{aligned} k_{\boldsymbol{\beta}}(t_n, t_m) &= \sigma_n^2 \delta_{nm} + \sum_{j=1}^{J} \frac{1}{2} \left[(a_j + i b_j) e^{-(c_j + i d_j)\tau_{nm}} \right. \\ &+ \left. (a_j - i b_j) e^{-(c_j - i d_j)\tau_{nm}} \right] \end{aligned} \tag{3.31}$$

where $\boldsymbol{\beta} = (a_1 ... a_J, b_1 ... b_J, c_1 ... c_J, d_1 ... d_J)$, σ_n^2 is the variance of the Gaussian-distributed white noise, and $\tau_{nm} = |t_n - t_m|$ with $n, m \in 1, \ldots, N$. This kernel defines a `celerite` model with J terms.

For a kernel function of this form, the covariance matrix is a symmetric, semiseparable matrix with semiseparability rank $P = 2J$. A matrix of this type can be written in terms of

two generator matrices U and V, both of size $(N \times P)$, along with a diagonal matrix A:

$$K = A + \mathrm{tril}(UV^{\mathrm{T}}) + \mathrm{triu}(VU^{\mathrm{T}}), \tag{3.32}$$

where tril is the lower-triangular operator which, when applied to a square matrix, preserves the entries below the diagonal and replaces all entries on and above the diagonal with zeros. The triu operator does the same for the upper-triangular entries in the matrix. In the case of our covariance matrix, the generator matrices are specified by:

$$\begin{aligned} U_{n,2j-1} &= a_j e^{-c_j t_n} \cos(d_j t_n) + b_j e^{-c_j t_n} \sin(d_j t_n), \\ U_{n,2j} &= a_j e^{-c_j t_n} \sin(d_j t_n) - b_j e^{-c_j t_n} \cos(d_j t_n), \\ V_{m,2j-1} &= e^{c_j t_m} \cos(d_j t_m), \\ V_{m,2j} &= e^{c_j t_m} \sin(d_j t_m), \end{aligned} \tag{3.33}$$

and A is given by:

$$A_{n,n} = \sigma_n^2 + \sum_{j=1}^{J} a_j. \tag{3.34}$$

We will soon see that the Cholesky decomposition for this covariance matrix can be computed in $\mathcal{O}(NJ^2)$ operations, allowing for the fast evaluation of the GP likelihood function.

The kernel function implemented by `celerite` is versatile in that by choosing appropriate coefficients it can be made to approximate a wide range of other kernel functions. Furthermore, Loper et al. (2020) demonstrated that `celerite` kernels provide a complete basis for one dimensional stationary covariance functions, meaning that these methods can, in principle, be used to approximate *any* stationary kernel, though there might be issues with numerical precision and computational cost when a large number of terms are required for accuracy. This versatility is demonstrated qualitatively in Fig. 3.9 which shows approximations of several popular kernels achieved by carefully choosing $\{a_j\}$, $\{b_j\}$, $\{c_j\}$ and $\{d_j\}$. Since each of these kernels may be approximated well by a `celerite` kernel, the products and sums of these component kernels are also `celerite` kernels, meaning that complex kernels

can still be approximated within the `celerite` kernel formalism.

The Cholesky factorization of the covariance matrix K is given by

$$K = LDL^{\mathrm{T}} \tag{3.35}$$

where L is the lower-triangular Cholesky factor and D is a diagonal matrix. Foreman-Mackey et al. (2017) begin their derivation of the Cholesky factorization algorithm with the ansatz that L can be represented in terms of U and a new (at this point unknown) matrix W with the same dimensions as U, as

$$L = I + \mathrm{tril}(UW^{\mathrm{T}}). \tag{3.36}$$

Then W and D can be found via the recursion relations

$$\begin{aligned} S_{n,j,k} &= S_{n-1,j,k} + D_{n-1,n-1} W_{n-1,j} W_{n-1,k} \\ D_{n,n} &= A_{n,n} - \sum_{j=1}^{P} \sum_{k=1}^{P} U_{n,j} S_{n,j,k} U_{n,k} \\ W_{n,j} &= \frac{1}{D_{n,n}} \left[V_{n,j} - \sum_{k=1}^{P} U_{n,k} S_{n,j,k} \right], \end{aligned} \tag{3.37}$$

where $S_{1,j,k}$ is a matrix of zeros and P is both the rank of the semiseparable covariance matrix and the number of columns in U and V, here equal to $2J$. In the original `celerite` paper it was found that, in order to avoid numerical stability issues caused by the exponential factors in equation 3.33, it was necessary to redefine the generator matrices U and V and to define an additional matrix ϕ of the same dimensions as U and V. The generators become

$$\begin{aligned} \tilde{U}_{n,2j-1} &= a_j \cos(d_j t_n) + b_j \sin(d_j t_n) \\ \tilde{U}_{n,2j} &= a_j \sin(d_j t_n) - b_j \cos(d_j t_n) \\ \tilde{V}_{m,2j-1} &= \cos(d_j t_m) \\ \tilde{V}_{m,2j} &= \sin(d_j t_m). \end{aligned} \tag{3.38}$$

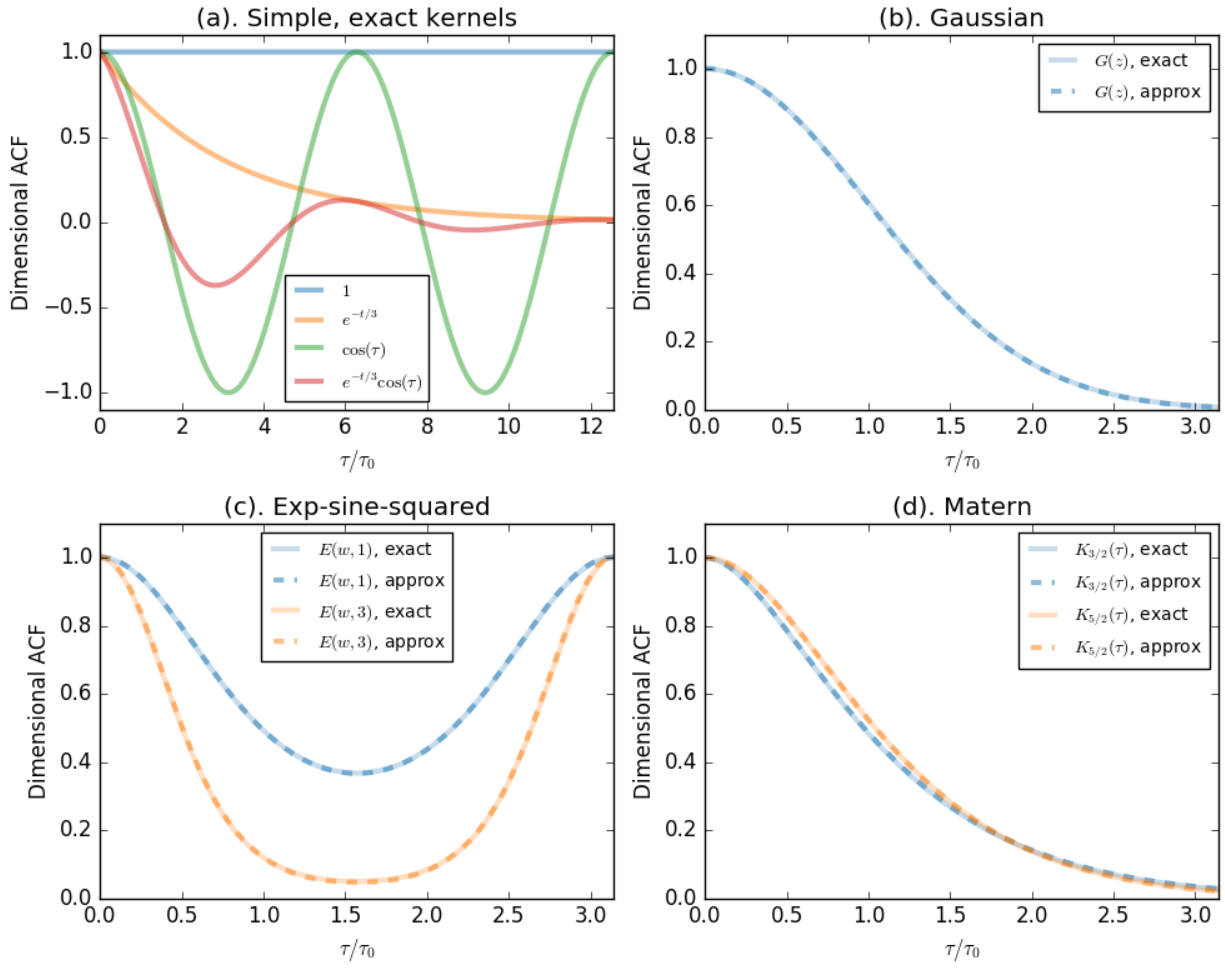

Figure 3.9 Approximation to various commonly used GP kernels (a) Simple kernels with an exact `celerite` representation: cosine, or exponential times cosine. (b) Approximation of a referred to as "exponential-squared" to distinguish it from sine-squared kernels. (d) Matern kernels.

The unknown matrix W becomes

$$\begin{aligned} \tilde{W}_{n,2j-1} &= e^{-c_j t_n} W_{n,2j-1} \\ \tilde{W}_{n,2j} &= e^{-c_j t_n} W_{n,2j}. \end{aligned} \tag{3.39}$$

And the new matrix ϕ is defined

$$\phi_{n,2j-1} = \phi_{n,2j} = e^{-c_j (t_n - t_{n-1})}. \tag{3.40}$$

The algorithm for decomposing the covariance matrix becomes

$$S_{n,j,k} = \phi_{n,j}\phi_{n,k}\left[S_{n-1,j,k} + D_{n-1,n-1}\tilde{W}_{n-1,j}\tilde{W}_{n-1,k}\right] \tag{3.41}$$

$$D_{n,n} = A_{n,n} - \sum_{j=1}^{P}\sum_{k=1}^{P}\tilde{U}_{n,j}S_{n,j,k}\tilde{U}_{n,k} \tag{3.42}$$

$$\tilde{W}_{n,j} = \frac{1}{D_{n,n}}\left[\tilde{V}_{n,j} - \sum_{k=1}^{P}\tilde{U}_{n,k}S_{n,j,k}\right]. \tag{3.43}$$

This completes our recap of the one-dimensional version of `celerite`; next we describe our novel 2D version.

3.4.2 Computing the two-dimensional GP

We now consider the Cholesky decomposition of the covariance matrix for a two-dimensional GP when the covariance in the second dimension can be written as the outer product of a vector with itself. This form of the covariance applies when the correlated component of the noise has the same shape along the first large dimension (of size N) and varies proportionally in amplitude along the second small dimension (of size M), as is the case for the multiwavelength stellar variability problem discussed above.

This covariance matrix is given by equation 3.16, reproduced here:

$$K = \Sigma + T \otimes R, \tag{3.44}$$

which has size $N' \times N' = NM \times NM$. Here Σ is a diagonal matrix containing the white noise components for each datapoint, which may be heteroscedastic, T is the covariance matrix in the first dimension, which must be defined by a `celerite` kernel, and R is the covariance matrix for the second dimension which must be an outer product of the form

$$R = \boldsymbol{\alpha}\boldsymbol{\alpha}^{\mathbf{T}}, \tag{3.45}$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_M)^{\mathrm{T}}$ is a vector of length M.

Writing K in terms of the celerite generator matrices from equation 3.32:

$$\begin{aligned} K &= \Sigma + \left[A_0 + \mathrm{tril}(UV^{\mathrm{T}}) + \mathrm{triu}(VU^{\mathrm{T}})\right] \otimes R \\ &= \Sigma + \mathrm{diag}(A_0 \otimes R) \\ &+ \mathrm{tril}(UV^{\mathrm{T}} \otimes R) + \mathrm{triu}(VU^{\mathrm{T}} \otimes R), \end{aligned} \tag{3.46}$$

where A_0 is the diagonal component of T obtained by setting $\sigma_n = 0$ for all $n \in 1, \ldots, N$ in equation 3.34. Substituting the outer product $\boldsymbol{\alpha}\boldsymbol{\alpha}^{\mathrm{T}}$ for R inside the upper and lower triangular operators we have

$$\begin{aligned} K = \Sigma &+ \mathrm{diag}(A_0 \otimes R) \\ &+ \mathrm{tril}(UV^{\mathrm{T}} \otimes \boldsymbol{\alpha}\boldsymbol{\alpha}^{\mathrm{T}}) \\ &+ \mathrm{triu}(VU^{\mathrm{T}} \otimes \boldsymbol{\alpha}\boldsymbol{\alpha}^{\mathrm{T}}). \end{aligned} \tag{3.47}$$

Applying the formula for mixed Kronecker and matrix products,

$$(AB) \otimes (CD) = (A \otimes C)(B \otimes D), \tag{3.48}$$

we can rewrite the covariance matrix as

$$\begin{aligned} K = \Sigma &+ \operatorname{diag}(A_0 \otimes R) \\ &+ \operatorname{tril}((U \otimes \boldsymbol{\alpha})(V \otimes \boldsymbol{\alpha})^{\mathrm{T}}) \\ &+ \operatorname{triu}((V \otimes \boldsymbol{\alpha})(U \otimes \boldsymbol{\alpha})^{\mathrm{T}}). \end{aligned} \tag{3.49}$$

We now see that the two-dimensional covariance matrix has exactly the same semi-separable structure as the one-dimensional covariance matrix with new definitions of the generator matrices in terms of their Kronecker products with $\boldsymbol{\alpha}$:

$$\begin{aligned} A' &= \Sigma + \operatorname{diag}(A_0 \otimes R) \\ U' &= U \otimes \boldsymbol{\alpha} \\ V' &= V \otimes \boldsymbol{\alpha} \end{aligned} \tag{3.50}$$

The components of the refactored generator matrices, corresponding to equation 3.38, are now given by

$$\begin{aligned} \tilde{U}'_{M(n-1)+p,2j-1} &= \alpha_p(a_j \cos(d_j t_n) + b_j \sin(d_j t_n)) \\ \tilde{U}'_{M(n-1)+p,2j} &= \alpha_p(a_j \sin(d_j t_n) - b_j \cos(d_j t_n)) \\ \tilde{V}'_{M(m-1)+p,2j-1} &= \alpha_p \cos(d_j t_m) \\ \tilde{V}'_{M(m-1)+p,2j} &= \alpha_p \sin(d_j t_m), \end{aligned} \tag{3.51}$$

and ϕ' is given by

$$\phi'_{M(n-1)+p,:} = \begin{cases} e^{-c_j(t_n - t_{n-1})} & p = 1 \\ 1 & p > 1 \end{cases}, \tag{3.52}$$

with $n, m \in 1, \ldots, N$, $p \in 1, \ldots, M$, and the colon indicating that the element is identical for every entry of that row. For these definitions of the generator matrices the recursive

Cholesky decomposition algorithm becomes

$$\begin{aligned} S_{n,j,k} &= \phi'_{n,j}\phi'_{n,k}\left[S_{n-1,j,k} + D_{n-1,n-1}\tilde{W}_{n-1,j}\tilde{W}_{n-1,k}\right], \\ D_{n,n} &= A'_{n,n} - \sum_{j=1}^{P}\sum_{k=1}^{P}\tilde{U}'_{n,j}S_{n,j,k}\tilde{U}'_{n,k}, \\ \tilde{W}_{n,j} &= \frac{1}{D_{n,n}}\left[\tilde{V}'_{n,j} - \sum_{k=1}^{P}\tilde{U}'_{n,k}S_{n,j,k}\right], \end{aligned} \tag{3.53}$$

where again P is the number of columns in U' and V'.

The recursive algorithm defined above requires one pass through each of the $N' = NM$ rows of $\tilde{U}'$ and $\tilde{V}'$. At each step we compute a double sum over the P columns of these matrices. The resultant scaling is thus $\mathcal{O}(NMP^2)$. For the outer-product definition of R we have $P = 2J$ and the method scales as $\mathcal{O}(NMJ^2)$ (see figure A.1 for benchmarks).

As shown in appendix A.1, we can come up with similar definitions of $\tilde{U}'$, $\tilde{V}'$, and ϕ' for arbitrarily defined R which yield $P = 2JM$, allowing us to compute the Cholesky decomposition in $\mathcal{O}(NJ^2M^3)$. Algorithms for computing the likelihood function, computing predictions or extrapolations from the GP, and sampling the GP are given in appendices A.2, A.3, and A.4 respectively for both outer-product and arbitrary definitions of R.

For this two-dimensional GP the set of observations used to compute the GP likelihood is also two-dimensional. Actual computation of the likelihood however requires that the input be reduced to one dimension. The Kronecker structure of the covariance matrix determines the form of the vector of observations. For input defined on a grid of size $\boldsymbol{t}\times\boldsymbol{r}$ where $\boldsymbol{t}$ represents the dimension along which the covariance is described by a celerite kernel and $\boldsymbol{r}$ represents the second small dimension, we have a 2D matrix of observations:

$$Y_{i,j} = y(r_i, t_j). \tag{3.54}$$

We define the observation vector to be

$$\boldsymbol{y} = \mathrm{vec}(Y), \tag{3.55}$$

where $\mathrm{vec}(Y)$ is the concatenation of the rows of Y. In other words,

$$\boldsymbol{y} = (Y_{:,1}, Y_{:,2}, \ldots Y_{:,N})\,. \tag{3.56}$$

With the description of our computational methods completed, we now turn to the results of transit simulations.

3.5 Results

We have carried out an analysis of simulated transit light curves with a wide range of noise amplitudes, timescales, and ratios of correlated to white noise, which we summarize the results of here. We start with a discussion of the results from a case study of seven examples with different ratios of correlated to white noise (§3.5.1), and then expand the discussion to a wider range of simulations for which we compare the Information matrix, analytic, and MCMC error analyses (§3.5.2).

3.5.1 Case studies

To start with, Figure 3.10 shows seven examples of our simulations for two bands with correlated noise amplitudes which differ by a ratio of two. These were made with moderate signal-to-noise and with $\omega_0\delta = 100$, which corresponds to a characteristic timescale of the correlated noise which is shorter than the transit duration and the ingress/egress timescales (regime III). In this case we held the white noise in the two bands to be identical in amplitude (corresponding to an identical photon count rate in both bands), and we compared a joint analysis of the two bands (we refer to this as "polychromatic") with an analysis of a single band consisting of the sum of the same simulated light curves from the two bands (this analy-

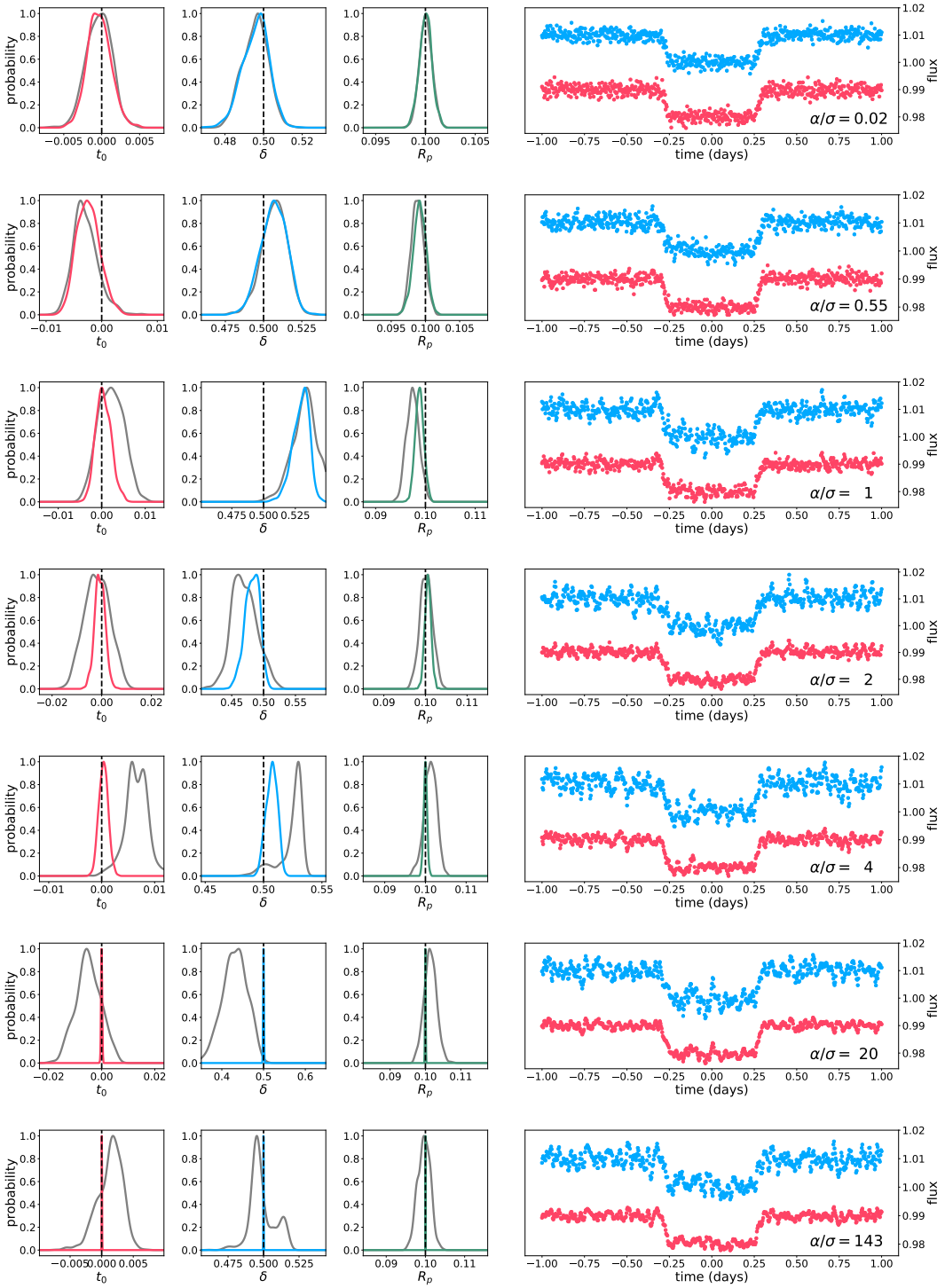

Figure 3.10 **Left:** Posteriors for three transit parameters estimated by MCMC analysis on the two band (colored) and single band (gray) data. Posteriors are smoothed using Gaussian kernel density estimation for $\omega_0\delta = 100$ (corresponding to the final panel of Figure 3.11). From left to right: the center of transit t_0, transit duration δ, and radius ratio R_p/R_*. For $\alpha/\sigma = 20$ and $\alpha/\sigma = 143$ the posterior distributions for the two-band case are too sharply peaked to be visible. **Right:** Representative light curves for each value of the noise amplitude ratio α/σ zoomed in on the transit signal (the input light curves have a duration of 10 days).

sis we refer to as "monochromatic"). Across these simulations we have varied the ratio of the total correlated noise to the white noise, α/σ, over seven values, $\{0.02, 0.55, 1, 2, 4, 20, 143\}$, to examine the precision of the two-band analysis compared with a monochromatic analysis.

For the first two simulations, $\alpha = 0.02\sigma$ and $\alpha = 0.55\sigma$, the variance of the correlated noise is smaller than that of the white noise. At this low ratio of α/σ we find that the measurement of the transit depth and timing parameters is about the same in the two-band case as in the monochromatic case (top panel, Figure 3.10). In the third panel where the white and correlated noise amplitudes are equal, we see a slight improvement in the measurement of the transit time and depth. In the remaining panels (bottom four panels of 3.10), we find an increasing degree of improvement in all the measured parameters as α increases relative to σ. As we approach the small white noise limit the improvement in all parameters between the single-band and two-band analyses is dramatic, with the transit depth improving by a factor of 18 at $\alpha = 20\sigma$ and by a factor of 118 at $\alpha = 143\sigma$. The transit time measurement improves by a factor of 21 and 65 respectively for these simulations. This improvement results from the ability to distinguish correlated noise variations from the transit signal when two bands are utilized, thanks to the different amplitudes of the correlated noise in the two bands; the correlated noise variations are measured to high precision in this case due to the small photon noise. Even so, the precision of the transit parameters is worse than it would be if there were no correlated noise by a factor of $\sqrt{10}$. This is an astrophysical limitation, and yet it still demonstrates a dramatic improvement in the analysis which splits the photons into two bands versus a single summed band.

The intermediate values of $\alpha/\sigma = \{1, 2, 4\}$ shown in Figure 3.10 have a behavior which is intermediate between the high white noise and low white noise limits that we discuss above: a monotonic improvement in all of the measurements with the increase in α/σ.

The general trends of these simulations hold over a broader range of parameters. To examine a larger number of cases, we summarize the uncertainties of the monochromatic cases and polychromatic cases based on the measurement precision as a function of the noise parameters, which amounts to measuring the breadth of the posterior distributions

inferred for each parameter (left-hand panels of Figure 3.10). We also compare these to the uncertainty estimates using the Information matrix approach and the analytic estimates given in §3.3.5 and §3.3.6, which we discuss next.

3.5.2 Noise comparison

We have carried out a much broader parameter study, varying the ratio of α/σ over a wide range of values for three values of the timescale: $\omega_0\delta = 0.1, 10$, and 100. We compare the Information matrix analysis against the MCMC anlysis in the monochromatic case with the two-band case, also with $\alpha_2 = 2\alpha_1$, in Figure 3.11. The MCMC uncertainty estimates agree closely with the Information uncertainty curves for almost all of our simulations, as demonstrated by Figure 3.11 for moderate signal-to-noise.

In regime I, $\omega_0\delta = 0.1$, in which the characteristic variability timescale is longer than the transit duration, the uncertainties on the transit parameters are nearly identical between the monochromatic and multiband simulations up to $\alpha/\sigma \approx 10$, where the multiband uncertainties begin to diverge slightly from the monochromatic uncertainties. Since the transit signal is distinguishable from the noise on the basis of its duration alone, the amount of additional information contained in the inter-band correlation is insignificant and both models perform similarly well.

We now skip to regime III, with $\omega_0\delta = 100$ (the same as the case studies in the prior subsection), in which the characteristic variability timescale is smaller than the transit duration. Because the SHO power spectrum allocates equal power to all oscillations on timescales longer than $1/\omega_0$, the transit signal is not distinguishable from the variability on the basis of its duration. In this case the inter-band correlation contains the additional information necessary to correctly infer transit parameters. Both models perform similarly when the correlated noise amplitude is small compared to the white noise, but when the correlated noise amplitude α begins to dominate over the white noise σ the monochromatic model does a poor job of inferring parameters (as evidenced by the large uncertainties) while the multiband model infers more and more precise values as the white noise decreases relative to the

correlated noise.

The results for regime II, here represented by $\omega_0\delta = 10$, fall intermediately between regimes I and III. In regime II, the characteristic timescale of the variability falls between the transit duration and the ingress/egress timescale so that measurements of the transit duration must contend with correlated noise on the same timescale, whereas measurements of the ingress and egress are affected primarily by white noise rather than correlated noise. Since the transit time is constrained by the ingress and egress times rather than by the transit duration, measurements of t_0 are also primarily affected by white noise. This is why we see significant improvement in the measurement of the transit depth at high α and low σ between the single-band and two-band simulations, while the timing parameters show much less improvement until we reach the low white noise limit. At this point the white noise amplitude is small enough compared to the correlated noise amplitude that the relatively low correlated noise on the timescale of the ingress/egress duration does begin to interfere with timing measurements in the single-band case.

In Figure 3.12 we plot Information uncertainty curves in regime III for the multiband model against Information uncertainty curves for the monochromatic model having the same transit parameters but with only a white noise component — the correlated noise amplitude is set to zero. The colored curves representing the Information uncertainties for the full noise model (white and correlated noise) match the white noise-only uncertainty in the limit that the correlated noise component is very small, as expected. As we increase the relative amplitude of the correlated noise component the uncertainty for the full model jumps from the white noise-only curve with the same white noise amplitude to the white noise-only curve with $\sqrt{10}$ times greater amplitude. As the correlated noise amplitude further increases, the Information uncertainty for the full model behaves as though we're doing inference on an equivalent model with the correlated noise component exchanged for a larger white noise amplitude.

The behavior seen here is explained by the analytical model outlined in Section 3.3.6. In particular, Equation 3.28 explains why the uncertainty scales as the white noise-only

uncertainty with $\sqrt{10}\sigma$ in the large correlated noise limit for two bands with amplitudes related by $\alpha_2 = 2\alpha_1$.

This completes our description of the simulated light curves and the results from these simulations. We next discuss the implications of these results.

3.6 Discussion

We have demonstrated the application of our method to the problem of fitting a transit observed in multiple bands in the presence of correlated noise. We now revisit and summarize the results of that demonstration before outlining some other potential applications of our method.

Monochromatic transit observations are ill-equipped to deal with correlated noise, as the wavelength-integrated flux does not provide enough information to distinguish between transits and noise features except when the correlated noise amplitude is low on the timescale of the transit duration. When transits occur on timescales similar to or longer than the variability timescale we must rely on the spectral dimension to provide the information necessary to distinguish between the two.

We use the Information matrix to explore the difference between inference on a monochromatic noise model and a multiband model with wavelength-dependent variability. We construct sets of monochromatic and multiband models with identical noise properties by splitting a given number of photons per wavelength into different spectral bins. We find that our results depend strongly on the timescale of the noise with respect to the transit duration. When the timescale of the correlated variability is much longer than the transit duration the monochromatic and multiband models perform similarly, though the multiband model still allows us to infer slightly more precise parameters in the limit that the correlated noise amplitude is much larger than the white noise amplitude (see Figure 3.11).

For the noise regime in which the correlated variability timescale is similar to or shorter than the transit duration we summarize our results as follows:

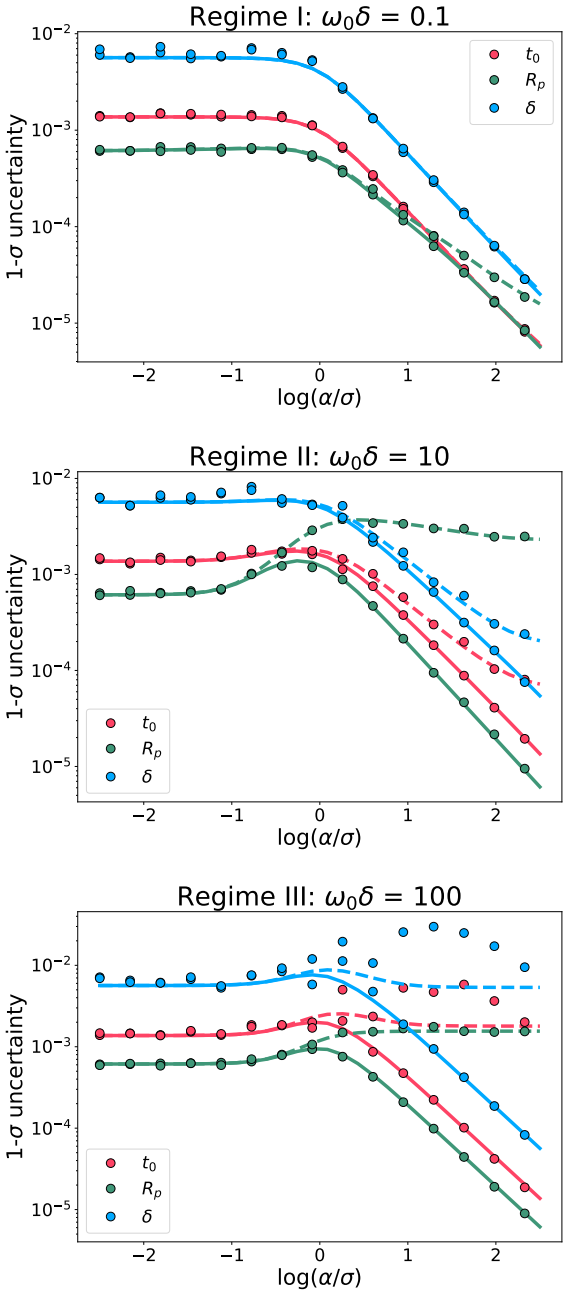

Figure 3.11 Information uncertainty curves overlayed with MCMC uncertainty estimates for trapezoidal transit parameters. Dashed lines show results for the monochromatic noise model and solid lines show results for the two-band noise model. Circles represent the MCMC uncertainty for distinct realizations of the noise and transit.

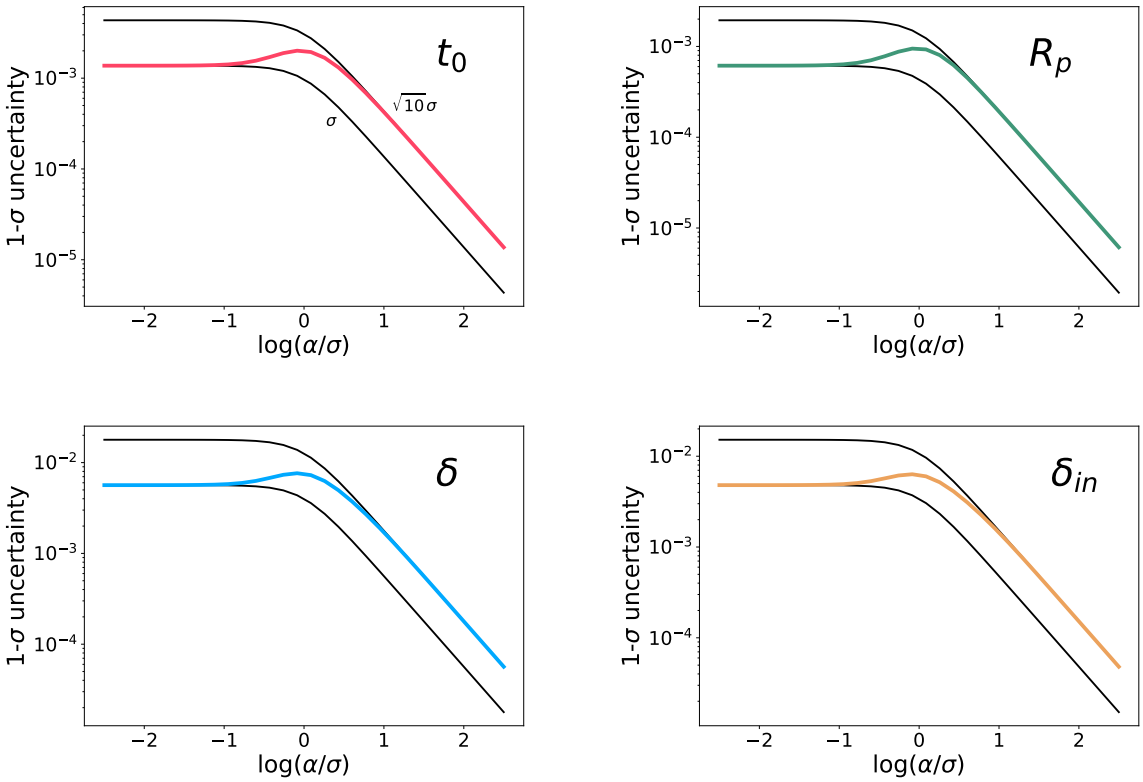

Figure 3.12 Information uncertainty curves (colored lines) for the two-band model compared to the white noise-only versions of the corresponding monochromatic noise model (black lines) in regime III. For the white noise-only models we set the correlated noise amplitude to zero and leave all other parameters the same as the monochromatic model. As we transition from the white noise dominated to the correlated noise dominated regimes the Information uncertainty curves for the two-band model transition from following the white noise model with $\sigma' = \sigma$ to the white noise model with $\sigma' = \sqrt{10}\sigma$. In effect perfect knowledge of the two-band correlated noise hyperparameters allows us to recover transit parameters at the same precision as if the correlated noise were simply white noise with a $\sqrt{10}$ larger amplitude.

- As the white noise amplitude decreases and the correlated noise amplitude increases, the precision inferred by the monochromatic noise model stays approximately constant, getting slightly worse for the radius ratio but improving slightly for the timing parameters δ and t_0. In contrast, the precision inferred by the multiband noise model improves as the white noise amplitude decreases even with increasing correlated noise amplitude. The increase in precision scales the same as if the correlated noise were held constant. The presence of correlated noise simply decreases the precision of the parameters by a constant factor which is related to the form of the variability as a function of wavelength.

- Most of the benefits of the multiband noise model can be realized by splitting the monochromatic variability into just two bands, but more bands achieve slightly better precision (see Figure 3.8)

- In the limit that we approach an infinitely high-resolution spectrum we can derive the factor by which the precision of the transit parameters is worse than the case where there is no correlated variability. Using equation 3.26 we find that the precision inferred in the presence of correlated noise is worse than in the white noise-only case by a factor of 2 when the variability amplitude scales linearly with cumulative photon counts with wavelength and 2.2 when the variability amplitude is distributed according the the blackbody distribution. In other words, in the presence of linearly scaling correlated variability amplitudes, we need four times as many photons to achieve the same precision in the presence of correlated noise as can be achieved when there is only white noise, provided we use a multiband noise model to do our inference.

3.6.1 Low Transit SNR Limit

The limit where the transit depth is small compared to the correlated noise amplitude is important if we are interested in detecting planets with small radii, or rocky planets around

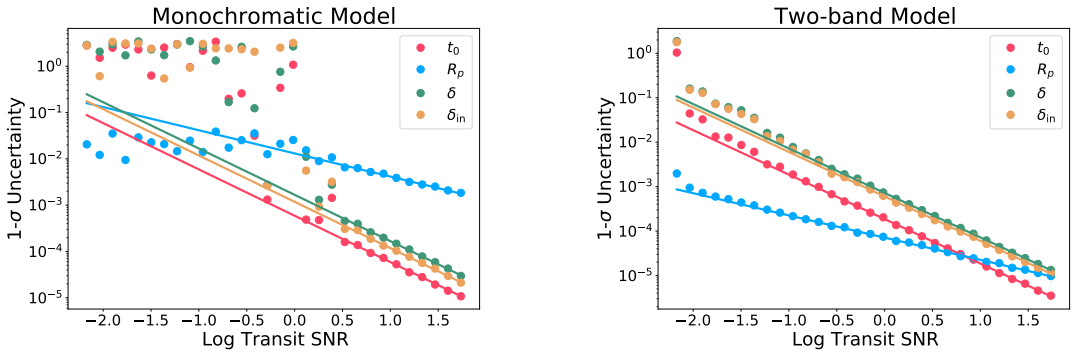

Figure 3.13 MCMC uncertainties (dots) and Information matrix uncertainties (lines) for monochromatic and two-band noise models as a function of the transit SNR with $\alpha_2 = 2\alpha_1$ for the two-band simulations. For these simulations the correlated noise is held constant at 150 times the amplitude of the white noise component and the total noise defined to be the sum in quadrature of the white noise and correlated noise amplitudes is conserved. The variability timescale $1/\omega_0 = \delta/10$, placing these simulations in regime II. For the monochromatic model, the Information and MCMC uncertainties correspond down to an SNR of about 10, which is the point at which the MCMC simulations no longer converge to the correct transit solution, as evidenced by the scatter in MCMC uncertainties at lower SNR. For the two-band simulations the Information and MCMC uncertainties correspond down to an SNR of 1/100.

sun-like stars. The Information matrix analysis above was done in the high SNR limit, because that is the limit in which the Information matrix can be shown to approximate the uncertainty on model parameters. We now include results on the correspondence between the Information matrix and MCMC uncertainties in the low SNR limit. Since we are primarily interested in the correlated noise component, we use SNR to refer to the ratio of the transit depth to the correlated noise amplitude.

Figure 3.13 shows the MCMC-derived uncertainties and the Information uncertainties for our four trapezoidal transit parameters in both the monochromatic and two-band cases. We use a correlated noise to white noise amplitude ratio (α/σ) of 150 for this portion of the analysis.

When we use a monochromatic model the Information uncertanties diverge from the MCMC uncertainties at a SNR of about 10. This corresponds to the point at which the MCMC uncertainties jump to very high values for the timing parameters, indicating that

the MCMC fails to converge to the correct solution.

This contrasts strongly with the two-band model. Using two bands the Information analysis finds the same uncertainty as the MCMC analysis down to an SNR of about 1/100, for which the ratio of the transit depth to the white noise is near unity.

In Figure 3.14 we repeat the analysis for $\alpha/\sigma = 10$. With a larger white noise component the MCMC uncertainties diverge from the Information uncertainties at a higher SNR. However, the two-band model still outperforms the monochromatic model with the MCMC corresponding to the Information and converging to the correct solution down to an SNR of about 1/10, where again the transit depth is comparable to the white noise.

These results imply that the improvement resulting from multiple bands applies only when the signal is larger than the white noise, and in this limit, the Information matrix provides an adequate estimate of the uncertainties on the model parameters, assuming that the Gaussian Process parameters are well constrained as these were not varied in our analysis. This approach may be used to estimate sensitivity and detection of transiting bodies, such as exomoons, discussed next.

3.6.2 Other applications

Exomoons, or moons of exoplanets, are are an oft-theorized but thus far undetected object of interest both for their ability to inform understandings of planetary formation and for their potential habitability. While one candidate exomoon, Kepler-1625b-i (Teachey & Kipping 2018), has been identified it remains unconfirmed (Teachey et al. 2019; Kreidberg et al. 2019). The saga of Kepler-1625bi illustrates one of the primary barriers to observing exomoons: their small size and correspondingly shallow transits. An additional complication is that exomoon transits will not be strictly periodic, due to orbital motion about their planets. This means that folding the light curve on the planet's orbital period to increase the signal-to-noise for a detection will not be effective.

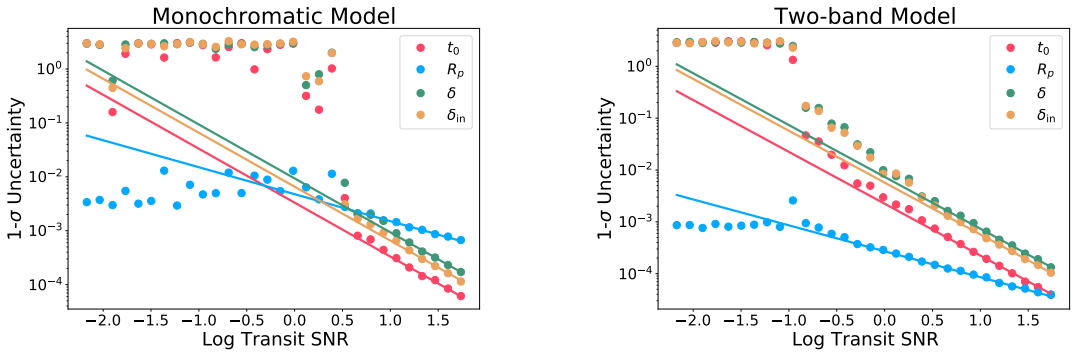

Figure 3.14 MCMC uncertainties (dots) and Information uncertainties (lines) for monochromatic and two-band noise models as a function of the transit SNR with wavelength dependence specified $\alpha_2 = 2\alpha_1$ for the two-band simulations. For these simulations the correlated noise is held constant at 10 times the amplitude of the white noise component and the total noise defined to be the sum in quadrature of the white noise and correlated noise amplitudes is conserved. The variability timescale $1/\omega = \delta/10$, placing these simulations in regime II. The larger white noise component compared to figure 3.13 pushes the SNR limit below which the MCMC and Information uncertainties diverge to higher SNR. As before, there is an abrupt transition at this limiting SNR where the MCMC suddenly fails to converge to the correct transit solution.

Observations designed for detecting transiting exomoons may likely need to consist of very high signal-to-noise photometry of more than one transit of a known exoplanet. In the near future JWST will be the observatory best suited to these observations (Beichman et al. 2014). It has the ability to observe timeseries spectra of bright objects via the NIRSpec instrument (Bagnasco et al. 2007). Our method is well-suited to model these observations and we believe it may end up being the optimal method of identifying an exomoon transit signal. Simulating JWST observations of transiting planet systems with realistic noise (**?**), while applying our multiwavelength GP model to the results, would reveal what sensitivity JWST would have to shallow transiting bodies such as exomoons.

Transit transmission spectroscopy aims to measure the transmission spectrum of an exoplanet by measuring the effective radius of the planet as a function of wavelength. This

is typically accomplished by varying the transit depth in the fit to the timeseries photometry at each wavelength as in Berta et al. (2012) and Mandell et al. (2013). In studies like these the effects of stellar variability have been minimal and largely ignored. However in the future high precision observations of bright stars at optical and NIR wavelengths will likely have to contend with variability resulting from stellar granulation and/or pulsations (Sarkar et al. 2018).

Our method offers an elegant means of measuring the transmission spectrum. Given a sufficiently long time baseline, the wavelength dependence of a star's variability can be arbitrarily well-determined. In this case any "leftover" variability — variations in transit depth that aren't explained by the wavelength-dependence of the star's variability — can be attributed to the planet's transmission spectrum. By allowing the GP mean function to vary in transit depth across wavelength during MCMC analysis we can recover an estimate of the transmission spectrum with uncertainties in the presence of stellar variability. As such, this is a straightforward extension of our model as the only change involves varying the depth and limb-darkening as a function of wavelength, while the covariance remains the same as in the examples we have already shown.

Transit timing variations occur when the gravitational interaction between planets in a multi-planet system perturbs a transiting planet away from a Keplerian orbit (Agol et al. 2005; Holman 2005). The perturbed planet will transit earlier or later than the Keplerian solution would dictate based on the relative position of the transiting planet and perturbing planet. Observations of these transit timing variations over the course of many orbits help to constrain the orbital parameters of the perturber as well as the masses of both the perturber and the transiting planet. A notable application of this technique is to the seven-planet TRAPPIST-1 system (Gillon et al. 2017; Grimm et al. 2018). Correlated noise on timescales similar to the ingress/egress time of a transit can substantially affect measurements of the transit time (Agol & Fabrycky 2018).

At present correlated noise is observable for transiting planets around evolved stars. A notable example is Kepler-91b (Barclay et al. 2015), a hot Jupiter orbiting a red giant. Individual transits of Kepler-91b are nearly undetectable due to correlated noise on similar timescales and amplitudes to the transit signal. While most main-sequence Kepler targets don't show significant correlated variability, we expect that this variability will become observable in the near future with the advent of larger space-based telescopes such as JWST. This means that accurate transit timing measurements for small planets transiting main-sequence stars will require the use of methods like ours to overcome the effects of correlated noise.

Variable phenomena: While we are primarily interested in the transiting planet problem, our multiwavelength GP implementation is likely to be useful for studies of other astronomical objects displaying time-correlated, stochastic variations. Many subfields in astronomy make use of GP variability models, or stochastic models which are equivalent to a Gaussian process, including the study of eclipsing binaries (Mahadevan et al. 2019), pulsating binaries (Hey et al. 2020), X-ray variability of the logarithm of the flux of X-ray binaries and AGN (Uttley et al. 2005; Kelly et al. 2014), the study of transient phenomena such as supernovae (Kim et al. 2013), quasar variability (Kelly et al. 2009; MacLeod et al. 2010), reverberation mapping (Zu et al. 2011; Pancoast et al. 2014), and gravitational lensing time delays (Press et al. 1992; Hojjati et al. 2013; Hojjati & Linder 2014). Multiwavelength data may be exploited to better characterize these systems. For example, Boone (2019) found much better characterization of transients with multiwavelength Gaussian process modeling, while (Peters et al. 2015) use the color-dependence of the time-correlation of quasar variability to better characterize their physical properties. It is our hope that some of these fields may benefit from applying our new multiwavelength GP implementation to study the wavelength-dependence of these various phenomena.

3.6.3 Limitations of the method

When the second dimension's covariance matrix can be represented in terms of an outer product between a vector and itself, our method has a fast scaling with the number of data points. If the second dimension cannot be described as an outer product, then we obtain a poor scaling with the size of this dimension cubed. For the method to be computationally efficient in this case, the non-`celerite` dimension should be small compared to the size of the dimension along which the covariance is specified by a `celerite` kernel function. For problems where the second dimension is comparable in size to the first and where R must be arbitrarily defined, approximate methods such as the HODLR method (Ambikasaran et al. 2015), `KISS-GP` Wilson & Nickisch (2015), or the black box methods implemented in `GPyTorch` Gardner et al. (2018) may be more efficient.

The `celerite` method is a stationary GP method, meaning that the covariance kernel is constant the 1D coordinate. In other words, $k(x_i, x_j) = k(|x_i - x_j|)$. Our 2D method inherits this limitation. What this means is that our method is not suited to modeling variability which changes substantially in amplitude or timescale over the time period in question. For instance, while our method is well-suited to modeling stellar variability over relatively short time periods, it would not be able to model solar variability across an entire solar cycle because the changing amplitude of the Sun's variability could not be captured by our stationary kernel.

If non-stationarity is required for a particular application, we refer users to other methods such as the sparse GP method of Almosallam et al. (2016) or the tree-structured GP of Bui & Turner (2014), both of which are approximate methods. It also may be possible to extend the `celerite` algorithm to non-stationary kernels by, for example, allowing the kernel coefficients to vary explicitly as a function of time, but we have yet to implement this.

Another limitation is fundamental to the Gaussian process framework — our method, like all GP methods, does a poor job of modeling outliers. When analyzing observational data, outliers are often dealt with by discarding them prior to analysis. However, in some

cases outliers may represent useful information and should be included in a model. One method of dealing with outliers without discarding them is to adopt a Student-t likelihood (Vanhatalo et al. 2009; Jylänki et al. 2011; Shah et al. 2014; Tang et al. 2017). The wider Student-t likelihood better accommodates outliers than the Gaussian likelihood, decreasing their influence on the regression. Similarly, a Gaussian mixture likelihood may be adopted, again increasing the robustness of the method to outliers (Daemi et al. 2019). We consider that a Student-t process (TP) may be an even better model for datasets containing outliers as well as having other advantages, especially with regards to TP prediction (Tracey & Wolpert 2018). We leave to future investigation the prospects for implementing a TP or TP likelihood version of `celerite` and evaluating the performance of these models on transit photometry. Unfortunately, Gaussian mixture likelihoods appear not to be compatable with the `celerite` formalism.

3.6.4 Limitations of the multiband photometric noise model

We make several assumptions in the construction of our multiband noise model which likely do not hold in all cases. First and foremost, a Gaussian process assumes that the noise is stationary and Gaussian. This does not apply to some sources of noise, such as stellar flares, or sources which undergo outbursts in which the amplitude and/or shape of the power spectrum change dramatically. Likewise our method does not apply if there is a significant time delay between the bands, if one band involves a time convolution of the other, nor if the correlated components of the bands have no correlation with one another.

Second, the specific form we've chosen for the wavelength covariance assumes that the wavelength-dependence of the flux is due to varying covering fractions of a hot and cold component in a two-component photosphere. We expect that this model will work under different assumptions; for instance, small-amplitude temperature variations should have a similar behavior as area fluctuations. However, different sources of variability will result in different forms for the covariance in the wavelength dimension.

Additionally, if there are more than two components to the photosphere, then we must

consider the possibility that each component's covering fraction varies with a different characteristic timescale. In this case rather than pairing a single wavelength covariance matrix T with a single time covariance matrix R to form the full covariance matrix $K = T \otimes R$, we should pair multiple wavelength covariance matrices with corresponding time covariance matrices, each having different characteristic timescales:

$$K = \sum_{i=1}^{N} T_i \otimes R_i. \tag{3.57}$$

Our code accepts multiple kernel components, each with a unique T matrix. We have limited ourselves to the case of a single kernel component in this chapter for the sake of clarity and simplicity.

In the examples in this chapter we chose to fix the kernel parameters. In practice the kernel parameters will need to be measured alongside the parameters of the mean model. This brings up the question of how long of a timeseries is required to produce a sufficiently strong constraint on the kernel parameters that the inference of a transit is unambiguous. We also defer this question to future work.

Finally, our formulation assumes that the observations are complete; i.e. in the multiband timesseries example, every time of observation contains data in every band. In principle this assumption could be relaxed, and in equation (3.50) the Kronecker products with $\boldsymbol{\alpha}$ could be replaced with an $\boldsymbol{\alpha}$ (and corresponding R matrix) which varies with time stamp, and only contains the amplitudes of the bands observed at each time stamp. This would also require modifying the indexing in equations (3.51) and (3.52), but the rest of the method would remain the same.

3.7 Conclusions

We have extended the `celerite` method for fast one-dimensional GP computations to two dimensions. Our method inherits the $\mathcal{O}(N)$ scaling of `celerite` in one of the two dimensions while incurring a computational cost of $\mathcal{O}(M)$ for a grid with size M in the second dimension.

Computing the 2D GP on an $N \times M$ grid thus costs $\mathcal{O}(NM)$ using our method, compared to $\mathcal{O}(N^3M^3)$ for the direct solution (i.e. inverting the full $NM \times NM$ covariance matrix). This scaling applies only when the amplitude of correlated noise varies across the bands; a more general dependence on the second dimension has a poorer scaling, yet still improves upon direct solution.

This extension may have many possible applications, among them simultaneous modeling of stellar variability across wavelength. This application is of particular interest to us, as we would like to mitigate the effects of stellar variability on detecting transiting exoplanets and measuring their properties. We demonstrate that we can improve the precision of transit depth, time, and duration measurements by modeling the transit in multiple wavelengths when compared to the monochromatic case.

When the signal-to-noise is high, we have shown that a precision which is proportional to the photon noise limit is achievable. For instance, in the two-band case in which the correlated noise in one band is twice that in the second band, one can achieve $\sqrt{10}$ of the photon-noise limit. This means that to reach the same precision as the no correlated noise case requires 10 times as many photons, or a telescope which has a collecting area ten times larger. In the limit of a blackbody which is photon-noise dominated, with a large number of bands, one can reach 2.2 times the photon-noise limit in which the correlated noise is absent. Hence, one needs to use a telescope which has $2.2^2 = 4.8$ times the collecting area. Thus, in general one can achieve a precision of measurement which is comparable to the pure photon-noise limit, but this requires about an order of magnitude more photons to do so.

In future work, we plan to extend our variability model to model more realistic stellar variability by including terms in the covariance kernel function that capture variability on different timescales with different wavelength dependencies. We suggest that the SOHO spacecraft's three-channel sunphotometer data may be a useful starting point for exploring the wavelength dependence of variability in sun-like stars. This dataset consists of measurements of the Sun's irradiance in three visible-light bands at one-minute cadence (Frohlich et al. 1995).

We are additionally interested in applying our method to RV observations of exoplanet host stars, following the method demonstrated by Rajpaul et al. (2015). This requires us to compute linear combinations of the GP and its time derivatives, which in principle should be feasible.

Our code is available in the form of a `pip` installable python package called `specgp`. `specgp` extends `exoplanet`[2] to enable 2D Gaussian process computations. Interested users can find instructions and tutorials at `https://github.com/tagordon/specgp`.

[2] `https://github.com/exoplanet-dev/exoplanet`

Chapter 4

GAUSSIAN PROCESSES FOR OUTLIER-PRONE DATASETS

4.1 Summary

Standard Gaussian process methods used in astronomy suffer from a number of limitations. One is that Gaussian process regression is sensitive to outliers in a dataset. A frequent work-around for this problem is to remove outliers before applying the GP model using a procedure like sigma-clipping in which datapoints that diverge from the mean by a set number of standard deviations are eliminated from the dataset.

In this chapter we outline an extension to the Gaussian process methods described in chapters 2 and 3 that enables their application to datasets plagued by outliers. We begin by using a Student-t likelihood function in place of the ordinary Gaussian likelihood in order to reduce the influence of outliers on the regression. We then introduce two Gaussian process models that incorporate outlying datapoints based on the models developed in Park et al. (2022): a fixed bias model in which the mean function is modified to include an additional offset for outlier observations, and a random bias model in which the white noise component of the covariance matrix is modified to accomodate outlying observations.

We develop an analytic expression for the optimal fixed and random bias terms for an individual datapoint and demonstrate how this expression can be used to identify likely outliers. Finally, we discuss prospects for applying this work to transit lightcurves in order to conduct accurate inference in the presence of outliers.

4.2 Introduction

The presence of photometric outliers is a persistent problem in transit detection and characterization. In addition to causing systematic errors in transit fits (Benneke et al. 2017; Zhang et al. 2018; Csizmadia et al. 2021) they can confound Gaussian process regression, resulting in poor detrending of stellar variability. Outliers can occur for a variety of reasons, but in data from space-based observatories they are commonly the result of cosmic ray hits on the detector (Smith et al. 2012; Quintana et al. 2010). These outliers are frequently low in amplitude and difficult to identify for removal by sigma-clipping. While cosmic ray hits can

be identified in individual cases (see Benneke et al. 2017; Kipping et al. 2015), this typically requires human intervention and substantial computational resources and is therefore not scalable to large datasets.

Robust GP methods are often implemented by replacing the Gaussian likelihood function

$$\ln\ \mathcal{L} = -\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})^{\mathrm{T}}K^{-1}(\boldsymbol{y}-\boldsymbol{\mu}) - \frac{1}{2}\ln\ \det(K) - \frac{N}{2}\ln(2\pi) \tag{4.1}$$

with a new, heavier-tailed, likelihood function. Here $\boldsymbol{y}$ is the data vector, $\boldsymbol{\mu}$ is the mean function evaluated at the datapoints, K is the covariance matrix of the process, and N is the length of the vector $\boldsymbol{y}$. While other heavy-tailed distributions have been used for this purpose (for example, Algikar & Mili (2023)), a common choice is the Student-t likelihood:

$$\ln\ \mathcal{L} = -\frac{1}{2}\frac{\nu+N}{\nu}\ln\left(1+(\boldsymbol{y}-\boldsymbol{\mu})K^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\right) - \frac{1}{2}\ln\ \det(K) - \frac{1}{2}N\ln(\nu\pi) + \ln\Gamma\left(\frac{\nu+d}{2}\right) - \ln\Gamma\left(\frac{\nu}{2}\right) \tag{4.2}$$

where ν is the degrees of freedom of the distribution and Γ is the Gamma function. We note that the Student-t likelihood contains the same dependence on K as the Gaussian likelihood. This means that the `celerite` algorithm can be employed to compute those factors in $\mathcal{O}(\mathcal{N})$ time just as for the Gaussian likelihood. This is not true of all choices of heavy-tailed likelihoods, which is one of the primary advantages of choosing the Student-t likelihood. See figure 4.2 for a comparison of the univariate Gaussian and Student-t likelihoods, and figure 4.2 for an comparison between two fits to a simulated transit observation containing outliers, one with parameters determined by optimizing a standard GP and another with parameters found by optimizing a GP with a Student-t likelihood.

While the use of a Student-t likelihood function does reduce the impact of outliers it does not completely eliminate their impact, which is why some robust GP methods still resort to deleting outliers (see Li et al. 2021). We take a different approach by using the Student-t likelihood to conduct an initial optimization of the GP hyperparameters, identifying any outliers by Bayesian model comparison, and then constructing a GP bias model (Park et al.

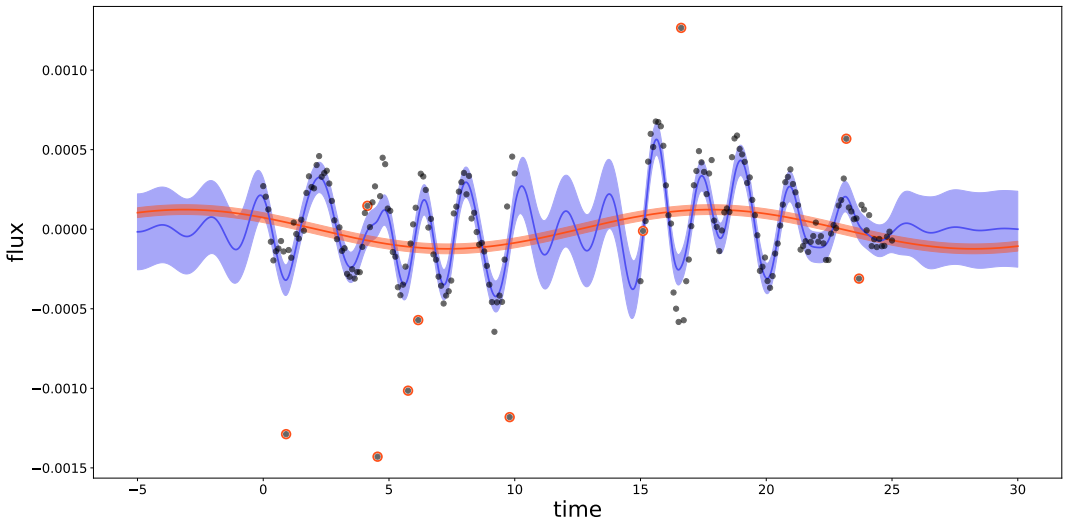

Figure 4.1 A standard GP fit compared to GP with Student-t likelihood fit to a simulated dataset containing moderate outliers. Each model was optimized using the BFGS minimization algorithm provided by `scipy`Virtanen et al. (2020) initialized to the input parameters of the simulation. The fits shown here represent the most common result of this experiment: an acceptable fit with the Student-t likelihood (blue) and a failed fit with the Gaussian likelihood (red). The outliers added to the data are the points highlighted in red. Other possible but less likely outcomes are that both fits fail or both are acceptable, with each of those outcomes happening approximately one time in five to ten simulations.

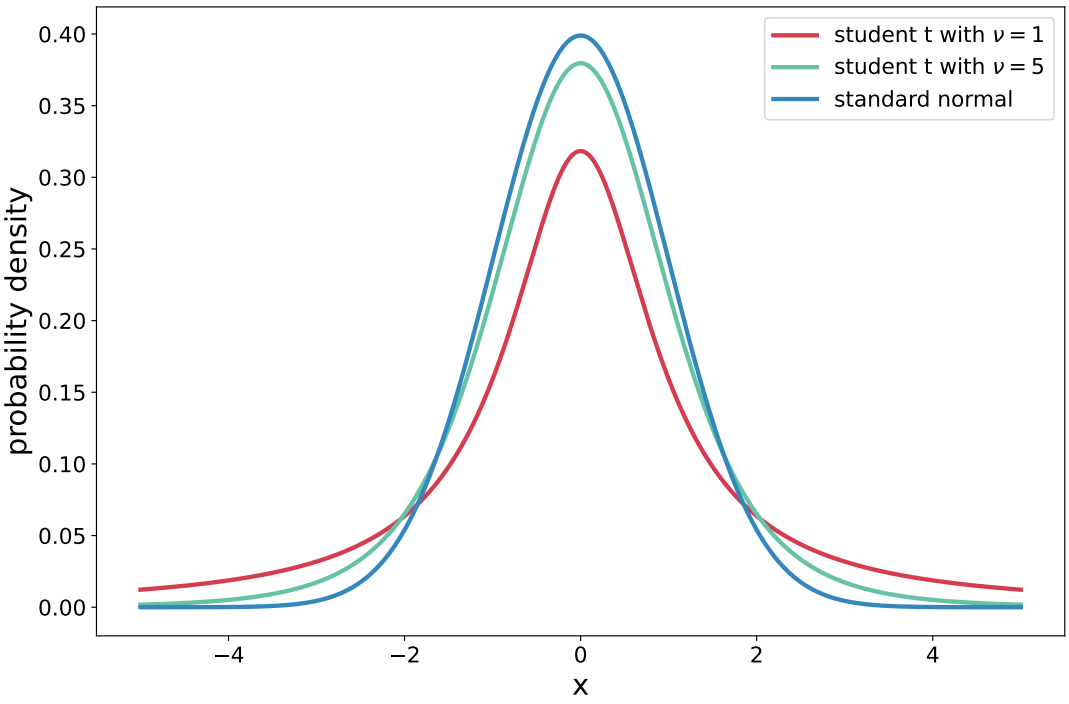

Figure 4.2 Comparison between a standard Gaussian distribution and Student-t distributions with two different values of the degree-of-freedom parameter ν. The Student-t distribution has heavier tales than the Gaussian distribution, as seen by the larger probability density at very large and very small values of the coordinate x. The parameter ν controls the weight in the tales of the distribution with the Student-t distribution approaching the Gaussian in the limit $\nu \to \infty$.

2022). In section 4.3 we present an algorithm for identifying outliers in a timeseries using a GP bias model. In section 4.4 we demonstrate the application of the two bias models to a set of simulated transit observations and compare these to straightforward outlier removal and inference with a student-t likelihood function. In section 4.5 we discuss our results and suggest some directions for future work.

Before we describe the algorithms for outlier identification it will be helpful to review the bias models described in Park et al. (2022). The constant bias model modifies the standard Gaussian process likelihood by the addition of a constant offset for each datapoint. The likelihood is written:

$$\ln\ \mathcal{L} = -\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu}-\boldsymbol{b})^{\mathrm{T}}K^{-1}(\boldsymbol{y}-\boldsymbol{\mu}-\boldsymbol{b}) - \frac{1}{2}\ln\ \det(K) - \frac{N}{2}\ln(2\pi) \tag{4.3}$$

where $\boldsymbol{b}$ is a vector containing the constant bias terms for each datapoint and all other parameters have the same definitions as in equation 4.1. It is only in the inclusion of this bias vector that the likelihood differs from the standard GP likelihood. This addition is equivalent to modifying the data vector by an independent offset for each datapoint. The model is susceptible to overfitting, which can be corrected by the addition of a regularization term to the likelihood:

$$\ln \mathcal{L}_{\mathrm{reg}} = \ln \mathcal{L} + \lambda \sum_{i}^{N} |b_i| \tag{4.4}$$

where the parameter λ controls the strength of the regularization.

The random bias model is similar in principle, except that it is the covariance rather than the mean function that is modified. The likelihood for the random bias model is:

$$\ln\ \mathcal{L} = -\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})^{\mathrm{T}}(K+I\boldsymbol{\tau}^2)^{-1}(\boldsymbol{y}-\boldsymbol{\mu}) - \frac{1}{2}\ln\ \det(K+I\boldsymbol{\tau}^2) - \frac{N}{2}\ln(2\pi) \tag{4.5}$$

where $\boldsymbol{\tau}$ is a vector of bias terms that has been added to the diagonal of the covariance matrix. Note that this definition differs from that given by Park et al. (2022) in two ways. First, we add the bias term to the existing white noise term along the diagonal of the covariance

matrix rather than replacing the diagonal with the bias vector. This alteration only effects the definition of the bias vector and should have no effect on the model itself. Second, we neglect the mean of the bias which is included in Park et al. (2022) as an offset to the data vector similar to the constant bias case. The effect of this choice is not fully explored in this this chapter but may be the subject of continuing work on this topic.

Similar to the for the constant bias model, a regularization term can be added to prevent overfitting:

$$\ln \mathcal{L}_{\text{reg}} = \ln \mathcal{L} + \sum_{i}^{N} \left(\lambda_1 \tau_i^2 + \frac{\lambda_2}{\tau_i^2} \right). \tag{4.6}$$

4.3 Identifying Outliers with Bayesian Model Selection

We propose the following algorithm for outlier detection. Given a dataset containing N observations in a vector $\boldsymbol{y}$:

1. Optimize a Gaussian process model with a Student-t likelihood to obtain an initial estimate of the kernel and mean hyperparameters.

2. For each y_i in $\boldsymbol{y}$:

   (a) Construct a constant bias GP model in which only b_i, the i^{th} entry in the bias vector $\boldsymbol{b}$, is allowed to vary. Set all other entries to 0. We refer to this as a "single-point bias model"

   (b) Optimize the constant bias model with respect to b_i.

   (c) Compute the logarithm of the likelihood ratio between the optimized bias model and an unbiased GP model.

3. After the previous step we will have a value of the log of the likelihood ratio associated with each datapoint y_i. Choose a cutoff value for the likelihood ratio and identify the datapoints for which the ratio is above that cutoff as outliers.

This same procedure can be performed for the random bias model, where instead of optimizing the likelihood with respect to b_i in step 2 we optimize with respect to τ_i.

The optimization in step 2(b) can be computed analytically, at a computational cost of one matrix inversion for each datapoint. With the `celerite` algorithm this can be accomplished at a total cost of $\mathcal{O}(N^2)$ where N is the length of the data vector $\boldsymbol{y}$. The optimal bias terms for for the single-point constant bias model are given by

$$b_i = \frac{(K^{-1}\boldsymbol{y})_i}{(K^{-1})_{ii}}, \tag{4.7}$$

where $(K^{-1}\boldsymbol{y})$ is the i^{th} entry of the vector $K^{-1}\boldsymbol{y}$ and $(K^{-1})_{ii}$ is the i^{th} element along the diagonal of K^{-1}. The numerator of this expression can be computed using a single pass of the `celerite` algorithm for all b_i while the denominator requires an application of the `celerite` algorithm for each entry, resulting in the $\mathcal{O}(N^2)$ scaling mentioned previously. The likelihood ratio between the single-point constant bias model and the unbiased GP model is given by

$$\Delta \ln \mathcal{L} = \frac{1}{2}\left(K^{-1}\boldsymbol{y}\right)_i b_i. \tag{4.8}$$

which can be computed at no extra cost by re-using b_i and the marix-vector product from 4.7. The derivation of these formulas is given in appendix B

For the random bias model the single-point optimal bias term is:

$$\tau_i = \frac{\boldsymbol{y}^{\mathrm{T}}(K^{-1})_i(K^{-1}\boldsymbol{y})_i - (K^{-1})_{ii}}{(K^{-1})_{ii}^2} \tag{4.9}$$

where $(K^{-1})_i$ is a column vector constructed from the i^{th} column of K^{-1} and the rest of the definitions are the same as in equation 4.7. The likelihood ratio between the single-point random bias model and the unbiased GP is:

$$\Delta \ln \mathcal{L} = \frac{1}{2}(K^{-1})_{ii}\tau_i + \frac{1}{2}\ln\left((K^{-1})_{ii}\tau_i + 1\right). \tag{4.10}$$

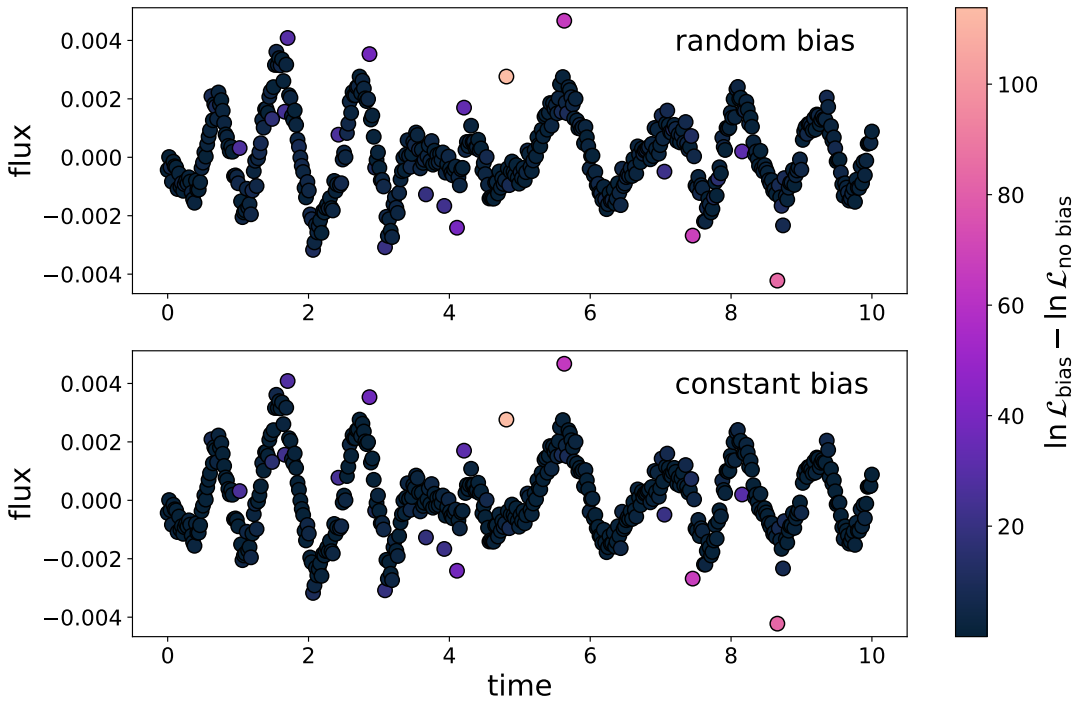

Figure 4.3 Demonstration of outlier detection using the algorithm described in the text. The top panel corresponds to the random bias model and the bottom panel to the constant bias model. Points are colored by the likelihood ratio between a single-point bias model for the point in question and an unbiased GP model. We can see that the results are very similar for both models.

The derivation of these formulas are given in appendix B.

Figure 4.3 demonstrates the application of these algorithms to a simulated lightcurve with outliers drawn from a Gaussian mixture model.

Choosing the cutoff value above which points will be considered outliers is an important and nontrivial task. The likelihood ratio does a good job of separating outliers from non-outliers for outliers of all levels of outlier severity, but the cutoff value that marks the separation changes depending on the strength of the outliers. For low-amplitude outliers a smaller cutoff value is warranted than for high-amplitude outliers, which typically require

a larger cutoff value. Choosing too small a cutoff value for a case with severe outliers may result in non-outlier points located near the outlier on the time axis time to be incorrectly identified as outliers themselves.

An optimal cutoff value for a given dataset can be found by maximizing the likelihood ratio between a bias model with N outliers and an unbiased GP model with respect to the cutoff value, where the N outliers are the points above the cutoff. In practice this optimization is expensive, and in most cases we recommend choosing a cutoff value by eye. This should be feasible for most cases because for a given application the strength of the anticipated outliers will often be similar. For low amplitude outliers a cutoff value of between 5 and 15 is often a good choice.

This procedure for identifying outliers has several attractive properties when compared to outlier clipping. The first is that it provides the likelihood ratio which can be interpreted as a probability that a specific point is an outlier, given a model for the noise covariance in the form of the GP. The second is that it provides the bias term, which in the constant bias case can be interpreted as the shift along the y-axis necessary to bring the point into agreement with the noise covariance and in the random bias case can be interpreted as the amplitude of the standard deviation that would be required to explain the observed datapoint. All of this information may be useful depending on the application.

This procedure is also useful in cases where the presence of correlated noise would require the use of a smoothing algorithm prior to outlier clipping. In cases where the timescale of the variability is very short, smoothing the data prior to outlier removal may not be feasible. A smoothing window that is too short will itself be affected by the outliers, and a smoothing window that is too long will fail to capture the variability. The result is either that outliers will be over-fit and not recognized as such, or that peaks and troughs in the variability will be interpreted as outliers. Figure 4.3 demonstrates how our procedure solves this problem.

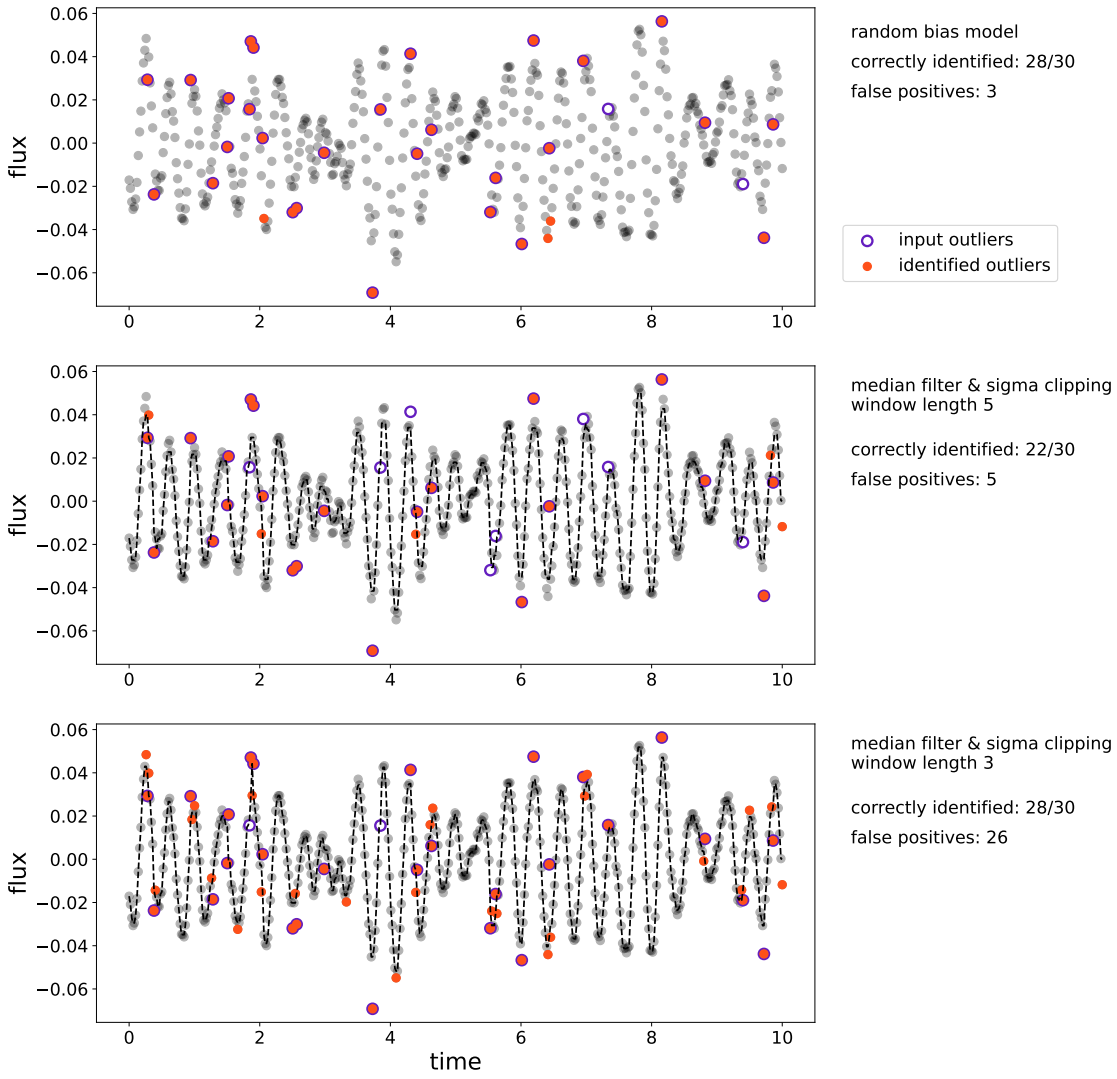

Figure 4.4 Application of our outlier detection algorithm to datasets with high-frequency variability or sparse sampling. **Top:** Outliers detected using the random bias GP model. **Middle:** Outliers detected by smoothing with a median filter followed by sigma-clipping using the sigma clipping algorithm implemented by `astropy` (Astropy Collaboration et al. 2013). Here the median filter uses a window length of 5 datapoints. **Bottom:** Same as middle panel but with a window length of 3 datapoints. In both cases we clip outliers beyond four sigma from the trend. In experimenting with different sigma values we found this value to be nearly optimal, with larger sigma values resulting in very few outliers selected and smaller sigma values resulting in large numbers of false positive outlier selections.

4.4 Inference with GP Bias Models

Once the outliers have been identified the GP bias model can be applied to infer the hyperparameters of the GP. Because the primary motivation for this work is to model outliers in transit light curves, we use simulated transit observations as an example to demonstrate the utility of these models. The procedure we employ is as follows: First we use the algorithm outlined in the previous section to identify the outliers using our chosen bias model. We then initialize an MCMC simulation using the estimate of the bias vector from above. In this simulation we allow the bias terms to vary along with the hyperparameters of the covariance and the mean function.

As an example of this procedure, figure 4.5 shows a lightcurve with noise drawn from a `celerite` Gaussian process described by the SHO kernel described in Foreman-Mackey et al. (2017). To this we have added a transit model mean and 50 simulated outliers, all of which are above the mean function in order to replicate the effect of cosmic ray strikes on the detector of a Kepler or TESS-like telescope. Although the outliers are known, for the sake of our demonstration we pretend that we do not know which points are outliers and instead identify them using the procedure described in section 4.3.

Having identified the outliers, we now construct a random bias GP model to use to infer the transit parameters. We compare our results to an identical inference procedure using an unbiased GP model with the identified outliers removed, and another in which we retain the outliers. The results of this experiment is shown in figures 4.6 and 4.7. We find that both outlier removal and the GP bias model perform similarly, and both significantly outperform the inference without outlier identification or removal.

4.5 Conclusions

Despite the similarity in performance between the GP bias model and outlier removal, the use of the bias model may still be warranted by the increased flexibility it offers. For instance, because the bias model includes the bias terms for each outlier as parameters, it is possible

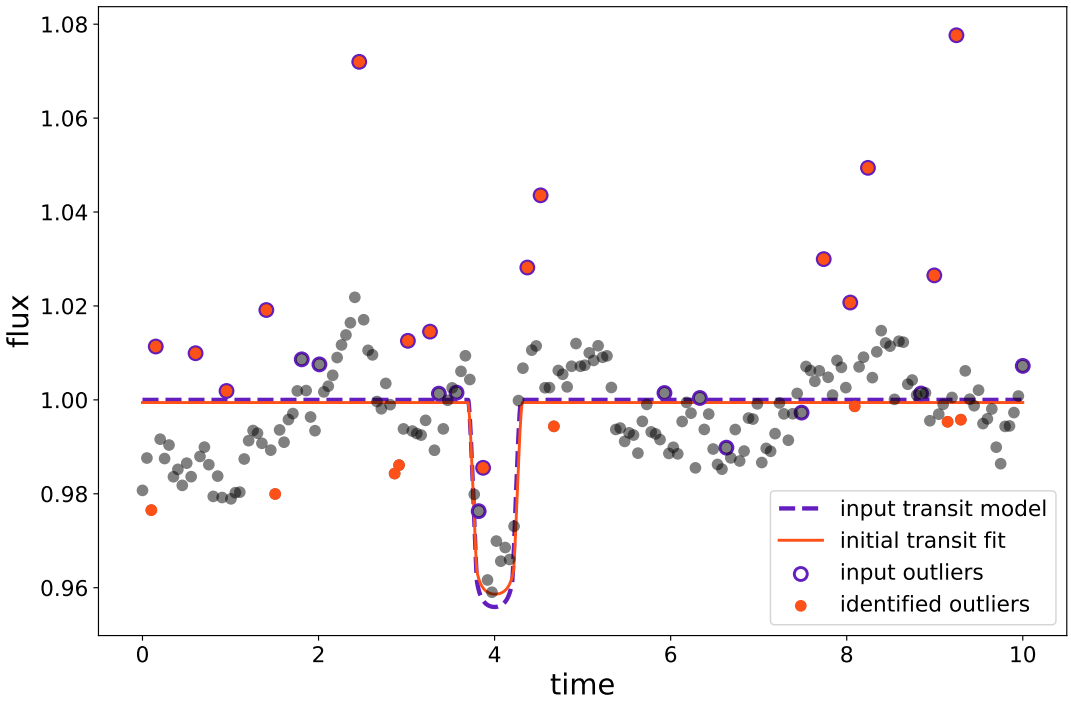

Figure 4.5 A simulated transit lightcurve with outliers. Outlier selection was conducted without knowledge of the simulated outliers. The outliers were selected using a GP with the same parameters as the GP used initially to simulate the noise. The initial transit fit was optimized using the input transit parameters as a starting point. This initial fit was then subtracted off in order to apply the outlier detection algorithm. The outliers were selected using the random bias GP model with a cutoff log likelihood ratio of $(\ln \mathcal{L}_{\text{bias}} - \ln \mathcal{L}_{\text{no bias}}) = 10$.

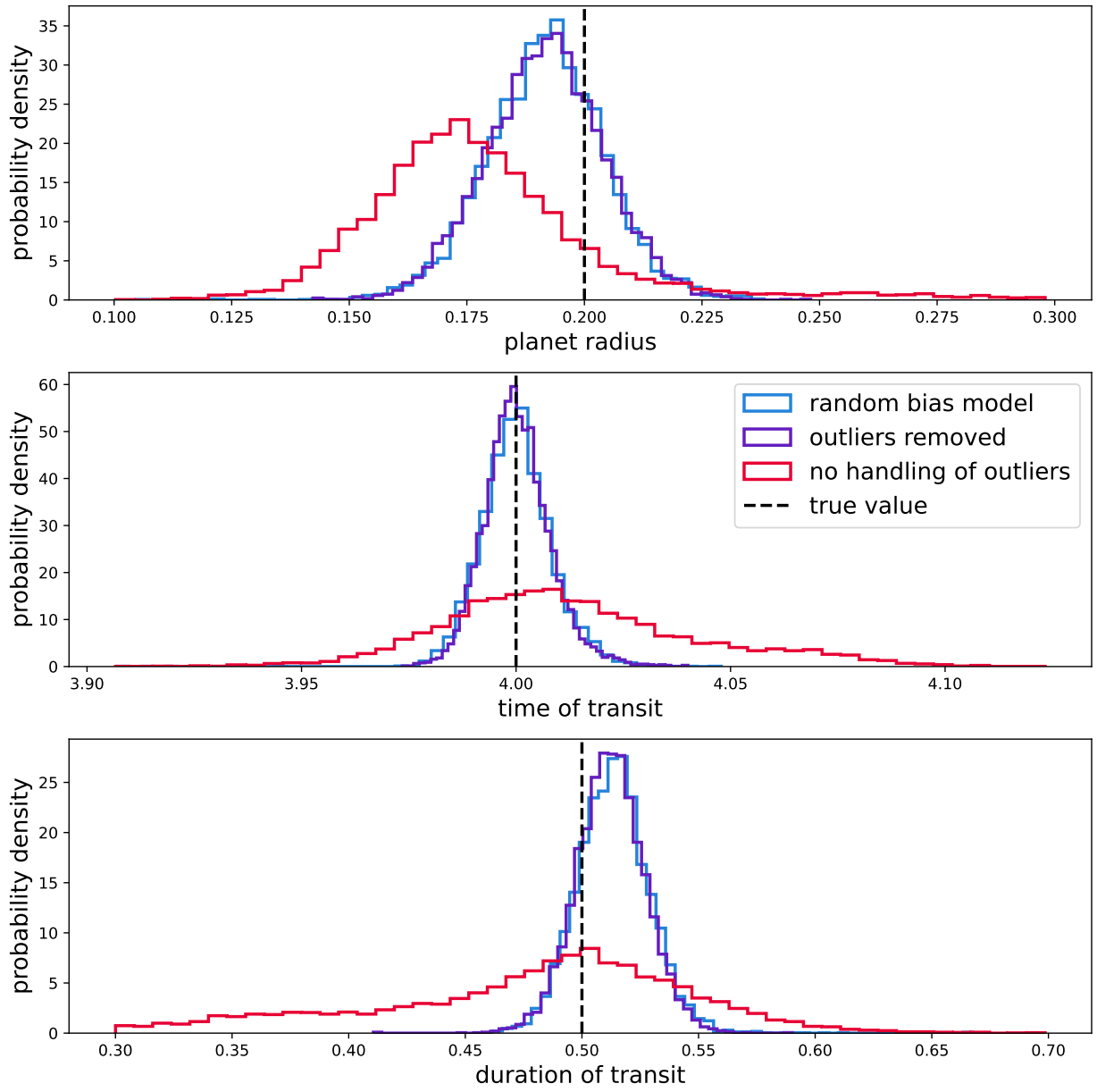

Figure 4.6 Posterior distributions for the transit parameters inferred using the random bias GP model (blue), a standard GP model with the outliers removed (purple) and a standard GP model with no handling of outliers (red). We note that the random bias model and standard GP model with outliers removed perform similarly well, and both perform much better than the standard GP with no handling of outliers.

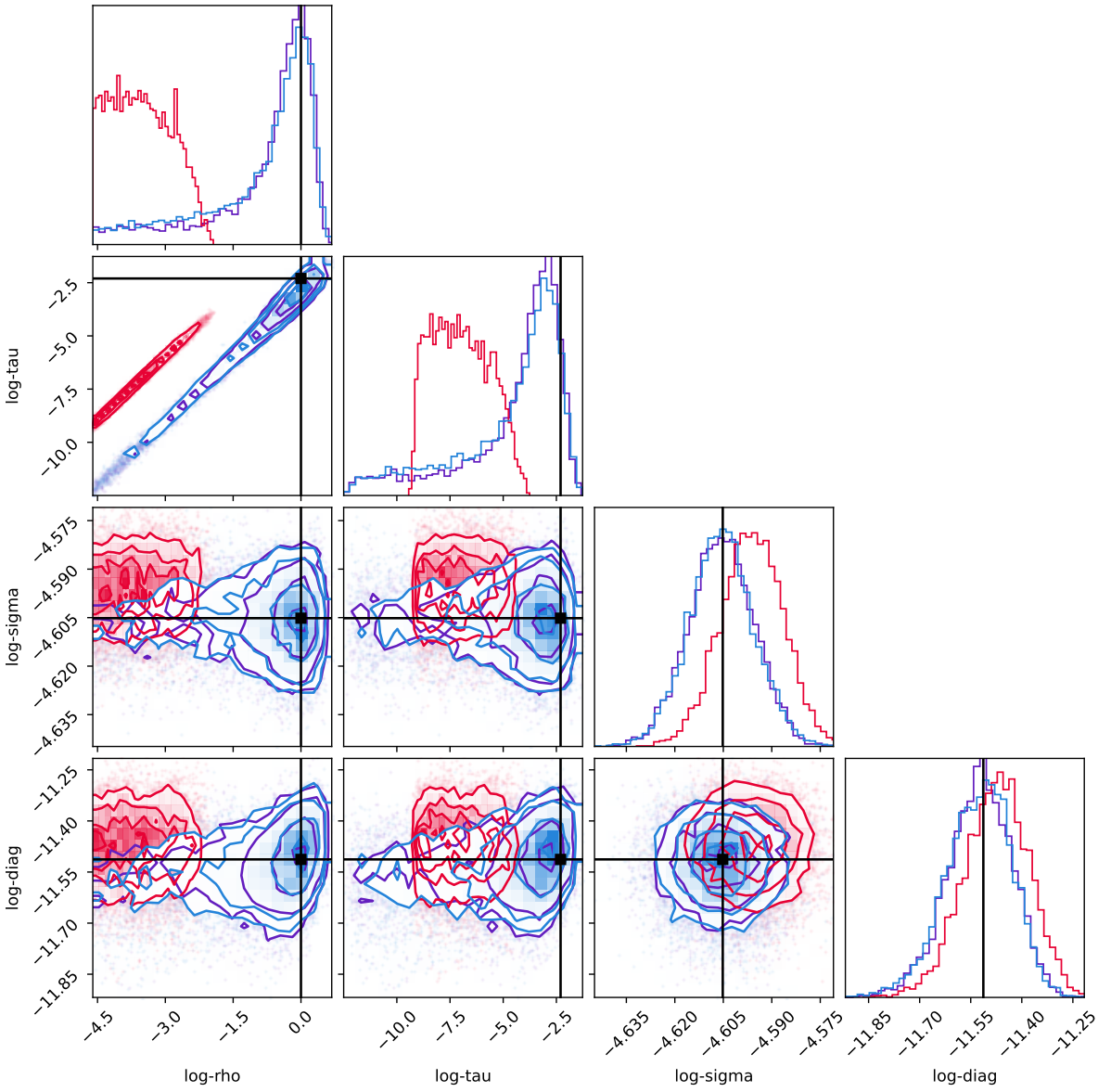

Figure 4.7 Posterior distributions for the hyperparameters of the SHO kernel of the random bias GP model. The input values are marked by black lines. The colors are the same as for figure 4.6.

to place priors on the bias term in order to reflect uncertainty about the nature of the data point. For instance a prior centered at a small value for b_i or τ_i would cause the model to weigh that outlier more heavily than if the prior were less informative or was centered on a larger value. In this way the user can make a choice about how much influence to allow each outlier to have on the result of the inference.

Another advantage of the bias model is that it allows for more aggressive outlier selection. Since the information present in the outliers is retained, incorrectly identifying a valid datapoint as an outlier does not have as great an effect as it would if the outlier were simply deleted. This allows the user to err on the side of selecting too many points as outliers without as much concern about losing information about an underlaying trend or mean function.

There are many directions for future work in this area to take. First, the algorithm for outlier detection can likely be further refined. Because the hyperparameters of the GP kernel are held constant during the optimization for the bias terms, the minima found are not the global minima of the model. Furthermore, by considering only one potential outlier at a time we ignore the effect that other outliers have on the quality of the fit. A better procedure may be one in which groups of candidate outliers are considered simultaneously.

Another consideration for future work is the effect of priors on the bias vectors in the GP bias models. Adding priors allows for a level of control over the influence of outliers and offers a natural way to incorporate uncertainty on whether individual points are best described as outliers or not.

Chapter 5

A LIGHTCURVE MODEL FOR JOINT EXOMOON/EXOPLANETARY TRANSITS

Portions of this chapter were originally published in collaboration with Eric Agol in the August 2022 issue of The Astronomical Journal (Tyler A. Gordon and Eric Agol 2022 AJ 164 111) and are reproduced here with permission of the American Astronomical Society.

5.1 Summary

We present a solution for the light curve of two bodies mutually transiting a star with polynomial limb darkening. The term "mutual transit" in this chapter refers to a transit of the star during which overlap occurs between the two transiting bodies. These could be an exoplanet with an exomoon companion, two exoplanets, an eclipsing binary and a planet, or two stars eclipsing a third in a triple star system. We include analytic derivatives of the light curve with respect to the positions and radii of both bodies. We provide code that implements a photodynamical model for a mutual transit. We include two dynamical models, one for hierarchical systems in which a secondary body orbits a larger primary (e.g. an exomoon system) and a second for confocal systems in which two bodies independently orbit a central mass (e.g. two planets in widely separated orbits). Our code is fast enough to enable inference with MCMC algorithms, and the inclusion of derivatives allows for the use of gradient-based inference methods such as Hamiltonian Monte Carlo. While applicable to a variety of systems, this work was undertaken primarily with exomoons in mind. It is our hope that making this code publicly available will reduce barriers for the community to assess the detectability of exomoons, conduct searches for exomoons, and attempt to validate existing exomoon candidates. We also anticipate that our code will be useful for studies of planet-planet transits in exoplanetary systems, transits of circumbinary planets, and eclipses in triple-star systems.

5.2 introduction

Mutual transits, which have also been referred to as "planet-planet eclipses", "planet-planet transits", and "overlapping double transits" (Luger et al. 2017; Hirano et al. 2012; Masuda et al. 2013; Masuda 2014), occur when two bodies transit a star at the same time and ex-

hibit some degree of overlap during the simultaneous transit. This can happen between an exoplanet and its moon, between two exoplanets, between a planet and a star in a circumbinary system, or between two stars in an eclipsing triple-star system. In past work, different techniques have been used to model the light curves resulting from this three-body overlap depending on the type of system being investigated.

In this chapter we present a solution for the flux blocked by two mutually transiting bodies. Our solution follows the method outlined by Pál (2012) in applying Green's theorem to transform the integral over the area of the star blocked by the transiting bodies into a line integral along the boundary of the unocculted area of the star, but unlike Pál (2012) which gives a general solution for N overlapping bodies, our solution is for exactly two bodies. This allows us to write down an analytic solution for the flux in each of the 16 possible geometric cases shown in Figure 5.1. We also expand on previous work by computing derivatives with respect to the limb-darkening parameters and the positions and radii of each body.

We use our solution to implement a photodynamical model which takes as input the dynamical parameters of two bodies, computes the positions of each at an array of times, and then calculates the flux observed at those times. We call this model `gefera`[1]. `Gefera` includes two dynamical models: a hierarchical model that can be used to simulate either an exomoon system or a hierarchical triple-star system, and a confocal model which is applicable to mutual transits of two exoplanets. In Figure 5.2 we present a sample light curve for a mutual exomoon/exoplanet transit computed with `gefera`. In addition to computing the light curve itself, `gefera` provides derivatives of the light curve with respect to all of the input parameters. These derivatives can help speed up fitting algorithms and is essential for gradient-based MCMC methods such as Hamiltonian MCMC. Derivatives are also necessary to compute the information matrix, which gives insight into the information content of the light curve with respect to each parameter under given noise assumptions (Vallisneri 2008).

[1] *gefera* is an Old English word with the literal meaning "fellow traveller," which might be translated today as companion or comrade. We feel that this is an appropriate name for a code developed to search for exomoons, companions and fellow travellers to their exoplanet hosts.

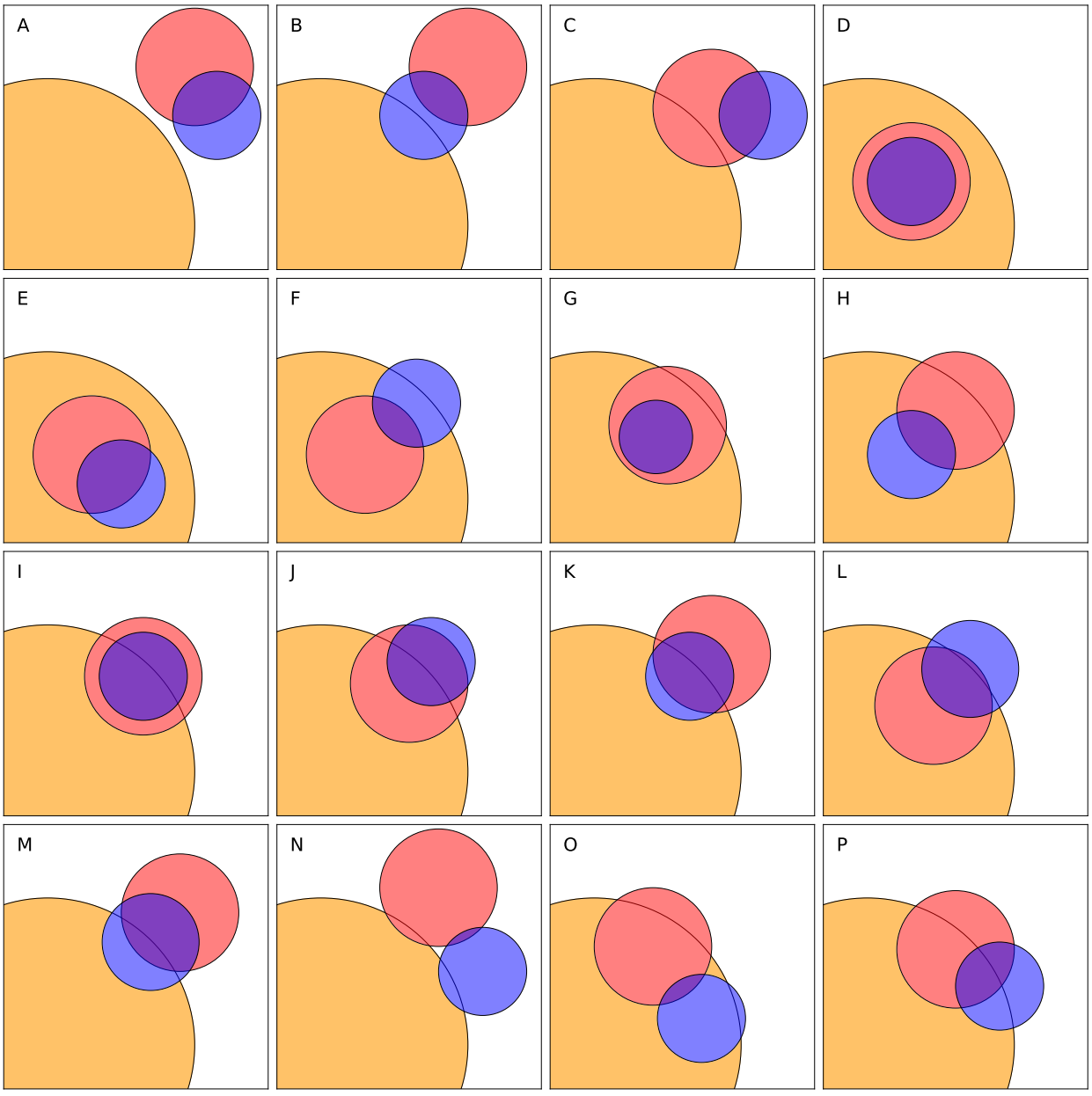

Figure 5.1 Representative examples of the different possible configurations of the planet-moon transit for the case where the planet and moon overlap each other. The letters in the top right of each box refer to the outcomes of the decision tree in Figure 5.4.

Figure 5.3 shows the derivatives of the simulated light curve in Figure 5.2. `Gefera` has been developed openly in a public github repository and is provided for use by the community under a GPL license [2].

In the remainder of the introduction we discuss existing models for mutual transits and the types of systems to which they have been applied, highlighting the differences between these models and `gefera`.

5.2.1 Exomoon Transits

Despite a lack of observed exomoon transits, it is likely that these are the most frequently occurring, though not the most easily observable, mutual transit events. Several candidate exomoons have been identified in Kepler photometry (see Teachey & Kipping 2018; Kipping et al. 2022). The analyses that produced these candidates used the code `LUNA` (Kipping 2011) which employs the small-planet approximation (or in this case the small-moon approximation) from Mandel & Agol (2002) to compute the flux blocked by the moon while applying the exact analytic solution (also from Mandel & Agol 2002) for the planet.

In contrast to `LUNA`, our method is exact for both bodies. While the small-planet approximation is very accurate for reasonably-sized exomoons, this does limit the utility of `LUNA` for very large satellites or binary planet pairs, especially around M-dwarfs with small radii. `LUNA` does not compute derivatives of the transit light curve. Finally, `LUNA` is proprietary whereas `gefera` is being made publicly available.

After this work was initially published the pubicly available exomoon transit code `Pandora` (Hippke & Heller 2022) was released. Like `LUNA`, this code implements an approximate solution to the mutual transit based on the small-planet approximation. However unlike `LUNA` the small-planet approximation is only applied to the area of overlap between the planet

[2]The code is available at https://github.com/tagordon/gefera and is pip-installable under the name gefera. The Gnu Public License (GPL) generally permits both non-commericial and commerial use of the code under the stipulation that any works making use of the code are licensed in such a way that they grant users the same freedoms as received under the Gnu Public License. A permanent snapshot of the code is available on Zenodo (Gordon 2022)

and the moon, so that the flux is exact when there is no overlap between the two bodies. `Pandora` does not compute derivatives of the light curve.

5.2.2 Mutual Transits of Exoplanets

A far less likely but more easily observed scenario is that in which two exoplanets overlap during a simultaneous transit (Ragozzine & Holman 2010). Brakensiek & Ragozzine (2016) gives a full treatment of the geometric probability of these events. Such an event is thought to have been observed in the Kepler-94 system (Hirano et al. 2012; Masuda et al. 2013). Additionally, one observed simultaneous transit of Kepler-51 b and d shows a feature which may be explained by planet-planet overlap during the transit (Masuda 2014, to be further explored in §5.10.3). These authors model the mutual transit event as a superposition of two single transits computed using the Mandel & Agol (2002) solution, with the addition of a "bump" function that depends on the area of the overlap between the two bodies and takes into account stellar limb-darkening in an approximate fashion similar to the small-planet approximation put forth in Mandel & Agol (2002). The difference between the methods used by Masuda et al. (2013); Masuda (2014) and Kipping (2011) is that in the latter method the small-planet approximation for the entire area of the smaller body, whereas Masuda et al. (2013) and Masuda (2014) only approximate the flux blocked by the region of overlap between the two bodies – the non-overlapping regions are computed exactly.

The code `planetplanet` described in Luger et al. (2017) is also capable of modeling planet-planet mutual transits. This code implements a somewhat novel integration scheme in which the limb-darkened stellar surface is approximated by a series of annuli of constant radiance. The region of overlap between these annuli and the occulting body can then be computed analytically. This method is then generalized for overlapping bodies. This treatment of the overlapping transit is analytic in the sense that the integral for each annulus is computed analytically, but relies on an approximation of the limb-darkening profile and requires a potentially large number of evaluations of the analytic integrals to accurately approximate the limb-darkened light curve.

`Gefera` again distinguishes itself in that it computes the exact light curve and includes derivatives, although in the case of the Masuda et al. (2013) method the light curve will be exact for any times during which the two bodies are non-overlapping. The derivatives enable faster optimization and inference than can be accomplished otherwise.

5.2.3 Eclipsing Triple-star Systems

The type of mutual transit event for which we have the most examples are those that take place in triple-star systems when two stars transit the primary. These eclipses have most frequently been modeled using the Wilson-Devinney (Wilson & Devinney 1971) code or more recent codes based on that work (see Borkovits et al. 2020, 2022 and Mitnyan et al. 2020 for some recent examples that make use of `lightcurvefactory`, Borkovits et al. 2013, 2019). These codes compute light curves numerically, accounting for tidal distortion and light travel time, among other effects, and often including radial velocities and spectral information in their output. While slower to compute than an analytical model, the inclusion of tidal distortion and light travel time effects is important for accurately modeling many multi-star systems.

When spectral data is not available and when tidal distortion effects are not considered important, much faster analytic methods can be employed. Carter et al. (2012) use an analytic method developed by Pál (2012) which employs Green's theorem to transform the integral over the occulted region of the star into a line integral along the boundary of the occulter. The code `photodynam`, based on that method, has been used to model the Kepler-126 system, a hierarchical triple-star system in which two low mass stars transit a third star (Carter et al. 2011). `Photodynam` has also been used in the analysis of circumbinary planets Kepler-34 b and Kepler-35 b (Welsh et al. 2012), Kepler-16 b (Doyle et al. 2011) and Kepler-36 b and c (Carter et al. 2012). Short et al. (2018) also model Kepler-126 and Kepler-16 by numerically evaluating the line integral given in Pál (2012). This allows for the use of arbitrary limb-darkening laws and enables them to model the Rossiter-McLaughlin effect using the same principle of numerically evaluating the line integral over a map of the

radial velocity.

Although `gefera` is not a viable alternative to the full spectral model implemented by Wilson & Devinney (1971) and similar codes, it may be preferable to `photodynam` in cases where the more complete model is not required. While both `photodynam` and `gefera` compute an exact light curve, `photodynam` does not include derivatives and was found to be slightly slower than `gefera` in testing. Finally, whereas `photodynam` is available only as a `c/c++` code and is coded to write output to a file, `gefera` is designed from the ground up for Bayesian inference and includes an easy-to-use `python` interface. `Gefera` allows for one or both of the transiting bodies to be larger than the star. For transiting triple-star systems the user may wish to add the flux from the transit between the two foreground stars, which can be accomplished by running `gefera` a second time with the radius of one of the transiting bodies set to zero and adding this to the flux from the three-body computation. Details on this procedure are given in section 5.10.2

As TESS and CHEOPS continue to collect light curves the small number of mutually transiting multi-planet systems is likely to expand. Furthermore, upcoming transit missions such as PLATO and ARIEL represent new opportunities to detect transiting exomoons, binary planets, or other interesting and exotic systems. These prospects all point towards the need for a fast, accurate, robust, and easy-to-use mutual transit model that is compatible with cutting-edge inference techniques. It is this need that we hope `gefera` will fulfill.

The organization of this chapter is as follows: in §5.3 we outline our approach to computing the integral of the limb-darkening profile over the area of the occulting bodies. This follows closely the approach of Pál (2012), Luger et al. (2019), and Agol et al. (2020). We also show how we use the chain rule to write the derivatives with respect to the radii and positions of the two transiting bodies. In §5.4.1 we give the solution and its derivatives for the uniform term of the limb-darkening profile. In §5.4.2 we give the solution and derivatives for the linear term, and in §5.4.3 we do the same for the quadratic term. In §5.4.4 we extend our model to polynomial limb-darkening laws. In §5.5 we give formula for computing the locations of the intersections between each transiting body and the limb of the star, and the

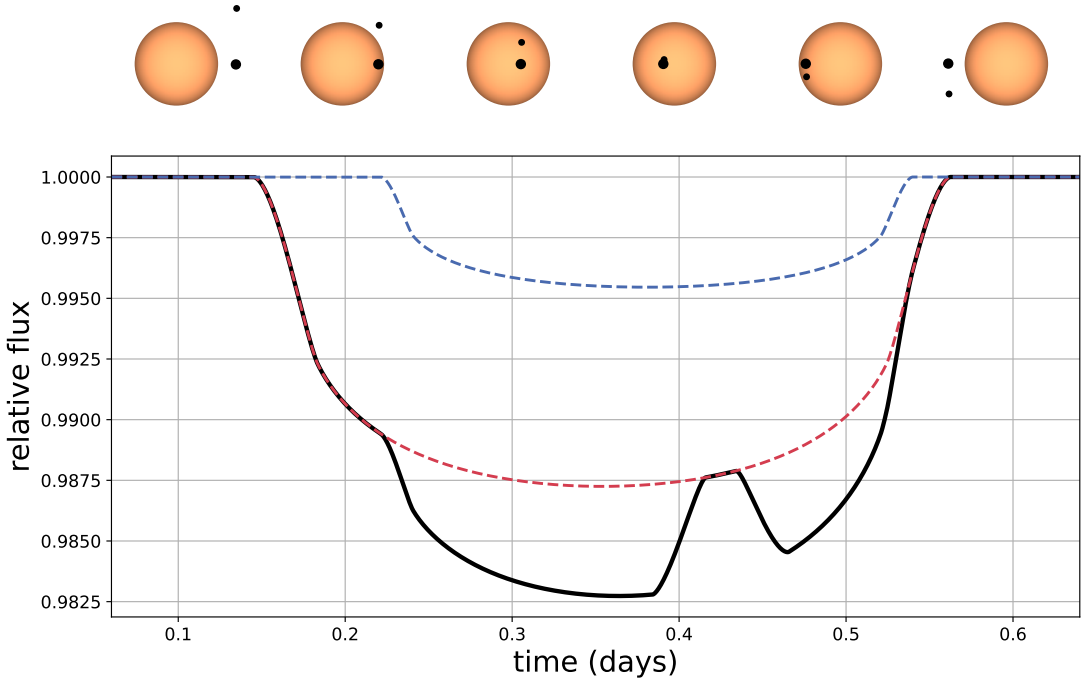

Figure 5.2 A simulated exoplanet/exomoon transit in which the exomoon and exoplanet fully overlap while transiting the star. The snapshots across the top of the plot show the configuration of the system at the points in time indicated along the x-axis of the light curve plot. The blue dashed line shows the transit of the moon as it would appear if it were the only transiting body, and the red dashed line shows the same for the planet.

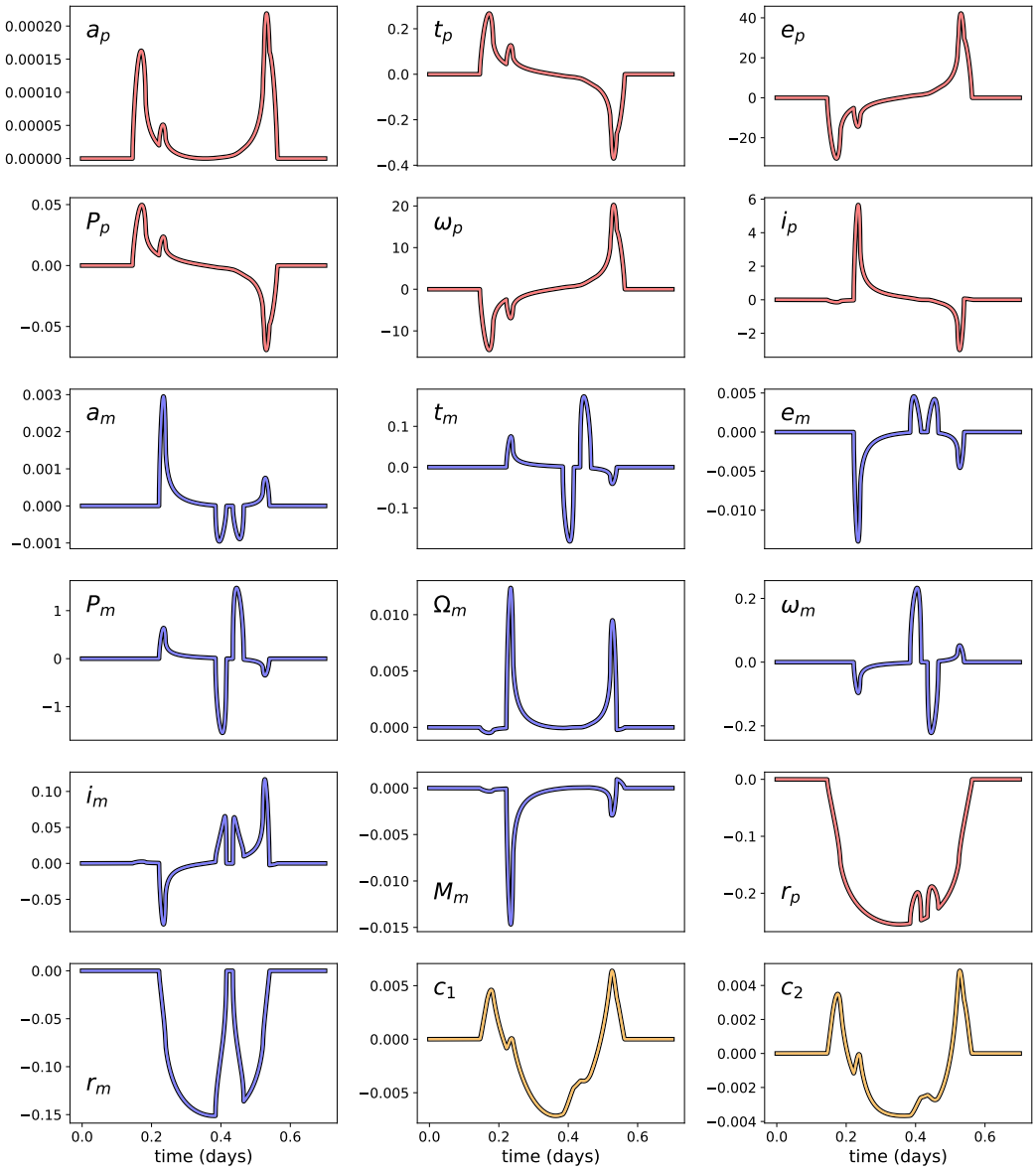

Figure 5.3 Derivatives of the simulated transit in Figure 5.2 with respect to all model parameters. In order from left to right, top to bottom the parameters are the semimajor axis, epoch, eccentricity, period, longitude of periastron, and inclination of the planet, the semimajor axis, epoch, eccentricity, period, longitude of the ascending node, longitude of periastron, and inclination for the moon, the moon mass as a fraction of the planetary mass, the planetary radius, the moon's radius, and the two limb-darkening parameters. The units along the y-axes are in units of flux per the relevant unit for the parameter in question. For example, if the semimajor axis of the planet is in terms of stellar radii then the units for the upper left plot are flux per unit stellar radius. The lines are color-coded as follows: parameters related to the planet are shown in red, those related to the moon are shown in blue, and the limb-darkening parameters for the star are shown in yellow.

planet or moon and the star's limb. The formulae we give are crafted to be numerically stable and to mitigate roundoff error. In §5.6 we show how to choose the integration limits for each arc depending on the geometry of the system at that moment. This section references the flowchart in Figure 5.4 which shows the procedure for determining which geometric case the system is in at a given time. Figure 5.1 gives examples of the geometry for each of the cases in the flowchart, and Table 5.6.3 gives the evaluated integral for each case in terms of the solutions for G_N given in §5.4.1, §5.4.2, §5.4.3, and §5.4.4.

In §5.8 we benchmark our code, both with and without the gradient computation, against `photodynam`, and in §5.9 we compare the accuracy of our code with `LUNA`. Finally, in §5.10 we discuss several applications of our code, focusing on conducting MCMC simulations to infer parameters of the moon/planet transit and demonstrating the use of our code by re-analysing the potential mutual transit of Kepler-51 b & d. Lastly, we include as an appendix a minor correction to Kipping (2011) and Fewell (2006) that has to do with an incorrect expression for the area of overlap of three circles.

5.3 Our approach

We consider the flux from a star eclipsed by two bodies, here taken to be an exoplanet and its moon, though they could also represent two exoplanets, two stars, or a combination of these. For clarity and consistency throughout the chapter we refer to the two transiting bodies as the planet and moon, with the planet being the larger body and the moon the smaller. We frequently use "p" and "m" as subscripts to refer to parameters that pertain to the planet and moon (or the larger and smaller bodies) respectively.

We initially consider the star's radial surface brightness profile, $I(r)$, to be described by a quadratic limb-darkening law before expanding to consider polynomial limb-darkening. The quadratic limb-darkening profile is given by

$$I(r) = 1 - c_1(1-\mu) - c_2(1-\mu)^2, \tag{5.1}$$

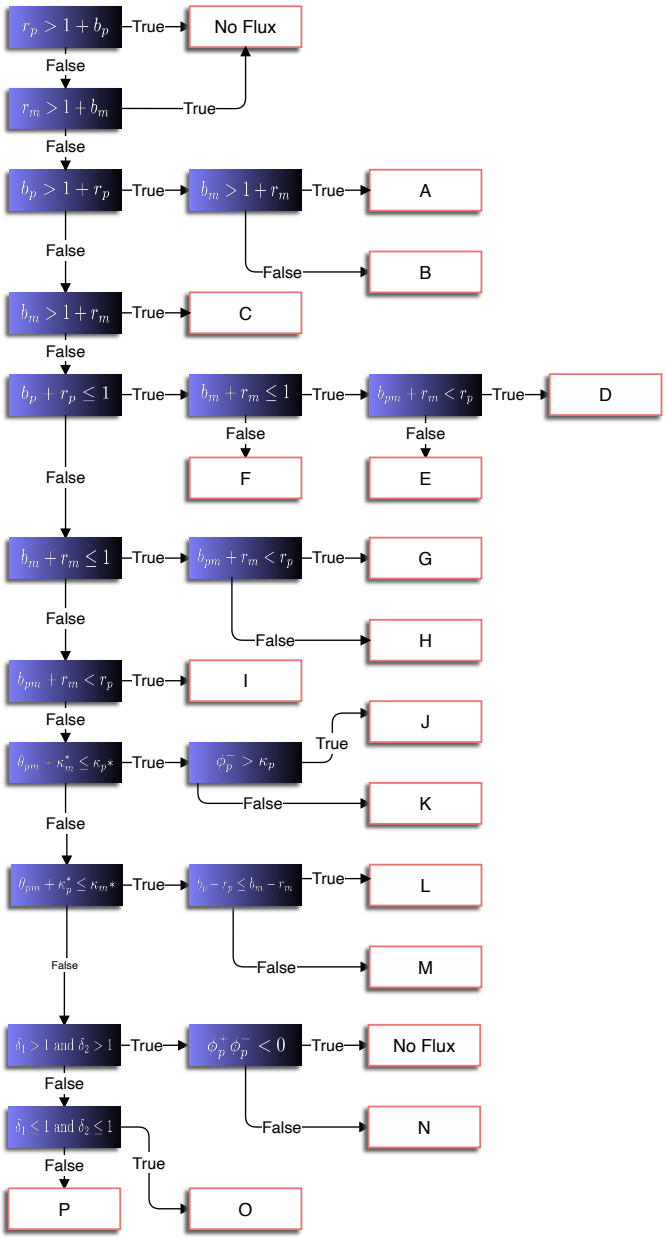

Figure 5.4 Decision tree when the planet and moon overlap each other ($b_{pm} < r_p + r_m$). Blue cells represent a test on the state of the system, the outcomes of which determine the cases represented by white boxes. The letters in the white boxes refer to the configurations shown in Figure 5.1, and the arcs to integrate and their integration limits for each case are listed in Table 5.6.3. The angles are defined in section **??**. The boxes labeled "No Flux" indicate cases where the star is completely obscured by one or both of the transiting bodies. These outcomes are only possible when either r_p, r_m, or both are greater than 1, and thus would likely only occur for eclipsing binary or triple-star systems.

where $\mu = \sqrt{1-r^2}$ with r being the radial distance from the center of the star projected onto the sky plane. This intensity profile can be rearranged as follows:

$$\begin{aligned} I(r) &= (1-c_1-2c_2)+(c_1+2c_2)\mu+c_2(x^2+y^2) \\ &= (1-c_1-2c_2)g_0+(c_1+2c_2)g_1+c_2g_2 \\ &= u_0g_0+u_1g_1+u_2g_2, \end{aligned} \tag{5.2}$$

where we have defined $g_0 = 1$, $g_1 = \mu = \sqrt{1-x^2-y^2}$ and $g_2 = r^2 = x^2+y^2$. We consider each term of this equation separately, referring to the g_0 term as the constant or uniform term, g_1 as the linear term, and g_2 as the quadratic term. We define a coordinate system in which the planet is centered on the positive x-axis at $x = b_p$ and the moon is a distance b_{pm} from the planet. The line connecting the centers of the moon and planet is at an angle θ from the x-axis as shown in Figure 5.5.

Following the approach of Pál (2012), we apply Green's theorem to transform the integral from a surface integral over the eclipsed region of the star into a line integral over a vector field along the closed loop bounding the unocculted region. The integration along the boundary is split into arcs along the boundaries of the star, moon, and planet. An example configuration is shown in figure 5.5.

We consider the arcs belonging to the planet and the moon separately. The path along the arc of the planet is parameterized as

$$x = b_p - r_p\cos\phi', \tag{5.3}$$

$$y = r_p\sin\phi', \tag{5.4}$$

where ϕ' is the angle measured from the negative x-axis.

Once we have completed the integration along the arcs belonging to the planet, we rotate our coordinate system through an angle θ_{pm}, where θ_{pm} is the angle between the center of the planet and the center of the moon relative to the center of the star, so that the moon

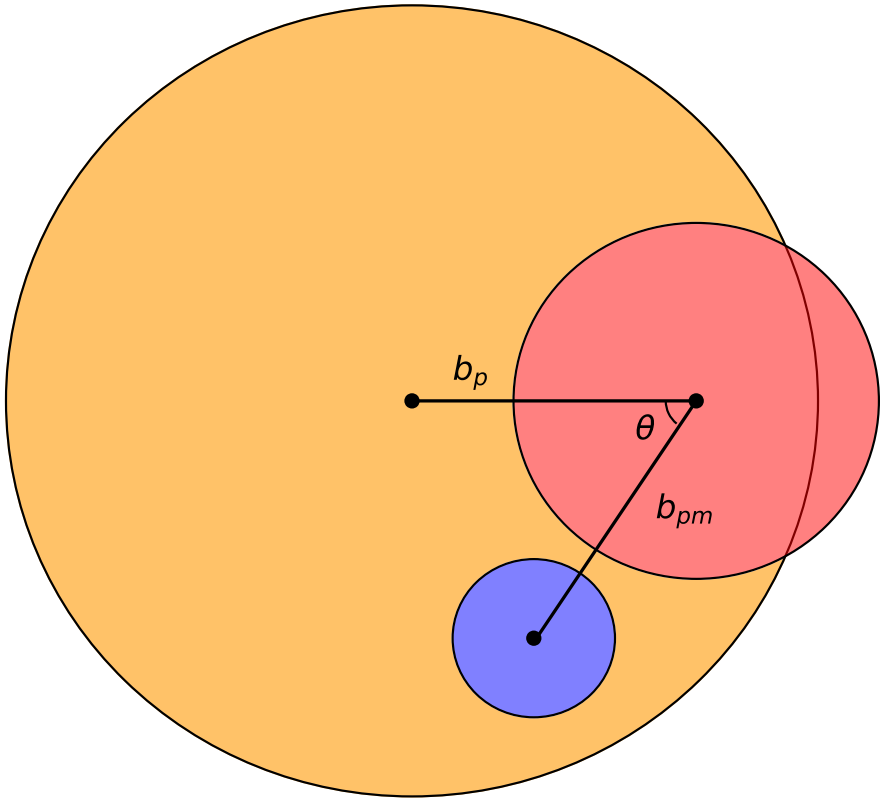

Figure 5.5 A sample configuration of the transiting exoplanet (red circle) and exomoon (blue circle). We use the input coordinates $\{b_p, b_{pm}, \theta\}$. Each of these coordinates are independent of the other two and together with the radii r_p and r_m they completely define the state of the system.

sits on the positive x-axis at $(0, b_m)$. The path along an arc belonging to the moon can now be parameterized as

$$x = b_m - r_m \cos\phi', \tag{5.5}$$

$$y = r_m \sin\phi'. \tag{5.6}$$

Green's theorem does not technically allow us to transform coordinate systems mid-way through the integration, as that breaks the closed loop to which the theorem applies. Fortunately, for radial limb-darkening laws it is possible to choose a vector field for Green's theorem which is invariant to rotations about the center of the star so that the integral along the arc of a circle centered at $(b, 0)$ is identical to the integral along the arc of a circle centered at $(b\cos\theta_{pm}, b\sin\theta_{pm})$ for any value of θ_{pm}.

For each term in Equation (5.2) we define the *primitive integral*

$$G_n(\phi, r, b) = -\int_0^{\phi} \left[f_{n,y}(x, y)c_{\phi'} + f_{n,x}(x, y)s_{\phi'}\right] r d\phi', \tag{5.7}$$

where $c_{\phi'}$ and $s_{\phi'}$ stand in for $\cos\phi'$ and $\sin\phi'$ respectively, and where $\boldsymbol{f}_n$ is a vector field chosen such that

$$\frac{df_{n,y}}{dx} - \frac{df_{n,x}}{dy} = g_n. \tag{5.8}$$

Table 5.6.3 shows how these integrals can be combined to compute the total flux. As discussed above, we also require that the integrand be invariant with respect to rotation about the origin as this condition allows us to transform between coordinate systems where each body is positioned on the positive x-axis. While this isn't strictly necessary (we could complete the integration in a single coordinate system), it simplifies the algebra significantly and reduces the required number of computations by allowing us to compute an integral of the same form

for every arc. The vector fields $\boldsymbol{f}_0$ and $\boldsymbol{f}_1$ are given by Pál (2012) as

$$\begin{aligned} \boldsymbol{f}_0 &= \frac{1}{2}(-y\hat{\boldsymbol{x}} + x\hat{\boldsymbol{y}}), \\ \boldsymbol{f}_1 &= \frac{1-(1-x^2-y^2)^{3/2}}{3(x^2+y^2)}(-y\hat{\boldsymbol{x}} + x\hat{\boldsymbol{y}}). \end{aligned} \tag{5.9}$$

These fields meet our requirement of being invariant to rotations about the center of the star. However the expression given by Pál (2012) for the quadratic term does not meet this requirement. Therefore we adopt the following for $\boldsymbol{f}_2$:

$$\boldsymbol{f}_2 = \frac{1}{4}(x^2+y^2)(-y\hat{\boldsymbol{x}} + x\hat{\boldsymbol{y}}). \tag{5.10}$$

Finally, for the polynomial terms we adopt the fields given by Agol et al. (2020), which already meet our requirement:

$$\boldsymbol{f}_n = (1-x^2-y^2)^{n/2}(-y\hat{\boldsymbol{x}} + x\hat{\boldsymbol{y}}), \tag{5.11}$$

for $n \geq 2$.

The angle ϕ is the endpoint of the arc as measured from the negative x-axis. This angle depends on the geometry of the system and can correspond to either an intersection between the planet and moon, or an intersection between the planet or moon and the limb of the star. Thus ϕ is a function of r_p, r_m, b_p, and b_m. With this dependence made explicit, G_n becomes a function of r_p, r_m, b_p, and b_m as well. We can now use the chain rule to write the derivatives of G_n with respect to each of these variables:

$$\begin{aligned} \frac{dG_n(\phi, b_p, r_p)}{dr_p} &= \frac{\partial G_n}{\partial r_p} + \frac{\partial G_n}{\partial \phi}\frac{\partial \phi}{\partial r_p}, \\ \frac{dG_n(\phi, b_m, r_m)}{dr_p} &= \frac{\partial G_n}{\partial \phi}\frac{\partial \phi}{\partial r_p}, \end{aligned} \tag{5.12}$$

$$\begin{aligned}
\frac{dG_n(\phi, b_p, r_p)}{dr_m} &= \frac{\partial G_n}{\partial \phi}\frac{\partial \phi}{\partial r_m}, \\
\frac{dG_n(\phi, b_m, r_m)}{dr_m} &= \frac{\partial G_n}{\partial r_m} + \frac{\partial G_n}{\partial \phi}\frac{\partial \phi}{\partial r_m},
\end{aligned} \tag{5.13}$$

$$\begin{aligned}
\frac{dG_n(\phi, b_p, r_p)}{db_p} &= \frac{\partial G_n}{\partial b_p} + \frac{\partial G_n}{\partial \phi}\frac{\partial \phi}{\partial b_p}, \\
\frac{\partial G_n(\phi, b_m, r_m)}{\partial b_p} &= \frac{\partial G_n}{\partial \phi}\frac{\partial \phi}{\partial b_p},
\end{aligned} \tag{5.14}$$

$$\begin{aligned}
\frac{\partial G_n(\phi, b_p, r_p)}{\partial b_m} &= \frac{\partial G_n}{\partial \phi}\frac{\partial \phi}{\partial b_m}, \\
\frac{\partial G_n(\phi, b_m, r_m)}{\partial b_m} &= \frac{\partial G_n}{\partial b_m} + \frac{\partial G_n}{\partial \phi}\frac{\partial \phi}{\partial b_m}.
\end{aligned} \tag{5.15}$$

Because we locate the moon using the parameters b_{pm} and θ rather than using b_m directly, we now need to show how the derivatives with respect to the set of input coordinates $\{b_p, b_{pm}, \theta\}$ are recovered from the integrals with respect to b_p and b_m using the chain rule. For the derivatives of G_n we have:

$$\begin{aligned}
\frac{dG_n(\phi, b_m, r_m)}{db_p} &= \frac{\partial G_n}{\partial b_p} + \frac{\partial G_n}{\partial b_m}\frac{\partial b_m}{\partial b_p}, \\
\frac{dG_n(\phi, b_m, r_m)}{db_{pm}} &= \frac{\partial G_n}{\partial b_m}\frac{\partial b_m}{\partial b_p}, \\
\frac{dG_n(\phi, b_p, r_p)}{db_{pm}} &= \frac{dG_n}{d\phi}\frac{\partial \phi}{\partial b_{pm}}, \\
\frac{dG_n(\phi, b_m, r_m)}{d\theta} &= \frac{\partial G_n}{\partial b_m}\frac{\partial b_m}{\partial \theta}, \\
\frac{dG_n(\phi, b_p, r_p)}{d\theta} &= \frac{\partial G_n}{\partial b_m}\frac{\partial b_m}{\partial \theta}.
\end{aligned} \tag{5.16}$$

The derivatives of the angle ϕ which appear in Equations (5.12), (5.13), (5.14), and (5.15)

will be dealt with in section 5.5. The derivatives of b_m are

$$\begin{aligned} \frac{\partial b_m}{\partial \theta} &= \frac{b_p b_{pm} \sin\theta}{b_m}, \\ \frac{\partial b_m}{\partial b_{pm}} &= \frac{b_{pm} - b_p \cos\theta}{b_m}, \\ \frac{\partial b_m}{\partial b_p} &= \frac{b_p - b_{pm} \cos\theta}{b_m}, \end{aligned} \tag{5.17}$$

where b_m is computed using the formula

$$b_m = \sqrt{(b_p - b_{pm})^2 + 2b_p b_{pm}(1 - \cos\theta)}, \tag{5.18}$$

which is a re-arranged version of the familiar law of cosines constructed to eliminate round-off error (Kahan 2000).

5.4 Evaluating the Primitive Integrals

With the description of the primitive integral complete, we now evaluate the integral and its derivatives for each value of n, starting with $n = 0$; i.e. a uniformly emitting source.

5.4.1 Uniform Limb-Darkening

For the uniform term the primitive integral is

$$G_0(\phi, r, b) = \int_0^{\phi} \frac{r}{2}(r - b\cos\phi')d\phi', \tag{5.19}$$

which gives us

$$G_0(\phi, r, b) = \frac{r}{2}(r\phi - b\sin\phi). \tag{5.20}$$

This has the derivatives

$$\frac{\partial G_0}{\partial r} = r\phi - \frac{b}{2}\sin\phi, \tag{5.21}$$

$$\frac{\partial G_0}{\partial b} = -\frac{r}{2}\sin\phi, \tag{5.22}$$

$$\frac{\partial G_0}{\partial \phi} = \frac{r}{2}(r - b\cos\phi). \tag{5.23}$$

5.4.2 Linear Limb-Darkening

For the linear term ($n = 1$) the primitive integral is

$$G_1(\phi, r, b) = \frac{r}{3}\int_0^{\phi} \frac{1 - (\mu^2 - b^2 + 2brc_{\phi'})^{3/2}}{b^2 + r^2 - 2brc_{\phi'}}(r - bc_{\phi'})d\phi'. \tag{5.24}$$

When this integral is evaluated it yields a combination of incomplete elliptic integrals of the first, second, and third kind which we express in the form:

$$\begin{aligned} G_1(\phi, r, b) &= \alpha(r, b)E(s, k) + \beta(r, b)F(s, k) \\ &+ \gamma(r, b)\Pi(s, n, k) + R(\phi, r, b), \end{aligned} \tag{5.25}$$

where $s = s(\phi)$ is the amplitude of the incomplete elliptic integrals, $n = n(r, b)$ is the parameter, and $k = k(r, b)$ is the modulus, following the conventions of Byrd & Friedman (1954). The functions $\alpha(r, b)$, $\beta(r, b)$ and $R(\phi, r, b)$ are defined as

$$\alpha(r, b) = \begin{cases} \frac{1}{9}\sqrt{1 - (b - r)^2}\ (7r^2 + b^2 - 4) & b + r \le 1 \\ \frac{2}{9}\sqrt{br}\ (7r^2 + b^2 - 4) & b + r > 1, \end{cases} \tag{5.26}$$

$$\beta(r, b) = \begin{cases} \frac{r^4 + b^4 + r^2 - b^2(5 + 2r^2) + 1}{9\sqrt{1 - (b - r)^2}} & b + r \le 1 \\ -\frac{3 + 2r(b^3 + 5b^2 r + 3r(r^2 - 2) + b(7r^2 - 4))}{18\sqrt{br}} & b + r > 1, \end{cases} \tag{5.27}$$

$$\gamma(r,b) = \begin{cases} \frac{b+r}{3(b-r)} \frac{1}{\sqrt{1-(b-r)^2}} & b+r \leq 1 \\ \frac{b+r}{6(b-r)\sqrt{br}} & b+r > 1, \end{cases} \tag{5.28}$$

and

$$\begin{aligned} R(\phi,r,b) = \frac{\phi}{6} &- \frac{1}{3}\tan^{-1}\left(\frac{b+r}{b-r}\tan\left(\frac{\phi}{2}\right)\right) \\ &- \frac{2br}{9}\sin\phi\sqrt{1-b^2-r^2+2br\cos\phi}. \end{aligned} \tag{5.29}$$

The formula for $R(\phi,r,b)$ is the same in the $b+r>1$ case as in the $b+r<1$ case.

The amplitude, characteristic, and parameter of the incomplete elliptic integrals is given by:

$$s = \begin{cases} \frac{\phi}{2} & b+r \leq 1 \\ \sin^{-1}\left(\frac{2\sqrt{br}\sin(\phi/2)}{\sqrt{1-(r-b)^2}}\right) & b+r > 1 \end{cases}, \tag{5.30}$$

$$k^2 = \begin{cases} \frac{4br}{1-(r-b)^2} & b+r \leq 1 \\ \frac{1-(r-b)^2}{4br} & b+r > 1 \end{cases}, \tag{5.31}$$

and

$$n = \begin{cases} -\frac{4rb}{(b-r)^2} & b+r \leq 1 \\ 1-\frac{1}{(b-r)^2} & b+r > 1 \end{cases}. \tag{5.32}$$

When $b=0$ and $r=1$ the solution simplifies to $G_1(\phi,b,r) = \frac{\phi}{3}$, which is useful when integrating along the boundary of the star.

When ϕ corresponds to the intersection between the arc and the limb of the star, the amplitude of the incomplete elliptic integrals becomes $\pi/2$ and the solution can be computed

with complete rather than incomplete elliptic integrals. In this case the integral evaluates to

$$\begin{aligned} G_1(\phi, r, b) = & \quad \alpha(r,b)E(k) + \beta(r,b)F(k) + \gamma(r,b)\Pi(n,k) \\ + & \quad \frac{\phi}{6} - \frac{1}{3}\tan^{-1}\left(\frac{b+r}{b-r}\tan\left(\frac{\phi}{2}\right)\right), \end{aligned} \tag{5.33}$$

where α, β, and γ are given by the expressions for the $b + r > 1$ case and ϕ is the angle to the intersection of the arc with the star's limb.

Derivatives for the Linear Limb-Darkening Case

The partial derivatives of $G_1(\phi, r, b)$ with respect to b and r are

$$\begin{aligned} \frac{\partial G_1}{\partial r} &= u_r(r,b)E(s,k) + v_r(r,b)F(s,k) + p_r(\phi,r,b), \\ \frac{\partial G_1}{\partial b} &= u_b(r,b)E(s,k) + v_b(r,b)F(s,k) + p_b(\phi,r,b). \end{aligned} \tag{5.34}$$

The coefficients by which the elliptic integrals are multiplied are given by

$$u_r(r,b) = \begin{cases} 2r\sqrt{1-(b-r)^2} & r+b \le 1 \\ 4\sqrt{br^3} & r+b > 1 \end{cases}, \tag{5.35}$$

$$v_r(r,b) = \begin{cases} 0 & b+r \le 1 \\ -\sqrt{\frac{r}{b}}((b+r)^2-1) & b+r < 1 \end{cases}, \tag{5.36}$$

$$u_b(r,b) = \begin{cases} \frac{1-(b-r)^2}{3b}(b^2+r^2-1) & b+r \le 1 \\ \frac{2}{3}\sqrt{\frac{r}{b}}(b^2+r^2-1) & b+r > 1 \end{cases}, \tag{5.37}$$

$$v_b(r,b) = \begin{cases} \frac{b^4-(r^2-1)^2-2b^2(r^2+1)}{3b\sqrt{1-(b-r)^2}} & b+r \le 1 \\ -\frac{1}{3}\sqrt{\frac{r}{b}}((b+r)^2-1) & b+r > 1 \end{cases}, \tag{5.38}$$

and the functions $p_r(\phi, r, b)$ and $p_b(\phi, r, b)$ are given by

$$\begin{aligned} p_r(\phi, r, b) &= \frac{b \sin\phi}{3\xi}\left((1-\xi)^{3/2} - 1\right) \\ p_b(\phi, r, b) &= \frac{r \sin\phi}{3\xi}\left(1 - (1+\xi)\sqrt{1-\xi}\right) \end{aligned}$$

where $\xi = b^2 + r^2 - 2br\cos\phi$.

The partial derivative with respect to ϕ is

$$\frac{\partial G_1(\phi, r, b)}{\partial \phi} = \frac{1-(1-\xi)^{3/2}}{\xi}(r^2 - br\cos\phi). \tag{5.39}$$

5.4.3 Quadratic Limb-darkening

For the quadratic term ($n = 2$), the primitive integral is

$$\begin{aligned} G_2(\phi, r, b) = \int_0^{\phi} & \tfrac{r}{4} \quad [\ r^3 + b^2 r(2 + \cos(2\phi')) \\ & - \quad b(b^2 + 3r^2)c_{\phi'}\ ]\, d\phi', \end{aligned} \tag{5.40}$$

which evaluates to

$$\begin{aligned} G_2(\phi, r, b) = & \tfrac{r}{4} \quad [\ (2b^2 r + r^3)\phi \\ & + \quad (b^2 r\cos\phi - 3br^2 - b^3)\sin\phi\ ]\,. \end{aligned} \tag{5.41}$$

This has the derivatives

$$\begin{aligned} \frac{\partial G_2}{\partial r} &= \frac{b}{4}(b^2 + 9r^2 - 2br\cos\phi)\sin\phi + rb^2 + r^2, \\ \frac{\partial G_2}{\partial b} &= \frac{r}{4}br(4\phi + \sin(2\phi) - 3(b^2 + r^2)\sin\phi), \\ \frac{\partial G_2}{\partial \phi} &= \frac{r}{4}\left(2b^2 r + r^3 + b(br\cos(2\phi) - (b^2 + 3r^2)\cos\phi)\right). \end{aligned} \tag{5.42}$$

5.4.4 Polynomial Limb-darkening

We now give the general solution for polynomial limb-darkening laws of the form

$$I(r) = 1 - \sum_{i=1}^{N} c_i (1-\mu)^i, \tag{5.43}$$

where again $\mu = \sqrt{1-r^2} = \sqrt{1-x^2-y^2}$. We follow Agol et al. (2020) in writing this limb-darkening law in Green's basis as

$$I(r) = \mathbf{g}^{\mathrm{T}} \mathbf{u}, \tag{5.44}$$

where $\mathbf{g}$ is defined

$$\mathbf{g} = (1, \mu, 4\mu^2 - 2, 5\mu^3 - 3\mu, \ldots, (N+2)\mu^N - N\mu^{N-2})^{\mathrm{T}}, \tag{5.45}$$

and $\mathbf{u}$ is a vector with elements defined recursively by

$$u_n = \begin{cases} \frac{p_n}{n+2} + u_{n+2} & N \geq n \geq 2 \\ p_n + (n+2)u_{n+2} & n = 1, 0 \end{cases}, \tag{5.46}$$

with $u_{N+1} = g_{N+2} = 0$, and where the elements of $\mathbf{p}$ are related to the elements of $\mathbf{c} = (c_0, c_1, c_2, \ldots, c_N)$ by:

$$p_n = (-1)^{n+1} \sum_{m=0}^{n} \binom{m}{n} c_m, \tag{5.47}$$

with $c_0 = -1$. It is important to note that while this basis shares the first two terms with Equation (5.1), the quadratic and higher order terms are distinct. Therefore the solution for the quadratic term given above should not be re-used when computing light curves for higher order limb-darkening laws.

Our goal is now to compute the integral over each term in Equation (5.45) along an

arbitrary arc. The appropriate primitive integral in this case is

$$G_n(\phi, r, b) = \int_0^{\phi} r(r - bc_{\phi'})(1 - r^2 - b^2 + 2brc_{\phi'})^{n/2} d\phi', \tag{5.48}$$

for $n \geq 2$, which we can rewrite in the form

$$G_n(\phi, r, b) = (1 + r^2 - b^2)\mathcal{M}_n - \mathcal{M}_{n+2}, \tag{5.49}$$

where

$$\mathcal{M}_n = \int_0^{\phi} \frac{1}{2}(1 - r^2 - b^2 + 2br\cos\phi')^{n/2} d\phi'. \tag{5.50}$$

This integral obeys the recursion relation

$$\begin{aligned} n\mathcal{M}_n &= 2(n-1)(1 - b^2 - r^2)\mathcal{M}_{n-2} \\ &+ (n-2)(1 - (b-r)^2)((b+r)^2 - 1)\mathcal{M}_{n-4} \\ &+ 2br\sin\phi(1 - b^2 - r^2 + 2br\cos\phi)^{\frac{n}{2}-1}. \end{aligned} \tag{5.51}$$

Given the $\mathcal{M}_1$ through $\mathcal{M}_4$ we can then use this recursion relation to compute $\mathcal{M}_n$ for any n, which can then be plugged into Equation (5.49) to find the integral over any term in $\mathbf{g}$.

We can re-use $\mathcal{M}_n$ to write the derivatives of the polynomial terms:

$$\begin{aligned} \frac{\partial G_n}{\partial r} = \; & 2 \quad r\left[(2+n)\mathcal{M}_n - n\mathcal{M}_{n-2}\right] \\ & - \quad b\sin\phi(1 - b^2 - r^2 + 2br\cos\phi)^{\frac{n}{2}}, \end{aligned} \tag{5.52}$$

$$\begin{aligned} \frac{\partial G_n}{\partial b} = \; & \frac{n}{b} \quad \left[(r^2 + b^2)(\mathcal{M}_n - \mathcal{M}_{n-2}) + (r^2 - b^2)^2\mathcal{M}_{n-2}\right] \\ & - \quad r\sin\phi(1 - b^2 - r^2 + 2br\cos\phi)^{\frac{n}{2}}, \end{aligned}$$

and

$$\frac{\partial G_n}{\partial \phi} = r(r - b\cos\phi)(1 - r^2 - b^2 + 2br\cos\phi)^{\frac{n}{2}}.$$

Finally, we give the first four terms of the recursion which are necessary to compute subsequent terms:

$$\begin{aligned} \mathcal{M}_0 &= \frac{\phi}{2}, \\ \mathcal{M}_1 &= \alpha_1 E(s,k) + \beta_1 F(s,k), \\ \mathcal{M}_2 &= \frac{1}{2}(1 - b^2 - r^2)\phi + br\sin\phi, \\ \mathcal{M}_3 &= \alpha_3 E(s,k) + \beta_3 F(s,k) + R_3(\phi, r, b), \end{aligned} \tag{5.53}$$

where the coefficients for $\mathcal{M}_1$ are

$$\begin{aligned} \alpha_1 &= \begin{cases} \sqrt{1-(b-r)^2} & b + r \le 1 \\ 2\sqrt{br} & b + r > 1 \end{cases}, \\ \beta_1 &= \begin{cases} 0 & b + r \le 1 \\ \frac{1-(b+r)^2}{2\sqrt{br}} & b + r > 1 \end{cases}, \end{aligned} \tag{5.54}$$

and for $\mathcal{M}_3$,

$$\begin{aligned} \alpha_3 &= \begin{cases} \frac{4}{3}(1-b^2-r^2)\sqrt{1-(b-r)^2} & b + r \le 1 \\ \frac{8}{3}(1-b^2-r^2)\sqrt{br} & b + r > 1 \end{cases}, \\ \beta_3 &= \begin{cases} \frac{2r^2(b^2+1)-(b^2-1)^2-r^4}{3\sqrt{1-(b-r)^2}} & b + r \le 1 \\ \frac{1-(b+r)^2}{6\sqrt{br}}(3(1-b^2-r^2)-2br) & b + r > 1 \end{cases}. \end{aligned} \tag{5.55}$$

The function R_3 is given by

$$R_3(\phi, r, b) = \frac{2}{3} br \sin\phi \sqrt{1 - b^2 - r^2 + 2br\cos\phi}. \tag{5.56}$$

The definitions of k and s are the same as for the linear limb-darkening case and are given in Equations (5.30) and (5.31).

We do not yet include polynomial limb-darkening laws in `gefera`, as in most cases quadratic limb-darkening will be sufficiently accurate to model an observed light curve. As a result, we have not tested these equations for numerical stability to the extent that we have for uniform, linear, and quadratic expressions. Anyone interested in implementing this solution should therefore take care to check for stability and accuracy. Interested readers might find it useful to refer back to Agol et al. (2020) as a guide.

5.5 Computing ϕ

The angle ϕ is the one-sided integration limit as an angle along the arc, measured from the negative x-axis. It can take on values in the range $(-\pi, \pi)$. For $\phi < 0$ the integrals can be computed using the property $G(-\phi, r, b) = G(\phi, r, b)$. Depending on the geometry of the system, ϕ will represent either the intersection of the planet and moon, or the intersection of the planet or moon with the limb of the star. In the case that body we're integrating along the arc of does not intersect either the other body or the limb of the star, we will have $\phi = \pm\pi$. In this section we show how to compute ϕ when the endpoint of the arc is an intersection with the other transiting body or with the stellar limb.

5.5.1 Intersections between planet and moon

When the planet and moon overlap we need to compute four angles, these being the angles to the two points of intersection with respect to the centers of each body. Figure 5.6 shows the geometry of the planet/moon overlap. The two points of intersection with respect to the planet are measured from the line connecting the center of the star to the center of the

planet. These are given by

$$\begin{aligned} \phi_p^+ &= \theta + \phi_p, \\ \phi_p^- &= \theta - \phi_p, \end{aligned} \tag{5.57}$$

Similarly, the two points of intersection with respect to the moon are measured from the line connecting the center of the star to the center of the moon and are given by

$$\begin{aligned} \phi_m^+ &= \theta_m + \phi_m, \\ \phi_m^- &= \theta_m - \phi_m, \end{aligned} \tag{5.58}$$

where θ_m is the angle between the line connecting the centers of the planet and moon and the line connecting the centers of the star and moon, as seen in Figure 5.6. This angle is computed as

$$\theta_m = \text{Atan2}(b_p \sin\theta, b_{pm} - b_p \cos\theta).^3 \tag{5.59}$$

We use the method outlined in Agol et al. (2020) to compute ϕ_p and ϕ_m to high accuracy. First we compute a quantity we call Δ, which is four times the area of the triangle formed by r_p, r_m, and b_{pm}, using the modified version of Heron's formula by Kahan (2000):

$$\Delta = \sqrt{(a + (b + c))(c - (a - b))(c + (a - b))(a + (b - c))}, \tag{5.60}$$

where a, b and c are r_p, r_m, and b_{pm} ordered from highest to lowest. In this formula the exact placement of the parentheses is essential to preserving numerical accuracy and should

[3]The function Atan2(y, x) is equivalent to Atan(y/x). It is available as a built-in math function in most programming languages. We have found that this alternative formulation of the arctangent function is more numerically stable in the limit $y \gg x$.

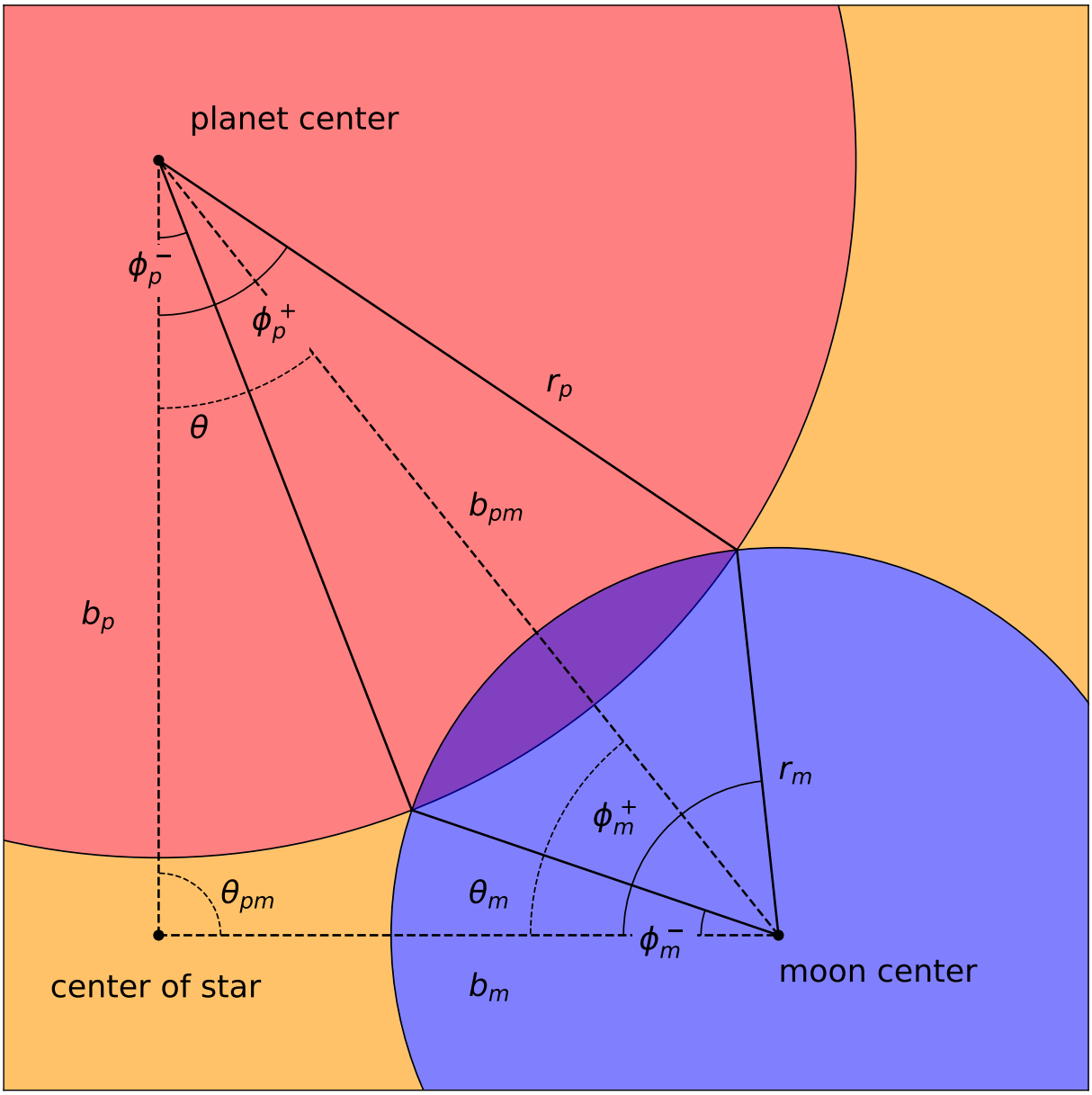

Figure 5.6 Geometry of the planet-moon overlap showing the angles $\phi_p^{\pm}$ and $\phi_m^{\pm}$ from Equations (5.57) and (5.58). The dashed lines show the triangle formed by the sides b_p, b_m and b_{pm}.

not be altered. The angles are then given by

$$\begin{aligned}\phi_p &= \text{Atan2}(\Delta, (r_p - r_m)(r_p + r_m) + b_{pm}^2),\\ \phi_m &= \text{Atan2}(\Delta, (r_m - r_p)(r_m + r_p) + b_{pm}^2).\end{aligned} \tag{5.61}$$

For the derivatives of ϕ_p and θ we have

$$\begin{aligned}\frac{\partial \phi_p}{\partial r_p} &= \frac{(b_{pm} - r_p)(b_{pm} + r_p) - r_m^2}{r_p \Delta},\\ \frac{\partial \phi_p}{\partial r_m} &= \frac{2r_m}{\Delta},\\ \frac{\partial \phi_p}{\partial b_{pm}} &= \frac{(r_p - r_m)(r_p + r_m) - b_{pm}^2}{b_{pm}\Delta},\\ \frac{\partial \phi_p}{\partial b_p} &= \frac{\partial \phi_p}{\partial \theta} = 0.\end{aligned} \tag{5.62}$$

The derivatives of θ are all equal to zero except for the derivative of θ with respect to itself, which is 1. For ϕ_m and θ_m we have

$$\begin{aligned}\frac{\partial \phi_m}{\partial r_p} &= \frac{2r_p}{\Delta},\\ \frac{\partial \phi_m}{\partial r_m} &= \frac{(b_{pm} - r_m)(b_{pm} + r_m) - r_p^2}{r_m \Delta},\\ \frac{\partial \phi_m}{\partial b_{pm}} &= \frac{(r_m + b_{pm})(r_m - b_{pm}) - r_p^2}{b_{pm}\Delta},\\ \frac{\partial \phi_m}{\partial b_p} &= \frac{\partial \phi_m}{\partial \theta} = 0,\end{aligned} \tag{5.63}$$

$$\begin{aligned}
\frac{\partial\theta_m}{\partial r_p} &= \frac{\partial\theta}{\partial r_m} = 0, \\
\frac{\partial\theta_m}{\partial b_p} &= \frac{b_{pm}\sin\theta}{b_m^2}, \\
\frac{\partial\theta_m}{\partial b_{pm}} &= -\frac{b_p\sin\theta}{b_m^2}, \\
\frac{\partial\theta_m}{\partial\theta} &= \frac{(b_{pm}-b_m)(b_{pm}+b_m)-b_p^2}{2b_m^2}.
\end{aligned}$$

5.5.2 Intersections between the planet/moon and stellar limb

When the planet or moon intersects the limb of the star we again must compute four angles. The first two are the angles to the two points of intersection with respect to the planet or moon, which are used to integrate along the arc of that body, and the second two are the points of intersection with respect to the center of the star which are used to integrate along the limb of the star. The angles to the intersections with respect to the center of the planet or moon are represented by κ_p or κ_m, and those with respect to the center of the star are represented by κ_p^* or κ_m^*. Figure 5.7 illustrates this geometry. All of these angles are computed relative to the line connecting the center of the star to the body in question. For the planet we have

$$\begin{aligned}
\kappa_p &= \text{Atan2}(\Delta_p, (r_p-1)(r_p+1)+b_p^2) \\
\kappa_p^* &= \text{Atan2}(\Delta_p, (1-r_p)(1+r_p)+b_p^2)
\end{aligned} \tag{5.64}$$

where Δ_p is given by equation (5.60) with a, b and c equal to 1, r_p, and b_p ordered from highest to lowest, and for the moon we have

$$\begin{aligned}
\kappa_m &= \text{Atan2}(\Delta_m, (r_m-1)(r_m+1)+b_m^2) \\
\kappa_m^* &= \text{Atan2}(\Delta_m, (1-r_m)(1+r_m)+b_m^2)
\end{aligned} \tag{5.65}$$

where Δ_m is again given by (5.60) but with a, b and c equal to 1, r_m, and b_m ordered from highest to lowest. The geometry of the intersection between star and transiting body is shown in figure 5.7

For the derivatives we have

$$\begin{aligned}
\frac{\partial \kappa_p}{\partial r_p} &= \frac{(b_p + r_p)(b_p - r_p) - 1}{r_p \Delta_p}, \\
\frac{\partial \kappa_p}{\partial r_m} &= \frac{\partial \kappa_p}{\partial b_{pm}} = \frac{\partial \kappa_p}{\partial \theta} = 0, \\
\frac{\partial \kappa_p}{\partial b_p} &= \frac{(r_p + b_p)(r_p - b_p) - 1}{b_p \Delta_p},
\end{aligned} \tag{5.66}$$

$$\begin{aligned}
\frac{\partial \kappa_p^*}{\partial r_p} &= \frac{2 r_p}{\Delta_p}, \\
\frac{\partial \kappa_p^*}{\partial r_m} &= \frac{\partial \kappa_p^*}{\partial b_{pm}} = \frac{\partial \kappa_p^*}{\partial \theta} = 0, \\
\frac{\partial \kappa_p^*}{\partial b_p} &= \frac{(1 + b_p)(1 - b_p) - r_p^2}{b_p \Delta_p},
\end{aligned} \tag{5.67}$$

$$\begin{aligned}
\frac{\partial \kappa_m}{\partial r_p} &= 0, \\
\frac{\partial \kappa_m}{\partial r_m} &= \frac{(b_m + r_m)(b_m - r_m) - 1}{r_m \Delta_m}, \\
\frac{\partial \kappa_m}{\partial b_p} &= \frac{(r_m + b_m)(r_m - b_m) - 1}{b_m^2 \Delta_m}(b_p - b_{pm} \cos\phi), \\
\frac{\partial \kappa_m}{\partial b_{pm}} &= \frac{(r_m + b_m)(r_m - b_m) - 1}{b_m^2 \Delta_m}(b_{pm} - b_p \cos\theta), \\
\frac{\partial \kappa_m}{\partial \theta} &= \frac{(r_m + b_m)(r_m - b_m) - 1}{b_m^2 \Delta_m} b_p b_{pm} \sin\theta,
\end{aligned} \tag{5.68}$$

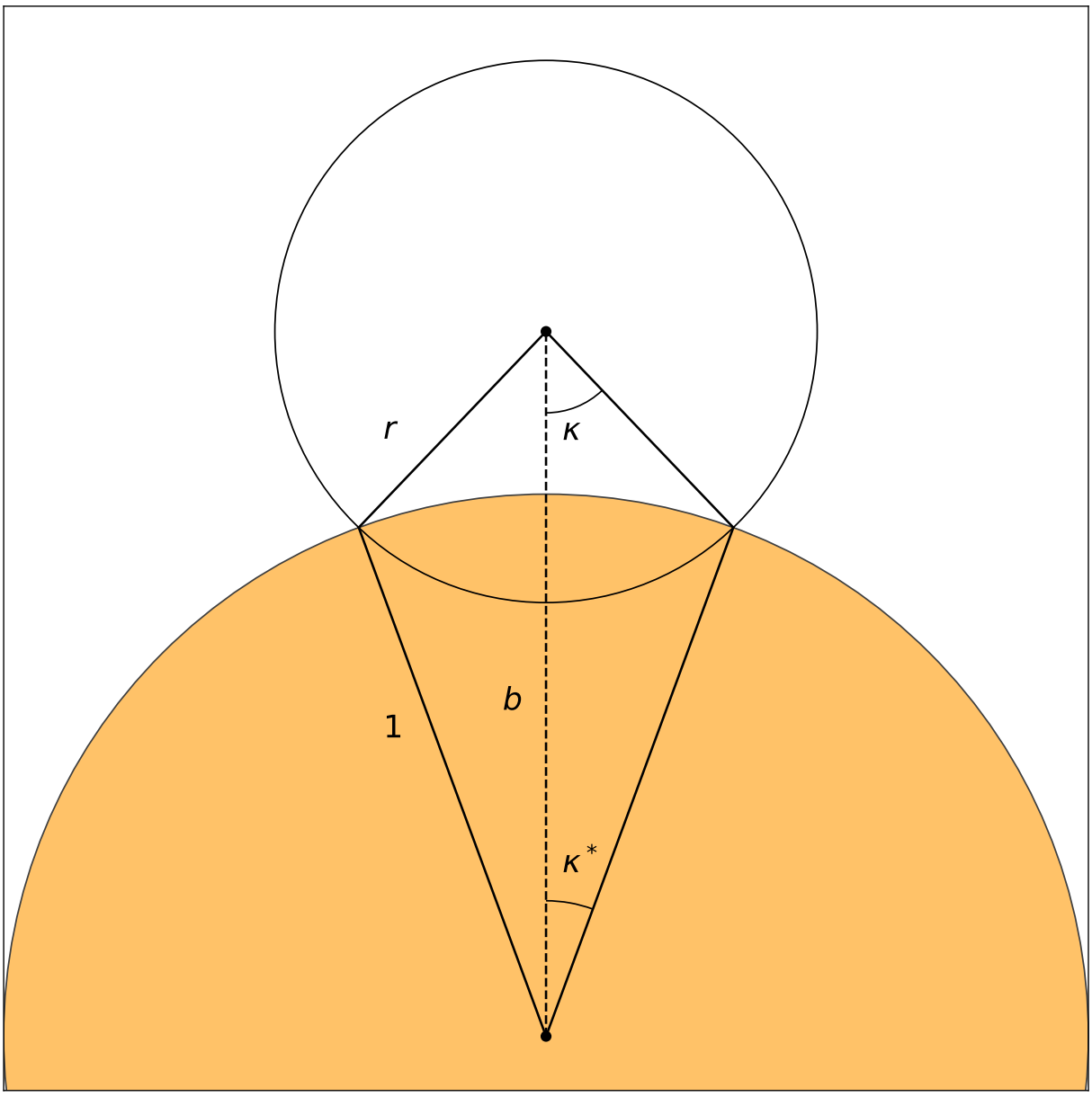

Figure 5.7 Geometry of the overlap between the star and a transiting planet or moon showing the angles κ and κ^* defined in Equations (5.64) and (5.65). Here the filled yellow circle represents the star and the smaller un-filled circle represents either the planet or the moon. In the text κ and κ^* are subscripted with m or p for the moon and planet respectively.

and

$$\begin{aligned}
\frac{\partial \kappa_m^*}{\partial r_p} &= 0, \\
\frac{\partial \kappa_m^*}{\partial r_m} &= \frac{2r_m}{\Delta_m}, \\
\frac{\partial \kappa_m^*}{\partial b_p} &= \frac{(1+b_m)(1-b_m)-r_m^2}{b_m^2 \Delta_m}(b_p - b_{pm}\cos\phi), \\
\frac{\partial \kappa_m^*}{\partial b_{pm}} &= \frac{(1+b_m)(1-b_m)-r_m^2}{b_m^2 \Delta_m}(b_{pm} - b_p\cos\theta), \\
\frac{\partial \kappa_m^*}{\partial \theta} &= \frac{(1+b_m)(1-b_m)-r_m^2}{b_m^2 \Delta_m} b_p b_{pm}\sin\theta.
\end{aligned} \tag{5.69}$$

5.6 Finding the Integration Limits

5.6.1 No planet-moon overlap

When $b_{pm} > b_p + b_m$ the planet and moon do not overlap each other. In this case we integrate counter-clockwise around the region of overlap between the transiting body and the star, and then to subtract this from the total flux of the star. For each body we determine whether or not it is entirely outside the perimeter of the star ($b > 1 + r$), overlapping the limb of the star ($1 - r < b < 1 + r$), or completely overlapping the star ($b < 1 - r$). In the first case the flux from the star is unobscured. In the second case we integrate from κ to $-\kappa$ along the boundary of the transiting body, where κ is defined in Equation (5.65) for the moon and Equation (5.64) for the planet. To this we add the integral along the limb of the star from $\kappa^* - \pi$ to $\pi - \kappa^*$ where κ^* is defined alongside κ in Equations (5.65) and (5.64). For each body we can write an expression for the obscured flux:

$$\begin{aligned}
F_{\text{obscured}} &= \sum_{n=0}^{N} u_n [G_n(\kappa, r, b) + G_n(-\kappa, r, b) \\
&+ G_n(\kappa^* - \pi, 1, 0) + G_n(\pi - \kappa^*, 1, 0)] \\
&= 2\sum_{n=0}^{2} u_n \left[G_n(\pi - \kappa^*, 1, 0) - G_n(\kappa, r, b)\right]
\end{aligned} \tag{5.70}$$

where we have used the property that $G_n(-\phi, r, b) = G_n(\phi, r, b)$ to simplify the expression. The expression for the flux received from the star can now be given as

$$F = F_0 - F_{\text{obscured}} \tag{5.71}$$

where F_0 is the unobscured stellar flux and is given by Equation (5.77). This solution for non-overlapping bodies is equivalent to the solution given in Agol et al. (2020).

5.6.2 Planet-moon overlap

When the planet and moon overlap each other there are a number of possible geometries that can occur, each of which require us to compute integrals along a different set of arcs. Figure 5.4 is a flowchart demonstrating how we identify the correct geometry given the radii r_p and r_m, radial distances b_p and b_m and the planet-moon distance b_{pm}. The outcomes in the flowchart, represented by the white boxes, indicate the corresponding system configuration shown in Figure 5.1. Table 5.6.3 gives the formula for the flux in each configuration. In Figure 5.4 and Table 5.6.3 the quantities δ_1 and δ_2 represent the distance from the center of the star to each of the planet-moon intersections. These are given by

$$\begin{aligned} \delta_1^2 &= r_m^2 + b_m^2 - 2 r_m b_m \cos\phi_m^+, \\ \delta_2^2 &= r_m^2 + b_m^2 - 2 r_m b_m \cos\phi_m^-. \end{aligned} \tag{5.72}$$

where the angles $\phi_m^\pm$ are defined in Equation (5.58). The angle θ_{pm} is the angle between the line connecting the center of the star to the planet, and the line connecting the center of the star to the moon. It is computed as

$$\theta_{pm} = \text{Atan2}(\Delta, (b_m - b_{pm})(b_m + b_{pm}) + b_p^2), \tag{5.73}$$

where Δ is computed from Equation (5.60) with a, b, c equal to b_p, b_m, and b_{pm} in order from highest to lowest. The derivatives of this angle are

$$\frac{\partial \theta_{pm}}{\partial b_p} = -\frac{b_{pm} \sin\theta}{b_m^2}, \tag{5.74}$$

$$\frac{\partial \theta_{pm}}{\partial b_{pm}} = (b_p \cos\theta - b_{pm}) \frac{b_{pm}}{b_m^2}, \tag{5.75}$$

$$\frac{\partial \theta_{pm}}{\partial \theta} = \frac{b_p \sin\theta}{b_m^2}. \tag{5.76}$$

The factor of b_m^2 in the denominator of these expressions may cause concern when b_m approaches zero. This singularity occurs because the angle θ_{pm} changes by $\pi/2$ instantaneously when the moon crosses the center of the star. We have no reason to expect any of the derivatives of the flux to have a discontinuity here, however, which points to the fact that this singularity must not exist in the derivatives of the flux itself, but is rather an artifact of our specific formulation. In appendix C.1 we demonstrate that in the limit $b_m \to 0$ two factors of b_m arise in the numerator of the ϕ-derivatives of the flux which cancel out with the factors of b_m in the denominator such that the limit of the derivative as $b_m \to 0$ is well-defined.

5.6.3 Normalization

In most cases we are interested in computing the normalized light curve where the unobscured flux is set to 1 and the light curve is interpreted as a measure of the fraction of the total stellar flux observed as a function of time. Equation (5.71) and the entries in table 5.6.3 include F_0, which we define as the unnormalized stellar flux. To find the normalized light curve we simply divide out F_0, which can be found by integrating the primitive integrals along the boundary of the star from $-\pi$ to π. This integral evaluates to

$$F_0 = u_0 \pi + u_1 \frac{2\pi}{3} + u_2 \frac{\pi}{2}. \tag{5.77}$$

for both the quadratic and polynomial limb-darkening laws.

Table 5.1. Expressions for the total flux in each of 16 geometric cases corresponding to the diagrams in Figure 5.1. To use this table, first consult Figure 5.4 to determine the correct case, then look up the flux for that case. The expressions in this table are unnormalized, but can be normalized to an out-of-transit flux of 1 by dividing out F_0 which is given in Equation (5.77).

Case Label	Flux
A	F_0
B	$2\sum_{n=0}^{N} u_n[G_n(\pi-\kappa_m^*,1,0)+G_n(\kappa_m,r_m,b_m)]$
C	$2\sum_{n=0}^{N} u_n[G_n(\pi-\kappa_p^*,1,0)+G_n(\kappa_p,r_p,b_p)]$
D	$F_0-2\sum_{n=0}^{N} u_n G_n(\pi,r_p,b_p)$
E	$F_0-\sum_{n=0}^{N} u_n[(G_n(\phi_p^-,r_p,b_p)+G_n(\phi_p^+,r_p,b_p))-(G_n(\phi_m^-,r_m,b_m)-G_n(\phi_m^+,r_m,b_m))]$
F	$\sum_{n=0}^{N} u_n[2G_n(\kappa_m^*,1,0) - (G_n(\phi_m^-,r_m,b_m)-G_n(-\kappa_m,r_m,b_m)) - (G_n(\kappa_m,r_m,b_m)-G_n(\phi_m^+,r_m,b_m)) - (G_n(\phi_p^-,r_p,b_p)-G_n(\phi_p^+,r_p,b_p))]$
G	$2\sum_{n=0}^{N} u_n[G_n(\pi-\kappa_p^*,1,0)+G_n(\kappa_p,r_p,b_p)]$
H	$\sum_{n=0}^{N} u_n[2G_n(\pi-\kappa_p^*,1,0) - (G_n(\phi_p^-,r_p,b_p)-G_n(-\kappa_p,r_p,b_p)) - (G_n(\kappa_p,r_p,b_p)-G_n(\phi_p^+,r_p,b_p)) - (G_n(\phi_m^-,r_m,b_m)-G_n(\phi_m^+,r_m,b_m))]$
I	$2\sum_{n=0}^{N} u_n[G_n(\pi-\kappa_p^*,1,0)+G_n(\kappa_p,r_p,b_p)]$
J	$2\sum_{n=0}^{N} u_n[G_n(\pi-\kappa_p^*,1,0)+G_n(\kappa_p,r_p,b_p)]$

Table 5.1 (cont'd)

Case Label	Flux
K	$\sum_{n=0}^{N} u_n[2G_n(\kappa_p^*,1,0) - (G_n(\phi_p^-,r_p,b_p) - G_n(-\kappa_p,r_p,b_p)) - (G_n(\kappa_p,r_p,b_p) - G_n(\phi_p^+,r_p,b_p)) - (G_n(\phi_m^-,r_m,b_m) - G_n(\phi_m^+,r_m,b_m))]$
L	$\sum_{n=0}^{N} u_n[2G_n(\kappa_m^*,1,0) - (G_n(\phi_m^-,r_m,b_m) - G_n(-\kappa_m,r_m,b_m)) - (G_n(\kappa_m,r_m,b_m) - G_n(\phi_m^+,r_m,b_m)) - (G_n(\phi_p^-,r_p,b_p) - G_n(\phi_p^+,r_p,b_p))]$
M	$2\sum_{n=0}^{N} u_n[G_n(\pi-\kappa_m^*,1,0) + G_n(\kappa_m,r_m,b_m)]$
N	$2\sum_{n=0}^{N} u_n[G_n(\pi-\kappa_p^*-\kappa_m^*,1,0) - G_n(\kappa_p,r_p,b_p) - G_n(\kappa_m,r_m,b_m)]$
O	$\sum_{n=0}^{N} u_n[2G_n(\pi-\kappa_p^*-\kappa_m^*,1,0) - (G_n(\phi_m^+,r_m,b_m) - G_n(-\kappa_m,r_m,b_m)) - (G_n(\kappa_m,r_m,b_m) - G_n(-\phi_m^-,r_m,b_m)) - (G_n(\kappa_p,r_p,b_p) - G_n(\phi_p^+,r_p,b_p)) - (G_n(\phi_p^-,r_p,b_p) - G_n(-\kappa_p,r_p,b_p))]$
P	$\sum_{n=0}^{N} u_n[2G_n(\pi-(\kappa_p^*+\kappa_m^*+\theta_{pm})/2,1,0) - (G_n(\kappa_m,r_m,b_m) - G_n(-\phi_m^-,r_m,b_m)) - (G_n(\phi_p^-,r_p,b_p) - G_n(-\kappa_p,r_b,b_p))]$

5.7 Implementation Details

5.7.1 Keplerian Dynamics

We implement two separate dynamics models which apply to different physical systems in which mutual transits might be observed. The first is a hierarchical model in which a primary body is orbited by a secondary body which together orbit the central mass. This model is applicable to exomoons, binary planets, and hierarchical triple-star systems. The second is a confocal model in which two bodies independently orbit a central mass, which is applicable to multi-planet systems in which multiple planets transit the host star. We

describe each of these models in detail below. For both models we solve Kepler's equation using the solver from Foreman-Mackey et al. (2021), which is itself based on the Kepler solver described in Raposo-Pulido & Peláez (2017). We propagate derivatives of the mean and eccentric anomalies with respect to each of the input parameters through the formula for the Cartesian coordinates of each body using the chain rule. For both types of system our dynamics module outputs b_p, b_{pm}, θ, and the derivatives of each with respect to all input parameters. These coordinates are then used as input to the photometry module. The derivatives of the flux with respect to each of the input parameters are then given by the chain rule as

$$\frac{dF}{dq_i} = \frac{\partial F}{\partial b_p}\frac{db_p}{dq_i} + \frac{\partial F}{\partial b_{pm}}\frac{db_{pm}}{dq_i} + \frac{\partial F}{\partial \theta}\frac{d\theta}{dq_i} \tag{5.78}$$

where q_i is the i^{th} input parameter to the Kepler module.

Hierarchical Systems

For hierarchical systems we adopt a simplified model of the dynamics in which we assume the system consists only of the star, planet, and moon. We further assume no interaction between the moon and the star. This simplifies the computation significantly because it allows us to model two Keplerian systems, one consisting of the star and the planet/moon combined mass and the other consisting of the planet and moon separately, rather than requiring us to simulate the three-body dynamics of the system.

The approximation of the system as a hierarchy of two-body Keplerian systems is valid in the case that the moon orbits well within the planet's Hill sphere. Because a moon orbiting too close to the outer edge of the Hill sphere is unstable over long timescales we would expect that any exomoon we observe would meet this criteria. Our approximation excludes non-Keplerian forces that occur in three-body systems, so care should be taken in cases where the precession rate is expected to be high relative to the duration of observations. The nodal precession period due to three-body moon-planet-star interactions is given by Martin et al.

(2019) as

$$P_{\rm precess} = \frac{4}{3}\frac{M_p + M_*}{M_*}\frac{P_p^2}{P_m}\frac{1}{\cos i_m}. \tag{5.79}$$

If we assume that the mass of the planet is negligible compared to that of the star, then we can use the following criterion to determine when precession is not expected to impact our ability to model the system:

$$\frac{P_{\rm precess}}{P_p} \approx \frac{4}{3}\frac{P_p}{P_m}\frac{1}{\cos i_m} \gg 1. \tag{5.80}$$

When this criterion is not met we may observe nodal precession on the timescale of the planetary period, causing the parameters of the moon's orbit to change from transit to transit and thus invalidating our dynamical model. In this case the user may wish to provide a dynamics module of their own which takes into account three-body interactions.

Confocal Systems

For confocal systems we assume no interactions between the two bodies and solve Kepler's equation for each body separately. For the first body we set the longitude of the ascending node, Ω, equal to 180 degrees following Winn (2010). For the second body we allow Ω to vary since this gives us the mutual inclination between the two orbits.

Because we assume no interaction our code is not able to model transit timing variations due to planet-planet gravitational interactions, though it does include timing variations of the outer planet due to the motion of the star in its orbit about the mutual center of gravity with the inner planet. However, since our primary purpose is to model mutual transits and the timing variation due to the other body is small when the bodies transit simultaneously, this should not be a problem in most cases. One scenario in which this might present an issue is when other planets are present in the system and contribute to transit timing variability.

5.7.2 Photometric Model

Given the positions of each of the two occulting bodies at a point in time we use the decision tree in Figure 5.4 to identify the appropriate geometric case. This determines the formula for the flux integral which can be looked up in Table 5.6.3.

The linear and polynomial terms of the flux integral include incomplete elliptic integrals. Instead of computing these integrals directly, which can result in numerical instabilities for some values of the inputs, we compute them in terms of the Bulirsch elliptic integrals $el1$, $el2$, and $el3$, which are related to the standard elliptic integrals F, E, and Π as follows:

$$\begin{aligned} F(\phi, k) &= el1(x, k_c), \\ E(\phi, k) &= el2(x, k_c, 1, k_c^2), \\ \Pi(\phi, n, k) &= el3(x, k_c, p), \end{aligned} \tag{5.81}$$

where $x = \arctan\phi$, $k_c = \sqrt{1-k^2}$, and $p = n + 1$ (Bulirsch 1965, 1969a). When derivatives are not required we use the following relationship to compute a linear combination of $F(\phi, k)$ and $E(\phi, k)$ with only one evaluation of $el2$ (Bulirsch 1969b):

$$\alpha E(\phi, k) + \beta F(\phi, k) = el2(x, k_c, \alpha + \beta, \beta + \alpha k_c^2). \tag{5.82}$$

When computing the derivatives of the model all three of the incomplete elliptic integrals are required in different linear combinations. This requires us to compute all three functions individually, but the integrals can be computed simultaneously for computational efficiency by using some components of the computation which are shared in common.

5.7.3 Exposure time integration

Telescope observations do not represent instantaneous snapshots of a system, but are instead integrated over the duration of the exposure. Failing to account for finite integration times can negatively affect our ability to measure parameters of the system (Kipping 2010). To

capture this effect, we include in our model an option to numerically integrate the light curve using either the trapezoidal rule or Simpson's rule.

For the trapezoidal rule with arbitrary exposure times we up-sample the light curve by a factor of two and compute the observed flux at a time t as

$$F_{\rm obs}(t) = \frac{1}{2}\left[F\left(t - \frac{\Delta t}{2}\right) + F\left(t + \frac{\Delta t}{2}\right)\right], \tag{5.83}$$

where Δt is the exposure time. This computation comes at the cost of twice the number of evaluations of the flux. However, for the common case where the exposure time is equal to the time between observations the start times and end times for adjacent exposures are the same. In this case we need only compute $N + 2$ flux evaluations where N is the number of input times and the 2 arises from the need to evaluate the flux at the beginning of the first integration and the end of the last integration. If derivatives are required these can be integrated in the same way as the flux. Because the trapezoidal rule computes the integral of a linear spline interpolation of the true light curve the result can be inaccurate at points where the derivative of the light curve with respect to time changes rapidly. This occurs at the contact points between the star and planet, star and moon, and planet and moon. A more accurate option is Simpson's rule. For Simpson's rule in the case of arbitrary exposure times we up-sample the light curve by a factor of three and compute the observed flux as

$$F_{\rm obs}(t) = \frac{1}{6}\left[F\left(t - \frac{\Delta t}{2}\right) + 4F\left(t\right) + F\left(t + \frac{\Delta t}{2}\right)\right]. \tag{5.84}$$

This costs three times the number of flux evaluations as the un-integrated light curve, and 50% more than for the trapezoidal rule, but it will be much more accurate than the trapezoidal rule for long integration times. Figure 5.8 shows the effect of exposure time integration for an exposure time of 0.03 days on a light curve identical to the light curve in Figure 5.2. The left panel of the figure shows the full light curve while the right panel zooms in on the point where the moon transits the planet to show the effect of the sharp transition in the

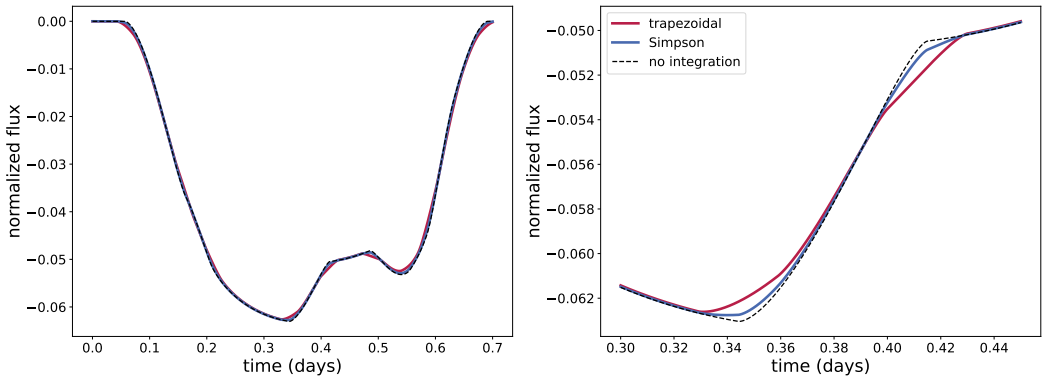

Figure 5.8 An exposure-time integrated light curve. The left panel shows the full light curve with the dashed line being the un-integrated curve. The red line is the integrated curve produced by the trapezoidal rule and the blue line is the integrated curve produced by Simpson's rule, both with an exposure time of 0.03 days. On the right we have zoomed in to the part of the transit where the moon crosses the limb of the planet. At this point the flux changes sharply which produces errors in the time-integration, which accounts for the difference between the trapezoidal curve and the Simpson's curve. Simpson's method is a higher order method and is more accurate, especially at these contact points.

flux on the integral produced by the two methods.

5.8 Speed

We have benchmarked `gefera` on an Intel Macbook Pro from early 2015 with a 3.1 GHz Dual-Core Intel Core i7 CPU. Because the number of integrals needed to compute the flux depends on the geometry of the planet and moon, the time to compute the model is dependent on the geometry of the transit. We have therefore benchmarked two models, a simple model and a complex model. The simple model simulates an event in which the planet and moon do not overlap each other at any point during the transit, whereas the complex model simulates a transit in which the moon and planet overlap while crossing the limb of the star. This is the most computationally expensive case, in which the full set of three incomplete elliptic integrals must be computed for as many as four separate arcs at some timesteps. In comparison the simple model will only require two complete elliptic integrals to be computed

for each timestep during which the moon and planet both overlap the star.

Figure 5.9 shows our benchmarks for a number of observations ranging from ten to one-million for both the simple (dashed lines) and complex (solid lines) models. In the left panel we compute our model without gradients, and in the right we compute the model with gradients. We also include benchmarks for the `photodynam` code (Pál 2012; Carter et al. 2011) which implements a similar method to our code but is written to be more general and can include any number of bodies (as opposed to `gefera` which only allows for two bodies in addition to the star). While `photodynam` does not include gradients, we have plotted the benchmarks in both panels for the sake of comparison. We find that `gefera` is about 22 times faster than `photodynam` without computing gradients for the simple model, and 10 times faster for the complex model without gradients. When we include the gradient computation, `gefera` is about 15 times faster than `photodynam` for the simple model and 7 times faster for the complex model. These results hold for all runs with more than about 100-1,000 timesteps. For very small numbers of timesteps `photodynam` performs similarly to, but still slightly slower than, `gefera`.

We benchmark `gefera` with and without the inclusion of the dynamics module. While the `photodynam` code does include dynamics, we have chosen to time only the photometry module. This is because the dynamics module provided with `photodynam` is not well-suited to reproducing the hierarchical sun-planet-moon system that we implement in `gefera` and as a result a comparison between the two would not be very informative. The benchmarks for `photodynam` should therefore be compared to the results for `gefera` without dynamics (the dark blue lines) rather than with dynamics.

We were unable to include benchmarks for `LUNA` as we do not have access to the fully optimized `LUNA` code. Our implementation of the `LUNA` algorithm was written in order to test accuracy, not speed, and is likely not as efficient as the original implementation. Therefore it does not seem fair or useful to provide a comparison.

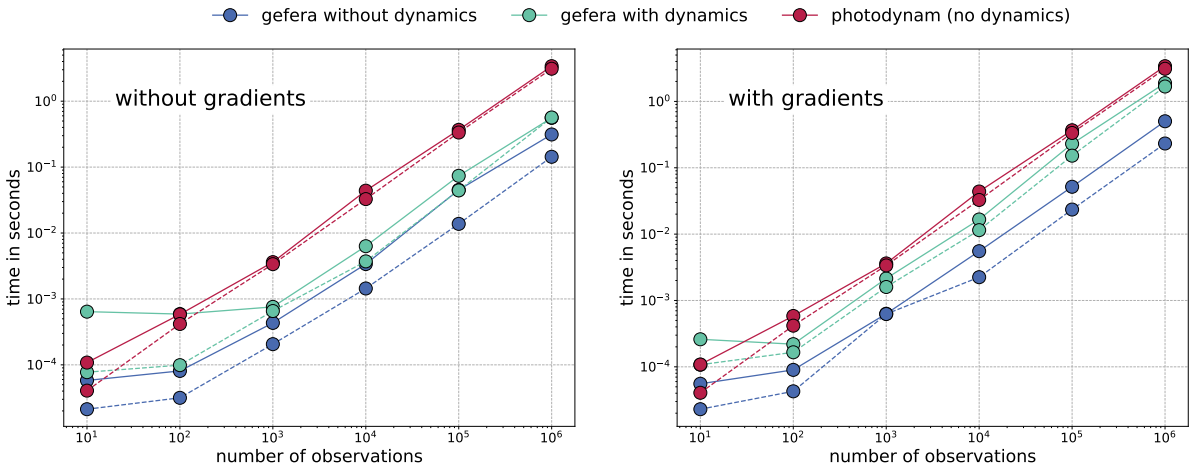

Figure 5.9 Benchmarks for `gefera` and `photodynam`. The right panel includes the gradient calculation whereas the left panel does not. Since `photodynam` does not include gradients these benchmarks are the same for both panels. Dashed lines indicate that a "simple" transit was simulated in which the star and planet do not overlap at any point during the transit. Solid lines indicate a "complex" model where the moon and planet overlap during their egress. For `gefera` we include benchmarks both with and without the inclusion of the dynamics module, whereas for `photodynam` we include benchmarks only for the photometry module.

5.9 Comparison to LUNA

LUNA (Kipping 2011) is an algorithm for computing exoplanet/moon transits which applies the small-planet approximation (Mandel & Agol 2002) for the flux blocked by the area of the moon's overlap with the star while using the full analytic transit solution for the area of the planet/star overlap. While the original implementation of LUNA has not been made publicly available, we have implemented our own version of the algorithm as published in Kipping (2011) for purposes of comparison.

As a test, we model the transit of an exoplanet and accompanying moon with parameters consistent with the parameters inferred for the exomoon candidate Kepler-1701b i as published in Kipping et al. (2022). We adopt the mean of the posterior distribution for the planet-star radius ratio, 0.0818 and the moon-planet radius ratio 0.263. We find a maximum discrepancy of about 4 ppm between LUNA and our model, which demonstrates the applicability of the small-moon approximation for this case. The first panel of Figure 5.10 shows the LUNA and gefera light curves with the absolute error between the two models. In the second panel we increase the radii of the planet and moon to 0.2 and 0.15 R_* respectively. Here the small-moon approximation begins to break down with the LUNA model differing by about 500 ppm from the gefera model. In the third panel we again increase the radii to 0.7 and 0.6 R_* respectively. While these parameters are unlikely to occur for a realistic binary planet system, it serves to demonstrate the breakdown of the small-planet approximation for large bodies.

We conclude that LUNA is likely sufficiently accurate for the vast majority of realistic planet/moon systems, but the difference between the approximate light curve computed by LUNA and the exact light curve computed by gefera may become important for large moons or binary planet systems, or in the case that very high precision photometry is available.

Another outcome of the use of the small-moon approximation in LUNA is that a LUNA light curve does not have a continuous derivative with respect to the position or radius of the moon. The derivatives of the LUNA model diverge at the contact points between the

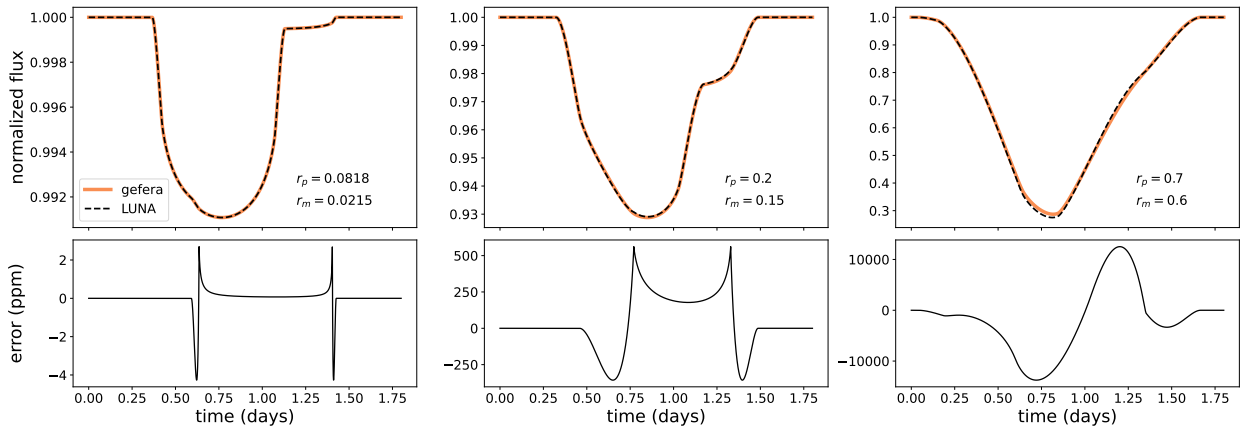

Figure 5.10 `LUNA` and `gefera` light curves and error in ppm, labeled by the radii of the planet and moon in units of the stellar radius. **Left:** Radii for the moon and planet are taken from the maximum likelihood solution for Kepler-1708 b i given in Kipping et al. (2022). In this case the difference between the `LUNA` solution and the `gefera` solution is only about four parts-per-million at most. **Center:** The same transit with a much larger planet and moon. Here the maximum difference between the two solutions is 500 parts-per-million. **Right:** The same transit again with a "planet" and "moon" approaching the radius of the star. These parameters may be reasonable for an eclipsing triple-star system. In this case the small-planet approximation breaks down with the error approaching 1%.

moon and the inner edge of the stellar limb. To illustrate this we compute the derivative of the `LUNA` and `gefera` models with respect to the period of the moon's orbit using finite differences using the parameters from the central panel of Figure 5.10. The upper panel of Figure 5.11 shows the derivative itself while the lower panel shows the difference between the `LUNA` and gefera derivatives. The configuration of the system at the points of divergence are illustrated in Figure 5.12. The lack of continuous derivatives may create issues for gradient-based inference algorithms such as Hamiltonian MCMC and any other analysis that requires the use of model derivatives.

We are unable to benchmark the computational speed of `LUNA` since we don't have access to the fully optimized version of the code, but we note that `gefera` is fast enough to enable MCMC inference for most datasets on a laptop computer, so we argue that any reduction in computational speed between `LUNA` and `gefera` is likely justified by the increase in accuracy. Additionally, since `gefera` provides derivatives of the transit it can be used with Hamiltonian MCMC which, carried out with No U-turn Sampling (NUTS), may be more efficient than standard MCMC.

One final note on `LUNA` is that in the process of implementing the algorithm we identified an error in equation 37 of Kipping (2011) that appears to have been carried over from an incorrect formula published in Fewell (2006) for the area of overlap between three circles in the case that the region of overlap includes more than half of the area of the smallest circle. In appendix C.2 we provide an explanation of the error and a corrected formula.

5.10 Applications

We envision our model being used to search for exomoons in observations from current and upcoming missions including JWST, TESS, and PLATO, as well as archival data from the Kepler and CoRoT missions. It also may be useful for analyzing simultaneous mutual transits of eclipsing binary stars and circumbinary planets or more complicated multi-star systems (Carter et al. 2011). In the next section we briefly discuss the options for conducting Bayesian inference using our model. Because we provide derivatives of the planet/moon transit, our

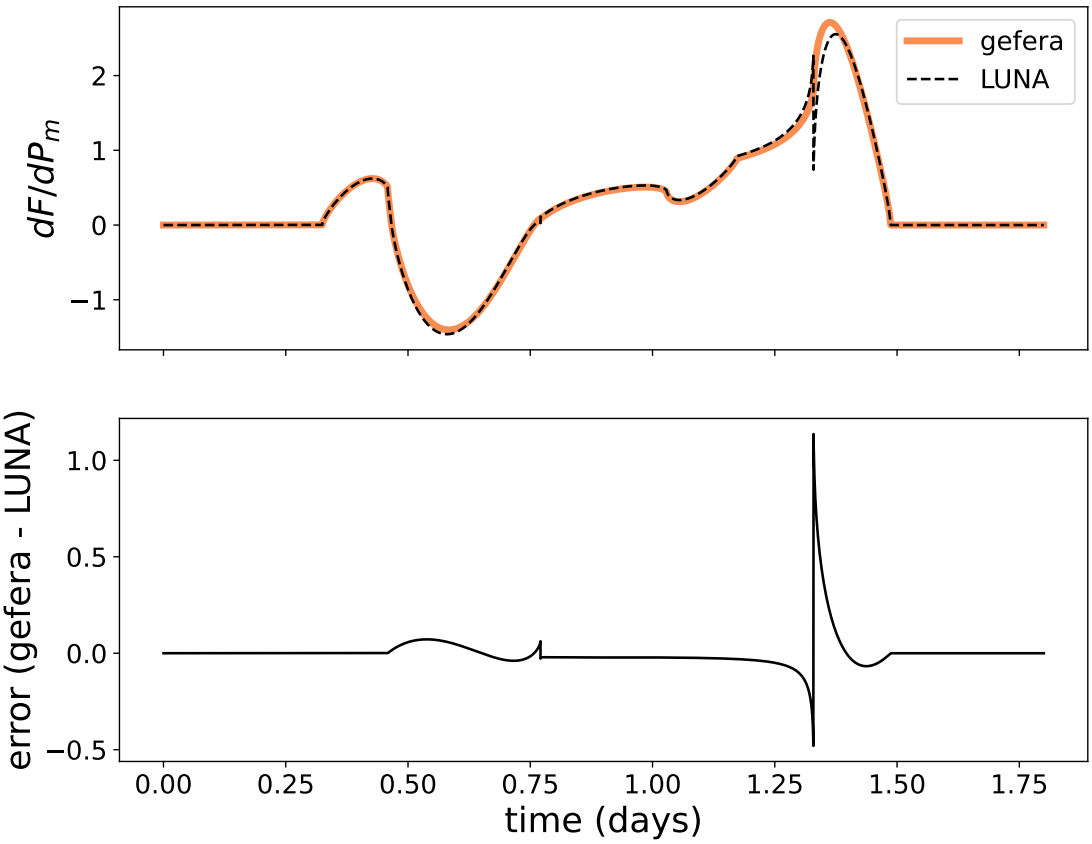

Figure 5.11 `LUNA` and `gefera` derivatives and the difference between them. The orbital parameters are identical to those used in Figure 5.10 and the radii of the planet and moon are $0.2R_*$ and $0.15R_*$ respectively, matching the center panel of Figure 5.10. The error in the derivative is larger than for the light curve itself, crossing the one percent threshold at the point where the moon contacts the inner edge of the stellar limb.

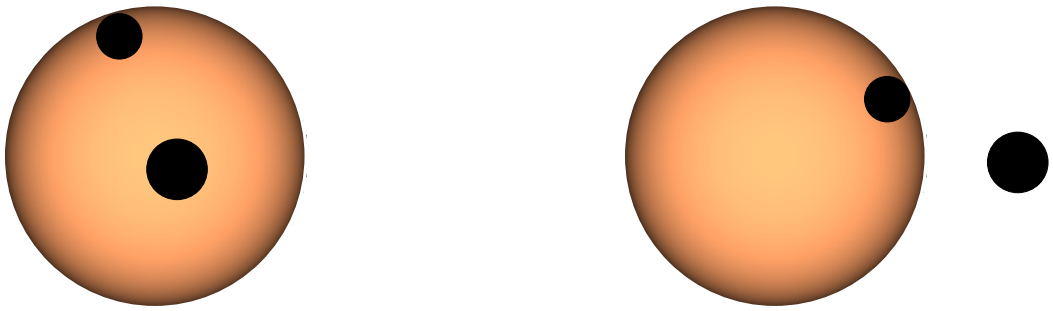

Figure 5.12 System configuration at the points where the derivative diverges in Figure 5.11. The left panel shows the first point of divergence and the right the second.

model also may be used to conduct an information analysis along the lines of Carter et al. (2008) or Price & Rogers (2014).

It is also our hope that providing a publicly available exomoon transit code will make it easier for the community to conduct independent analyses of exomoon candidates such as Kepler-1625b i (Teachey & Kipping 2018) and Kepler-1708b i (Kipping et al. 2022). These candidates raise the exciting possibility that exomoon detections lay within the grasp of existing datasets, but the lack of non-proprietary exomoon transit models has made it difficult for other groups to conduct their own searches or to follow up on existing candidates.

5.10.1 Inference with `gefera`

The `gefera` codebase is provided as a pip-installable python package. It is written to have a simple python interface which allows the user to define the orbit of the moon-planet system about the star and the orbit of the moon about the planet using standard keplerian

orbital parameters. The user can then compute a light curve with or without derivatives. This interface is useful both for quickly computing and plotting sample light curves as well as for plugging into inference packages for MCMC or nested sampling. With access to the derivatives users may also make use of gradient-based inference algorithms such as Hamiltonian Monte-Carlo with No-U-Turn Sampling (NUTS; Hoffman & Gelman 2014). In testing the code we found that the streamlined python interface used with `emcee` (Foreman-Mackey et al. 2019) is the simplest and most efficient way to conduct inference for a moon/planet transit model. However, in the future we intend to implement a set of `aesara` ops (Willard et al. 2021) which will allow `gefera` to be used with `pymc` (Salvatier et al. 2016). We will also explore the possibility of adding interfaces for other popular inference packages in future versions of our `gefera`. Making use of the more advanced gradient-based samplers provided by `pymc` may also be advantageous for larger models combining multiple planets or binary planet/moon systems in which the number of parameters can become very large.

5.10.2 Using `Gefera` to Model Binary or Triple-star Systems

We have briefly mentioned that `gefera` may be used to model stellar transits in binary or triple-star systems in addition to moon-planet and planet-planet transits. We now explain how `gefera` can be used to model such a transit.

When one of the transiting bodies emits its own flux, we must first consider whether that body is itself being transited (i.e. is the body emitting flux positioned between the two other bodies, or is it the nearest to the observer). If the second flux-emitting body is the closest body to the observer and is not itself being transited, then there is no need for a special procedure. This type of transit can be directly modeled in `gefera` using the same prescription outlined for planet-planet or planet-moon transits. However, when the second flux-emitting body is positioned between the two other bodies in the system, as in the case of a coplanar circumbinary system, we must consider the second transit of this luminous body by the frontmost transiting body. We can do this by first computing the mutual transit under the assumption that neither transiting body is luminous. To this flux we can add the

flux from a single transit between the second star and the front-most transiting body. The total flux can be expressed as

$$F_{\text{total}} = F_{\text{non-luminous}} + F_{\text{two bodies}} \tag{5.85}$$

where $F_{\text{non-luminous}}$ refers to the `gefera` model computed under the assumption that the two transiting bodies are non-luminous and $F_{\text{two bodies}}$ refers to the ordinary two-body transit between the second star and the front-most star or planet.

5.10.3 Test Case: Kepler-51 b & d

As a test of our code we have conducted a short re-analysis of a possible mutual transit event involving Kepler-51 b & d, previously studied by (Masuda 2014, hereafter M14). Mutual transit events are characterized by a "bump" – an increase in flux during the transit resulting from the decrease in the obstructed area of the stellar disk during the planet-planet overlap. Unfortunately bumps may also result from starspot crossings making the interpretation of such a feature ambiguous. Ultimately M14 concludes that a starspot crossing is the likely cause of this feature, basing their conclusion on several arguments. First, fitting the mutual transit event, M14 infers impact parameters for both planets near zero and a large mutual inclination of -25 degrees between the two orbits. The impact parameters conflict with the values inferred from other transits of planets b and d which are close to 0.25 for both planets, despite agreeing with the Kepler team's determination of $b = 0.03$ for planet b and $b = 0.061$ for planet d[4]. These large impact parameters reduce the likelihood of a mutual transit, while impact parameters close to zero increase the likelihood of the event. Second, large mutual inclination is a problem for the mutual transit interpretation as such a configuration is a priori unlikely to have both planets transit and would further reduce the chance of planet-planet overlap during the simultaneous transit.

M14 models the mutual transit bump by computing the area of overlap between the two

[4]Data retrieved from the NASA Exoplanet Archive (NASA Exoplanet Archive YYYY)

planets, then multiplying that by a factor related to the surface brightness of the star at the location of the overlap as specified by a quadratic limb-darkening law, and finally adding this flux back to the light curve (since it is double-subtracted when the planet-planet interaction is not accounted for). This is similar to the approach taken by Kipping (2011) based on the small-planet approximation of Mandel & Agol (2002), except that the approximation is applied only to the area of overlap rather than to the entirety of the smaller body.

We begin our own analysis by using the built-in methods from `lightkurve` (Lightkurve Collaboration et al. 2018) to detrend the short-cadence PDCSAP Kepler flux using a Savitzky-Golay filter with a polynomial of order 2 and a window length of 1.65 hours. We mask the in-transit points prior to detrending. We then use the truncated Newtonian minimization routine provided by `scipy` (Virtanen et al. 2020) to minimize a two-planet transit model with respect to the Keplerian orbital parameters $\{a, t_0, e, P, \omega, i\}$ for each body where a is the semimajor amplitude, t_0 is the reference epoch, e the eccentricity, P the period, ω the longitude of periapse, and i the inclination. In addition to these we include Ω, which is the mutual inclination between the two orbits, as well as the radii of both bodies and the two quadratic limb-darkening parameters (c_1 and c_2 in Equation (5.1)). The truncated Newtonian minimization routine accepts the derivatives of the likelihood with respect to the parameters which we find speeds up the minimization routine by a factor of 10 or more. The best-fit model is shown in Figure 5.13, and Figure 5.16 shows the best-fit transit as a series of snapshots forming a "movie" of the event. These snapshots can be compared to Figure 3 in M14, with the caveat that our snapshots are rotated counterclockwise through an angle Ω relative to the figure in M14. This rotation does not affect the appearance of the light curve.

We also demonstrate the use of our model to estimate the posterior distribution for the model given the Kepler observations with MCMC. We use `emcee` Foreman-Mackey et al. (2019) to run 1,000 walkers for 44,000 steps. Figure 5.14 shows a random selection of 100 light curves sampled from the posterior distribution of the model given the flux (here shown as blue points). Figure 5.15 shows the posterior distribution for the mutual inclination marginalized over all other parameters (right) alongside the joint distribution of the impact

parameters for each planet. Our results are consistent with M14. In particular we infer a mutual inclination of $-33.0^{+21.5}_{-29.5}$ deg compared to the M14 value of $-25.3^{+6.2}_{-6.8}$ degrees. Because M14 do not describe the details of their MCMC algorithm it is difficult to explain the discrepancy between our relatively large error bars and their small error bars. In particular, M14 does not discuss what priors (if any) were used in their analysis. It should be noted however that in contrast to M14 our posterior admits models with low (though not zero) mutual inclination.

Like M14 we find a low likelihood for the large impact parameters inferred from the phase-folded light curves of planets b and d. Figure 5.15 shows the joint posterior distribution for the impact parameters with the line $b_{K51b} = b_{K51d}$ and the values determined by the Kepler team. Since 2014, the Kepler-51 system has been observed by *HST* which has yielded additional estimates of the impact parameters for Kepler-51b & d (Libby-Roberts et al. 2020). We include these values in the figure as well, and note that these measurements are consistent with the impact parameters found by M14, and that the mean estimates of the impact parameters place both models in the low-probability region of the posterior. Figure 5.15 can be compared to Figure 4 in M14. The multiple modes seen in Figure 5.15 present an issue for the default `emcee` sampler that we used for this analysis as the chains have a low probability of crossing the low-likelihood valleys. As a result the chains tend to get stuck in the local high probability regions and do not fully mix between the modes. Given this, caution should be taken when interpreting our posteriors. In particular the apparent relative densities of each mode should not be taken to indicate a preference for one mode over another. We might partially rectify this limitation of our analysis by using Umbrella Sampling (Torrie & Valleau 1977; Gilbert 2022) to efficiently sample the low-likelihood regions between modes, or by employing an adaptive MCMC method designed for multi-modal likelihoods such as the Jumping Adaptive Multimodal Sampler (Pompe et al. 2018). Such an analysis is beyond the scope of this chapter but may be the subject of future work.

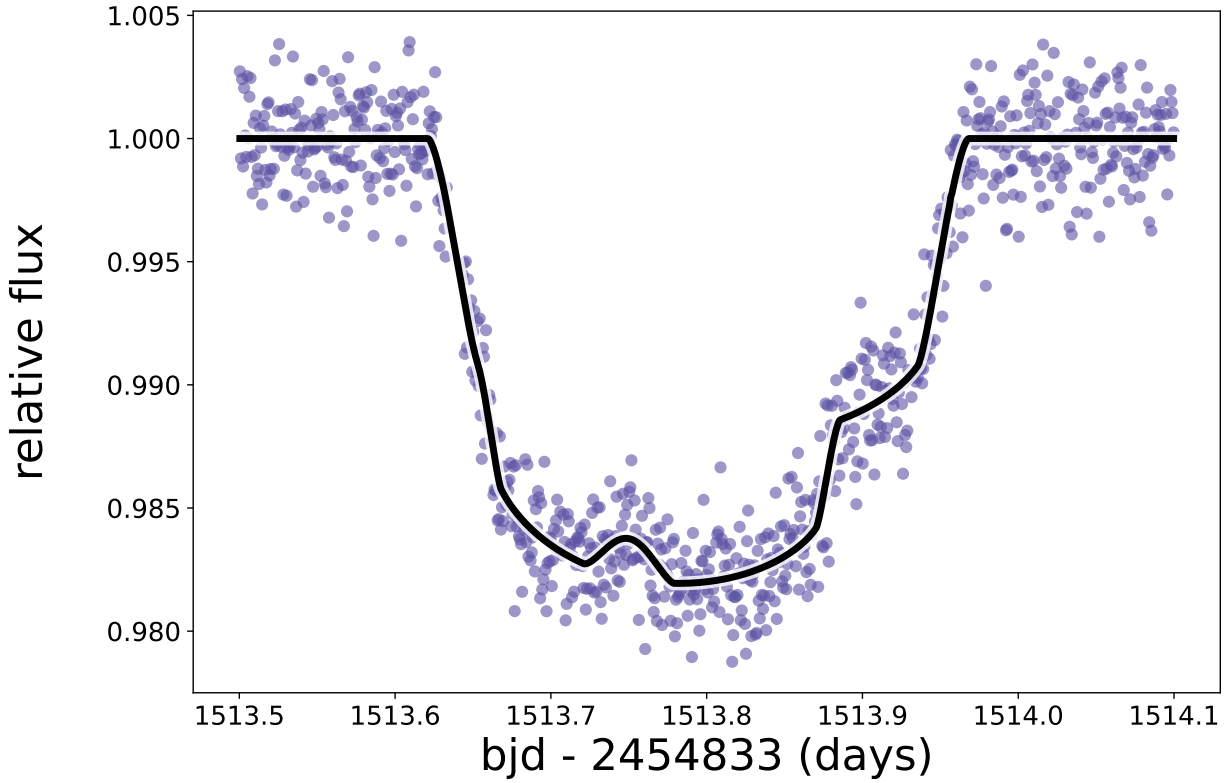

Figure 5.13 Best fit to the Kepler observations of the Kepler-51 system. The fit was found using the truncated Newtonian algorithm provided by `scipy` with derivatives.

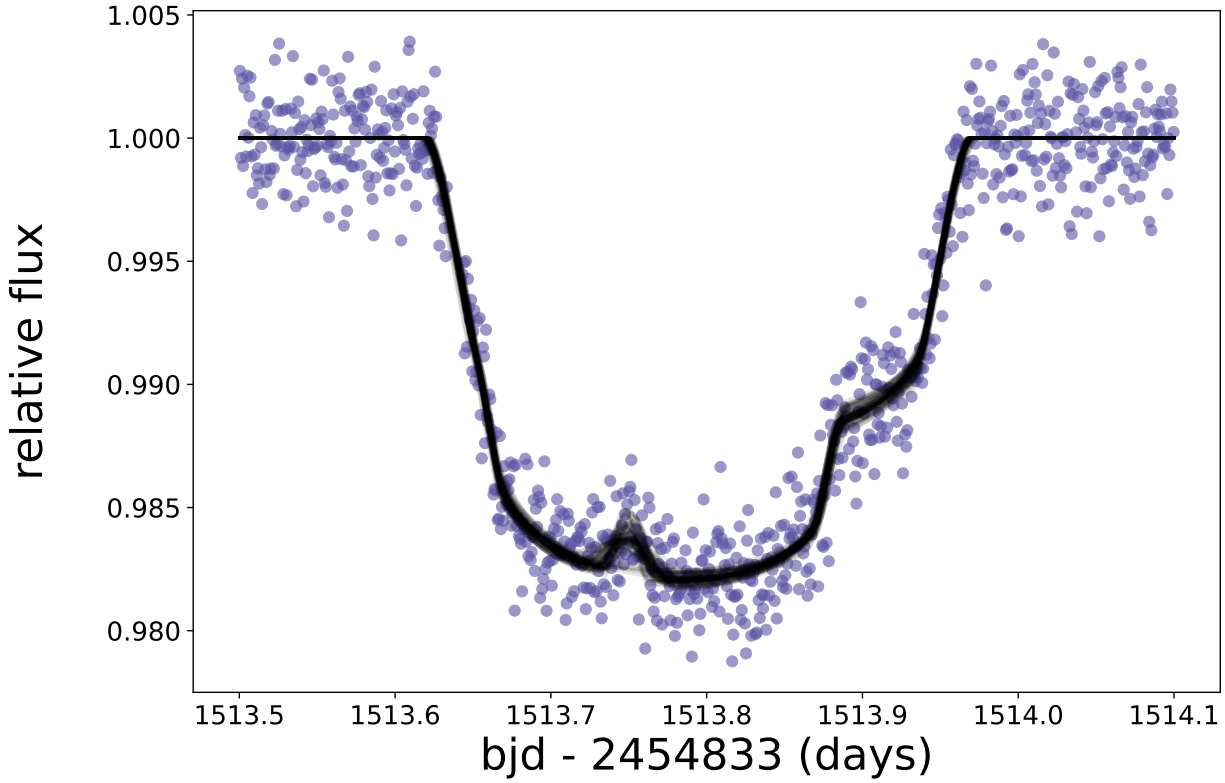

Figure 5.14 Kepler-51 photometric time series observed by Kepler (purple points) with 100 randomly drawn samples from the posterior distribution (black lines).

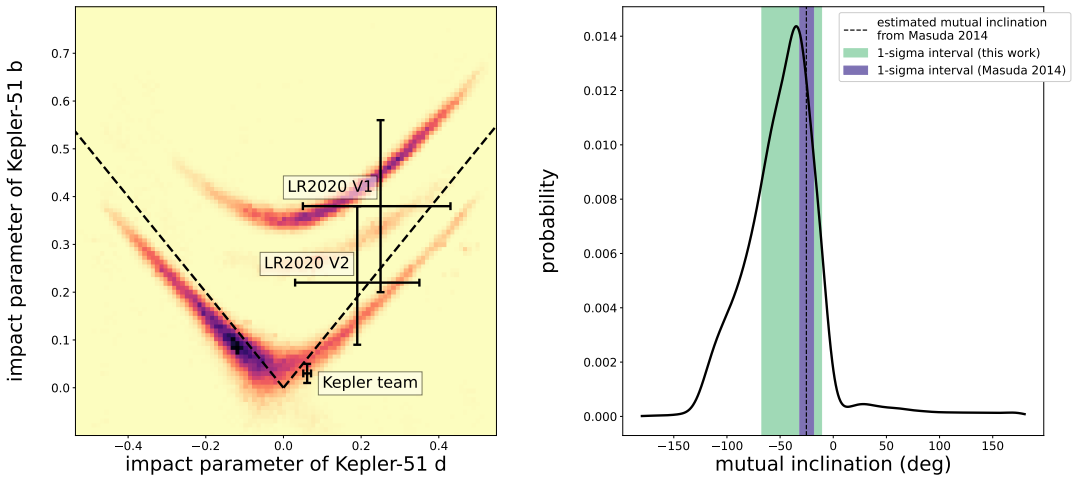

Figure 5.15 Posterior plots for the impact parameter of each planet versus the mutual inclination. **Left:** Joint posterior distribution for the impact parameters of planets b & d with darker colors indicating higher probability. The distribution is multi-modal, which presents a difficulty for MCMC methods. Dashed lines represent the line along which the impact parameters for both planets are equal, and the error bars represent the impact parameters found by the Kepler team and for each of the visits (V1 and V2) with *HST* from Libby-Roberts et al. (2020). This figure can be compared to Figure 4 in Masuda (2014). We have made no attempt to account for the possible failure of the MCMC chains to fully explore all of the modes. Therefore the apparent densities of each mode relative to the others should not be interpreted as a preference for one mode over another. We merely intend to show that multiple modes exist and to demonstrate the use of our model with an MCMC sampler. **Upper right**: Kernel density estimate of the posterior for the mutual inclination between planets b & d marginalized over all other parameters. The dashed line is the estimate from Masuda (2014) and the purple shaded region shows the 1-sigma confidence interval from that paper. The green shaded region represents the 1-sigma confidence interval from this work. We find agreement between our determination of the mutual inclination and the value found by M14, with the exception that our posterior admits a wider range of mutual inclination angles than was found by M14. Again, we caution that the multi-modality of the posterior presents an issue for our MCMC analysis and thus the relative heights of the modes in this plot may not represent the actual relative probabilities of true transit parameters having been drawn from each mode.

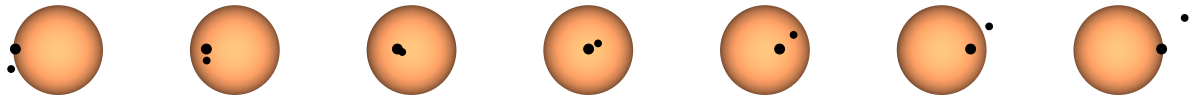

Figure 5.16 A series of snapshots showing a transit with the best-fit parameters from figure 5.13.

5.11 Conclusions

We have presented a solution for the flux during a mutual transit event in which two bodies overlap during their simultaneous transit of a star. It is mathematically equivalent to the solution given by Pál (2012), but is parameterized differently and includes derivatives with respect to the positions and radii of both bodies. While Pál (2012) gives a general method for computing the flux blocked by N overlapping bodies, our solution only applies for exactly two bodies. As a result we are able to give an explicit analytic solution for each of 16 different possible geometric cases.

We couple our solution to a dynamical model to build a full photodynamical model in a code which we call `gefera`. We give the user an option to model either a hierarchical system such as an exomoon/exoplanet system or hierarchical triple-star system or a confocal system such as two exoplanets orbiting the same star. We compare the speed of our implementation to `photodynam` and find that it performs several times faster even with the inclusion of the derivative computation. We also discuss the differences between our photodynamical model and `LUNA`, but because `LUNA` computes only an approximate solution and has not been made publicly available, a direct comparison is difficult to make.

As a demonstration of our method we conduct a brief re-analysis of a simultaneous transit of Kepler-51 b & d, which was originally studied by Masuda (2014). We come to the conclusion that Masuda (2014) is correct in assessing that an overlapping planet-planet transit is an unlikely explanation of the data, we further agree with their conclusion that an overlapping transit explanation is inconsistent with the measured impact parameters of the

two planets.

The code `gefera` is made publicly available as a pip-installable python package. It is our hope that the release of this package will lower barriers to conducting exomoon searches, carrying out independent analyses of putative exomoon transits and planet-planet mutual transits, and modeling transiting triple-star or circumbinary systems.

Chapter 6

EXOMOON TRANSIT OBSERVATION WITH JWST

6.1 Summary

In the previous chapter we described a transit model for exomoon/exoplanet mutual transits. The motivation for that work was in part to enable future searches for exomoons. At present, JWST is the only instrument likely to be capable of detecting an exomoon. Even with its vast capabilities, an exomoon detection will likely require a combination of advanced analysis techniques and a careful (or lucky) selection of an appropriate exoplanetary system. In this chapter we outline the prospects for exomoon detection with JWST, taking into account the sensitivity of the instruments, the effects of stellar variability, and the prospects for the existence of exomoons in known transiting exoplanet systems. The ultimate product of this chapter is a list of transiting exoplanets that may be suitable targets for exomoon searches. We include information about the accessibility of these systems to JWST in each of its first four campaigns. We also present a recommended strategy for exomoon searches and a framework for estimating the minimum radius of a detectable exomoon in each system.

6.2 Introduction

Astronomers have long imagined exoplanets to be accompanied by exomoons, yet searches for these companions have so far turned up very little. Attempts to detect exomoons are motivated by their importance both for characterizing exoplanets and assessing their habitability and for understanding the formation and evolution of exoplanetary systems (Heller et al. 2014).

Many previous papers have dealt in extensive detail with a wide range of interesting questions surrounding exomoon formation, evolution, characterization, and habitability. Among these are the impact of exomoon tides on the evolution of exoplanet/moon systems (Sasaki & Barnes 2014; Tokadjian & Piro 2020), the habitability of tidally heated exomoons (Tjoa et al. 2020; Reynolds et al. 1987; Scharf 2006; Forgan & Dobos 2016), the formation of satellites (Canup & Asphaug 2001; Agnor & Hamilton 2006; Canup & Ward 2006), and the impact of a system's dynamical evolution on its exomoon population (Kane 2017; Ramírez

& Sucerquia 2018; Trani et al. 2019, 2020).

To this day only a single exomoon candidate, Kepler-1625 b i (Teachey & Kipping 2018), has been proposed and has so far eluded confirmation. The primary difficulty in detecting exomoons is that they are either too small to be detected with current instruments and techniques, or they are rare and we have not been lucky enough to observe them yet. The truth is likely to be some combination of these — exomoons are not numerous at radii and orbital separations that make them easy to observe with current techniques. This is supported by the work of Hippke (2015), in which the author used a superposition of thousands of Kepler transits to infer a population of unseen exomoons with a mean radius of 0.8 R_{Ganymede}. The small dip in the superimposed transit signals that leads to this inference is most prominent in transits with periods longer than 35 days, which is consistent with the idea that exomoons of close-in planets are dynamically unstable due to the small Hill spheres of their host planets (Barnes & O'Brien 2002). The observations of Hippke (2015) underscore the issue with exomoon detection: a body with a radius of 0.8 R_{Ganymede} yields a transit signal of only 9 ppm around a sun-like star, and a period of longer than 35 days means that it is difficult to gather a large number of transit observations. Even if suitable photometric precision could be achieved, the presence of stellar variability with amplitudes similar to or greater than the depth of the transit signal would frustrate attempts at detection.

While transits are not the only method proposed for detecting exomoons, other methods face similar challenges. Proposed methods include the detection of transit timing and duration variations induced by an exomoon on the exoplanet that it orbits (Kipping 2009), measurements of the combined effect of multiple exomoon transits on stacked transits of a single exomoon (Heller 2014; Simon et al. 2012), measuring the effects of an exomoon on the radio emissions of a giant planet (Noyola et al. 2014), direct detection of an exomoon via direct imaging (Peters & Turner 2013), Doppler shifting of directly imaged exoplanets (Vanderburg et al. 2018), and the separation of the spectral signatures of an exoplanet and its moon (Agol et al. 2015; Forgan 2017). The probability of detecting an exomoon by any of these methods scales inversely with the size of the moon, the distance of the moon from

its host planet, and the distance of the host planet from the star.

In this chapter we show that JWST has the capability to detect exomoons as small as Ganymede for a number of planetary systems. A factor that may present an obstacle to such a detection is stellar flicker, the photometric analogue to jitter in radial velocity observations. Flicker is a poorly constrained phenomenon that consists of low amplitude correlated noise that is observed in high precision light curves arising from granulation on the star's surface (Bastien et al. 2013; Cranmer et al. 2014). We show that the effect of flicker on a transit light curve can be mitigated by modeling noise with a wavelength-dependent Gaussian process as discussed in Chapter 3.

6.3 Challenges to Exomoon Detection

We can divide the host of challenges to exomoon detection into three categories: challenges having to do with the size of exomoons, challenges related to the abundance (or lack of abundance) of exomoons, and challenges related to the dynamics of an exomoon.

While the absence of any detected exomoons means that we know very little about the size distribution of moons outside of the solar system, solar system observations and theoretical considerations suggest that exomoons will be small. The largest moon in the solar system, Jupiter's satellite Ganymede, is approximately four tenths the radius of the Earth. This places it intermediate in size between Mercury and Mars, with a transit depth that is only 17% that of an Earth-radius body. To detect transits this shallow it is necessary to have either a very small star in order to maximize the radius ratio between the star and the moon and thus to maximize the moon's transit depth, or to have a very bright star in order to decrease shot noise. The combination of these traits is often necessary to detect the very smallest transits. For some bright stars a further issue arises: correlated noise arising from stellar granulation and pulsations can mimic or obscure shallow transit signals.

The second challenge is the abundance of exomoons around known exoplanets. The observational biases of transit surveys are such that the sample of known transiting exoplanets is not optimal from a standpoint of exomoon stability. Dynamical arguments suggest that

a planet with a large Hill sphere has more room for a moon to evolve dynamically without becoming unbound from the planet. The Hill sphere of a planet increases with increasing semi-major axis of the planet's orbit. Unfortunately, the probability of a planet transiting its star decreases with semi-major axis. Furthermore, planets orbiting farther from their stars transit less frequently and are therefore observed less frequently as well. The most abundant population of transiting exoplanets, the hot Jupiters, are thought to be unlikely exomoon hosts due to their proximity to their stars and their dynamical history. It is thought that hot Jupiters must undergo some form of migration from beyond the snowline to get to their short period orbits, and planetary migration scenarios are unfavorable to the stability of exomoons in these systems (Trani et al. 2020; Namouni 2010).

The third challenge is that the dynamics of an exomoon pose difficulties for transit detections. The fact that a planet transits its star does not guarantee that any moons it possesses will also transit. Highly inclined moons have a substantial probability of failing to transit alongside their planets (Martin et al. 2019). Additionally, exomoon transits are not strictly periodic due to their orbit about the planet. Any transit detection algorithm that relies on periodicity, for instance methods that fold the light curve on trial periods, may be unable to detect exomoon transits.

One way to overcome the problem of a low abundance of detectable exomoons is to simply increase the number of observed exomoon targets. While TESS has already observed over 200,000 stars at two minute cadence during its primary mission and continues to observe as it moves into its extended mission. Unfortunately, TESS's photometric precision is lower than that of Kepler and, due to its shorter baseline, is biased towards detecting short period planets. Together these factors make detections of exomoons unlikely (Sucerquia et al. 2020). CHEOPS achieves comparable photometric precision to Kepler when accounting for its brighter target list, and a study by Simon et al. (2015) suggests that it is likely to achieve similar or better detection limits than Kepler. Another promising mission for exomoon detection in the near future is the PLATO mission (Rauer et al. 2014), another ESA operated mission which will focus on detecting exoplanets (in contrast to CHEOPS

which aims to characterize known planets). PLATO is expected to achieve a lower noise floor than CHEOPS for its target sample of bright, sunlike stars when it launches in 2026.

Targeted observations with JWST offer an alternative path towards exomoon detection. The unprecedented noise floor achievable with JWST (see, for instance, Rustamkulov et al. 2022) enables the detection of Lunar to Ganymede-sized moons for a number of known transiting systems. The primary remaining obstacle to these observations is the possibility of stellar flicker on the transit timescale which can lead to missed detections or false positives if not carefully accounted for.

For sun-like stars the main sources of variability on timescales of minutes to hours is granulation, which induces correlated variations in the light curve known as flicker. Spot and facula-induced variability occurs on the timescale of the star's rotation period, which is typically several days or longer for main sequence F, G, and K stars, and therefore does not interfere with transit detection. Stellar pulsations, in contrast, occur on timescales of seconds to a few minutes with a power spectrum that does not extend to longer timescales. As a result, pulsations add a source of white noise to lightcurves and interfere with the measurement of transit ingress and egress timesclaes, but they do not tend to prevent transit detection and have less impact on measurements of the transit depth than other correlated noise sources. Figure 6.1 shows the power spectrum of the Sun as observed by the SOHO spacecraft with labels to indicate the contributions of different sources to the overall variability.

Because flicker due to granulation is the main contributor to the variability spectrum on transit timescales we ignore pulsations and sunspots for our analysis. The Gaussian process model we use to handle flicker can, however, account for these sources of variability as additional terms in the Gaussian process kernel, each with their own amplitude and characteristic timescale.

Flicker in the Kepler sample has been extensively studied both for its impact on transit detection (Cranmer et al. 2014) and for its correlation with a star's surface gravity (Bastien et al. 2013; Pande et al. 2018). While much of this work has focused on subgiants due to the fact that flicker amplitudes increase as stellar surface gravity decreases, granulation and

pulsation variability has been detected for F, G, and K stars as well. Karoff et al. (2013) finds granulation variability amplitudes of $60 - 80$ ppm for these stars, which is near the limit of Kepler short-cadence precision for even the brightest targets.

In this chapter we investigate the capabilities of JWST to detect exomoons by simulating realistic JWST observations of transiting exomoons including all known sources of JWST instrumental noise as well as stellar granulation and pulsations. We use the `JexoSim` code (Sarkar et al. 2020) to simulate JWST observations for several targets chosen for their potential to host exomoons and their amenability to JWST observation. We then undertake an analysis of these simulated observations to understand the limits of our ability to detect exomoon transits. Finally, we demonstrate the recovery of an exomoon signal in a simulated light curve. We finish with a discussion of strategies for choosing exomoon targets and present a table of exoplanets with potential as exomoon targets for JWST.

6.4 Simulations

6.4.1 Observation Model

We use JexoSim to compute the noise amplitude on the transit timescale, taking into consideration photon noise, zodiacal noise, dark current, optical surface emission, read noise, and spatial and spectral jitter. Of these photon noise, read noise, and combined spatial/spectral jitter are most important at NIRSpec wavelengths. NIRSpec employs an up-the-ramp pixel sampling scheme wherein the charge is read off the detector in m consecutive frames averaged into n groups, for a total of mn read-outs before a pixel is cleared of its charge. As a result, photon noise is correlated across the n groups. This means that the fractional photon noise does not average down according to $\sigma_{\rm photon} \propto \sqrt{n}$ as would be expected for uncorrelated noise. Additionally, jitter is correlated over timescales set by the power spectral density (PSD) of the pointing timeline (Sarkar et al. 2020). While the pointing timeline PWD for JWST is not known to the author of this chapter, JexoSim adopts the PSD of the Herschel space telescope, scaling the amplitude of the pointing error to match that expected

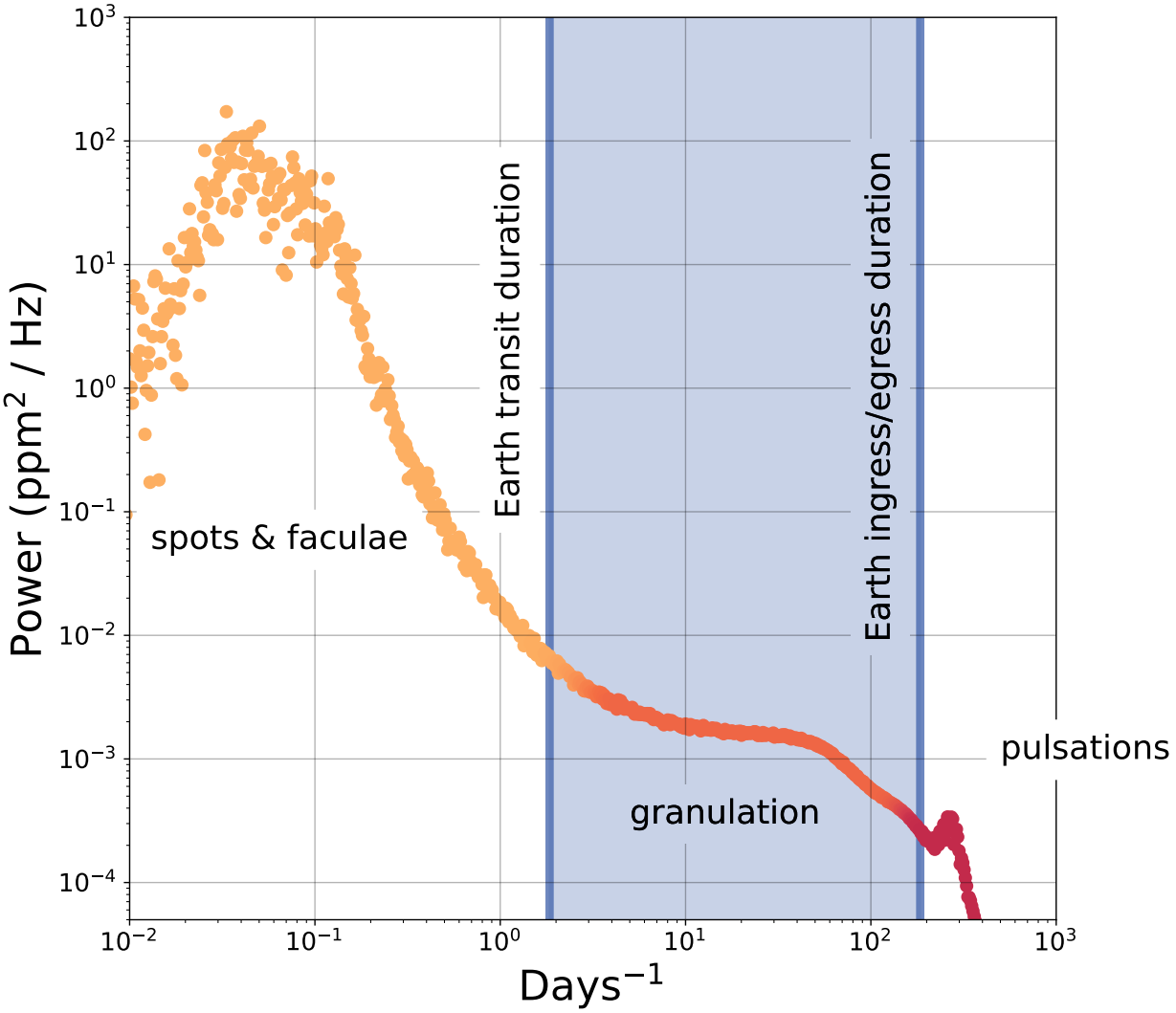

Figure 6.1 Power spectrum of the sun as seen by the SOHO spacecraft. The primary components of the power spectrum are labeled and the Earth's transit and ingress/egress duration across the sun is marked by the vertical lines. The main source of variability on the transit timescale is granulation, with pulsation noise occuring on timescales just shorter than the ingress/egress duration. For some stars pulsation noise may interfere with the ingress and egress, while the transit timescale should always be dominated by granulation noise for main sequence F, G, and K stars. (data downloaded from https://soho.nascom.nasa.gov/data/archive.html)

for JWST. Figure 6.3 shows how these correlated noise sources effect the scaling of the noise amplitude with respect to the transit duration. We find that the fractional noise amplitude scales approximately as

$$\mathrm{SNR} \propto \delta^{\,0.43} \tag{6.1}$$

where the SNR represents the total signal-to-noise for all noise sources and δ is the transit duration.

6.4.2 Flicker Model

We model flicker arising from granulation noise as the result of a trade-off between the covering fraction of hot and cold components of the star's photosphere. The hot component corresponds to the center of the granules, where the upward flow of hot material from deeper in the star is visible, and the cold component corresponds to the cooler intergranular material. We model the covering fraction of the cool component, x_c as a Gaussian process with a simple harmonic oscillator kernel, as described in Foreman-Mackey et al. (2017) and Chapter 3 of this work. Since there are only two components, the covering fraction of the cold component determines the covering fraction of the hot component via the equation $x_h = 1 - x_c$. The wavelength-dependence of the variability is then determined by the spectra of each of these components.

We use PHOENIX model spectra (Husser et al. 2013) for each component with temperatures calculated by assuming a temperature contrast of 1.03. This value are taken from the Stagger-grid simulations for solar metallicity $\log_{10}(g) = 4.5$ (Magic & Asplund 2014). We take the temperature of the cold component to be equal to the star's effective temperature, and the temperature of the hot component is then determined by the contrast ratio. Since the PHOENIX models are only available on a grid with $\Delta T = 100K$ between grid points, we use a linear interpolation between adjacent spectra in the grid to determine the spectrum of the hot component, while the cool component is taken directly from the PHOENIX model with no interpolation.

The flicker amplitude is related to the star's surface gravity, and is often measured as F_8, which is defined to be the scatter in the light curve taken in a broad-band, on time scales 8 hours and less, once shot noise has been removed (Bastien et al. (2013), hereafter B13). We use the $\log_{10}(g)$-F_8 relationship from B13 to scale the flicker amplitude observed for the sun to main-sequence stars with different values of $\log_{10}(g)$.

We model the power spectrum of the flicker noise is modeled as

$$S(\omega) = \sqrt{\frac{2}{\pi}} \frac{S_0}{(\omega/\omega_0)^4 + 1}, \tag{6.2}$$

which describes a Gaussian process with kernel function

$$k(|t_i - t_j|) = S_0 \omega_0 e^{-\frac{\omega_0 |t_i - t_j|}{\sqrt{2}}} \cos\left(\frac{\omega_0 |t_i - t_j|}{\sqrt{2}} - \frac{\pi}{4}\right). \tag{6.3}$$

This power spectrum has been used to model flicker by Kallinger et al. (2014), and is discussed as a model of stellar variability in Foreman-Mackey et al. (2017) and Chapter 3 of this work. In practice, adding additional kernel terms may improve the model fit (Barros et al. 2020). For our purposes, using a single term simplifies the analytic analysis in section 6.5.2 without altering our conclusions.

6.5 Results

6.5.1 Flicker-free analysis

We initially assume no stellar variability and a rectangular transit model defined

$$\mu(t) = \begin{cases} 1 - \frac{R_p}{R_\star}^2 & -\frac{\delta}{2} \leq t \leq \frac{\delta}{2} \\ 1 & \text{otherwise} \end{cases} \tag{6.4}$$

where δ is the transit duration and the transit is centered at $t = 0$. We further assume that the out-of-transit baseline is sufficiently long to permit an arbitrarily precise determination

of the out-of-transit flux. We consider a transit to be detectable if the depth, $(R_P/R_*)^2$, is distinguishable from the out-of-transit baseline at a level of 5-sigma.

For a transit duration of 10 hours, we find that that exomoons with sub-Mars radii are detectable for stars with radii up to about $1R_\odot$ for stars brighter than $J = 13$. Exomoons as small as a Ganymede radius are detectable around M and K dwarfs for $J < 12$, and exomoons with sub-Lunar radii may be detectable around M-dwarfs. These results are summarized in Figure 6.2.

The 10 hour transit duration that we used for these calculations is representative of the range of transit durations for temperate planets around G and K dwarfs. Since the orbital velocity of the exomoon will be small compared to that of the planet around its star, an exomoon should have a transit duration similar to the host planet, provided its orbit is relatively coplanar. For temperate planets around M dwarfs, as well as for shorter period planets around G and K dwarfs, the transit duration will be shorter and the transit SNR will be lower. Because the transit signal-to-noise depends on the white noise in the light curve binned to the transit duration, a shorter transit duration means lower transit signal-to-noise according to the scaling law in equation 6.1. In terms of the orbital period, P, this becomes

$$\mathrm{SNR} \propto P^{\,0.143} \sim P^{1/7} \tag{6.5}$$

where we have assumed a circular orbit. For reference, the 10 hour transit duration used in Figure 6.2 corresponds to a period of 167 days around the Sun, assuming an impact factor of $b = 0$. The second panel of Figure 6.2 shows the same detection limits for a 5 hour transit, illustrating the effect of an additional factor of $\sqrt{2}$ in white noise amplitude. The limits can be seen to be qualitatively similar but shifted to smaller stellar radii. For instance, detecting a Mars-sized body would require a star smaller than about 0.9 solar radii rather than 1.1 solar radii in the 10 hour case, with smaller exomoons correspondingly more difficult to detect.

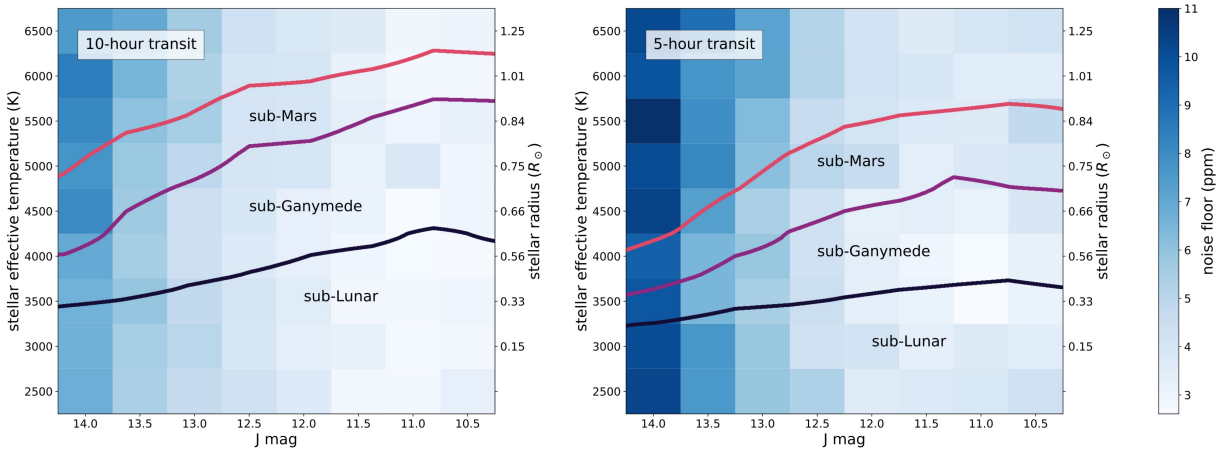

Figure 6.2 **Left:** Detection limits for Mars, Ganymede, and Moon-sized exomoons assuming a white noise-only noise model and a transit duration of 10 hours observed with the NIRSpec prism. The color of each cell indicates the noise floor for the corresponding JexoSim simulation in parts per million. **Right:** Identical to the first panel but assuming a shorter 5 hour transit.

6.5.2 Impact of flicker

A range of techniques have been applied to model correlated astrophysical variability in transit light curves, including various wavelet filtering (Carter et al. 2008), Kalman filtering (Kelly et al. 2014), sliding mean or median filters, sums of sines and cosines (Kipping et al. 2013; Mazeh & Faigler 2010), and Gaussian process methods (Barclay et al. 2015; Foreman-Mackey et al. 2017). In order to avoid detrending the transit signal, transits must typically be masked, or filters must be designed that avoid detrending on the transit timescale. When applying Gaussian process methods, the transit and GP noise model are fit simultaneously. As long as the shape of the transit is not well-modeled by the GP, a transit model will be preferred and the transit can be detected and its parameters measured. However, if the location of the transit is unknown it cannot be masked, and if the depth and duration of the transit is such that it isn't distinguishable from noise, then neither careful filtering or simultaneous GP fitting can avoid detrending the transit.

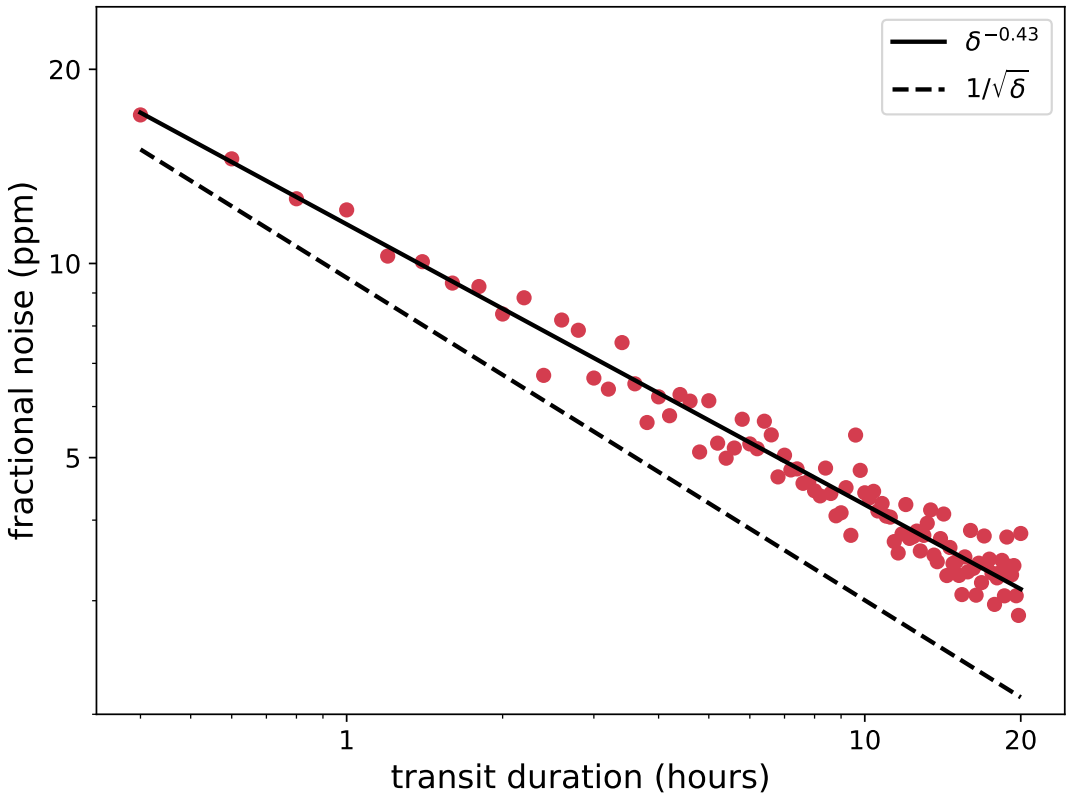

Figure 6.3 Scaling of the fractional total noise amplitude with transit duration. Red points represent individual realizations of a JexoSim noise simulation, and the solid black line represents the best fit power law model. The dashed line is provided for reference and represents the scaling that would be expected for pure, uncorrelated white noise.

For monochromatic light curves this problem is insurmountable. However, if we have access to spectral information the differing wavelength dependence between the stellar variability features and an authentic transit signal allows us to distinguish between the two. This is the basis of the wavelength-dependent GP model described in Chapter 3, which allows for modeling a light curve split into M spectral bins to take advantage of variations in the noise amplitude across wavelength. That chapter develops an Information analysis of the uncertainty on the transit depth for a rectangular transit model in the presence of noise drawn from a wavelength-dependent Gaussian process. The uncertainty on transit depth is given by

$$\rho_M^2 = \frac{\sigma^2 + \sum_{i=1}^{M} \alpha_i^2}{M\left(1 + \sigma^{-2}\sum_{i=1}^{M}\alpha_i^2\right) + \left(\sum_{i=1}^{M}\alpha_i\right)^2} \tag{6.6}$$

where ρ is the standard deviation of the estimated transit depth, σ is the white noise which is assumed to be the same in each band, $i = 1, 2 \ldots M$ indexes the spectral bins, and α_i is the variability amplitude in the i^{th} bin on the timescale of the transit. This can be compared to the equivalent monochromatic case:

$$\rho_1^2 = \frac{\sigma^2}{M} + \left(\sum_{i=1}^{M}\frac{\alpha_i}{M}\right)^2. \tag{6.7}$$

For the granulation noise we consider here, the variability amplitudes, α_i, depend on the temperature contrast of the hot and cold components of the photosphere as described in section 6.4.2. Because α_i represents the noise amplitude on the timescale of the transit, it also depends on the transit duration δ as well as the parameters of the power spectrum of the flicker noise:

$$\alpha_i(T_{\text{eff}}, \Delta T_{\text{gran}}, \delta, \tau) = a_i(T_{\text{eff}}, \Delta T_{\text{gran}}) k_{\delta,\tau,S_0}(t_i = t_j) \tag{6.8}$$

where $a_i(T_{\text{eff}}, \Delta T_{\text{gran}})$ is the coefficient for the amplitude in the i^{th} bin, which depends on the effective temperature of the star and on the temperature contrast between the granules and

the inter-granular photosphere, $k_{\delta,\tau,S_0}(t_i = t_j)$ is the covariance of the flicker on the transit timescale, δ, with characteristic timescale τ, evaluated at $t_i = t_j$. For the power spectrum defined in Equation 6.2 this is

$$k_{\delta,\tau,S_0}(t_i = t_j) = \frac{S_0}{\delta}\left(2 - \frac{\tau}{\sqrt{2}\pi\delta} + \frac{\tau}{\pi\delta}e^{-\frac{\sqrt{2}\pi\delta}{\tau}}\cos\left(\frac{\sqrt{2}\pi\delta}{\tau} + \frac{\pi}{4}\right)\right). \tag{6.9}$$

This formula can be found by convolving the kernel function in Equation 6.3 with a window function

$$W_\delta(t_i, t_j) = \begin{cases} 1 & -\delta/2 < t_i, t_j < \delta/2 \\ 0 & \text{otherwise} \end{cases} \tag{6.10}$$

and evaluating at $t_i = t_j$ in order to recover the covariance at zero time separation.

Figure 6.4 shows the dependence of the 5-sigma transit detection limit on the flicker amplitude and the timescale τ. We calculate the detection limits for a single spectral bin, in which the wavelength-dependence of the flicker noise has been effectively integrated out by summation across wavelength, as well as for three bins. We find that for transit durations longer than a few hours and for flicker amplitudes comparable to the shot noise the difference between the noise model with three bins and with a single bin is not significant. However at large flicker amplitudes and short transit durations the difference between the two models becomes much more significant. The usefulness of the spectral information also increases when the flicker timescale, τ is closer to the transit duration. For exomoons we expect the transit duration to be on the order of 10 hours because of the preference for long period planets, while τ will vary from star to star. Sulis et al. (2020) calculates the cut-off flicker frequency f_g, which is related to our τ by $f_g = 1/\tau$, for a sample of main-sequence Kepler stars and finds that it ranges over nearly an order of magnitude for F and G dwarfs, decreasing from $f_g \sim 100\mu Hz$ to $f_g \sim 800\mu Hz$. This corresponds to a range in τ from 20 minutes to 2.7 hours. In order to represent the bounds of this range, the detection limits in the upper

panels of Figure 6.4 are calculated at $\tau = 30$ minutes while the lower panels are calculated at $\tau = 2$ hours. For small flicker timescales the improvement between the the single bin and three bin cases is small over most of the parameter space, whereas for longer flicker timescales the difference is more noticeable, especially at short transit durations.

In Figure 6.5 we have computed the 5-sigma transit detection limit over the same parameter space as in Figure 6.2 for flicker amplitudes of 10 ppm (upper panels) and 20 ppm (lower panels) and flicker timescales of 30 minutes (left panels) and 2 hours (right panels).

6.6 Target Selection

Beginning with a sample consisting of all confirmed transiting exoplanets as of May 2021, as catalogued in the NASA Exoplanet Archive, we forecast their masses using the probabilistic mass-radius forecasting method of Chen & Kipping (2017). We then estimate the Hill radius of each planet based on this forecasted mass, rejecting any planets with a Hill radius smaller than 0.005 au (for reference, the Earth's Hill radius is about 0.01 au, and the Moon orbits at a quarter of this distance). By rejecting planets with small Hill radii, we avoid those for which tidal evolution is likely to have driven any moons either outside of the Hill sphere or inside of the planet's Roche limit (Sasaki & Barnes 2014). This still leaves a fraction of massive hot Jupiters in our sample, which are not thought to be likely to host exomoons on the basis that most migration scenarios would destabilize any moons that may have originally formed alongside the planet (Trani et al. 2020; Sucerquia et al. 2020; Spalding et al. 2016). We therefore exclude planets with orbital periods less than 35 days from consideration. Figure 6.6 shows the orbital periods of the full sample of transiting exoplanets against our Hill radius estimates, with the size and color of the points representing the planetary radius and equilibrium temperature respectively. In this representation planets in the upper right-hand corner of the plot, with long orbital periods and large Hill spheres, are those that we consider to be the best exomoon hosts. Unfortunately, our preference for planets with long periods and large Hill spheres is in tension with observational constraints, as their infrequent transits make scheduling observations difficult and their long duration transits and large Hill spheres

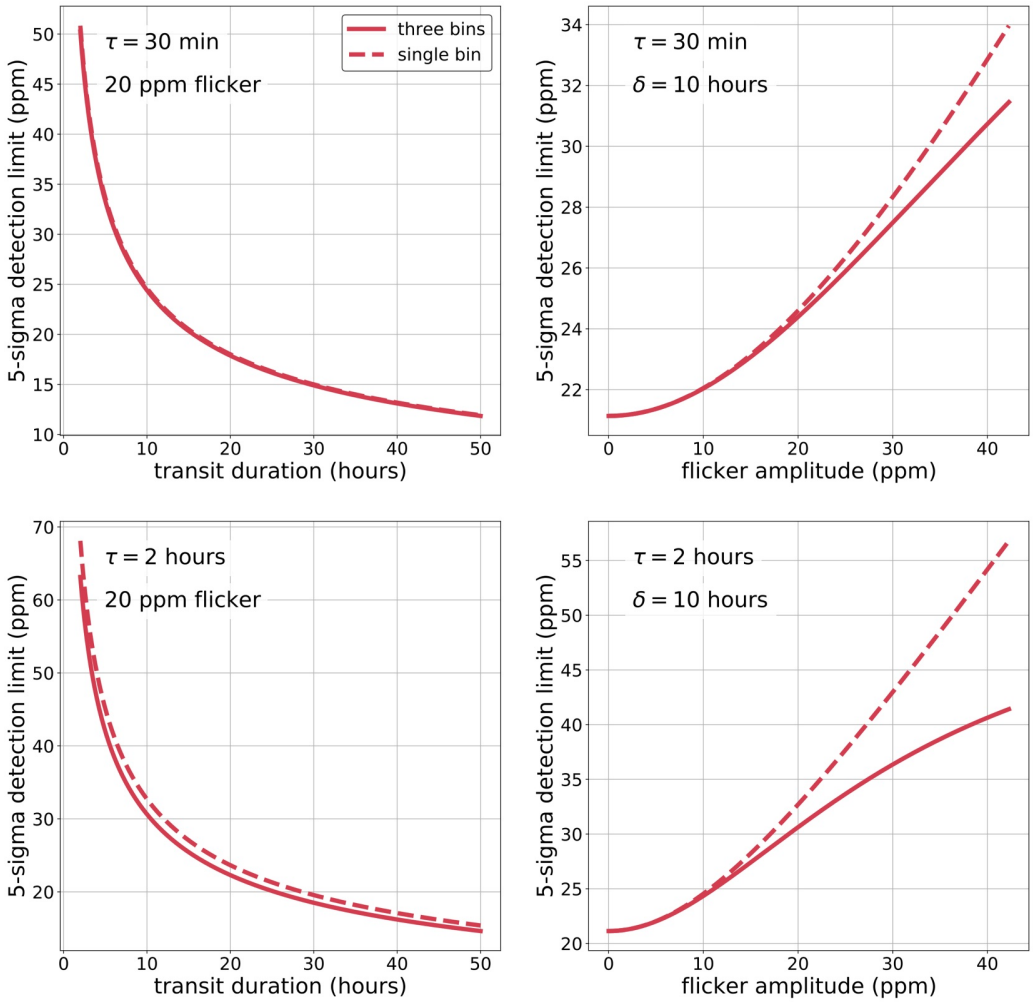

Figure 6.4 Detection limits in the presence of flicker with a power spectrum described by Equation 6.2. **Left panels:** Transit depth required for a 5-sigma detection for transit durations greater than 2 hours, assuming a rectangular transit model and a flicker amplitude of 16.8 ppm. **Right panels:** Transit depth required for a 5-sigma detection vs. flicker amplitude for a transit duration of 10 hours. The upper panels are calculated for a flicker timescale of 30 minutes while the lower panels are calculated for a flicker timescale of 2 hours. For all panels we consider a star with $T_{\rm eff} = 5300$ K, granule temperature contrast $\Delta T_{\rm gran} = 200$ K, and a J-magnitude of 13, which gives a white noise floor of 3.64 ppm over 10 hours. The horizontal black line in both plots represents the 5-sigma detection limit in the absence of flicker. We include curves for the case of a single spectral bin, in which the wavelength-dependency has been integrated out by summation across wavelength, as well as for three spectral bins. Using more than three bins improves the detection limit only slightly.

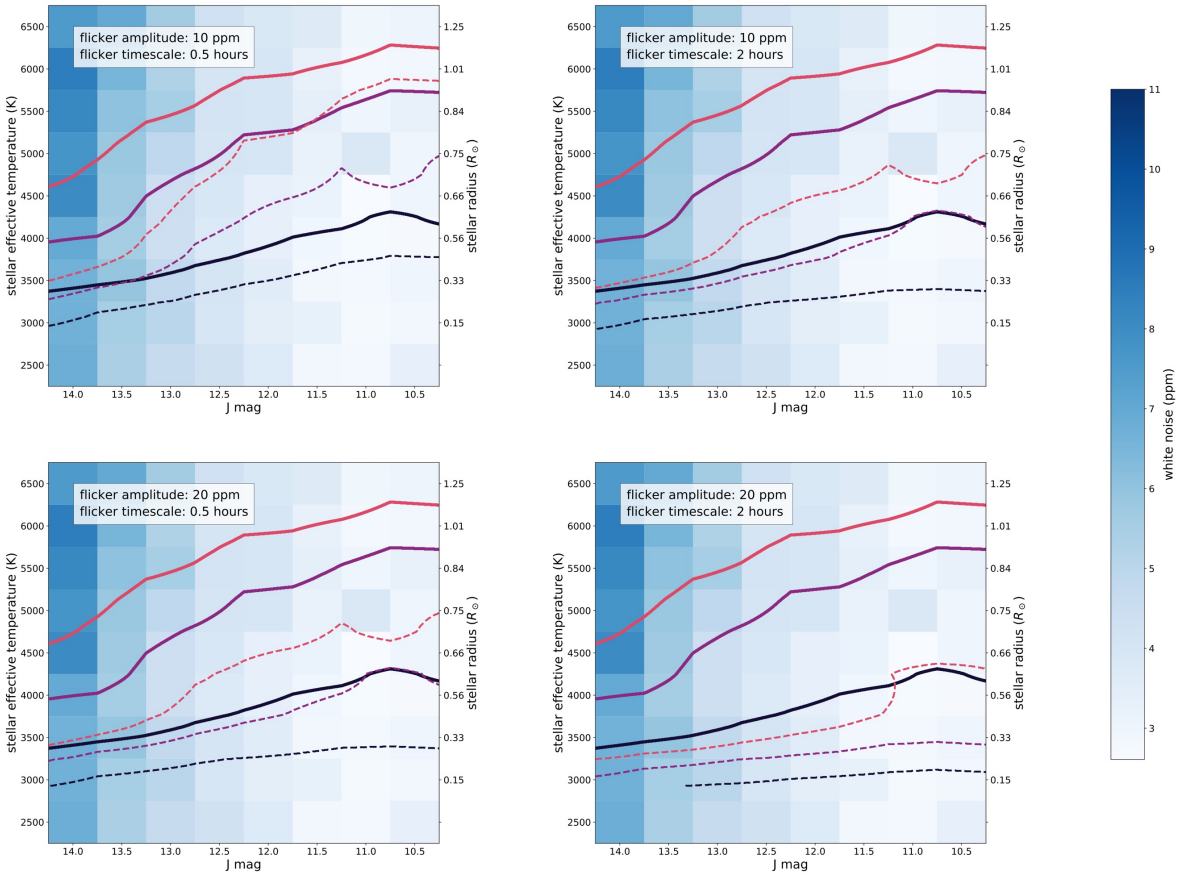

Figure 6.5 Detection limits in the presence of flicker with a power spectrum described by Equation 6.2. Solid lines represent the white-noise-only 5-sigma detection limit and dashed lines represent the same detection limit in the presence of flicker assuming we divide the flux into three spectral bins and use the flicker model described in section 6.4.2. **Upper panels:** Flicker amplitude of 10 ppm **Lower panels:** Flicker amplitude of 20 ppm. In the left panels we consider a flicker timescale of 30 minutes, similar to that which is observed for the Sun, and near the lower limit expected for this parameter for F and G dwarfs. In the right panels the flicker timescale is set to 2 hours, which is approximately the upper limit expected for main-sequence stars and is close to the upper limit expected for F and G dwarfs. The flicker timescale tends to increase with increasing mass for main-sequence stars (see Sulis et al. (2020)). Note that the upper right and lower left plots appear identical. This is not a mistake. It is likely due to the relatively course grid on which we have computed our simulations that these two panels appear the same.

require many hours of observations.

Selecting systems for a targeted search means considering not only the potential for a system to host an exomoon, but also the detectability of such moons. We consider a moon to be detectable if the signal-to-noise of an exomoon transit is high enough to make a statistically significant detection with a single transit. The detectability therefore increases as the brightness of the host star increases, as a brighter host star allows for more precise flux measurements. Since the depth of the transit of an exomoon with a given radius increases with decreasing stellar radius, the overall detectability of an exomoon is higher around a smaller host star. In this way the considerations for exomoon detectability are no different than those for an exoplanet. The left panel of Figure 6.7 shows the stellar radii of stars in our subsample plotted against the J band magnitude, for stars brighter than J mag = 14 and radii smaller than 1.4 $R_{\odot}$. The size of the circles is again proportional to the radius of the planet, and the color represents the planet's equilibrium temperature. In this representation planets in the lower right-hand corner offer the highest potential transit SNR. We note that many of the planets in this quarter are hot or warm giant planets which have likely undergone significant migration. Since migration is likely to destabilize exomoons, we exclude planets with periods shorter than 35 days, as shown in the right panel of Figure 6.7. To avoid crowding, Figure 6.7 shows only those planets for which the Hill radius is estimated to be greater than 0.007 au, while Table 6.7 includes planets with Hill radii down to 0.005 au.

6.7 Conclusions

In this chapter we perform realistic simulations of JWST observations using JexoSim and combine them with a theoretical treatment of the effects of stellar flicker in order to study the detectability of exomoons. We find that exomoons between the size of Earth's moon and Ganymede should be within the grasp of JWST for sufficiently bright, low-mass host stars. We further find that a careful treatment of stellar flicker may be essential for avoiding false positives and for finding moons at the edge of detectability. Finally, we present a list of known transiting exoplanet systems for targeted exomoon searches with JWST based on

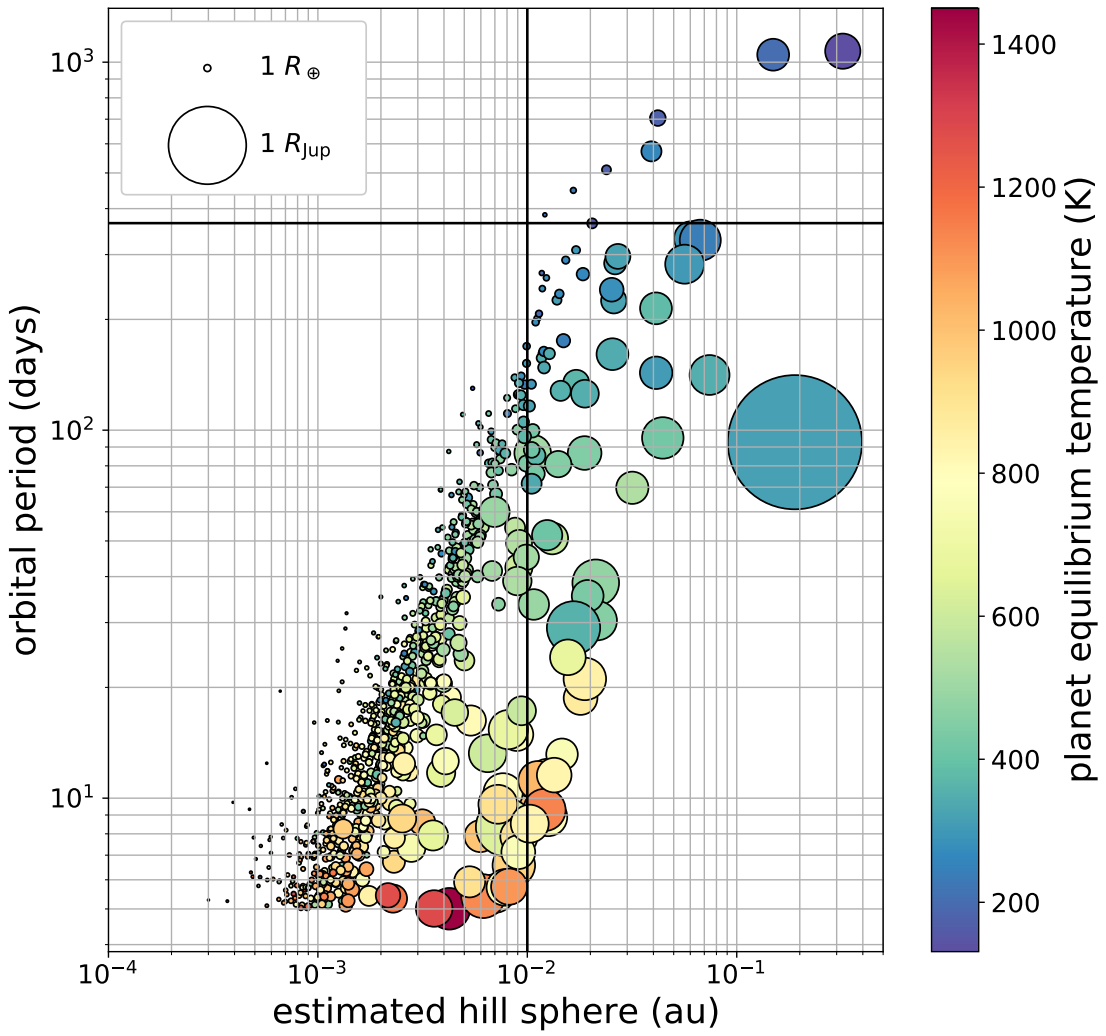

Figure 6.6 Orbital period vs Hill radius estimate for the full sample of transiting exoplanets. The size of each point is proportional to the size of the planet, and the color is indicative of the planet's equilibrium temperature. The vertical black line indicates the Earth's Hill sphere and the horizontal black line marks the Earth's orbital period of one year. Planets occupying the upper right-hand portion of the plot are those which are most conducive to the stability of an exomoon over long time periods, absent any additional information about their composition and dynamical histories.

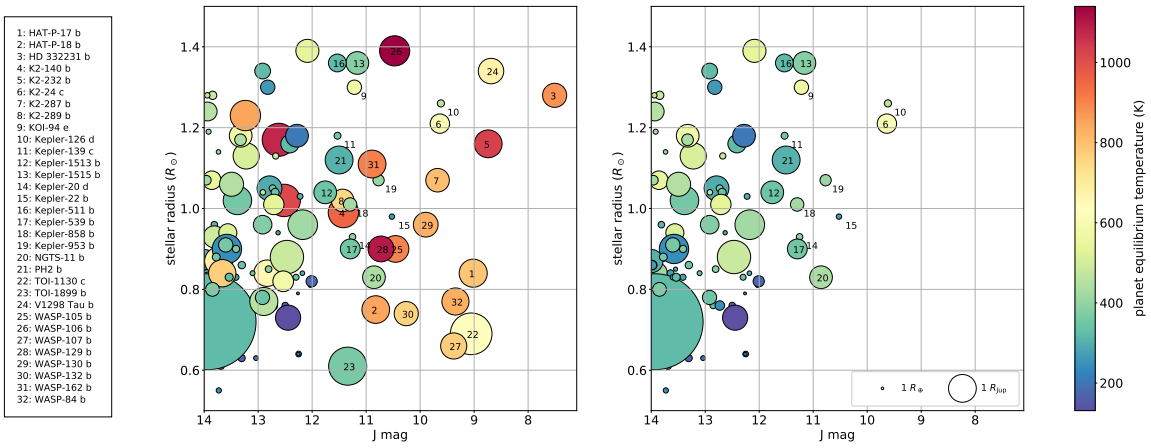

Figure 6.7 Stellar radii vs. J band magnitude for the subsample of transiting exoplanets. Since brighter, smaller host stars yield the highest transit signal-to-noise, planets in the lower right-hand corner of these plots are most amenable to exomoon searches. **Left:** Planets with estimated Hill radii greater than 0.007 au. **Right:** Same as the left panel, but excluding planets with orbital periods less than 35 days. While the hot, massive planets in the left panel make up some of the best targets when considering observational constraints only, they are likely to have undergone migration which would destabilize any exomoons that may have been present upon formation of the planets.

Table 6.1. Systems with potential for targeted exomoon observations

planet name	M_J	$R_\odot$	$T_{\rm eff}$ (K)	$P_{\rm orb}$	$R_\oplus$	estimated $R_{\rm Hill}$(au)	vis. in cycles
HD 95338 b	7.099	0.87	5212	55.1	3.89	0.006	1 2 3 4
V1298 Tau e	8.687	1.34		60.0	8.74	0.007	1 3 4
Kepler-126 d	9.616	1.26	6164	100.3	2.89	0.009	1 2 3 4
K2-24 c	9.635	1.21	5743	42.4	7.82	0.009	1 2 3 4
Kepler-16 b	9.815	0.65	4450	228.8	8.45	0.037	1 2 3 4
Kepler-538 b	10.026	0.87	5451	81.7	2.29	0.007	1 2 3 4
Kepler-22 b	10.523	0.98	5642	289.9	2.17	0.015	1 3 4
Kepler-953 b	10.770	1.07	5611	88.4	4.55	0.010	1 2 3 4
NGTS-11 b	10.855	0.83	5050	35.5	9.16	0.019	1 2 3 4
Kepler-517 b	11.020	1.06	5598	60.9	2.87	0.007	1 2 3 4
Kepler-1515 b	11.163	1.36	6471	214.3	9.17	0.041	1 2 3 4
KOI-94 e	11.218	1.3	6098	54.3	5.79	0.009	1 2 3 4
Kepler-20 d	11.252	0.93	5465	77.6	2.62	0.007	1 2 3 4
Kepler-539 b	11.294	0.9	5838	125.6	8.00	0.019	1 2 3 4
Kepler-858 b	11.297	1.01	5565	76.1	5.41	0.011	1 2 3 4
Kepler-1185 b	11.386	0.87	5650	104.4	1.39	0.006	1 2 3 4
PH2 b	11.501	1.12	6007	282.5	11.24	0.056	1 2 3
Kepler-139 c	11.530	1.18	5622	157.1	2.80	0.012	1 2 3 4
Kepler-511 b	11.534	1.36	5951	296.6	7.18	0.027	1 2 3
Kepler-104 d	11.558	1.0	5979	51.8	2.61	0.005	1 2 3 4
Kepler-1524 b	11.622	1.0	5900	71.0	2.86	0.007	1 2 3 4

the estimated Hill radius of the planet and ordered by J-band magnitude.

Table 6.1 (cont'd)

planet name	M_J	$R_\odot$	$T_{\rm eff}$ (K)	$P_{\rm orb}$	$R_\oplus$	estimated $R_{\rm Hill}$(au)	vis. in cycles
Kepler-1513 b	11.758	1.04	5650	160.9	9.16	0.025	1 2 3 4
Kepler-888 b	11.915	1.18	6190	70.7	2.22	0.006	1 2 3 4
Kepler-140 c	11.989	1.06	6167	91.4	1.54	0.006	1 2 3 4
Kepler-421 b	12.009	0.82	5305	705.0	4.53	0.042	
Kepler-1632 b	12.033	0.81	6070	448.3	1.66	0.017	1 2 3
Kepler-419 b	12.088	1.39	6293	69.7	9.30	0.032	1 2 3 4
Kepler-482 b	12.173	0.67	4616	56.4	2.20	0.005	1 2 3 4
Kepler-1047 b	12.174	0.93	5760	56.2	2.20	0.005	1 2 3 4
CoRoT-9 b	12.175	0.96	5625	95.3	11.95	0.044	1 2 4
Kepler-1040 b	12.181	0.84	5804	201.1	1.84	0.011	1 2 3 4
Kepler-1533 b	12.224	1.03	6363	308.5	2.36	0.017	1 2 3 4
Kepler-1540 b	12.250	0.64	4366	125.4	2.30	0.009	1 2 3 4
Kepler-62 f	12.256	0.64	4925	267.3	1.41	0.012	2 3 4
Kepler-452 b	12.263	0.79	5578	384.8	1.13	0.012	1 2 3 4
Kepler-1654 b	12.281	1.18	5597	1047.8	9.18	0.149	
Kepler-965 b	12.298	0.83	5973	134.3	2.42	0.009	1 2 3 4
Kepler-379 c	12.341	1.14	6273	62.8	2.12	0.006	1 2 3 4
Kepler-453 b	12.349	0.83	5527	240.5	6.20	0.024	
Kepler-978 b	12.363	1.18	5938	49.6	3.33	0.006	1 2 3 4
Kepler-111 c	12.412	1.16	5952	224.8	7.34	0.026	1 2 3 4
Kepler-803 b	12.415	0.94	5642	50.3	3.20	0.006	1 2 3 4

Table 6.1 (cont'd)

planet name	M_J	$R_\odot$	$T_{\rm eff}$ (K)	$P_{\rm orb}$	$R_\oplus$	estimated $R_{\rm Hill}$(au)	vis. in cycles
Kepler-215 e	12.417	1.01	5741	68.2	1.75	0.005	1 2 3 4
Kepler-198 c	12.421	0.93	5569	49.6	2.91	0.005	1 2 3 4
Kepler-167 e	12.446	0.73	4890	1071.2	10.15	0.321	
Kepler-199 c	12.449	0.88	5750	67.1	3.51	0.007	1 2 3 4
Kepler-983 b	12.456	0.99	6201	60.1	2.19	0.006	1 2 3 4
Kepler-468 b	12.467	0.88	5510	38.5	13.44	0.021	1 2 3 4
Kepler-186 f	12.473	0.46	3761	129.9	1.06	0.005	1 2 3 4
Kepler-1630 b	12.499	0.76	5037	510.0	2.62	0.024	3 4
Kepler-986 b	12.513	0.88	5734	56.4	2.09	0.005	1 2 3 4
Kepler-361 c	12.526	1.23	6321	55.2	2.55	0.005	1 2 3 4
Kepler-1675 b	12.532	1.09	6085	63.0	2.73	0.007	1 2 3 4
Kepler-11 f	12.548	1.1	5680	46.7	2.61	0.005	1 2 3 4
Kepler-1620 b	12.559	1.05	6210	102.0	1.36	0.006	1 2 3 4
Kepler-69 c	12.631	0.94	5637	242.5	1.74	0.012	1 2 3 4
Kepler-268 c	12.675	1.13	6322	83.4	2.52	0.007	1 2 3 4
Kepler-630 b	12.686	1.04	5761	161.5	3.25	0.013	1 2 3 4
Kepler-9 c	12.710	1.01	5777	38.9	8.22	0.009	1 2 3 4
Kepler-279 d	12.724	1.07	6781	54.4	3.20	0.006	1 2 3 4
Kepler-1661 b	12.729	0.76	5100	175.1	3.87	0.015	1 2 3 4
Kepler-218 d	12.732	1.05	5497	124.5	2.78	0.009	1 2 3 4
Kepler-146 c	12.739	1.06	6197	76.7	2.91	0.007	1 2 3 4

Table 6.1 (cont'd)

planet name	M_J	$R_\odot$	$T_{\rm eff}$ (K)	$P_{\rm orb}$	$R_\oplus$	estimated $R_{\rm Hill}$(au)	vis. in cycles
Kepler-430 c	12.743	1.24	5831	111.0	1.48	0.007	1 2 3 4
Kepler-324 c	12.783	0.8	5355	51.8	3.25	0.006	1 2 3 4
KOI-351 e	12.790	1.07	6330	91.9	2.25	0.007	1 2 3 4
KOI-351 d	12.790	1.2	5970	59.7	2.84	0.006	1 2 3 4
KOI-351 f	12.790	1.05	6238	124.9	2.46	0.009	1 2 3 4
KOI-351 h	12.790	1.05	6238	331.6	10.01	0.061	1
Kepler-209 c	12.797	1.04	5514	41.7	3.40	0.005	1 2 3 4
Kepler-1093 c	12.807	1.26	6429	89.7	2.08	0.007	1 2 3 4
Kepler-1316 b	12.807	0.85	5821	88.0	2.70	0.007	1 2 3 4
KIC 9663113 b	12.823	1.3	6226	572.4	5.80	0.039	1 3
EPIC 212737443 c	12.824	0.67	4684	65.5	2.69	0.005	1 3 4
Kepler-152 c	12.853	0.76	5262	88.3	2.56	0.007	1 2 3 4
Kepler-289 d	12.861	1.0	5990	66.1	2.68	0.007	1 2 3 4
Kepler-79 e	12.913	1.04	6388	81.1	2.23	0.007	1 2 3 4
Kepler-627 b	12.916	1.36	6419	40.7	3.36	0.005	1 2 3 4
KOI-1783.02	12.917	1.34	5791	284.1	6.36	0.026	1 3 4
KOI-1783.01	12.917	0.96	6215	134.5	7.50	0.017	1 2 3 4
Kepler-951 b	12.920	0.78	4855	71.5	5.76	0.010	1 2 3 4
Kepler-1227 b	12.963	0.97	5840	94.3	2.13	0.007	1 2 3 4

Chapter 7

CONCLUSIONS

The purpose of this thesis is to outline our efforts to advance the field of exoplanetary transit science in two directions. The first is the development of Gaussian process models for stellar variability, in particular for application to multiwavelength lightcurves and lightcurves containing photometric outliers. The second is the development of computational tools for exoplanet detection, and the analysis of the prospects for making such a detection with JWST.

In Chapter 2 we reviewed the concept of a Gaussian process, which for our purposes is a mathematical model of a timeseries containing random correlated and uncorrelated noise. While in the field of explanet science Gaussian processes are often used to marginalize over the noise in a lightcurve in order to estimate parameters of some underlaying mean function (often a transit signal), we demonstrate the application of a GP model to a slightly different problem: that of measuring the rotation period of a star based on the variability induced by the appearance of starspots on the stellar disk and their subsequent rotation into or out of view. For this application we were interested not in the parameters of some deterministic mean function but rather in the parameters that describe the noise; in particular the period of the variability which we associated with the rotation period of the star.

In Chapter 3 we reviewed the application of GP variability models to transit lightcurves. One limitation of these models for single-band or "white light" lightcurves is that they cannot help us to distinguish a transit signal from correlated variability when the variability has a similar amplitude and timescale to the transit itself. However, when multiwavelength observations are available the difference in the spectral dependence of a transit feature compared to the stellar variability can be used to distinguish between the two. In order to carry out this analysis we developed a multiwavelength Gaussian process noise model based on the fast and efficient `celerite` GP framework. We demonstrated that this model can be used to substantially improve our ability to detect and characterize transits in multiwavelength data compared to modeling each wavelength independently with a single-band GP model.

Another limitation of Gaussian process models is that they perform very poorly in the presence of outlying datapoints in the timeseries. In Chapter 4 we presented a method of

identifying outliers given a GP model of the timeseries. We then demonstrated the use of biased GP models to conduct inference in cases where ordinary GP models would fail due to the outliers. We compared the use of biased GP models to a more typical procedure in which outliers are identified and subsequently deleted from the timeseries. We found that both methods achieve similar results, but argued that the bias models provide more flexibility in outlier handling and may still be preferable to outlier deletion in some cases.

In Chapter 5 we developed a model for the simultaneous overlapping transit of two bodies such as a moon and a planet or two planets. The advantage of our model is that it is exact and differentiable with respect to all input parameters. The exactness means that the model is applicable in all cases regardless of the relative sizes of the planet(s), moon, and star. The differentiability of the model means that it can be used with inference methods that require access to the gradient of the likelihood function such as Hamiltonian Monte Carlo and nested sampling with gradients. Because exomoon detection is a difficult problem that is likely to require significant computational resources, the advantages of these methods in terms of computational speed and efficiency may be important for carrying out a detection.

In Chapter 6 we analyzed the prospects for exomoon transit detection with JWST. We used realistic time domain simulations of JWST observations from `JexoSim` (**?**) to understand the noise properties of the telescope. We then applied a theoretical treatment of the effects of stellar flicker derived from the results of Chapter 3. We found that large exomoons (Ganymede-sized or larger) are likely detectable for some of the brighter exoplanet host stars observable in the first 4 JWST cycles, and we presented a list of candidate systems for an exomoon transit search.

It is our hope that the research presented in this thesis will help to drive forward the field of exoplanet and exomoon transit science. We expect that as telescope technology advances and even larger successors to JWST are developed and launched, correlated stellar variability will increasingly become a limiting factor in our ability to detect and measure transits. It is therefore of crucial importance that we begin to develop techniques for handling this variability now in order to inform the design and predict the performance of future

instruments. The beginning of JWST's exoplanet observation program offers a potential test bed for these methods, and the detection of exomoons with JWST may require their application.

BIBLIOGRAPHY

Adibekyan, V., Dorn, C., Sousa, S. G., et al. 2021, Science, 374, 330, doi: http://doi.org/10.1126/science.abg879410.1126/science.abg8794

Agnor, C. B., & Hamilton, D. P. 2006, Nature, 441, 192, doi: http://doi.org/10.1038/nature0479210.1038/nature04792

Agol, E., & Fabrycky, D. C. 2018, in Handbook of Exoplanets (Cham: Springer International Publishing), 797–816. http://link.springer.com/10.1007/978-3-319-55333-7_7

Agol, E., Jansen, T., Lacy, B., Robinson, T. D., & Meadows, V. 2015, ApJ, 812, 5, doi: http://doi.org/10.1088/0004-637X/812/1/510.1088/0004-637X/812/1/5

Agol, E., Luger, R., & Foreman-Mackey, D. 2020, AJ, 159, 123, doi: http://doi.org/10.3847/1538-3881/ab4fee10.3847/1538-3881/ab4fee

Agol, E., Steffen, J., Sari, R., & Clarkson, W. 2005, MNRAS, 359, 567, doi: http://doi.org/10.1111/j.1365-2966.2005.08922.x10.1111/j.1365-2966.2005.08922.x

Agol, E., Dorn, C., Grimm, S. L., et al. 2021, , 2, 1, doi: http://doi.org/10.3847/PSJ/abd02210.3847/PSJ/abd022

Agüeros, M. A., Bowsher, E. C., Bochanski, J. J., et al. 2018, ApJ, 862, 33, doi: http://doi.org/10.3847/1538-4357/aac6ed10.3847/1538-4357/aac6ed

Ahrer, E.-M., Stevenson, K. B., Mansfield, M., et al. 2023, Nature, 614, 653, doi: http://doi.org/10.1038/s41586-022-05590-410.1038/s41586-022-05590-4

Algikar, P., & Mili, L. 2023, arXiv e-prints, arXiv:2301.07858, doi: http://doi.org/10.48550/arXiv.2301.0785810.48550/arXiv.2301.07858

Almosallam, I. A., Lindsay, S. N., Jarvis, M. J., & Roberts, S. J. 2016, MNRAS, 455, 2387, doi: http://doi.org/10.1093/mnras/stv242510.1093/mnras/stv2425

Amard, L., Roquette, J., & Matt, S. P. 2020, MNRAS, 499, 3481, doi: http://doi.org/10.1093/mnras/staa303810.1093/mnras/staa3038

Ambikasaran, S., Foreman-Mackey, D., Greengard, L., Hogg, D. W., & O'Neil, M. 2015, IEEE Transactions on Pattern Analysis and Machine Intelligence, 38, 252, doi: http://doi.org/10.1109/TPAMI.2015.244808310.1109/TPAMI.2015.2448083

Anderson, E. R., & Jefferies, S. M. 1990, 364, 699

Angus, R., Morton, T., Aigrain, S., Foreman-Mackey, D., & Rajpaul, V. 2018, MNRAS, 474, 2094, doi: http://doi.org/10.1093/mnras/stx210910.1093/mnras/stx2109

Angus, R., Beane, A., Price-Whelan, A. M., et al. 2020, arXiv e-prints, arXiv:2005.09387. https://arxiv.org/abs/2005.09387https://arxiv.org/abs/2005.09387

Asimov, I. 1978, Lucky Starr and the oceans of Venus (Fawcett)

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33

Astropy Collaboration, Price-Whelan, A. M., SipH ocz, B. M., et al. 2018, aj, 156, 123, doi: http://doi.org/10.3847/1538-3881/aabc4f10.3847/1538-3881/aabc4f

Avduevskij, V. S., Avduevsky, V. S., Marov, M. Y., et al. 1971, Journal of Atmospheric Sciences, 28, 263, doi: http://doi.org/10.1175/1520-0469(1971)028¡0263:SLOVOT¿2.0.CO;210.1175/1520-0469(1971)028¡0263:SLOVOT¿2.0.CO;2

Bagnasco, G., Kolm, M., Ferruit, P., et al. 2007, in Cryogenic Optical Systems and Instruments XII, ed. J. B. Heaney & L. G. Burriesci (SPIE). https://doi.org/10.1117/12.735602

Barclay, T., Endl, M., Huber, D., et al. 2015, The Astrophysical Journal, 800, 46, doi: http://doi.org/10.1088/0004-637X/800/1/4610.1088/0004-637X/800/1/46

Barnes, J. W., & O'Brien, D. P. 2002, ApJ, 575, 1087, doi: http://doi.org/10.1086/34147710.1086/341477

Barnes, S. A. 2003, ApJ, 586, 464, doi: http://doi.org/10.1086/36763910.1086/367639

Barros, S. C. C., Demangeon, O., Díaz, R. F., et al. 2020, A&A, 634, A75, doi: http://doi.org/10.1051/0004-6361/20193608610.1051/0004-6361/201936086

Bastien, F. A., Stassun, K. G., Basri, G., & Pepper, J. 2013, Nature, 500, 427, doi: http://doi.org/10.1038/nature1241910.1038/nature12419

Batalha, N. M., Rowe, J. F., Bryson, S. T., et al. 2013, ApJS, 204, 24, doi: http://doi.org/10.1088/0067-0049/204/2/2410.1088/0067-0049/204/2/24

Bedding, T. R. 2014, in Asteroseismology, ed. P. L. Pallé & C. Esteban, 60

Beichman, C., Benneke, B., Knutson, H., et al. 2014, Publications of the Astronomical Society of the Pacific, 126, 1134, doi: http://doi.org/10.1086/67956610.1086/679566

Benneke, B., Werner, M., Petigura, E., et al. 2017, ApJ, 834, 187, doi: http://doi.org/10.3847/1538-4357/834/2/18710.3847/1538-4357/834/2/187

Berman, L. 1932, Lick Observatory Bulletin, 443, 24, doi: http://doi.org/10.5479/ADS/bib/1932LicOB.16.24B10.5479/ADS/bib/1932LicOB.16.24B

Berta, Z. K., Charbonneau, D., Désert, J.-M., et al. 2012, ApJ, 747, 35, doi: http://doi.org/10.1088/0004-637X/747/1/3510.1088/0004-637X/747/1/35

Boone, K. 2019, The Astronomical Journal, 158, 257, doi: http://doi.org/10.3847/1538-3881/ab518210.3847/1538-3881/ab5182

Borkovits, T., Derekas, A., Kiss, L. L., et al. 2013, MNRAS, 428, 1656, doi: http://doi.org/10.1093/mnras/sts14610.1093/mnras/sts146

Borkovits, T., Rappaport, S., Kaye, T., et al. 2019, MNRAS, 483, 1934, doi: http://doi.org/10.1093/mnras/sty315710.1093/mnras/sty3157

Borkovits, T., Rappaport, S. A., Tan, T. G., et al. 2020, MNRAS, 496, 4624, doi: http://doi.org/10.1093/mnras/staa181710.1093/mnras/staa1817

Borkovits, T., Mitnyan, T., Rappaport, S. A., et al. 2022, MNRAS, 510, 1352, doi: http://doi.org/10.1093/mnras/stab339710.1093/mnras/stab3397

Borucki, W. J., & Summers, A. L. 1984, Icarus, 58, 121, doi: http://doi.org/10.1016/0019-1035(84)90102-710.1016/0019-1035(84)90102-7

Borucki, W. J., Koch, D., Basri, G., et al. 2010, Science, 327, 977, doi: http://doi.org/10.1126/science.118540210.1126/science.1185402

Bouvier, J. 2008, A&A, 489, L53, doi: http://doi.org/10.1051/0004-6361:20081057410.1051/0004-6361:200810574

Bradbury, R. 2017, The Martian chronicles (New York London Toronto Sydney New Delhi: Simon Schuster Paperbacks)

Brakensiek, J., & Ragozzine, D. 2016, ApJ, 821, 47, doi: http://doi.org/10.3847/0004-637X/821/1/4710.3847/0004-637X/821/1/47

Broeg, C., Fortier, A., Ehrenreich, D., et al. 2013, in European Physical Journal Web of Conferences, Vol. 47, European Physical Journal Web of Conferences, 03005

Brown, T. M., & Charbonneau, D. 2000, in Astronomical Society of the Pacific Conference Series, Vol. 219, Disks, Planetesimals, and Planets, ed. G. Garzón, C. Eiroa, D. de Winter, & T. J. Mahoney, 584

Brown, T. M., Charbonneau, D., Gilliland, R. L., Noyes, R. W., & Burrows, A. 2001, ApJ, 552, 699, doi: http://doi.org/10.1086/32058010.1086/320580

Bui, T. D., & Turner, R. E. 2014, in Advances in Neural Information Processing Systems 27, ed. Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, & K. Q. Weinberger (Curran Associates, Inc.), 2213–2221. http://papers.nips.cc/paper/5459-tree-structured-gaussian-process-approximations.pdf

Bulirsch, R. 1965, Numerische Mathematik, 7, 78

—. 1969a, Numerische Mathematik, 13, 305

—. 1969b, Numerische Mathematik, 13, 266

Byrd, P. F., & Friedman, M. 1954, Mitteilungen der Astronomischen Gesellschaft Hamburg, 5, 99

Campbell, B., Walker, G. A. H., & Yang, S. 1988, ApJ, 331, 902, doi: http://doi.org/10.1086/16660810.1086/166608

Canny, J. 1986, IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-8, 679, doi: http://doi.org/10.1109/TPAMI.1986.476785110.1109/TPAMI.1986.4767851

Canup, R. M., & Asphaug, E. 2001, Nature, 412, 708, doi: http://doi.org/10.1038/3508901010.1038/35089010

Canup, R. M., & Ward, W. R. 2006, Nature, 441, 834, doi: http://doi.org/10.1038/nature0486010.1038/nature04860

Canup, R. M., Righter, K., Dauphas, N., et al. 2021, arXiv e-prints, arXiv:2103.02045, doi: http://doi.org/10.48550/arXiv.2103.0204510.48550/arXiv.2103.02045

Carr, M. H., Masursky, H., Baum, W. A., et al. 1976, Science, 193, 766, doi: http://doi.org/10.1126/science.193.4255.76610.1126/science.193.4255.766

Carter, J. A., Yee, J. C., Eastman, J., Gaudi, B. S., & Winn, J. N. 2008, The Astrophysical Journal, 689, 499, doi: http://doi.org/10.1086/59232110.1086/592321

Carter, J. A., Fabrycky, D. C., Ragozzine, D., et al. 2011, Science, 331, 562, doi: http://doi.org/10.1126/science.120127410.1126/science.1201274

Carter, J. A., Agol, E., Chaplin, W. J., et al. 2012, Science, 337, 556, doi: http://doi.org/10.1126/science.122326910.1126/science.1223269

Casertano, S., Lattanzi, M. G., Sozzetti, A., et al. 2008, A&A, 482, 699, doi: http://doi.org/10.1051/0004-6361:2007899710.1051/0004-6361:20078997

Chakrabarty, A., & Sengupta, S. 2019, AJ, 158, 39, doi: http://doi.org/10.3847/1538-3881/ab24dd10.3847/1538-3881/ab24dd

Charbonneau, D., Brown, T. M., Latham, D. W., & Mayor, M. 2000, ApJL, 529, L45, doi: http://doi.org/10.1086/31245710.1086/312457

Chen, J., & Kipping, D. 2017, ApJ, 834, 17, doi: http://doi.org/10.3847/1538-4357/834/1/1710.3847/1538-4357/834/1/17

Choi, J., Dotter, A., Conroy, C., et al. 2016, ApJ, 823, 102, doi: http://doi.org/10.3847/0004-637X/823/2/10210.3847/0004-637X/823/2/102

Cochran, W. D., & Young, B. W. 1985, in Bulletin of the American Astronomical Society, Vol. 17, 704

Cranmer, S. R., Bastien, F. A., Stassun, K. G., & Saar, S. H. 2014, ApJ, 781, 124, doi: http://doi.org/10.1088/0004-637X/781/2/12410.1088/0004-637X/781/2/124

Crowe, M. J., & Dowd, M. F. 2013, in Astrobiology, History, and Society, ed. D. A. Vakoch, 3

Csató, L., & Opper, M. 2002, Neural Computation, 14, 641, doi: http://doi.org/10.1162/089976602317250933

Csizmadia, S., Smith, A. M. S., Cabrera, J., et al. 2021, arXiv e-prints, arXiv:2108.11822, doi: http://doi.org/10.48550/arXiv.2108.1182210.48550/arXiv.2108.11822

Curtis, J. L., Agüeros, M. A., Douglas, S. T., & Meibom, S. 2019, ApJ, 879, 49, doi: http://doi.org/10.3847/1538-4357/ab239310.3847/1538-4357/ab2393

Curtis, J. L., Agüeros, M. A., Matt, S. P., et al. 2020, arXiv e-prints, arXiv:2010.02272. https://arxiv.org/abs/2010.02272https://arxiv.org/abs/2010.02272

Daemi, A., Kodamana, H., & Huang, B. 2019, Journal of Process Control, 81, 209

Danielson, D. 2013, in Astrobiology, History, and Society, ed. D. A. Vakoch, 57

Davenport, J. R. A. 2017, ApJ, 835, 16, doi: http://doi.org/10.3847/1538-4357/835/1/1610.3847/1538-4357/835/1/16

Davenport, J. R. A., & Covey, K. R. 2018, ApJ, 868, 151, doi: http://doi.org/10.3847/1538-4357/aae84210.3847/1538-4357/aae842

Dawson, R. I., Johnson, J. A., Fabrycky, D. C., et al. 2014, ApJ, 791, 89, doi: http://doi.org/10.1088/0004-637X/791/2/8910.1088/0004-637X/791/2/89

Deeg, H. J., Doyle, L. R., Kozhevnikov, V. P., et al. 1998, A&A, 338, 479, doi: http://doi.org/10.48550/arXiv.astro-ph/980637110.48550/arXiv.astro-ph/9806371

Deisenroth, M. P., & Ng, J. W. 2015, arXiv e-prints, arXiv:1502.02843, doi: http://doi.org/10.48550/arXiv.1502.0284310.48550/arXiv.1502.02843

Denissenkov, P. A., Pinsonneault, M., Terndrup, D. M., & Newsham, G. 2010, ApJ, 716, 1269, doi: http://doi.org/10.1088/0004-637X/716/2/126910.1088/0004-637X/716/2/1269

Dorn, C., Khan, A., Heng, K., et al. 2015, A&A, 577, A83, doi: http://doi.org/10.1051/0004-6361/20142491510.1051/0004-6361/201424915

Dorn, C., Venturini, J., Khan, A., et al. 2017, A&A, 597, A37, doi: http://doi.org/10.1051/0004-6361/20162870810.1051/0004-6361/201628708

Dotter, A. 2016, ApJS, 222, 8, doi: http://doi.org/10.3847/0067-0049/222/1/810.3847/0067-0049/222/1/8

Douglas, S. T., Agüeros, M. A., Covey, K. R., & Kraus, A. 2017, ApJ, 842, 83

Douglas, S. T., Curtis, J. L., Agüeros, M. A., et al. 2019, ApJ, 879, 100, doi: http://doi.org/10.3847/1538-4357/ab246810.3847/1538-4357/ab2468

Doyle, L. R., Carter, J. A., Fabrycky, D. C., et al. 2011, Science, 333, 1602, doi: http://doi.org/10.1126/science.121092310.1126/science.1210923

Durney, B. 1972, Evidence for Changes in the Angular Velocity of the Surface Regions of the Sun and Stars - Comments, ed. C. P. Sonett, P. J. Coleman, & J. M. Wilcox, Vol. 308, 282

Endal, A. S., & Sofia, S. 1981, ApJ, 243, 625, doi: http://doi.org/10.1086/15862810.1086/158628

Fewell, M. P. 2006, Area of common overlap of three circles, Tech. rep., Defence Science and Technology Group

Foreman-Mackey, D., Agol, E., Ambikasaran, S., & Angus, R. 2017, celerite: Scalable 1D Gaussian Processes in C++, Python, and Julia. http://ascl.net/1709.008http://ascl.net/1709.008

Foreman-Mackey, D., Czekala, I., Luger, R., et al. 2019, dfm/exoplanet v0.2.3, doi: http://doi.org/10.5281/zenodo.199844710.5281/zenodo.1998447. https://doi.org/10.5281/zenodo.1998447

Foreman-Mackey, D., Farr, W., Sinha, M., et al. 2019, The Journal of Open Source Software, 4, 1864, doi: http://doi.org/10.21105/joss.0186410.21105/joss.01864

Foreman-Mackey, D., Luger, R., Agol, E., et al. 2021, The Journal of Open Source Software, 6, 3285, doi: http://doi.org/10.21105/joss.0328510.21105/joss.03285

Forgan, D., & Dobos, V. 2016, MNRAS, 457, 1233, doi: http://doi.org/10.1093/mnras/stw02410.1093/mnras/stw024

Forgan, D. H. 2017, MNRAS, 470, 416, doi: http://doi.org/10.1093/mnras/stx121710.1093/mnras/stx121

Frohlich, C., Romero, J., Roth, H., et al. 1995, Solar Physics, 162, 101, doi: http://doi.org/10.1007/BF0073342810.1007/BF00733428

Gaia Collaboration, Katz, D., Antoja, T., et al. 2018, A&A, 616, A11, doi: http://doi.org/10.1051/0004-6361/20183286510.1051/0004-6361/201832865

Gallet, F., & Bouvier, J. 2013, A&A, 556, A36, doi: http://doi.org/10.1051/0004-6361/20132130210.1051/0004-6361/201321302

Garcia-Mejia, J., Charbonneau, D., Fabricant, D., et al. 2020, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 11445, Ground-based and Airborne Telescopes VIII, ed. H. K. Marshall, J. Spyromilio, & T. Usuda, 114457R

Gardner, J. R., Pleiss, G., Bindel, D., Weinberger, K. Q., & Wilson, A. G. 2018, CoRR, abs/1809.11165. https://arxiv.org/abs/1809.11165https://arxiv.org/abs/1809.11165

Garraffo, C., Drake, J. J., Dotter, A., et al. 2018, ApJ, 862, 90, doi: http://doi.org/10.3847/1538-4357/aace5d10.3847/1538-4357/aace5d

Gatewood, G., & Eichhorn, H. 1973, AJ, 78, 769, doi: http://doi.org/10.1086/11148010.1086/111480

Gatewood, G. D., & Stein, J. W. 1986, A survey of nearby stars for the perturbatative effects of extrasolar planetary systems, In NASA, Washington Reports of Planetry Astronomy, 1985 p 145-146 (SEE N87-12407 03-89)

Gelman, A., & Rubin, D. B. 1992, Statistical Science, 7, 457, doi: http://doi.org/10.1214/ss/117701113610.1214/ss/1177011136

Gilbert, G. J. 2022, AJ, 163, 111, doi: http://doi.org/10.3847/1538-3881/ac45f410.3847/1538-3881/ac45f4

Gillon, M., Triaud, A. H. M. J., Demory, B.-O., et al. 2017, Nature, 542, 456, doi: http://doi.org/10.1038/nature2136010.1038/nature21360

Gordon, T. 2022, tagordon/gefera: Gefera 0.1, v0.1, Zenodo, doi: http://doi.org/10.5281/zenodo.681720010.5281/zenodo.6817200. `https://doi.org/10.5281/zenodo.6817200`

Greene, T. P., Bell, T. J., Ducrot, E., et al. 2023, Nature, 618, 39, doi: http://doi.org/10.1038/s41586-023-05951-710.1038/s41586-023-05951-7

Grimm, S. L., Demory, B.-O., Gillon, M., et al. 2018, A&A, 613, A68, doi: http://doi.org/10.1051/0004-6361/20173223310.1051/0004-6361/201732233

Gruner, D., & Barnes, S. A. 2020, A&A, 644, A16, doi: http://doi.org/10.1051/0004-6361/20203898410.1051/0004-6361/202038984

Heller, R. 2014, ApJ, 787, 14, doi: http://doi.org/10.1088/0004-637X/787/1/1410.1088/0004-637X/787/1/14

Heller, R., Williams, D., Kipping, D., et al. 2014, Astrobiology, 14, 798, doi: http://doi.org/10.1089/ast.2014.114710.1089/ast.2014.1147

Hershey, J. L. 1973, AJ, 78, 421, doi: http://doi.org/10.1086/11143610.1086/111436

Hey, D. R., Murphy, S. J., Foreman-Mackey, D., et al. 2020, The Astronomical Journal, 159, 202, doi: http://doi.org/10.3847/1538-3881/ab7d3810.3847/1538-3881/ab7d38

Hippke, M. 2015, ApJ, 806, 51, doi: http://doi.org/10.1088/0004-637X/806/1/5110.1088/0004-637X/806/1/51

Hippke, M., David, T. J., Mulders, G. D., & Heller, R. 2019, AJ, 158, 143, doi: http://doi.org/10.3847/1538-3881/ab398410.3847/1538-3881/ab3984

Hippke, M., & Heller, R. 2022, A&A, 662, A37, doi: http://doi.org/10.1051/0004-6361/20224312910.1051/0004-6361/202243129

Hirano, T., Narita, N., Sato, B., et al. 2012, ApJL, 759, L36, doi: http://doi.org/10.1088/2041-8205/759/2/L3610.1088/2041-8205/759/2/L36

Hoffman, M. D., & Gelman, A. 2014, J. Mach. Learn. Res., 15, 1593

Hojjati, A., Kim, A. G., & Linder, E. V. 2013, Physical Review D, 87, doi: http://doi.org/10.1103/physrevd.87.12351210.1103/physrevd.87.123512

Hojjati, A., & Linder, E. V. 2014, Physical Review D, 90, doi: http://doi.org/10.1103/physrevd.90.12350110.1103/physrevd.90.123501

Holman, M. J. 2005, Science, 307, 1288, doi: http://doi.org/10.1126/science.110782210.1126/science.1107

Husser, T. O., Wende-von Berg, S., Dreizler, S., et al. 2013, A&A, 553, A6, doi: http://doi.org/10.1051/0004-6361/20121905810.1051/0004-6361/201219058

Impey, C. 2013, in Astrobiology, History, and Society, ed. D. A. Vakoch, 201

Jacob, W. S. 1855, MNRAS, 15, 228, doi: http://doi.org/10.1093/mnras/15.9.22810.1093/mnras/15.9.228

Jylänki, P., Vanhatalo, J., & Vehtari, A. 2011, Journal of Machine Learning Research, 12

Kahan, W. 2000, Tech. rep., University of California, Berkeley

Kallinger, T., De Ridder, J., Hekker, S., et al. 2014, Astronomy & Astrophysics, 570, A41, doi: http://doi.org/10.1051/0004-6361/20142431310.1051/0004-6361/201424313

Kane, S. R. 2017, ApJL, 839, L19, doi: http://doi.org/10.3847/2041-8213/aa6bf210.3847/2041-8213/aa6bf2

Karoff, C., Campante, T. L., Ballot, J., et al. 2013, ApJ, 767, 34, doi: http://doi.org/10.1088/0004-637X/767/1/3410.1088/0004-637X/767/1/34

Kasting, J. F., Whitmire, D. P., & Reynolds, R. T. 1993, Icarus, 101, 108, doi: http://doi.org/10.1006/icar.1993.101010.1006/icar.1993.1010

Kelly, B. C., Bechtold, J., & Siemiginowska, A. 2009, The Astrophysical Journal, 698, 895, doi: http://doi.org/10.1088/0004-637x/698/1/89510.1088/0004-637x/698/1/895

Kelly, B. C., Becker, A. C., Sobolewska, M., Siemiginowska, A., & Uttley, P. 2014, The Astrophysical Journal, 788, 33, doi: http://doi.org/10.1088/0004-637X/788/1/3310.1088/0004-637X/788/1/33

Kim, A. G., Thomas, R. C., Aldering, G., et al. 2013, The Astrophysical Journal, 766, 84, doi: http://doi.org/10.1088/0004-637x/766/2/8410.1088/0004-637x/766/2/84

Kipping, D., Bryson, S., Burke, C., et al. 2022, Nature Astronomy, doi: http://doi.org/10.1038/s41550-021-01539-110.1038/s41550-021-01539-1

Kipping, D. M. 2009, MNRAS, 392, 181, doi: http://doi.org/10.1111/j.1365-2966.2008.13999.x10.1111/j.1365-2966.2008.13999.x

—. 2010, MNRAS, 408, 1758, doi: http://doi.org/10.1111/j.1365-2966.2010.17242.x10.1111/j.1365-2966.2010.17242.x

—. 2011, MNRAS, 416, 689, doi: http://doi.org/10.1111/j.1365-2966.2011.19086.x10.1111/j.1365-2966.2011.19086.x

Kipping, D. M., Hartman, J., Buchhave, L. A., et al. 2013, ApJ, 770, 101, doi: http://doi.org/10.1088/0004-637X/770/2/10110.1088/0004-637X/770/2/101

Kipping, D. M., Huang, X., Nesvorný, D., et al. 2015, ApJL, 799, L14, doi: http://doi.org/10.1088/2041-8205/799/1/L1410.1088/2041-8205/799/1/L14

Kreidberg, L., Luger, R., & Bedell, M. 2019, ApJL, 877, L15, doi: http://doi.org/10.3847/2041-8213/ab20c810.3847/2041-8213/ab20c8

Lanzafame, A. C., & Spada, F. 2015, A&A, 584, A30, doi: http://doi.org/10.1051/0004-6361/20152677010.1051/0004-6361/201526770

Laskar, J., Joutel, F., & Robutel, P. 1993, Nature, 361, 615, doi: http://doi.org/10.1038/361615a010.1038/361615a0

Li, Z.-Z., Li, L., & Shao, Z. 2021, Astronomy and Computing, 36, 100483, doi: http://doi.org/10.1016/j.ascom.2021.10048310.1016/j.ascom.2021.100483

Libby-Roberts, J. E., Berta-Thompson, Z. K., Désert, J.-M., et al. 2020, AJ, 159, 57, doi: http://doi.org/10.3847/1538-3881/ab5d3610.3847/1538-3881/ab5d36

Lightkurve Collaboration, Cardoso, J. V. d. M., Hedges, C., et al. 2018, Lightkurve: Kepler and TESS time series analysis in Python, Astrophysics Source Code Library. http://ascl.net/1812.013http://ascl.net/1812.013

Lissauer, J. J., Barnes, J. W., & Chambers, J. E. 2012, Icarus, 217, 77, doi: http://doi.org/10.1016/j.icarus.2011.10.01310.1016/j.icarus.2011.10.013

Loper, J., Blei, D., Cunningham, J. P., & Paninski, L. 2020, arXiv e-prints, arXiv:2003.05554. https://arxiv.org/abs/2003.05554https://arxiv.org/abs/2003.05554

Lowell, P. 1897, Bulletin de la Societe Astronomique de France et Revue Mensuelle d'Astronomie, de Meteorologie et de Physique du Globe, 11, 220

Luger, R., Agol, E., Foreman-Mackey, D., et al. 2019, AJ, 157, 64, doi: http://doi.org/10.3847/1538-3881/aae8e510.3847/1538-3881/aae8e5

Luger, R., & Barnes, R. 2015, Astrobiology, 15, 119, doi: http://doi.org/10.1089/ast.2014.123110.1089/ast.2014.1231

Luger, R., Kruse, E., Foreman-Mackey, D., Agol, E., & Saunders, N. 2018, AJ, 156, 99, doi: http://doi.org/10.3847/1538-3881/aad23010.3847/1538-3881/aad230

Luger, R., Lustig-Yaeger, J., & Agol, E. 2017, ApJ, 851, 94, doi: http://doi.org/10.3847/1538-4357/aa9c4310.3847/1538-4357/aa9c43

Lustig-Yaeger, J., Fu, G., May, E. M., et al. 2023, arXiv e-prints, arXiv:2301.04191, doi: http://doi.org/10.48550/arXiv.2301.0419110.48550/arXiv.2301.04191

MacGregor, K. B., & Brenner, M. 1991, ApJ, 376, 204, doi: http://doi.org/10.1086/17026910.1086/170269

MacLeod, C. L., Ivezić, Ž., Kochanek, C. S., et al. 2010, ApJ, 721, 1014, doi: http://doi.org/10.1088/0004-637X/721/2/101410.1088/0004-637X/721/2/1014

Magic, Z., & Asplund, M. 2014, arXiv e-prints, arXiv:1405.7628. https://arxiv.org/abs/1405.7628https://arxiv.org/abs/1405.7628

Mahadevan, S., Bender, C. F., Hambleton, K., et al. 2019, The Astrophysical Journal, 884, 126, doi: http://doi.org/10.3847/1538-4357/ab379310.3847/1538-4357/ab3793

Mandel, K., & Agol, E. 2002, ApJL, 580, L171, doi: http://doi.org/10.1086/34552010.1086/345520

Mandell, A. M., Haynes, K., Sinukoff, E., et al. 2013, ApJ, 779, 128, doi: http://doi.org/10.1088/0004-637X/779/2/12810.1088/0004-637X/779/2/128

Mann, A. W., Gaidos, E., & Aldering, G. 2011, PASP, 123, 1273, doi: http://doi.org/10.1086/66264010.1086/662640

Martin, D. V., Fabrycky, D. C., & Montet, B. T. 2019, ApJL, 875, L25, doi: http://doi.org/10.3847/2041-8213/ab0aea10.3847/2041-8213/ab0aea

Masuda, K. 2014, ApJ, 783, 53, doi: http://doi.org/10.1088/0004-637X/783/1/5310.1088/0004-637X/783/1/53

Masuda, K., Hirano, T., Taruya, A., Nagasawa, M., & Suto, Y. 2013, ApJ, 778, 185, doi: http://doi.org/10.1088/0004-637X/778/2/18510.1088/0004-637X/778/2/185

Mayor, M., & Queloz, D. 1995, Nature, 378, 355, doi: http://doi.org/10.1038/378355a010.1038/378355a0

Mazeh, T., & Faigler, S. 2010, A&A, 521, L59, doi: http://doi.org/10.1051/0004-6361/20101555010.1051/0004-6361/201015550

McMillan, R. S., Frecker, J. E., Smith, P. H., Roland, E. H., & Merline, W. J. 1982, in Bulletin of the American Astronomical Society, Vol. 14, 625

McQuillan, A., Aigrain, S., & Mazeh, T. 2013, MNRAS, 432, 1203, doi: http://doi.org/10.1093/mnras/stt53610.1093/mnras/stt536

McQuillan, A., Mazeh, T., & Aigrain, S. 2014, ApJS, 211, 24, doi: http://doi.org/10.1088/0067-0049/211/2/2410.1088/0067-0049/211/2/24

Meibom, S., Barnes, S. A., Latham, D. W., et al. 2011, ApJL, 733, L9

Messeri, L. 2016, Placing outer space: an Earthly ethnography of other worlds, Experimental futures: technological lives, scientific arts, anthropological voices (Durham: Duke University Press)

Mitnyan, T., Borkovits, T., Rappaport, S. A., Pál, A., & Maxted, P. F. L. 2020, MNRAS, 498, 6034, doi: http://doi.org/10.1093/mnras/staa276210.1093/mnras/staa2762

Morin, J., Donati, J. F., Petit, P., et al. 2010, MNRAS, 407, 2269, doi: http://doi.org/10.1111/j.1365-2966.2010.17101.x10.1111/j.1365-2966.2010.17101.x

Morris, B. M., Bobra, M. G., Agol, E., Lee, Y. J., & Hawley, S. L. 2020, Monthly Notices of the Royal Astronomical Society, 493, 5489, doi: http://doi.org/10.1093/mnras/staa61810.1093/mnras/staa618

Namouni, F. 2010, ApJL, 719, L145, doi: http://doi.org/10.1088/2041-8205/719/2/L14510.1088/2041-8205/719/2/L145

NASA Exoplanet Archive. YYYY, Planetary Systems, Version: YYYY-MM-DD HH:MM, NExScI-Caltech/IPAC, doi: http://doi.org/10.26133/NEA1210.26133/NEA12. `https://catcopy.ipac.caltech.edu/dois/doi.php?id=10.26133/NEA12`

National Academies of Sciences, E., & Medicine. 2021, Pathways to Discovery in Astronomy and Astrophysics for the 2020s, doi: http://doi.org/10.17226/2614110.17226/26141

Newton, E. R., Irwin, J., Charbonneau, D., et al. 2016, ApJ, 821, 93, doi: http://doi.org/10.3847/0004-637X/821/2/9310.3847/0004-637X/821/2/93

Nickson, T., Gunter, T., Lloyd, C., Osborne, M. A., & Roberts, S. 2015, Blitzkriging: Kronecker-structured Stochastic Gaussian Processes. https://arxiv.org/abs/1510.07965https://arxiv.org/abs/1510.07965

Nielsen, M. B., Gizon, L., Schunker, H., & Karoff, C. 2013, A&A, 557, L10

Noyola, J. P., Satyal, S., & Musielak, Z. E. 2014, ApJ, 791, 25, doi: http://doi.org/10.1088/0004-637X/791/1/2510.1088/0004-637X/791/1/25

Osborn, J., Föhring, D., Dhillon, V. S., & Wilson, R. W. 2015, MNRAS, 452, 1707, doi: http://doi.org/10.1093/mnras/stv140010.1093/mnras/stv1400

Osborn, J., Wilson, R. W., Dhillon, V. S., Avila, R., & Love, G. D. 2011, MNRAS, 411, 1223, doi: http://doi.org/10.1111/j.1365-2966.2010.17759.x10.1111/j.1365-2966.2010.17759.x

Pál, A. 2012, MNRAS, 420, 1630, doi: http://doi.org/10.1111/j.1365-2966.2011.20151.x10.1111/j.1365-2966.2011.20151.x

Pancoast, A., Brewer, B. J., Treu, T., et al. 2014, Monthly Notices of the Royal Astronomical Society, 445, 3073, doi: http://doi.org/10.1093/mnras/stu141910.1093/mnras/stu1419

Pande, D., Bedding, T. R., Huber, D., & Kjeldsen, H. 2018, MNRAS, 480, 467, doi: http://doi.org/10.1093/mnras/sty186910.1093/mnras/sty1869

Park, C., Borth, D. J., Wilson, N. S., Hunter, C. N., & Friedersdorf, F. J. 2022, Pattern Recognition, 124, 108444, doi: http://doi.org/10.1016/j.patcog.2021.10844410.1016/j.patcog.2021.108444

Paxton, B., Bildsten, L., Dotter, A., et al. 2011, ApJS, 192, 3, doi: http://doi.org/10.1088/0067-0049/192/1/310.1088/0067-0049/192/1/3

Paxton, B., Cantiello, M., Arras, P., et al. 2013, ApJS, 208, 4, doi: http://doi.org/10.1088/0067-0049/208/1/410.1088/0067-0049/208/1/4

Paxton, B., Marchant, P., Schwab, J., et al. 2015, ApJS, 220, 15, doi: http://doi.org/10.1088/0067-0049/220/1/1510.1088/0067-0049/220/1/15

Pereira, F., Campante, T. L., Cunha, M. S., et al. 2019, MNRAS, 489, 5764, doi: http://doi.org/10.1093/mnras/stz240510.1093/mnras/stz2405

Perryman, M., Hartman, J., Bakos, G. Á., & Lindegren, L. 2014, ApJ, 797, 14, doi: http://doi.org/10.1088/0004-637X/797/1/1410.1088/0004-637X/797/1/14

Peters, C. M., Richards, G. T., Myers, A. D., et al. 2015, The Astrophysical Journal, 811, 95, doi: http://doi.org/10.1088/0004-637x/811/2/9510.1088/0004-637x/811/2/95

Peters, M. A., & Turner, E. L. 2013, ApJ, 769, 98, doi: http://doi.org/10.1088/0004-637X/769/2/9810.1088/0004-637X/769/2/98

Piaulet, C., Benneke, B., Almenara, J. M., et al. 2023, Nature Astronomy, 7, 206, doi: http://doi.org/10.1038/s41550-022-01835-410.1038/s41550-022-01835-4

Piro, A. L. 2018, AJ, 156, 54, doi: http://doi.org/10.3847/1538-3881/aaca3810.3847/1538-3881/aaca38

Pompe, E., Holmes, C., & Łatuszyński, K. 2018, arXiv e-prints, arXiv:1812.02609. https://arxiv.org/abs/1812.02609https://arxiv.org/abs/1812.02609

Press, W. H., Rybicki, G. B., & Hewitt, J. N. 1992, The Astrophysical Journal, 385, 404, doi: http://doi.org/10.1086/17095110.1086/170951

Price, E. M., & Rogers, L. A. 2014, The Astrophysical Journal, 794, 92, doi: http://doi.org/10.1088/0004-637x/794/1/9210.1088/0004-637x/794/1/92

Quintana, E. V., Jenkins, J. M., Clarke, B. D., et al. 2010, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7740, Software and Cyberinfrastructure for Astronomy, ed. N. M. Radziwill & A. Bridger, 77401X

Rackham, B. V., Apai, D., & Giampapa, M. S. 2018, ApJ, 853, 122, doi: http://doi.org/10.3847/1538-4357/aaa08c10.3847/1538-4357/aaa08c

—. 2019, AJ, 157, 96, doi: http://doi.org/10.3847/1538-3881/aaf89210.3847/1538-3881/aaf892

Radica, M., Welbanks, L., Espinoza, N., et al. 2023, arXiv e-prints, arXiv:2305.17001, doi: http://doi.org/10.48550/arXiv.2305.1700110.48550/arXiv.2305.17001

Ragozzine, D., & Holman, M. J. 2010, arXiv e-prints, arXiv:1006.3727. https://arxiv.org/abs/1006.3727https://arxiv.org/abs/1006.3727

Rajpaul, V., Aigrain, S., Osborne, M. A., Reece, S., & Roberts, S. 2015, MNRAS, 452, 2269, doi: http://doi.org/10.1093/mnras/stv142810.1093/mnras/stv1428

Ramírez, V., & Sucerquia, M., A. J. Z. J. 2018, in Diversis Mundi: The Solar System in an Exoplanetary Context, 26

Raposo-Pulido, V., & Peláez, J. 2017, MNRAS, 467, 1702, doi: http://doi.org/10.1093/mnras/stx13810.1093/mnras/stx138

Rasmussen, C. E., & Williams, C. K. I. 2006, Gaussian Processes for Machine Learning

Rauer, H., Catala, C., Aerts, C., et al. 2014, Experimental Astronomy, 38, 249, doi: http://doi.org/10.1007/s10686-014-9383-410.1007/s10686-014-9383-4

Rebull, L. M., Stauffer, J. R., Cody, A. M., et al. 2018, AJ, 155, 196, doi: http://doi.org/10.3847/1538-3881/aab60510.3847/1538-3881/aab605

Rebull, L. M., Stauffer, J. R., Bouvier, J., et al. 2016, AJ, 152, 113, doi: http://doi.org/10.3847/0004-6256/152/5/11310.3847/0004-6256/152/5/113

Reinhold, T., Bell, K. J., Kuszlewicz, J., Hekker, S., & Shapiro, A. I. 2019, A&A, 621, A21, doi: http://doi.org/10.1051/0004-6361/20183375410.1051/0004-6361/201833754

Reinhold, T., & Hekker, S. 2020, A&A, 635, A43, doi: http://doi.org/10.1051/0004-6361/20193688710.1051/0004-6361/201936887

Reinhold, T., Reiners, A., & Basri, G. 2013, A&A, 560, A4, doi: http://doi.org/10.1051/0004-6361/20132197010.1051/0004-6361/201321970

Reuyl, D., & Holmberg, E. 1943, ApJ, 97, 41, doi: http://doi.org/10.1086/14448910.1086/144489

Reynolds, R. T., McKay, C. P., & Kasting, J. F. 1987, Advances in Space Research, 7, 125, doi: http://doi.org/10.1016/0273-1177(87)90364-410.1016/0273-1177(87)90364-4

Ricker, G. R., Winn, J. N., Vanderspek, R., et al. 2015, Journal of Astronomical Telescopes, Instruments, and Systems, 1, 014003, doi: http://doi.org/10.1117/1.JATIS.1.1.01400310.1117/1.JATIS.1.1.014003

Rogers, L. A., & Seager, S. 2010, ApJ, 712, 974, doi: http://doi.org/10.1088/0004-637X/712/2/97410.1088/0004-637X/712/2/974

Rosenblatt, F. 1971, Icarus, 14, 71, doi: http://doi.org/10.1016/0019-1035(71)90103-510.1016/0019-1035(71)90103-5

Rustamkulov, Z., Sing, D. K., Liu, R., & Wang, A. 2022, ApJL, 928, L7, doi: http://doi.org/10.3847/2041-8213/ac5b6f10.3847/2041-8213/ac5b6f

Rybicki, G. B., & Press, W. H. 1992, The Astrophysical Journal, 398, 169, doi: http://doi.org/10.1086/17184510.1086/171845

—. 1995, Physical Review Letters, 74, 1060, doi: http://doi.org/10.1103/physrevlett.74.106010.1103/phys

Sagan, C., & Lederberg, J. 1976, Icarus, 28, 291, doi: http://doi.org/10.1016/0019-1035(76)90039-710.1016/0019-1035(76)90039-7

Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. 2016, PeerJ Computer Science, 2, e55, doi: http://doi.org/10.7717/peerj-cs.5510.7717/peerj-cs.55

Sarkar, S., Argyriou, I., Vandenbussche, B., Papageorgiou, A., & Pascale, E. 2018, MNRAS, 481, 2871, doi: http://doi.org/10.1093/mnras/sty245310.1093/mnras/sty2453

Sarkar, S., Madhusudhan, N., & Papageorgiou, A. 2020, MNRAS, 491, 378, doi: http://doi.org/10.1093/mnras/stz295810.1093/mnras/stz2958

Sasaki, T., & Barnes, J. W. 2014, International Journal of Astrobiology, 13, 324, doi: http://doi.org/10.1017/S147355041400018410.1017/S1473550414000184

Scharf, C. A. 2006, ApJ, 648, 1196, doi: http://doi.org/10.1086/50525610.1086/505256

Schneider, J., & Doyle, L. R. 1995, Earth Moon and Planets, 71, 153, doi: http://doi.org/10.1007/BF0061287510.1007/BF00612875

See, T. J. J. 1896, AJ, 16, 17, doi: http://doi.org/10.1086/10236810.1086/102368

Shah, A., Wilson, A., & Ghahramani, Z. 2014, in Artificial intelligence and statistics, 877–885

Short, D. R., Orosz, J. A., Windmiller, G., & Welsh, W. F. 2018, AJ, 156, 297, doi: http://doi.org/10.3847/1538-3881/aae88910.3847/1538-3881/aae889

Simon, A. E., Szabó, G. M., Kiss, L. L., Fortier, A., & Benz, W. 2015, PASP, 127, 1084, doi: http://doi.org/10.1086/68339210.1086/683392

Simon, A. E., Szabó, G. M., Kiss, L. L., & Szatmáry, K. 2012, MNRAS, 419, 164, doi: http://doi.org/10.1111/j.1365-2966.2011.19682.x10.1111/j.1365-2966.2011.19682.x

Skumanich, A. 1972, ApJ, 171, 565, doi: http://doi.org/10.1086/15131010.1086/151310

Smith, J. C., Stumpe, M. C., Van Cleve, J. E., et al. 2012, PASP, 124, 1000, doi: http://doi.org/10.1086/66769710.1086/667697

Snelson, E., & Ghahramani, Z. 2006, in Advances in neural information processing systems, 1257–1264

Somers, G., & Pinsonneault, M. H. 2016, ApJ, 829, 32, doi: http://doi.org/10.3847/0004-637X/829/1/3210.3847/0004-637X/829/1/32

Spada, F., & Lanzafame, A. C. 2020, A&A, 636, A76, doi: http://doi.org/10.1051/0004-6361/20193638410.1051/0004-6361/201936384

Spalding, C., Batygin, K., & Adams, F. C. 2016, ApJ, 817, 18, doi: http://doi.org/10.3847/0004-637X/817/1/1810.3847/0004-637X/817/1/18

Stefansson, G., Mahadevan, S., Hebb, L., et al. 2017, ApJ, 848, 9, doi: http://doi.org/10.3847/1538-4357/aa88aa10.3847/1538-4357/aa88aa

Strand, K. A. 1943, PASP, 55, 29, doi: http://doi.org/10.1086/12548410.1086/125484

Struve, O. 1952, The Observatory, 72, 199

Sucerquia, M., Ramírez, V., Alvarado-Montes, J. A., & Zuluaga, J. I. 2020, MNRAS, 492, 3499, doi: http://doi.org/10.1093/mnras/stz354810.1093/mnras/stz3548

Suissa, G., & Kipping, D. 2018, Research Notes of the American Astronomical Society, 2, 31, doi: http://doi.org/10.3847/2515-5172/aac32f10.3847/2515-5172/aac32f

Sulis, S., Lendl, M., Hofmeister, S., et al. 2020, Astronomy & Astrophysics, 636, A70, doi: http://doi.org/10.1051/0004-6361/20193741210.1051/0004-6361/201937412

Tang, Q., Niu, L., Wang, Y., et al. 2017, in IJCAI, 2822–2828

Teachey, A., Kipping, D., Burke, C. J., Angus, R., & Howard, A. W. 2019, arXiv e-prints, arXiv:1904.11896. https://arxiv.org/abs/1904.11896https://arxiv.org/abs/1904.11896

Teachey, A., & Kipping, D. M. 2018, Science Advances, 4, eaav1784, doi: http://doi.org/10.1126/sciadv.aav178410.1126/sciadv.aav1784

Tinetti, G., Drossart, P., Eccleston, P., et al. 2018, Experimental Astronomy, 46, 135, doi: http://doi.org/10.1007/s10686-018-9598-x10.1007/s10686-018-9598-x

Tjoa, J. N. K. Y., Mueller, M., & van der Tak, F. F. S. 2020, A&A, 636, A50, doi: http://doi.org/10.1051/0004-6361/20193703510.1051/0004-6361/201937035

Tokadjian, A., & Piro, A. L. 2020, arXiv e-prints, arXiv:2007.01487. https://arxiv.org/abs/2007.01487https://arxiv.org/abs/2007.01487

Torrie, G. M., & Valleau, J. P. 1977, Journal of Computational Physics, 23, 187, doi: http://doi.org/10.1016/0021-9991(77)90121-810.1016/0021-9991(77)90121-8

Tracey, B. D., & Wolpert, D. H. 2018, arXiv e-prints, arXiv:1801.06147. https://arxiv.org/abs/1801.06147https://arxiv.org/abs/1801.06147

Trani, A. A., Hamers, A., & Geller, A. 2019, in AAS/Division for Extreme Solar Systems Abstracts, Vol. 51, AAS/Division for Extreme Solar Systems Abstracts, 318.18

Trani, A. A., Hamers, A. S., Geller, A., & Spera, M. 2020, MNRAS, 499, 4195, doi: http://doi.org/10.1093/mnras/staa309810.1093/mnras/staa3098

Uttley, P., McHardy, I. M., & Vaughan, S. 2005, Monthly Notices of the Royal Astronomical Society, 359, 345, doi: http://doi.org/10.1111/j.1365-2966.2005.08886.x10.1111/j.1365-2966.2005.08886.x

Vallisneri, M. 2008, Physical Review D, 77, doi: http://doi.org/10.1103/physrevd.77.04200110.1103/phys

van de Kamp, P. 1963, AJ, 68, 515, doi: http://doi.org/10.1086/10900110.1086/109001

—. 1969a, AJ, 74, 238, doi: http://doi.org/10.1086/11079910.1086/110799

—. 1969b, AJ, 74, 757, doi: http://doi.org/10.1086/11085210.1086/110852

—. 1975, AJ, 80, 658, doi: http://doi.org/10.1086/11179110.1086/111791

—. 1982, Vistas in Astronomy, 26, 141, doi: http://doi.org/10.1016/0083-6656(82)90004-610.1016/0083-6656(82)90004-6

van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., et al. 2014, PeerJ, 2, e453, doi: http://doi.org/10.7717/peerj.45310.7717/peerj.453

van Saders, J. L., Pinsonneault, M. H., & Barbieri, M. 2019, ApJ, 872, 128, doi: http://doi.org/10.3847/1538-4357/aafafe10.3847/1538-4357/aafafe

Vanderburg, A., Rappaport, S. A., & Mayo, A. W. 2018, AJ, 156, 184, doi: http://doi.org/10.3847/1538-3881/aae0fc10.3847/1538-3881/aae0fc

Vanhatalo, J., Jylänki, P., & Vehtari, A. 2009, in Advances in neural information processing systems, 1910–1918

Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, Nature Methods, 17, 261, doi: http://doi.org/https://doi.org/10.1038/s41592-019-0686-2https://doi.org/10.1038/s41592-019-0686-2

Ward, P., & Brownlee, D. 2000, Rare earth : why complex life is uncommon in the universe

Welsh, W. F., Orosz, J. A., Carter, J. A., et al. 2012, Nature, 481, 475, doi: http://doi.org/10.1038/nature1076810.1038/nature10768

Wiechert, U., Halliday, A. N., Lee, D. C., et al. 2001, Science, 294, 345, doi: http://doi.org/10.1126/science.106303710.1126/science.1063037

Willard, B. T., Osthege, M., Ho, G., et al. 2021, pymc-devs/aesara:, rel-2.0.7, Zenodo, doi: http://doi.org/10.5281/zenodo.469533110.5281/zenodo.4695331. https://doi.org/10.5281/zenodo.4695331

Wilson, A., & Adams, R. 2013, in International conference on machine learning, 1067–1075

Wilson, A. G., & Nickisch, H. 2015, arXiv e-prints, arXiv:1503.01057. https://arxiv.org/abs/1503.01057https://arxiv.org/abs/1503.01057

Wilson, R. E., & Devinney, E. J. 1971, ApJ, 166, 605, doi: http://doi.org/10.1086/15098610.1086/150986

Winn, J. N. 2010, in Exoplanets, ed. S. Seager, 55–77

Wolszczan, A., & Frail, D. A. 1992, Nature, 355, 145, doi: http://doi.org/10.1038/355145a010.1038/355145a0

Young, E. D., Kohl, I. E., Warren, P. H., et al. 2016, Science, 351, 493, doi: http://doi.org/10.1126/science.aad052510.1126/science.aad0525

Zhang, Y., Leithead, W. E., & Leith, D. J. 2005, in Proceedings of the 44th IEEE Conference on Decision and Control, IEEE, 3711–3716

Zhang, Z., Zhou, Y., Rackham, B. V., & Apai, D. 2018, AJ, 156, 178, doi: http://doi.org/10.3847/1538-3881/aade4f10.3847/1538-3881/aade4f

Zu, Y., Kochanek, C. S., & Peterson, B. M. 2011, The Astrophysical Journal, 735, 80, doi: http://doi.org/10.1088/0004-637x/735/2/8010.1088/0004-637x/735/2/80

Appendix A

CHAPTER 3

A.1 `Celerite` *algorithm for arbitrary covariance matrix in second dimension*

In this section we assume that the covariance in the second dimension, defined by the covariance matrix R, is arbitrary, subject to the constraint that the full covariance matrix K must be positive definite.

We start by rewriting T in terms of the celerite generator matrices A, U, and V from equation 3.32:

$$\begin{aligned} K &= \Sigma + \left[A_0 + \mathrm{tril}(UV^{\mathrm{T}}) + \mathrm{triu}(VU^{\mathrm{T}})\right] \otimes R \\ &= \Sigma + \mathrm{diag}(A_0 \otimes R) \\ &+ \mathrm{tril}(UV^{\mathrm{T}} \otimes R) + \mathrm{triu}(VU^{\mathrm{T}} \otimes R) \end{aligned} \tag{A.1}$$

We rewrite R as RI_M where I_M is the $M \times M$ identity matrix, which allows us to write K as

$$\begin{aligned} K = \Sigma &+ \mathrm{diag}(A_0 \otimes R) \\ &+ \mathrm{tril}((U \otimes R)(V \otimes I_M)^{\mathrm{T}}) \\ &+ \mathrm{triu}((V \otimes I_M)(U \otimes R)^{\mathrm{T}}) \end{aligned} \tag{A.2}$$

where we have again applied equation 3.48. As for the outer product case, we now have a semi-separable matrix defined by a new set of generators:

$$\begin{aligned} A' &= \Sigma + \mathrm{diag}(A_0 \otimes T) \\ U' &= U \otimes T \\ V' &= V \otimes I_M. \end{aligned} \tag{A.3}$$

In terms of the `celerite` coefficients the refactored generator matrices are defined element-

wise as follows:

$$\begin{aligned}
A'_{(n-1)M+p,(n-1)M+p} &= \sigma^2_{(n-1)M+1} + R_{p,p}\sum_{j=1}^{J} a_j \\
\tilde{U}'_{(n-1)M+p,(2j-1)M+q} &= R_{p,q}\tilde{U}_{n,2j-1} \\
\tilde{U}'_{(n-1)M+p,2jM+q} &= R_{p,q}\tilde{U}_{n,2j} \\
\tilde{V}'_{(n-1)M+p,(2j-1)M+q} &= \delta_{p,q}\tilde{V}_{n,2j-1} \\
\tilde{V}'_{(n-1)M+p,2jM+q} &= \delta_{p,q}\tilde{V}_{n,2j},
\end{aligned} \tag{A.4}$$

where $\tilde{U}$ and $\tilde{V}$ are the refactored generator matrices defined in equation 3.38, n ranges over $(1, N)$, p and q range over $(1, M)$, and $\delta_{p,q}$ is the Kronecker delta function:

$$\delta_{p,q} = \begin{cases} 1 & p = q \\ 0 & p \neq q \end{cases}. \tag{A.5}$$

The recursive algorithm for carrying out the Cholesky decomposition is identical to the outer-product case. Starting with $D_{1,1} = A'_{1,1}$ and $\tilde{W}_{1,j} = \tilde{V}_{1,j}/D_{1,1}$, we then recursively define:

$$\begin{aligned}
S_{n,j,k} &= \phi'_{n,j}\phi'_{n,k}\left[S_{n-1,j,k} + D_{n-1,n-1}\tilde{W}_{n-1,j}\tilde{W}_{n-1,k}\right], \\
D_{n,n} &= A'_{n,n} - \sum_{j=1}^{P}\sum_{k=1}^{P}\tilde{U}'_{n,j}S_{n,j,k}\tilde{U}'_{n,k}, \\
\tilde{W}_{n,j} &= \frac{1}{D_{n,n}}\left[\tilde{V}'_{n,j} - \sum_{k=1}^{P}\tilde{U}'_{n,k}S_{n,j,k}\right],
\end{aligned} \tag{A.6}$$

for $n = 2, ..., N'$, $N' = NM$, with $P = 2JM$ the number of rows in $\tilde{U}'$ and $\tilde{V}'$. This additional factor of M accounts for the relatively poorer scaling of the method for arbitrary R over the outer-product case. For arbitrary definitions of R, $P = 2JM$ and the Cholesky decomposition thus scales as $\mathcal{O}(NJ^2M^3)$.

A.2 Computing the log-likelihood

The log-likelihood is given by

$$\ln\ \mathcal{L} = -\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})^{\mathrm{T}} K^{-1}(\boldsymbol{y}-\boldsymbol{\mu}) - \frac{1}{2}\ln\ \det(K) - \frac{N'}{2}\ln(2\pi), \tag{A.7}$$

which incorporates both the inverse and log-determinant of the covariance matrix, K. We therefore begin by describing the algorithms for each of these computations separately. The following algorithm comes directly from the original `celerite` paper, but with our modified definitions of the semi-separable matrix components, $\tilde{U}', \tilde{V}'$ and $\tilde{W}$, and $\phi'_{n,j}$ rather than $\phi_{n,j}$ (see §3.4.2).

The product of the inverse covariance matrix with a vector, $\boldsymbol{z} = K^{-1}\boldsymbol{y}$, is computed with a two-part algorithm. We first compute the intermediary $\boldsymbol{z'}$, setting $z'_1 = y_1$, and then using the recursion relation

$$f_{n,j} = \phi'_{n,j}\left[f_{n-1,j} + \tilde{W}_{n-1,j} z'_{n-1}\right] \tag{A.8}$$

$$z'_n = y_n - \sum_{j=1}^{P} \tilde{U}'_{n,j} f_{n,j}, \tag{A.9}$$

for $n = 2, ..., N'$, where $N' = NM$ and $f_{0,j} = 0$ for all j. We then use $\boldsymbol{z'}$ to compute $\boldsymbol{z}$ in the second step of the algorithm, first setting $z_{N'} = z'_{N'}/D_{N',N'}$, and then using downward recursion

$$g_{n,j} = \phi'_{n+1,j}\left[g_{n+1,j} + \tilde{U}'_{n+1,j} z_{n+1}\right] \tag{A.10}$$

$$z_n = \frac{z'_n}{D_{n,n}} - \sum_{j=1}^{P} \tilde{W}_{n,j} g_{n,j} \tag{A.11}$$

for $n = N' - 1, ..., 1$, where $g_{N',j} = 0$ for all j and P is the number of columns in $\tilde{U}'$, $\tilde{V}'$, and $\tilde{W}$.

The log-determinant of K is given by

$$\ln(\det K) = \sum_{n=1}^{N'} \ln(D_{n,n}). \tag{A.12}$$

Putting these two steps together we can compute the log-likelihood. Because the algorithm for taking products of the inverse requires $\mathcal{O}(NMP)$ operations, whereas the log-determinant can be computed in only $\mathcal{O}(NM)$ operations, the log-likelihood computation as a whole scales as $\mathcal{O}(NMP)$. In practice, the bottleneck for applications such as maximizing the likelihood or MCMC is computing the Cholesky factor rather than computing the log-likelihood, since the log-likelihood computation itself is faster by $\mathcal{O}(P)$. Again we have $P = 2J$ when R is an outer product and $P = 2JM$ when R is any arbitrary covariance matrix.

A.3 Prediction Algorithm

A Gaussian process prediction is an interpolation or extrapolation of the observed data using with the GP model. A prediction evaluated at each datapoint can also be thought of as a smoothing operation as it yields an estimate of the function with white noise removed.

The predictive distribution of a Gaussian process is a multivariate normal with a mean μ^* and covariance K^* evaluated at the input coordinates $\boldsymbol{x}^*$. For a GP with no white noise component the mean is constrained to pass directly through each observation of the data points $\boldsymbol{y}$. For a GP with a non-zero white noise component the GP will act as a filter such that when the mean is subtracted from the data the residuals will be distributed according to a Gaussian distribution whose width is given by the GP white noise.

The predictive mean and covariance are computed as follows

$$\boldsymbol{\mu^*} = \boldsymbol{\mu_\theta}(\boldsymbol{x^*}) + K(\boldsymbol{x^*}, \boldsymbol{x})K(\boldsymbol{x}, \boldsymbol{x})^{-1}[\boldsymbol{y} - \boldsymbol{\mu_\theta}(\boldsymbol{x})] \tag{A.13}$$

$$K^* = K(\boldsymbol{x^*}, \boldsymbol{x^*}) - K(\boldsymbol{x^*}, \boldsymbol{x})K(\boldsymbol{x}, \boldsymbol{x})^{-1}K(\boldsymbol{x}, \boldsymbol{x^*}) \tag{A.14}$$

where $K(\boldsymbol{x^*}, \boldsymbol{x})$ and $K(\boldsymbol{x}, \boldsymbol{x^*})$ are the covariance kernel evaluated between the input coor-

dinates and the data coordinates. If the input coordinates consist of N^* points in the first dimension and M^* points in the second then these matrices have dimensions $(M^*N^* \times NM)$ and $(NM \times N^*M^*)$ respectively.

For the 2D Kronecker-structured covariance matrix $K = T \otimes R$, we can rewrite equation A.13 as

$$\begin{aligned} \boldsymbol{\mu}^* &= \boldsymbol{\mu}_{\boldsymbol{\theta}}(\boldsymbol{x}^*) + [T(\boldsymbol{x}^*, \boldsymbol{x}) \otimes R(\boldsymbol{x}^*, \boldsymbol{x})]\, K(\boldsymbol{x}, \boldsymbol{x})^{-1}[\boldsymbol{y} - \boldsymbol{\mu}_{\boldsymbol{\theta}}(\boldsymbol{x})] \quad \text{(A.15)} \\ &= \boldsymbol{\mu}_{\boldsymbol{\theta}}(\boldsymbol{x}^*) + [T(\boldsymbol{x}^*, \boldsymbol{x}) \otimes R(\boldsymbol{x}^*, \boldsymbol{x})]\, \boldsymbol{z} \quad \text{(A.16)} \end{aligned}$$

where $\boldsymbol{z} = K(\boldsymbol{x}, \boldsymbol{x})^{-1}\left[\boldsymbol{y} - \boldsymbol{\mu}_{\boldsymbol{\theta}}(x)\right]$ Writing the second term of equation A.15 in terms of the vectorization operator we have

$$[T(\boldsymbol{x}^*, \boldsymbol{x}) \otimes R(\boldsymbol{x}^*, \boldsymbol{x})]\, \boldsymbol{z} = [T(\boldsymbol{x}^*, \boldsymbol{x}) \otimes R(\boldsymbol{x}^*, \boldsymbol{x})]\, \mathrm{vec}(Z) \quad \text{(A.17)}$$

where $Z = Y - \mu_\theta(X)$ with X and Y matrices of size $N \times M$ defined by $\boldsymbol{x} = \mathrm{vec}(X)$ and $\boldsymbol{y} = \mathrm{vec}(Y)$ respectively. For matrices A, B, and C of sizes $(n \times m)$, $(m \times p)$, and $(p \times q)$ respectively there is an identity that states the following:

$$\mathrm{vec}(ABC) = (A \otimes C^{\mathrm{T}})\mathrm{vec}(B). \quad \text{(A.18)}$$

Applying this to equation A.17 gives

$$[T(\boldsymbol{x}^*, \boldsymbol{x}) \otimes R(\boldsymbol{x}^*, \boldsymbol{x})]\, \boldsymbol{z} = \mathrm{vec}(TZR). \quad \text{(A.19)}$$

The full expression for the predictive mean is now

$$\boldsymbol{\mu}^* = \boldsymbol{\mu}_{\boldsymbol{\theta}}(\boldsymbol{x}^*) + \mathrm{vec}(TZR). \quad \text{(A.20)}$$

The matrix product TZR can be computed via a modified version of the `celerite`

prediction algorithm presented in (Foreman-Mackey et al. 2017).

First, we compute the product ZR at a computational cost of $\mathcal{O}(NM)$ when R is outer product and $\mathcal{O}(NM^2)$ for arbitrary R. We then compute

$$\mu^*_{p,m} = \sum_{n=1}^{N}\sum_{j=1}^{J} e^{-c_j|t^*_p - t_n|}\left[a_j \cos(d_j|t^*_p - t_n|) + b_j \sin(d_j|t^*_p - t_n|)\right][ZR]_{n,m}. \tag{A.21}$$

in two parts. Here p and m index the elements of the predicted mean matrix. The first part consists of a forward pass through $n_0 = 1, \ldots, N$ where we define:

$$G^-_{n,m,k} = \left[G^-_{n-1,p,k} + [ZR]_{n,m}\tilde{V}'_{n,k}\right] e^{-c_{k//2}(t_{n+1}-t_n)} \tag{A.22}$$

$$H^-_{p,n,k} = e^{-c_{k//2}(t^*_p - t_{n+1})}\tilde{U}'^*_{p,k}, \tag{A.23}$$

[1] and the second consisting of a backward pass through $n_0 = N, ..., 1$ where we define

$$G^+_{n,m,k} = \left[G^+_{n+1,p,k} + [ZR]_{n,m}\tilde{U}'_{n,k}\right] e^{-c_{k//2}(t_n - t_{n-1})} \tag{A.24}$$

$$H^+_{p,n,k} = e^{-c_{k//2}(t^*_{n-1} - t_p)}\tilde{V}'^*_{p,k}, \tag{A.25}$$

where $t_0 = t_1$, $t_{N+1} = t_N$, $G^-_{0,m,k} = 0$, and $G^+_{N+1,m,k} = 0$ for $k = 1, ..., 2J$ and for all m. The expressions for $\tilde{U}'^*_{p,i}$ and $\tilde{V}'^*_{p,i}$ are evaluated at t^*_p. For each value of p, $G^{\pm}$ are evaluated recursively from n to n_0 and then the prediction $\mu^*_{p,m}$ is computed from

$$\mu^*_{p,m} = \sum_{k=1}^{P}\left[G^-_{n_0,m,k}H^-_{p,n_0,k} + G^+_{n_0+1,p,k}H^+_{p,n_0+1,k}\right]. \tag{A.26}$$

This two part computation scales as $\mathcal{O}(nN + n^*N^*)$ where n and n^* are constants. The overall scaling is therefore determined by the cost of the matrix multiplication step.

---

[1] $k//2$ denotes integer division of k by 2. In other words, $k//2 = \text{floor}(k/2)$.

A.4 Sampling from the GP

A sample $\boldsymbol{y}$ can be drawn from a Gaussian process by computing

$$\boldsymbol{y} = \boldsymbol{\mu} + L\boldsymbol{n} \tag{A.27}$$

where $\boldsymbol{\mu}$ is the mean function and $\boldsymbol{n}$ is a vector of draws from a normal distribution

$$n_i \sim \mathcal{N}(0, D_{i,i}^{1/2}) \tag{A.28}$$

for each entry n_i in $\boldsymbol{n}$. The ordering of entries in $\boldsymbol{\mu}$ and consequently $\boldsymbol{y}$ is determined by the structure of K. For the Kronecker structured covariance matrix given in equation 3.16, $\boldsymbol{\mu}$ is the concatenation of the N length-M vectors containing the mean function evaluated at each point in the second dimension at a given point in the first. In other words,

$$\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \dots \boldsymbol{\mu}_N) \tag{A.29}$$

where $\boldsymbol{\mu}_i = (\mu_{i,1}, \mu_{i,2}, \dots \mu_{i,M})$ is the mean function evaluated at the i^{th} point in the first dimension.

Thus $\boldsymbol{\mu}$ is a one-dimensional vector of length $N' = NM$ where N is the size of the first dimension and M the size of the second. The sample vector $\boldsymbol{y}$ then has the same structure. Most users will wish to either unpack the sample into M separate vectors obtained by taking every M^{th} entry in $\boldsymbol{y}$ or reshape it into an $N \times M$ array before displaying or examining the sample.

A.5 Proof of the Positive-definiteness of the 2D Kernel

To begin, we give the following equivalent definitions of positive-definite and positive-semidefinite matrices:

1. A square matrix $A \in \mathbb{R}^{n \times n}$ is **positive-definite** if and only if all of its eigenvalues are

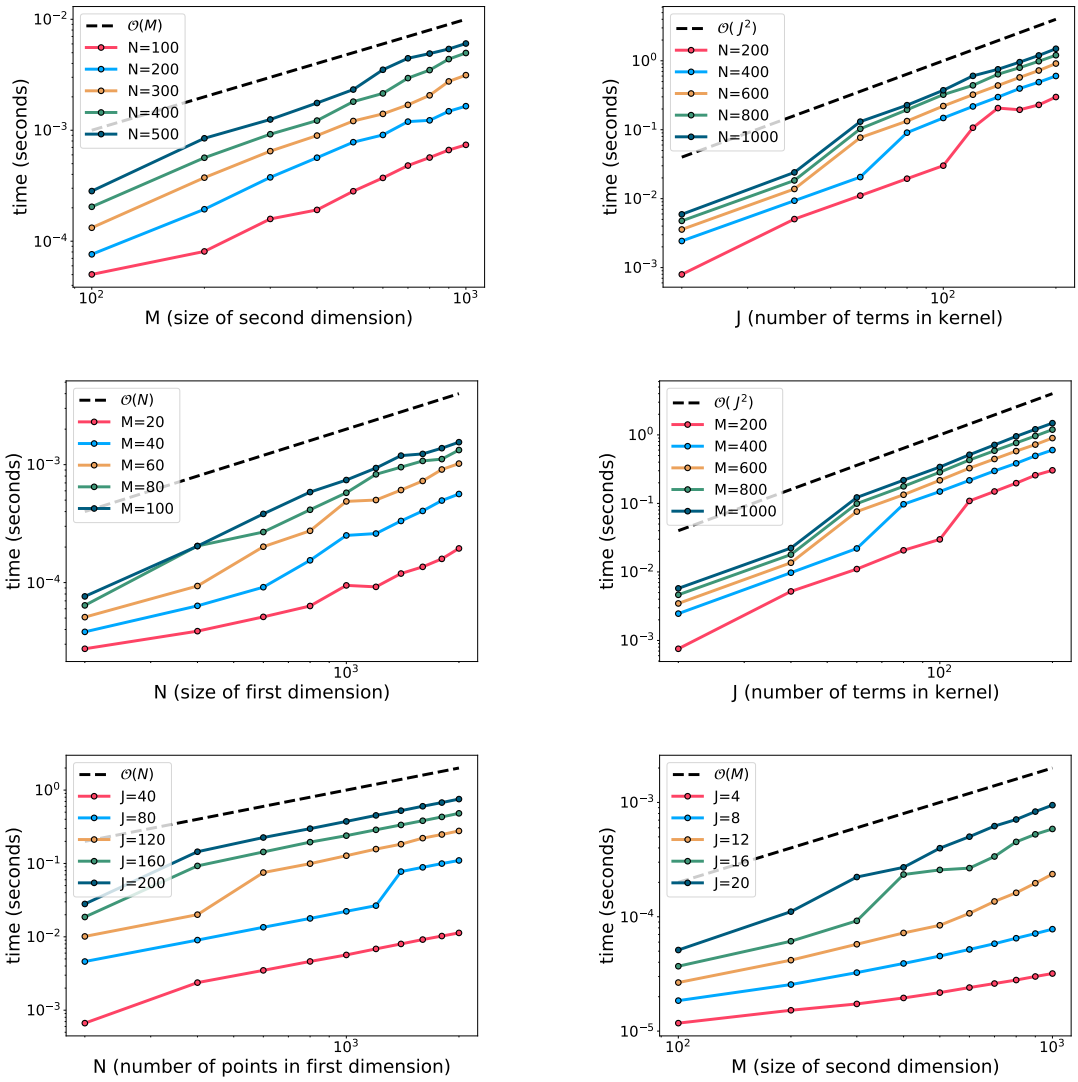

Figure A.1 Benchmarks for the two-dimensional celerite implementation with outer-product covariance in the second dimension. We recover the anticipated linear scaling with respect to both N and M, and the quadratic scaling with respect to J.

positive.

2. Equivalently, a square matrix $A \in \mathbb{R}^{n\times n}$ is **positive-definite** if and only if the scalar $\boldsymbol{x}^T A\boldsymbol{x}$ is positive for all real-valued vectors $\boldsymbol{x} \in \mathbb{R}^{\kappa}$.

3. A square matrix $A \in \mathbb{R}^{n\times n}$ is **positive-semidefinite** if and only if all of its eigenvalues are nonnegative (they may be zero).

4. Equivalently, a square matrix $A \in \mathbb{R}^{n\times n}$ is **positive-semidefinite** if and only if the scalar $\boldsymbol{x}^T A\boldsymbol{x}$ is nonnegative for all real-valued vectors $\boldsymbol{x} \in \mathbb{R}^{\kappa}$.

For a square matrix $A \in \mathbb{R}^{n\times n}$ with eigenvalues given by λ_i for $i = 1, \ldots, n$ and $B \in \mathbb{R}^{m\times m}$ with eigenvalues μ_j for $j = 1, \ldots, m$, the eigenvalues of the Kronecker product $A \otimes B$ are given by $\lambda_i \mu_j$ for all values of i and j. To see this, note that for eigenvectors u and v corresponding to eigenvalues λ and μ of A and B respectively, $A\boldsymbol{u} \otimes B\boldsymbol{v} = (A \otimes B)(\boldsymbol{u} \otimes \boldsymbol{v}) = \lambda\mu(\boldsymbol{u} \otimes \boldsymbol{v})$.

We now consider a positive-definite covariance matrix $T \in \mathbb{R}^{N\times N}$ and a positive-semidefinite square matrix $R \in \mathbb{R}^{M\times M}$. We consider the Kronecker product $T \otimes R$: since the eigenvalues of T are positive and the eigenvalues of R are nonnegative, the products of their eigenvalues that make up the eigenvalues of $T \otimes R$ will also be nonnegative. Therefore $T \otimes R$ is a positive-semidefinite matrix.

Similarly, if R is positive-definite (rather than positive-semidefinite), the eigenvalues of $T \otimes R$ will be uniformly positive and $T \otimes R$ will be a positive-definite matrix.

We now consider the effect of adding a real, positive-valued diagonal matrix $\Sigma \in \mathbb{R}^{NM\times NM}$ to the Kronecker product $T \otimes R$. Using

First consider the case that R is positive-definite. In this case $T \otimes R$ is positive-definite. Using the definition of positive-definiteness that states that a matrix A is positive definite if and only if $\boldsymbol{x}^T A\boldsymbol{x}$ is a positive scalar for all $\boldsymbol{x} \in \mathbb{R}^n$, we compute

$$\boldsymbol{x}^T(T \otimes R + \Sigma)\boldsymbol{x} = \boldsymbol{x}^T(T \otimes R)\boldsymbol{x} + \boldsymbol{x}^T\Sigma\boldsymbol{x} \tag{A.30}$$

Assuming that R is positive-semidefinite we have already established that $T \otimes R$ is positive-semidefinite as well. The left term above is therefore nonnegative. Since the matrix Σ is a diagonal matrix with positive entries, its eigenvalues, which are given by the diagonal entries, are positive, and therefore Σ is positive-definite. Consequently $\boldsymbol{x}^T \Sigma \boldsymbol{x}$ is a positive scalar and the matrix $T \otimes R + \Sigma$ is proven to be positive-definite. This completes the proof that the covariance matrix $K = T \otimes R + \Sigma$ is positive-definite in the case that T is a positive-definite kernel function, R is positive-definite, and Σ is a diagonal matrix with positive entries.

If we assume that R is positive-definite rather than positive-semidefinite then, as shown previously, $T \otimes R$ is itself a positive-definite matrix. In this case the covariance matrix $K = T \otimes R$ is positive-definite even without the addition of Σ. By the same logic as above, the addition of Σ will preserve the positive-definiteness of the covariance and $K = T \otimes R + \Sigma$ will be positive-definite as well.

A.6 Notation

Notation and symbols, in order of appearance:

K: covariance matrix

k: kernel function corresponding to K

x: general independent variable for the GP

$\mathcal{L}$: GP Likelihood

$\boldsymbol{\mu}$: GP mean vector

$\boldsymbol{y}$: vector of observations

N': number of observations corresponding to the length of vector $\boldsymbol{y}$

ω_0: characteristic frequency of simple harmonic oscillator (SHO) term

ϵ: stochastic force term, driving force of SHO

t: an independent variable used to represent time

Q: quality factor of SHO

ω: an independent variable used to represent frequency in expressions for the power spectral density of a process

S_0: Amplitude of the SHO

τ: an independent variable used to represent time lag, as in $\tau_{|i-j|} = |t_i - t_j|$

n, m, i, j, p, q: integers used to index independent variables and matrices

x_h, x_c: covering fractions of a hot and cold component of the stellar photosphere

R_*: stellar radius

d: distance from star to observer

F: flux

B_n: n^{th} spectral band

λ: independent variable representing wavelength

$S_c(\lambda), S_h(\lambda)$: spectra of the hot and cold components of a stellar photosphere

$\mathcal{R}_{B_i}(\lambda)$: Response curve for band B_i

α_i: variability amplitude integrated over band B_i

σ_c, σ_h: $\mathrm{var}(x_c)^{1/2}$, $\mathrm{var}(x_h)^{1/2}$ respectively; the RMS of the cold and hot covering fraction.

Σ_i: diagonal matrix containing the white noise variances for each wavelength at the ith time index.

T: covariance matrix representing the first dimension or time dimension of the 2D GP. T will always be described by a `celerite` kernel function.

R: covariance matrix representing the second dimension or wavelength dimension of the 2D GP. R may be an arbitrary covariance matrix or an outer-product.

N: Length of the first dimension, equal to the number of times in our example application of multiband time series

M: Length of second dimension, equal to the number of bands in our example application of multiband time series

J: number of `celerite` terms in kernel function

P: rank of `celerite` generator matrices

$\boldsymbol{\alpha}$: Vector of correlated noise amplitudes in the second dimension.

Σ: diagonal matrix containing the white noise variances for each observation; the white noise component of the GP covariance matrix

σ_i^2: white noise variance for i^{th} datapoint

R_p: planetary radius

t_0: time of center of transit

$\delta_{\rm in}$: duration of transit ingress/egress

δ: Transit duration (mid-ingress to mid-egress).

$\boldsymbol{\theta}$: Vector of transit parameters.

$\mu_{\rm trap}$: Transit mean model.

f_0: Characteristic frequency of the correlated noise model.

$\bar{\alpha}$: Weighted mean of $\boldsymbol{\alpha}$ used to represent the total amplitude of the correlated variability component of the GP summed over all bands ("monochromatic").

σ: Mean of $\boldsymbol{\sigma}$, the vector of white noise terms; used to represent the total amplitude of the uncorrelated variability component of the GP

$\mathcal{I}$: Information matrix.

N_θ: Number of mean-model parameters, equal to the length of $\boldsymbol{\theta}$

$\sigma_{R_p^2}$: Uncertainty on the transit depth (with "poly" and "mono" to indicate the polychromatic and monochromatic values).

$\boldsymbol{\beta}$: Vector of coefficients used in defining the `celerite` kernel (Foreman-Mackey et al. (2017) use $\boldsymbol{\alpha}$).

a, b, c, d: `celerite` coefficients

A: diagonal component of full kernel function; $K = A + \mathrm{tril}(UV^{\rm T}) + \mathrm{triu}(VU^{\rm T})$

U, V: `celerite` generator matrices

L: Lower triangaular matrix used in LDLT Cholesky decomposition.

tril, triu: lower and upper triangular matrix operators

D: Diagonal matrix used in decomposition.

W: Matrix used in semi-separable LDLT Cholesky decomposition.

I: Identity matrix.

S: intermediary matrix used in the `celerite` decomposition algorithm

D: diagonal matrix in the Cholesky decomposition of K

A_0: diagonal component of K with white noise amplitude set to zero; $A_0 = A - \Sigma$

U', V', A': Kronecker products of U, V, and A taken with $\vec{\alpha}$ or R and I_M

$\tilde{U}, \tilde{V}, \tilde{W}$: refactored `celerite` matrices corresponding to U, V, and W

$\tilde{U}', \tilde{V}', \tilde{W}'$: refactored `celerite` matrices corresponding to U', V', and A'.

ϕ, ϕ': matrices used in the refactored version of `celerite`

$F^{\pm}, G^{\pm}$: intermediary matrices for prediction algorithm

$f_{n,j}, g_{n,j}$: intermediary vectors used to compute the likelihood of the GP model

$\boldsymbol{\mu}^*$: predictive mean model

K^*: predictive covariance

$\boldsymbol{x}^*$ independent variable used to represent the points at which the predictive mean and covariance of the GP are evaluated

$\boldsymbol{z}$: the product between the observed vector $\boldsymbol{y}$ and the inverse of the covariance matrix K^{-1} used to compute the GP likelihood

$\boldsymbol{z}'$: intermediary vector used to compute $\boldsymbol{z}$

X, Y, Z: $\boldsymbol{x} = \mathrm{vec}(X)$, $\boldsymbol{y} = \mathrm{vec}(Y)$, and $\boldsymbol{z} = \mathrm{vec}(Z)$ respectively; the matrix versions of $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{z}$ for the two-dimensional GP.

t^*: independent variable used to represent the points at which the predictive mean and covariance of the GP are evaluated; same as x^* when the independent variable is time.

$H^{\pm}$: intermediary matrix used to compute the GP prediction ($Q^{\pm}$ in Foreman-Mackey et al. (2017))

N^*: the number of points at which the prediction is evaluated in the first dimensions

n^*: constant on which the computational scaling of the prediction algorithm depends

$\boldsymbol{n}$: vector of random draws from a standard normal distribution used to draw a sample from the GP

Appendix B

CHAPTER 4

B.1 Derivation of the Optimal Bias Vector and Likelihood Ratio for the Constant Bias GP Model

We begin with the likelihood for the constant bias model. For simplicity we assume zero mean, but the same derivation can be performed with a mean function by substituting $\boldsymbol{y}$ for $\boldsymbol{y}$ at every step.

$$\ln\ \mathcal{L} = -\frac{1}{2}(\boldsymbol{y}-\boldsymbol{b})^{\mathrm{T}}K^{-1}(\boldsymbol{y}-\boldsymbol{b}) - \frac{1}{2}\ln\ \det(K) - \frac{N}{2}\ln(2\pi) \tag{B.1}$$

We replace the vector $\boldsymbol{b}$, which contains a bias term for each datapoint, with a single-point bias vector $\boldsymbol{b}_i$ for which only the i^{th} entry has a non-zero value:

$$\boldsymbol{b}_i = (0, 0, \ldots, b_i, \ldots, 0, 0)^{\mathrm{T}} \tag{B.2}$$

Rewriting the likelihood with this substitution and expanding the first term we find:

$$\begin{aligned}\ln\ \mathcal{L} = \ & -\ \frac{1}{2}\boldsymbol{y}^{\mathrm{T}}K^{-1}\boldsymbol{y} - \frac{1}{2}\boldsymbol{b}_i^{\mathrm{T}}K^{-1}\boldsymbol{b}_i + \frac{1}{2}\boldsymbol{y}^{\mathrm{T}}K^{-1}\boldsymbol{b}_i + \frac{1}{2}\boldsymbol{b}_i^{\mathrm{T}}K^{-1}\boldsymbol{y} \\ & -\ \frac{1}{2}\ln\ \det(K) - \frac{N}{2}\ln(2\pi)\end{aligned} \tag{B.3}$$

Using the fact that K^{-1} is symmetric, we have

$$\frac{1}{2}\boldsymbol{b}_i^{\mathrm{T}}K^{-1}\boldsymbol{y} = \frac{1}{2}\boldsymbol{y}^{\mathrm{T}}K^{-1}\boldsymbol{b}_i \tag{B.4}$$

Which allows us to simplify equation B.3:

$$\ln\ \mathcal{L} = -\frac{1}{2}\boldsymbol{y}^{\mathrm{T}}K^{-1}\boldsymbol{y} - \frac{1}{2}\boldsymbol{b}_i^{\mathrm{T}}K^{-1}\boldsymbol{b}_i + \boldsymbol{y}^{\mathrm{T}}K^{-1}\boldsymbol{b}_i - \frac{1}{2}\ln\ \det(K) - \frac{N}{2}\ln(2\pi) \tag{B.5}$$

Exploiting the fact that $\boldsymbol{b}_i$ only has a single entry, we rewrite the second and third terms as scalar quantities:

$$\frac{1}{2}\boldsymbol{b}_i^{\mathrm{T}}K^{-1}\boldsymbol{b}_i = \frac{1}{2}b_i^2(K^{-1})_{ii} \tag{B.6}$$

and

$$\boldsymbol{y}^{\mathrm{T}} K^{-1} \boldsymbol{b}_i = b_i (K^{-1}\boldsymbol{y})_i \tag{B.7}$$

Which gives us the following for the likelihood:

$$\ln\ \mathcal{L} = -\frac{1}{2}\boldsymbol{y}^{\mathrm{T}} K^{-1}\boldsymbol{y} - \frac{1}{2} b_i^2 (K^{-1})_{ii} + b_i (K^{-1}\boldsymbol{y})_i - \frac{1}{2}\ln\ \det(K) - \frac{N}{2}\ln(2\pi) \tag{B.8}$$

At this point we can differentiate with respect to b_i to find the critical points of the likelihood:

$$\frac{\partial \ln \mathcal{L}}{\partial b_i} = -b_i \left(K^{-1}\right)_{ii} + (K^{-1}\boldsymbol{y})_i \tag{B.9}$$

Setting this equal to zero and solving for b_i gives us the value of b_i at the critical point:

$$b_i = \frac{(K^{-1}\boldsymbol{y})}{(K^{-1})_{ii}}. \tag{B.10}$$

We can check that this critical point is a maximum by computing the second derivative of equation B.9 with respect to b_i:

$$\frac{\partial^2 \ln \mathcal{L}}{\partial b_i^2} = -(K^{-1})_{ii}. \tag{B.11}$$

Since the entries of K^{-1} are positive (because K is a positive-semidefinite matrix), the second derivative is always negative and the point b_i represents a maximum of the log-likelihood.

We can find the maximized log-likelihood by plugging the solution for b_i back into the B.3, from which we obtain:

$$\ln \mathcal{L}_{\max} = \frac{1}{2}(K^{-1}\boldsymbol{y})_i b_i - \frac{1}{2}\boldsymbol{y}^T K^{-1}\boldsymbol{y} - \frac{1}{2}\det(K) - \frac{N}{2}\ln(2\pi). \tag{B.12}$$

When we consider the difference in likelihood between the bias model and an unbiased GP

model the terms involving only $\boldsymbol{y}$ and K cancel out and we obtain

$$\Delta \ln \mathcal{L} = \ln \mathcal{L}_{\text{bias}} - \ln \mathcal{L}_{\text{no bias}} = \frac{1}{2}(K^{-1}\boldsymbol{y})_i b_i \tag{B.13}$$

B.2 Derivation of the Optimal Bias Vector and Likelihood Ratio for the Random Bias GP Model

The likelihood for the random bias model is

$$\ln \mathcal{L} = -\frac{1}{2}\boldsymbol{y}^{\mathrm{T}}(K+\Sigma)^{-1}\boldsymbol{y} - \frac{1}{2}\ln \det(K+\Sigma) - \frac{N}{2}\ln(2\pi) \tag{B.14}$$

where Σ is a diagonal matrix with the vector of random bias variance terms along its diagonal. As in the derivation for the constant bias model, we consider only a single bias term at a time. Setting all the other bias terms to zero, we can rewrite the single-term Σ matrix as

$$\Sigma_i = \boldsymbol{o_i}\boldsymbol{o_i}^{\mathrm{T}} \tag{B.15}$$

where

$$\boldsymbol{o_i} = (0, 0, \ldots, \sigma_i, \ldots, 0, 0)^{\mathrm{T}} \tag{B.16}$$

We can use the Sherman-Morrison formula and the Matrix Determinant Lemma to compute the factors $(K + \boldsymbol{o_i}\boldsymbol{o_i}^{\mathrm{T}})^{-1}$ and $\det(K + \boldsymbol{o_i}\boldsymbol{o_i}^{\mathrm{T}})$:

$$(K + \boldsymbol{o_i}\boldsymbol{o_i}^{\mathrm{T}})^{-1} = K^{-1} - \frac{K^{-1}\boldsymbol{o_i}\boldsymbol{o_i}^{\mathrm{T}}K^{-1}}{1 + \boldsymbol{o_i}^{\mathrm{T}}K^{-1}\boldsymbol{o_i}} \tag{B.17}$$

and

$$\det(K + \boldsymbol{o_i}\boldsymbol{o_i}^{\mathrm{T}}) = (1 + \boldsymbol{o_i}^{\mathrm{T}}K^{-1}\boldsymbol{o_i})\det(K) \tag{B.18}$$

The likelihood can now be written

$$\ln \mathcal{L} = -\frac{1}{2}\boldsymbol{y}^{\mathrm{T}} K^{-1}\boldsymbol{y} + \frac{1}{2}\boldsymbol{y}^{\mathrm{T}}\frac{K^{-1}\boldsymbol{o_i}\boldsymbol{o_i}^{\mathrm{T}}K^{-1}}{1+\boldsymbol{o_i}^{\mathrm{T}}K^{-1}\boldsymbol{o_i}}\boldsymbol{y} - \frac{1}{2}\ln\left[(1+\boldsymbol{o_i}^{\mathrm{T}}K^{-1}\boldsymbol{o_i})\det(K)\right] - \frac{N}{2}\ln(2\pi). \tag{B.19}$$

When we rewrite the vector matrix-vector products in terms of scalar quantities by exploiting the structure of $\boldsymbol{o_i}$ this becomes

$$\ln \mathcal{L} = -\frac{1}{2}\boldsymbol{y}^{\mathrm{T}} K^{-1}\boldsymbol{y} + \frac{1}{2}\frac{\nu_i (K^{-1}\boldsymbol{y})_i^2}{1+\nu_i (K^{-1})_{ii}} - \frac{1}{2}\ln\ \det(K) - \frac{1}{2}\ln(1+\nu_i (K^{-1})_{ii}) - \frac{N}{2}\ln(2\pi), \tag{B.20}$$

where we have made the substitution $\nu_i = \sigma_i^2$, so that ν represents the bias in the variance of the i^{th} datapoint. At this point we can differentiate with respect to the variance in order to find the critical points of the likelihood function:

$$\frac{\partial \ln \mathcal{L}}{\partial \sigma_i} = \frac{1}{2}\frac{(K^{-1}\boldsymbol{y})_i}{(1+\nu_i (K^{-1})_{ii})^2} - \frac{1}{2}\frac{(K^{-1})_{ii}}{1+\nu_i (K^{-1})_{ii}} \tag{B.21}$$

Setting this equal to zero we obtain the value of ν_i that maximizes equation B.20

$$\nu_i = \frac{(K^{-1}\boldsymbol{y})_i - (K^{-1})_{ii}}{(K^{-1})_{ii}}. \tag{B.22}$$

Plugging this back into equation B.20 to obtain the maximum log-likelihood we find:

$$\ln \mathcal{L}_{\max} = -\frac{1}{2}\boldsymbol{y}^{\mathrm{T}} K^{-1}\boldsymbol{y} + \frac{1}{2}\left(\frac{(K^{-1}\boldsymbol{y})_i^2}{(K^{-1})_{ii}} - 1\right) - \frac{1}{2}\ln\det(K) - \frac{1}{2}\ln\left(\frac{(K^{-1}\boldsymbol{y})_i^2}{(K^{-1})_{ii}}\right) - \frac{N}{2}\ln(2\pi). \tag{B.23}$$

The difference between the likelihood of the random bias model and an unbiased GP is given by

$$\Delta \ln \mathcal{L} = \ln \mathcal{L}_{\text{bias}} - \ln \mathcal{L}_{\text{no bias}} = \frac{1}{2}\left(\frac{(K^{-1}\boldsymbol{y})_i^2}{(K^{-1})_{ii}} - 1\right) - \frac{1}{2}\ln\left(\frac{(K^{-1}\boldsymbol{y})_i^2}{(K^{-1})_{ii}}\right). \tag{B.24}$$

Appendix C

CHAPTER 5

C.1 Derivative of the flux with respect to ϕ in the limit of small center-of-star-moon separations

To see how the two factors of b_m cancel out to resolve the apparent singularity in the limit $b_m \to 0$, let us begin by identifying the factors of b_m that appear in the series expansion of the numerator. The first of these comes from the factor of $\sin\theta$ in the numerator of Equations (5.74) and (5.76) for which the first term of the Taylor series expansion is b_m/b_p. For Equation (5.75) we first note that in the limit $b_m \to 0$ we have $b_{pm} \to b_p$, which allows us to factor out b_p from the term $b_p \cos\theta - b_{pm}$ leaving a factor of $\cos\theta - 1$ in the numerator. The first term of the series expansion for $\cos\theta - 1$ is $b_m^2/(2b_p^2)$, which fully cancels out the factors of b_m in the denominator.

While this takes care of the singularity in Equation (5.75), we still need to find another factor of b_m in the numerator of Equations (5.74) and (5.76). To see where this second factor comes from we look to the Taylor series expansion of the ϕ-derivatives of each of the primitive integrals (Equations (5.19), (5.24), (5.40), and (5.48)). Because each integral is evaluated from 0 to ϕ taking the derivative with respect to ϕ is simple – by the fundamental theorem of calculus we recover the integrand itself evaluated at $\phi' = \phi$. The first term in the Taylor series expansion for small b can be found by setting $b = 0$. We immediately note that for each of these integrands setting $b = 0$ eliminates all factors of ϕ and trigonometric functions thereof from the expressions. In other words we can define

$$H_{m,n}(r) = \lim_{b_m \to 0} \frac{\partial G_n(\phi, r_m, b_m)}{\partial \phi} \tag{C.1}$$

to be the constant term of the Taylor series where $H_{m,n}$ is a function only of r. The integral along an arc of the moon intersected by the planet is computed from two evaluations of the primitive integrals as

$$\sum_{n=0}^{N} u_n \left[G_n(\phi_m^-, r_m, b_m) - G_n(\phi_m^+, r_m, b_m) \right]. \tag{C.2}$$

This term appears in cases E, F, H, and K of Table 5.6.3, which represent all the cases that can arise when $b_m = 0$. Using the definition from Equation (C.1) we can write the first term of the Taylor expansion for the derivative of this expression with respect to ϕ as

$$\sum_{n=0}^{N} u_n \left[H_{m,n}(r_m) - H_{m,n}(r_m) \right] = 0. \tag{C.3}$$

We now see that the constant terms cancel out when the integral is evaluated along any arc of the moon in the limit that $b_m \to 0$. All remaining terms of the Taylor expansion have at least one factor of b_m in their numerators which is sufficient to cancel the factor of b_m in the denominator of Equations (5.74) through (5.76) when the chain rule is applied.

We can apply this same reasoning to the arc of the planet. In this case the only variable in the primitive integrals which depends on b_m is the integration limit, ϕ, since the b in the integrands refers to b_p. In this case taking the limit as $b_m \to 0$ is equivalent to taking the simultaneous limit as $\theta \to 0$ and $b_{pm} \to b_p$. As we did for arcs along the moon we define

$$H_{p,n}(\phi, r_p, b_p) = \lim_{\theta \to 0, b_{pm} \to b_p} \frac{\partial G_n(\phi, r_p, b_p)}{\partial \phi} \tag{C.4}$$

Note that this expression is a function of ϕ, unlike the corresponding limit for arcs along the moon in Equation (C.1). When $\theta \to 0$ the integration limits along the arc become $\phi = \pm\phi_p$ (see Equation (5.57)). We again consider cases E, F, H, and K as the relevant cases for $b_m = 0$. Each of these cases includes a term

$$\sum_{n=0}^{N} u_n \left[G_n(\phi_p^-, r_p, b_p) + G_n(\phi_p^+, r_p, b_p) \right]. \tag{C.5}$$

In some cases the expressions in Table 5.6.3 must be rearranged to reveal this term, but it does arise in each case. Taking the limit of the ϕ-derivative of this term and using the

definition of $H_{p,n}$ from (C.4) we have

$$\begin{aligned} & \sum_{n=0}^{N} u_n \left[H_{n,p}(-\phi_p, r_p, b_p) + H_{n,p}(\phi_p, r_p, b_p) \right] \\ = & \sum_{n=0}^{N} u_n \left[-H_{n,p}(\phi_p, r_p, b_p) + H_{n,p}(\phi_p, r_p, b_p) \right] \\ = & \ 0 \end{aligned} \tag{C.6}$$

where we have used the fact that the derivative of an even function (such as G_n which is even with respect to ϕ) is an odd function to make the substitution $H_{n,p}(-\phi_p, r_p, b_p) = -H_{n,p}(\phi_p, r_p, b_p)$. Again we see that the constant terms of the expansion are eliminated when the integral is taken along a complete arc, and the remaining terms contain at least one factor of b_m to cancel the factor of b_m in the denominator of Equations (5.74) through (5.76).

None of the remaining terms in cases E, F, H, or K depend on θ_m, and therefore are not subject to the singularity in the derivatives of θ_m. Therefore we do not need to account for factors of b_m in any of these terms.

Unfortunately some numerical error is introduced into the calculation by these subtractions and cancellations. In our testing we have found that this error is limited to the region in which b_m is less than 10^{-6} stellar radii and is generally smaller than 10^{-5}. It occurs only in the derivatives of parameters that control the position of the moon.

C.2 Correction to Equations 37 in Kipping (2011) and 16 in Fewell (2006)

This is a correction to the formula for the area of overlap of three circles when the chord connecting the intersections of the smallest circle with each of the larger circles spans more than half of the smallest circle. This configuration is shown in Figure C.1 where the three circles are labeled 1 through 3 with circles 1 and 2 having the same radius and circle 3 having a smaller radius. The intersections demarcating the area of overlap are labeled I_{12}, I_{13} and I_{23} with the subscript indicating the pair of circles that are intersecting. In Figure C.1 the region to the left of the vertical dashed line contains more than half of circle 3 (i.e. Equation

(34) in Kipping 2011 is not true).

Fewell (2006) (and therefore Kipping 2011 as well) gives the area of the full region as

$$A_{\text{fewell}} = \frac{\Delta}{4} + \sum_{i=1}^{3} r_i^2 \arcsin\frac{c_i}{2r_i} - \frac{c_1}{4}\sqrt{4r_1^2 - c_1^2} - \frac{c_2}{4}\sqrt{4r_2^2 - c_2^2} + \frac{c_3}{4}\sqrt{4r_3^2 - c_3^2} \tag{C.7}$$

where Δ is given by

$$\Delta = \sqrt{(c_1 + c_2 + c_3)(c_1 + c_2 - c_3)(c_1 - c_2 + c_3)(c_2 + c_3 - c_1)} \tag{C.8}$$

and $\Delta/4$ is Heron's formula for the area of a triangle with side lengths c_1, c_2, and c_3.

The above formula is incorrect. The correct expression can be arrived at by breaking down the region of overlap into three sub-regions and considering the area of each separately. Figure C.2 highlights each of these regions. The first region, labeled A in Figure C.2 is formed by the two minor segments of circles 1 and 2 formed by the chords connecting I_{12} to I_{23} and I_{13} to I_{23}. The general formula for the area of a minor segment is $A_{\text{segment}} = r^2(\theta - \sin\theta)$ where θ is the angular length of the chord, which in this case yields the formula

$$A = \sum_{i=1}^{2} \left(r_i^2 \arcsin\frac{c_i}{2r_i} - \frac{c_i}{4}\sqrt{4r_i^2 - c_i^2} \right), \tag{C.9}$$

where r_i is the radius of circle i and c_i is the length of the chord segmenting circle i. This formula agrees with Fewell (2006), as does the formula for the area of the next region, region B in Figure C.2 which is the triangle formed by connecting the three points of intersection between the three circles. The area of this triangle is given by Heron's formula:

$$B = \frac{\Delta}{4} \tag{C.10}$$

where Δ is defined as in Equation (C.7)

The disagreement between our formula and Fewell (2006) arises for the third sub-region, labeled region C in C.2, which is the remainder of the area of circle 3 after the minor segment

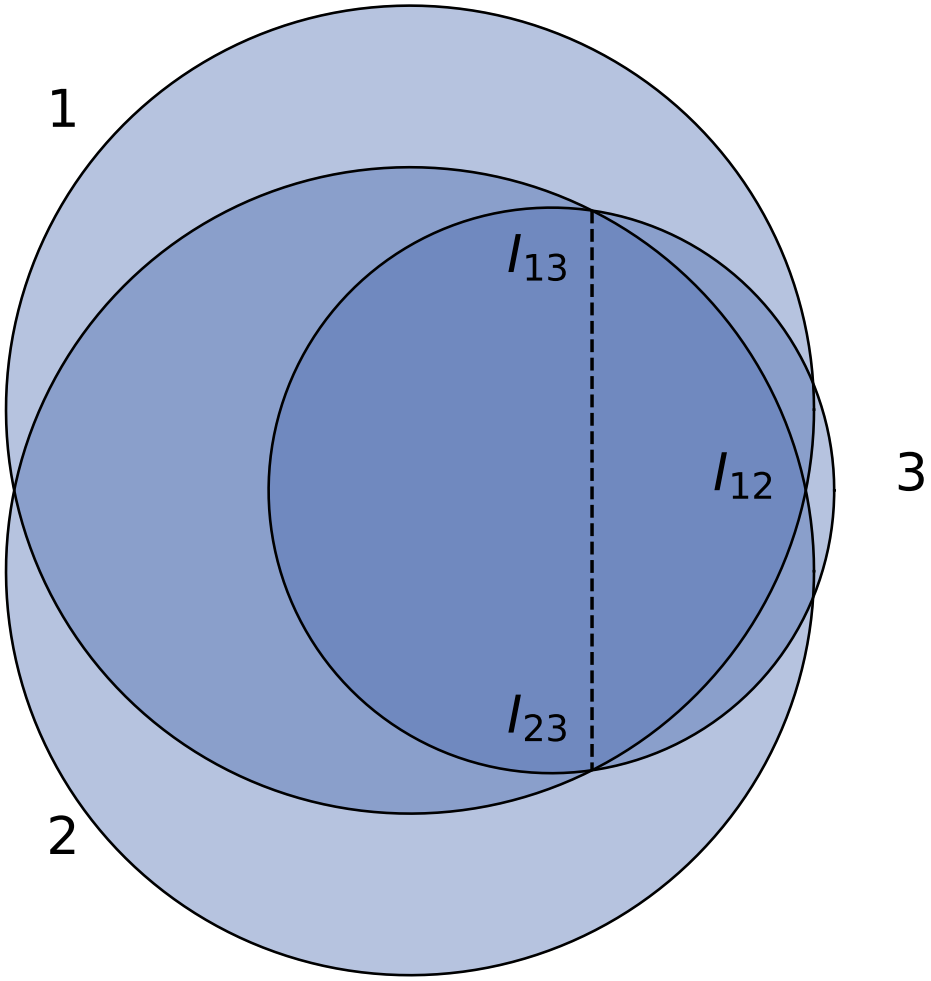

Figure C.1 Three overlapping circles meeting the condition that more than half of the area of circle 3 is to the left of the chord connecting I_{12} and I_{13}.

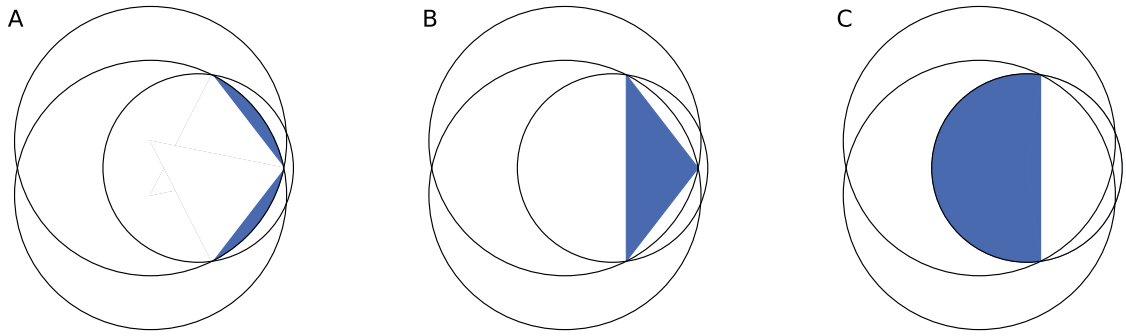

Figure C.2 Subdivision of the region of overlap between three circles. The total area of overlap is found by adding the areas of each of these sub-regions. The expressions for each area are given in the text.

formed by the chord connecting the intersections with circles 1 and 2 is subtracted. This can be expressed as

$$\begin{aligned} C &= \pi r_3^2 - (r_3^2 \arcsin\frac{c_3}{2r_3} - \frac{c_3}{4}\sqrt{4r_3^2 - c_3^2}) \\ &= r_3^2\left(\pi - \arcsin\frac{c_3}{2r_3}\right) + \frac{c_3}{4}\sqrt{4r_3^2 - c_3^2}. \end{aligned} \tag{C.11}$$

This differs from Fewell (2006) and Kipping (2011) in the inclusion of the πr_3^2 term as well as in the sign of the arcsin term. The corrected total area is then

$$\begin{aligned} A_{\text{total}} &= \frac{\Delta}{4} + \sum_{i=1}^{2}\left(r_i^2 \arcsin\frac{c_i}{2r_i} - \frac{c_i}{4}\sqrt{4r_i^2 - c_i^2}\right) \\ &+ r_3^2\left(\pi - \arcsin\frac{c_3}{2r_3}\right) + \frac{c_3}{4}\sqrt{4r_3^2 - c_3^2}. \end{aligned}$$