Understanding Starspot Lightcurves With The Kepler Satellite

by
Oliver Dewey

Professor Jorge Moreno, Advisor
Professor Leslie Hebb, Advisor

A thesis submitted in partial fulfillment
of the requirements for the
Degree of Bachelor of Arts with Honors
in Physics

POMONA COLLEGE
Claremont, California
April 19, 2023

Abstract

This thesis explores a new method of starspot analysis which takes Kepler transit lightcurves and analyzes the variation of in-transit flux as compared to out-of-transit flux as a proxy for starspot crossings. I do this through describing a pipeline that consists of 3 separate programs all of my own creation. The first of these programs takes the Kepler data and normalizes it to then be used in a statistical comparison of in-transit to out-of-transit flux variation. The second program simulates a transit with the same parameters as the chosen system but with spots of known sizes, longitudes, and latitudes randomly placed within the path of transit. And finally, a third program is described which either (1) takes the Kepler data and performs a statistical analysis of the flux variations or (2) takes the simulated data and performs an analysis and mapping of those variations and stellar surface features. A final code then compares starspot simulations of different numbers, sizes, and contrasts in efforts of understanding optimal spot parameters which statistically match those seen in the data. Such a result demonstrates new information about the general spottiness of features which have such starspot crossing signatures, information that for many of these objects is completely unknown.

With this analysis, my thesis shows how accurate information (that is to say, the findings match current literature) is gathered for KIC 1061656571. For this object, I find that (1) there are more than 10 spots which need to be simulated at a time, (2) the sunspots are roughly 2.5 times that of the Sun, and (3) their contrasts are roughly 0.6. For KIC 8672910, I find that the simulation matches the data at a starspot number of 16, a size of 2.75, and a contrast of 0.7. For KIC 7767559, I find that that the simulation matches the data at a starspot number of 16, a size of 3, and a contrast of 0.7, numbers which take into account peculiarities in that system's data. And finally, using code to inspect KIC 7447200, I find the possible existence of stellar plages and a long-lived polar spot. Along the way, discussion is held of limitations and caveats of the programs as well as the math and science behind such work.

Executive Summary

Despite aiding in understanding the magnetic dynamics of convective stars, starspots are a mysterious stellar feature. Unlike sunspots, which can be observed visually, there are only a few ways to resolve these surface irregularities on other stars, given the impossibility of visual resolution. Such ways include photometric approaches, Doppler imaging, spectroscopic analysis, and Zeeman-Doppler imaging. This thesis, however, demonstrates a fifth method of observing and mapping such a phenomena.

Similar to photometric means, a lightcurve of the system is analyzed in this fifth method. In this case, we specifically use transit lightcurves from the Kepler data, which takes highly precise and sensitive brightness data for thousands of objects where a secondary object (star or planet) is crossing in front of view. With each lightcurve, we then look at the cumulative variations in and outside of the in-transit part of the lightcurve, whereby larger in-transit variations suggest that the transiting object is passing over some surface irregularity. This is the case because when an object transits in front of its star, the total flux from the star periodically decreases by a certain amount. This decrease in light allows natural surface irregularities that are normally lost in the net-integrated light outside of transit to appear. For the case of starspots, which are cooler and therefore less bright, this manifests as an increase in brightness, given that the transiting object is covering a relatively dimmer portion of the total flux. Furthermore, by canceling out the general shape of the lightcurve using lightcurve modeling with known parameters and then comparing such residual variations for the in-transit and out-of-transit flux points, we can outline the frequency, intensity, and location of starspot crossings.

My thesis comes in by doing exactly this and going a step further by simulating starspots with known parameters on systems with matching characteristics to those being observed. This provides the capability to understand what the crossing features actually look like compared to the Sun and describe in greater detail the systems besides that it bears starspot signatures. For the objects that are selected, the system of three codes that I developed are able to take down and clean Kepler data, model lightcurves to generate residual plots, identify starspot crossings, run statistical tests to gather a quantitative understanding of the system, simulate similar systems with known spot parameters until an optimized system appears, then map these systems for a visual representation of such a system.

As stated in my Abstract, with this analysis, my thesis shows how accurate information (that is to say, the findings match current literature) is gathered for KIC 1061656571. For this object, I find that (1) there are more than 10 spots which need to be simulated at a time,

(2) the sunspots are roughly 2.5 times that of the Sun, and (3) their contrasts are roughly 0.6. A similar analysis is performed on KIC 8672910, KIC 7767559, and KIC 7447200, each with their own interesting features and characteristics. This is just a couple objects describing such a use, but theoretically all the objects in the Kepler archive can undergo such an analysis. From this, future steps are outlined, including updating the specifications and capabilities of the code as well as creating groupings of starspot total area coverage as it varies with size, temperature, or radius of the stars. Such work is a step towards understanding the magnetic workings (stellar dynamos) of other stars, particularly those in star-planet systems, and how they compare with our own Sun.

Acknowledgments

I would like to express my gratitude to all the people who helped me in the construction of my thesis and were there by my side throughout its making. Particularly, I would like to thank my mentor and advisor Professor Leslie Hebb of the William and Hobart Smith Colleges, who showed me how exciting astronomy research was and made this particular work one of the most rewarding experiences of college. I would also like to thank my on campus advisor Professor Jorge Moreno, who always was there to ensure that my work was of utmost quality and happening on a correct timeline (which was much needed)!

I would also like to thank everybody else in my life–professors, mentors, advisors, family, friends, partner, and all those wishing me the best and being there through a process that so often felt like a slug. Their company and love was the essential drive to have been able to get anything done at all.

Without any of these people, and so many more whose work I was building upon, whose emotional and professional help I was relying upon, I would never have been able to see this thesis through. Thank you all!

Contents

Chapter 1

Introduction

1.1 Starspots through the Ages

Starspots are not phenomena that are far and irrelevant to our world here on Earth. In fact, humans over centuries have been studying starspots, though much closer than this thesis pursues: on our Sun. Just like other stars, our Sun has dark spots on the surface that are the local equivalent of starspots. This knowledge has been known to exist in many cultures around the world, many before the features were even noted and scientifically studied by Galileo in 1612 (Thomas and Weiss 2012). In fact, it is the case that ancient astronomers in China, Japan, Greece, India, Egypt, and Arabia all knew of this phenomenon (Thomas and Weiss 2012). However, the first mention of sunspots in Western literature comes from Theophrastus of Athens in 325 BCE, referring to the meteorological impacts of such stellar features as a precursor for rain (Thomas and Weiss 2012). A similar observation was written even earlier in 165 BCE in China, which states that in the "within the Sun there was a black spot" as well as another veiled mention that writes "the Sun was orange in color. Within it there was a black vapour like a flying magpie. After several months it dispersed" (Thomas and Weiss 2012).

Still, it is not just these ancient civilizations that knew of the spots on our Sun. Many narratives of indigenous astronomy also highlight humanity's continuous observation and connection to this phenomenon. Inuit elders in Iglulik, for example, talk in their Sun-moon creation myth of how the Sun-woman, Siqiniq, formed spots on her body as an act of self-harm, whereas her brother, Taqqiq, the Moon-man, formed spots on his face by falling onto a candle (Hamacher 2022). This narrative describes the presence of both dark patches on the Sun (spots) and on the Moon (maria). Aztec tradition has a similar myth that explains the dark patches on our star. In this myth, Tecciztecatl and Nanhuatzin are chosen to take the role of the Sun-god, but to do so by jumping into a fire. While Tecciztecatl is too afraid to jump in, Nanhuatzin jumps without hesitation. Enraged at the cowardice of Tecciztecatl, one god throws a rabbit at his face, dimming him and creating a rabbit-like dark spot—the creation-story of the Moon (Hamacher 2022). Nanhuatzin, on the other hand is rewarded by becoming the Sun, but as the name implies in its language–"full of sores"—his face, deformed

by syphilis, has pockmarks, thus explaining a spotted Sun (Hamacher 2022). Just further South, researchers have noted how Sun-deity pictograms are shown with and without spots on cheeks (Hamacher 2022). The presence of these spots has been correlated to years of high sunspot activity whereas the lack of these spots has been correlated with low sunspot activity (Hamacher 2022). Thus, indigenous knowledge and art similarly demonstrate the constant knowledge of starspots on our local Sun.

If we take into consideration other phenomena correlated with these spots, which correspond to active and unpredictable stellar areas, then, such a feature becomes even more ingrained with our human story. It is because of the chaotic stellar nature that creates starspots that we have auroras (the ionization of the outflow of solar material as it hits the Earth's magnetic field). These in turn have become interpreted as the resting place of the ancestors to many cultures from the Arctic all the way to New Zealand. It is also because of this stellar weather/irregularities that we have aurora-crackling or "hissing"—only recently considered by Western science in 2016–which has been understood to be ancestors communicating to their loved ones, or in the Arctic people playing a traditional ballgame (MacDonald 2022).

This thesis thus carries on the legacy that has been with us all along to understand and integrate stellar phenomena into our own cosmovision. The only difference is that finally, aided by telescopes, we breach the boundaries of our solar system, where we may find new ways to continue the study of our ancestors.

As a quick layout as to what will follow, I first introduce my project by explaining the canonical Western history of starspot discovery. This is intended to describe their core features, as well as to reflect the way in which the process of understanding starspots has happened (that is, by centuries of observation which culminated in a singular model.) I will then explain the model of starspots' creation and propagation, commonly known as the stellar dynamo. From this, I will describe some of the ways in which starspot study can move beyond the solar system and what new knowledge we have only recently learned about doing so. In the same vein, I then introduce my project, which details another way in which we may look out into other stellar surface phenomena using the Kepler satellite. Here, I outline the data as well as relevant knowledge for the work I have done (notably in the factors that create and model transit curves). I then go on to explain the workings of my code, model code, and their combined outcomes, which reveal new mappings of pock-marked features on other Sun-like bodies.

1.2 Sunspots: The Beginning

It is hard to imagine how people were studying sunspots before the telescope was invented and indirect observation of sunspots was conducted. However, it *is* in fact possible to observe this phenomena during times of high particulate matter in the air (dust, ash, snow), which reduces glare and allows for the resolution of the dark solar features (Thomas and Weiss 2012). To get an idea of what this looks like, such an occurrence is seen below on a dusty

day in New Mexico (Figure 1.1). It is by this way of observation that starspots first came into modern consciousness in Europe, as Christophe Cheiner of Bavaria noticed the phenomena through smoke-filled skies and asserted that there were small bodies orbiting closely to our star (Thomas and Weiss 2012).

Figure 1.1: Sunspots seen on a dusty day in New Mexico taken by David Tremblay and gathered from Universe Today.

These observations were sent in letters to Galileo, who birthed today what would be considered as the beginning of scientific study and systematization of starspots. Galileo was the first person to make detailed observations and drawings of this phenomena, avoiding eye-damage of direct telescopic observation by using a projection technique whereby light was projected through the telescope in a dark room onto a piece of paper. In doing so, he was able to record sunspots (Figure 1.2), noting their location "in a narrow zone of the solar globe corresponding to the space in the celestial sphere that lies within the tropics" (Thomas and Weiss 2012). The scientist was also the first to confirm that due to a limb foreshortening-effect, these spots were likely not orbiting bodies but were rather "imperfections" on what was deemed a "perfect body" (Thomas and Weiss 2012). Using the apparent motion of sunspots over the course of the year, in fact, was one of the many arguments laid out by Galileo in his argument for a heliocentric orbit (Thomas and Weiss 2012). Though these observations were not accepted at the time, it began the modern study of starspots that continues all the way to today.

The next major discovery in starspots happened in Glasglow in 1769 by Alexander Wilson, who noted how the penumbra on the side farthest form the limb decreases faster than the width of the penumbra on the side nearest to the limb (Thomas and Weiss 2012). This led to the conclusion that sunspots are "a vast excavation in the luminous matter of the Sun", now called the Wilson Effect (Thomas and Weiss 2012). While this was interpreted by William Hershel—another sunspot astronomer—as the possibility that there was a cool world below luminous clouds ("places where the luminous clouds of the Sun are removed") (Figure 1.3), the Wilson Effect is now known to be because the radiation in the umbra comes from a deeper level within the Sun (Thomas and Weiss 2012).

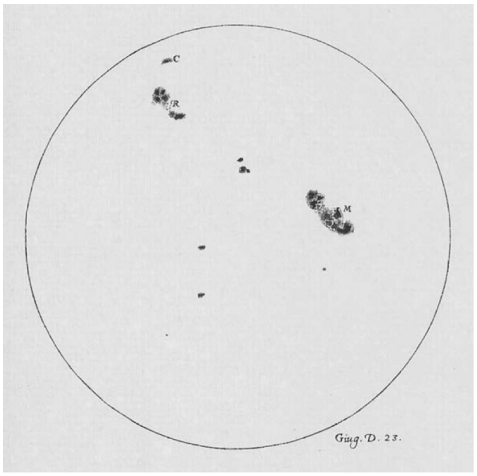


Figure 1.2: Drawing of Sunspots by Galileo from Thomas and Weiss 2012.

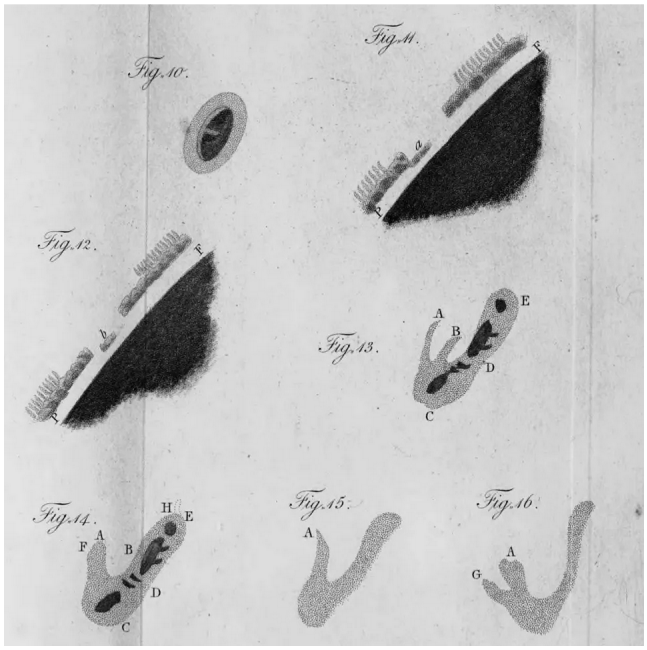


Figure 1.3: Drawing of the theory that Sunspots are gaps in luminous clouds by William Hershel gathered from Astronomical Society of the Pacific.

Continuous studies of spots throughout the next century began to reveal other important phenomena that today inform our idea of the formation of similar surface irregularities on other stars. In 1826, amateur astronomer Heinrich Shwabe, thinking that there could be another planet within the orbit of Mercury, studied the spots over 43 years uncovering a sunspot-intensity periodicity of 10 years (Thomas and Weiss 2012). While this had been noticed before by indigenous populations–notably in the periodic presence of auroras and, in one example, even in story by the Zambezi in Mozambique (who describe the jealous woman-

moon throwing mud at the face of the man-Sun on the same cyclic period) (Hamacher 2022)–this was the first published account of the phenomena, occurring in 1851. Just a sunspot cycle later, three astronomers (Edward Sabine, Rudolf Wolf, and Alfred Gautier), all independently discovered the relation between sunspot cycles and magnetic events on Earth (hence, the appearance of the auroral cycle in the cosmovision of many indigenous populations) (Thomas and Weiss 2012).

It was then discovered quickly thereafter by Richard Carrington the equator-ward shift of spots over cycles, from 40 degrees to 5 degrees, as well as the differential rotation between these spots depending on latitude. This finding demonstrated that the Sun was a fluid body (Thomas and Weiss 2012). This famous equator-ward drift and sunspot cycle are now referred to as the butterfly diagram, shown below (Figure 1.4)

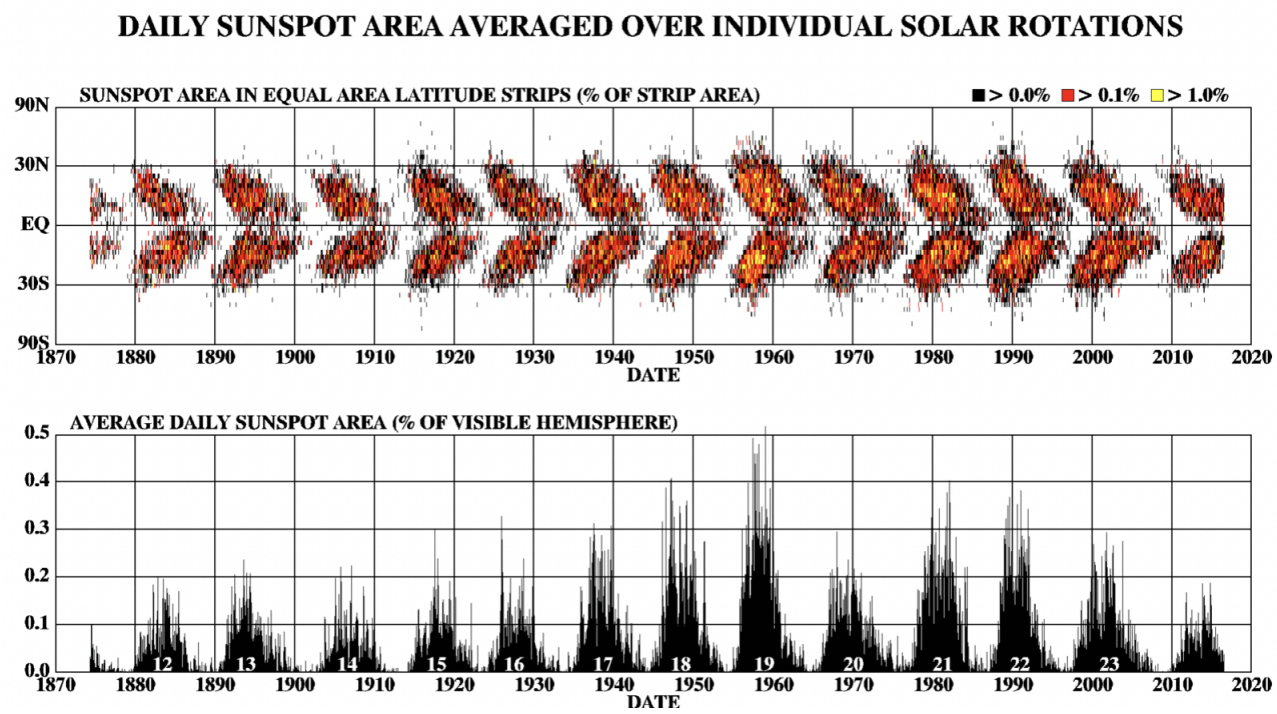


Figure 1.4: The Butterfly Diagram. Depicts sunspot cycles from 1878 to 2005. The top portion shows sunspot occurrence across latitudes, along with their migration equatorward over time. The lower portion depicts the percentage of the hemisphere covered by sunspots as a function of time. Image gathered from NASA.

Later in the century, Ernest Maunder drew attention to the correlation between the lack of sunspots in the 17th century–the Maunder minimum–and the little ice age that occurred at that time thereby linking solar activity with effects on life on orbiting planets (Thomas and Weiss 2012). Tracing solar cycles back to the medieval period, people have posited that such a connection could have been the cause for the medieval warm period—the same period when Greenland was settled (Thomas and Weiss 2012)–which occured in a time of intense solar activity. Such a sunspot-cycle-climate effect notably was known before by indigenous

populations, specifically the Dharwal of Australia who used the appearance of the tangara (the aurora) to announce the start of the Mudong weather cycle. This cycle was associated with hot, dry weather (Hamacher 2022).

As the 20th century was ushered in and the discovery of spectroscopy began altering the understanding of our Universe, more discoveries about the nature of the spots came to light, filling in the picture of this mysterious solar feature. Using Zeeman splitting in magnetically sensitive spectral lines, George Hale found there to be an intense magnetic field in sunspots—the first extraterrestrial magnetic field discovered (Harvey 1999). He also noted the systematic polarity pattern of the magnetic field, whereby sunspots occurred in pairs of opposite polarities with the leading spot polarity determined by the hemisphere (now known as Hale's polarity law) (Harvey 1999). This hemispherical lead-spot opposite polarity then switches hemispheres at the beginning of each cycle, revealing a dynamic magnetic mechanic happening with our star.

As photography came into popular use in the 1800s, it became the first time that high resolution structures of the sunspot could be depicted, making it possible to map the topography of the feature. Limited by atmospheric resolution, these photographs continued to improve all the way until the first unmanned balloon flight caught the amazingly detailed picture below (Thomas and Weiss 2012) (Figure 1.5).

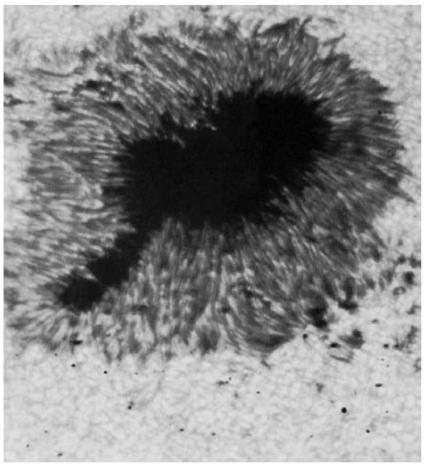


Figure 1.5: Sunspot image taken by Project Stratoscope in 1961. Image gathered from Solanki and taken by O. von der Luhe, M. Sailer, T. Rimmele.

Finally, in higher resolution study of our Sun, the presence of grid like cells, now known as "convective cells", were uncovered. These cells hinted at a new dynamic, convective nature of our star (Figure 1.6). Thus birthed a new age of solar physics.

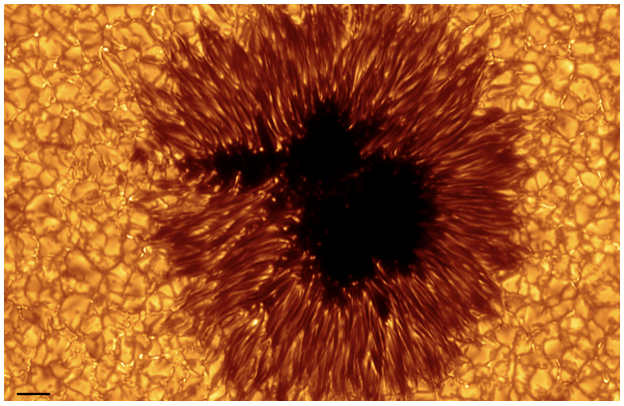


Figure 1.6: Detailed Sunspot image that depicts the granular nature of our convective Sun, image gathered from Wikipedia, originally taken by the Swedish 1-m Solar Telescope.

1.3 The Mystery of Starspots: What Are They?

The history of sunspots reveals interesting features but still does not necessarily answer the key question: what is a sunspot (starspot in general)? Until only recently in the 20th century, all that we knew was that that it was less bright and thus less hot than the Sun; that it migrated equator-ward over the course of a cycle that lasted 10/11 years; that it was a result of some dynamic within a fluid Sun; and that it was paired with the presence of intense magnetic fields with opposing-hemisphere-spot-pairs possessing opposite polarities. We also knew that it affected life here on Earth with high activity leading to warmer climactic periods (due to increased solar weather around the spots) and low activity leading to colder periods. But what exactly was a starspot/sunspot?

Succinctly, a starspot is a region on the photosphere of intense magnetic activity due to the dynamic nature of the star that suppresses convection by nature of tangled magnetic fields. This results in a relatively dark spot—like Nanhuatzin's pockmarks—on the face of a convective surface. In order to understand it more in depth, as well as the solar/stellar dynamo that produces such an occurrence, it is important to begin with the basics of stellar physics. In this elaboration, I will take note of starspot history and use our Sun as a guide. Later, I will highlight what differences do or may occur on other stars.

1.4 The Stellar Dynamo: The Sun As A Model

The magnetic nature of the Sun begins at the core (Figure 1.7), where the act of nuclear fusion takes place due to the high pressure and density of the center. This area takes up about 25 percent of the body of the Sun, and it is here where hydrogen is turned into helium that then powers the stellar body (*Solar Physics* 2022). In this process, two hydrogen protons (electrons are stripped away at the core), overcome their electrical repulsion due to the high density and temperature, and in a three-step process become helium. The excess

mass within this conversion is converted into energy in the form of photons which is then used to power the Sun, according to $E = mc^2$. The exact details of this three-step process called the pp chain are as as follows: (1) two protons collide to create deuterium, a positron, and a neutrino; (2) a proton collides with the deuterium to produce helium-3 nucleus and a gamma ray; and (3) two helium-3 nuclei collide to produce a helium-4 nucleus and two protons (Figure 1.8). The energy produced by this process can be on the order of 26.73 MeV, which then gets passed on from the core as high-energy radiation to the next layer of the Sun: the radiative zone.

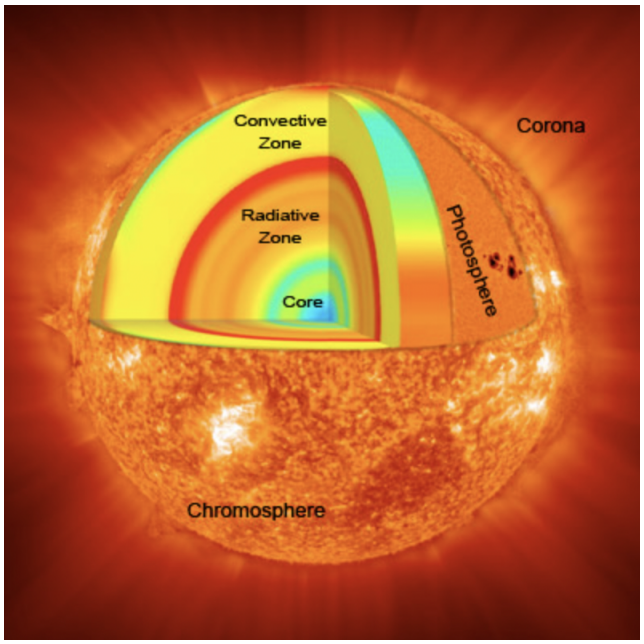


Figure 1.7: Diagram of the Sun depicting the most prominent features, particularly the fusion-burning core, the radiation zone, the convection zone, and sunspots. Image gathered from UCAR center for science education.

In the radiative zone (Figure 1.7), the high energy lightwaves interact with a slurry of particles, such as protons and electrons (*Solar Physics* 2022). In this interaction, the light is absorbed, bringing the particles to higher energy states, and then the light is re-emitted, later to be absorbed by other particles. This is known as scattering and is the next layer of how the core's high-energy radiation gradually leaves the star and loses energy to become visible light by the time it reaches the convective layer. It is notable to add that this process is incredibly slow. Due to the high density of the radiative zone and the amount of interactions taking place at this level, it can take up to millions of years for the radiation to get through from the core to the convective zone.

Once the sunlight leaves the radiative zone, it passes through an area called the tachocline, which refers to the transitive area between radiation and convection (Figure 1.7). This is the region where magnetic activity is thought to begin as radiation arrives to a cooler area of the Sun that hosts heavier ions like carbon and oxygen. Here, radiation cannot escape through

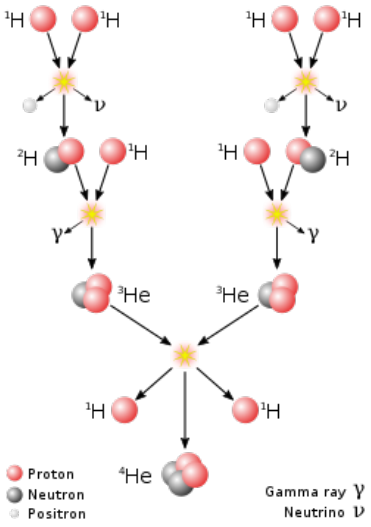


Figure 1.8: Diagram of the proton-proton chain. Image gathered from Wikipedia.

the thick plasma; instead, the heat is trapped and causes the plasma to begin to "boil" or convect. This occurs when the temperature gradient becomes larger than the adiabatic gradient, thus allowing for a bubble of material that is warmer than the surrounding fluid to move upwards (*Solar Physics* 2022). This bubble rises quickly, expanding along the way, until it breaks at the visible surface–the photosphere–having dropped temperature significantly (from 2 million Kelvin to 5700 Kelvin) (*Solar Physics* 2022). At the photosphere, these convective cells can be seen as granules, which are constantly in motion, coming and going, akin to boiling water.

Whereas in the radiative zone there is no flow of charged particles, from the tachocline onwards, there is a flow of ions–in many notable ways–which is the source of magnetic features. This flow of charged particles creates what are known as "flux tubes" or, rather, magnetic field lines, that cancel out and create a global magnetic field around the Sun. This global magnetic field looks very similar to a dipole whereby field lines emerge from the negative end of the Sun into the positive end of the Sun (depending on the cycle this corresponds to the South and North pole respectively or vice-versa). It is such a magnetic field we see below (Figure 1.9).

However, on the smaller-scale, magnetic field lines have a much more turbulent and chaotic nature. This can be related to the magnetic flux tubes becoming twisted as the Sun convects and rotates (Charbonneau 2014). In fact, the differential rotation that was noted by history's starspot researchers is the primary force which leads to this twisting. As the magnetic field lines move faster along the equator and slower towards the poles, they can become entangled within themselves, creating so-called "kinks" (notably at latitudes around the equator, Figure 1.10) These "kinks" then rise through the convection, "mushrooming" and splitting into a vortex pair with a north and south pole (Charbonneau 2014). These vortices produce a force on the surrounding charged material, suppressing convective movement, which in turn suppresses heat-transfer (Ruzmaikin 2001). As such, when these tangled

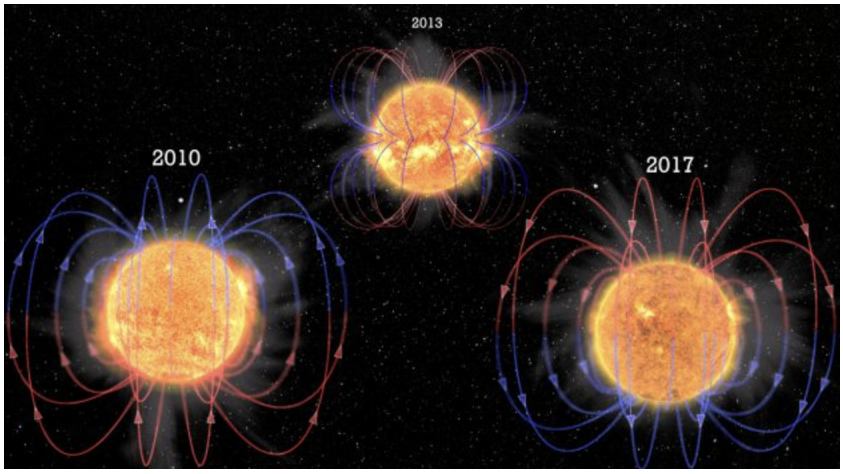


Figure 1.9: Diagram of the Sun's global magnetic field over three sunspot cycles in which the polarity flips. Image gathered from Smithsonian Magazine.

magnetic lines burst through the photosphere, there appears dark spots in pairs–known as Sun/starspots–that are cooler than the surrounding plasma (Ruzmaikin 2001). It is the creation and propagation of such magnetic fields and features–both local and global–that we call the stellar dynamo. Additionally, starspots are one of the few ways into understanding this chaotic stellar magnetism.

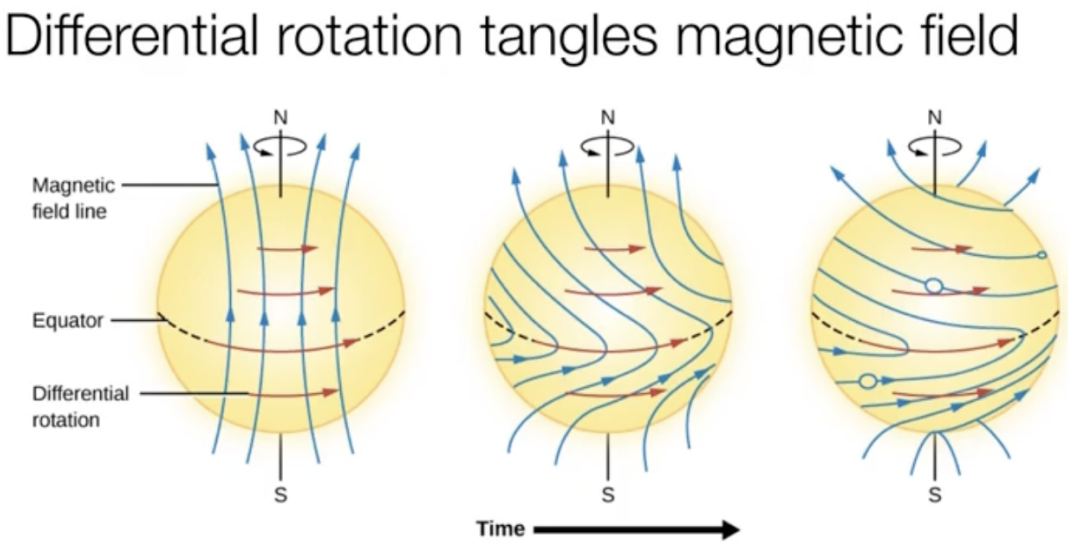


Figure 1.10: Diagram of magnetic field lines becoming twisted into "kinks" which then rise and suppress heat-transfer, creating Sunspots. Image gathered from University of Central Florida.

Thus, the story comes together: the differential rotation that was noted by way of sunspot

observation creates the tangled, intense, and polarized magnetic fields that Hale later realized characterized the dark patches of Sun. Through other phenomena noted in sunspots, we have come to realize more complex inner workings of the stellar dynamo. For example, the cyclic nature within which sunspot numbers increase to a maximum and then decrease and reverse polarities reveals that magnetic field lines can get tangled only up to a maximum before they begin to unwind themselves. Similarly, the fact that sunspots migrate from high latitudes to lower latitudes (equator-ward) demonstrates other complex flows within the Sun, in this case the meridional flow (conveyor-belt like flow that moves equator-wards within the Sun and pole-wards at the surface) (Nandy 2004) (Figure 1.11). Even the occurrence of certain climactic changes and geo-magnetic events can be tied to starspots/sunspot activity (and thus magnetic activity). In these cases, the strong magnetic fields send charged, energetic materials outwards towards earth in coronal mass ejections and solar flares. This likewise increases radiation on Earth and affects life (Thomas and Weiss 2012).

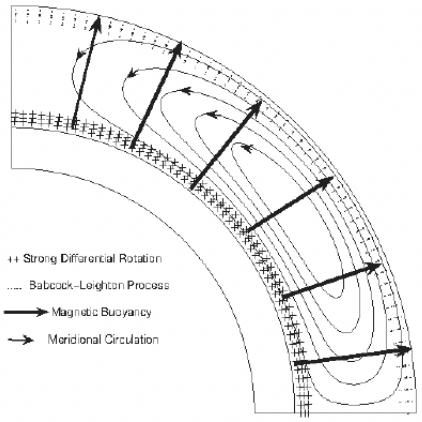


Figure 1.11: Diagram of meridional flow within the Sun. Note that the flow transfers energy equator-wards, which results in the migration of spots over a cycle.

With this, one can see why it would be necessary to understand starspots/sunspots in order to understand (1) more about stellar activity and (2) the effects on orbiting bodies (such as our Earth, particularly when considering the interactions with the bodies own magnetic fields). It is only through the study of such phenomena that we are able to access the complex microscale magnetic features of convective stars and thus probe the inner workings and interactions beyond what our eye can see. This continues to be an important field of study, even in our familiar Sun. Though we understand the basic inner-workings of the stellar dynamo as it exists in our host star there are still magnetic mysteries that abound, such as the odd appearances of the aforementioned centuries-cycle "Maunder Minumum's" or the extreme heating of the Sun's corona (atmosphere), which is millions of degrees hotter than the Sun's surface (*The Enduring Mysteries of the Sun* 2007). There is still much to learn, even here at home.

1.5 Starspots: Beyond Our Sun

When we get further away from our Sun, the mystery of magnetic features increases drastically. Not only is there a wide variety of different stars, which have possibly foreign dynamos of their own, but also the ability to study them becomes more difficult. The probing of starspots on our Sun (and thus the magnetic features) primarily is possible through visual observation, detailed cataloguing of evolution in number and location, and precise probing of the close, small-scale magnetic fields. As distance becomes greater to other stars, however, it becomes nearly impossible to visually resolve the variations on the surface, and all the small-scale magnetic fields cancel out for only big-picture ideas of the star. New methods of observation need to be sought out in order to understand anything more about the odd world of stellar bodies. What new magnetic phenomena may exist? How may it vary compared with our Sun? What new models of the stellar dynamo could this unveil? How could this inform our knowledge of convective stars, even our own Sun? And what may this mean for the features of orbiting planets? This is where the new field of starspot research enters the picture.

Currently, there are four general ways of observing starspots. This thesis offers a fifth possibility, only now possible through the long-cadence, highly precise observations of Sun-like stars. Before launching into this new method, it is worth noting the current methods that are in use. It is the utilization of various methods, including what is presented in the thesis, that provide the fullest picture of starspot phenomena. The existing ways of observing are photometric imagery, Doppler imaging, spectroscopy, and Zeeman Doppler imaging.

1.6 Starspots: Traditional Methods for Observation

1.6.1 Photometry

The photometric method for observing starspots consists of hunting for photometric lightcurve variations (variations within the brightness of the star as a function of time), where the variation is of the same period as that of the star. This hints at a surface variation on the star, rather than an orbital variation (Thomas and Weiss 2012). Models are then utilized to reproduce the lightcurve with various combinations of starspot properties, thereby narrowing the possible features of the surface phenomena (Figure 1.12) (Thomas and Weiss 2012). To determine the temperature ranges of these spots, it is also necessary to utilize the difference of v-r (Thomas and Weiss 2012). Because different temperatures have distinct black body radiation curves with varying intensities depending on wavelength, such a measurement can give access into the temperature. Likewise, by comparing the wavelengths of a spotted area in comparison to the photosphere, one can get access into the particular spot temperature. Though this is the most accessible of all the methods, it is also the most general: solutions are not unique, temperature is general, magnetic features are absent, and it is impossible to visualize any detailed information/evolution of the system.

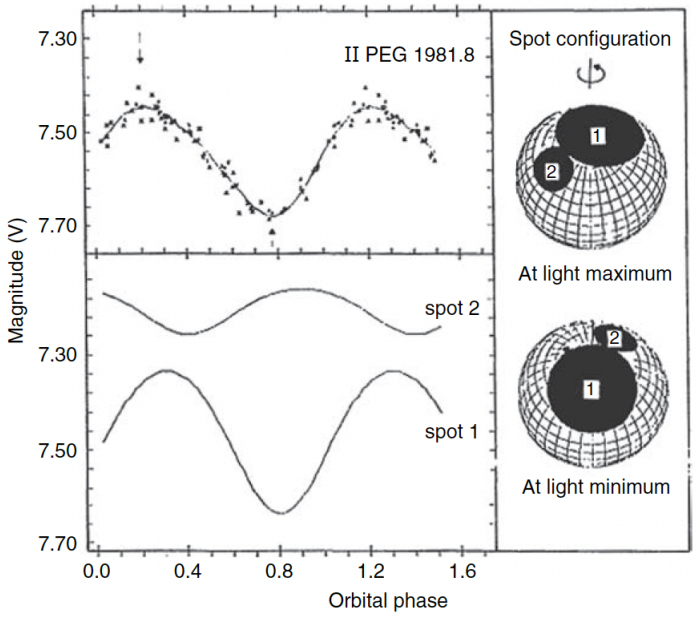


Figure 1.12: Process of mapping starspots through photometry. Two spots are modeled on the right to produce a lightcurve which matches the data for II Peg. Image gathered from Thomas and Weiss.

1.6.2 Spectroscopy

The next common way to resolve spots is spectroscopy, most notably utilized to determine spot temperature. Here, the dark spot reduces the depth of the absorption lines of certain temperature sensitive elements, allowing for the possibility to calculate the magnitude difference and thus the temperature. Similarly, another spectroscopic method of observing starspot properties is through molecular lines, typically TiO and OH (Thomas and Weiss 2012). This is because molecular spectroscopic lines are only possible within cooler environments where they can form, thereby presenting themselves only on dark starspots. These two methods allow us not only to access the temperature but also the location of these regions.

1.6.3 Doppler Imaging

The most notable of the methods for mapping distant starspots with locational accuracy is Doppler imaging (Thomas and Weiss 2012). Whereas photometric methods are imprecise, Doppler imaging allows for a way to understand where on a star a spot may reside. Doppler refers to the effect that happens when a fast star has wavelengths shifted redder or bluer depending on the direction it rotates as seen by the observer. That is, when the star spins towards the observer, the wavelengths "bunch" together as to be shorter and bluer, whereas when the star spins away from the observer, the wavelengths are "stretched out" as to appear longer and redder. Knowing this, a broadened spectral line profile provides a mapping of longitude, where a "bump" can be followed as it moves along the line of sight (Thomas and Weiss 2012). This "bump" represents a relative lack of spectroscopic absorption at that

wavelength, corresponding to a starspot moving along the rotation of the star (Figure 1.13) (Thomas and Weiss 2012). This can be quite accurate for individual spots, especially when enough spectral lines are observed and a high signal-to-noise ratio can be achieved. Such a process allows for the probing of latitude, longitude, and size. However, though accurate in location, this method fails in providing information about slow-rotating stars when these Doppler features do not arise (Thomas and Weiss 2012). This limits the bodies studied to primarily close binaries–notably not like our Sun. Additionally, this method works best for spots around the equator, where line profiles are wider. As we observe towards the poles, we find this method unhelpful, and as of recently, there is substantial evidence to determine that polar spots are a common feature of our Universe. Lastly, this method, as with the others previously mentioned, provide no new information on the magnetic nature of the star.

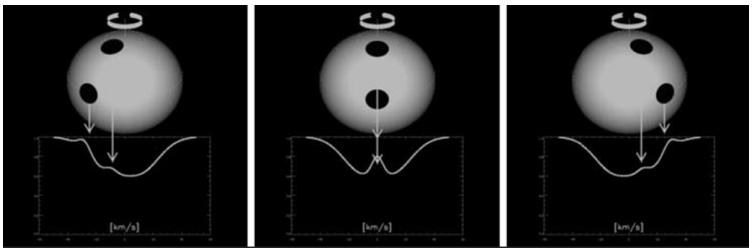


Figure 1.13: Illustration of the motion of bumps across broadened spectral line profiles (due to Doppler effect) for starspots. Image gathered from Thomas and Weiss.

1.6.4 Zeeman Doppler Imaging

Whereas the previous methods probe position, size, and temperature of the spots, the only known method of measuring the small-scale magnetic features of these spots is through a technique called Zeeman Doppler imaging. This is a modified version of the technique first used by Hale to discover the extraterrestrial magnetic fields of the Sun. The basic way in which Zeeman imaging works is that it looks at the splitting of electron energy states observed. When electrons are heated, they jump to a higher energy state. They then fall back to a lower energy state emitting energy as a photon emission line that can be measured. However, in the presence of a magnetic field the emission line can become split into many emission lines, varying by the intensity of the magnetic field (Donati and Brown 1997). This is because electrons, being spin one-half nuclei, have a magnetic dipole, which can align itself either along the line of the magnetic field (that is, alpha spin state) or against it (beta spin state) (Gomez 2015). Such an effect results in a change to the polarization of the light as well as differing results in the splitting of emission lines (Figure 1.14.) These differences can then be interpreted to determine field strength and direction for a distant extraterrestrial body (Donati and Brown 1997), particularly the changes in polarization given that in these fast-rotating stars, the split energy levels are lost in the broadening of spectral lines.

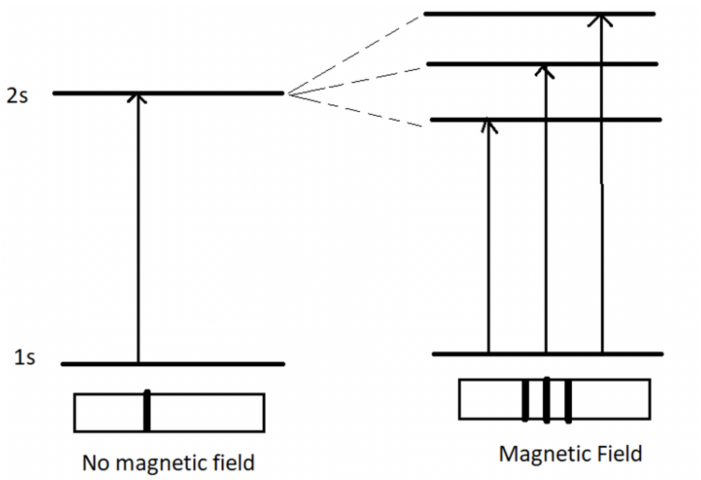


Figure 1.14: Illustration of spectral splitting in the presence of a magnetic field. Image gathered from Sanfoundry global education and learning.

Zeeman Doppler imaging takes this basic effect and complicates it by the relative velocities of starspot pairs on the rotating star. If the body was not rotating, it would be the case that the circular polarization signals of the two opposite polarity spots would cancel; however, because of their different location and line-of-sight velocity, the Doppler effect separates the contribution to the polarization signal by wavelength (Donati and Brown 1997). This allows for the subsequent line-of-sight magnetic field component to be calculated. Combining reading the polarization signals with Doppler imaging of the spot finally allows the viewer to pin point and track the movement of magnetic fields in the spot. It is necessary, however, in such small signals to use this technique over various spectral lines for meaningful signal-to-noise. Once successful, it is ultimately possible to get an idea of the magnetic mapping of the star, such as is shown on Peg II below (Figure 1.15) (Thomas and Weiss 2012).

1.6.5 Outcomes

It is predominately through these four methods that astronomers have able to begin to uncover new information about the mysterious magnetic nature of other stars. And in this research it has already conclusively been revealed that starspots are just as prevalent on other convective-driven stars as they are on the Sun. Following are some examples of the findings that have thus far been published about starspot (magnetic) phenomena on other stars.

1. It has been seen through photometric analysis that like the Sun, which has a sunspot cycle of 11 years, starspots commonly undergo cyclic variation.

2. It has also been found that cool stars with more rapid rotations demonstrate a higher level of magnetic activity and can have starspot coverage of up to 20% of the star

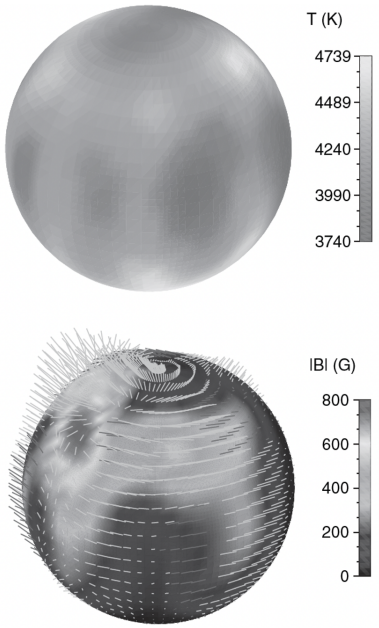


Figure 1.15: Starspots and magnetic fields mapped on Peg II. The top panel depicts starspot locations on the star observed by Doppler imaging. The bottom panel superimposes a magnetic field found by Zeeman-Doppler imaging. Notably, the magnetic field features align with the starspots. Image gathered from Thomas and Weiss.

(Berdyugina 2005). Furthermore, t is evidenced that these large spots can survive for many years on stellar surfaces, despite dynamic inner workings of the stellar body (Berdyugina 2005).

3. It has also been found, and much contested, that polar spots are a common feature amongst the convective stars observed. In fact, amongst these stars, which are cool and rapidly rotating (perfect for Doppler imaging), starspots preferentially form at 30 degrees and up, rather than towards the equator (Berdyugina 2005). This points to different dynamo workings and has been suggested to result from the extreme Coriolis force (due to rotation rate) making the flux tubes nearly parallel to the axis of rotation (Berdyugina 2005).

Observing other stars has not only highlighted possible differences, but it also has uncovered new findings about our own Sun. For example, with existing research, it seems that much of the same principles of the solar dynamo match that of the stellar dynamo, with complications due to the natural differences in evolution, rotation, and environment. This has led to new discoveries such as the existence of "flip-flop" cycles, whereby active longitudes switch hemispheres in a cyclic manner (Berdyugina 2005). Taking this finding, Berdyugina and Usoskin were able to uncover a 3.8/3.65 year cycle for our Sun that exhibited the same active-longitude-switching phenomena (Berdyugina 2005)

Looking at other stars that are similar to our Sun but for evolutionary stage also creates

the opportunity to learn about how magnetic phenomena can change with evolution. This shows how our Sun once was and how it might evolve to be. Thus, by means of starspot study, we work towards both unlocking the mysteries of our own pock-marked Sun as well as populating the sky with varied and unique stellar siblings.

Chapter 2

Data and Methods

2.1 A New Method: Transit Residuals

Analytical work has already been well underway on the former four methods of understanding starspots (photometry, spectroscopy Doppler imaging, and Zeeman Doppler imaging). This thesis works upon a novel fifth method of starspot observation–with its own unique benefits and drawbacks–that may add to this body of knowledge of the magnetic features on other stellar bodies. The basic mechanism of this process involves analyzing high cadence, high quality brightness measurements of a system of a star and an orbiting body as a function of time. Upon accessing the photometric data, we highlight the portion that depicts the transit of the orbiting body in front of the star–the "transit dip". This portion of the lightcurve is characterized by a periodic dip in the flux, given that some of the star is blocked. This is shown below in a characteristic planet transit (Fig 2.1). Though it may appear as though little information is gleaned from this plot (Fig 2.1), qualitatively comparing the scattering between the in-transit fluxes and the out-of-transit fluxes can provide important information about starspots. This is because as a body transits in front of its host star and the total brightness *decreases*, when it crosses over a starspot–a cooler darker patch on the star–the relative brightness *increases*. It may seem counter-intuitive that flux would increase by blocking a part of the star; however, when photometric data is obtained, it is taking into account a total integrated incident flux from the source. This includes the summed variations in surface flux that can be caused by those features that increase solar activity (flares, coronal mass ejections) and those that suppress it (spots). The variation as these features rotate along with the star is quite low, except in extreme cases, and nearly indistinguishable from noise. Yet, when the incident flux decreases due to a transiting body, the variations suddenly become much more noticeable. Their absence from the total integrated flux as the satellite eclipses increases distinctly more than that outside of the transit dip, given their relative negative contribution to the total lightcurve is now absent. Thus, by observing a notable difference in the flux scattering in the transit curve versus outside of the transit curve, one can deduce the presence of surface irregularities on distant stars too far away to resolve.

Additionally, using the orbital parameters from the satellite itself, one can map the

latitudinal locations of the starspot contributions, and using statistical comparisons with known spotted stars, one can gain insight into the number and size distributions. Whereas any multiple-body orbiting system works for this type of analysis, in this method a planet satellite is preferred, given the desired study and lack of knowledge of solar-like stars. This also allows for the added benefit of increasing the body of knowledge of light coming from the planet itself, as knowing the magnetic nature of a host star–and how active it is–suggests a more complicated spectroscopic analysis of its planets.

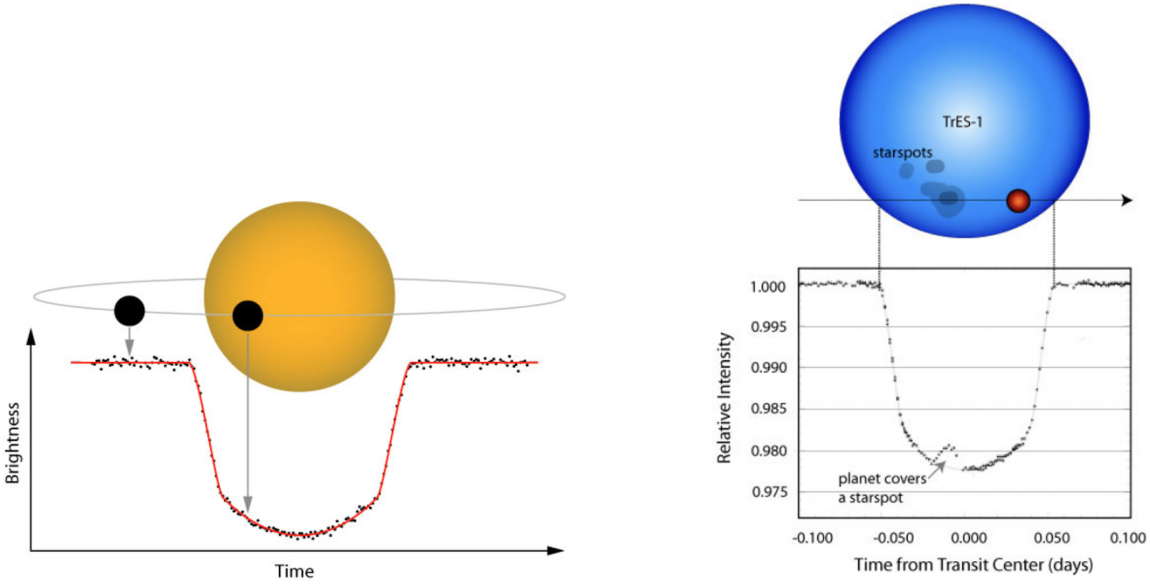
Unlike former methods of analysis, the transit-curve photometric scheme provides more detailed starspot information than the pure photometric approach and allows for the study of slower spinning bodies where Doppler imaging does not apply. It also permits the observing of spots wherever the latitude of transit may occur, meaning once again, that we are not limited to study around the equator only, which is another problem that affects Doppler imaging. For orbiting bodies that do not line in the same plane as the stars rotation, this method additionally allows the researcher to probe starspot crossings at multiple latitudes and compare.

On the flip side, it is important to acknowledge that this method provides non-discrete mappings of the starspot population and works towards giving general ideas about the spotted areas of stars (percentage of total area coverage, average number of spots, average size of spots, etc.) rather than specifics of a particular spot or spots. Still, when combined with the other methods of understanding surface irregularities of stars, it gives a quick and accessible way of noticing magnetic signatures and understanding generally what they look like–particularly their size, number, and latitude distribution. The crux of this thesis is the code that I have created which acts as a pipeline for this type of analysis. As we will see, my code allows the user to identify signatures from a given dataset, create simulations with known parameters to statistically match the given information, and understand what magnetic features may lie on these other stars.

2.2 Kepler: The Perfect Instrument

Satellite-crossing analysis requires incredibly precise photometric information that can distinguish incredibly minute flux differences. It also requires consistent observation over short intervals in order to catch the transit and consistently observe changes in spot signatures. Luckily, we have at our disposal this type of technology and, just as important, have access to this data. The data that this work focuses on is from the Kepler satellite, though it is also applicable to the new TESS (Transiting Exoplanet Survey Satellite) as well.

Kepler was created with the intention of discovering and studying potential Earth-like



(a) *(b)*

Figure 2.1: On the left a diagram and characteristic of a lightcurve without starspot signitures in-transit (image gathered from NASA). On the right, a diagram and characteristic lightcurve with starspot signitures in-transt (image gathered from Gomez.)

planets orbiting Sun-like stars. The method in which this was done was using transit lightcurves, that is, searching for those aforementioned periodic flux dips when the planet eclipses the star. When Kepler was launched in 2009, it was a groundbreaking point in time in exoplanet studies. Very little was known about planets outside of the solar system and very few had been observed. In addition, most of these planets were incredibly dissimilar to Earth, either fast rotators, and/or big Jupiter-like planets, discovered through greater flux dips or radial velocity measurements. In fact, before Kepler's launch in 2009, only around 300 discoveries had been reported, of which only around 70 were transiting systems (Batalha 2014). However, upon the launch of Kepler, this quickly jumped to over 3,500 viable planet candidates, and though not all of these came to be confirmed planets, these lightcurves are still supremely useful in potential starspot analysis. In reality, this was data from only a small amount of stars observed with Kepler, given the precise head-on orbital alignment that is necessary for observing a transit. Yet, these findings nonetheless opened up a whole new field of exoplanet/solar-system studies and uncovered a host of near-Earth planets orbiting Sun-like stars (Figure 2.2).

What made this transit study so successful was Kepler's astounding photometric precision for its time. The photometer was in fact engineered to observe 20–ppm relative precision in 6.5 hours for a 12th magnitude G-type main sequence star. To put this into perspective, Earth orbiting the Sun would produce an 84-ppm signal lasting about 13 hours. In addition to this high photometric precision, Kepler also operated continuously under a 30-minute cadence, releasing data in both short cadences (58.9 second exposure times) and long-cadences

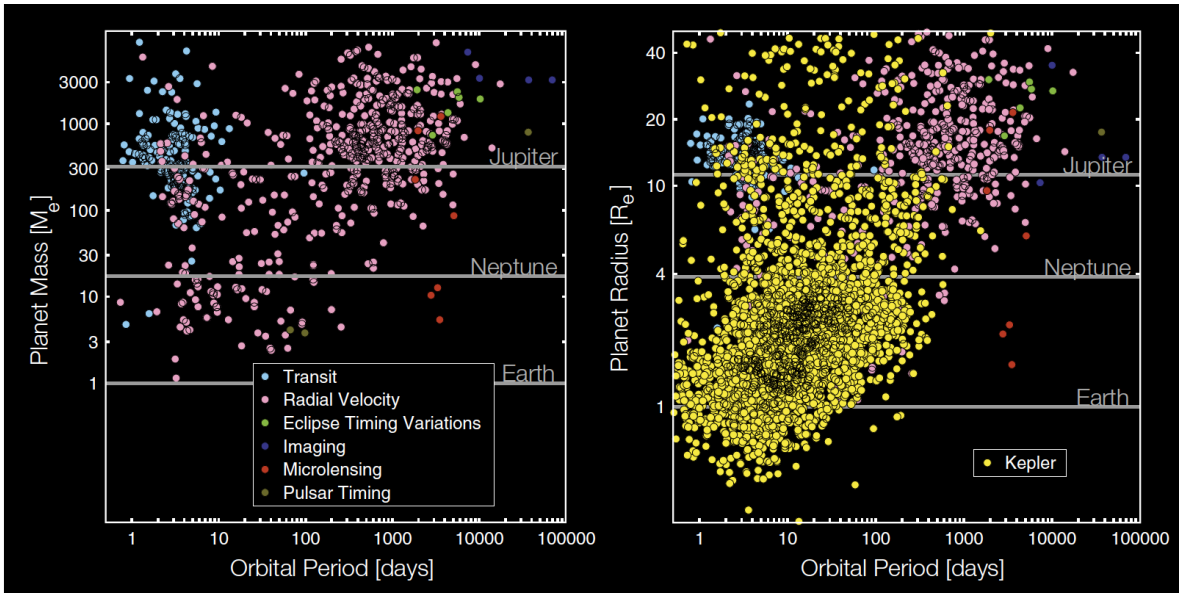


Figure 2.2: The findings of the Kepler satellite, demonstrating the small body of exoplanets known of before the transit-object survey and the large amount afterwards, particularly of smaller size. Image gathered from Batalha 2014.

(29.9 minute exposure times) (Batalha 2014). This was necessary in exoplanet studies as many of the transits can be quite short but with long orbital periods; hence, the continuous precise observation over short intervals. As for starspot studies, this provides the information necessary to construct the necessary high signal-to-noise ratios for minute starspot signatures.

The initial efforts of Kepler were incredibly successful, adding to the total wealth of exoplanet knowledge up until the degradation of the satellite's pointing stability by the loss of two reaction wheels. These wheels helped orient Kepler in the position to observe its commissioned patch of sky as it orbited around the Sun just behind Earth (Fig 2.3). Pointing outside of the ecliptic and into the Galaxy, this orientation had to be adjusted throughout the year, and so with the loss of the wheels, it became impossible to maintain a precise orientation that would result in the proper photometric data on the thousands of stars. However, the mission lived on and morphed into K2, a subsequent initiative which transitioned from the photometric hunt for transiting exoplanets and looked towards our solar system and local phenomena. This was until fuel ran out in 2018, and the mission was formally decommissioned.

In all the time Kepler was active, it was staring at a 115-square-degree field in the northern constellations of Cygnus and Lyra (Fig 2.3), gathering data for over 190,000 stars (Batalha 2014). This location was picked due to its location within the Milky Way, residing in the same direction as our solar system's velocity. This location also provided data from stars at roughly the same distance from the galactic center as us and in the galactic plane– important qualities in elaborations of the Rare Earth Hypothesis. This orientation to the

northern constellations also permitted pointing consistently far from the ecliptic and thus far from the Sun's glare over the course of the year.

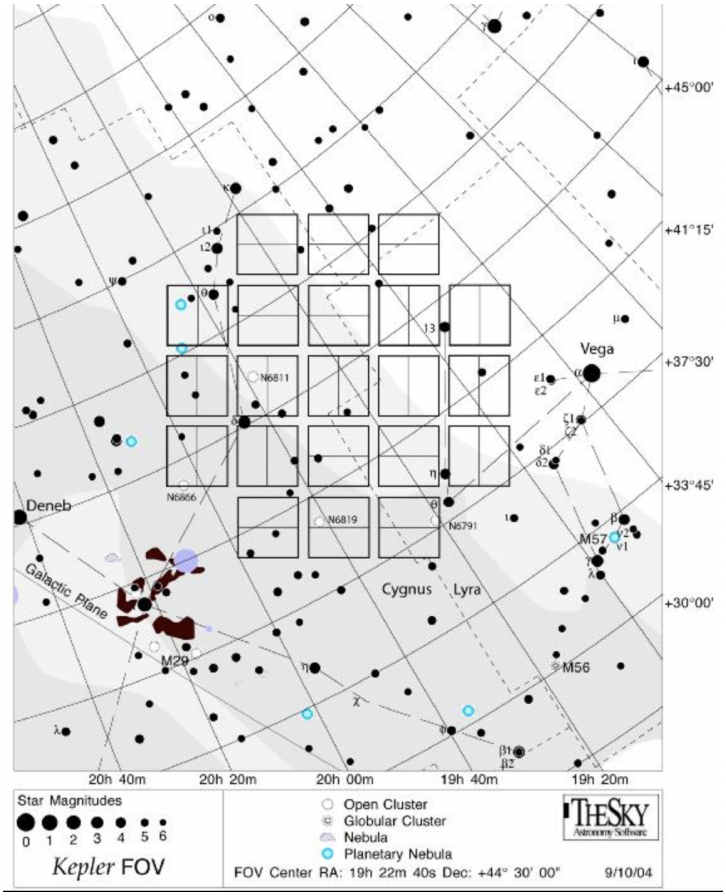


Figure 2.3: Portion of the sky near constellation Cygnus observed by Kepler and thus the patch of sky observed for the starspot analysis of this project. Image gathered from NASA.

Finally, of these stars which have transiting satellites, orbital parameters have been derived from the ground and recorded in the KIC (Kepler input catalog). The easy access to this information is also helpful for our own analysis, as such a database readily provides the vital information to calculate location and star properties for proper simulation. With these parameters and access to Kepler lightcurves, we have all the tools needed to operate the pipeline that I have made for starspot analysis.

2.3 On Lightcurves And Their Properties

The core of the pipeline that I created in this project to analyze starspot signatures is the lightcurves taken from Kepler's long cadence (exposure every 30 minutes), multi-year photometry. However, before we can better understand the features of starspot signatures

that can help identify magnetically-active stars and their phenomena, it is important to describe what factors influence a lightcurve in general and how we understand the geometry of the system. This section is also useful for defining and familiarizing ourselves with variables and processes that will be important later in my code. Note that the figures that follow as well as the geometric explanations were gathered from "Transiting Exoplanets: Measuring the Properties of Planetary Systems" by Carole A. Haswell, a book that outlines exoplanet discovery and science (Haswell 2010).

2.3.1 Period and Epoch

The two fundamentals to the lightcurve are their epoch and period. The epoch is simply defined as the middle point (or maximum depth) of the first transit observed in units of time; the period is the time between each star-occultation. The period can be easily discerned by calculating the time between each maximum transit depth. This information can further be used to estimate the semi-major axis a of the system and uncover further geometric orbital parameters, which I will go into next. This is done by Kepler's 3rd Law as shown below, where the mass of the star plus the mass of the planet is roughly equal to the mass of the star alone–a known quantity for a known stellar type:

$$\frac{a^3}{P^2} = \frac{G(M_\star + M_P)}{4\pi^2}. \tag{2.1}$$

$$a \approx (GM_*(\frac{P}{2\pi})^2)^{1/3}. \tag{2.2}$$

2.3.2 Impact Parameter

The impact parameter is the variable that defines the geometry of how we see a two-body system. It classifies the distance from the center of the star to the center of the planet/object that is orbiting. This is related to the inclination of the system, which defines the angle between the plane of the sky (our eye line to the star) and the plane of the orbiting body. It is also related to the semi-major axis, which as mentioned, can be easily calculated once a transit is observed multiple times. This geometry can be seen in Fig 2.4.

In Fig 2.4, one can see through trigonometric relations that the impact parameter b would be defined as

$$b = a\cos i. \tag{2.3}$$

If the impact parameter provides the distance from the center that the planet orbits, then this parameter also provides us with information regarding the latitude of transit of the planet. For example, if the impact parameter is 0, it is defining a system with a transit of zero inclination, that is, in the orbital plane of the sky. However, if the impact parameter is 1, this is defining a system with the transit 90 degrees from the orbit of the sky, that is, a non-transiting system. This is shown in the graphic Fig 2.5.

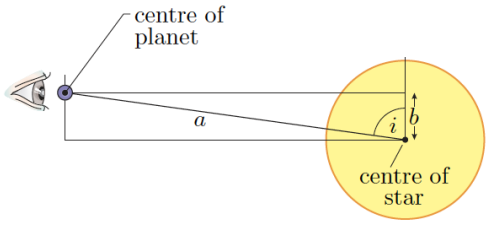


Figure 2.4: Diagram of the geometry of the impact parameter and planet-star occultations. Here, we can see the inclination as i, the semi-major axis as a, and the impact parameter as b. Image gathered from Haswell 2010.

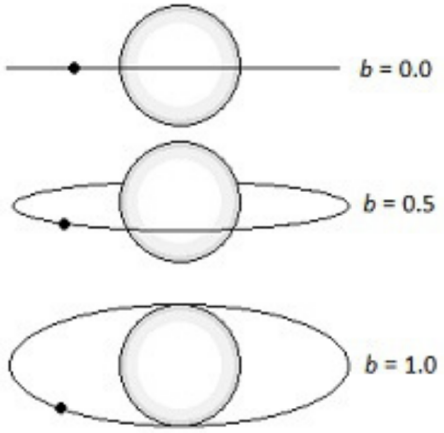


Figure 2.5: In this figure, we see the different possibilities for the impact parameter at its two extremes (1 and 0) as well as how it defines the latitude of transit for the planet. Image adapted from Gomez 2015.

2.3.3 Time of Duration

Upon knowing the impact parameter, one can then uncover the duration of a transit. This is an important parameter in the starspot pipeline as it allows one to highlight and model proper transit lengths (how long there is the flux dip.) We begin in understanding the duration of the occultation by looking again at the geometry of a system at fourth contact (just passing out of in front of the star) as shown in Fig 2.6.

We can see that by this geometry we can use trigonometric relations to define the length l as

$$l = \sqrt{(R_\star + R_P)^2 - a^2(\cos i)^2}. \tag{2.4}$$

Knowing this parameter, we can then turn our attention to Fig 2.7, which shows the full picture of the transit beyond what is just seen by the observer.

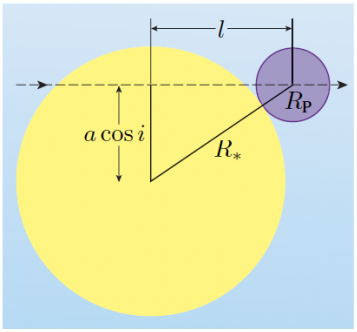


Figure 2.6: This diagram helps show and define the parameter l, which is the distance between the center longitude of the star and the center of the planet at fourth contact as seen by the observer. Image gathered from Haswell 2010.

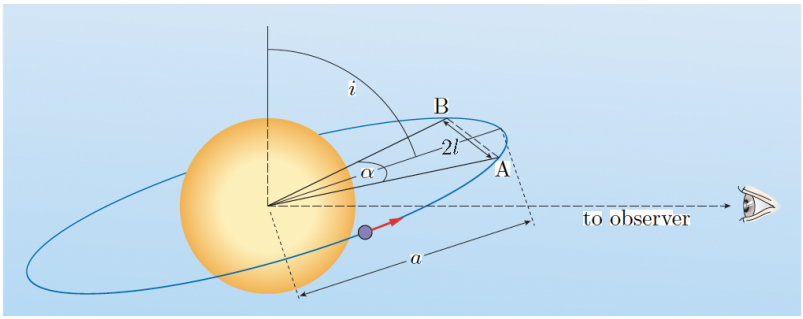


Figure 2.7: Diagram gathered from Haswell 2010 which shows the geometry of the orbiting system, particularly how we can present an angle subtended by the planet/satellite in terms of known quantities l and a.

Fig 2.7 shows how with l defined and the impact parameter and planet-to-star-radius known, we can find the duration of transit. Here, the planet moves from A to B, subtending an angle α. We can see from the triangle formed by A and B and the center of the star that $\sin \frac{\alpha}{2} = \frac{l}{a}$. Additionally, looking at the diagram we can define the time of duration as

$$T_{dur} = P\frac{\alpha}{2\pi} = \frac{P}{\pi} * \arcsin \frac{l}{a}. \tag{2.5}$$

Therefore, we can define the time of duration in terms of known values as

$$T_{dur} = \frac{P}{\pi} \arcsin \frac{\sqrt{(R_\star + R_P)^2 - a^2(\cos i)^2}}{a}. \tag{2.6}$$

2.3.4 Transit Depth

The transit depth of a lightcurve is simply defined by the amount of light blocked when the planet eclipses the star. This can be related to the ratio between the radii of the two bodies. Thus, the change in flux is

$$\frac{\Delta F}{F} = \frac{R_p^2}{{R_\star}^2}. \tag{2.7}$$

Note that this definition provides the maximum change that can occur. Therefore, it defines the depth of the transit. This ratio of planet to star size can result in minute transit signatures or very clear transit signatures (Fig 2.8). This is important for us in the recreation of proper model lightcurves, though typically the process works the other way around: looking at the flux-change upon observation is how one initially constrains planet radii.

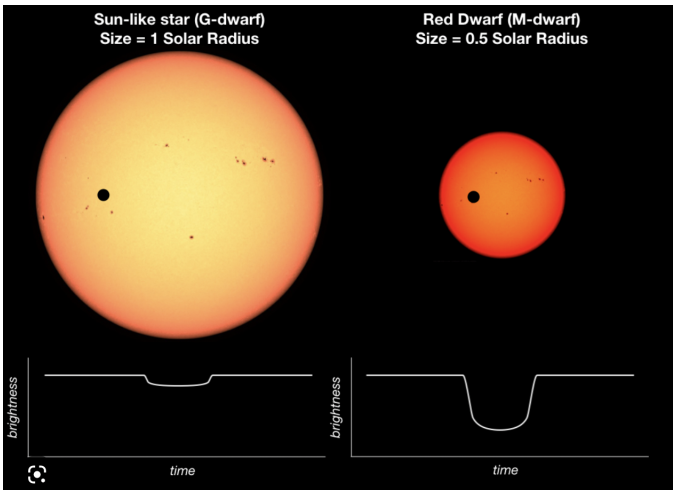


Figure 2.8: This figure demonstrates how the change in flux in the lightcurve is directly related to the ratio of the radii of the planet and its host star. This affects both how deep a transit curve can go and its starspot signitures, given that starspots will negatively contribute to a greater amount of relative flux in smaller stars (greater signatures) than larger stars (smaller signatures). Image gathered from Andrew Vanderburg.

2.3.5 Limb Darkening

Just looking at the ratio of the radii to one another, one would think that a transit would be box-like, immediately arriving at maximum flux upon eclipsing in front of the star and staying there until egress (outwards contact). However, this is not the case. Instead, transits have a characteristic u-shape. This shape can be attributed to limb darkening.

Limb darkening is a phenomena that occurs at the edge of stars that makes this area appear less bright than the center. This can be shown in the diagram 2.7 or many of the other diagrams/pictures in this thesis. Limb darkening occurs because depending on where a light is being seen from in a star, the photons have to travel through differing amounts of plasma. As the photon travels through the plasma, it undergoes various scatterings, absorption, and emissions as previously described when outlining the structure of our Sun. This changes the nature of the light that is seen from what has been sent through the radiative zone and

limits the probability of escape from the star itself. This probability of escape is called the optical depth, which is defined below as τ_ν. In the following equation, κ refers the changing opacity and the ρ refers to the changing density of the fluid.

$$\tau_\nu = \int_X^\infty \rho(s)\kappa_\nu ds. \tag{2.8}$$

In addition to this definition, the probability that a photon travelling along a path without being absorbed or scattered is $e^{-\tau_\nu}$, relating strictly to the optical depth. Therefore, the ratio of the intensity measured by us to the intensity emitted is

$$\frac{I}{I_{\text{emitted}}} = e^{-\tau_\nu}. \tag{2.9}$$

Now, even though the photons are emitted from the same radial location within the radiative zone, as mentioned, it is the case that they travel through different amounts of plasma as they travel through the star's plasma. Below is a diagram of this scenario (Fig 2.9. We can see that a ray that is emitted on the left travels through a path length of s that is greater than the path length of h, though they are emitted at radially the same distance.

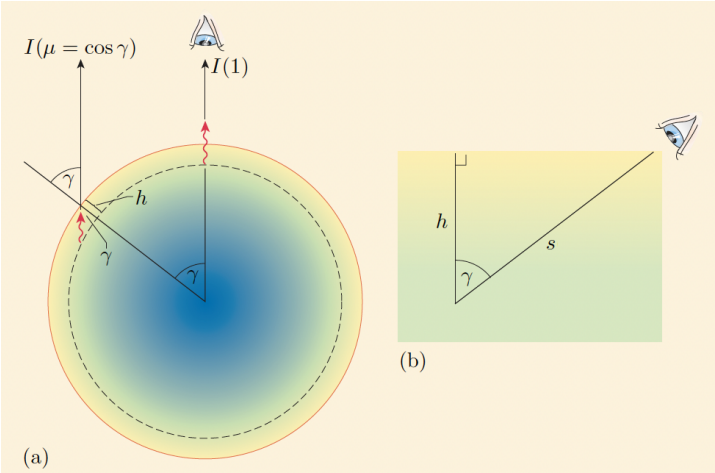


Figure 2.9: Image gathered from Haswell depicting the geometry of emitted light. Note the difference in the path length of h and s. At the center of the star, the light only moves through a path length of h to reach the observer, whereas on the side it needs to go through a path length of s

In greater detail, if the light is emitted at an angle γ from the radial line that points to the observer, the light would travel through a distance

$$s = \frac{h}{\cos\gamma} = \frac{h}{\mu}. \tag{2.10}$$

This shows that the path-length through which the photons travel near the limb is greater than that path-length through the center. Looking back at our equation for optical depth, the increase of path-length increases optical depth, which in turn decreases the amount of

photons that reach us (as seen in our equation for intensity). Fewer photons reach the observer and an effectively darker part of the star is produced.

Limb darkening affects the transit lightcurve, giving it a shape that is not boxy, but rather a type of "U" as the planet passes from lower-flux radial-rings to higher-flux radials-rings. This shape can vary depending on how the limb darkening evolves, and so various limb-darkening laws can be attributed to the transit evolution, going from simple linear darkening law to a cubic law. The typical limb-darkening laws that can be applied to the shape and used in a model are below, all depending on μ, defined in the previous equation, along with limb darkening coefficients that match the data. Below I define various limb darkening coefficients of interest.

Linear limb darkening

$$\frac{I(\mu)}{I(1)} = 1 - u(1-\mu). \tag{2.11}$$

Logarithmic limb darkening

$$\frac{I(\mu)}{I(1)} = 1 - u_l(1-\mu) - \nu_l \mu \ln \mu. \tag{2.12}$$

Quadratic limb darkening

$$\frac{I(\mu)}{I(1)} = 1 - u_q(1-\mu) - \nu_q(1-\mu)^2. \tag{2.13}$$

Cubic limb darkening

$$\frac{I(\mu)}{I(1)} = 1 - u_c(1-\mu) - \nu_c(1-\mu)^3. \tag{2.14}$$

2.3.6 Time of Transit, Ingress, and Egress

The final parameter that is necessary for the construction of the starspot-mapping pipeline is calculating the time of transit, ingress (the time of crossing fully in front of the star), and egress (the time of crossing fully out from in front of the star). In order to calculate this time, however, we first need to calculate the separation between the planet and the star at any time. As shown in the figure Fig 2.10, the geometry of the system demonstrates that this separation is equal to

$$s^2(t) = (a \sin \omega t)^2 + a(\cos i \cos \omega t)^2. \tag{2.15}$$

$$s(t) = a\sqrt{\sin \omega t)^2 + (\cos i)^2 (\cos \omega t)^2}. \tag{2.16}$$

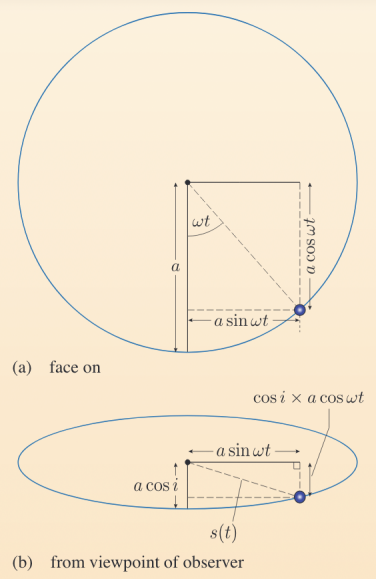


Figure 2.10: Image gathered from Haswell 2010 showing the geometric separation between the satellite and its host star over time.

Note that the extra $\cos i$ term can be demonstrated in part b of Fig 2.10. Because the orbital system is foreshortened by a factor of $\cos i$ it is necessary to multiply this by the displacement in the line-of-sight direction. Thus, it enters into the equation.

With this separation defined, we can look back at the system's geometry to see that when the separation is between $R_* + R_p$ and $R_* - R_p$ we are describing as system of ingress or egress. This is because any separation that is below $R_* + R_p$ would define a system with at least a partial transit (separation lies partially within the radius of the star). On the other hand, $R_* - R_p$ would define a system with at least part of the part of the planet *not* blocking the star. Any separation in-between is by definition ingress or egress.

Chapter 3

Overview of the Three Programs

Now that the different aspects of a lightcurve model have been detailed out and described, one can understand the inner-workings of a seemingly black-box of code. Moving onto the computational part of the project, the pipeline that I created gathers Kepler data and creates simulated lightcurves with known starspot properties to compare against the real data. This consists of three separate programs, each of which I designed and wrote. The intention here is that each may be used separately for other functions, such as isolated reduction of the data or isolated simulation.

The first program is entitled Starspot Program 1. In this program, data is gathered from Kepler and the routine cleaning, normalizing, extracting transit information, aligning transits, and exportation is conducted. The second program, called Starspot Program 2, takes the data from Kepler and runs it through a previously developed c program named STSP (STarSPot). STSP takes the flux of each time point, as well as other characteristics of the orbital system, to develop a new simulated ideal lightcurve (no inherent noise) with starspots overlaid. In addition to this, it prints out a data table of the simulated flux and starspots. The final program, named Starspot Program 3, takes either the original Kepler data or the simulated data (it operates with this dual functionality) to perform a starspot analysis. This is why there are two sections dedicated to this program later on in this thesis. Starspot Program 3 compares the scatterings in and out of the transit and runs a statistical measurement to provide an understanding of the comparison. When it is specified that the data is coming from the simulation, the program additionally provides a visualization of the generated starspot locations and relative sizes, including those that were crossed and those that were not in the simulation.

3.1 Starspot Candidates

According to Michelle Gomez's prior thesis, which looked at the body of Kepler data and sought to categorize and compare possible starspot signatures, a table of promising Kepler starspot stars are reproduced in the appendix (Gomez 2015). It is from this table that I chose kepid (keplerID) 10616571 as a working example in the elaboration of my pipeline,

though the program itself is applicable to any kepler object and theoretically TESS objects as well. This object is what will be presented in all the data, though my results will highlight features of other systems gathered through the same process.

The code I describe below provides a different analysis to starspots than that which was conducted for Gomez's thesis. I provide an updated code to Gomez's that can take down the Kepler lightcurves, clean them, normalize them, gather and overlay all important transits, and create a data table with this relevant for future codes. I also provide the functionality of reproducing a similar data table of candidates above, though this is not the aim of this project–merely an additional functionality for further work. The primary function that I will outlay is to garner more understanding about the spot features of other stellar systems. This would not have been nearly as possible without the groundwork Gomez laid down (Gomez 2015).

3.2 Starspot Program 1

As previously mentioned, the goal of Starspot Program 1 is to gather and clean the kepid (keplerID) data so that it provides the best version of the information necessary for further analysis. This consists of first gathering the data from the Kepler data archive. This is done using a Python program called lightkurve, which permits the user an accessible pipeline to download Kepler and TESS data (Lightkurve Collaboration et al. 2018). I then use the program's own command to clean the data of outliers and stitch the lightcurves together. A sample of the cleanest type of long-cadence lightcurve provided is displayed below (Fig 3.1). Here, I am choosing long-cadence data for its accuracy and consistency, despite the need for precise time stamps that the short-cadence data provides. The gaps in the data represent time periods in which Kepler was not taking data, and the line connecting the two data collections is an artifact of the plot. As is evidence in the graph, there is a high quantity of information; however, it is also clear that it is not yet usable and that it needs to be broken down in order to be helpful for transit analysis. We need to find some way to turn this mess into usable transit scatterings.

We do this by using the previously mentioned NASA exoplanet archive to download the pertinent information regarding the orbital characteristics of the observed Kepler system. This information is helpful in a number of ways, providing the duration of the transit, the epoch, the limb-darkening coefficients, the impact parameter, the period of orbit, the planet-to-star-radius, the duration of the transit, the planet-to-star-distance, and the inclination. By using the parameters for epoch and period, we are able to highlight from the total Kepler lightcurve the first transit and each subsequent transit. This was done by the calculation below, where the period and epoch are from the Exoplanet Archive and the transit number is an integer which represents the numbered transit since the epoch.

$$\text{Transit} = \text{Epoch} + \text{Period} \times \text{Transit Number}. \tag{3.1}$$

Examples of some of these transits over the course of Kepler's observing period of this

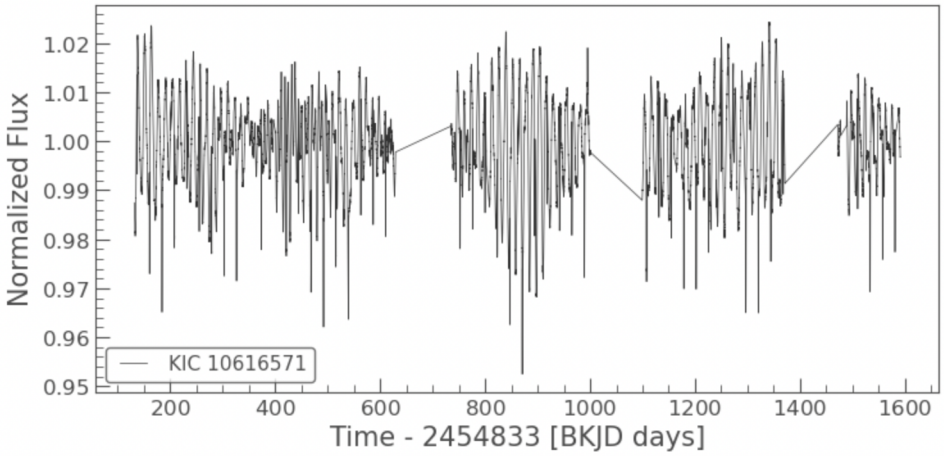
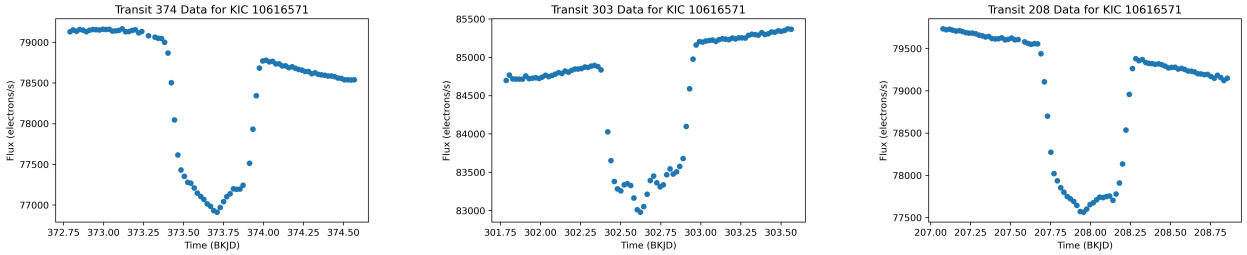


Figure 3.1: A sample plot for KIC 10616571 that is representative of the cleaned data downloaded from the Kepler archive. The bottom axis represents Barycentric Kepler Julian Date (BKJD) and the y-axis is the normalized flux–that is, the variance from the stars typical flux.

star are shown below, where the x-axis is in time units of BKJD (Barycentric Kepler Julian Date) (Fig 3.2.)



(a) Extracted transit BKJD 374 *(b) Extracted transit BKJD 303* *(c) Extracted transit BKJD 208*

Figure 3.2: Three examples of extracted transit, with the same amount of phase taken from both sides of the calculated epoch.

As is exhibited in the transits, the out of transit data tends to skew diagonally due to starspots on the surface moving in and out of view as the star rotates. This inhibits a quick and easy process for overlaying all transits and getting an idea of cumulative scatter/general properties of the curves. An overlaid plot of all the extracted transits is shown in Fig 3.3. Note that in order to overplot all the transits, we plot "phase" on the x-axis instead of time. Phase is a number between -0.5 and +0.5 and represents the fraction of the way a planet is around its orbit. Here, in Fig 3.3 we can see that in addition to a skewing of the data, there is also the problem that each transit does not lie on the same level of flux. This is a typical problem that can occur in such sensitive flux data due to photometric cross-contamination.

Before comparing the data to a model transit lightcurve, some manner is called for that allows for the "flattening" and normalizing of these given lightcurves. Such a data-cleaning

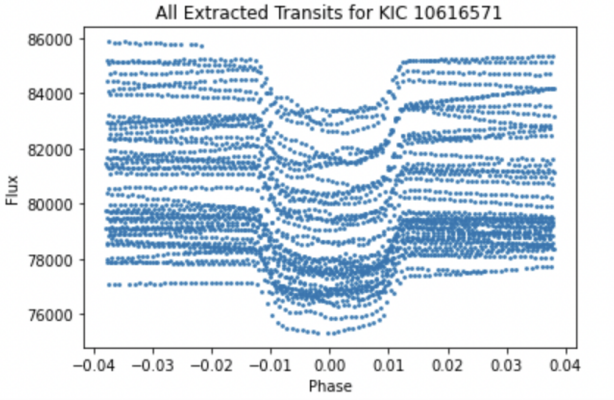


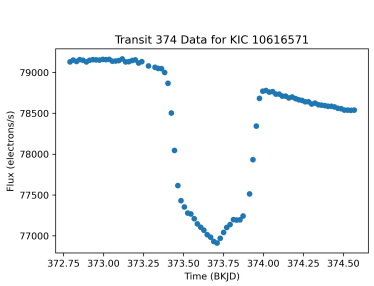
Figure 3.3: Transits overlaid without any normalization steps. Note the different levels of flux and how the transits do not initially align.

can be conducted by dividing by the maximum flux of each transit. As exhibited in the problem graph, all transits appear at different average fluxes; however, once the points are divided by the maximum flux, the y-values become normalized and relative to their maximum flux. This collapses the scale of all transits into one in which all data points are on the same level and overlaid.

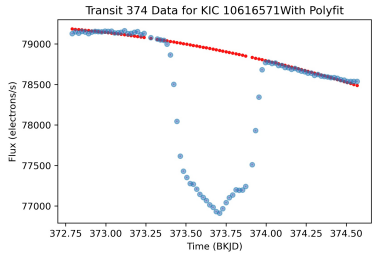
In order to get rid of the curving, we fit a polynomial to the out-of-transit data and divide all points by this fitted polynomial according to the equation below. Such a method of normalization and flattening preserve the transit depth even when the overall flux is changing.

$$\text{Normalized Flux} = \frac{\text{Flux} - \text{Polynomial Fit} + \text{Max Flux}}{\text{Max Flux}}. \tag{3.2}$$

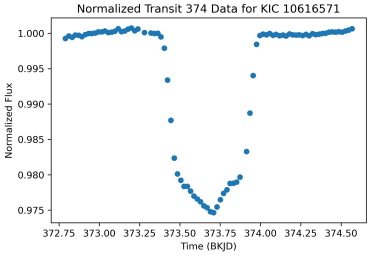
Such a process ultimately allows for the desired statistical "flattening" of all the data so that they are in relative flux units, by which we can compare the observed lightcurves to the theoretical model. A visual representation of this process is shown in Fig 3.4:



(a) Extracted transit BKJD 374



(b) Transit BKJD 374 with polynomial fitted



(c) Normalized transit for transit BKJD 374

Figure 3.4: Normalization method whereby a polynomial fit (in red) is fitted to the out-of-transit data, subtracted from the actual data (in and out of transit), then finally summed with and subsequently divided by the maximum flux.

3.3 Starspot Program 3: Data

Overlaying all the plots with its normalized data, we are able to see the proper aligning has been performed (Fig 3.5.)

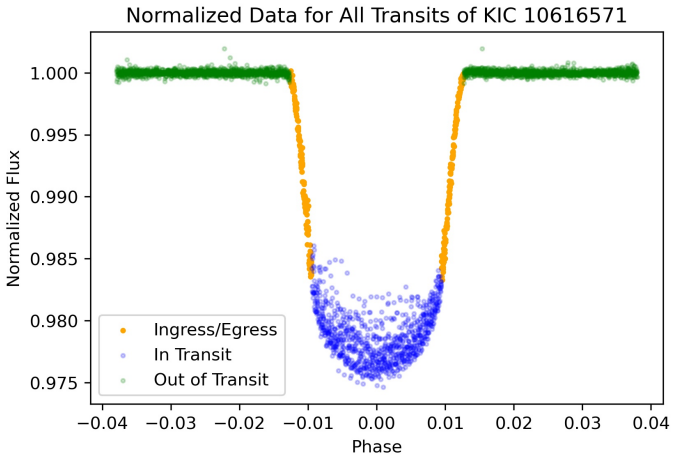


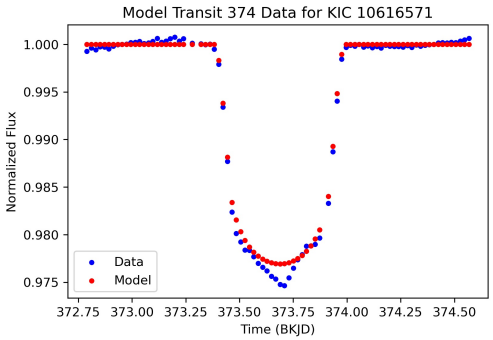
Figure 3.5: Transits overlaid after conducting normalization steps. The three different colors separate at which part of the transit the flux point is occurring, either in-transit, ingress or egress, or out-of-transit.

It is important to note in this figure that the distinct color features derive from the different parts of the transit that can occur at each flux point. These three points are in-transit (blue), which means that the satellite is transiting in front of the star; out-of-transit (green), which means that the satellite is not passing in front of the star; and ingress/egress (yellow), which refer to the satellite either entering or leaving the transit. Each datum is calculated to lie within one of these areas by the criteria highlighted in section 2.3.6. To do this, I calculate a "separation value" s at each timestamp using the parameters known, and then I split these up into when s is less than the sum of the planet and star radii; greater than the difference of the planet and star radii; and when it is between these two values (egress/ingress).

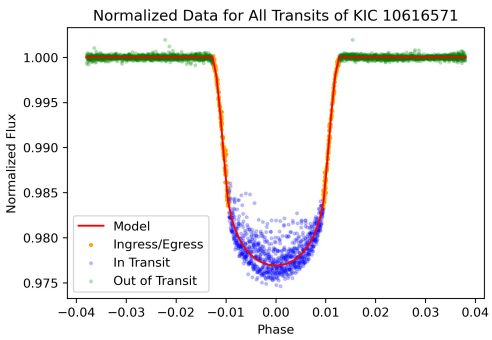
Already in this figure, one can see the difference between the scatterings within the transit and that outside of the transit. This is a visual representation of our method earlier described in Section 2.1 that describes the increase in flux-scatter within the transit as a starspot signature. Such a high scattering is a direct indication that there are surface features of KIC 10616571 that can only be resolved through this indirect method of observation. Additionally, because these flux-points are increasing as the satellite transits, the points most probably represent features that are darker than the average surface of the star as explained at the beginning of the methods section. This establishes KIC 10616571 as possessing starspots.

In order to fully quantify these spots, it is necessary to understand how much they truly diverge from the "idealized" transit. To do this, I use the known orbital parameters of the star and its satellite to implement an "idealized" model lightcurve. I say here *idealized*

because it does not take into account noise, stellar features, problems in the data, or shapes that are not perfect radiant circles. The model I use was created with the batman Python package, an open-source package which takes the epoch, period, planet-to-star radius, planet-to-star distance, inclination, eccentricity, longitude of periastron, limb darkening coefficients, and limb darkening model to generate a corresponding lightcurve of the system in such a way that reproduces my discussion in "On Lightcurves and Their Properties" as well as the modeling done by Mandel and Algol in 2002 (Kreidberg 2015). For the sake of brevity, I will not go into more specifics of the batman package, though the details can be found in the paper listed in the references of this thesis. For the most part, the primary functions of batman's modeling are nearly identical to the discussion of lightcurves and their properties. With such a model implemented, we thus have a version of this specific transit in a "perfect" world. Such an instance of the difference between model and data can be seen below which exhibits the model's exclusion of minute variation.



(a) Transit data (blue) overlaid with calculated model (red).



(b) All the data for KIC 10616571 with model overlaid

In order to highlight the scattering and the starspot signature and quantify the difference, we eliminate out the general shape of the model/the transit. To do this, we subtract the model flux, which, being based on the parameters of the orbit, approximately matches the data when not including starspots/noise (as observed in Fig 3.6b). This produces a graph of only the residuals, that is, that data that represents flux differences not accounted for by the satellite transit itself. For the case of KIC 10616571, this is represented in Fig 3.7. Here, ingress and egress are eliminated from view, given that it is hard to model these limb-darkening areas perfectly and a little variation can contribute to statistically skewed data when comparing in-transit vs. out-of-transit. Additionally, the effect of crossing starspots is more pronounced when comparing the in-transit, out-transit extreme.

These residuals are then statistically analyzed using a cumulative sum distribution and the Kolmogorov-Smirnov Test. These statistical tools will be expanded on in the analysis section of this thesis.

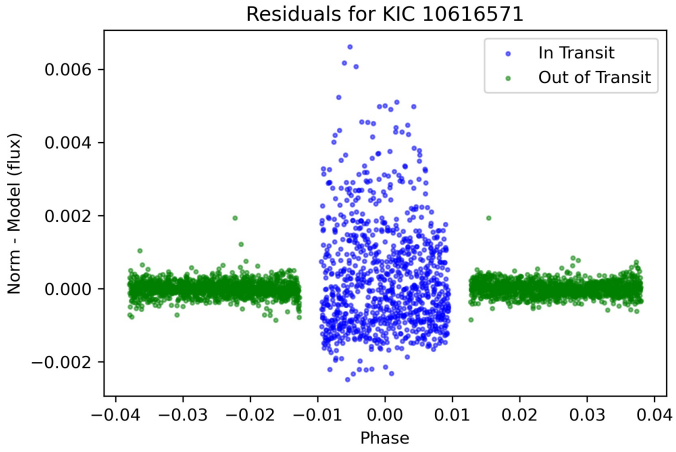


Figure 3.7: Residual of the lightcurve after dividing through by the created model. Note the extreme difference in variation in-transit as compared to out-of-transit.

3.4 Starspot Program 2

In order to glean any information from the data beyond the probable existence of starspots–and quantifying this probability through statistical tests–there must be something to which one can compare the data signatures. It is thus necessary to create a model for this very purpose. This is where new information may arise about the actual nature of the starspots on the other stars, particularly size and location distribution. To do this, I created another program, called Starspot Program 2, which writes a parameter file for use in another program called STSP (STarSPot) that creates synthetic lightcurves with starspot crossings in the transit. A sample of the in-file needed to be read by STSP is not provided in this thesis, but nonetheless, it is important to note that the model is created in conjunction with the lightcurve file provided beforehand, which contains flux and phase/time information.

To create the parameters of the spots crossed, I provide the program with a radius, a latitude, and a longitude for each simulated spot. The longitude is selected by dividing 2π by the number of spots specified by the program (with a program-limited maximum of 10 spots for each given simulated transit, later to be updated to 16 at the culmination of this thesis's creation.) This step can later be changed into a larger distribution of longitudes, which more accurately models the randomness of starspot location that may occur. However, for elaboration of this program, I keep the longitude selection simple at only 10 (later, 16) possible and known locations.

The latitude is chosen randomly from a normal distribution between either $\pm$ 20° or $\pm$ 10° away from the latitude of transit. 10° is sometimes opted for because by limiting the latitude of spot creations, we are able to maximize the number of starspots crossed (later to be talked about.) The latitude of transit is calculated by the equation below as agrees with the orbital geometry of the system previously mentioned.

$$\text{Latitude} = \arcsin(\text{Impact Parameter}). \tag{3.3}$$

Finally, the size is chosen according to the distribution of sunspot sizes, which acts as our base spot size to provide a sense of scale to our simulated data. This distribution is log-normal (Fig 3.8) as gathered from Choudhary and Hathaway where the mean is 4.95 and the standard deviation is 1.11 (Hathaway and Choudhary 2008). This gives us a selection of area sizes; however, in order to be compatible with STSP, we require a conversion into stellar radii. We thus choose the radii in a multi-step process. First, we create the log-normal distribution and select randomly from it. Then, we omit every instance that is below 15 micro-hemispheres (10^{-6} solar hemispheres) in agreement with research done by Baumann and Solanki which establishes this as the minimum size of sunspots (Baumann, I. and Solanki, S. K. 2005). We finally convert it back into a usable radius by the process below, where 1000 micro-hemispheres is 3043×10^6 km^2 and 1 solar radii is 695, 700 km:

$$\text{Spot radius} = \frac{1}{695700}\sqrt{\frac{e^{Area} \times 3043.7 \times 10^6}{1000\pi}}. \tag{3.4}$$

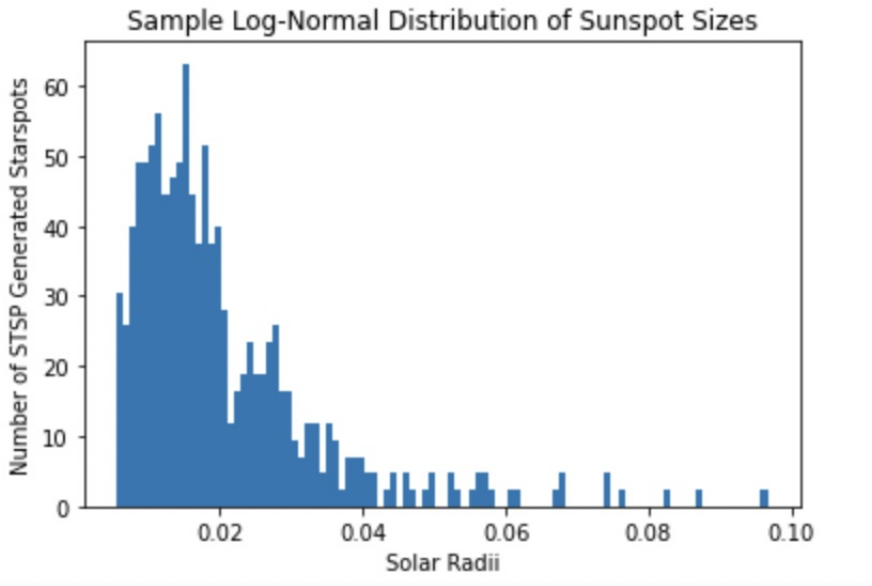


Figure 3.8: Sample graph of a log-normal distribution of starspot radius sizes. Sizes were randomly selected from a distribution as shown.

This code is also able to adjust the spot area parameter by a constant (while all else remains the same), thus allowing the user to compare size distributions to understand generally the size of the starspots that would match the starspot signature in the data.

Once the parameters of the spots as well as the orbital system are provided for STSP, this program is externally called to produce a model lightcurve with spots at the provided locations. What is then read back to the written code is the produced lightcurve, which

simulates an idealized lightcurve similar to that aforementioned with batman but with the addition of starspots that would be seen if the idealized system were observed with specified spots. A random example of one of these lightcurves is reproduced below (Fig 3.9.)

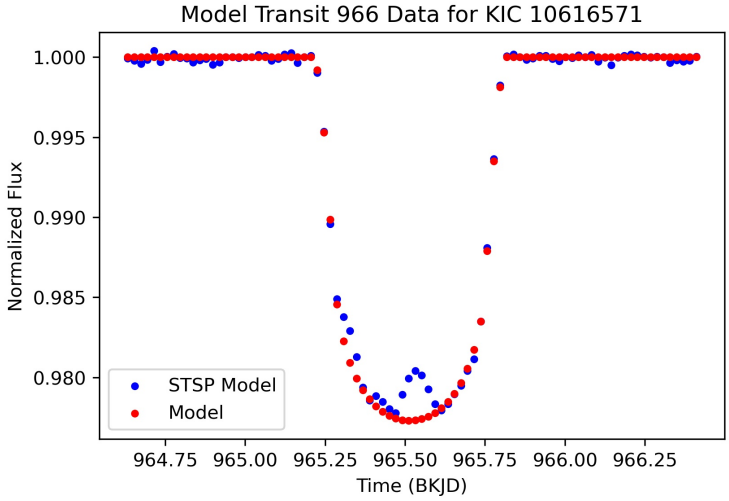


Figure 3.9: STSP model lightcurve with 2.5x sunspot-size spots on KIC 10616571. Note the upwards bump at the bottom of the transit. Also note the agreement with batman model flux plotted on top.

3.5 STarSPot: STSP

While STSP is its own program which would require a whole other paper to describe, here, I will highlight some of the basics of its functioning that allows us to produce lightcurve-matching simulated data. This program permits the user to create model lightcurves with only an in-file which supplies a well-defined lightcurve, orbital parameters, and starspot parameters (Hebb et al. in prep). The orbital parameters that are needed–and which can be found in the NASA exoplanet archive–are: the epoch of the midpoint of the transit (T_0), the orbital period (P_0), the orbital inclination angle (i_0), the planet to star radius (R_p/R_s), and the mean stellar density, which can be calculated to be equivalent to the dimensionless stellar radius (R_s/a). Other additional though not required parameters are the orbital eccentricity (e), the argument of periastron (ω), the projected spin-orbit angle (λ) and the tilt of the star's spin-axis (θ_s). All these parameters allow the calculation of the position of the planet within its orbit as a function of time. This location then determines the amount of light blocked from the planet from its projection onto its host star.

Instead of modeling the complete and complex morphologies of starspots–which include darker umbras and lighter penumbras–all spots are defined by their brightness contrast to the star's photosphere, that is, $c = \frac{I_{spot}}{I_{phot}}$, where I defines the regions' respective intensities. When this value is chosen–and we generally use a contrast of 0.7 to match that of a typical sunspot (Morris et al. 2017)– the brightness of the starspot is able to be calculated for proper

photometric simulations. More discussion on this contrast number will come in the results section as it will be important to fully understand the findings of this work.

The primary star in the program operates as a rotating sphere of unified brightness adjusted for limb darkening and the creation of N_s starspots on its surface. The limb darkening effect is recreated by overlapping concentric filled disks of increasing brightness (Hebb et al. in prep). This is shown below (Fig 3.10), where two methods are outlined–one in which the radii increase by a constant increment and the other where the radii increase by a constant increment in the brightness. In the creation of this program, it was found that $N = 40$ was a sufficient number of concentric circles to accurately model the basic transit shape given known limb-darkening coefficients (Hebb et al. in prep). While the program is created with a 4-coefficient limb-darkening, through algebraic conversion, we are able to adjust this law to linear, quadratic, or square-root functions and properly model the lightcurves given by Kepler (Hebb et al. in prep).

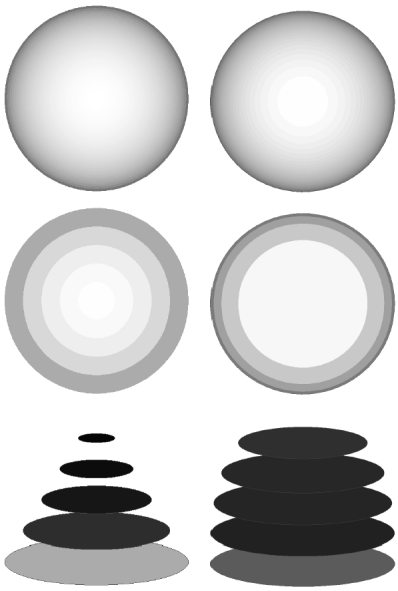


Figure 3.10: A demonstration of how STSP recreates limb darkening by the overlapping of concentric circles of increasing brightness (darker circles are brighter in this case. On the left is shown how the model looks if radii are chosen at a consecutive difference in length for each circle. The right shows how the model looks if the radii are chosen at consecutive and constant brightness increments for each circle. The top row shows each process with $N_s = 40$; the middle row shows each process when $N_s = 5$; and the last row shows the overlapping concentric circles of brightness. Image from Hebb et al. in prep

Knowing that the transiting planet will occlude an area of light on the star equal to its projection on the star, as well as the brightness for starspots and portions of the star previously mentioned through limb-darkening, one can finally calculate the total brightness by

$$B_j^{'} = B_j * [A_j - O_{p_j} - (\sum_{i=1}^{N_s} O_{sji}) * c]. \tag{3.5}$$

where B_j is the brightness of the disk of radius, r_j, $A_j = \pi r_j^2$ is the area of the disk, O_{pj} is the intersection of the jth disk and the planet, O_{sji} is the intersection of the jth disk with the ith spot, and c is the spot contrast (Hebb et al. in prep). The total brightness B is thus the sum of the brightness of each concentric disk, $B = \sum_{j=1}^{N_d} B_j^{'}$. The exact contributions of each portion is determined through calculations of the intersections of the spot and the planet on a 2-dimensional plane. This process is shown below (Fig 3.11) in an approximate sense, where each concentric circle is displayed at its brightness with various crossing features that form shapes of affected flux (Hebb et al. in prep). These areas can be either circular or elliptical, or consist of arcs cutting off these shapes. Through such contributions at different times–which are calculated by the parameters given to the program–STSP is able to simulate a lightcurve for the system Hebb et al. in prep.

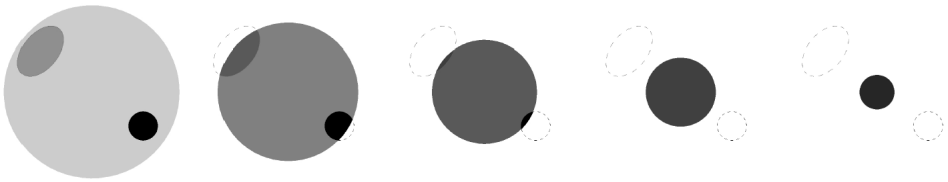


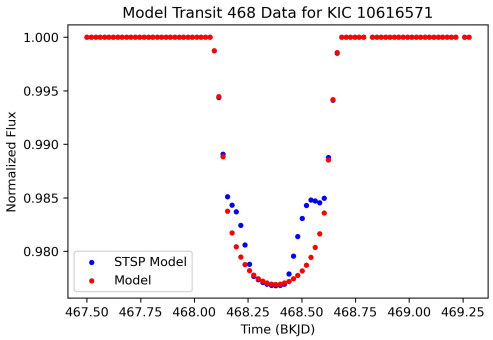
Figure 3.11: A demonstration of how total brightness is calculated by the STSP program. Here, the brightness of each concentric circle is added up including the flux-affected areas produced by crossings of darker-limbs, starspots, planets. Image from Hebb et al. in prep

3.6 Starspot Program 3: Simulation

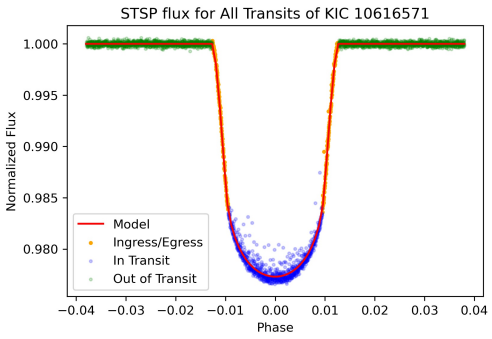
With an STSP-produced lightcurve, we can now revisit the aforementioned Starspot Program 3, which offers the functionality of providing a similar analysis on generated data as was conducted on the Kepler data. As before, this program takes the data given to it (in this case, simulated) and divides them all by their model "idealized" flux created by the batman model to highlight in-transit vs. out-of-transit scattering. Then, it creates a residual graph, where ingress/egress, in-transit, and out-of-transit are separated. Finally, it calculates the statistical difference to understand how well it relates to the actual data.

For the sake of brevity, a series of images depicting this process with the simulated data is shown below (Fig 3.12). For the most part, the process of analysis is the exact same for the simulated data as it is for the observed data, and so the general process formerly outlined also applies for this process of data analysis.

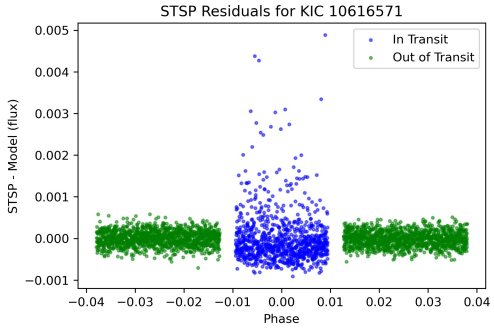
There are two distinctions, however, that though not displayed easily on the graph (Fig 3.12), are important in the proper statistical analysis of the simulated data. The first is



(a) Extracted simulated transit with STSP showing crossing with no simulated noise added.



(b) Simulated data for each transit overlaid with simulated noise added. Note the variation of flux in and out of transit due to noise not seen in figure (a).



(c) Residuals for simulated data with noise added.

Figure 3.12: Process of simulating spots on a lightcurve and finding residuals, similar to that outlined in section "3.3 Starspot Program 3: Data" except with starspots simulated onto the model lightcurve.

adding noise to the simulation in order to replicate the level of noise variation inherent in the data. This is created by making a normal distribution from the standard deviation of the out-of-transit Kepler data. Random selections from this distribution are then added onto all datapoints to reflect a similar statistical noise apparent in the actual Kepler satellite information.

The second difference is adjustments to the $R_p/R_\star$ parameter. The reason for this adjustment is a bit more complicated, and it is necessary to understand how this parameter ($R_p/R_\star$) is initially decided upon by the Kepler team. When lightcurves are analyzed, the $R_p/R_\star$ parameter is found by taking the average depth of the points within the transit. For a large scatter transit, such is found with this object, this parameter becomes slightly adjusted to account for the inherent variation due to starspots themselves. Thus, the $R_p/R_\star$ that we use as a parameter for the simulation *already* is taking into account the presence of starspot crossings. That is to say, it is not the idealized, zero-crossing $R_p/R_\star$. The effect of this result is apparent in the statistical analysis (later explained) whereby the model

lightcurve was always below the simulated lightcurve with spot crossings, which affected the statistical outcome of the KS Test. To adjust for this, I run the batman model, which is that model used in the isolation of the residuals from both the data and the simulation, with various adjustments in $R_p/R_\star$. The precise number of adjustments is 6000 between $R_p/R_\star \pm \frac{1}{20}R_p/R_\star$. For each of these trials, I perform a chi-squared test–which quantifies how closely two data sets match–between the simulated data with starspots and the batman model data with the adjusted parameter. We then pick out the adjusted parameter with the lowest chi-squared, that is, the model data that best fits the simulated data when accounting now for the starspots. While this is nearly always close to the actual value of the parameter as obtained by the exoplanet archive, when starspot crossings become significant, these adjustments are necessary and useful to not double account for spot crossing and provide a better statistical picture of what is happening.

3.7 Statistical Analysis

3.7.1 Cumulative Sum Distribution (CDF)

The final function that is performed in program 3 is a short statistical analysis of the starspot crossings, quantifying the scatter within the transit to that without. The first step in measuring this scatter variation is performing a cumulative sum distribution. The cumulative sum distribution is defined as

$$F(x) = P(X \leq x), \tag{3.6}$$

where $F(x)$ represents the probability of being less than or equal to a given observed x from the probability density function $f(x)$. To give an example of what this means, we can turn our attention to the classic six-sided die. There are only six options as to what a a roll could be (1,2,3...). As defined by our cumulative sum distribution (CDF), we infer that because the least value possible on the dice is one, there is only a 1/6 chance that the value rolled may be below or equal to one. For six, however, there is a 100% chance that the value will be six or below. Thus, our CDF can be represented as a graph starting from 0 (no probability of a number below one) to 1 (a guaranteed probability that there is a number below six) where the probability density function $f(x)$ increases by $\frac{1}{6}$ for each increment from 1 to 6.

This is relevant in quantifying our compared scatters because in essence it demonstrates numerically the amount of "scatter", or variation. That is to say, if there is a large range of possible values in the residuals, such that there is in in-transit scatters of crossed starspots, the cumulative sum distribution will less sharply arrive at 1. The case is opposite for the out of transit section. Such an example is shown in Fig 3.13.

Again, as with all of Starspot Program 3, this function is also reproducible with simulated data. Below is a graph of one of these cumulative sum distribution for sun-sized starspots (Fig 3.14). The coding of this portion involved simply graphing the criteria previously

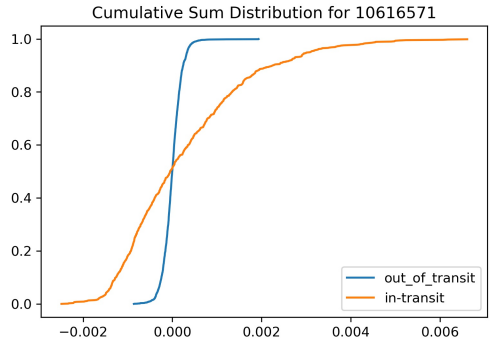


Figure 3.13: The cumulative sum distribution for the Kepler data of object 10616571 (i.e. not simulated data). Note the difference between the in-transit and out-of-transit line and how quickly they approach 0. The longer approach to zero signifies greater residuals and thus starspot crossings. Keep in mind that the x-axis represents the residual number whereas the y-axis represents the fraction of spots below that residual number.

displayed describing the CDF. It is important to note that though not displayed, the axes of these CDF graphs will always have the residual number on the x-axis and the fraction of spots below that residual number on the y-axis .

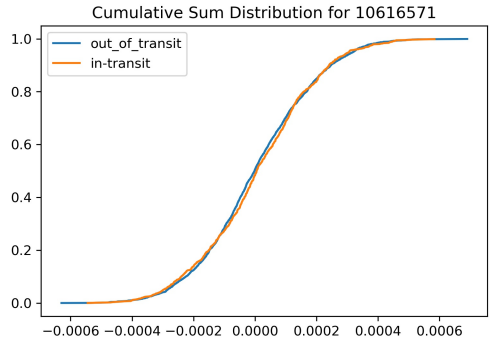


Figure 3.14: The cumulative sum distribution for the STSP data of object 10616571 (i.e. simulated data). Note the lack of difference between the in-transit and out-of-transit line for Sun-sized starspots. This signifies that the residuals in and out of transit for Sun-sized starspots are indistinguishable. Keep in mind that the x-axis represents the residual number whereas the y-axis represents the fraction of spots below that residual number.

3.7.2 Kolmogorov-Smirnov Test

However, the cumulative sum distribution is prone to its own systematic errors. For example, when working over small data sets which often occur within the Kepler Satellite archive–transits absent of points or with only a few points–the test can be heavily bias towards quantifying high scatter, when in reality that scatter may be only created by a few

systematic outliers. Thus, as was noted in Michelle's thesis, it can often happen where the CDF shows what may be considered a starspot crossing on an object when in reality the residuals themselves clearly do not show any crossing. Rather, they show a lack of data. This requires a further quantification and statistical check named the Kolmogorov-Smirnov Test.

This test essentially looks at the maximum distance between the two distribution functions. This is to quantify whether they belong to the same distribution or not, and if not to determine what quantity we can attach to their dissimilarity. A number of 0 means that the two distributions are practically from the same set, and a number of 1 means that they are not at all. The test values can range between these two numbers (Wall and Jenkins 2012).

Unlike the chi-squared test, the Kolmogorov-Smirnov Test does not necessitate a large sample size. Rather, it is independent of this factor and only relative to the distributions involved (Wall and Jenkins 2012). Thus, it is preferred for the range of sample sizes as the variation in data for each Kepler object is not a factor in the varying Kolmogorov-Smirnov Test numbers.

3.7.3 Visualizing The Spotted Surface of Stars

The final component of the code, which I place here under statistical analysis, is the mapping of the starspots crossed along with the starspots simulated. This component was a check that the simulated lightcurves were creating starspots in the correct locations and properly. Without this check there was the possible risk that low/high scatter simulations were numerical artifacts of the code rather than the data itself. Additionally, the mapping of the starspots permitted the understanding of what my code is visually representing, particularly when variables have been changed (as will become important in the results section), again providing not only checks to the black-box code but also comprehension as to the physical characteristics simulated. With this, the maps provide an understanding of what the rudimentary superficial features are on other stars. They recreate the same solar phenomena that have fascinated and been integrated into the cosmic belief of people's around this world. In this way may we begin to similarly understand, familiarize, and connect with some of these other Sun-like stars–many with their own planets–in our galaxy.

In order to get these maps, I identify the spots crossed in each simulated transit. STSP records the spots that are crossed by the planet at each point in the lightcurve using a single indicator for each point. If no spots are being crossed by the planet at that time, the indicator is zero. To ensure that STSP can record more than one spot crossing at a single time, it stores the information as the sum of every spot crossed in a simulation, where each spot represents a power of 2 (i.e. 2^0, 2^1, 2^2, 2^3, 2^4...). For example, a number displayed as 5, holds within it the information of two spots crossed, that is spot 2 and spot 0, given that $2^2 + 2^0 = 5$. Because we simulate 10 to 16 spots in each trial–the previous and later limits as to what can be simulated in STSP for each transit–these exponents and thus spot numbers can range from 0 to 15.

After recording which spot numbers are crossed in each transit by basis of this provided

"cross number", I cross-reference these transit-specific spots with the previously stored data regarding spot radius, latitude, longitude, etc. The total spots simulated are then recorded in one data file, while the total spots crossed are recorded in another. This all happens in Starspot Program 2.

This information is then read into the mapping feature of Starspot Program 3, the same program which produces the residuals. I use the *cartopy* graphing package (an open access Python package designed for mapping features onto Earth-like circular plots) to create a circular plot, changing the radius involved so that it corresponds to units of Earth radius, as used by *cartopy* (Met Office 2010 - 2015). This involves multiplying the radius ratio by 6371 km. I then mark the latitude of transit by means of the impact parameter (Equation 2.19) along with the limits of occultation of the transit. These limits are calculated by adding or subtracting the ratio of the planet-star radii from the impact parameter as such

$$\text{Limits} = \arcsin\left(\text{b} \pm \frac{\text{R}_\text{p}}{\text{R}_\text{s}} \times \frac{180}{\pi}\right). \tag{3.7}$$

To depict the front-side, the longitude of display was set at 0°, and for the backside, this was set at 180°. The spot data was then able to be graphed onto the circular plot resulting in starspot maps such as those below (Fig. 3.15).

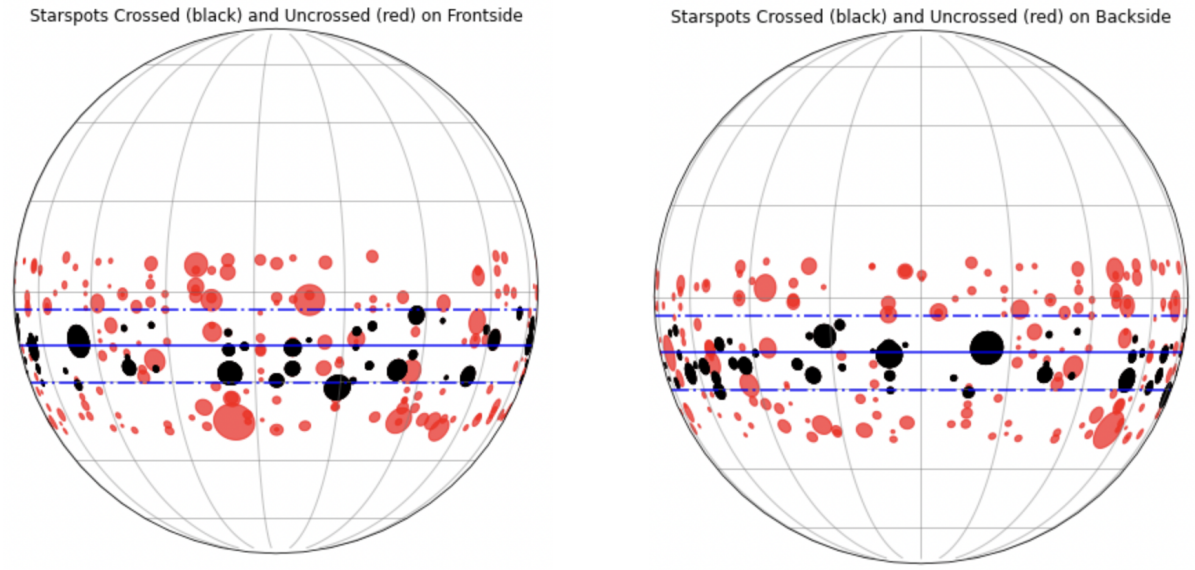


Figure 3.15: Sample transit maps for the frontside (left) and backside (right) of the KIC 10616571 with sun-sized starspots. Spots are spread out over 100 sections from 0 to 2/pi. Spots that are crossed by the planet are colored in red, whereas those that are simulated, but not crossed by the planet are in black.

It is important to note here again that the simulations conducted in this work as well as those maps which depict the simulations demonstrate a statistical understanding of what

the starspot distribution may look like on other stars. It is not a unique solution or the full solution, but it provides a greater understanding towards the general magnetic features that these stars have that otherwise would have been inaccessible to us on Earth.

Chapter 4

Results

4.1 Degeneracy

Before looking at the results of the starspot programs on KIC 1061656571 (that object which was used in the description of my code) and other objects, it is important first to note some crucial findings regarding the inherent degeneracy of this method of understanding starspot coverage on distant stars. When this program was completed and utilized, it was found that there was a degeneracy between the starspot spot contrast (0-1) and the starspot size increase from sunspot sizes. This makes sense intuitively, given that starspot contrast describes the fractional brightness of the spot compared to the star itself. For example, if the spot was given only a contrast of 0.1 this would mean that the spot is only 10% of the brightness of the photosphere. Likewise, when the satellite is not transiting, this incredibly dim spot would contribute to the net brightness of the star. Therefore, when the planet *is* transiting in front of this spot, there would be a fractional blockage of an area that is *90%* less bright than the star. That is to say, if we are looking at the residual graph, we would expect to see an in-transit scatter that is nearly reaching out-of-transit level, given that the light variation is approximately on the level of that if the planet was not transiting at all (which would be the case at a contrast of 0.)

This variation is similar to the effect that one would expect if the sizes of the starspots were increased significantly. By increasing the size of the starspot, we find longer amounts of time where larger portions of the transiting satellite is passing over spotted areas. This means that the brightness is more consistently and more dramatically increasing within the transit. Note that these two parameters do not appear one in the same completely: for example, a high contrast but small spot will not cover the same amount of time-coverage as a larger but lower contrast spot. In this way one can slightly identify what is causing the variation; however, picking apart this degeneracy completely would require a time accuracy that we do not always have with the Kepler information, as well as an in-depth knowledge of exactly what the dividing line is between the intensity of the variation and the time duration of the variation. All that being said, the following analyses will attempt to provide the most holistic picture of the spotted features while taking into consideration this caveat. In the

future, however, with more information on realistic sizes and contrasts of starspots on other stars, this program can easily be used again to provide more accurate pictures of said Kepler objects.

The last reminder before talking about each individual object is that the simulation we take as our "normal" consists of a starspot contrast of 0.7, as is found in the limited literature (Morris et al. 2017); starspot sizes comparable to the Sun; and a starspot number of 10 for KIC 1061656571–the prior STSP limit at the time when this thesis was first being written– and 16 for the others. In both cases, the spots are placed at equal longitudes around the star with a randomized starting longitude. Our "control" that we are comparing to, by means of graphs (residuals and cumulative sum distribution) and statistical analysis (KS test), is the actual data taken by Kepler.

4.2 KIC 1061656571

The data as taken by Kepler for KIC 1061656571 is reproduced below (Fig 4.1). The KS test calculated from the cumulative sum distribution was 0.3808. This number and the graphs shown are the goal this section hopes to replicate in intensity, frequency, and statistic with known parameters.

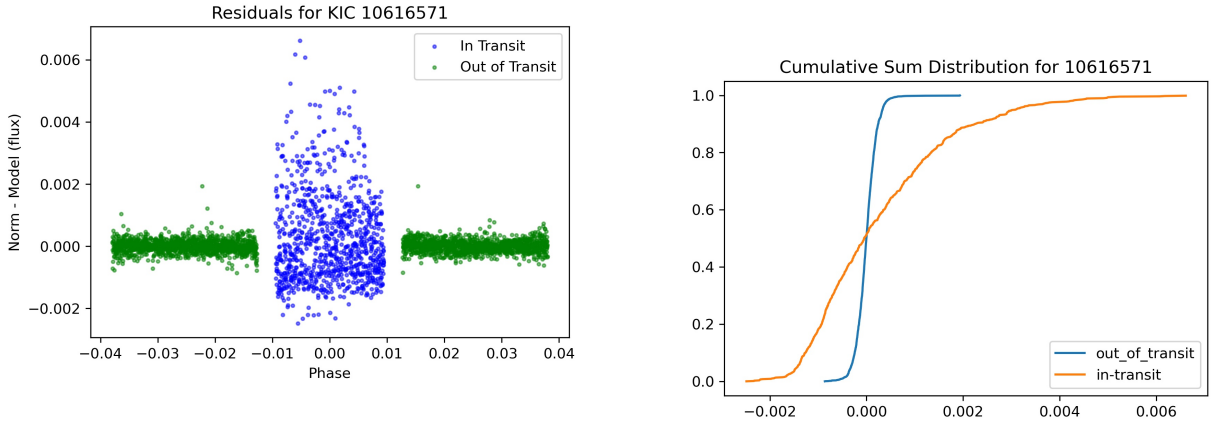


Figure 4.1: Residuals (left) and cumulative sum distribution (right) for the Kepler data of KIC 1061656571. The amplitude of the in-transit variation is 0.006.

When we compare this data with the simulation of KIC 1061656571 using (1) starspots the size of sunspots, (2) a contrast of 0.7, and (3) a maximum of 10 spots, we obtain the results in Fig 4.2. The KS number calculated from this simulation was 0.0525. Note that this number will change with each simulation as each iteration regenerates randomized parameters (according to the distributions previously mentioned.)

To get an understanding of what this system is like visually, the maps of these simulated starspots over all transits are reproduced in Fig 4.3.

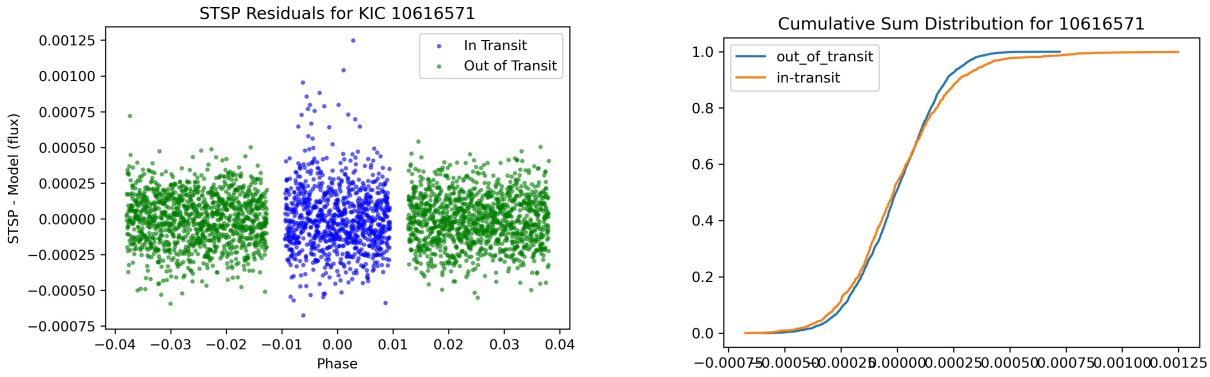
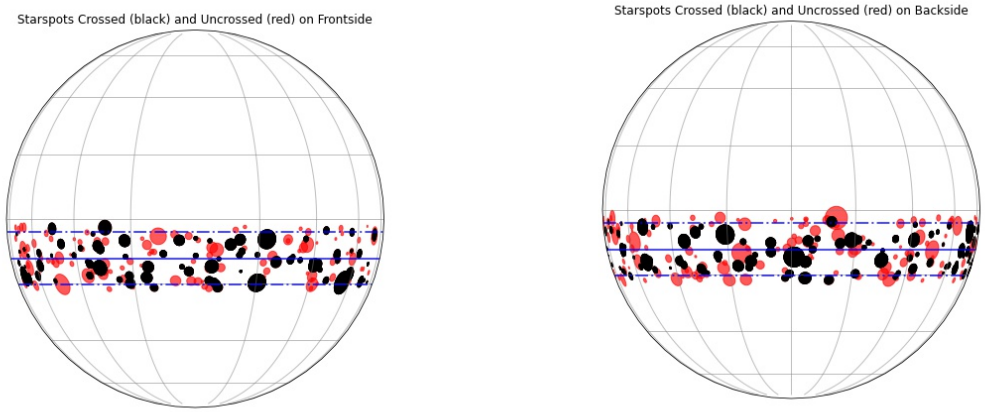


Figure 4.2: Residuals (left) and cumulative sum distribution (right) for the simulated data of KIC 1061656571 with "normal" parameters (10 spots, 0.7 contrast, and sunspot sizes). The amplitude of the in-transit variation is 0.00125, much less than the real data.



(a)

Figure 4.3: Simulated starspots on KIC 1061656571 on the frontside (left) and backside (right) of the star with "normal" parameters (10 spots, 0.7 contrast, and sunspot sizes.) Red represents those spots that went uncrossed, whereas black represents those that were crossed.

It is clear from the scale of the variations of the residuals, as well as the KS test and cumulative sum distribution that these "normal" parameters are not adequate to reproduce what is observed for KIC 1061656571. The data undergoes a peak in-transit variation of ~ 0.00600, whereas this simulation undergoes an in-transit variation of only 0.00125. The CDF of the two datasets also shows the dissimilarity in variation, ultimately quantified by the large difference in KS numbers. From this alone, we can determine that the spots on KIC 1061656571 are either larger or higher contrast than what we simulated with, or possibly both.

The next question is then what combination of starspot number, spot size, and contrast provides the most optimal representation of the spotted area of the star? To do this, I designed another code to perform and apply the KS analysis to simulated lightcurves with varying contrasts and spot sizes, with the spot number kept constant. The KS test for each combination of size and contrast was performed 5 times to minimize error (though there is the functionality of adding more tests to further reduce the error) and error bars are plotted along with the points themselves. This graph is shown below (Fig 4.4.)

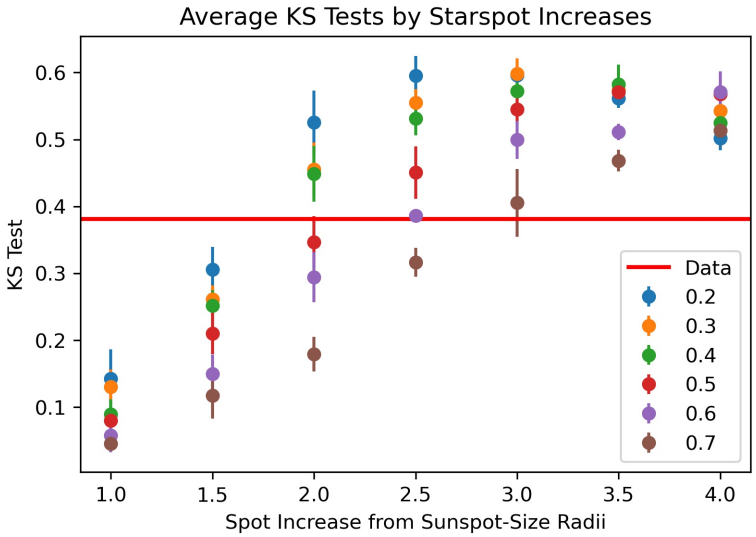


Figure 4.4: KS test analysis of different starspot size increases (from sunspot size) and varying contrasts. The contrasts are grouped by color, whereas the size appears on the x-axis. Note how the graph provides an idea of optimal options for the parameters that could have created the data seen by Kepler. Also note the degeneracy previously mentioned between spot contrast and spot size.

The general features of such a KS analysis (Fig 4.4) demonstrates that no matter the contrast of the features, the starspot sizes are larger than that of the sunspot features by at least 2 times. The exact details of this particular increase varies, as shown, by chosen contrast. Nonetheless, we may assume that within error the starspot sizes are ~ 2.5 times that of sunspot sizes at a contrast of ~ 0.6.

Representing this combination using the aforementioned codes we obtain the graphs below (Fig 4.5.)

The last notable finding of the information gathered from the simulation is that we are lacking in starspot number. Looking at the actual data of this stellar object (Fig. 4.1), one can see that there are many more points of variation within the transit than in the simulation (Fig. 4.5.) This results in the smoother shape of the CDF in comparison to the CDF produced from the simulated data, which has a sharp increase–points with no crossings–then a tall and long tail–crossed spots. This tail derives from the fact that there are relatively few crossed spots within the transit and most of them tend to be of greater flux variation (less partial crossings, less small spot crossings, etc.) That is to say, with more

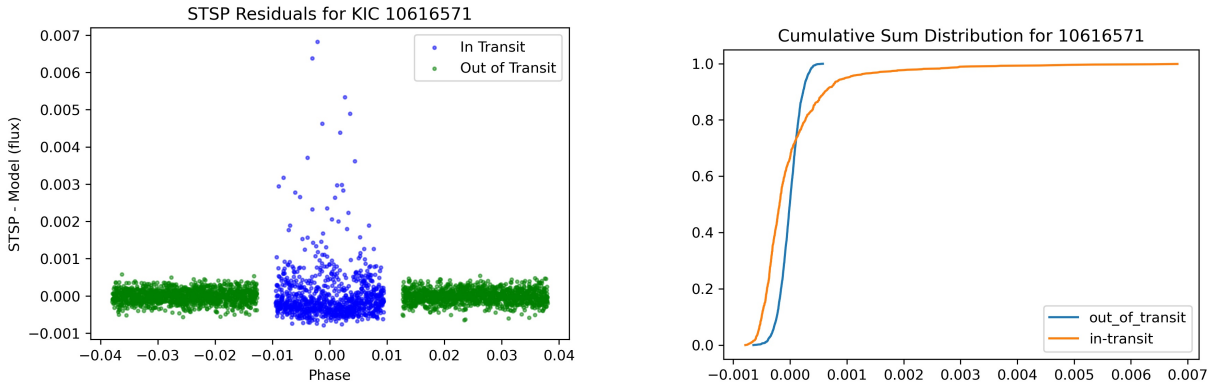


Figure 4.5: Residuals (left) and cumulative sum distribution (right) for the Simulated Data of KIC 1061656571 with "optimal" parameters (10 spots, 0.6 contrast, and 2.5x sunspot sizes.)

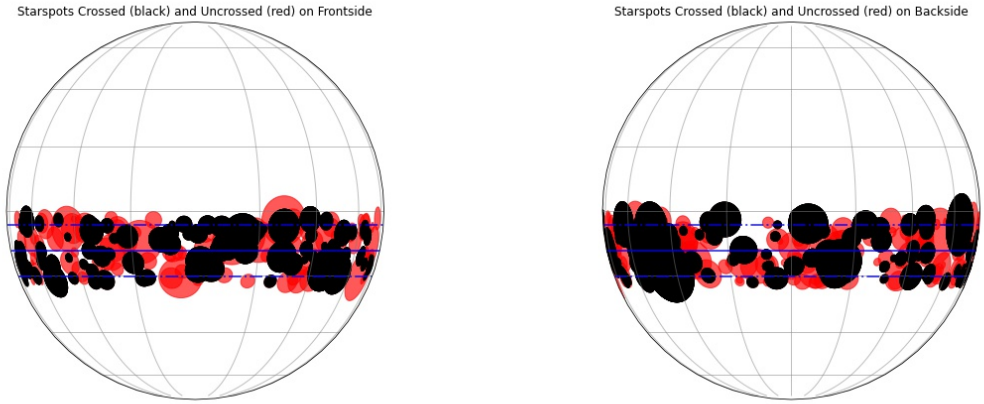
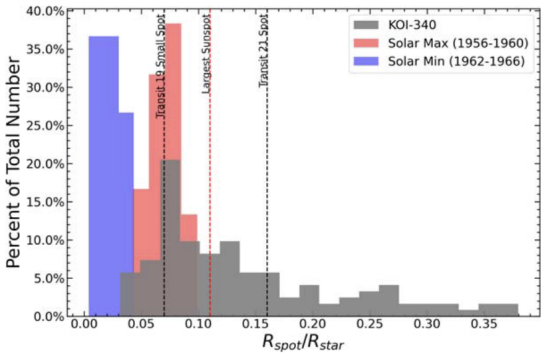


Figure 4.6: Simulated starspots on KIC 1061656571 on the frontside (left) and backside (right) of the star with "optimal" parameters (10 spots, .6 contrast, and 2.5x sunspot sizes)

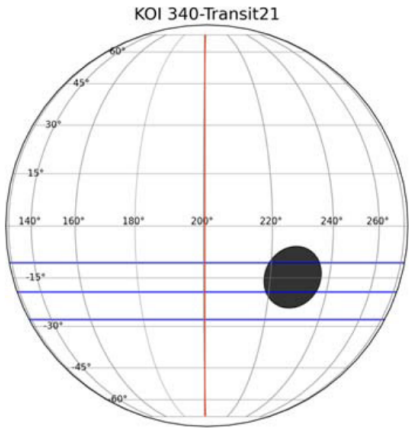
spots, there would be a larger sample of crossed spots, both partial and smaller, that would smooth out the simulated curve. Thus, a further iteration of this program would repeat this analysis with more starspots, which would most likely provide a better picture of the nature of this faraway star. Nonetheless, we may definitely say two things about this star: the spots are greater than sunspot sizes, at ~ 2.5 times, and there is a greater number of spots than the simulated 10 spots located around the star. We may say that it is quite a spotty star, at least compared to our own pock-marked Sun!

The next question is whether this analysis holds up to what may already be understood about the clearly spotted nature of KIC 1061656571. Does this finding agree with more in-depth, curated analyses of this star? According to Maria Schutte, who has previously

performed an in-depth analysis on this object using STSP, the findings of our code agrees. Using many more trials than I have used, more in-depth statistics, degeneracies regarding longitudinal effects on flux occultation, and considerations for artificially-large spots simulated by STSP, Schutte found on average that half of the starspots on this Kepler object were around 3.4 times larger than that of the Sun's (Schutte et al. 2022) with a contrast of 0.7. In a further analysis, she found that this number was more likely around 2.6 (still with the contrast of 0.7), with room for error (Schutte et al. 2022). While she did not focus on changes in spot number, the findings of the spot radii are shown below as compared to historical spots on the Sun.



(a) Distribution of starspot radii on KIC 1061656571 as found through Schutte's detailed analysis of spot crossing features. The dotted vertical lines refer to notable starspot sizes, where the furthest to the left is one of the smallest simulated radii found, the furthest to the right is an average radii simulated on the star, and the red line is the largest sunspot found on the Sun by Newton.



(b) Transit 21 simulated spot mapped. This spot represents the typical spot size found by Schutte to represent the data found in KIC 1061656571. This is also the furthest dotted black line depicted on the graph to the left. Note the similarity in size of Schutte's spot against the spot found with this thesis's code.

Comparing Schutte's findings with that found by the elaborated program, we can confirm both the success in the operation of the program and the likelihood that what is being found is scientifically sound and even more important, likely. Even though we settle on a contrast of 0.6, when considering a contrast of 0.7 like Shutte did, we reproduce similar results to that of Schutte. This is particularly so when we consider the need for greater spot numbers, which would likely increase the scatter to this level.

4.3 KIC 8672910

Here, the data as taken by Kepler for KIC 8672910 is reproduced (Fig 4.8.) The KS test calculated from the cumulative sum distribution was 0.256. This number and the graphs

shown, as with the last KIC object, are what this section seeks to replicate in intensity, frequency, and statistic with known parameters.

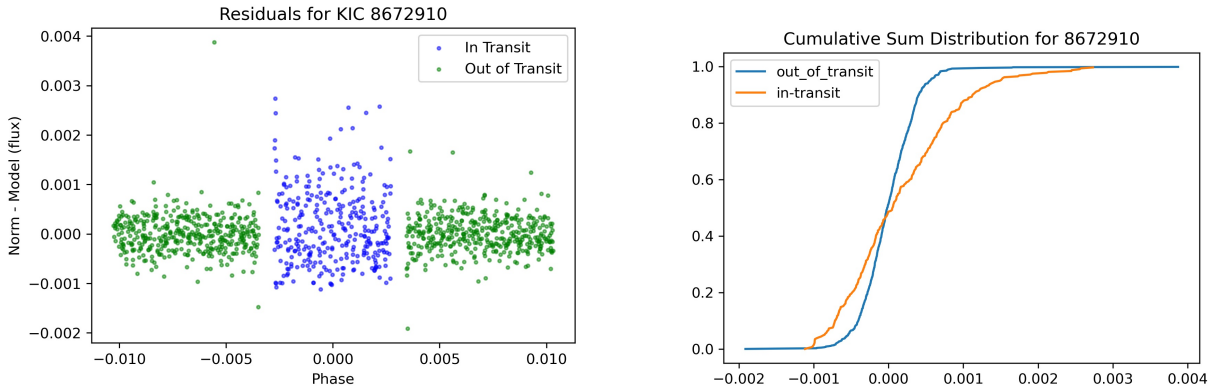


Figure 4.8: Residuals (left) and cumulative sum distribution (right) for Kepler data of KIC 8672910.

Following the analysis conducted for the previous KIC object, I display the simulated Sun-like starspots onto the system (Fig 4.9.)

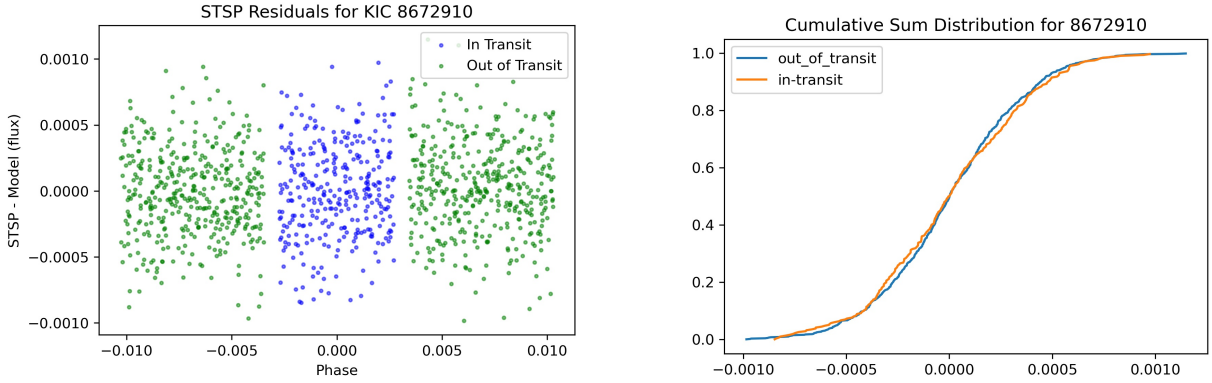


Figure 4.9: Residuals (left) and cumulative sum distribution (right) for the simulated Data of KIC 8672910 with Sun-like parameters but maximum spots (16 spots, 0.7 contrast, and sunspot size.)

Even with more limited data than KIC 10616571, it is seen from this preliminary simulation that Sun-like starspots cannot account for the variation that is seen due to the spotted features on KIC 8672910. This is supported by the difference in the intensity of the variation, as well as the largely different KS numbers (0.045 in this simulated case). Thus, similar to KIC 10616571, KIC 8672910 is also more spotted than our Sun, with both likely more spots and larger spots than what can be found locally.

To understand what the optimal combination of starspot size, contrast, and number is, I reach back to my master code to make a figure which compares these factors. The results are

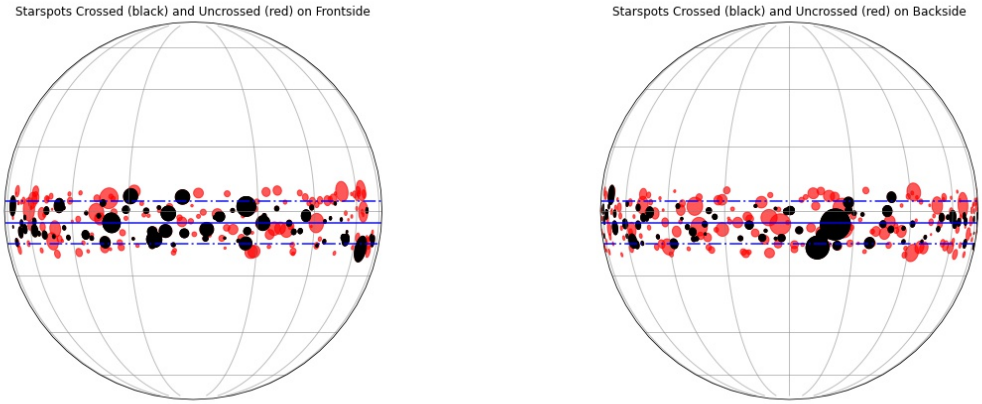


Figure 4.10: Simulated starspots on KIC 8672910 on the frontside (left) and backside (right) of a star with Sun-like parameters but maximum spots (16 spots, 0.7 contrast, and sunspot sizes.)

displayed, following the same process that employed on KIC 10616571 (that is, five trials for each average KS number, with uncertainties shown.) Fig 4.11 shows the work for 16 spots placed randomly around the star, the updated limit to the previous 10 spots to which STSP used to be limited.

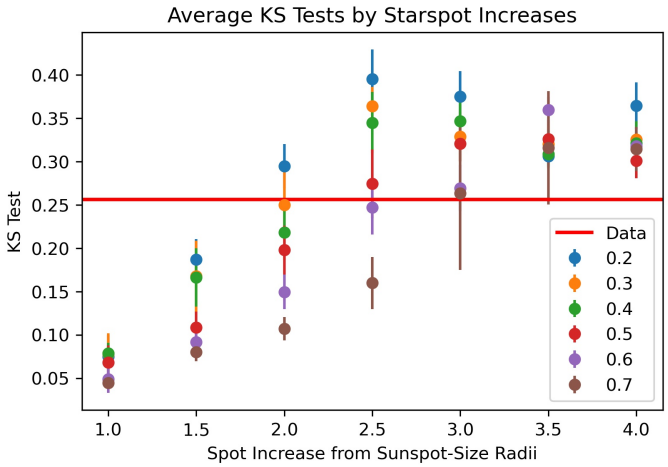


Figure 4.11: KS test analysis of different starspot size increases (from sunspot size) and varying contrasts. The contrasts are grouped by color, whereas the size appears on the x-axis. Note how the graph provides an idea of optimal options for the parameters that could have created the data seen by Kepler. Also note the degeneracy previously mentioned between spot contrast and spot size.

As shown in Fig 4.11, one can see that no matter the different contrasts of the spots, the spots on KIC 8672910 are much larger than sunspot features, in this case by roughly

2.5-3 times. While further work would again be needed to flesh out the exact details within this spotted system, we can say approximately that these are the characteristics of the size distribution of KIC 8672910. Inspecting these characteristics visually, we observe the graphs below.

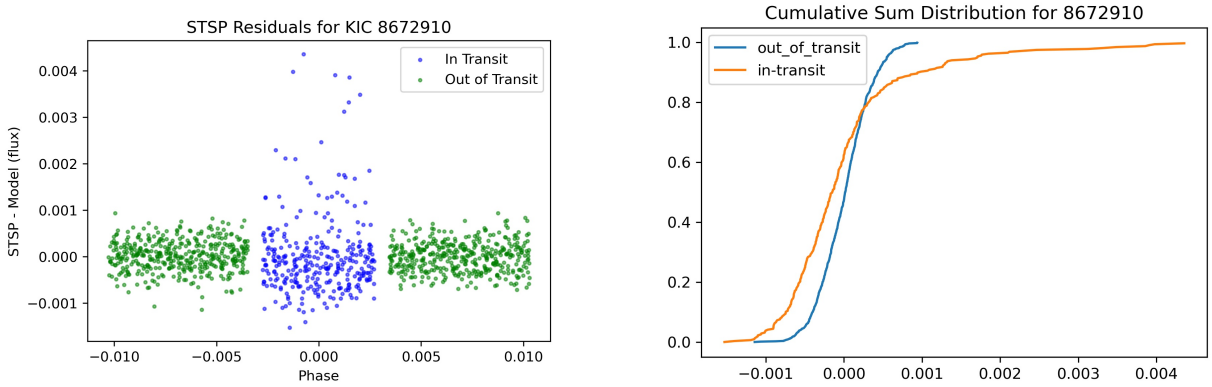


Figure 4.12: Residuals (left) and cumulative sum distribution (right) for the simulated data of KIC 8672910 with "optimal" parameters (16 spots, 0.7 contrast, 2.75x sunspot sizes)

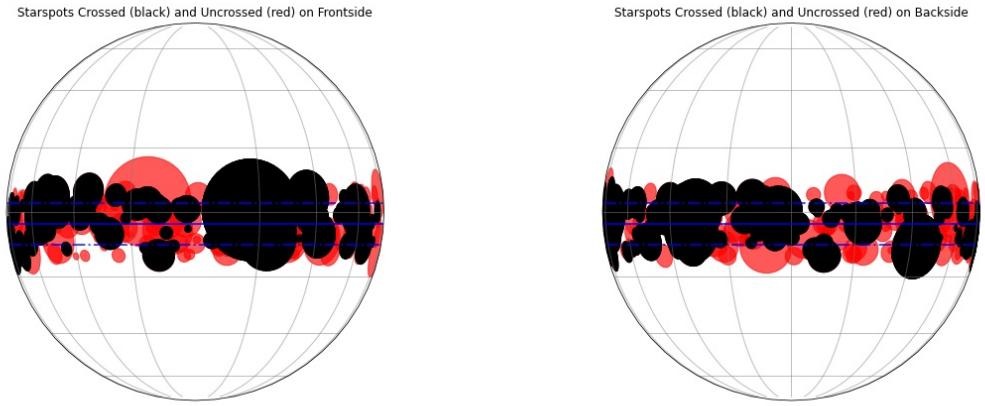


Figure 4.13: Simulated starspots on KIC 8672910 on the frontside (left) and on the backside (right) of the star with "optimal" parameters (16 spots, 0.7 contrast and, 2.75x sunspot sizes)

Comparing these graphs (Fig 4.12) with the data graphs for KIC 8672910 (Fig 4.8) demonstrates a remarkable similarity between the two. One only needs to look at the scale of variation and the shapes of the CDF's to validate this. Additionally, by using 16 spots rather than 10 spots with this KIC, it is observed that we have avoided the "tail" problem that still existed in the analysis of KIC 1061656571. The matching of these tails and the

shapes of the CDF and residual graphs demonstrate that this star has many more spots than 10 and that the spots that do exist are much larger than those on the Sun. Unlike with KIC 1061656571, there are no other in-depth studies that have been conducted on the spotted features of KIC 8672910, and so we cannot confirm whether this analysis matches other analyses of different codes/methods. However, we may say that this preliminary analysis that I have performed with my code provides perhaps one of the first insights into the characteristics of KIC 8672910's spotted surface. Given the accuracy in modeling starspots of KIC 1061656571, we may consider this model similarly accurate, and the first attempt to uniquely characterize such stellar features.

4.4 KIC 7767559

The data as taken by Kepler for KIC 7767559 is reproduced below, produced again by my Starspot Programs 1 and 3. The KS test is 0.274, and it is once more this number, as well as the intensity, frequency, and statistic of variation, that we seek to replicate in this analysis.

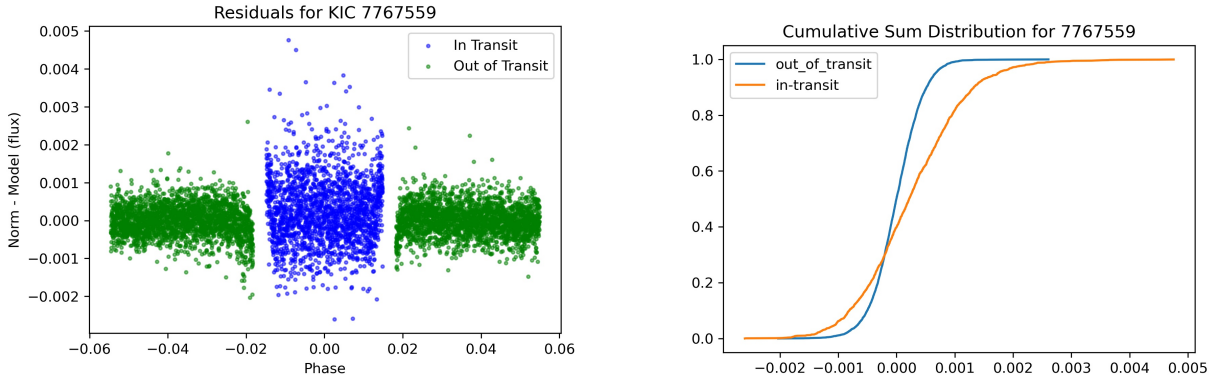


Figure 4.14: Residuals (left) and cumulative sum distribution (right) for Kepler data of KIC 7767559. Note the points still left in ingress/egress for the residuals.

Unlike the previous two objects, we already see that the residuals (Fig 4.14) do not perfectly represent the in-transit vs. out-of-transit portions of the data. Rather, it demonstrates that the criterion for selecting and omitting the points of ingress and egress have not fully captured those points. While in the extreme case, this can result in artificial statistics, here the scale is small enough and the variation in-transit is large enough for this to remain a caveat in the analysis instead of a total hindrance. This can be demonstrated by inspecting the cumulative sum distribution and noting that its shape is not incredible different from those other cumulative sum distributions previously shown.

KIC 7767559 is not alone in this problem; in fact such an effect occurs with various of the KIC objects. As the derivation of selecting ingress/egress points is purely geometrical and unlikely the problem, one can suppose then that the parameters used are not quite the

correct parameters for these systems or that their geometries are more complicated than what the parameters are able to capture. Thus, my program incidentally also works as a check of Kepler system parameters. In the case of KIC 7767559, the failure to capture all ingress/egress for omission can only result from a slightly imprecise inclination, planet-to-star radius, or planet-to star-distance as outlined in the discussion on transit lightcurves.

Following along in the analysis outlined with the previous two objects, below I reproduce the images of Sun-like starspots simulated on this system (Fig 4.15.)

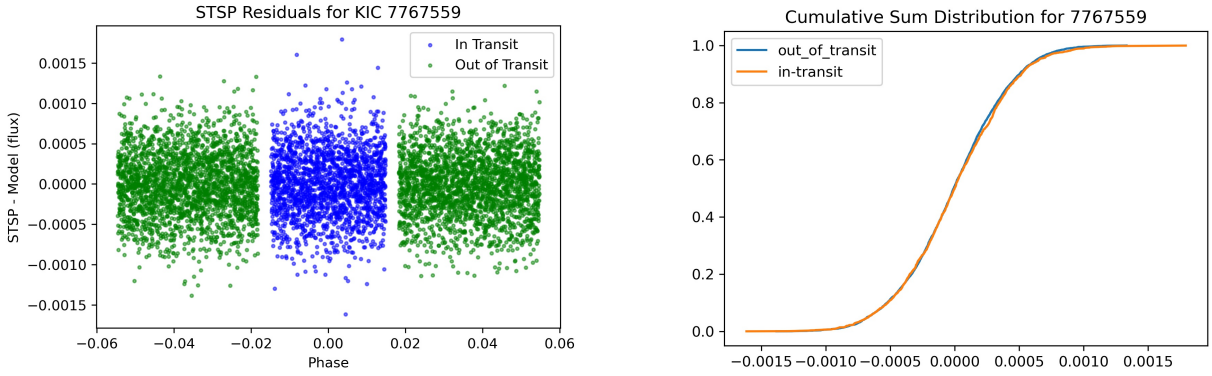


Figure 4.15: Residuals (left) and cumulative sum distribution (right) for the simulated Data of KIC 7767559 with Sun-like parameters but maximum spots (16 spots, 0.7 contrast, and sunspot size.) Note the proper emission of ingress/egress points.

Before displaying the maps, one feature to point out in the graphs above is the proper omission of ingress/egress points. This occurs because the transit that is being simulated is that which is *exactly* described by the parameters gathered from the Kepler exoplanet archive. For the previous two objects, these parameters were precise enough to simulate a lightcurve that exactly matched the data, while for this object, they are slightly off, leading to a simulated lightcurve that is slightly off from what is seen in the data. That is, we are simulating a lightcurve that would have perfectly described and omitted the ingress/egress points.

Again, as with the other displayed objects, we find that Sun-like starspots are inadequate to describe the in-transit variation/surface variation of this Kepler object. This result hints that a further analysis for optimal starspot parameters is required. This difference is quantified in the extremely different KS number of 0.027.

The graph of parameter optimization is now shown (Fig 4.17), as constructed with 16 spots placed randomly around the star.

Because of the slight over count of points that occurs in the data KS test due to the counting of ingress/egress points, we may assume that our KS data test is a slight *overestimate*. Likewise, let us choose the optimal parameters to be roughly 3× the size of sunspots with a 0.7 contrast. A system with these parameters produces the graphs and maps below (Fig 4.18 and Fig 4.19.)

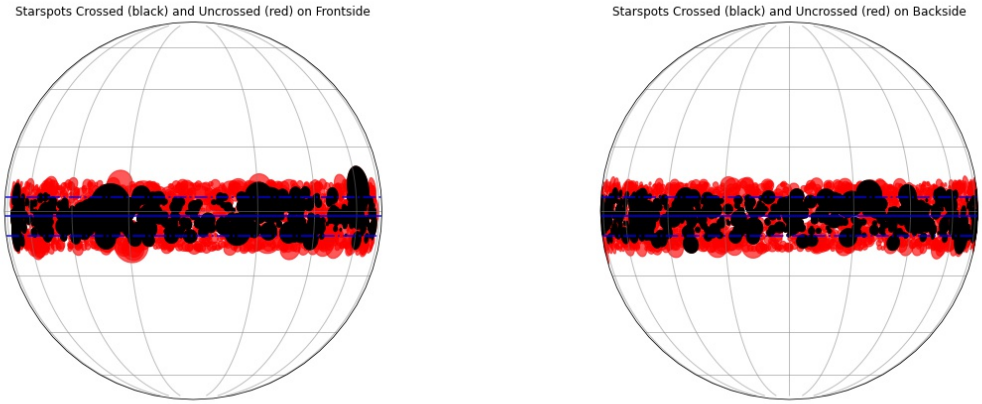


Figure 4.16: Simulated starspots on KIC 7767559 on the frontside (left) and backside (right) of a star with Sun-like parameters but maximum spots (16 spots, 0.7 contrast, and sunspot sizes.)

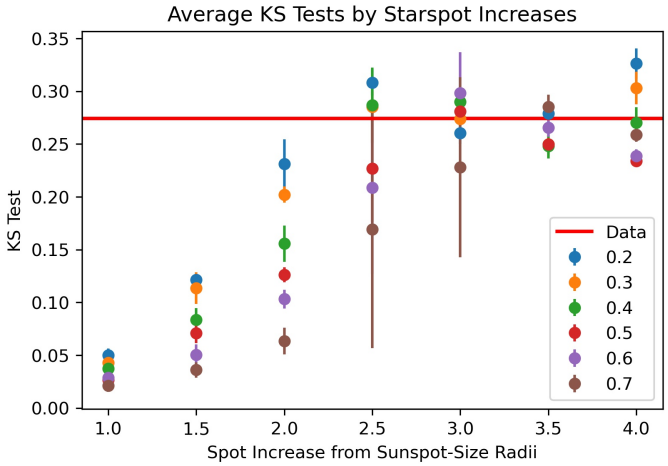


Figure 4.17: KS test analysis of different starspot size increases (from sunspot size) and varying contrasts. The contrasts are grouped by color, whereas the size appears on the x-axis. Note how the graph provides an idea of optimal options for the parameters that could have created the data seen by Kepler. Also note the degeneracy previously mentioned between spot contrast and spot size and the wide error bars for many points.

While, the KS test number may still be different from the data's KS test number, which takes into account parameter error, we can see that a starspot size of 3× sunspot sizes is still not a terrible estimation, given that the residuals graph both roughly matches the data (in frequency and intensity) as well as the cumulative sum distribution (in its general shape). In fact, it seems that the simulation slightly underestimated the starspot sizes, given that the intensity of the in-transit residual is just under 0.005 for the data and hovers around 0.004

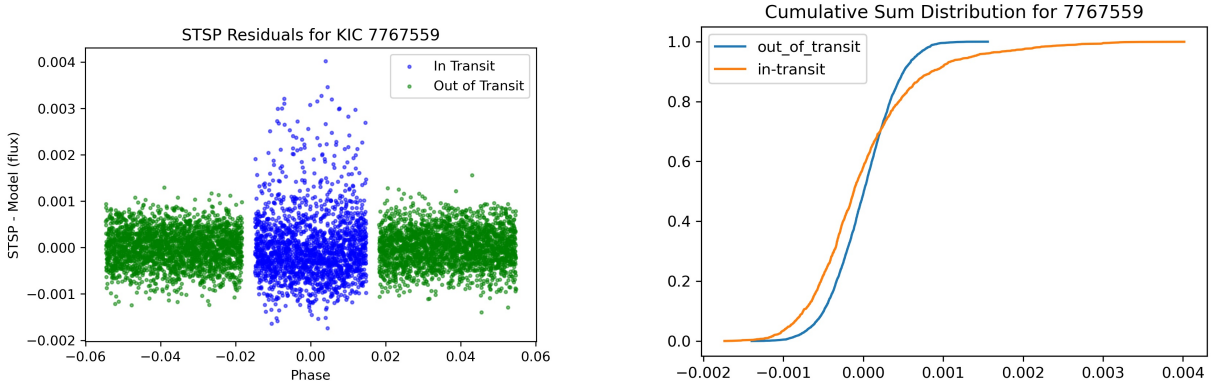


Figure 4.18: Residuals (left) and cumulative sum distribution (right) for the simulated data of KIC 7767559 with "optimal" parameters (16 spots, 0.7 contrast, 3x sunspot sizes)

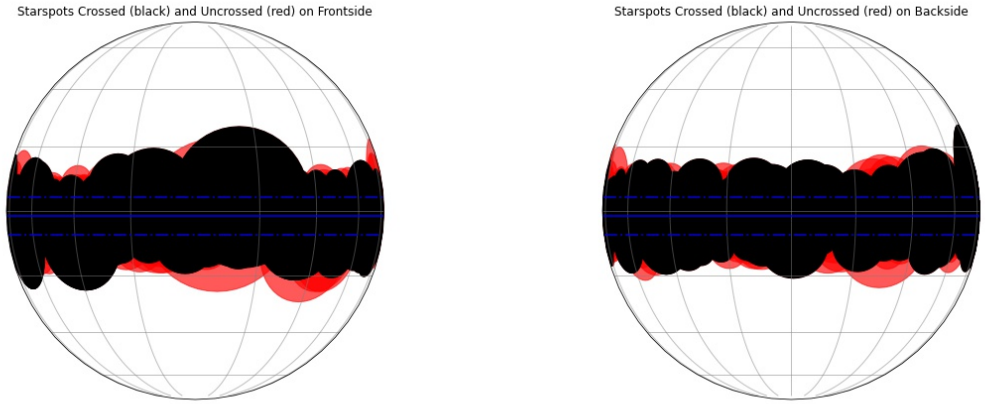


Figure 4.19: Simulated starspots on KIC 8672910 on the frontside (left) and on the backside (right) of the star with "optimal" parameters (16 spots, 0.7 contrast and, 3x sunspot sizes.) As these maps display the cumulative crossed/uncrossed spots for all the transits involved, in this Kepler case particularly one can see how the map begins to be saturated in cross spots. This is due to the fact that these are large spots being simulated and that there is much more data than the previous objects.

for the simulation. This could (1) hint that the starspots are likely 3-3.5× as big as sunspots or (2) the simulation randomly erred on smaller spots/less crossings given the semi-random nature of parameter selection. Further checks could easily be done for this object, but for the sake of time and space, we settle with saying that for KIC 7767559, starspots cover a much more significant area than they do on the Sun, with numerous spots being on average 3-3.5× as big as those on our Sun.

4.5 KIC 7447200

The final object that I will present is KIC 7447200 as reproduced in Fig 4.20. Instead of displaying the residuals and CDF, I present the lightcurve and residuals of the system to highlight some interesting features.

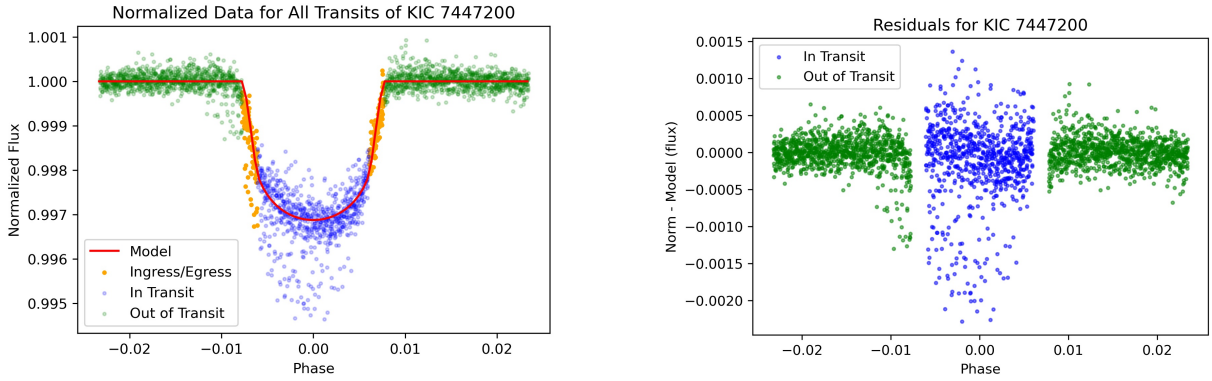


Figure 4.20: Data (left) and residuals (right) for KIC 7447200. Here note the difference in residuals with this object in comparison with the other objects, notably the quantity of flux points below the idealized lightcurve. Also note the consistent variation in the bottom left of the in-transit which points to the possibility of a permanent feature.

As seen in Fig 4.20, there are two interesting features found within the data that separate it from the starspot analysis formerly done. These two characteristics are (1) flux points below the idealized lightcurve and (2) consistent variation patterns within the lightcurve.

Beginning with the first feature–points below the lightcurve–we may take this as a sign that there is a process underway that produces the opposite effect of sunspot crossings, that is to say, crossing over particularly bright spots. While it is the case that these effects may come from errors in data acquisition/processing, the consistent, particular, and large variation hints that the effect comes from a surface feature with more intense flux. What may have been discovered on this system is thus a phenomena known as plages, which are more diffuse and brighter spots on the surface of a convective star that often appear in conjunction with starspot activity. Similarly produced by intense magnetic fields, plages are thought to be created when the flux tubes that suppress convection and lead to starspots emerge themselves from the photosphere and become a sight of intense coronal activity. While STSP does not have the capability of defining and simulating these regions, the plot that my starspot programs developed from the Kepler data of KIC 7447200 demonstrates the capability to search for and find these interesting stellar variations.

As mentioned, plages are typically produced by the same phenomena that creates starspots. Likewise, the two features often appear in conjunction. As we can tell by Fig 4.20, this is the case with KIC 7447200, which has both variations above and below the model lightcurve. What is interesting about this particular object is that we see in the lightcurve a diagonal

portion of higher increased flux in the left half (ingress) of the transit. Such a strong feature, where there are flux points repeated in a similar pattern, demonstrates what may be a spot that is not moving, as there is evidence of in former starspot research. This is particularly interesting in the case of KIC 7447200 because the path of transit is close to the Southern pole of the system. This confirms that (1) there are starspots that exist in the poles–something that only recently scientists have confirmed and which differs from the nature of our Sun; and (2) there is possibly a long-lasting polar spot. Again, this secondary finding is interesting because it is dissimilar to what we know about the Sun and may offer new insights into the convective/magneto dynamics of this star. Here, for sake of time and being repetitive, I do not undergo another starspot analysis of this system, leaving the section at pointing out only interesting and unique features that my programs are able to find and highlight.

4.6 Code Applications and Caveats

While the three codes are in functioning order, it is important to outline their possible functions clearly–as well as their current limitations and caveats. Recall that this code is primarily useful for (1) hunting for starspot signatures and other interesting stellar variations amongst the Kepler and TESS archive, (2) quantifying these spotted signatures, (3) quickly simulating similar systems, and (4) understanding and optimizing starspot parameters in terms of number, size (as compared to sunspot sizes), and contrast in the path of transit. An ancillary use of the code is that it is also helpful for checking the NASA exoplanet archival parameters and outlining what parameters may need updating. With all this information and particularly application 4, the work done with these codes can help aid to understand surface variations and the magnetic inner workings of distant convective stars like our Sun.

Despite these many applications, there are some important caveats to this work. One is that the data produced and outcomes gathered from the objects are only as good as the data and parameters themselves. When performing this analysis on a larger body of objects, there are plenty of objects that lack clear transits, have only partial V-shaped transits (in which case all points are ingress/egress), or have parameters that do not fit the data well enough to avoid artificial scatter be it from the limb darkening coefficients, the ratio of radii, or the planet-to-star-distances. Any of these parameter problems will lead to a less than optimal use of the code, but why this is the case varies. For one, it is possible that some of the transits are more grazing (less central latitudes) which in turn makes ingress and egress less hard to define and the model less precise. It could also be the case that error occurs from light pollution from nearby stars which haven't fully been filtered out. This would cause slightly different parameters for characteristics such as the ratio of the radii. Lastly, it could also be the case that the parameters are not perfectly well-defined, given that any slight variation could cause deviation from the model lightcurve. Nonetheless, the code is still useful to outline the problems that derive from elsewhere and point to work that needs to be done to further aid in starspot analysis.

I emphasize again that the solutions gathered from my simulations are not unique but rather general. When optimization is explored and the starspots show to be 3× as large

as sunspots, this is only to say that it is roughly in the vicinity of 3× the size of sunspots, and this is particularly so when considering possible degeneracies that have not yet been fully outlined or explored. The last caveat to mention is that the simulations rely on code that is also currently under development, particularly in updating the amount of starspots that can be simulated on a star at once (as evidenced on the basic starspot simulation number changing from 10 to 16.) This is because the program STSP lacks the capability to calculate fluxes from starspots that overlap, leading to problems when many starspots are simulated and there is overlapping between them. While in the future the number of simulated starspots may be updated, the current limit as of writing this thesis is 16 in order to avoid the breakdown of STSP. Furthermore, these spots need to be sequentially spaced. I attempt to circumvent this problem in all my trials by randomly picking the starting longitude of the star and then sequentially spacing the spots, as well as limiting the latitudes by 10 degrees North and South. Both of these solutions attempt to optimize number crossings and randomized crossings. Nonetheless, such a problem is worth noting and provides a cap on the number of starspot crossings possibly simulated.

Chapter 5

Conclusions and Future Steps

In this thesis, I have explained how my three programs work and the uses that they can be put towards. In that same explanation, I have highlighted four systems of interest, selected as a sample of what the programs can do. For KIC 106165657, it was seen that more than 10 randomly placed spots needed to be simulated (ten was insufficient to reproduce the scatter) and that the starspots were around 2.5 times the size of sunspots at around a contrast of 0.6. This is in alignment with what has already been uncovered of this system in literature. For KIC 8672910, it was seen that the simulation worked well with 16 randomly spaced spots, a contrast of 0.7, and a starspot increase of 2.75. Similarly, KIC 7767559 was found to work well with 16 randomly simulated spots, a contrast of 0.7, and a spot increase of 3, though this system does not appear to have been simulated with the most accurate parameters possible (more accurate parameters have not yet been found as of writing this thesis.) Finally, a graph and residual analysis of KIC 7447200 uncovered that there is the possible existence of two unique stellar magnetic features, namely plages and long-lasting polar spots. However, just as important as the scientific results garnered from using the code, the greatest outcome of this thesis is the creation of a pipeline that is multi-functional and provides a quick and accessible way to gather starspot-crossing information for Kepler and TESS systems.

With this pipeline in place, various further explorations into surface variations and the magnetic world of convective stars could be conducted. One is a more updated list of starspot crossing candidates than that which was provided by Michelle Gomez, given that such a catalogue was created before Kepler had fully undergone data processing. As such, a new catalogue gathered from my code may be more accurate and indeed more comprehensive due to this extra processing. The next possible step for future work could then be performing a catalogue-wide starspot analysis similar to that done in the results section of this thesis. This would give an idea as to how the KS numbers gathered from data residuals compare with the average starspot coverage on Kepler objects and further quantify just how common it is to see spotted Sun-like stars.

On the other end of things, a more detailed analysis of spot contrast and an updating of the maximum simulated spot amount on STSP could make the results from this code more accurate and pinpoint the spot nature of these stars. Here, another avenue for possible

exploration would be including the obliquity of the system. That is to say, the Kepler parameters do not currently take into account the alignment of the axis of stellar spin to that of the planet's orbit. This means that the latitudes of inspection for these analysis could in fact be describing not a strip of latitude but rather a full strip of the star at multiple latitudes. Though our Sun is not one of these systems, such an analysis could uncover new insights into how planet-star system formations correspond to the magnetic nature of the host star as well as how such spots latitudinally shift in other systems

Finally, after a full starspot analysis of all possible objects, it would be an interesting project to group starspot area coverage (as gathered from the KS tests), with the general features of each star, whether it be temperature, radius, age, type, etc. This could provide incredibly useful information into how the magnetic nature of convective stars change with these features and help flesh out the possible theories behind different stellar dynamos. Such work could greatly contribute to the body of knowledge of how stars form, behave, and affect planetary systems.

And so the story of the mysterious features that we call starspots continues onwards. From the Inuit people of Igloovik to the Aztec people of Mexico and the Zambezi people of Mozambique; from Theophrastus to Galileo, Wilson, and Hale; this investigation follows and continues the age long study of just what the deceptively dark patches on stars are and what they may mean to the systems in which they are found. What may come to light in the future, only time can tell, but one conclusion remains evident from this project: our pock-marked Sun is not alone in this galaxy but rather in the good company of family, each of which possesses their own exciting surface features. Each of which possesses their own unique face and life.

Figure 5.1: Picture taken by Juan Manuel Pérez Rayego of Active Region 1429 near Mérida, Spain depicting sunspots as seen from Earth.

Appendix A

An appendix

A.1 Starspot Candidates

Fig A.1 shows the different KIC candidates that were selected for initial analysis of my starspot program. The table that is reproduced and which was used as a guide post for the objects in my code construction gives the different systems that Gomez found to possibly contain starspot crossings as gathered by her code. While this may not be an exhaustive list, particularly when considering the state of the Kepler data when this table was constructed, it was extremely helpful with understanding which objects were the best to use as testers for my own starspot analysis.

A.2 Code Use

This section describes how to use the code that I have described in this thesis so that it may be utilized for other projects and investigations. All users can do what they will with this program and adjust the program as they like. All code can be found at the link http://people.hws.edu/hebb/ObservatoryWebpage/research/oliverdewey.html.

A.2.1 Starspot Program 1

In order to use Starspot Program 1, it is necessary to have the starspot parameter document entitled "total_KOI_rot.csv" in the same folder. Other than this, no extra document is necessary; however, the pathway for saving any document created will need to be changed for the pathway of the user's computer. For example, any pathway used to save figures/data tables, such as "/Users/oliverdewey/desktop/Starspots/STSPtable_transit" would need to be changed to "/Users/[Username]/[Folder location]/[Folder with Starspot files]/STSPtable_transit." The program get_transit_data(kepid, STSP=False, figures=False) only needs the kepid number to be passed into the program. All other arguments are optional, but in order to create the data file necessary for the STSP program, STSP needs to be set to True. To create the figures of each transit of the data, figures can be set to True. To run the same object with

Table 1: Objects With Starspot Crossing Features

Object	KS	Diff	Ratio	Prot (days)	Temp(K)
K01074.01	0.13407	4.5196e-07	2.7030	4.0925	6302.0
K00895.01	0.15299	5.2795e-07	3.5289	5.1448	5600.0
K01176.01	0.15733	1.7478e-06	4.0368	18.293	3806.0
K01353.01	0.16533	1.6187e-07	4.3919	8.6480	6279.0
K00806.02	0.16569	9.3152e-07	3.9685	16.004	5485.0
K00918.01	0.16852	4.1132e-07	4.6201	17.452	5552.0
K00676.01	0.18517	2.4080e-07	8.6638	12.290	3914.0
K00774.01	0.18716	9.4279e-07	2.5028	11.235	6108.0
K00217.01	0.19494	6.4174e-07	6.1088	20.461	5543.0
K00003.01	0.21721	2.3645e-08	23.446	29.472	5850.0
K00372.01	0.22047	8.9544e-08	8.4347	11.9	5838.0
K00883.01	0.22420	1.7588e-06	6.0798	9.1287	4809.0
K01784.01	0.23302	1.9027e-07	10.933	4.4940	5936.0
K01786.01	0.25997	1.9909e-06	2.2585	66.289	4461.0
K00203.01	0.28423	7.2625e-07	16.455	12.159	5624.0
K00063.01	0.28559	4.0582e-08	12.810	5.3985	5650.0
K00340.01	0.34334	1.8739e-06	47.107	12.936	5774.0

Figure A.1: Image of the data table as created by Michelle Gomez which highlights the interesting features found in her analysis.

the figures parameter set to True, another folder of the same name but empty will need to be created in the host folder, and the original folder renamed.

A.2.2 Starspot Program 2

In order to use Starspot Program 2, it is necessary to again have the starspot parameter document entitled "total_KOI_rot.csv" in the same folder. Note that this document has

the rotations for all the objects; however, it is the case that not all kepid objects will have recorded rotations, or rotations that are mistakenly zero. If this happens, the rotation will need to be manually put into the 'total_KOI_rot.csv" document, and a warning will appear saying that it is either Michelle's object (because this was the case with her object I analyzed) or another. It is also necessary to have the STSP program "stsp".

Again as with the last program, all paths will need to be adjusted to the pathway on each user's computer. This is necessary for reading and saving the right files for use of the program. This program will write a starspot parameter program for each simulated transit, entitled "[kepid]_ [transit]+starspot_parameter.txt'"; an "in-file" entitled "[kepid]+_+ [transit]+.in" for use in STSP; and produce the STSP-simulated lightcurves entitled "[kepid]+_+[transit]+_lcout.txt." Total combined files of each transit, simulated spot, and crossed spot are also created here, entitled "[kepid]_total_lcout.txt", "All_spots.txt", and "All_crossings.txt" respectively. The program get_STSP(kepid, planet_number=1, spot_number=1, stay=False, spot_increase=1, contrast=.7) currently has the functionality to have one planet (though more can be added in the future), spot numbers up to 16 (which generally provides a picture similar to the data), and a contrast from 0-1. The spot_increase argument will multiply the sunspot sizes by the number chosen. The stay function will use whatever previously written files for that object were used before instead of simulating new randomly chosen spots. This has been underused and so may hold a bug that came up somewhere in development. Lastly, the kepid argument defines what KIC object is being looked at.

A.2.3 Starspot Program 3

As with the other programs, this last of the 3 programs requires the "total_KOI_rot.csv" file as well as changing the pathway to the user's pathway at every instance a pathway shows up in the code. The files required for the use are those created either by STSP or Starspot Program 1, that is "[kepid]_total_lcout.txt" or "kepidfile_'+[kepid]+.txt" respectively. For mapping the starspots onto a circular map the files "All_spots.txt", and "All_crossings.txt" are also needed. The program get_residuals(kepid, STSP=False, figures=False, contrast = 0) only needs the Kepler ID number to be passed into the program. If STSP is False, then the residual analysis is conducted on the Kepler data. If it is True, then the analysis is conducted on the simulated STSP files. Finally, if figures is True, plots will be saved of all steps. The contrast arguement was implemented to attempt to display the relative contrasts on the maps, but is not useful in the overall use of the program.

A.2.4 Final Starspot Program

Though there was no section in this thesis on the final program, given that it was constructed of snippets of the last three programs, here, I will briefly explain how to work it. This is the program that produces the KS analysis and explores the combinations of contrast, size, and number for the simulated starspots. This program again requires the documents "total_KOI_rot.csv" and the STSP program "stsp", as well as changing any pathways to the

user's pathways. Other than this, the only additional requirement is that, if there is not data on the object, to perform the get_transit_data program first with STSP as True, so that the proper files have been created. From here, the only other necessary thing to do is to run the program Starspot_test(kepid, test_num, contrasts, sizes). In this program, the user can specify the kepler ID number of the object; say how many tests they want performed (integer number); enter the contrasts desired to explore (as a list); and those sizes to explore (as a list). This program will only save the final plot, which looks at how the KS test changes with these various variables. Note that this program is quite slow in comparison with all the other programs and gets significantly slower the more number of tests the program is asked to do.

Bibliography

Batalha, Natalie M. (2014). "Exploring exoplanet populations with NASA's Kepler Mission". In: *Proceedings of the National Academy of Sciences* 111.35, pp. 12647–12654. DOI: 10.1073/pnas.1304196111. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.1304196111. URL: https://www.pnas.org/doi/abs/10.1073/pnas.1304196111.

Baumann, I. and Solanki, S. K. (2005). "On the size distribution of sunspot groups in the Greenwich sunspot record 1874-1976". In: *A&A* 443.3, pp. 1061–1066. DOI: 10.1051/0004-6361:20053415. URL: https://doi.org/10.1051/0004-6361:20053415.

Berdyugina, S (2005). "Starspots a Key to the Stellar Dynamo". In: *Living Reviews in Solar Physics* 2.

Charbonneau, Paul (Aug. 2014). "Solar Dynamo Theory". In: 52, pp. 251–290. DOI: 10.1146/annurev-astro-081913-040012.

Donati, J-F and SF Brown (1997). "Zeeman-Doppler imaging of active stars. V. Sensitivity of maximum entropy magnetic maps to field orientation." In: *Astronomy and Astrophysics* 326, pp. 1135–1142.

Gomez, M. (2015). "Starspot crossing transits in long-cadence Kepler data: a search for correlation between starspots and stellar properties".

Hamacher, D. (2022). *The First Astronomers: How Indigenous Elders read the stars*. Allen & Unwin. ISBN: 9781761063800. URL: https://books.google.com/books?id=x3lREAAAQBAJ.

Harvey, John (Nov. 1999). "Hale's Discovery of Sunspot Magnetic Fields". In: 525C, p. 60.

Haswell, C (2010). *Transiting Exoplanets: Measuring the Properties of Planetary Systems*. Cambridge: Cambridge University Press.

Hathaway, David H. and Debi Prasad Choudhary (Aug. 2008). "Sunspot Group Decay". In: 250.2, pp. 269–278. DOI: 10.1007/s11207-008-9226-4.

Hebb, Leslie et al. (in prep). "STSP: An eclipse mapping program for modeling transiting light curves with starspot crossing events". In: *Unpublished*.

Kreidberg, Laura (Nov. 2015). "ttbatman/tt: BAsic Transit Model cAlculatioN in Python". In: *Publications of the Astronomical Society of the Pacific* 127.957, pp. 1161–1165. DOI: 10.1086/683602. URL: https://doi.org/10.1086%2F683602.

Lightkurve Collaboration et al. (Dec. 2018). *Lightkurve: Kepler and TESS time series analysis in Python*. Astrophysics Source Code Library. ascl: 1812.013.

Met Office (2010 - 2015). *Cartopy: a cartographic python library with a Matplotlib interface*. Exeter, Devon. URL: https://scitools.org.uk/cartopy.

Morris, Brett M. et al. (Sept. 2017). "The Starspots of HAT-P-11: Evidence for a Solar-like Dynamo". In: *The Astrophysical Journal* 846.2, p. 99. DOI: 10.3847/1538-4357/aa8555. URL: https://doi.org/10.3847%5C%2F1538-4357%2Faa8555.

Nandy, D. (Oct. 2004). "Meridional Circulation and the Solar Magnetic Cycle". In: *SOHO 14 Helio- and Asteroseismology: Towards a Golden Future*. Ed. by D. Danesy. Vol. 559. ESA Special Publication, p. 241.

Ruzmaikin, Alexander (Jan. 2001). "Origin of Sunspots". In: 95, pp. 43–53. DOI: 10.1023/A:1005290116078.

Schutte, Maria C. et al. (June 2022). "Modeling Stellar Surface Features on a Subgiant Star with an M-dwarf Companion". In: *The Astronomical Journal* 164.1, p. 14. DOI: 10.3847/1538-3881/ac70ca. URL: https://doi.org/10.3847%5C%2F1538-3881%5C%2Fac70ca.

Solar Physics (2022). NASA. URL: http://accuglassproducts.com/home.php.

The Enduring Mysteries of the Sun (2007). Space. URL: https://www.space.com/4522-enduring-mysteries-sun.html.

Thomas, J.H. and N.O. Weiss (2012). *Sunspots and Starspots*. Cambridge Astrophysics. Cambridge University Press. ISBN: 9781107410558. URL: https://books.google.com/books?id=cHR6LgEACAAJ.

Wall, J.V. and C.R. Jenkins (2012). *Practical Statistics for Astronomers*. Cambridge Observing Handbooks for Research Astronomers. Cambridge University Press. ISBN: 9780521732499. URL: https://books.google.com/books?id=TeUG0sZ47EYC.