Hobart and William Smith Colleges

Identifying objects with starspot crossing characteristics from Kepler K2 data

Thesis submitted in partial fulfillment of the requirement for Honors in

Physics

Misty Chien

Advisor: Prof. Leslie Hebb

April 11, 2023

Contents

1 Introduction **3**

1.1 Magnetic Field on the Sun 3

1.2 How about other stars? 10

1.3 Transiting Planets 13

1.3.1 Transits and Lightcurves 14

1.4 Kepler/K2 Mission 20

1.4.1 Overview 20

1.4.2 Instrumentation 21

2 Methods **25**

2.1 Getting individual transits from lightcurve 27

2.2 Normalize data on transit-by-transit basis 29

2.3 Rejecting bad data 31

2.4 Generate batman model 33

2.5 Using model and normalized transits to get residuals 34

3 Results **35**

3.1 Statistics 35

3.2 Likely objects with starspot crossings 37

4 Conclusion and Future Directions **40**

References **43**

A Additional Figures **44**

B Python code **49**

Abstract

Sunspots are observable manifestations of the Sun's magnetic field. The Sun is close to us compared to other stars in the Universe. This means that with a telescope, we can see the dark sunspot features in much more detail than a star that is both fainter and much farther away. What if we want to learn about the magnetic field of other stars? A lot less is known about these stars, and the tools with which to learn about them are more limited. A technique that is used by astronomers to circumvent this problem is using transiting planets as a probe. The Kepler spacecraft monitors the brightness of stars and detects orbiting planets through an apparent dip in the brightness when a planet crosses in front of the star. Planets sometimes cross in front of starspots on the star; this has a unique effect on the dip in brightness. The planet is now blocking less light, so there are bumps in the dip of the brightness. Therefore, analyzing in-transit variation of brightness of stars can help identify objects with starspot crossing events. The goal of this project is to identify objects with starspot crossing features from a list of K2 candidates. The planet samples are obtained from Campaigns 1-8 and 10-18 of the K2 mission (Zink et al., 2021). Given an object, the program extracts its long cadence lightcurve data from the Kepler/K2 catalogs of the Mikulski Archive for Space Telescopes (MAST), normalizes individual transits, and computes a model based on its stellar parameters using BAsic Transit Model cAlculatioN in Python (batman). The program then calculates the difference between the data and the model and performs the Kolmogorov-Smirnov (K-S) test on the residuals to determine whether they are statistically from different distributions. Out of 196 objects, 10 objects are identified by visual inspection of their lightcurves to have starspot crossing features. These objects have K-S statistics ranging from 0.15 to 0.39. By compiling a short list of objects with starspot crossing features, this allows for further study on these stars.

1 Introduction

1.1 Magnetic Field on the Sun

The sun is a giant ball of boiling gas. The center of the Sun is the core that is full of nuclear fusion reactions of hydrogen into helium, at a temperature of 15.7 million Kelvin (K) [7]. This is where all of the Sun's energy is generated. All of the energy is trying to escape the core, but it can't all escape at the same rate. This creates layers in the Sun, where energy escapes through different mechanisms, such as convection and radiation. The layer surrounding the core is the radiative zone, at a temperature of 2-7 million K, depending on the distance to the core [7]. This layer absorbs a lot of the core's energy. The following convection zone is at a lower temperature, at about 5780 K on the surface, where heat exchange occurs through convection [7]. The boundary between the radiation and convection zone is called the tacholine. Past the tachocline, the convection zone extends into the photosphere, which is the visible, surface layer of the Sun. Photons from this layer escape the solar atmosphere, and reach us in the form of solar radiation and sunlight. The atmosphere, like the interior, have several layers. The innermost layer of the Sun's atmosphere is the chromosphere. Interestingly, the atmosphere of the Sun is increasingly hotter than the surface of the Sun with increasing altitude, starting from about 20,000 K at the chromosphere, to 1-20 million K at the corona [7]. The last layer, the heliosphere, is less like a layer or a sphere, but more like a fluid region around the Sun whose shape is dependent on the fluid interstellar medium and solar wind [7]. Fig. 1 illustrates the internal layers of the Sun, as well as the atmosphere and its features.

From the core, the radiation random walks through the radiation zone before reaching the convection zone much later in time. The gases of the Sun are in the form of plasma, or ionized gas. In the convection zone, heated plasma rises and cooled plasma falls, forming a

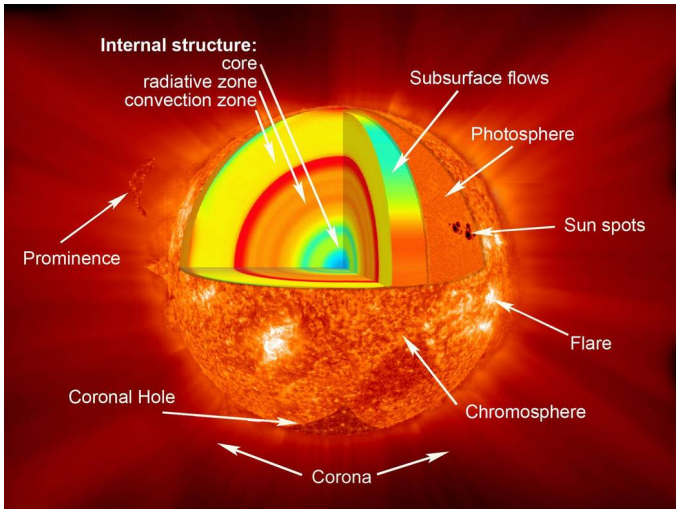


Figure 1: Layers of the sun. Image credit: NASA

convection current. The Sun rotates around an axis through its center, and the Sun experiences differential rotation due to the photosphere's varying rates of rotation at different latitudes. Sidereal time is a unit of time used in astronomy to find celestial objects. The time unit is based on Earth's rate of rotation relative to the fixed stars [1]. The photosphere rotates according to the sidereal rotational period, which converts to different velocities at different latitudes. The rotational motion is much faster than convection so it dominates the motion of the Sun's deep interior [7]. This internal spin is not entirely uniform. In the convection zone, the rotation is latitude-dependent, while further down closer to the core, the rotation is not dependent on the latitude. This sharp difference occurs at the tacholine, the boundary between the convection and radiation zone. The dramatic forces that occur here is the suspected solar dynamo location, where the Sun's magnetic field is generated [7]. This magnetic field needs to be amplified and continually regenerated. The convection zone provides a suitable environment for the generation of electric currents as the convection currents are good conductors of electricity. This electric current generates a magnetic field, which in turn generates electricity, and the cycle repeats [2]. In Fig. 2,

the magnetic field lines of the Sun, connecting the magnetic poles, are twisted by internal convection process and differential rotation. As the magnetic field lines wind up as the rotation continues, two belts are formed, and they twist into a magnetic loop that can burst and leave behind a dark spot on the Sun.

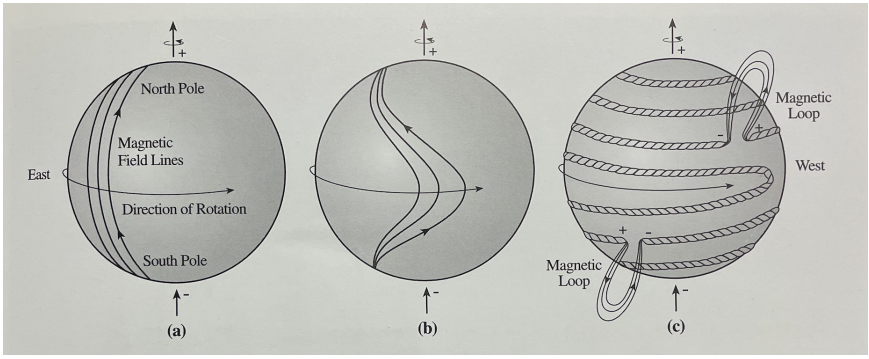


Figure 2: A model illustrating how polarity is generated in the Sun. Image a) shows the magnetic field lines at the solar minimum, at the beginning of the Sun's 11 year cycle. Because the rotation is faster around the equator, deformations in the field lines start to form in Image b). Furthermore, the conductive, rotating plasma winds up the magnetic fields, and two belts are formed at the solar maximum, in Image c). Image credit: Cambridge Encyclopedia of the Sun

Magnetic fields are vector fields that describe how magnetic force is distributed in space. They are produced by moving electric charges. In each atom, there are electrons orbiting its nucleus. The motion of the electrons produces a small magnetic field around each atom. Each magnetic field has a distinct orientation; this is called the magnetic moment. When taking a group of atoms whose magnetic fields have different magnetic moments arranged randomly, their net magnetization is zero, and there are no observable magnetic fields. However, when the direction of the magnetic fields of the atoms align, they produce a magnetization, and thus a measurable magnetic field [11]. This is the familiar bar magnet, as shown in Fig. 3. The lines connecting the north and south poles are the magnetic field lines. In the physical world, these field lines are invisible to our naked eye. Yet, when iron

filings are piled around the bar magnet, the iron filings form a pattern, obeying an invisible rule. This is the magnetic force in action. Magnetic force, $\vec{F}$, for a charge, q, follows the direction of the magnetic field, $\vec{B}$,

$$\vec{F} = q\vec{v} \times \vec{B} \tag{1}$$

Equations of magnetic force and magnetic fields quickly get complicated, but from Eqn. 1, we understand the components necessary for magnetic activity.

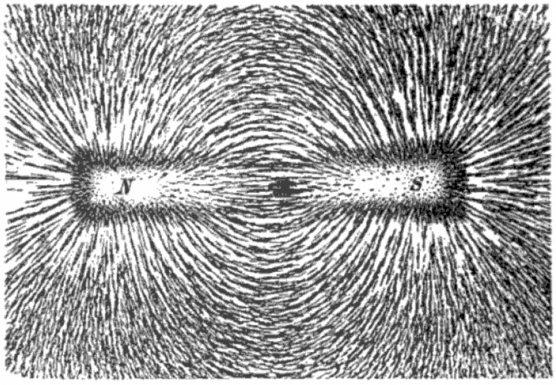


Figure 3: Magnetic field lines around a bar magnet. Image credit: Practical Physics, Newton Henry Black

Things that produce magnetic fields have two poles. In Fig. 3, magnetic field lines emanating from the north and south poles of the bar magnet. Similar to electric charges, opposite poles attract and identical poles repel. If the bar magnet is snapped in half, each piece would have a north and south pole, rather than a piece of the north pole, and a piece of the south pole. There are no magnetic monopoles, but one thing that the poles could do is flip their orientation.

Generally, magnetic fields are generated by moving charges. The plasma of the Sun is highly conductive. Unlike the bar magnet, the Sun is not composed of solid, magnetic material. Rather, the convection of plasma inside of the Sun in addition to its rotation

generates a magnetic field. Just like the bar magnet, the Sun has a north and a south pole from which the solar magnetic fields emanate. The solar cycle is based on solar magnetic activity where the poles of the magnetic field of the Sun flip every 11 years. During each cycle, solar magnetic activity fluctuates from a minimum, to a maximum, then back to the minimum. The level of magnetic activity correlates with the appearance of sunspots, solar flares, and coronal loops on the Sun's surface [7]. At a maximum, the magnetic loops in Fig. 2 bursts into explosive coronal mass ejections (CME), leaving behind dark spots which we call sunspots. Fig. 4 is an image of the Sun, captured by the Helioseismic and Magnetic Imager on NASA's Solar Dynamics Observatory in 2014. The sunspot in Fig. 4 is the largest sunspot observed in this solar cycle, spanning 80,000 miles across [16].

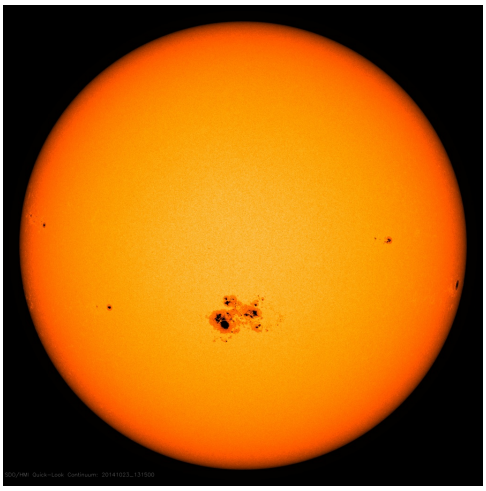
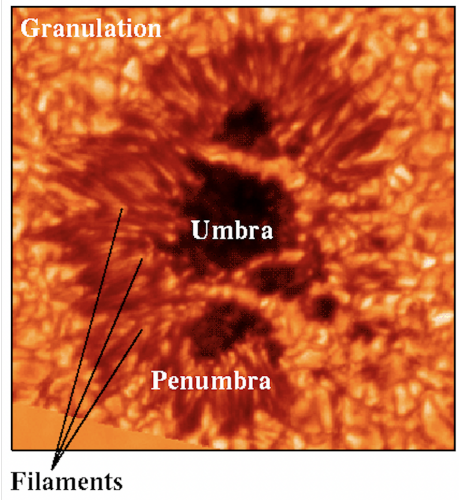


Figure 4: Image of the sun, with visible sunspots. Produced by solar surface velocity maps (dopplergrams) and broad-wavelength photographs on the Solar Dynamics Observatory. Image credit: NASA

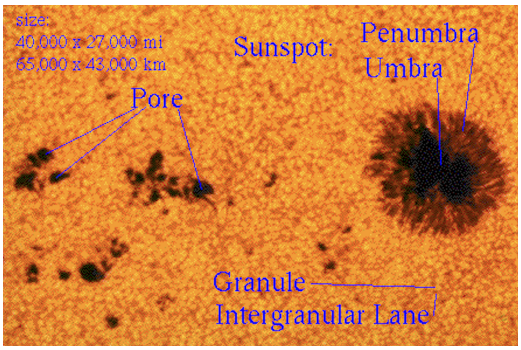
Sunspots are observable manifestations of the Sun's magnetic field. They are dark spots that rises from the inside of the Sun and disappear into the surface of the Sun. They are about 2000 K colder than their surrounding photosphere at 5780 K, which attributes to their dark color [7]. The darker center of the sunspot is called the umbra, and its lighter

surrounding is called the penumbra, as shown in Fig. 5a. A typical, developed sunspot has a penumbral diameter of 20 to 60 million meters [7]. More complicated, larger sunspots can have multiple umbrae and filamentary penumbra where matter flows out of. Sunspots can last from just a few hours up to months [7].

Much is unknown about exactly how sunspots form, but the general idea holds that they are manifestations of magnetic flux tubes from the convection zone, locating where magnetic flux tubes burst in Fig. 2. The rate at which sunspot emerges also correlate with the solar cycle, with slower rates at the end of the roughly 11 year cycle. Sunspots disperse when the magnetic field is no longer concentrated in that region [7].



(a) Sunspot with filaments



(b) Sunspot and pores

Figure 5: A detailed view of sunspots, with labelled umbra, penumbra, filaments, and pores. Image credit: NASA

People have been observing sunspots for centuries. In his time, Galileo saw sunspots through his telescope. Since 1874, the Royal Greenwich Observatory has been recording sunspot observations [14]. Fig. 6 is a time-latitude diagram of sunspot activity of the first recorded solar cycle to the current cycle. The first sunspot bands, representing the

sunspot areas, emerged at around 30 degrees latitude above and below the equator. Then, the sunspot band widens, and progresses towards the equator. From the bottom panel, we see that this is when the solar cycle is at a maximum. By the time the sunspots reaches the equator, the solar cycle retreats to a minimum, and another sunspot band reemerges at the same initial $\pm 30^\circ$ latitudes. Despite the varying cycle amplitude and cycle length, the pattern repeats itself. The periodicity in solar cycles creates a pattern that resembles the wing of a butterfly.

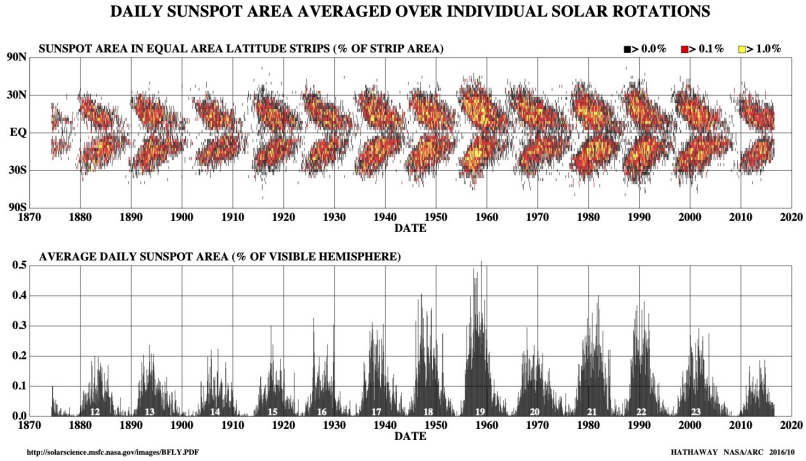


Figure 6: Butterfly diagram of the solar cycle. Top panel shows the location of sunspots; bottom panel shows the total sunspot area. Sunspots appear to form at around 30 degrees latitude and migrate towards the equator. Although the pattern is consistent, the cycle amplitude and cycle length have varied. Image credit: NASA

How can magnetic fields be measured? When an atom is placed in a magnetic field, the spectral lines, describing the energy level transitions of electrons, split into multiple closely spaced components [7]. This is called the Zeeman effect. The number and amount of Zeeman splitting increase with the strength of the magnetic field, thus allowing the magnetic field to be measured. In 1908, the American astronomer George Ellery Hale used

this technique to show that the magnetic fields inside of a sunspot is thousands of times stronger than that of the Earth's magnetic field [7]. It is later measured that sunspots have magnetic field strengths as strong as 0.3 Tesla (T), while the Earth's magnetic field is 0.00003 T at the equator [7].

Why do we study the Sun? From a practical perspective, magnetic activity on the Sun affects the Earth. Solar winds are expansions of the plasma from the solar corona into interplanetary space. The wind are streams of highly energetic charged particles, traveling at a velocity of 8×10^5 ms^{-1} [7]. The terrestrial magnetic field deflects most of the solar wind hitting the Earth and carves a cavity in which the planet is shielded from the solar wind. During the peak of the Sun's activity, intense geomagnetic storms can occur. The most powerful magnetic storms result from coronal mass ejections from the Sun's surface. During the peak, CME's can also generate intense auroras by affecting the Earth's magnetotail.

Our world have grown to be dependent on networking from geosynchronous satellites. These signals are vulnerable to space weather events, putting the world's communication, navigation, safety, and data transmission at risk. The strong electric current flow of millions of amps, from intense geomagnetic storms, can also induce a voltage difference between the ground and power grids, temporarily or permanently disrupting the power supplies of major cities [7]. By studying the magnetic field of the Sun, scientists could one day develop a model of predicting the Sun's explosive events and learn about the Sun's internal dynamo.

1.2 How about other stars?

Because the Sun is only 1 astronomical unit (AU), or 149.6 million kilometers, away from the Earth, sunspots can be observed in much greater detail with a telescope compared to

stars that are both fainter and farther away. To give an idea of the difficulty in observing other stars in detail, Proxima Centauri, the closest star to the solar system, is roughly 268,770 AU from the Sun [15].

Cosmic magnetic fields are the principle mechanisms behind many astrophysical phenomena. These events include the accretion, or particle acculumation, process of black holes, the accretion and acceleration of low to medium mass stars, and cosmic gamma-ray bursts [13]. These areas of astrophysics build upon our understanding of the magnetic fields of the Sun and other stars. Similar to the Sun, starspots are created by the magnetic field on the surface. Stellar rotation paired with the internal convection motion of the plasma generate strong magnetic fields that induce starspots, coronal loops, and flares [13]. However, Strassmeier (2009) suggests that even starspots on stars with the same mass and effective temperature as the Sun cannot simply be explained by scaling our knowledge about the Sun [13].

For astronomers studying the evolution of stars, rotational measurements are necessary for testing models of how stellar rotation changes with mass and age [13]. Starspots are useful for this endeavor. As stars rotate, if starspots are present, the dip in brightness from the dark spots would be carried across the star as it rotates. Differential rotation refers to when surface rotation is measured as a function of stellar latitude if multiple spots are present at different latitudes and one of the latitudes were known [13]. Doppler imaging is a method of detecting starspots, as well as measuring differential rotation and magnetic field strengths. As stars rotate, the side that is moving towards us is blue shifted, and the side that is moving away from us is red shifted. This is the Doppler effect in action, as frequency of light from the object changes due to the relative motion from the observer. Fig. 7 illustrates the how Doppler imaging can find starspots. Beneath the purple spot, there is an observable distortion in the absorption line directly below. As the star rotates,

the bump from the purple spot moves across the curve, and we begin to see the bump from the yellow spot as it comes into view. Using the polarization of light, Doppler imaging can also find the strength of the star's magnetic field. The bumps in the polarized absorption lines show that the magnetic field in stronger in areas with spots. As spots create observable distortions in the absorption line, Doppler imaging is a powerful technique to detect them.

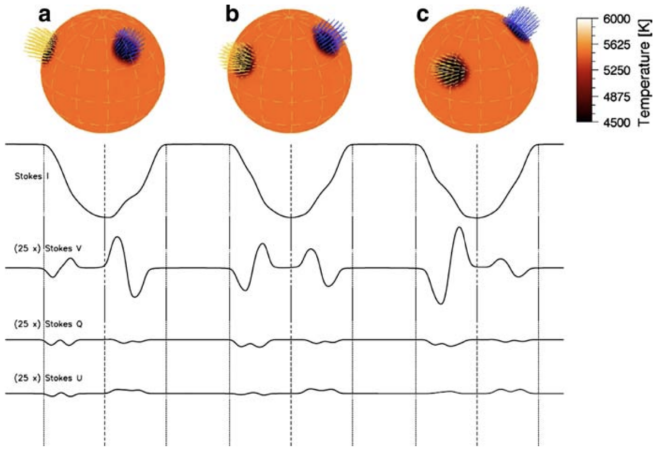


Figure 7: Doppler imaging of stars with starspots. Top row shows the brightness at a particular wavelength of the absorption line. The distortion caused by starspots moves across the the dip as the star rotates. The bottom three are the polarized absorption lines of the same wavelength in each direction. Image Credit: Klaus Strassmeier, 2009

By studying other stars, we are also gaining information on our Sun. The Sun is only one star, among about 100 billion stars in our galaxy [7]. Other stars provide more stellar parameters, as they have different masses and internal structures. This makes our understanding of the Sun more complete, as information about other stars allow us to see where our Sun fits in in the study of stars.

1.3 Transiting Planets

When a star is being orbited by a planet, the star moves slightly due to the gravitational pull of the orbiting planet. The radial velocity method of extra-solar planet detection is based on detecting this host star's change in velocity as it moves around its orbit. From a distance, the change in radial velocity can be measured using Doppler shift of the emitted light. Ground telescope with spectrometers can measure the star's spectrum, from which astronomers search for any periodic blue or red shifts. By measuring the amplitude of the velocity, the mass of the planet can be determined. One of the drawbacks of this method is when the star-planet orbit is orientated such that it doesn't move away from and towards the Earth so Doppler shift cannot be observed.

A novel way to detect the presence of a planet around a star is the transit method. The transit method measures the brightness of stars as periodic dips in brightness indicating planet crossings between the star and the observer on Earth. The size of the planet can be found from how much light is being blocked, and the shape, depth, and timing of the transit can indicate the orbit and period of the planet. Additionally, by observing the light passing through the planet's atmosphere through a spectrograph, astronomers can learn about the planet's atmospheric composition [2]. The transit method and Doppler imaging are powerful, complementary methods of discovering starspots, but the transit method has the additional benefit of discovering the latitudes of spots from known planet parameters. My program uses the transit method to find starspot crossing events among stars with transiting planets. Fig. 8 is an illustration of a planet transiting in front of a star and the corresponding lightcurve when the brightness of the star is measured.

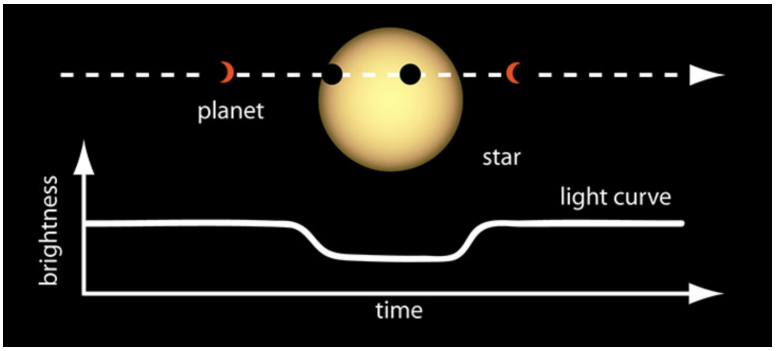


Figure 8: Lightcurve of a planet transiting its star. Image Credit: NASA Ames

1.3.1 Transits and Lightcurves

Lightcurves are plots of observed flux versus time, depicting variations in brightness of an object. Fig. 8 is an example of a transit lightcurve. The brightness axis plots the measured flux of the star with respect to time. As the planet crosses in front of the star, it occults some of the brightness from the star, resulting in a dip in brightness.

What determines the shape of a lightcurve? Transit lightcurves are characterized by a number of parameters, including transit depth, transit duration, ingress or egress duration, and limb darkening coefficients. The depth of a transit is proportional to the square of ratio of the planet-to-star distance [2],

$$\frac{\Delta F}{F} = \frac{R_P^2}{R_*^2} \tag{2}$$

Here, ΔF represents the change in flux, and R_P and R_* represents the radius of the planet and the star respectively. The ratio of the planet radius and the star radius estimates the fraction of light that the planet is blocking, by estimating how much of the star's area covered by the planet during a transit. Eqn. 2 does not account for the relative distance between the planet and the star. In general, the planet-to-star distance is much smaller than their distance to the observer, thus to a high degree, the planet distance and the star

distance are identical [2]. In catalogs, this parameter is denoted as ratio or planet-to-star radius, R_P/R_*.

Based on observations, the period of a transit can be found by the relationship:

$$P = \frac{T_{\text{elapsed}}}{N_{\text{cycles}}} \tag{3}$$

where T_{elapsed} is the duration of observation between two points, starting and stopping at the same point during the transit. The precision of the period is affected by the number of cycles being observed [2]. Once the orbital period is measured, other quantities of the orbit can be found from Kepler's Third Law:

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

where a is the semi-major axis of the orbit, M_* is the mass of the star, and M_P is the mass of the planet [2]. When $M_* \gg M_P$, Eqn. 4 can be simplified to approximate the semi-major axis [2]. Derived from Newton's laws and solar system observations, Kepler's third law characterizes the basis of planetary orbits.

Transit duration is found by calculating the fraction of the orbit that the planet is in front of the star [2]. Fig. 9 represents an orbit with radius a around a star with radius R_*. Since the light rays V and W are parallel, the arc length between them is about twice the radius of the star. The fraction of the orbit and the orbital period becomes:

$$T_{\text{dur}} \approx \frac{P \times 2R_*}{2\pi a} \approx \frac{PR_*}{\pi a} \tag{5}$$

Not all transits pass through the center of the star. To account for deviations from the center, the impact parameter, b, describes the shortest distance from center of the stellar

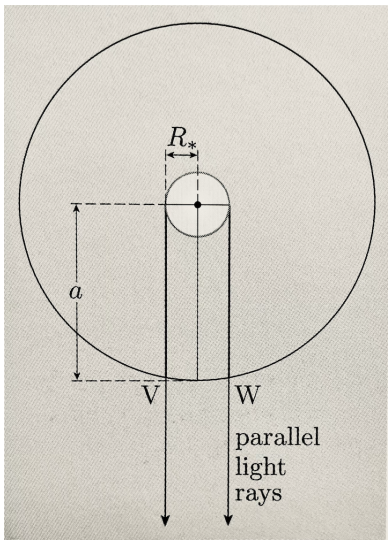
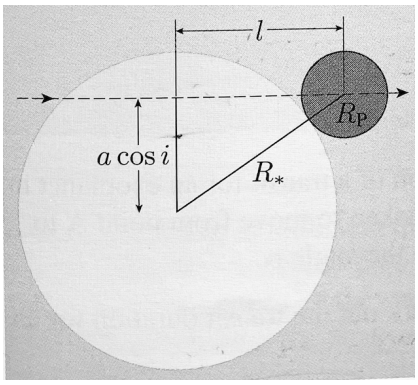
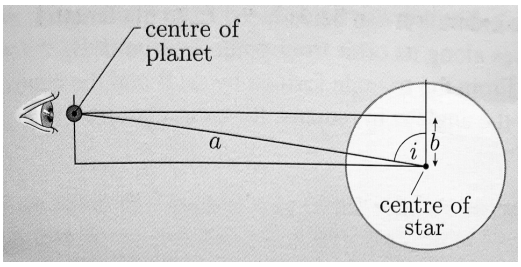


Figure 9: A circular orbit of radius a, around a star of radius R_*. The observer is at the bottom of the page, in the direction facing the light rays from V and W. Image credit: Transiting Exoplanets, C.A. Haswell



(a) The length l can be found using Pythagorean's theorem.



(b) The planet is at an impact parameter, b, above the center of the star.

Figure 10: Image credit: Transiting Exoplanets, C.A. Haswell

disc to the locus of the planet [2]. Eqn. 5 is an approximate equation for the duration, which only works for $b = 0.0$. From Fig. 10b, we find the impact parameter, b, to be:

$$b = a\cos i \tag{6}$$

$$= \frac{\mathrm{a\cos i}}{\mathrm{R_*}} \text{ (dimensionless)} \tag{7}$$

and the length, l, in Fig. 10a, is calculated from Pythagorean's theorem:

$$l = \sqrt{(R_* + R_P)^2 + a^2\cos^2 i} \tag{8}$$

As the planet moves from point A to point B on the orbit in Fig. 11, it is at an angle, α, from the center of the star [2]. Using the triangle formed by point A, B, and the center of the star, the distance from the center to point A or B is $\sin(\alpha/2) = l/a$. With the impact

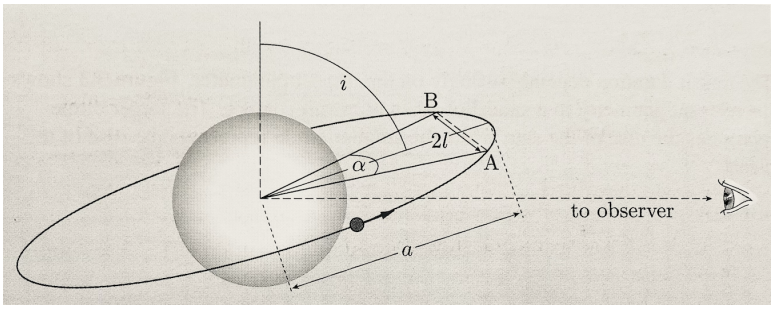


Figure 11: Transit of a planet with non-zero impact parameter. Image Credit: Transiting Exoplanets, C.A. Haswell

parameter correction, for circular orbits, the duration becomes:

$$T_{\mathrm{dur}} = P\frac{\alpha}{2\pi} = \frac{P}{\pi}\arcsin\frac{\sqrt{(R_* + R_P)^2 + a^2\cos^2 i}}{a} \tag{9}$$

Even though these values are typically provided in planet catalogs, a general idea of how these quantities are derived are useful to the understanding of how the geometry and planet

parameters relate to one another.

A transit begins when the planet comes in contact with the limb of the star. The period from which the planet edge first contact the limb of the star to the point where the other edge of the planet contacts in the inside edge of the limb is the ingress, t_{1-2}, of the transit. The period in which the planet starts and finish transiting in front of the star is the egress, t_{3-4}. Fig. 13 maps the planet transit with the corresponding points in the lightcurve, as well as labeling the properties: impact parameter b, transit duration T_{tran}, transit depth ΔF, limb darkening c, and the ingress t_{1-2}.

When the brightness of stars are measured, they exist in a gradient that decreases radially, with the limb of the star being fainter [7]. The study of limb darkening provides insight into temperature structures of the stars. Fig. 12 is an illustration of the temperature gradient. Limb darkening coefficients affect the shape of the bottom of the transit. The shape of the transit in Fig. 13 indicates that the flux is greatest at the center of the transit and gets progressively darker towards the limbs while entering and exiting the transit. From Eqn. 2, the depth of the transit is related to the ratio of planet radius to star radius. With transits not being flat-bottomed, this suggests that the planet-to-star radius varies smoothly during the ingress and egress of the transit. As the edges of stars are not sharp edges but plasma, light escapes from the atmosphere at different rates [2]. Limb darkening depends on the spectral type and composition of the star, and direct observations of limb darkening cannot currently be made on stars other than the Sun [2]. To model lightcurves, limb darkening laws are computational models used to approximate the limb darkening of stars. Limb darkening models can be linear, logarithmic, quadratic, or cubic [2]. The quadratic law, requiring two limb darkening coefficients, is used in my program, as it provides adequate flexibility to fit the K2 data.

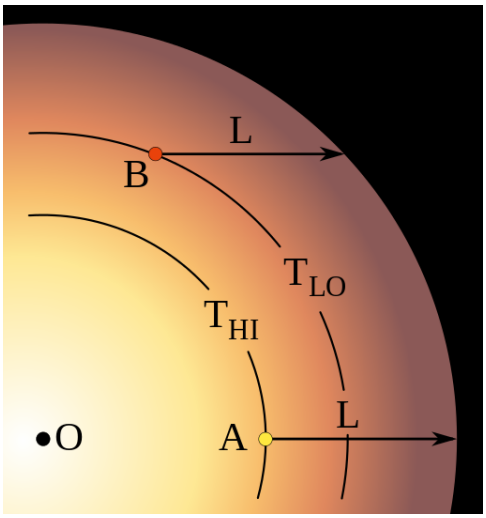


Figure 12: An idealized illustration of limb darkening on the Sun. Image credit: Wikipedia

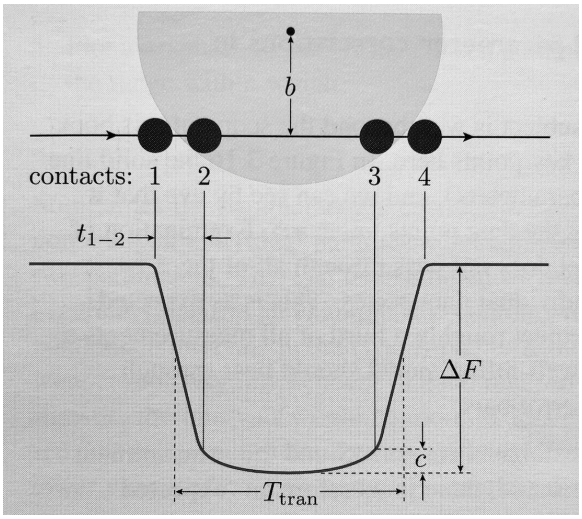


Figure 13: Lightcurve of a planet transiting its star, showing the properties: impact parameter b, ingress t_{1-2}, transit duration T_{tran}, transit depth ΔF, limb darkening coefficient c. Image credit: Transiting Exoplanets, C.A. Haswell

1.4 Kepler/K2 Mission

1.4.1 Overview

In 2009, the Kepler spacecraft was launched by NASA from Cape Canaveral, Florida. The spacecraft houses a 0.95m space telescope with photometric precision of up to 30 parts per million [3]. This means that Kepler was well suited for observing fainter stars as low as the 12th magnitude. The main science goal of the Kepler mission is to survey exoplanets in or near habitable zones in the Milky Way, and more specifically, determine whether these planets are in multiple-star systems, determine the variety in planetary sizes and orbits, and determine properties of the host star of these planets [3].

Kepler was mainly a statistical survey of the exoplanet population in the specified region. Designed to monitor about 100,000 main sequence stars, Kepler discovered more than 2,600 exoplanets [4]. The primary findings show that terrestial-sized planets are fairly common – with 20 to 50 percent of the stars in the sky that are likely to have these planets orbiting them [3]. Kepler also found a diverse population of planet and systems with different sizes. Besides planets, the amount of stars that Kepler observed provided a multitude of data for astronomers [3].

Kepler uses the transit method to detect exoplanets. Requiring a continuous view of one patch of sky in order to catch any changes in its brightness, Kepler's field of view must be fixed while operating. The ideal field of view would have an abundance of stars and exclude the Sun. For these reasons, the chosen field of view is a region of Cygnus and Lyra constellations in our galaxy. When Kepler lost a second reaction wheel in May 2013, Kepler continued to take data until it was reprogrammed to point at a different patch of the sky. This became the K2 mission, operating from May 2014 to October 2018 [3]

The K2 mission is an example of resilience and ingenuity. After the spacecraft had lost two reaction wheels, the spacecraft lost control of its roll axis, shown in Fig. 15. The

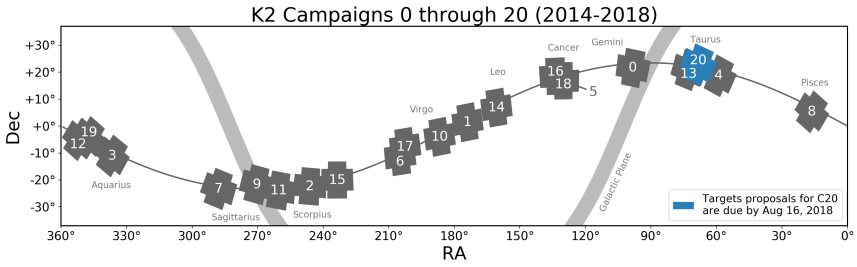


Figure 14: Map of Observations of K2 mission. Image Source: MAST

reaction wheels provided Kepler the ability to control its attitude, so K2 needed a new way to stabilize the roll axis. The engineering solution is to use the perturbing torque of the solar radiation pressure to stabilize the spacecraft, since it is mostly symmetrical about the solar panel ridges. This meant that the spacecraft must have its back facing the Sun at all times. A way of doing so is to for the spacecraft to follow the ecliptic, or the plane of the solar system. For an observer on the Earth, the Sun's path traces out the ecliptic against the fixed background stars [1]. This path is not a flat line, but it is a curved path due to the tilt of the Earth. Fig. 14 shows the right ascension and declination of the K2 observations. The curve that K2 campaigns follow is the ecliptic. Using solar radiation pressure, the Kepler spacecraft was able to accurately point in the orbital plane. K2 differs from Kepler in its attitude control and field of view scheme as a result [9].

1.4.2 Instrumentation

The instrument on the Kepler spacecraft is a 0.95 m Schmidt telescope. The telescope feeds the 94.6 million pixel CCD array, containing the Science and Fine Guidance sensor (FGS) CCD's. The local detector electronics (LDE) operates the CCD's and digitizes the output of 94.6 million pixels [5]. The LDE and the detector array assembly compose the Focal Plane Array Assembly (FPAA). In Fig. 16, the photometer is then the system containing the telescope and the FPAA.

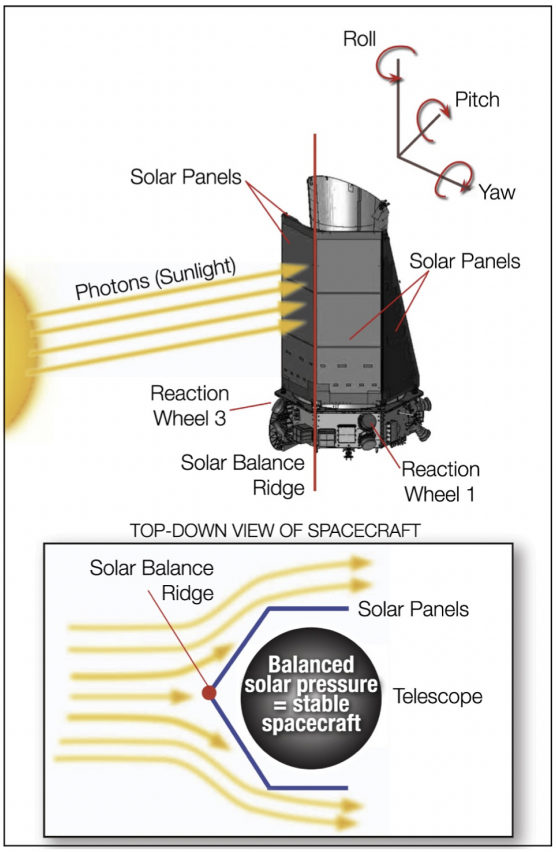


Figure 15: Illustration of the engineering solution for the K2 mission, using the remaining reaction wheels 1 & 3 to point the Kepler spacecraft to a field of view on the ecliptic plane. The roll axis of the spacecraft is balanced by the solar radiation pressure and small thruster firings at 6 hour intervals [9]. Image credit: NASA Ames

The wide field of view (FOV) Schmidt telescope has a primary mirror and a Schmidt corrector. Typical Schmidt-Cassegrain telescopes are spherical mirrors with a thin corrector lens to correct the spherical aberrations produced by the mirror in order to form a sharp image [1]. Fig. 17 illustrates the light path entering a simple Schimdt telescope. They are compound telescopes, where the optics include both reflection and refraction. The use of a spherical mirror allows for the telescope to achieve a wider field of view, which is the angular area of the sky seen through the eyepiece. The wide FOV makes them ideal for star surveys.

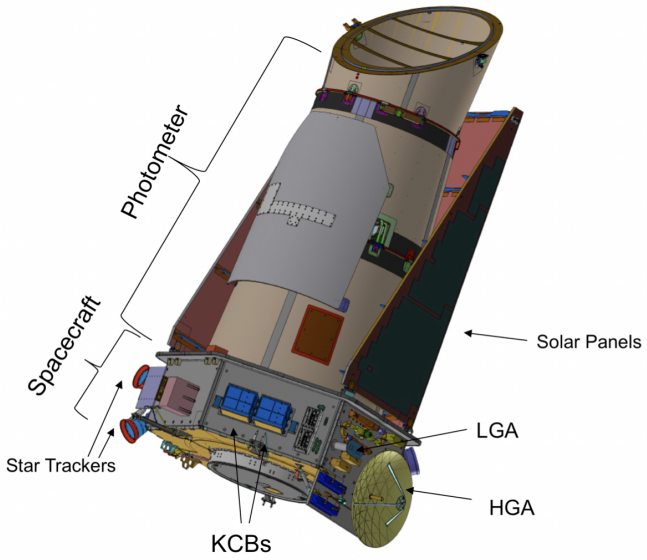


Figure 16: Kepler flight system, showing integrated photometer, spacecraft, and reaction wheels. KCB: Kepler Control Boxes, H/LGA: High/Low Gain Antennas. Image Credit: Kepler Instrumentation Handbook, NASA Ames

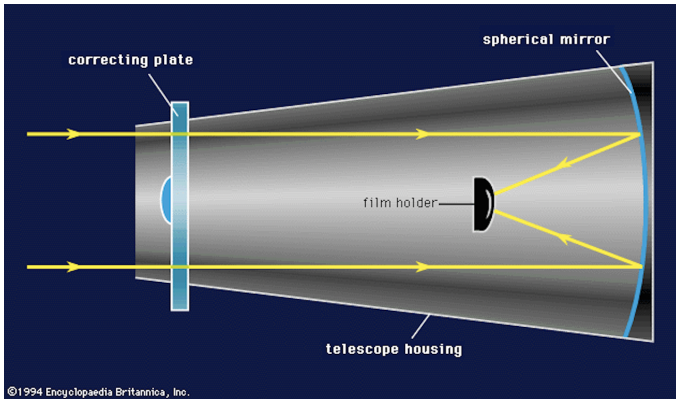


Figure 17: Lightpath through a Schmidt telescope. Image Credit: Encyclopedia Britannica

2 Methods

Photometric lightcurves can reveal details on stellar surfaces. As a planet traverses a star, the planet blocks light coming from the star, resulting in a dip in the lightcurve. Like sunspots, starspots are cooler, darker patches of the star. The light emitted from starspots will be less than the light emitted from areas without starspots. If starspots are present on the star, when a planet traverses in front of the starspot, less light will be blocked by the planet since there was less light to begin with. This results in in-transit variations of the lightcurve. In Fig. 18, when the planet covers a starspot, the amount of light available

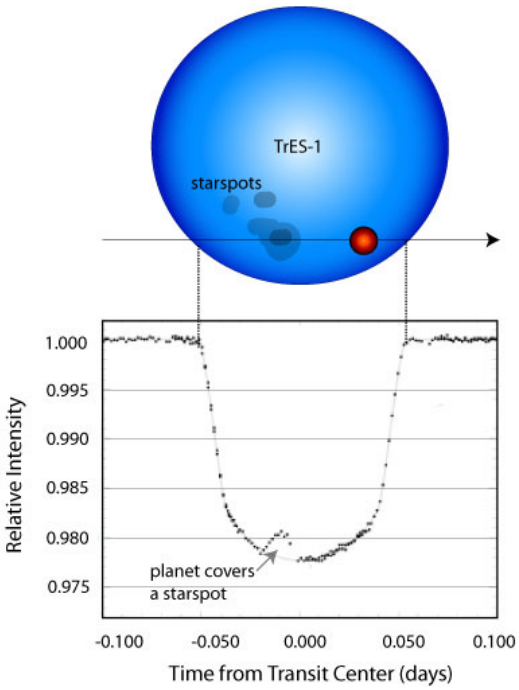


Figure 18: Hubble Space Telescope (HST) observations of the TrES-1 transit. Image credit: Brown et al., 2001

to be blocked by the planet is less, resulting in a bump in the transit dip. By looking for bumps in the transit lightcurves, starspot crossings can be observed. The program in this thesis aims to produce phase-folded lightcurves from a sample of K2 planet candidates.

Lightcurves are periodic. Phase-folded diagrams overlay the transits over the period of the star on the same lightcurve. Even though time is different for each lightcurve, the phase stays the same, thus by plotting flux against phase produces a single lightcurve with multiple transits. This is useful in observing variability in the transits. I wrote a program in Python that searches for stars that show in-transit starspot crossing events from the K2 data. To quantify variability, my program calculates the residual and performs the Kolgomorov-Smirnov test on the in-transit and out-of-transit samples of the residuals.

Fig. 19 is the lightcurve of a Kepler K2 object, EPIC 212351868. The plot is produced in Python, using the lightkurve package that extracts data from the Mikulski Archive of Space Telescopes (MAST). Fig. 20 is a zoomed-in image of the lightcurve. In the figure, the out-of-transit portion of the curve is sinusoidal. This is a result of starspots coming in and out of view due to the rotation of the star as Kepler observes the star over time. The sinusoidal pattern of the lightcurve is an indication of the presence of starspots, but the feature of interest is the bumps within the transit as the planet crosses in front of the star as that would allow the study of the starspot's latitude from the impact parameter.

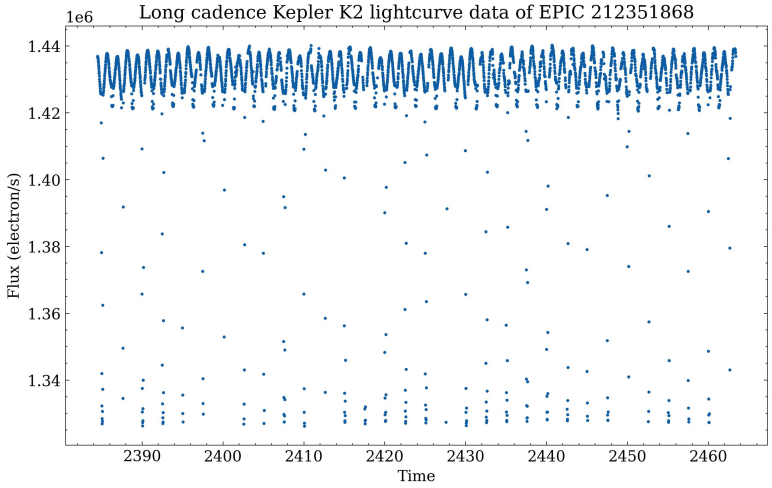


Figure 19: Lightcurve of EPIC 212351868. The data stretching down from the cluster of sinusoids are the individual transits that we want to study.

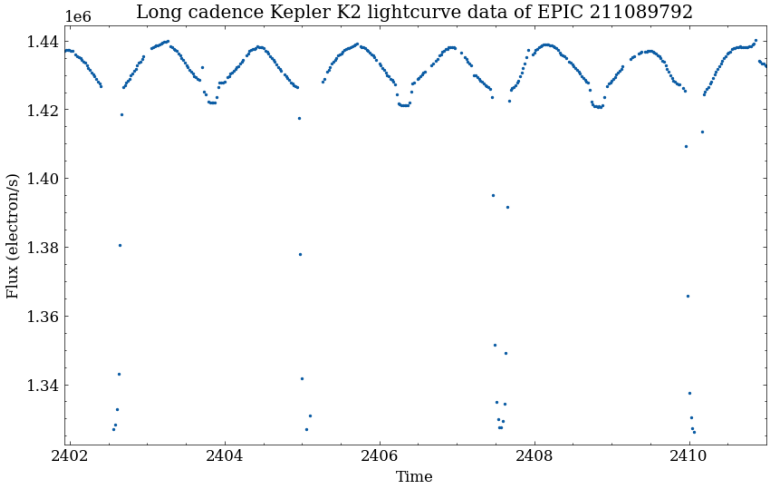


Figure 20: Zoomed in on the lightcurve of EPIC 212351868. The sinusoidal pattern in the out-of-transit portion is due to starspots coming in and out of view with the rotation of the star as Kepler stares at the patch of the sky over time.

2.1 Getting individual transits from lightcurve

The full list of K2 transiting planet candidates and their stellar parameters are obtained from Zink et al (2021). The list was derived using photometry from Campaigns 1-8 and 10-18 of the K2 mission. The catalog includes 747 unique planet candidates and 57 multiplanet systems, 366 of which have not been previously identified [17]. The table was downloaded from the VizieR Astronomical Server and used as an input for my program. The list was sorted by the candidate's planet-to-star radius in descending order, as we are interested in objects with deep transits.

The Python package, lightkurve, is a tool to search and download data from the Mikulski Archive for Space Telescopes (MAST) [8]. The public archive includes data for active and legacy missions, including Webb, Hubble, TESS, Kepler, K2, and more [10]. Given the object identification number from the catalog, lightkurve retrieves time-series brightness data of K2 objects. The long cadence images, with 30-minute observation times, are

selected. My program iterated through the first 196 objects with transit depths ranging from 0.2797 to 0.0498.

A test object, EPIC 211089792, was used for all preliminary analysis. The object was chosen because it has planet-to-star radius of 0.1388, indicating that its transits are fairly deep, and it has a high effective temperature of 6099 K. Table 1 presents the stellar properties of EPIC 211087972.

Parameter	**Object**
EPIC	211089792
Epoch (BJD)	2231.431407
Period (day)	3.258836
Duration (day)	2.424
Eccentricity	0
Planet-to-star radius	0.1388
Planet-to-star distance	8.39
Inclination (degrees)	84.5627431916225
Limb darkening coefficients	0.3354, 0.2949
Effective Temperature (K)	6099

Table 1: The properties of EPIC 211089792

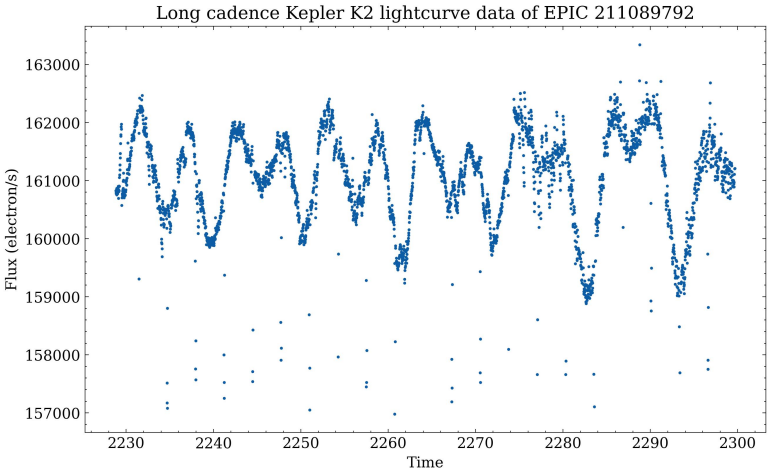


Figure 21: Lightcurve downloaded from MAST using lightkurve

The relevant stellar parameters provided in the table include object identification, epoch, period, duration, planet-to-star radius (ratio), and planet-to-star distance (major axis). These parameters are used as inputs into batman, a Python package that computes transit model values from the given parameters.

Some parameters that are required but are missing from the VizieR table are the limb darkening coefficients and inclination. Since limb darkening is related to temperature, the effective temperature of the K2 objects are used to estimate the limb darkening coefficients. From the table of existing limb darkening coefficients for Kepler objects, the K2 objects adopt the limb darkening coefficients of the Kepler object that is the closest match in temperature. The resulting model produced from the estimation did not produce outstanding deviations, so no further modifications were made to this estimation. The inclinations of the K2 candidates are calculated from the impact parameter using Eqn. 7.

To extract individual transits from the lightcurve data over a time range, we take the given epoch, or transit midpoint, of the transit, and add an integer number of periods to the epoch:

$$\text{transit midpoint} = n * \text{period} + \text{epoch} \tag{10}$$

The result is a set of values identifying the midpoints of each transit, which are appended to an empty vector in order normalize and calculate the model for each transit individually before creating a phase-folded diagram.

2.2 Normalize data on transit-by-transit basis

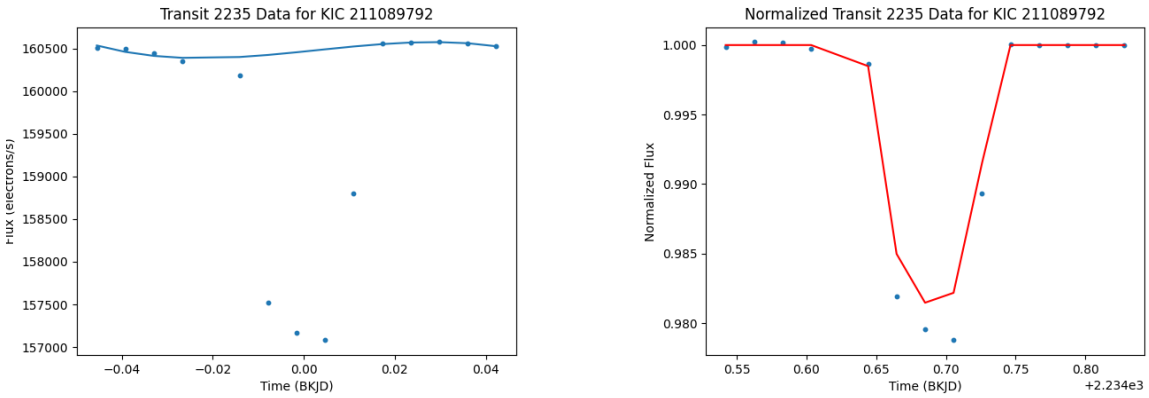
In order to compare to the model, individual transits need to be normalized. The K2 data are recorded in counts, whereas the model generates the relative flux, which plots the lightcurve values relative to one. Each transit, identified by its transit midpoint, is normalized through a second-order polynomial fit of the out of transit data. The fitted

values were obtained by using the polyfit function in Python on the out-of-transit flux and phase.

$$\text{normalized flux} = (\text{all flux} - \text{polyfit} + \text{peak})/\text{peak} \tag{11}$$

The goal of normalization is to remove the curvature in the out-of-transit portion and convert it to relative flux. In Fig. 21, it is visible that the out-of-transit flux for each transit reaches different min and max points. In order to account for this difference, the peak flux value of the total lightcurve is found. The difference between flux and the polyfit values produce a flat, normalized transit, but transits remain at different values depending on the out-of-transit values. By adding and dividing by the peak flux value, all transits are normalized on the same scale. Fig. 22 is an example of an individual transit. Fig. 22a is plotted with a second-order polynomial fit. After applying Eqn. 11, Fig. 22b is the normalized data plotted with the model.

All of the flux values are sorted, and the tenth largest value is used as the peak flux value, as the absolute maximum value can be an outlier. The same peak value is used for each normalization. The normalized flux is added to a column in the transit data table.



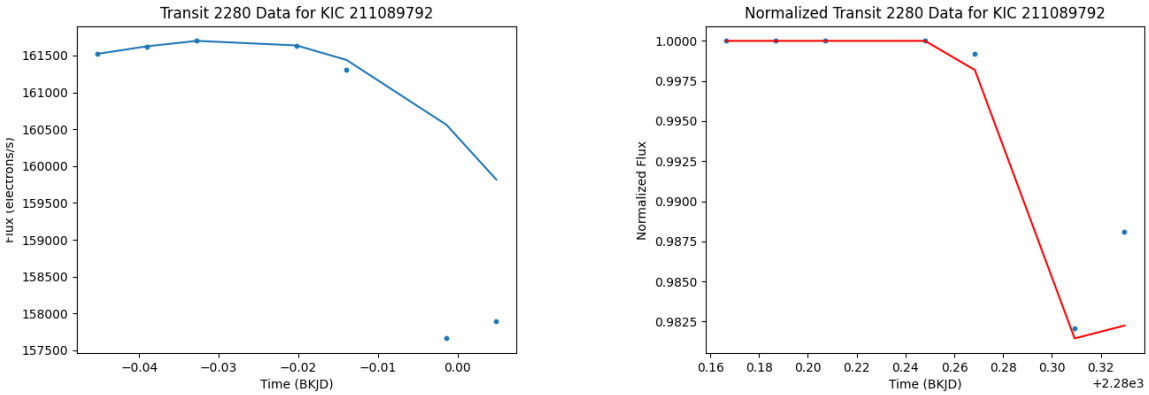
(a) Transit with polynomial fit

(b) Normalized transit with model

Figure 22: Example of a good transit

2.3 Rejecting bad data

Not all transits in the lightcurve data are useful in helping us identify starspot crossing features. Transits with missing data create normalizaitons that scatter the end result of the phase-folded plots of the object, deterring our ability to observe the residuals in detail, as in Fig. 24.



(a) Transit with polynomial fit

(b) Normalized transit with model

Figure 23: Example of a bad transit

There are four ways that a transit can be considered 'bad' in our case:

- if the data includes only half of a transit, as shown in Fig. 23
- if the χ^2 value of the data, calculated from Eqn. 12, is above a certain threshold that indicates the model is not a good fit
- if the quality flag in the original archive indicates an 'error' from the spacecraft at that data point
- or if there is no data at the expected time of the transit

To account for these errors, an overall flag is created to select indices of good transits for our phase-folded diagram. For the half transits, a loop was created to inspect whether

a transit had missing portions. The quality flag from the database is directly read as a criteria for the overall flag. If the transit is empty, the program moves onto the next transit. The χ^2 value is calculated for each data point and appended to a vector of the same length. A large value of χ^2 shows that the null hypothesis, where the probability that N_i is drawn from the population of n_i is unlikely [12].

$$\chi^2 = \sum_i (\frac{x_i - \mu_i}{\sigma_i})^2 \tag{12}$$

The likelihood of whether a good model is generated is of interest. In the program, x_i is the measured normalized flux, and μ_i is the model flux in Eqn. 12. Fig. 24 is an initial phase-folded diagram for EPIC 211089792, before data rejection. A number of bad transits for this object skewed the fitting and normalizing process, rendering the residual plots unhelpful.

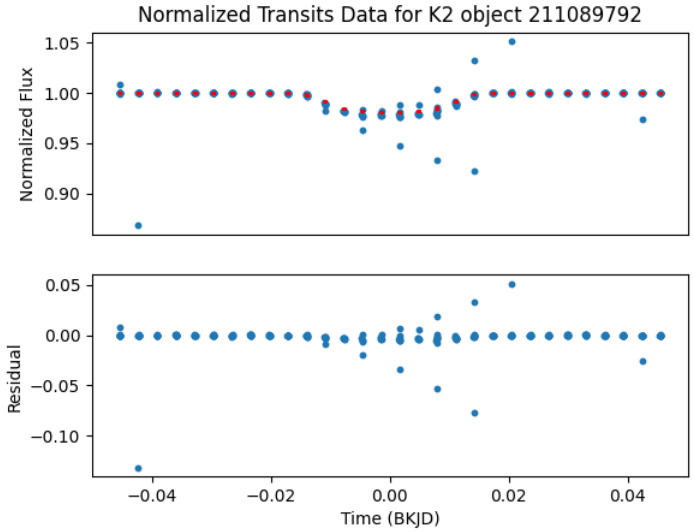


Figure 24: Bad transits induce scattering in phase-folded diagramr

Fig. 25 is the result after data rejection. The scattering has been removed, and the lightcurve resembles more to what we expect to see.

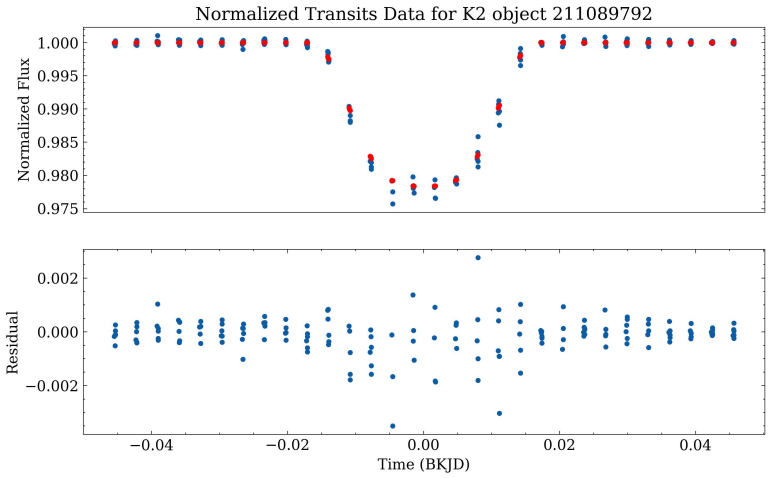


Figure 25: After data rejection, with planet-to-star-ratio adjusted to 0.15 for the model. The transit depth of the model was not matching the data very well, resulting in a shifted residual. The effect is still visible in the out of transit residuals.

2.4 Generate batman model

Batman stands for BAsic Transit Model cAlculatioN in Python. The package calculates the lightcurve for any radially symmetric stellar limb darkening law. The implementation is fast, as it uses an integration algorithm that improves the speed of models that are unable to be calculated analytically [6].

To use batman, the model is first initialized and the parameters defined. The same parameters from the data shown in Table 1, which define the shape of the transit, are used. The model is generated using the parameters and a given continuous time frame. For analysis, a model was generated using the given data points, as this creates a vector of the same length as the phase and flux data. The normalized transit data is plotted with the model values computed by batman, as shown in Fig. 26.

2.5 Using model and normalized transits to get residuals

The residual of the lightcurve is defined as the difference between measured data and model data. This difference illustrates any fluctuations between the measured and expected values, therefore highlighting data that may suggest starspot crossing features.

By comparing the actual data to the model data, we can see if the actual data deviates from the expected, which may be suggestive of starspot crossings. To produce the phase-folded diagram, each individual transit, with its phase and flux values, is appended to a data frame. Since the phase is the same for all transits, the plot of flux versus phase produces a single lightcurve overlaid with multiple transits.

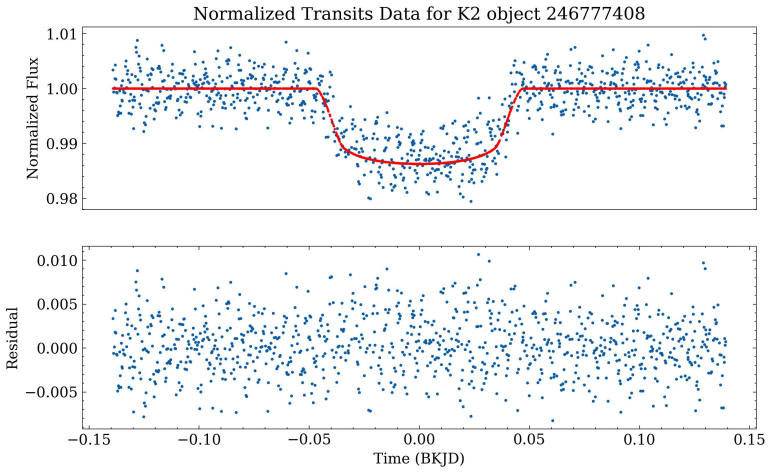


Figure 26: Example of a lightcurve with no starspot crossing events

3 Results

3.1 Statistics

To quantify how the in-transit residuals differ from the out-of-transit residuals, the statistical method is to conduct the Kolmogorov-Smirnov (K-S) test. This method calculates the maximum difference between two cumulative distribution functions, which are probability functions describing the probability that a variable will have a value less than or equal to the value of the probability density function at that point, as shown in Eqn. 14. The two distributions of interest for acquiring this statistic is the cumulative distribution functions of the in-transit and out-of transit residuals. If the in-transit residuals have a different distribution than the out-of-transit residuals, it may indicate an effect of starspots on the transit.

A separate program was written to create a histogram of the residuals of the in-transit and out-of-transit portions and calculate the probability density function and the cumulative distribution function. The probability density function is computed by the fraction of the number in each bin over the total number of points for each bin. The cumulative distribution function is then the sum of values up to each point in the probability distribution function. The K-S statistic is calculated as the maximum difference between two cumulative distribution functions. In SciPy, the ks_2samp function performs the test on two samples.

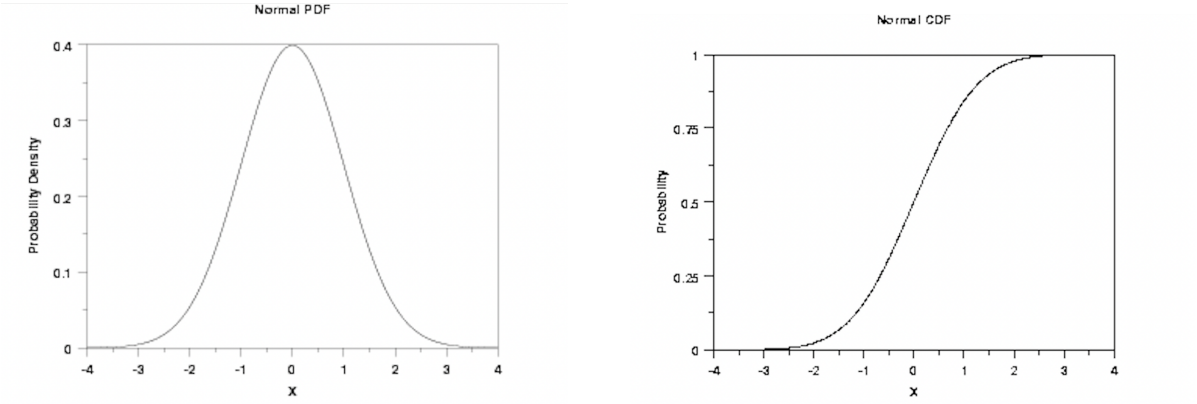
The probability density function (PDF) is the probability that a variate has a value of x [12]. It is used to gauge probability of a discrete outcome. For continuous distributions, the probability is expressed as an integral:

$$\int_a^b f(x)dx = P(a \leq X \leq b) \tag{13}$$

Cumulative distribution functions (CDF) are calucated by summing up the probability density function at a specific point. It shows the probability that the variable takes a value less than or equal to x:

$$\int_{-\infty}^{x} f(t)dt = P(X \leq x) \tag{14}$$

for $-\infty < x < \infty$ [12].



(a) Plot of a probability distribution function (b) Plot of a cumulative distribution function

Figure 27: Image credit: NIST

The K-S test is best used for distributions of functions of a single variable, where each data point can be defined as a single value, such as the flux of a star [12]. The statistics that K-S tests employ is the number D to denote the absolute maximum difference between two cumulative distribution functions. The maximum difference is represented by the black arrow in Fig. 28.

The K-S statistic allow us to compare one sample, $S_N(X)$, with the theoretical distribution, $P(X)$, or two samples, $S_{N1}(X)$ and $S_{N2}(X)$ [12]. The maximum difference, D, is found by:

$$D = \max_{-\infty<x<\infty} |S_N(X) - P(X)| \tag{15}$$

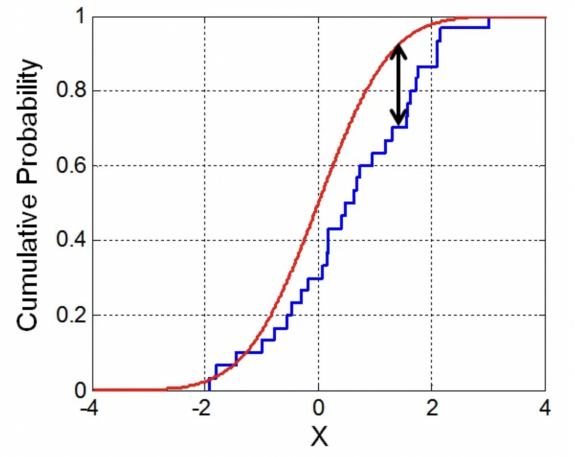
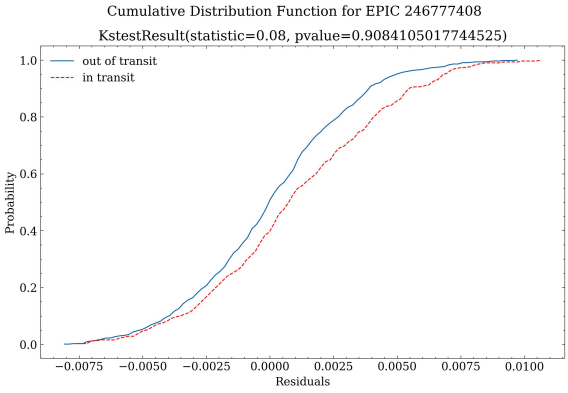


Figure 28: Example of K-S test performed on two samples. The statistic denoted by black arrow. The blue curve, $S_N(X)$ denotes one's data set, and the red curve, $P(X)$ denotes a known curve. Image Source: Wikipedia

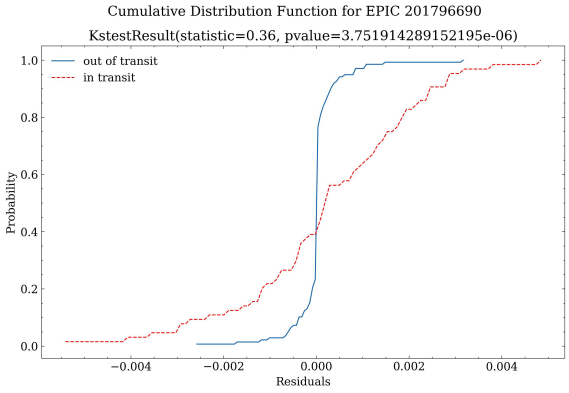
The two samples in our case is the in-transit and out-of-transit portions of each lightcurve. The K-S statistic is partially used to determine whether the in-transit distribution differs significantly from the out-of-transit distribution, which would suggest the appearance of starspot crossing for this transit.

3.2 Likely objects with starspot crossings

By visual inspection of the transit lightcurve and the KS value, a list of objects with noticeable starspot crossing features was curated. Out of 196 objects, 10 objects showed starspot crossing characteristics. The 10 objects were selected by visual inspection, as well as having a high K-S statistic of at least 0.15. Fig. 29a is an example of an object without starspot crossings. The K-S value is 0.08, and the shape of the in-transit and out-of-transit cumulative distribution functions are very similar. Fig. 29b is an example of an object with starspot crossing characteristics, with a K-S statistic of 0.36. The difference of 0.36 is visible in two cumulative distributive functions. This plot also does not show any odd,



(a) KS statistics for object in Fig. 26. Statistic=0.08, which is small, suggesting that the two samples are from the same distribution.



(b) KS statistics for EPIC 201796690, an object likely with starspot crossings. Statistics=0.36, considerably larger than EPIC 246777408

Figure 29: KS statistics of an object with and without starspot crossings

unexpected features.

There are also some objects that are not very clear. Some objects have multiple clusters of data. Some look like they have starspot crossing features but there are some scatter that could be outliers. Additional analysis, in conjunction with improving the rigor of my program's bad transit rejection process, is required in order to discern whether the features are related to starspot crossings, and that the fluctuations are not for other reasons such as noise.

The lightcurve of EPIC 211002562 in Fig. 30 display clear starspot crossing events. The residuals in the out-of-transit portion is much less than the residuals in the in-transit portion. The K-S test supports this, as in Fig. 31, the statistic is shown to be 0.39, the highest in our results in Table 2.

Object (EPIC)	KS	Ratio (d)	Teff (K)
211006562	0.39	0.1159	5826
246067459	0.37	0.0856	5605
201796690	0.36	0.0709	3749
246325464	0.34	0.0586	4940
204621597	0.33	0.1770	4103
206029314	0.29	0.1736	5474
201247497	0.29	0.0917	3699
246260670	0.26	0.0856	4459
212219881	0.17	0.1088	5372
205962305	0.15	0.0689	4386

Table 2: K2 Objects with starspot crossing features. The lightcurves and KS plots of these objects are in Appendix A.

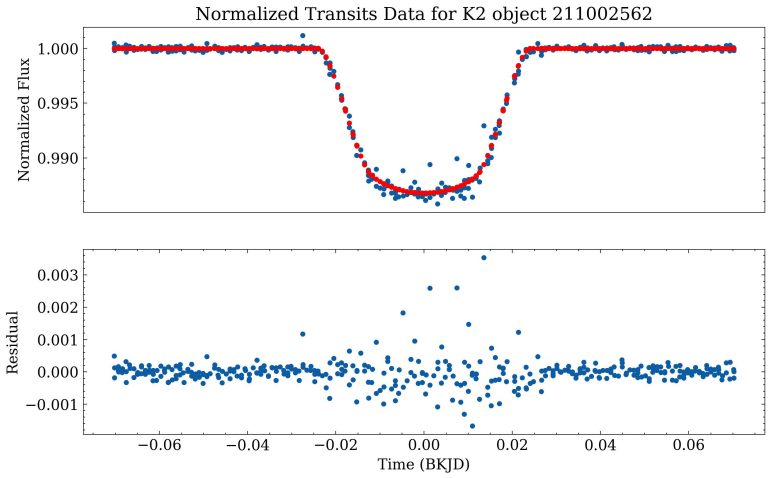


Figure 30: The residuals of EPIC 211002562 looks like there are starspot crossings during transits.

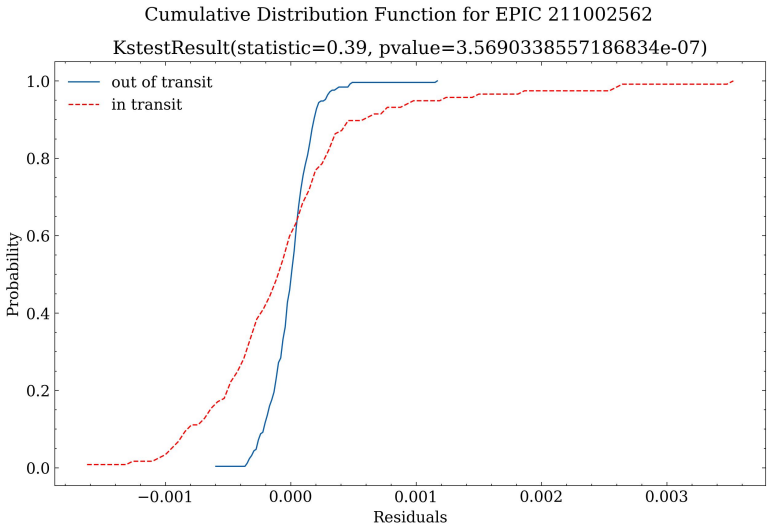


Figure 31: The K-S statistics of Fig. 30. The high K-S value of 0.39 suggests the presence of starspot crossings.

4 Conclusion and Future Directions

The goal of this project is to iterate through the list of K2 planet candidates and determine the objects with starspot crossing characteristics. The program developed during this project was able to generate lightcurves of 196 objects, with ratios ranging from 0.0498 to 0.2797, and compute the KS statistics of their in-transit and out-of-transit residuals.

More work is necessary to explore the correlation between stellar parameters and starspot crossing characteristics, as well as further clean up the data retrieved from MAST. For some K2 objects, there appear to be multiple sets of data. This affected the program's ability to obtain accurate residuals for those objects. For this project, my program skipped transits that appear more difficult to deal with. The odd features of some transits may be due to interesting reasons that weren't explored due to a shortage of time.

The method described by this paper could be explored with data from the Transiting Exoplanet Survey Satellite (TESS). For this project, initial difficulties in finding and un-

derstanding the planet candidate results for TESS, in addition to the shortage of time, limited our study to K2 objects. My program has a feature of selecting either the Kepler or K2 mission to retrieve data for. In the future, the program could be modified to work for TESS or other exoplanets mission, in order to produce a comprehensive list of objects with starspot crossing events using the transit method. The last figure, Fig. 32, shows the abundance of efforts to understand exoplanets, stars, and our place in the Universe.

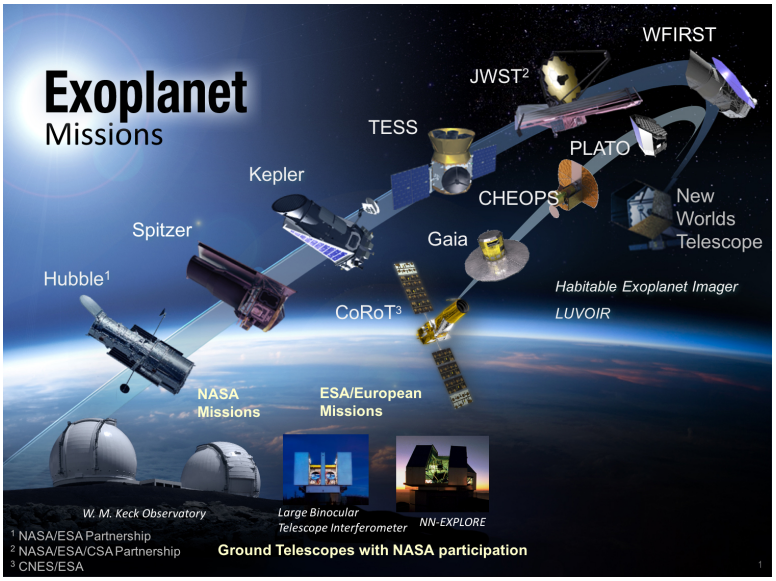


Figure 32: Exoplanet missions, from ground observatories to space telescopes. Image Credit: NASA

References

[1] D. Oesper D.S. Birney G. Gonzalez. *Observational astronomy*. Cambridge University Press, 2010.

[2] C.A. Haswell. *Transiting Exoplanets*. Cambridge University Press, 2010.

[3] S. Tremaine J. Lissauer R. Dawson. "Advances in exoplanet science from Kepler". In: *Nature* 513 (2014), pp. 336–344. DOI: http://dx.doi.org/10.1002/andp.19053221004.

[4] *Kepler In depth*. July 2019. URL: https://solarsystem.nasa.gov/missions/kepler/in-depth/.

[5] *Kepler: A search for terrestrial planets - plarchv1.stsci.edu*. URL: https://plarchv1.stsci.edu/files/live/sites/mast/files/home/missions-and-data/kepler/_documents/archive_manual.pdf.

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

[7] K.R. Lang. *The Cambridge Encyclopedia of The Sun*. Cambridge University Press, 2001.

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

[9] Kenneth Mighell and Jeffrey Van Cleve. *K2: Extending Kepler's Power to the Ecliptic - MAST*. URL: https://archive.stsci.edu/files/live/sites/mast/files/home/missions-and-data/k2/_documents/KSCI-19116-003.pdf.

[10] *Missions and data*. URL: https://archive.stsci.edu/missions-and-data.

[11] *More on Magnetic Fields*. URL: https://ece.northeastern.edu/fac-ece/nian/mom/magfields.html.

[12] William H. Press et al. *Numerical Recipes in C*. Second. Cambridge, USA: Cambridge University Press, 1992.

[13] Klaus G. Strassmeier. "Starspots". In: *The Astronomy and Astrophysics Review* 17.3 (May 2009), pp. 251–308. DOI: 10.1007/s00159-009-0020-6. URL: https://doi.org/10.1007/s00159-009-0020-6.

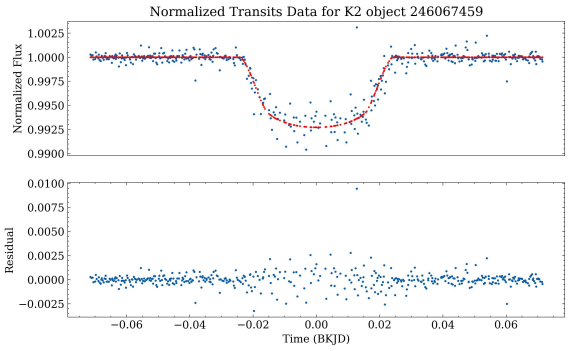
[14] *Sunspots and the Solar Max*. URL: https://earthobservatory.nasa.gov/features/SolarMax/solarmax_3.php.

[15] *The nearest neighbor star*. URL: https://imagine.gsfc.nasa.gov/features/cosmic/nearest_star_info.html.

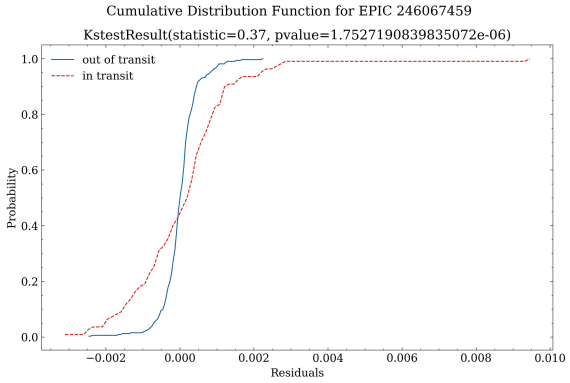
[16] Holly Zell. *NASA's SDO observes largest sunspot of the solar cycle*. Oct. 2014. URL: https://www.nasa.gov/content/goddard/sdo-observes-largest-sunspot-of-the-solar-cycle/.

[17] Jon K. Zink et al. "Scaling K2. IV. A Uniform Planet Sample for Campaigns 1–8 and 10–18". In: *The Astronomical Journal* 162.6 (Nov. 2021), p. 259. DOI: 10.3847/1538-3881/ac2309. URL: https://doi.org/10.3847%2F1538-3881%2Fac2309.

A Additional Figures

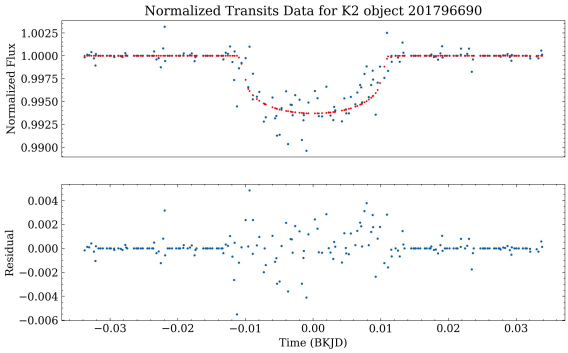


(a) Top panel: normalized transit with model, bottom panel: residuals

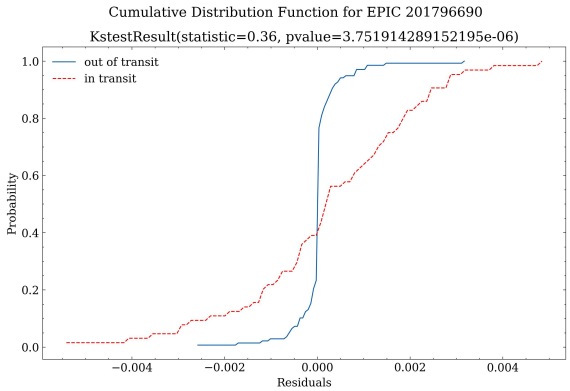


(b) KS statistics

Figure 33: EPIC 246067459

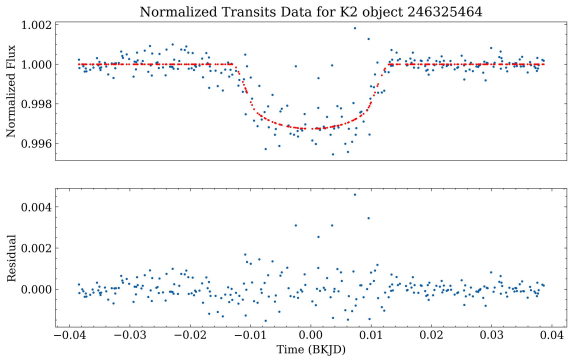


(a) Top panel: normalized transit with model, bottom panel: residuals

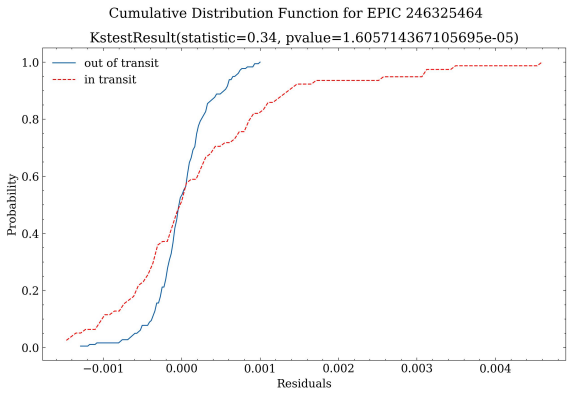


(b) KS statistics

Figure 34: EPIC 201796690

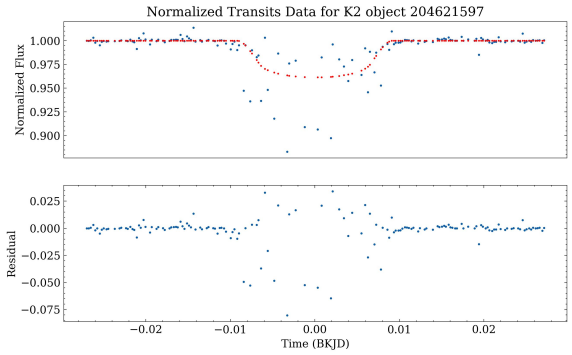


(a) Top panel: normalized transit with model, bottom panel: residuals

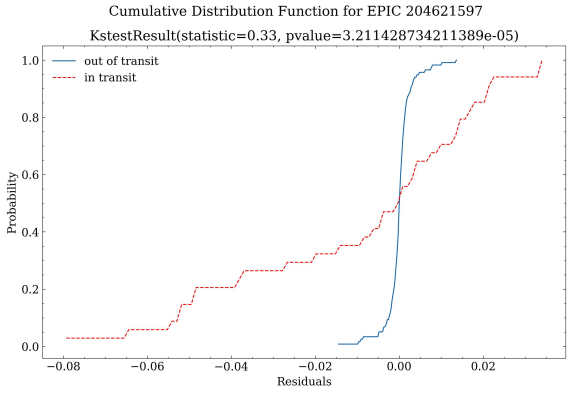


(b) KS statistics

Figure 35: EPIC 2463254640

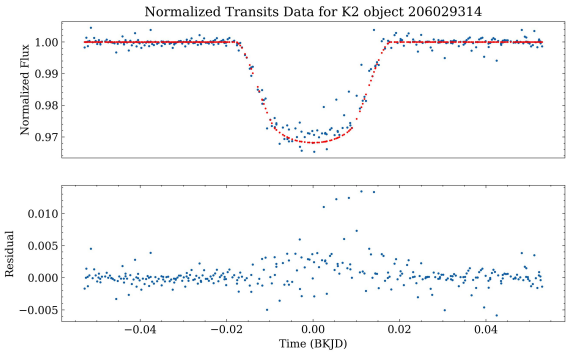


(a) Top panel: normalized transit with model, bottom panel: residuals

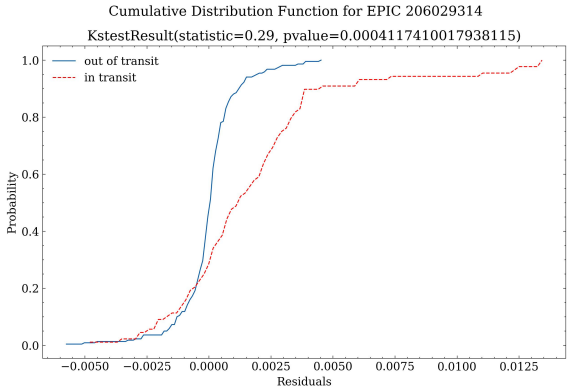


(b) KS statistics

Figure 36: EPIC 204621597

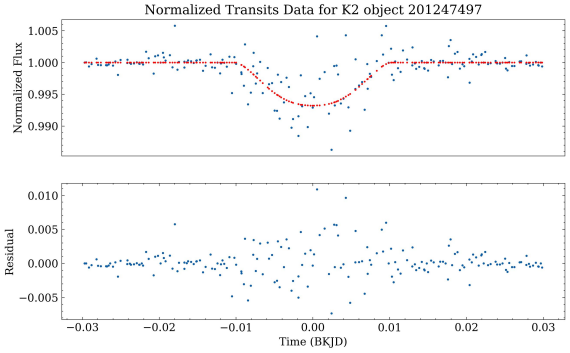


(a) Top panel: normalized transit with model, bottom panel: residuals

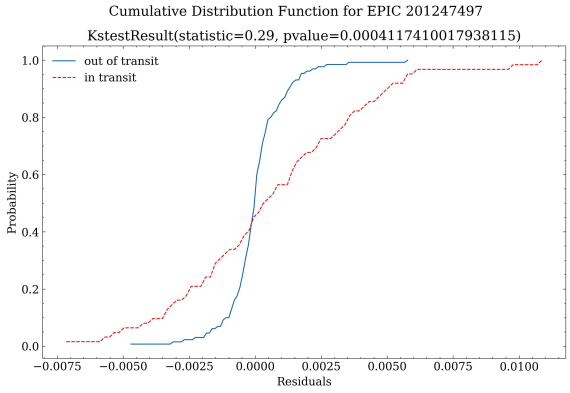


(b) KS statistics

Figure 37: EPIC 206029314

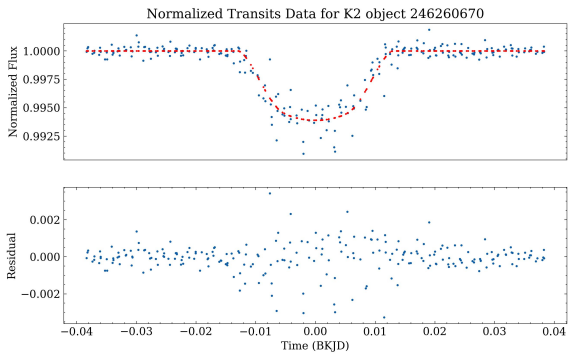


(a) Top panel: normalized transit with model, bottom panel: residuals

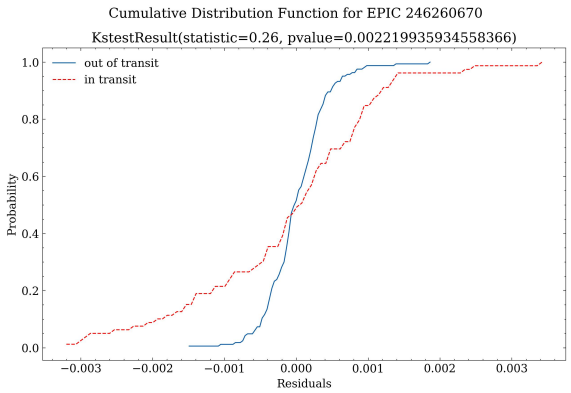


(b) KS statistics

Figure 38: EPIC 201247497

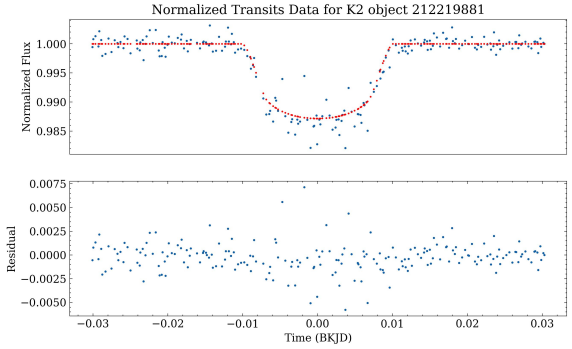


(a) Top panel: normalized transit with model, bottom panel: residuals

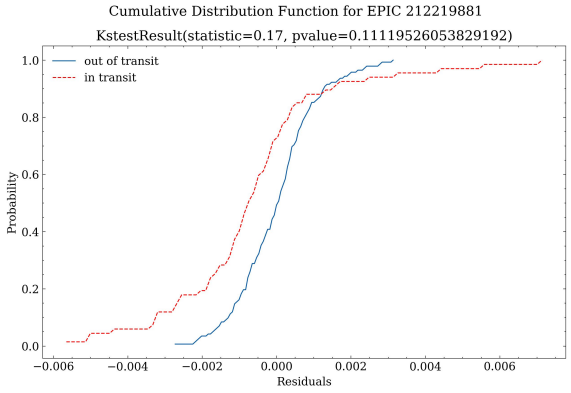


(b) KS statistics

Figure 39: EPIC 246260670

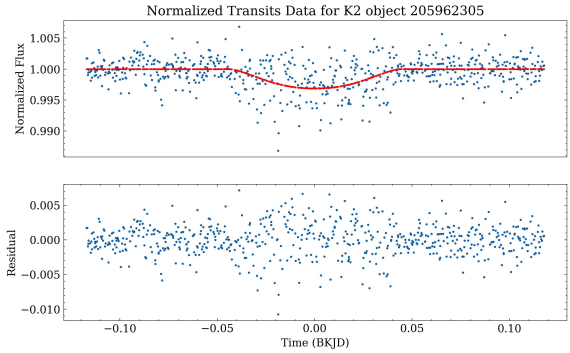


(a) Top panel: normalized transit with model, bottom panel: residuals

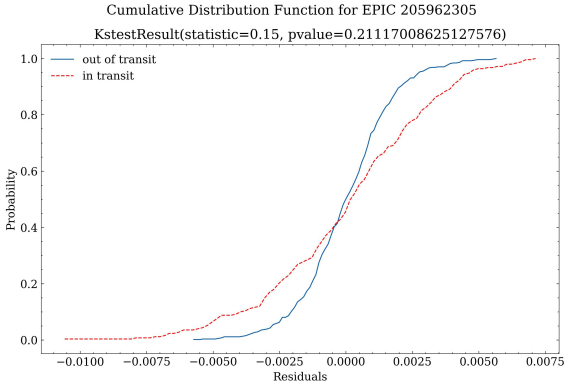


(b) KS statistics

Figure 40: EPIC 212219881



(a) Top panel: normalized transit with model, bottom panel: residuals



(b) KS statistics

Figure 41: EPIC 205962305

B Python code