Spring 6-10-2023

OPTICAL VARIABILITY OF GREEN PEA GALAXIES

Julissa Marie Sarmiento
JSARMIE3@depaul.edu

OPTICAL VARIABILITY OF GREEN PEA GALAXIES

A Thesis

Presented in

Partial Fulfillment of the

Requirements for the Degree of

MASTER OF SCIENCE

June, 2 0 2 3

BY

Julissa Marie Sarmiento

PHYSICS AND ASTROPHYSICS DEPARTMENT

College of Science and Health

DePaul University

Chicago, Illinois

ACKNOWLEDGEMENTS

I would like to begin by thanking my advisor at NASA, Dr. Sangeeta Malhotra, for welcoming me into her research team and assisting me every step along the way, enabling my research to reach a successful conclusion; this thesis is a testament to that. Next, I would like to thank Dr. Anuj Sarma for serving as chair of my thesis committee at DePaul, and especially for all the help getting my thesis into a final form that I am proud of. Thanks also go to Dr. Jesús Pando and Dr. Bernhard Beck-Winchatz for their generosity in giving their time to serve on my thesis committee and for their suggestions and guidance.

I would like to thank Drs. James Rhodes, Santosh Harish, and Mainak Singha for all the help they have provided during the entire course of my research. I really appreciate their help with everything, including how to navigate through basic steps in the analysis, advice on later steps, and helping me refine my presentation. I am grateful that they welcomed me into their research group, and I thank them for their patience in helping me learn about their research methods and procedures.

I would also like to thank everyone at the DePaul Physics and Astrophysics department that I have had the pleasure of working with over the past two years. I would like to thank all the faculty in the physics and astrophysics department that I have taken classes with and TA'd for; I have really enjoyed both learning physics and helping others learn physics with you all. I would also like to thank the other graduate students I have interacted with during my time at DePaul. I have been grateful for your help while doing classwork and homework, as well as your friendship outside of school. I would like to thank Mary Ann Quinn for always being so kind and helpful. I have really enjoyed my time at DePaul; the entire department is so welcoming and understanding and I am saddened to be leaving.

I would like to thank the Illinois Space Grant Consortium and the National GEM Consortium. They provided me with the opportunity to perform this research, and have allowed me to be able to focus on my research. These organizations not only enabled this thesis, but have set me on the career path that I have always wanted.

Finally, I would like to thank my family for their constant support and for being proud of me. I would especially like to thank my parents and my sister. I would like to thank my parents for always believing in me and supporting me no matter what. I would like to thank them for allowing me to pursue my dreams since I was a child and for doing everything in their power to support me in whatever ways they could. I would like to thank my sister, for being my greatest ally and best friend. She has provided me with so much support and advice (and the occasional cookie). She always has the most faith in me and has helped me in more ways than I could count. This thesis would not be possible without her.

TABLE OF CONTENTS

TABLE OF CONTENTS* – *Continued

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

In this thesis, I am investigating the optical variability of Green Pea galaxies (GPs). GPs are good analogs to high-redshift galaxies, enabling us to learn more about the first galaxies in the universe. One of their key properties is their strong emission lines, some of which indicate the presence of an active galactic nucleus (AGN). An effective way to identify AGN is to look for stochastic variability in the optical light from the galaxy. Finding AGN in these galaxies would help us learn more about the formation and evolution of the supermassive black holes that power AGN. In this thesis, I analyzed 317 observations of 216 GPs taken with the Transiting Exoplanet Survey Satellite (TESS). These data are comprised of nearly continuous observations of the GP on a 27-day timescale. I used these data to create light curves and used the inverse von Neumann ratio ($1/\eta$) to quantify the variability. Considering sectors with a value of $1/\eta > 0.588$ as having a high confidence of variability, 11 sectors showed indications of variability and 3 of those sectors had significant variability when compared to other nearby TESS pixels. Since those 3 sectors correspond to 2 GPs, we conclude that 2 GPs (GP# 41 and GP# 219) show significant optical variability on 30-day time scales, therefore likely containing AGN. This is supported by other independent methods for finding AGN such as mid-infrared colors and X-ray luminosity for one of these two GPs (GP# 219).

CHAPTER 1

Introduction

Green Pea galaxies (GPs) are a unique class of galaxies in our local universe. They are characterized by their compact physical size and their very strong [O III] emission lines. One of the most important aspects of these galaxies is that they have been shown to be good analogs to high-redshift (z) Lyman-alpha (Lyα) galaxies (Yang et al. 2016; Yang et al. 2017; Kim et al. 2021), which are some of the first galaxies in the universe. The ability of GPs to be good analogs for high-z galaxies is very beneficial. Since GPs are relatively local galaxies with low redshifts, they are easier to study than high-z galaxies. This allows us to use what we learn about GPs to try to have a better understanding of high-z galaxies and galaxy evolution.

In this thesis, I will be investigating the optical variability of GPs in order to look for Active Galactic Nuclei (AGN) in these galaxies. Most AGN have been found in galaxies with large masses ($> 10^{10}$ $M_\odot$), powered by supermassive black holes (SMBH) with masses on the order of $> 10^6$ $M_\odot$. The origin of these SMBH is not yet well understood. Finding AGN in lower mass galaxies can help us better understand the evolution of SMBH and their potential seeds. Since GPs are low mass galaxies, with masses between $10^8 - 10^{10}$ $M_\odot$ (Cardamone et al. 2009; Yang et al. 2017), we can use them to try to find intermediate mass black holes (IMBH), which would have masses $< 10^6$ $M_\odot$. GPs are also useful in finding the progenitors of SMBHs since they are analogs to high-z galaxies. So, they can give us a better understanding of what the first IMBHs looked like and maybe even what early SMBHs looked like. This can also help us better understand the formation and evolution of SMBHs and the galaxies around them.

In the rest of this chapter, I will provide background information on GPs and AGN. In Section 1.1, I will provide more info on GPs and elaborate on some of their key properties, including why they are considered as analogs to high-z galaxies. In Section 1.2, I will give some background specific to AGN and briefly talk about the methods used to find AGN. In Section 1.3, I will discuss why we believe there might be AGN in GPs. Finally in Section 1.4, I will discuss the goals of this thesis. Then, in Chapter 2, I will talk about the observations used in this thesis and the data reduction process. The results will be presented in Chapter 3. The results will be analyzed and discussed in Chapter 4. Finally, I will discuss the conclusions we can gather and the future work in Chapter 5.

1.1 Green Pea Galaxies (GPs)

GPs are a class of galaxies with very distinct properties. They were first discovered by participants of Galaxy Zoo (Cardamone et al. 2009). Galaxy Zoo is a citizen science project, where hundreds of thousands of volunteers helped classify over 10^6 images of galaxies observed with the Sloan Digital Sky Survey (SDSS). As part of the project, there were forums where users could discuss interesting objects. It was in one of these forums that these compact, round, and green objects were first discussed and given the name Green Peas. GPs got their distinct green color in the images because these galaxies have very strong emission lines, which were redshifted into the SDSS r band, which is colored green in the SDSS images. Cardamone et al. (2009) followed up on these objects, helping summarize their unique properties and marking the beginning of the research into these galaxies.

1.1.1 Properties of GPs

There are several properties that make GPs unique galaxies to study. One defining characteristic is that they are compact galaxies, meaning that they have very

small physical sizes. In general, the upper limit to their physical radius is < 5 kpc (Cardamone et al. 2009); however, these galaxies can appear even more compact when observed with different instruments. For example, recent images taken with the Hubble Space Telescope, which has sharper imaging than SDSS, have found that these galaxies have an upper limit of radii < 1.5 kpc (Yang et al. 2017; Kim et al. 2021). For comparison, the Milky Way Galaxy has a radius of roughly 15 kpc. Figure 1.1 shows some examples of GPs, taken in the UV with the Hubble Space Telescope. The figure demonstrates their compact nature.

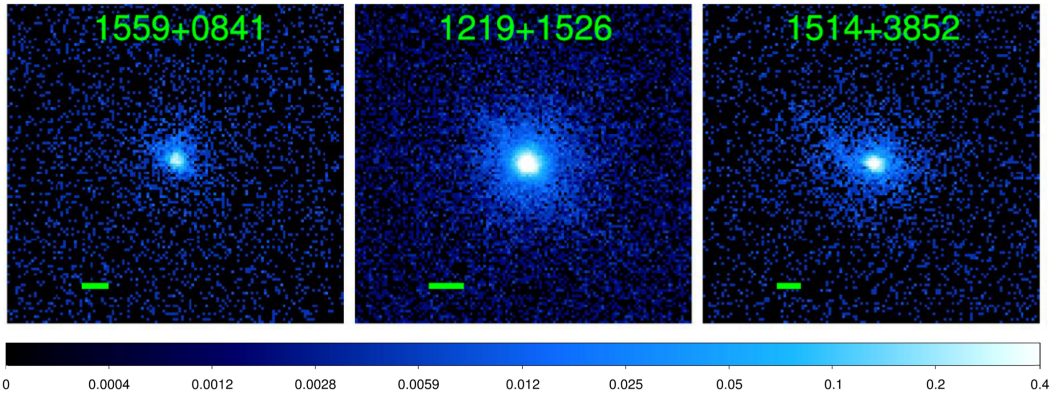

Figure 1.1: This figure shows three examples of GPs, observed in the UV with the Hubble Space Telescope. The color bar at the bottom of the image shows the count rate on a log scale. The count rate refers to the rate at which UV photons were collected from the source by the CCD pixels; so, the higher the rate the more photons were collected. The GPs are labeled with IDs, and the green bar in each panel is scaled to 1 kpc. Figure taken from Yang et al. (2017).

GPs are also low redshift galaxies. Redshift refers to the doppler shift of light towards the red part of the spectrum when the light originates from a source that is moving away from us. Since the universe is expanding, much of the light that arrives at the Earth is redshifted to some extent, especially from sources that are farther away. Since we know the rest wavelength, we can calculate how much the light has been redshifted. This is then used an indicator of distance or time. The more the light has been redshifted, the further away an object is, both physically and backwards in time. Galaxies located at redshift of $z \sim 3 - 6$ are considered

high-redshift galaxies. By comparison, GPs are very close. They are typically found at red shifts of $z \sim 0.1 - 0.4$.

GPs also have low masses, on the order of $10^8 - 10^{10}$ $M_\odot$ (Cardamone et al. 2009; Yang et al. 2017). By comparison, the Milky Way has a mass on the order of 10^{12} $M_\odot$. Interestingly, GPs also have high star formation rates (SFR). SFR is the rate of how many new stars are formed per year; it is given in terms of the total new stellar mass produced per year. GPs have a SFR on the order of 10 $M_\odot$ yr^{-1}, but some can have rates as high as 30 $M_\odot$ yr^{-1} (Cardamone et al. 2009). This is $10-100$ times higher than in typical compact galaxies at similar redshifts (Izotov et al. 2011). Because of their low mass and high star formation rate, this means that GPs also have very high specific star formation rates (sSFR), which is the star formation rate per total mass. For GPs, the sSFR can range from $\sim 10^{-9} - 10^{-7}$ yr^{-1}, which is very high. Merging galaxies, which typically have large outbursts of star formation have sSFR within $\sim 10^{-10} - 10^{-9}$ yr^{-1} (Izotov et al. 2011).

GPs are also extreme emission line galaxies. Emission lines are emitted by hot gas that have excited atoms. As the electrons in the atoms in the gas fall back down to lower energy states, they release photons of specific wavelengths, which we see as emission lines when we take the spectra of a galaxy. The particular wavelengths that are emitted are dependent on the element which was excited, as well as the particular transition states. Extreme emission line galaxies are galaxies that show very strong emission lines with large equivalent widths. The equivalent width refers to the width of the actual spectral feature, if the area under the curve were modeled as a rectangle; it is a way to measure the strength of the spectral feature. GPs have an especially strong [O III] λ5007 line with very large equivalent widths (up to $\sim$ 1000 Å); this strong emission line is one of the main characteristics of a GP (Cardamone et al. 2009). In fact, the [O III] line is the emission line that is responsible for their green color in the Sloan Digital Sky Survey images that were used in the Galaxy Zoo project.

GPs also have low metallicities. The metallicity (Z) of a galaxy refers to the amount of 'metals' present in the galaxy. Metals in this case refer to all the elements other than hydrogen and helium. Typically, the metallicity of an object is determined by looking at the ratio of the emission lines from its spectra. While the abundance of each element can be calculated separately based on the spectra, a general indication of the metallicity of a galaxy is usually presented as the oxygen abundance, given as 12 + log(O/H). Galaxies with 12 + log(O/H) < 8.5 are considered low metallicity galaxies (Jiang et al. 2019). GPs typically have 12 + log (O/H) ~ 8.1 and have been found to have metallicities ranging from $7.2 \leq 12 + \log$ (O/H) ≤ 8.6 (Izotov et al. 2011; Jiang et al. 2019). A metallicity of 12 + log (O/H) ~ 8.1 is equivalent to 0.2 $Z_\odot$, where $Z_\odot$ represents the metallicity of the Sun. This means that GPs have a metal abundance equal to roughly 20% of the metal abundances in our own Sun. So, these galaxies have very low metallicities.

1.1.2 GPs as analogs of High-Redshift Galaxies

While the properties of GPs discussed above are themselves interesting, a very important aspect of GPs is that they share many of these properties with high-redshift galaxies located at $z \sim 3 - 6$. These high-z galaxies are some of the first galaxies in the universe. In particular, GPs share many properties with high-z Lyα emitters. Lyα photons are produced in regions with lots of star formation. Usually, it is difficult for Lyα photons to escape galaxies, since dust can be a major inhibitor in their escape. There are still many questions regarding these Lyα emitters, especially the mechanism for the Lyα escape. However, it has been shown that the compact size of the high-z Lyα emitters is a key component to the Lyα escape (Malhotra et al. 2012; Kim et al. 2021). It has also been suggested that Lyα emitters might be responsible for the reionization of the universe, which is when much of the neutral opaque gas in the universe was ionized and the universe became transparent. Thus, studying Lyα emitters is crucial to our understanding of the universe.

There are several important properties GPs and Lyα emitters share. Both GPs and Lyα emitters are very compact, with Lyα emitters having a radius $\lesssim 1.5$ kpc (Kim et al. 2021). Like GPs, Lyα emitters have high sSFR ($\gtrsim 10^{-8}$ yr^{-1}) and low masses ($5 \times 10^8 - 10^{11}$ M$_\odot$; Kim et al. 2021). Most importantly, both galaxies show strong Lyα lines and very large Lyα equivalent widths (Yang et al. 2016; Yang et al. 2017). This is significant because even though there are other classes of local objects that share some properties with Lyα emitters, none have similar Lyα equivalent widths, which is an important property for the high-z Lyα emitters (Izotov et al. 2011). These shared properties allow us to use GPs as good analogs for these galaxies. This is a great advantage since they are significantly closer than the high-z Lyα emitters, thus they are much easier to observe. So, we can use GPs to try to learn more about these high-z Lyα galaxies.

1.2 Active Galactic Nuclei (AGN)

As mentioned earlier, one particular aspect of GPs I am interested in investigating is the presence of AGN in GPs. It is widely accepted that most galaxies have black holes at their center. For galaxies with masses $\gtrsim 10^{10}$ M$_\odot$, these black holes are known to be SMBH (Baldassare et al. 2018). SMBH, or supermassive black holes, have masses $\geq 10^6$ M$_\odot$ (Padovani et al. 2017). For some galaxies, these SMBH are actively accreting material onto the black hole. This causes these galaxies to emit large amounts of energy across the entire electromagnetic spectrum (Padovani et al. 2017). Some even have jets of material or lobes visible in the radio that extend far beyond the host galaxy. These galaxies are considered active, and their nuclei are called AGN.

AGN are thought to be comprised of several different components. The different wavelengths of light, as well as their subsequent variability which will be discussed in Section 1.2.1, come from different regions in the AGN. While AGN emit across the entire spectrum of electromagnetic radiation, I will focus my description on X-ray,

optical, and infrared (IR) wavelengths. The driving component of an AGN is the SMBH, which is located at the center. It is surrounded by an accretion disk. The accretion disk is comprised of material that is actively falling into the central black hole. Some of the X-rays that are emitted come from the inner most regions of the disk, possibly even from the corona around the black hole (Padovani et al. 2017). There is also contribution to the X-ray luminosity from the jets of those AGN that have such jets. Optical and UV radiation are also emitted from the accretion disk, also powered by the accretion onto the SMBH. Around the accretion disk is a dusty torus, which is where the IR light radiation is emitted. This light is the result of light from the accretion disk heating the dust particles in the dusty torus, which then emit in the infrared. In addition to the optical continuum emission from the accretion disk, there are also optical emission lines that can be detected for AGN. Broad emission lines are thought to be emitted from within the boundary of the inner wall of the dusty torus, from photoionized gas around the accretion disk; this gas is being ionized by the radiation emitted from the accretion disk (Padovani et al. 2017). Narrow emission lines come from the large regions above and below the plane of the dusty torus and accretion disk (Padovani et al. 2017). Figure 1.2 shows an image depicting the structure of an AGN, labeled with the different regions discussed.

1.2.1 Finding AGN

Even though all AGN generally have the same structure, they are selected by a wide variety of different techniques, such that there are several different 'types' of AGN. (Padovani et al. 2017). There are many theories as to explain why there are many different types of AGN that look so different. One of the most prevalent theories to explain the different 'appearances' is that they are just oriented differently in our line of sight. For example, if the orientation of the AGN is such that the dusty torus is directly in the line of sight between the observer and the SMBH, much of

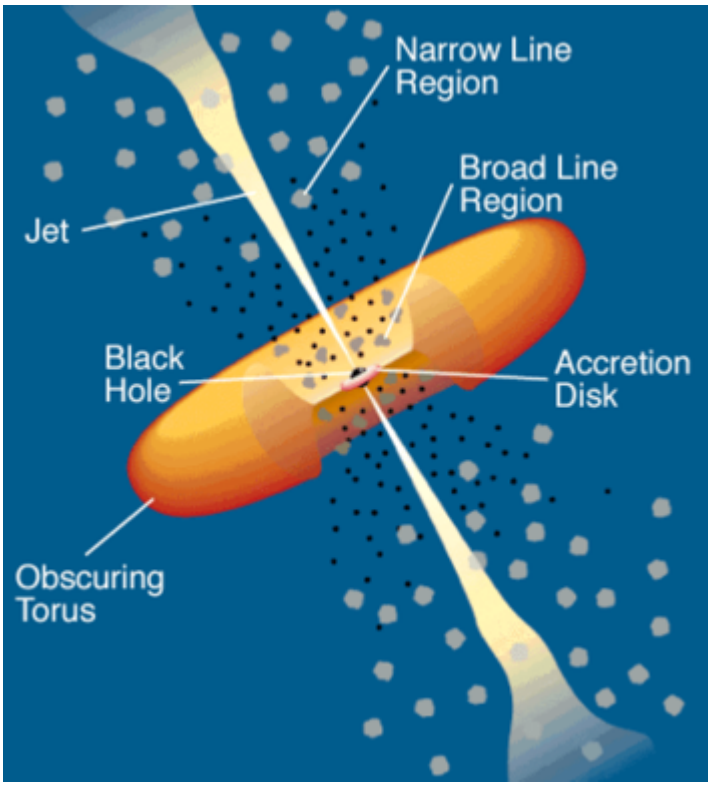

Figure 1.2: Labeled diagram of an AGN. Optical and X-ray light are thought to come from the accretion disk surrounding the SMBH. X-rays are also thought to come from the jets. IR radiation is thought to come from the dusty torus. Optical emission lines are believed to come from several regions. Broad emission lines are thought to be emitted from the region between the inside of the torus and the accretion disk, while the narrow emission lines are thought to be emitted from above and below the plane of the accretion disk. Modified from Figure 1 of Urry & Padovani (1995).

the optical light will be obscured, and the broad emission lines will be much weaker, so many of the optical selection techniques might not work. As such, there are several different methods that can be used to detect and verify AGN. One of the most effective ones is by looking at the emission lines of the galaxy. AGN typically have strong emission lines, and the emission line ratios can be used to identify AGN. For example, comparing [O III] λ5007/Hβ to [N II] λ6584/Hβ is one of the most common ways to select for AGN (Baldwin et al. 1981). In addition to being used to find AGN, emission lines are helpful as they can be used to figure out properties like the mass of the central black hole. However, as mentioned earlier, these can sometimes be at risk of being obscured by the dusty torus. In addition to emission lines, looking at the X-ray luminosity is another very effective way to find AGN. A low luminosity AGN can produce X-ray luminosities on the order of $\sim 10^{42}$ erg/s and higher luminosity AGN can produce X-rays $> 10^{44}$ erg/s (Padovani et al. 2017). AGN are the only sources that produce such high X-ray luminosities. Galaxies rarely produce X-ray luminosities above 10^{42} erg/s. Thus, using the X-ray luminosity is a great way to distinguish an AGN from a quiet, non-active galaxy.

Another widely used method of detecting AGN is by looking at the variability of the light (Padovani et al. 2017; Baldassare et al. 2018; Harish et al. 2023). There are several wavelengths where variability might indicate an AGN. The variable light from AGN is stochastic and can vary on a wide range of time scales, from hours to years (Padovani et al. 2017). The stochastic nature of the variability is due to the fact that the matter is accreting onto the SMBH in a stochastic, aperiodic matter. Since this accretion is taking place in the innermost regions of the accretion disk, where most of the X-ray light originates, this means that over time variability will be observed in the X-ray. In addition to this X-ray variability, AGN also display optical and UV variability that also originates from the accretion disk. Some of this variability is from reprocessing the X-ray variability, but it is not the only source of the optical variability, and the exact mechanisms are not yet known (Padovani et al. 2017). The time scale of this optical variability can be from days to months,

and sometimes there can be variability on longer timescales, on the scale of months to years. Like the optical emission lines, this optical variability can be obscured by the dusty torus. Finally, AGN also exhibit variability in the MIR. This light originates in the dusty torus, and its variability is believed to be powered by the optical variability from the accretion disk.

Any one of these methods alone can provide evidence to be able to classify an object as an AGN. There are advantages and limitations to each of the detection methods. And they can each pick out different types of AGN. However, different methods can be combined to help verify that a galaxy is an AGN.

1.3 AGN in GPs

Even though AGN are typically found in brighter, larger galaxies, there is some indication that AGN might be present in GPs. The strong emission lines that are characteristic of GPs can indicate hard ionizing sources like AGN. For example, GPs show hard ionizing lines like He II λ4686 and [Ne III] λ3869. They also have high line ratios, like [O III]/[O II], that also might indicate an AGN. Additionally, AGN candidates have been identified in GPs in the mid-IR (MIR). Harish et al. (2023) looked at a sample of GPs in the MIR and were able to identify AGN candidates. They were able to find 31 AGN candidates using the MIR colors, 2 of which also displayed stochastic MIR variability.

1.4 Using Optical Variability to Find AGN in GPs

For this thesis, I will be using the optical variability to try to find AGN in a sample of GPs. For GPs, using optical variability is more likely to yield results than trying to use the optical emission line selection rules. For example, [O III] is a common line used in AGN selection, but it is challenging to try to use that line to find AGN

in GPs. This is because GPs are characterized by their strong [O III] lines and large equivalent widths; the [O III] line in GPs is saturated with star formation and is therefore not a good discriminant to find AGN. Moreover, it has been shown that in general emission line selection rules, like those described by Baldwin et al. (1981), are less effective in galaxies with low metallicities and high star formation rates (Trump et al. 2015). Using variability of the light is a method that is better suited for low mass galaxies (Baldassare et al. 2018; Baldassare et al. 2020). Therefore, looking for stochastic optical variability would be a good way to try to identify AGN in GPs, since it is a common phenomenon in AGN and is not at risk of being contaminated by any stellar processes. Observing in the optical will allow us to look at shorter time scales, on the scale of days to months. Optical variability has proven successful in identifying AGN on both short timescales (Treiber et al. 2022) and long timescales (Baldassare et al. 2018); and it has been shown to be effective in identifying AGN in low mass galaxies (Baldassare et al. 2020). Thus, the goal of this thesis is to try to identify AGN candidates by observing the optical variability of GPs.

CHAPTER 2

Observations and Data Reduction

I am investigating the optical variability of GPs in order to find AGN. In this chapter, I will describe the observations of the GPs used in this thesis and describe reducing the data. In Section 2.1, I will describe the telescope from which the data comes from and why it can be used to find AGN. In Section 2.2, I will discuss how the sample of GPs was selected. In Section 2.3, I will discuss how I reduced the data, describing the main data processing package I used and the process I used to reduce the data.

2.1 Transiting Exoplanet Survey Satellite (TESS)

The data examined in this thesis were taken with the Transiting Exoplanet Survey Satellite (TESS). TESS is a space telescope in a highly elliptical orbit, coming as close as ∼17 Earth radii and reaching as far away as ∼59 Earth radii. It is an optical telescope; the detectors have a bandpass of roughly $600 - 1000$ nm. It was launched in 2018 with the goal of finding transiting exoplanets through a near all-sky survey. Its primary 2-year mission was to observe the majority of the north and south ecliptic hemispheres, excluding the area of the sky closest to the ecliptic, which is the plane of the Earth's orbit around the Sun. Its primary mission commenced on July 25th, 2018, and ran until July 4th, 2020. TESS is continuing to observe in its extended mission which involves re-observing the ecliptic hemispheres and parts of the ecliptic.

TESS observes a large patch of sky continuously for 27 days. Each 27-day observing cycle is called a sector. TESS has a ∼14-day orbit, and it completes its orbit twice

during each sector, downlinking data to the Earth when it reaches its closest point. The term sector is also used to refer to the specific patch of sky that TESS observed during that particular observing cycle. Therefore, a TESS sector can refer to both the actual observing cycle and the part of sky that was observed at that time. Each sector is assigned a number in the order in which it was observed. TESS has a very large rectangular field of view, $24° \times 96°$, which is achieved by four $24° \times 24°$ CCD cameras in a horizontally stacked orientation. The pixels on the TESS CCDs are very large, with a width of $21''$. By comparison, SDSS has $0.396''$ pixels. However, the tradeoff is that TESS has very high precision photometry. Because of the geometric constraints of mapping a rectangle onto a sphere, there are several portions of the sky that are observed in multiple observing cycles. In other words, the sectors overlap. Additionally, the areas around the south and north ecliptic poles were continuously observed during years 1 and 2, respectively, by having one camera during each sector in their respective years centered on the pole.

TESS provides two main data products from its pipeline. One data product is 2-minute cadence target pixel files, which are low cadence data that were taken for exoplanet candidate targets. The other data product, which is what I used in this thesis, is full frame images, which have a 30-minute cadence during TESS's primary mission. The full frame images are images of the entire $24° \times 96°$ field of view, allowing the user to gather data for any point in the sky that was observed during that sector. Both data products are processed through the TESS pipeline, which returns calibrated data. The calibrated data is simply the raw data that has had various systematic effects and cosmic rays removed (Vanderspek et al. 2016).

2.1.1 Using TESS for AGN

TESS has made great advancements in identifying transiting exoplanets. But although the primary role and design of TESS is to search for exoplanets, TESS has been used to observe other objects (Vallely et al. 2019; Burke et al. 2020; Treiber et

al. 2022). Not only does TESS have very high precision photometry, but it is also continuously observing a large section of the sky. This makes it an excellent telescope to use to observe different types of objects over a period of time. In fact, there have been studies that have shown that TESS can detect the optical variability of AGN. Burke et al. (2020) used TESS to observe the AGN in NGC 4395. While the AGN in the dwarf galaxy NGC 4395 had already been discovered and studied for many years, Burke et al. (2020) were able to demonstrate the ability of TESS to find AGN. They were able to demonstrate the optical variability of the AGN on a 27-day timescale. Treiber et al. (2022), however, used TESS to try to find AGN in a very large sample of galaxies. They were able to find 29 AGN candidates through their optical variability, 18 of which had not been identified before. Therefore, despite its name, I aim to use TESS to try to find optical variability in GPs on a 27-day timescale.

2.2 Sample Selection

The sample I used in this thesis comes from a larger sample of 1004 GPs. This sample was selected from SDSS Data Release 13. These GPs were selected because they were classified as 'star-forming' or 'starburst' galaxies in SDSS, have strong [O III] λ5007 and Hβ lines with signal-to-noise (S/N) > 5 and large equivalent widths, and are compact ($r < 3''$; Jiang et al. 2019). From this larger sample, we selected GPs that were isolated, meaning that they did not have an object as bright or brighter than the GP within $21''$. We chose the value of $21''$ since that is the size of a single TESS pixel; this will help reduce contamination from other sources that would be in the same pixel. This left us with 345 GPs. Then, we eliminated GPs that had an object brighter than 14th magnitude within $90''$. This was done to further reduce contamination from bright sources. This left us with a sample of 321 GPs.

2.3 Data Reduction Process

2.3.1 The Data Processing Package Lightkurve

In order to gather and process the data from TESS, I used the python package Lightkurve (Lightkurve Collaboration et al. 2018). Lightkurve features several built-in python functions and classes that help create light curves from data and process them. A light curve is a standard representation of data used in astronomy in which the intensity of a source is plotted as a function of time. While the features of the Lightkurve package can be used with any user-input data, it was specifically built for gathering and analyzing data from the Kepler and TESS missions in order to find exoplanets. However, its features can still be used for a wide variety of science, including trying to find AGN.

2.3.2 Photometry and Scattered Light Subtraction

I used the `search_tesscut` function in the python package Lightkurve to search through all the available TESS data for the 321 GPs. This function uses the Mikulski Archive for Space Telescopes (MAST) service TESScut (Brasseur et al. 2019). At the time, most of the data available through the TESScut service were only from the primary mission of TESS. Even though some GPs had data that had been observed during the extended mission of TESS, I did not include that data since the cadences were different in the extended mission and only some GPs had the new data. So, for consistency, my thesis will be limited to data observed during the TESS primary mission, which corresponds to the first 26 sectors and took place over 2 years. From the list of 321 GPs, only 217 GPs were observed in at least one TESS sector in the primary mission. This is the final sample of GPs I will be studying in this thesis. Several GPs were observed in multiple sectors, so in total, I analyzed 369 sectors.

With this list of 217 GPs observed with TESS, I used the Lightkurve functions to gather and process the data. I used the `search_tesscut` function in Lightkurve again to search for all the sectors a single GP was observed in, and then downloaded the data for all the sectors that GP was observed in. For the data download, Lightkurve allows users to select a quality data mask that will ignore cadences with certain quality flags which were assigned in the data pipeline. I chose a 'hard' quality mask; such a mask would exclude data that were flagged for things like scattered light or cosmic rays in collateral pixels. This mask is slightly stricter than the recommended mask in the Lightkurve documentation, as it can throw out good data. However, with the default mask, some bad data still slips through, and it was causing some errors in the rest of the data analysis. Most of these bad data were a result of scattered light from the earth, which is very noticeable the closer TESS gets to the earth. While there is still a lot of scattered light present in the data, the 'hard' data mask gets rid of the worst.

The data were downloaded as 5-pixel $\times$ 5-pixel cutouts from the full frame image, following which I used the `interact_sky` function to display the cutout and the location of the provided coordinate. I used this interactive image to visually determine the location of the GP in the cutout in order to determine where to place an aperture. This aperture tells the program where the object of interest is located. This is an important step that will be used in the aperture photometry step later on in the data reduction process. Once I determined which pixel in the cutout contained the GP, I custom-placed the aperture on that pixel. For all of the GPs and in every sector, I used a one-pixel aperture. If a GP was located near the boundary between two pixels, the aperture was placed on the pixel that contained most of the GP. In many cases, the GP was located in the center pixel; however, there were several instances where the GP was actually located in an adjacent pixel. If a GP was observed in multiple sectors, an interactive image was created for each sector and the aperture was placed in the optimal location for that sector. In other words, the apertures were custom-placed by hand for each sector of data I looked at.

After the aperture was created, I then created a light curve for each GP and each sector using the `to_lightcurve` function in Lightkurve. This function takes the downloaded data, automatically performs simple aperture photometry on the data, and returns a light curve. It uses the aperture I had created earlier for each specific sector. Aperture photometry is a method used to try to get the 'true' flux values from a source by subtracting the background flux, which could have contamination from nearby objects as well. When performing aperture photometry, the background flux calculated from the pixels outside the aperture is subtracted from the flux of the object within the aperture. Thus, the `to_lightcurve` function generates a light curve that is already background subtracted.

Even after the photometry was performed, there was still a lot of scattered light, so the next step was to correct for that scattered light. The method I used is detailed in the Lightkurve documentation but is repeated here. I used a linear regression method to get rid of the scattered light. The linear regression aims to get an idea of the general trends in the entire cutout and subtract that light from the light curve of our object. Lightkurve has several functions that allows users to do this. First, I created an array of the flux of all the pixels outside of the aperture over the entire time period. Then I converted that array into a design matrix, which is a python class in Lightkurve that makes it easier to use several of its correction methods. Then I used a principal component analysis function to reduce the matrix to 5 principal components. I then appended a column of ones to this design matrix, which is necessary for the regressor corrector to be able to fit the light curve properly.

The next step is to do the actual linear regression. First, the light curve that was created earlier is turned into a regressor corrector class, another python class in Lightkurve. Converting the light curve to this class will allow me to use the Lightkurve methods on it to get rid of the scattered light. Then that light curve is corrected with the design matrix; this is done automatically by a Lightkurve method that finds the best-fit correction for the light curve using the design matrix I provided. This method returns a corrected light curve. However, because the

original light curve had a lot of scattered light, this "corrected" light curve that was generated by setting to the mean level of the original light curve still contains scattered light in it. So, I needed to perform a few additional steps to get rid of that remaining scattered light. I used the model of the background that was created in the previous steps to get a model of the scattered light. However, the background model might have negative values of flux, which would not be possible since that would imply that the scattered light lowered the flux of the source, rather than added to it. These negative values are probably because the model was being set to the mean level of the original light curve, which will have a higher flux than the lowest flux value in the cutout. So, I calculated the 5th percentile of the model background and added it back to the model background, which would get rid of any negative fluxes. Then I subtracted this model background from the uncorrected light curve to get our final corrected light curve. This gives us our final, corrected light curve.

CHAPTER 3

Results

In this thesis, I am investigating the 27-day optical variability of Green Pea galaxies using TESS. In Chapter 2, I described the observations that were used in this thesis, and I described the processing of the data to create the light curves. In this chapter, I will go into more detail about creating the light curves for the GPs. In Section 3.1, I will describe how I created the light curves and the additional measures I took to get rid of bad data. I will also describe our limitations with GPs found in multiple sectors. In Section 3.2, I will walk through some sample light curves, presenting examples of light curves both with and without obvious optical variability.

3.1 Creating the Light Curves

Using the process described in Section 2.3.2, I created light curves for the 217 GPs. Since several GPs were observed in multiple sectors, this meant that I had a total of 368 light curves, one for each sector a GP was observed in. Unfortunately, even after the data cleaning described in Section 2.3.2, there were some sectors with bad flux measurements: they had many negative fluxes or fluxes very close to zero. Typically, sectors had flux measurements on the order of ~ 100 electrons/second. Sectors with many negative fluxes are due to bad data. Sectors with median fluxes close to 0 are also due to bad data since they would have many negative fluxes for the median to be near zero. To avoid using data from these bad sectors, I eliminated any sector where the median flux was less than 1. This also meant that I had to eliminate one GP whose only sector had these bad values. This left me with light curves for 317 sectors, which represented 216 GPs.

Another step I took was to remove specific sections of bad data from the final light curves. It was not until the actual light curves were inspected that it was clear that this step needed to be done for several GPs. As can be seen in the light curve figures below (Figures 3.1 - 3.4), all of the light curves, with the exception of light curves using data taken during Sector 03, have a gap in the middle. This gap represents the time during which TESS was near the earth to downlink the data. Because this is when the scattered light is the worst, the 'hard' data mask I used usually masks most of the data taken during this time. However, there were some instances where a GP sector had data in this gap. These data were usually very noisy and had a lot of outliers. Since these data were observed when TESS was closest to the earth, they were mainly comprised of scattered light and were considered as bad data, likely not from the source. Therefore, I decided to remove any data that were present in the data gap referenced above. I removed the days of data corresponding to that gap, including a few shoulder days, since that is when TESS was closest to the earth. In order to ensure consistency among my data, I removed the same dates of data from *all* the GPs observed during those dates. In other words, the same dates of data were removed across all GPs that had data taken during those dates.

3.1.1 Multi-sector GPs

Another important aspect to consider is the fact that several GPs were observed in multiple sectors. Therefore, we could theoretically observe the optical light of some of the GP across two or more months. However, it is only possible to analyze each sector on its own. There is a known issue with TESS where the relative flux between sectors changes significantly. Therefore, it is not possible to compare the flux of the same object in different sectors, since there will be a large artificial change in flux. So, I will be analyzing each sector on its own, even if the GP was observed in multiple sectors. This also means that I will only be able to look at the variability of these GPs on a 27-day timescale.

3.2 Light Curves

With these considerations in mind, I can create the finalized light curves for the 317 sectors. I created a simple, corrected light curve, with just the flux measurements versus the date. With this simple light curve, I then created normalized light curves. I normalized all the flux values in each sector to the median flux value in the entire sector. Then I plotted a running median. To do this I binned the data in 6-hour time bins, to smooth out the data. Then I selected the median data point in the 6-hour bin and plotted that point, creating a binned light curve. In addition to the running median, I also plotted an error shadow. I generated the error shadow by using the error from creating the binned light curve to plot a blue shadow extending from (median − error) to (median + error). Examples of these light curves are shown and described in Section 3.2.1 and Section 3.2.2 below.

3.2.1 Examples of Sectors Displaying Optical Variability

Of the 216 GPs examined for this thesis, 2 were determined to be variable using the von Neumann ratio described in Section 4.1.1. Figures 3.1 and 3.2 shown below are examples of light curves that were determined to display variability.

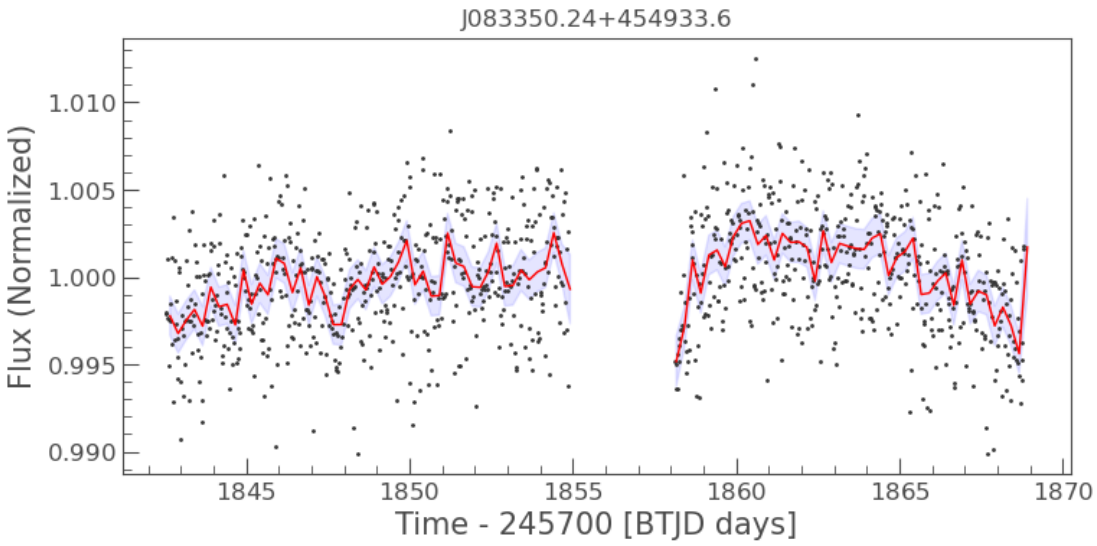

Figure 3.1: Light Curve for the GP with SDSS designation J083350.24+454933.6 referred to as GP# 41 in this data set. On the vertical axis is the normalized flux value, normalized to the median. On the horizontal axis is the time in Modified TESS Julian date, which is the current modified Julian date minus 245700. The black dots are each individual TESS observation. The solid red curve is the running median, binned with 6-hour bins. The blue shadow is the error shadow.

Figure 3.1 shows the light curve for GP# 41. This is one of the GPs that have a high von Neumann ratio that is interpreted to be an indication of variability (as discussed in Section 4.1.1). As can be seen, the overall light curve is fairly smooth, with a gradual general increase in flux before slowly decreasing again. There appears to be slow change in the flux of the GP over this time period. Even ignoring the large upward spike at the end of the time gap that is seen near Julian Day 1858 on the graph, there is a gradual increase, then decrease in flux over the course of this sector.

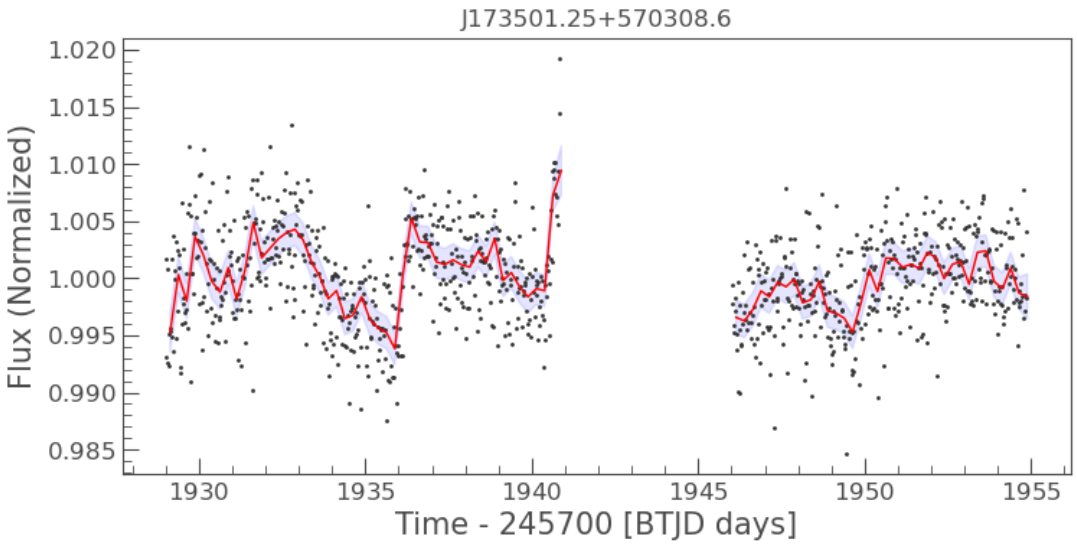

Figure 3.2: Light Curve for the GP with SDSS designation J173501.25+570308.6 referred to as GP# 219 in this data set. On the vertical axis is the normalized flux value, normalized to the median. On the horizontal axis is the time in Modified TESS Julian date, which is the current modified Julian date minus 245700. The black dots are each individual TESS observation. The solid red curve is the running median, binned with 6-hour bins. The blue shadow is the error shadow.

Figure 3.2 shows the light curve for GP# 219 taken during one sector. This is also one of the GPs that have a high von Neumann ratio (discussed in section 4.1.1). Even visually, one can see that there is some stochastic variability in the light curve. There is clearly a slight upward trend in the light curve before there is a large dip that increases again before the data gap. There is a large spike in flux right before the data gap, at around Julian Date 1940, but that can be ignored, since TESS was approaching the Earth by that point. After the data gap there is a very slight upward trend to the data, with a very brief dip.

3.2.2 Examples of Sectors without Obvious Optical Variability

Of the 216 GPs examined for this thesis, 214 were determined to exhibit no substantial variability. Figures 3.3 and 3.4 are light curves that did not display any variability.

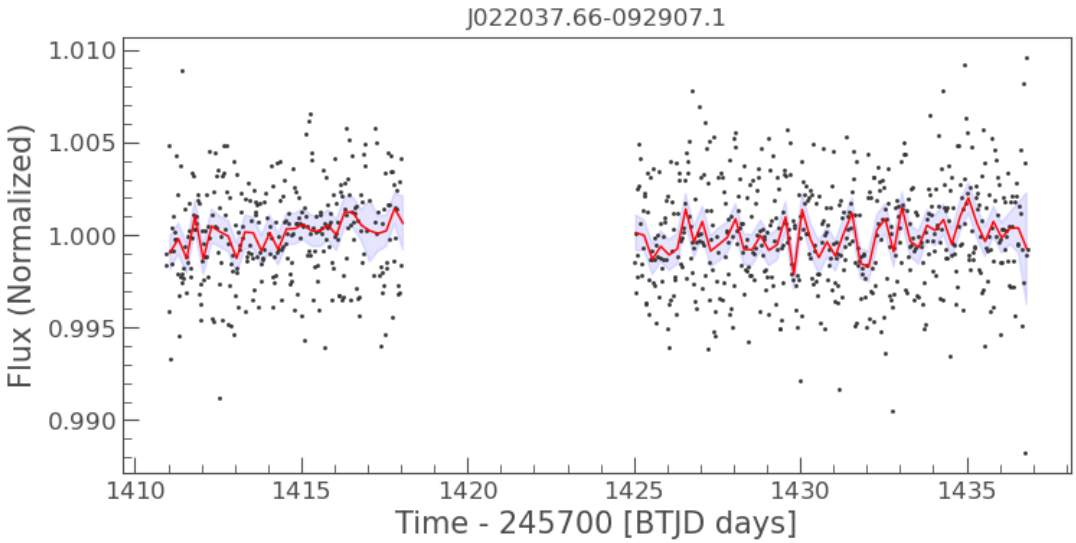

Figure 3.3: Light Curve for the GP with SDSS designation J022037.66-092907.1 referred to as GP# 18 in this data set. On the vertical axis is the normalized flux value, normalized to the median. On the horizontal axis is the time in Modified TESS Julian date, which is the current modified Julian date minus 245700. The black dots are each individual TESS observation. The solid red curve is the running median, binned with 6-hour bins. The blue shadow is the error shadow.

Figure 3.3 shows the light curve for GP# 18. This GP did not show variability, which can be seen in the light curve. Even though the running median looks very jagged, it is a result of the general noise and spread of the data and not because of any actual variability in the flux from the source. The overall trend of the data is very smooth, centered around the normalized value of 1. So, this GP does not show much, if any, variability in this sector.

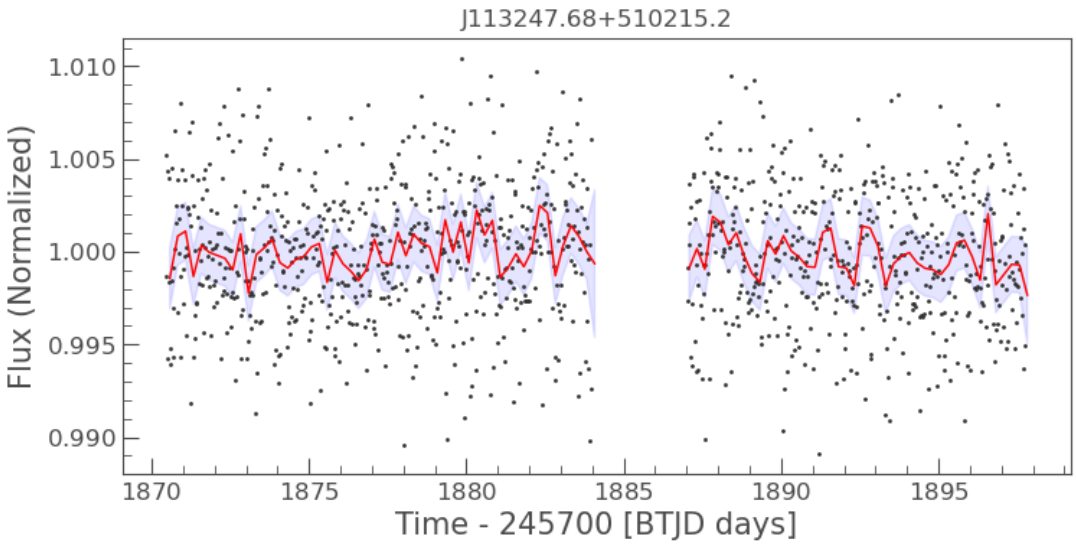

Figure 3.4: Light Curve for the GP with SDSS designation J113247.68+510215.2 referred to as GP# 96 in this data set. On the vertical axis is the normalized flux value, normalized to the median. On the horizontal axis is the time in Modified TESS Julian date, which is the current modified Julian date minus 245700. The black dots are each individual TESS observation. The solid red curve is the running median, binned with 6-hour bins. The blue shadow is the error shadow.

Figure 3.4 shows the light curve for GP# 96. This GP also did not show much variability. The running median is very jagged, with a fairly large error shadow, but still the overall trend over the course of the light curve is fairly steady and is centered around 1.

CHAPTER 4

Discussion

In this thesis, I am investigating the optical variability of Green Pea galaxies. In Chapter 2, I discussed the methods and techniques used in the inspection and processing of the data. I discussed and presented the resultant light curves in Chapter 3. In this chapter, I will discuss the quantitative analysis that I used to identify variability in the light curves of these galaxies. In Section 4.1, I will discuss the optical variability of these light curves. I will describe the variability index I chose (the inverse von Neumann ratio), why I chose it, and then demonstrate how I calculated it for this sample. I will compare the variability index to the brightness of the GP. Then I will compare the variability index of the GP to the variability index of the other pixels in the TESS cutout, and discuss my interpretations. Then in Section 4.2, I will go over other ways of looking for AGN and how they relate to this sample.

4.1 Optical Variability

The goal of this thesis is to find if the sample of GPs observed with TESS and selected as described in Chapter 2 have optical variability. There are several different ways to quantify such optical variability, and several different variability indices one could use. For example, one could look at the standard deviation, look at the covariance, or look at χ^2. One of the challenges with trying to quantify the optical variability is trying to figure out the best method for a specific dataset. Some methods are better suited for observations with only a few data points, and some are better suited for observations with a large number of data points. One also needs to take into consideration if the expected variability occurs on scales longer or shorter than the sampling time.

Sokolovsky et al. (2017) compared different variability detection techniques to try to find a good general-purpose technique. They looked at 18 different techniques to detect variability and compared their performance on 7 different data sets with a wide range of properties. Their test data sets contained both observed and simulated data and included not only a wide assortment of variable objects but also non-varying objects. They used two indices to quantify the accuracy of the detection technique, Completeness (C) and Purity (P):

$$C = \frac{\text{Number of selected variable objects}}{\text{Total number of confirmed variable objects}} \tag{4.1}$$

$$P = \frac{\text{Number of confirmed variable objects selected}}{\text{Total number of objects selected}} \tag{4.2}$$

They were able to use these indices since they knew beforehand which sources in their dataset were variable. The Completeness C measures if the technique is able to select all of the objects that are confirmed to be variable in the sample. The numerator is the number of objects that the technique selected as variable, and the denominator is the total number of confirmed variable objects in that dataset. If all of the confirmed variable objects were selected by the technique, then the numerator and denominator of 4.1 would be the same and then C would have a value of 1. Conversely, if no variable objects were selected by the technique, C would have a value of 0. The Purity P measures if the objects the technique selects are indeed variable; it can be thought of a measure of the amount of false positives selected. The numerator is the number of objects that were selected by the technique that were confirmed variable objects. The denominator is the total number of objects the technique selected. If the technique only selected confirmed variable objects, then P would have a value of 1. But if the technique only selected false positives, P would have a value of 0.

Sokolovsky et al. (2017) used these two indices to calculate a fidelity score (F), which is the harmonic mean of C and P, given by:

$$F = \frac{2(C \times P)}{(C + P)} \tag{4.3}$$

The fidelity score F has a value of 1 when all confirmed variables were selected by the variability detection technique and there were no false positives. Conversely, F has a value of 0 when none of the confirmed variables are selected. Thus, variability indices with higher values of F were better at selecting variable objects. They found that two indices, the von Neumann ratio, η (von Neumann 1942), and the interquartile range (IQR), had consistently high F values across all the test data sets. They used the inverse of the von Neumann, $1/\eta$, instead of η. Using the inverse of the ratio does not affect the statistic in any way, it just makes it a bit easier for humans to pick out variability: the larger value of $1/\eta$ is, the more variable the object. Since they suggested that correlation-based indices, such as $1/\eta$, were more efficient for data sets with hundreds of data points, I decided to use $1/\eta$ for my analysis. Thus, for this thesis, I will be using the inverse of the von Neumann ratio to quantitatively analyze the light curves.

Sokolovsky et al. (2017) did note two things, however. First, they noted that $1/\eta$ can only detect variability on timescales longer than the time between observations. This is not a problem for my dataset since the sampling time is 30 minutes and we expect variability to be on timescales of days or longer. They also noted that using any variability index alone is not enough to justify variability, and that this was best followed up with visual inspection.

4.1.1 von Neumann Ratio

The von Neumann ratio can be thought of as the 'smoothness' of the light curve, so its inverse indicates the presence of variability in the light curve by telling us its departure from smoothness. The (inverse) von Neuman ratio can be calculated as:

$$\frac{1}{\eta} = \frac{\sigma^2}{\delta^2} = \frac{\frac{1}{N-1}\sum_{i}^{N}(m_i - \langle m \rangle)^2}{\frac{1}{N-1}\sum_{i}^{N-1}(m_{i+1} - m_i)^2} \tag{4.4}$$

where N is the total number of observed points, m_i and m_{i+1} correspond to the magnitudes of the i^{th} and $(i+1)^{\text{th}}$ observed points, respectively, and $\langle m \rangle$ is the average magnitude of the entire light curve. Even though the von Neumann ratio does not take into account the errors of the measurements it is still accurate for picking out variability.

As described in Sokolovsky et al. (2017), there isn't a specific pre-determined value of the ratio that would indicate that the source was definitively variable. Instead, this ratio can pick out which sources are more variable relative to the other sources, which we can then investigate further. The 'cutoff' point to distinguish between not variable and potentially variable sources will depend on the dataset itself.

4.1.2 von Neumann Ratio of the GPs

To calculate the $1/\eta$ value for each of the light curves, I first took the flux measurements for each sector and converted the fluxes into magnitudes. I used the standard flux to magnitude (m) conversion equation:

$$m = -2.5 \log_{10}(f) + ZP \tag{4.5}$$

where f is the flux, and ZP is the zero point. The zero point will vary depending

on the instrument with which the data is taken. It helps ensure that the flux is properly scaled. It is calculated such that an object with known zero magnitude will have a magnitude of zero when observed with that instrument. For TESS, the zero point is 20.44 (Vallely et al. 2019). Using Equation (4.5) and the TESS zero point, I calculated the magnitude of each data point in a sector. Then I used these magnitudes in Equation (4.4) to calculate the $1/\eta$ of each of the 317 sectors.

Most of the light curves in our sample had values of roughly $1/\eta \approx 0.5$. In order to find a 'cutoff' value for $1/\eta$, above which we can select GPs as potentially varying, I used a method similar to the one used in Sokolovsky et al. (2017). I got the median value of $1/\eta$ for all 317 sectors, which is the expected value ($\overline{1/\eta}$) of the $1/\eta$ for our sample. This value was $\overline{1/\eta} = 0.511$. Then, following Sokolovsky et al. (2017), I calculated the expected dispersion, σ, as the median absolute deviation (MAD) scaled to a Gaussian distribution, which is:

$$\sigma = 1.4826 \times \text{MAD} = 1.4826 \times \text{median}(\textstyle\sum_i |(1/\eta)_i - \overline{1/\eta}|) \tag{4.6}$$

where $(1/\eta)_i$ refers to each i^{th} inverse von Neumann ratio and the MAD is multiplied by 1.4826 in order to scale it to a Gaussian distribution. Sokolovsky et al. (2017) prefer using MAD instead of the standard deviation from the mean because the standard deviation is relatively sensitive to outlier points, whereas MAD is mostly insensitive to outliers. The value of the dispersion σ obtained from equation (4.6) is then multiplied by a factor a; the value of $a\sigma$ determines the threshold of variability.

Sokolovsky et al. (2017) computed C, P, and F from equations (4.1)-(4.3) as functions of a, then adopted the value of a for which F reaches its maximum value as the optimal value. From their graphs, it is clear that the optimal value of a depends not only on the index used to measure the variability but also on the dataset itself. I could not use this method to select a since F is an index that requires already knowing how many variable objects are in a dataset. Instead, I used a different

method suggested by Sokolovsky et al. (2017); they suggested starting at $a = 3$ and adjusting the value of a to reach a manageable set of false positives. Since I do not know how many of these sectors are variable, I will have to adjust the value of a until I get a cutoff value that selects a reasonable sample of sectors.

I started with $a = 3$, which gave me a cutoff value of $1/\eta$ of 0.569. There were 19 sectors that fell above this value. With a value of $a = 4$, I got a cutoff value of 0.588 and 13 sectors were selected. At a value of $a = 5$, I got a cutoff value of 0.607 and 10 sectors were selected. At a value of $a = 14.5$ no sectors were selected. For this thesis, I decided to use a value of $a = 4$. I chose this value by comparing the sectors that were selected at $a = 4$ to the sectors that were selected at $a = 3$ and $a = 5$. At $a = 3$ there were $\sim 3 - 4$ sectors that did not appear to have much variability. These sectors were a bit scattered and there was not much change in the overall light curve, or there was only a very subtle trend. These sectors appeared a little bit 'flat' overall, but not as flat as other sectors with $1/\eta$ ratios close to the median value. When I increased the value of a to 4, these 'flat' sectors were among the 6 sectors that were eliminated. When I increased the value of a to 5, there were sectors eliminated that displayed just enough variability when examined visually that I could not justify eliminating them. Therefore, I will use a cutoff value of 0.588, corresponding to $a = 4$ to select for potentially varying sectors. These sectors will be the 'high confidence' sectors, where there is a high confidence of variability. The sectors with $1/\eta$ between 0.569 and 0.588 will be considered the low confidence sectors, since they did display some variability above 3σ, but not enough to be among the high confidence sectors. For this thesis, I will be mainly looking at the high confidence sectors to try to find optical variability.

Figure 4.1 shows a histogram with the distribution of the $1/\eta$ values for all of the sectors in our sample. The horizontal axis is the value of the $1/\eta$ ratio, and the high confidence cutoff value of $1/\eta = 0.588$ is marked in the figure by a dashed vertical line. It can be seen from Figure 4.1 that most of the observations had a $1/\eta$ ratio around 0.5, and there are very few sectors with a value above 0.588. Even

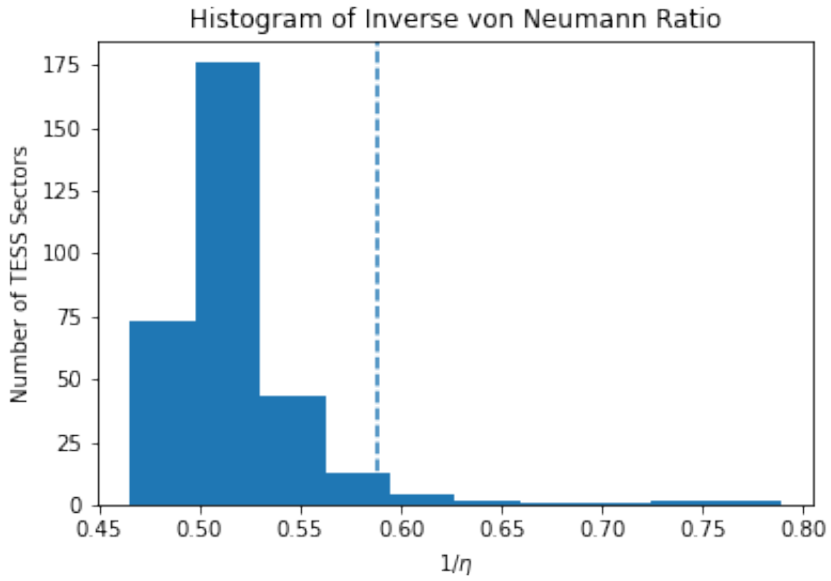

Figure 4.1: Histogram showing the different values of the inverse von Neumann ratio, $1/\eta$ for all sectors. The horizontal axis shows the value of the ratio. The dashed vertical line is placed at $1/\eta = 0.588$, the adopted high confidence cutoff value to select for potentially varying sectors.

though higher values of $1/\eta$ correspond to greater variability in a light curve, we cannot say that the sectors above the cutoff value definitely have optical variability. Instead, the sectors with values of $1/\eta$ above the cutoff have light curves that were less smooth when compared to the majority of the other light curves, so we have a high confidence that they are varying. So, we are using this as a way to narrow down which sectors we should inspect closer. Therefore, we will look at these 13 sectors in greater detail.

4.1.3 Brightness vs. von Neumann Ratio

Before continuing my analysis, I wanted to ensure that the high $1/\eta$ wasn't just because of noisy, faint GPs. I compared the brightness of all of the GPs to their inverse von Neuman ratios for the sectors they were observed in. Figure 4.2 shows the comparison of each sector's $1/\eta$ versus the SDSS z magnitude of the GP. The

dashed vertical line shows the high confidence cutoff value of 0.588. GPs that were observed in multiple sectors have the $1/\eta$ value plotted for each sector, with the same value of the z magnitude. The graph shows that the GPs with the high $1/\eta$ are not restricted to just the faintest GPs but are distributed throughout the range of the GP magnitudes. Thus, we can move forward knowing the high $1/\eta$ for these 13 high confidence sectors was not because of the low brightness of GPs.

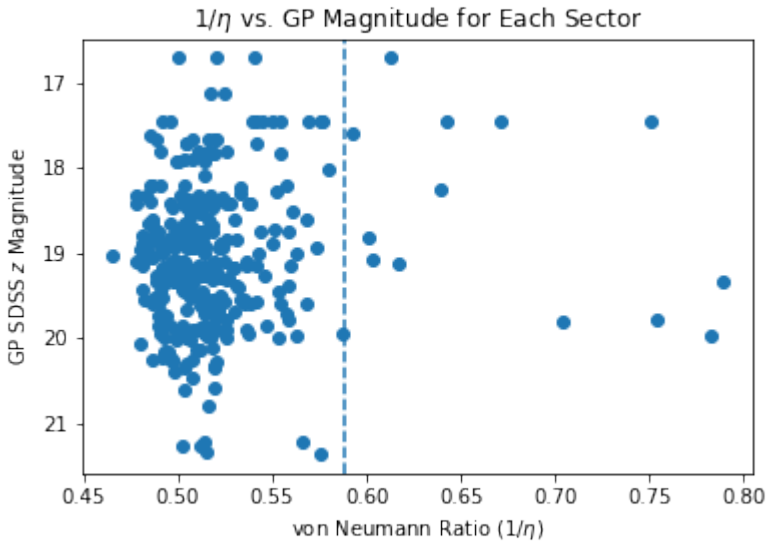

Figure 4.2: Plot showing the inverse von Neumann ratio versus the magnitude of the GP. The magnitude of the GP is given in the SDSS z filter. The dashed vertical line is placed at $1/\eta = 0.588$. GPs that are observed in multiple sectors have the von Neumann ratio for each sector plotted, against the same z filter magnitude. The GPs with the high $1/\eta$ ratios are distributed along the graph, not just concentrated on the faintest magnitudes.

4.1.4 Visual Inspection

After ensuring the high $1/\eta$ values were not because of faint GPs, I visually inspected the 13 high confidence sectors. Of the 13, I eliminated 2 sectors where it was very obvious the high ratio was caused by a few large outliers. Both of these sectors were generally smooth but had a few very large outliers in portions of their light curves where TESS was generally farther from the earth. Figures 4.3 and 4.4 show the light curves for those two sectors, where the large outliers can be seen in the in the middle of the 'good' data portions. Since the observations in these sectors were not taken during those intervals for which the data had been determined to be bad, there was no way I could choose segments of time for which to eliminate bad data. Instead, I chose to exclude both sectors from further analysis. This leaves me with 11 high confidence sectors that show either obvious variability, like the example in Figure 3.2, or very slight variability, comparable to the example in Figure 3.1.

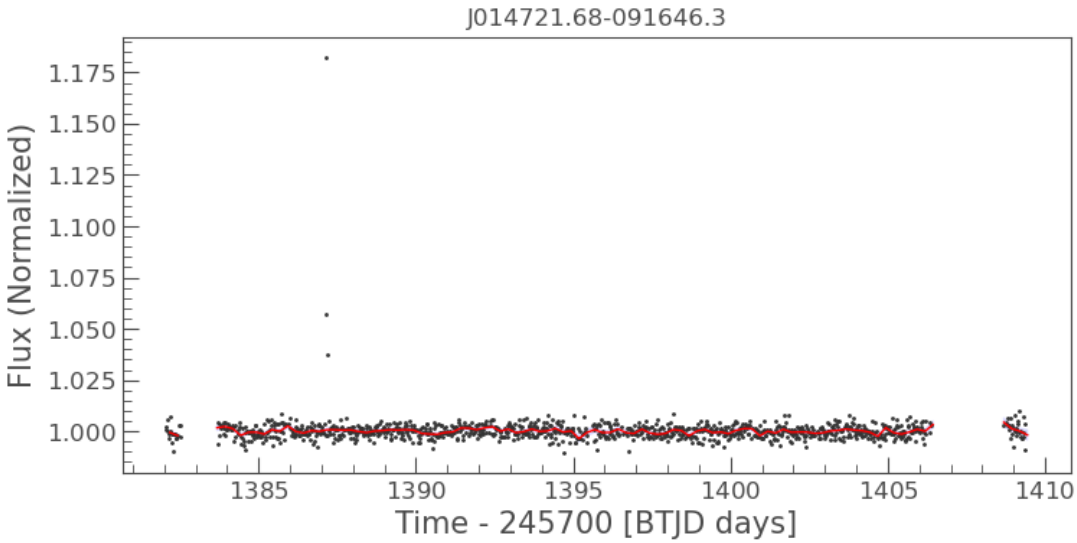

Figure 4.3: Light curve of the GP with SDSS designation J014721.68-091646.3, also referred to as GP# 13 in this data set. The vertical axis has the normalized flux value, normalized to the median. On the horizontal axis is the time in Modified TESS Julian date. The black dots are each individual TESS observation. The solid red curve is the running median, binned with 6-hour bins. The blue shadow is the error shadow. This sector had a generally smooth light curve except for 3 very large outliers. These outliers are not during the time when TESS is close to the earth (during the beginning and middle of the sector), thus this was not taken during a time of known bad data, and the whole sector was eliminated from the high $1/\eta$ analysis.

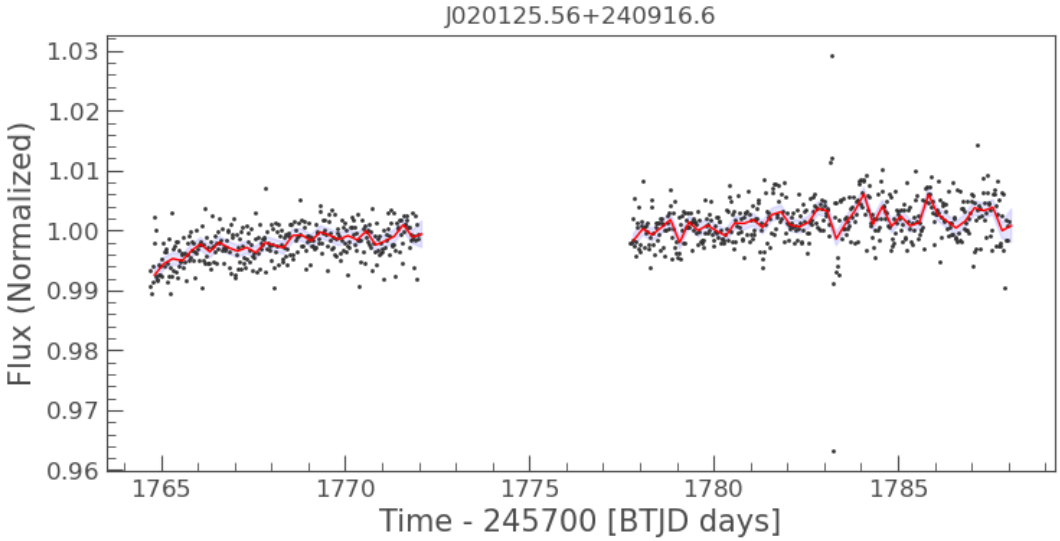

Figure 4.4: Light curve of the GP with SDSS designation J020125.56+240916.6, also referred to as GP# 16 in this data set. The vertical axis has the normalized flux value, normalized to the median. On the horizontal axis is the time in Modified TESS Julian date. The black dots are each individual TESS observation. The solid red curve is the running median, binned with 6-hour bins. The blue shadow is the error shadow. This sector had an overall smooth light curve but had several very large outliers. These outliers are not during the time when TESS is close to the earth (during the beginning and middle of the sector), thus this was not taken during a time of known bad data, and the whole sector was eliminated from the high $1/\eta$ analysis.

4.1.5 Pixel Comparisons

In order to further inspect these 11 high confidence observations, I looked at the whole TESS cutout that was created for each of these sectors. When I created the light curves for these sectors, I got a 5 x 5 cutout of the TESS full frame image, as described in Section 2.3.2. I wanted to check that the higher value of $1/\eta$ for these sectors was not just due to a particularly noisy section of the cutout or contamination from another source within the TESS cutout. So, I compared the $1/\eta$ of each of the other 24 pixels in the cutout to the $1/\eta$ value in the pixel containing the GP. To do so, I first created a light curve for every pixel in the cutout for each of the 11 sectors I wanted to further inspect. I did this using the same methods I used to create the light curve of the GP described in Section 2.3.2, but I moved the aperture to every pixel in the cutout.

I then used the same methods described in Section 3.2 to normalize the light curve to the median and create a binned light curve with an error shadow. For these pixel comparisons I only plotted the binned running median and the error shadow, not the full light curve which has all the individual observations, to reduce crowding in the final pixel comparison image. I plotted the light curves and error shadows in each of the 25 pixels in the cutout in its corresponding location, creating a grid of 5 x 5 light curves. I then calculated $1/\eta$ for every pixel in the cutout. Figure 4.5 shows an example of one of these light curve graph grids, with the $1/\eta$ value placed in its corresponding light curve. The red and green curves represent the normalized binned light curves, and the error shadow is shown in blue. The light curve shown in green and labeled 'GP' in the graph is the light curve of the Green Pea galaxy. The GP shown in Figure 4.5 is from the same sector shown in Figure 3.2.

Once I had the pixel comparison plot, I compared the $1/\eta$ values for all the other pixels with $1/\eta$ value for the pixel with the GP. I got the median value and standard deviation of the $1/\eta$ values for the 25 pixels within each cutout. I then compared these to the $1/\eta$ of the pixel with the GP. These results are presented in Table 4.1,

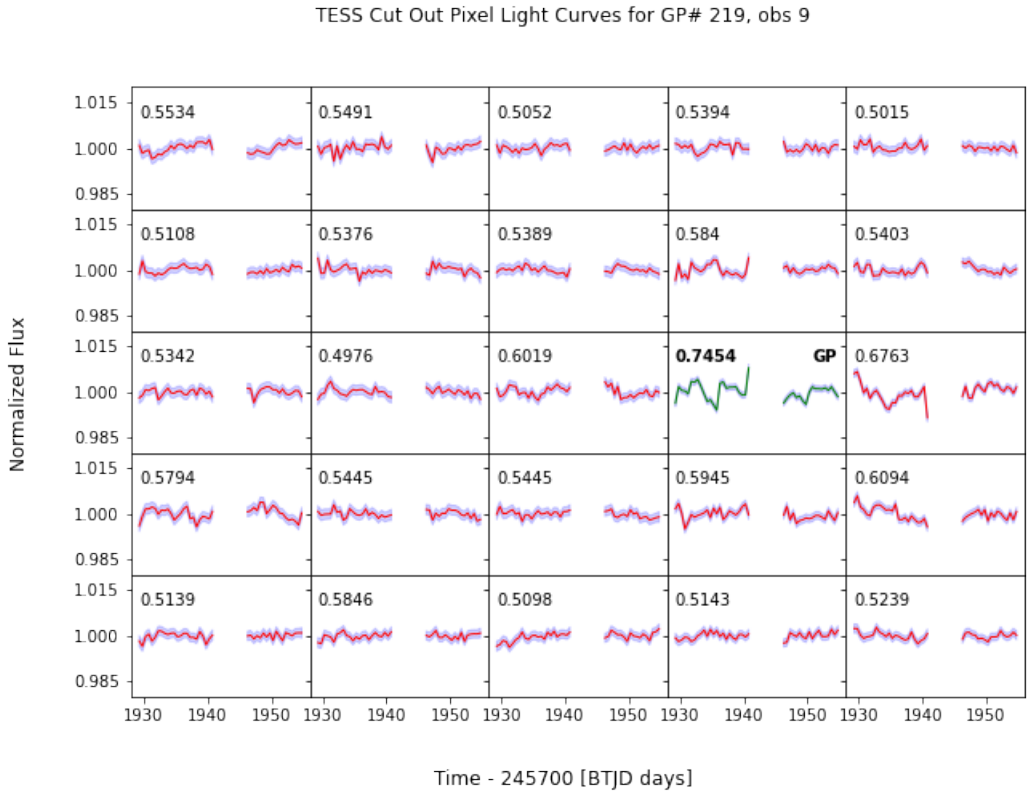

Figure 4.5: Example of a Light Curve grid. This grid is for GP# 219, observation 9. On the vertical axis is the normalized flux value, which has been normalized to the median. On the horizontal axis is the time in Modified TESS Julian Date. The curves on each graph in red/green represent the normalized and binned light curves. The blue shadow is the error shadow. The green curve, in the panel marked GP with labels in bold face, is from the pixel in which the GP is located and, therefore, is the light curve of the GP.

$1/\eta$ Statistics for Entire Cutout Light Curve Grid					
GP Number	Obs.	Median $1/\eta$	Median + 3σ	Max $1/\eta$	GP $1/\eta$
12	0	0.5684	1.0567	0.9272	0.7541
17	0	0.5225	0.8998	0.9600	0.7043
41	**0**	**0.5069**	**0.6047**	**0.6175**	**0.6175**
84	0	0.5127	0.6895	0.6821	0.6002
102	0	0.5143	0.6358	0.6735	0.6007
147	0	0.5673	1.1706	1.1258	0.6389
150	0	0.5120	0.9427	1.1559	0.5923
171	3	0.5236	0.6247	0.6110	0.6110
219	3	0.5444	0.7000	0.6821	0.6427
219	**9**	**0.5405**	**0.7323**	**0.7478**	**0.7478**
219	**10**	**0.5224**	**0.6539**	**0.6698**	**0.6698**

Table 4.1: Table with statistics on the whole cutout $1/\eta$ values for the 11 high confidence sectors values of $1/\eta > 0.588$. "GP Number" refers to the number assigned to that GP. "Obs." refers to the observation number of the sector, for example obs. 0 is the first sector the GP was observed. "Median $1/\eta$" is the median value of the $1/\eta$ values of the 25 pixels in the cutout. "Median + 3σ" is the median value plus 3σ. "Max $1/\eta$" is the maximum value of $1/\eta$ in the cutout for that sector. "GP $1/\eta$" is the value for $1/\eta$ calculated for that GP. Sectors that are boldface have the GP $1/\eta$ greater than the median + 3σ *and* have the highest value for $1/\eta$ in the cutout.

which shows the median, 3σ value, the GP ratio, and the highest ratio for each of the 11 high confidence sectors that I was analyzing. I found that 3 sectors had a GP with a von Neumann ratio at least 3σ above the median value. Those 3 sectors also had the pixel with the GP as the pixel with the highest von Neumann ratio. The 3 sectors with the von Neumann ratio above 3σ *and* the highest von Neumann ratio in the cutout are in bold face in Table 4.1. These 3 sectors, corresponding to 2 GPs, are the sectors that display the most optical variability in our sample on this 27-day timescale.

4.1.6 Upper Limit on the Number of AGN in our Sample

Based on the analysis in the previous section, I conclude that 3 sectors, representing 2 GPs, display optical variability on the 27-day timescale investigated for this thesis. As discussed in Section 1.2.1, the presence of this variability tells us that these GPs might contain AGN. This is promising; this tentative detection of 2 AGN also implies that there could be more AGN present in our sample, but that I haven't detected them because they vary on timescales longer than 27 days. AGN can vary in the optical on time scales of days to months. So, it is possible that any AGN present in our sample vary on time scales longer than 27 days. In fact, MacLeod et al. (2012) has shown that only about 10% of AGN vary up to 10% on 30-day timescales. So, this means that we would only find up to 10% of AGN present in our sample on this short time scale. It also means that any AGN we do find would only display low (subtle) variability, no more than 10%, on this short time scale.

Thus, we can place upper limits on the amount of AGN we can expect to find in this sample. If the 2 GPs in our sample that display some optical variability represent 10% of the AGN present in our sample, that means there might be up to 20 AGN present in our sample, with 90% of the AGN present varying on longer timescales.

Of course, the short time scale is a limitation in this thesis. Since only a small number of AGN vary on such short time scales, we can only find the 10% of AGN that vary on such short time scales. Another limitation is that even though we have data for some GPs for more than 30 days, we cannot compare the GPs across sectors, as mentioned in Section 3.1.1. So even though trying to find the variability on longer timescales would be ideal, it is not possible at this time.

4.2 Other methods of detecting AGN

Based on the analysis in the preceding section, we have reason to believe that there are 2 GPs in our sample that vary on 27-day time scales and that might, therefore, contain AGN. The next steps to help confirm the presence of AGN in these GPs would be to look at other data. While any one of these methods alone can help identify potential AGN candidates, it is important to note that the lack of detection using one method does not completely rule out the possibility of an AGN in that galaxy. Because different AGN can display vastly different properties, it is important to instead look at several different methods for AGN detection to confirm the presence of AGN.

4.2.1 Optical Spectra

One of the most common ways of finding AGN in galaxies is by looking at the spectra of the galaxies. Since AGN are very strong ionizing sources, they can produce distinct spectra from star forming galaxies, as discussed in Section 1.2.1. However, despite the issues surrounding using the spectra to identify AGN in GPs (discussed in Section 1.4), we can still look at the spectra of these galaxies to try to find evidence of AGN in them. One emission line that can be used to try to find AGN is the He II λ4686 line (Bär et al. 2017). Since SDSS has spectroscopy for all of the GPs, I can check if this line was detected in their spectra. Table 4.2 shows the spectroscopic data for the 11 high confidence sectors. Table 4.2 shows the flux of the He II emission line for the GPs and the Signal-to-Noise (S/N) ratio. The S/N ratio measures the ratio of how much of the flux is actually from the signal we are trying to measure compared to how much of the flux is coming from noise. A higher S/N can mean a higher quality measurement and can be used to help determine if the signal was actually detected. In this case, we can say that GPs with a S/N > 5 can be considered to have a He II detection. The 2 GPs that show significant variability are in bold face.

He II Emission Line Fluxes for High Confidence GPs		
GP Number	He II Flux	He II S/N
12	1.4906	1.7163
17	-1.4865	-0.7664
41	**1.1454**	**1.5033**
84	1.7911	0.8861
102	1.6778	1.404
147	2.8596	1.2562
150	9.4039	3.3892
171	35.249	6.3468
219	**26.9697**	**8.2667**

Table 4.2: Table with the He II λ4685 emission line flux and S/N for the 11 high confidence GPs. "GP Number" refers to the number assigned to that GP. "He II Flux" is the flux that was detected from that GP at a wavelength of 4685 Å. "He II S/N" is the S/N ratio for that GP. GPs that are in boldface are the two GPs that showed significant optical variability. Only 2 GPs listed in this table have a S/N > 5, which indicates the detection of the He II emission line. One of those, GP# 219, also showed significant optical variability.

Table 4.2 shows that one of the GPs that has demonstrated optical variability, GP# 219, has a S/N > 5. Therefore, the He II emission line was detected in this GP. This means it is likely there might be an AGN in this galaxy. Additionally, we see that only one other GP in our high confidence group had S/N > 5, GP# 171. Even though this GP did not have significant variability, the presence of this detection line means it is possible that it might contain an AGN as well.

4.2.2 Mid-Infrared Light

We can also look at these galaxies in MIR. Harish et al. (2023) searched for AGN in these galaxies using MIR light from the WISE telescope, as discussed in Section 1.3. They were able to find 31 AGN candidates by looking at the MIR colors and using color-color plots to select GPs as AGN candidates. They also found 2 AGN candidates by looking at the MIR variability of these galaxies.

If we compare the 2 groups of variable GPs in optical and MIR, we find that neither of the 2 GPs that displayed some optical variability were variable in MIR. In fact, the 2 GPs that were variable in the MIR were not among the high confidence nor low confidence optical variability sectors. Thus, the two GPs selected as potential AGN based on their MIR variability displayed no optical variability on a 27-day timescale, since they both had $1/\eta < 0.568$. This is an interesting result, and there are two possible reasons. Since the MIR light is caused by dust, it is possible that any optical variability in these GPs is obscured by the dust. The dust could completely absorb the optical light in the line of sight, thus obscuring any optical variability and reemitting it as MIR variability. Another reason is that the variation is on different time scales. Harish et al. (2023) looked at the sample of GPs over the course of more than 5 years, looking for long term variability. I was looking at the GPs over a period of $\sim$ 30 days. This very large difference in time scale could be a major factor as to why different GPs show different variability in the two wavelengths.

In addition to the MIR variability, we can also look at the MIR colors to select for AGN. Like the MIR variability, none of the 31 MIR color selected AGN candidates were among the sectors with high confidence of variability nor the low confidence group. This means that none of the GPs that Harish et al. (2023) indicated as potential AGN candidates display optical variability. As discussed above, this could be due to the fact that the MIR light is caused by the torus which can obscure the optical light. Another reason that these AGN candidates don't overlap is because Harish et al. (2023) looked at a smaller sample. In fact, one of the 2 GPs that displayed some optical variability does have MIR colors that select it as a potential AGN, even though it was not among the sample of MIR color selected AGN candidates in Harish et al. (2023). This GP is GP# 219. At this time, it is not known if the other GP that displayed optical variability has MIR colors that select it as a potential AGN.

4.2.3 X-Ray Luminosity

Finally, another way we can look at these GPs in more detail to find AGN is to look at the X-ray luminosities. As described in Section 1.2.1, X-rays are a good way to try to find AGN. AGN are one of the only sources that can produce the high X-ray luminosities that are used to identify them. So, I looked for these GPs in data taken with the Chandra X-ray Observatory. I looked through archival data from Data Release 2 (DR2). From the full sample of 216 GPs, only 4 GPs were detected with Chandra. The X-ray luminosities from this data appear to indicate that the X-rays might be coming from an AGN, not from X-ray binaries. This will be discussed in more detail in Singha et al. (2023; in preparation). One of these four GPs selected based on Chandra data is GP# 219, one of the two GPs that I found displays the highest optical variability in my sample.

The presence of the He II emission line, the MIR color selection, and the detection at X-ray wavelengths reinforces my conclusion that at least one of the GPs that displayed optical variability, GP# 219, contains an AGN. The other GP that displayed significant optical variability, GP# 41, likely also contains an AGN but will need to be confirmed.

CHAPTER 5

Conclusions and Future Work

In this thesis I have investigated the 27-day optical variability of GPs using TESS with the goal of finding AGN in these galaxies. In Chapter 1, I provided background on some of our current knowledge and my motivations for the project. In Chapter 2, I described the observations that these data come from as well as the photometry and light subtraction process. In Chapter 3, I presented the light curves that were created. Finally, in Chapter 4, I introduced the $1/\eta$ ratio and discussed my process of analyzing the light curves using this ratio. I then discussed some other methods that can be used to find AGN and how the GPs that showed the most promise for variability looked in those data. In the following chapter, I will present my final conclusions in Section 5.1 and describe potential future work in Section 5.2.

5.1 The Presence of AGN in GPs

In this thesis, I used observations from TESS to look at the light curves of 219 GPs, using 317 separate observations. After creating and cleaning the light curves, I analyzed all of the sectors using the inverse von Neumann ratio $1/\eta$, which was shown to be a fairly reliable index to quantify variability. After an iterative process, I adopted $1/\eta = 0.588$ as the cutoff value for variability with high confidence. I found that a total of 11 sectors had a high confidence for variability, with $1/\eta > 0.588$. There were 6 additional sectors that exhibited variability above 3σ, with $0.569 > 1/\eta > 0.588$; I consider these to have a low confidence of variability. Three of the high confidence sectors showed demonstrable variability. Since those 3 sectors correspond to 2 GPs, I conclude that those two GPs, GP# 41 (J083350.24+454933.6) and GP# 219 (J173501.25+570308.6), are reasonable candidates for AGN.

Not only do both these GPs show optical variability, but one of them, GP# 219, shows considerable evidence for AGN in several different AGN-detection techniques. The He II emission line is detected in its spectra. It has MIR colors that classify it as an AGN and it was detected with X-ray luminosities that are likely from an AGN (Singha et al. 2023, in preparation). Because one of the two GPs with optical variability has strong evidence for an AGN, I conclude that it is also likely that the other GP also contains an AGN.

Finally, since I found only two AGN candidates in the sample of 216 GPs, we can place an upper limit to the number of AGN in this sample. As discussed in Section 4.1.6, since only 10% of AGN show any variability on a 30-day timescale, I conclude that the upper limit of AGN in this sample is roughly 20 AGN.

5.2 Future Work

In a sample of 216 GPs, I have discovered optical variability in two GPs. One of these, GP# 219, has signatures of AGN in other tracers, like MIR colors, He II emission, and X-ray luminosity, and is therefore confirmed as an AGN with reasonably high confidence. Future work would include searching for these tracers in the second of these two candidates, GP# 41, to confirm the presence of an AGN in this source. In particular, it would be interesting to observe GP# 41 at X-ray wavelengths. It would also be beneficial to observe both the high and low confidence groups in the X-ray. Another next step would be to look at the optical spectra in more detail, to try to find other emission lines that could indicate the presence of an AGN for both the high and low confidence GPs. We could also observe the optical variability of these galaxies on longer time scales. This could involve either solving the issue with TESS or using another telescope. If we can observe some of these GPs in the optical for a longer amount of time, we would be able to check if there are any GPs in our sample that demonstrate optical variability on a longer time scale. This is very likely to yield results since 90% of AGN do vary on timescales longer than 30 days in the optical.

REFERENCES

Bär, R. E., Weigel, A. K., Sartori, L. F., et al. 2017, MNRAS, 466, 2879. doi:10.1093/mnras/stw3283

Baldassare, V. F., Geha, M., & Greene, J. 2018, ApJ, 868, 152. doi:10.3847/1538-4357/aae6cf

Baldassare, V. F., Geha, M., & Greene, J. 2020, ApJ, 896, 10. doi:10.3847/1538-4357/ab8936

Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, PASP, 93, 5. doi:10.1086/130766

Brasseur, C. E., Phillip, C., Fleming, S. W., et al. 2019, Astrophysics Source Code Library. ascl:1905.007

Burke, C. J., Shen, Y., Chen, Y.-C., et al. 2020, ApJ, 899, 136. doi:10.3847/1538-4357/aba3ce

Cardamone, C., Schawinski, K., Sarzi, M., et al. 2009, MNRAS, 399, 1191. doi:10.1111/j.1365-2966.2009.15383.x

Harish, S., Malhotra, S., Rhoads, J. E., et al. 2023, ApJ, 945, 157. doi:10.3847/1538-4357/acb99c

Izotov, Y. I., Guseva, N. G., & Thuan, T. X. 2011, ApJ, 728, 161. doi:10.1088/0004-637X/728/2/161

Jiang, T., Malhotra, S., Rhoads, J. E., et al. 2019, ApJ, 872, 145. doi:10.3847/1538-4357/aaee8a

Kim, K. J., Malhotra, S., Rhoads, J. E., et al. 2021, ApJ, 914, 2. doi:10.3847/1538-4357/abf833

Lightkurve Collaboration, Cardoso, J. V. de M., Hedges, C., et al. 2018, Astrophysics Source Code Library. ascl:1812.013

MacLeod, C. L., Ivezić, Ž., Sesar, B., et al. 2012, ApJ, 753, 106. doi:10.1088/0004-637X/753/2/106

Malhotra, S., Rhoads, J. E., Finkelstein, S. L., et al. 2012, ApJL, 750, L36. doi:10.1088/2041-8205/750/2/L36

von Neumann J. 1942, Ann. Math. Stat., 13, 86

Padovani, P., Alexander, D. M., Assef, R. J., et al. 2017, A&ARv, 25, 2. doi:10.1007/s00159-017-0102-9

Sokolovsky, K. V., Gavras, P., Karampelas, A., et al. 2017, MNRAS, 464, 274. doi:10.1093/mnras/stw2262

Treiber, H. P., Hinkle, J. T., Fausnaugh, M. M., et al. 2022, arXiv:2209.15019. doi:10.48550/arXiv.2209.15019

Trump, J. R., Sun, M., Zeimann, G. R., et al. 2015, ApJ, 811, 26. doi:10.1088/0004-637X/811/1/26

Urry, C. M. & Padovani, P. 1995, PASP, 107, 803. doi:10.1086/133630

Vallely, P. J., Fausnaugh, M., Jha, S. W., et al. 2019, MNRAS, 487, 2372. doi:10.1093/mnras/stz1445

Vanderspek, R., Doty, J. P., Fausnaugh, M., et al. 2018 https://archive.stsci.edu/files/live/sites/mast/files/home/missions-and-data/active-missions/tess/_documents/TESS_Instrument_Handbook_v0.1.pdf

Yang, H., Malhotra, S., Gronke, M., et al. 2016, ApJ, 820, 130. doi:10.3847/0004-637X/820/2/130

Yang, H., Malhotra, S., Gronke, M., et al. 2017, ApJ, 844, 171. doi:10.3847/1538-4357/aa7d4d