Title

Planets and their Hidden Companions: Expanding Detection Methods with Machine Learning

Author

Angelo, Isabel

Publication Date

2025

UNIVERSITY OF CALIFORNIA

Los Angeles

Planets and their Hidden Companions:
Expanding Detection Methods with Machine Learning

A dissertation submitted in partial satisfaction

of the requirements for the degree of

Doctor of Philosophy

in Astronomy and Astrophysics

by

Isabel Nora Angelo

2025

ABSTRACT OF THE DISSERTATION

Planets and their Hidden Companions:
Expanding Detection Methods with Machine Learning

by

Isabel Nora Angelo

Doctor of Philosophy in Astronomy and Astrophysics

University of California, Los Angeles, 2025

Professor Erik A. Petigura, Chair

In this thesis, I expand on existing methodologies for identifying distant planetary and stellar companions to known planet hosts. First, I outline my discovery of a distant planetary companion to Kepler-1656b, a highly eccentric sub-Saturn. To make this discovery, I used a combination of traditional radial velocity detection methods and dynamical modeling of the system to resolve the properties of an outer planetary companion. In doing so, I characterized the dynamical environment of the Kepler-1656 system and found that, surprisingly, the outer planet is exciting the inner planet eccentricity in situ, i.e. without inducing inward migration of the inner planet. I also identified signatures of similar companions, which I call "gentle giants", in the broader sub-Saturn and giant planet populations.

The second portion of my thesis explores more experimental methods of detecting stellar companions to known planet hosts using data-driven spectroscopy. Recent advances in data-driven spectroscopy have enabled more modeling of stellar spectra, which may in turn allow us to detect signatures of stellar companions to planet hosts that are missed by more traditional binary detection methods. I trained data-driven models on spectra from the Gaia mission and Keck-I telescope, and in doing so, developed a novel wavelet-based method for removing non-astrophysical contamination in stellar spectra. I found that while our models

are excellent at label transfer– a standard application of data-driven spectroscopy– they are limited in their ability to accurately model stellar spectra and identify hidden stellar companions. I also comment on the state of the field of data-driven spectroscopy and outline possible paths towards more accurate spectroscopic models and surveys for planet-hosting binaries.

The dissertation of Isabel Nora Angelo is approved.

Smadar Naoz

Bradley M. Hansen

Tuan H. Do

Erik A. Petigura, Committee Chair

University of California, Los Angeles

2025

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First and foremost, I'd like to thank my advisors. Most people will get 0-1 PhD advisors in a lifetime, and I was lucky enough to have three. To Erik Petigura (my primary advisor), Megan Bedell, and Smadar Naoz- you are all incredible scientists and mentors, and I am so lucky to have worked with and learned from you. I can only hope that the scientist I become is some linear combination of the three of you.

Second, I want to thank my parents, for their long and unwavering support. When I was four I said I wanted to be a plumber, my dad brought home pieces of metal pipes and gave them to me "for practice". When I was eight and wanted to be an architect, he bought me adult-level architecture books that probably went way over my head, but I remember reading them over and over nonetheless (or at least looking at the pictures). When I started studying astronomy in college, he read every astronomy and physics book I recommended him. I once joked that he should read my favorite textbook[1], only to find when I came home that he had taken it off my shelf and started to read it.

My mom always alludes to a brief period in my childhood where I was really into potions as the first signs that I wanted to be a scientist (to be honest I think I just liked the pretty bottles). Every time I wanted to make a new potion, she took me to Michael's in Glendale to pick out a new glass bottle to put it in. She was with me in Germany when I submitted my grad school applications- I was an absolute mess walking through the German Christmas markets and she never once accused me of ruining her vacation. Since I've been in grad school, she always asks me about my research. She even carried around a drawing I made of Kepler-1656b (see Chapter 2) in her wallet. To my mom and dad– I wouldn't be here without your support and encouragement!

Finally, I want to thank my siblings. I've turned to you in my most challenging times and

[1]Barbara Ryden's Introduction to Cosmology, and I highly recommend

you're always there. I love and admire you both, and I wouldn't be here or anywhere without you.

VITA

Isabel Nora Angelo

EDUCATION

M.S. – Astronomy and Astrophysics **2022**
UCLA *Los Angeles, CA*

B.S. – Astrophysics **2017**
University of California, Berkeley *Berkeley, CA*

RESEARCH EXPERIENCE

Graduate Research Assistant **2019–2025**
University of California, Los Angeles *Los Angeles, CA*

Post-baccalaureate Research Assistant **2018-2019**
University of California, Berkeley *Berkeley, CA*

Undergraduate Research Fellow **2017**
California Institute of Technology *Pasadena, CA*

Undergraduate Research Assistant **2017–2019**
SETI Institute *Mountain View, CA*

TEACHING EXPERIENCE

Teaching Assistant **2019–2025**
University of California, Los Angeles *Los Angeles, CA*

Undergraduate Student Instructor **2016–2020**
University of California, Berkeley *Berkeley, CA*

REFEREED JOURNAL PUBLICATIONS

Isabel Angelo, Petigura, E., and Bedell, M. (2025). Seeking Spectral Binaries with Data-Driven Models. *submitted to The Astrophysical Journal*

Isabel Angelo, Bedell, M., Petigura, E., and Ness, M. (2024). A Data-driven Spectral Model of Main-sequence Stars in Gaia DR3. *The Astrophysical Journal*, 974(1):43

Isabel Angelo, Duchene, G., Stapelfeldt, K., Telkamp, Z., Ménard, F., Padgett, D., Van der Plas, G., Villenave, M., Pinte, C., Wolff, S., Fischer, W. J., and Perrin, M. D. (2023). Demographics of Protoplanetary Disks: A Simulated Population of Edge-on Systems. *The Astrophysical Journal*, 945(2):130

Isabel Angelo, Naoz, S., Petigura, E., MacDougall, M., Stephan, A. P., Isaacson, H., and Howard, A. W. (2022). Kepler-1656b's Extreme Eccentricity: Signature of a Gentle Giant. *The Astronomical Journal*, 163(5):227

Isabel Angelo, Rowe, J. F., Howell, S. B., Quintana, E. V., Still, M., Mann, A. W., Burningham, B., Barclay, T., Ciardi, D. R., Huber, D., and Kane, S. R. (2017). Kepler-1649b: An Exo-Venus in the Solar Neighborhood. *The Astronomical Journal*, 153(4):162

Isabel Angelo and Hu, R. (2017). A Case for an Atmosphere on Super-Earth 55 Cancri e. *The Astronimical Journal*, 154(6):232

Chapter 1

Introduction

Where is everybody?

— Enrico Fermi, *Fermi's paradox*

In the past three decades, we have seen the emergence of a new field of astronomy: the study of exoplanets, or planets beyond our solar system. Since the discovery of the first known exoplanet in 1995 (51 Pegasi b, Mayor and Queloz, 1995), thousands of planets have been discovered, with over 5,000 known exoplanets at the time of this dissertation's publication. This is due in large part to space-based transit missions like NASA's *Kepler* mission, which has confirmed nearly 3,000 planets[1] since its launch in 2009. This mission was the first of its kind, and discovered the largest sample of exoplanets to date by monitoring the brightness of hundreds of thousands of stars in a patch of sky covering 115.6 square degrees- about the size of a human's hand extended at arm's length.

Kepler's legacy extends far beyond the sheer number of planets in its sample. It revealed classes of planets unlike any in our own solar system (e.g., hot Jupiters; Dawson and Johnson 2018, Super-Earths; Mayor et al. 2011; and Sub-Saturns; Petigura et al. 2017c), alluding to

[1] from from `http://exoplanetarchive.ipac.caltech.edu/`

a diversity of planet properties and origins. Its large sample size also enabled unprecedented demographic studies of exoplanets. In particular, spectroscopic follow-up efforts to classify *Kepler* planet hosts illuminated key trends that have informed our understanding of planet formation. For example, features like the 'Radius Valley' and 'Hot Neptune Desert' encode information about core accretion, mass loss, and migration processes (Fulton and Petigura, 2018; Petigura, 2020).

Still, the *Kepler* sample is far from complete. Small planets are lost to noise, and the mission's four-year baseline and bias towards edge-on systems means that we are less sensitive to planets beyond ∼1AU (Petigura et al., 2018b). Indeed, if we were to observe the solar system with *Kepler*, the probability of observing Earth transit even once is only ≈0.5%, and this value is nearly an order of magnitude for Jupiter (≈0.08%). This means that if we were to observe the solar system with just *Kepler*, we would miss more distant planets and our picture of the system would be incomplete. So too are the pictures of planetary systems we've observed with *Kepler*, which may host distant, non-transiting companions that play an important role in the systems' architectures and formation histories.

Beyond this, *Kepler* targets are subject to flux dilution, where spatially unresolved stellar binaries are mis-classified as single stars, leading to underestimated planet radii. This is particularly a problem for studying demographic trends like the 'Radius Valley' which depend sensitively on stellar radii from which the planet radii are inferred (see for example Fulton and Petigura, 2018; Berger et al., 2018). Beyond this, our insensitivity to unresolved stellar binaries in *Kepler* has led to competing estimates of the prevalence of close binaries among planet hosts. While some studies have shown that both disks and planets are suppressed for close binaries (e.g., Wang et al., 2014; Cieza et al., 2009), others have found no evidence of such a suppression (e.g., Horch et al., 2014; Matson et al., 2018). Additionally, the role binaries play in sculpting the orbital architectures of planets is unclear. For example, although binaries are thought to be responsible for phenomena such as retrograde hot Jupiters, tidally locked and eccentric planets (Naoz et al., 2011; Storch et al., 2014), correlations between binary companions and these planet classes are conflicted and limited by small number statistics (Wang et al., 2014; Ngo et al., 2015, 2016). Thus, constraining the prevalence of close binaries in *Kepler* is necessary to understand the occurrence, radius distribution, system architecture, and dynamical environments of the exoplanet population.

1.1 Signatures of Planetary and Stellar Companions to Exoplanets

Though *Kepler* is less sensitive to distant planetary and stellar companions, these objects leave signatures of their presence on the known planet population nonetheless. For example, the presence of hot Jupiters suggest that planets can form far from their host star and migrate inward, possibly due to gravitational perturbations of a companion (Dawson and Johnson, 2018). Gaps in resonance structures of multi-planet systems are thought to indicate the presence of "missing pea" planets orbiting between consecutive known planets, either too small or too inclined to produce a detectable transit signal (see Thomas et al., 2025, and references therein). Beyond this, several planets on highly eccentric, inclined, and even retrograde orbits have been observed, and may be imprints of a past stellar flyby or the presence of a third body in the system (see for example Naoz, 2016; Rasio and Ford, 1996, , etc.). These signatures of hidden planets offer opportunities for us to search for planetary and stellar companions missed by *Kepler* to paint a more complete picture of the exoplanet population as a whole.

Identifying these companions and understanding their prevalence is necessary to understanding how companions sculpt the exoplanet population more broadly. A sample of planets known to have or lack companions will allow us to test theories about whether and how planets form with companions. For example, a lack of close companions in the planet census would suggest that companions suppress planet formation (Kraus et al., 2016; Ziegler et al., 2018; Hirsch et al., 2021). Fewer or smaller planets in systems with companions would indicate that companions deplete planet formation reservoirs, as suggested by Quintana et al. (2007) and Jang-Condell (2015). Finally, correlations between companions and dynamically rich orbits or anomalous interior structures would confirm theories that these scenarios are the byproduct of perturbations from an outer companion (see for example Naoz, 2016; Li et al., 2014c; Stephan et al., 2019, etc.).

One of the main findings of this dissertation is the discovery of *Kepler*-1656c (Chapter 2, which paints a more complete picture of the *Kepler*-1656 system as one that is sculpted by the presence of a companion that was originally missed by *Kepler*. The prime *Kepler* mission detected a planet around *Kepler*-1656 whose size placed it firmly in the sparely populated sub-Saturn regime (Borucki et al., 2010). Radial velocity follow-up of the host star revealed a dense interior structure and a highly eccentric orbit for this planet, suggesting a possible history of gravitational perturbations, planet-planet scattering, and tidal migration (Brady et al., 2018; Millholland et al., 2020). I investigated this system using a combination of radial velocity observations and dynamical simulations to show that surprisingly, while there is indeed a planetary companion that sculpts the inner planet's eccentric orbit, it does so gently, i.e. without inducing tidal migration.

1.2 Searches for Stellar Companions to Known Planet Hosts

Kepler's insensitivity to unresolved binaries is due to it's large ≈4" pixels and considerable distance (∼1 kpc) to most targets in its field of view, which causes binaries with $\lesssim 100\,\mathrm{au}$

separations to appear as single stars. These unresolved binaries can produce transit-like lightcurves and cause false positive planet detections. Alternatively, even when a transit signal is indeed due to a planet, an unresolved companion star can still dilute the transit signal, which leads to overestimated stellar and derived planet radii (Ciardi et al., 2015) as well as systematic errors in the inferred stellar properties (El-Badry et al., 2018a). Constraining the prevalence of binaries among *Kepler* hosts is thus critical to the precision and purity of the sample, which is the backbone of most exoplanet demographic studies to date.

One of the major challenges facing previous searches for binaries in the *Kepler* sample is their limited sensitivity to close ($\lesssim$50 au) binaries. A majority of binary searches were conducted via direct imaging surveys, which have limited sensitivity at the separations smaller than 50 au and are thus blind to 50% of the parameter space where binaries reside. While a number of studies have employed advanced imaging techniques such as non-redundant masking (e.g., Kraus et al., 2016; Wang et al., 2014; Horch et al., 2014; Matson et al., 2018), these studies are still only able to probe down to separations of 20-30 au, and reach different conclusions about planet suppression in binaries. Separately, Hirsch et al. (2021) collected imaging and RV observations of a volume-limited sample of binary and single stars to search for planets in both populations. This study detected suppressed planet occurrences in close binaries, but their sample was insensitive to lower mass planets ($< 50\,M_{\oplus}$) and suffered from small number statistics due to the small total number of detected planets $\sim$ 30). Additional searches for double-lined spectroscopic companions (SB2, see Kolbl et al., 2015) showed that close binaries do exist among planet hosts, but limited sensitivity to wider binaries left open questions about their overall occurrence.

Binaries are typically identified using direct imaging or searching for offset sets of spectral features. These methods are expensive, making it difficult to systematically search many stars at once, rendering previous surveys incomplete. Additionally, these methods often miss close ($\sim$1-100 au) binaries that make up a significant fraction of the underlying star

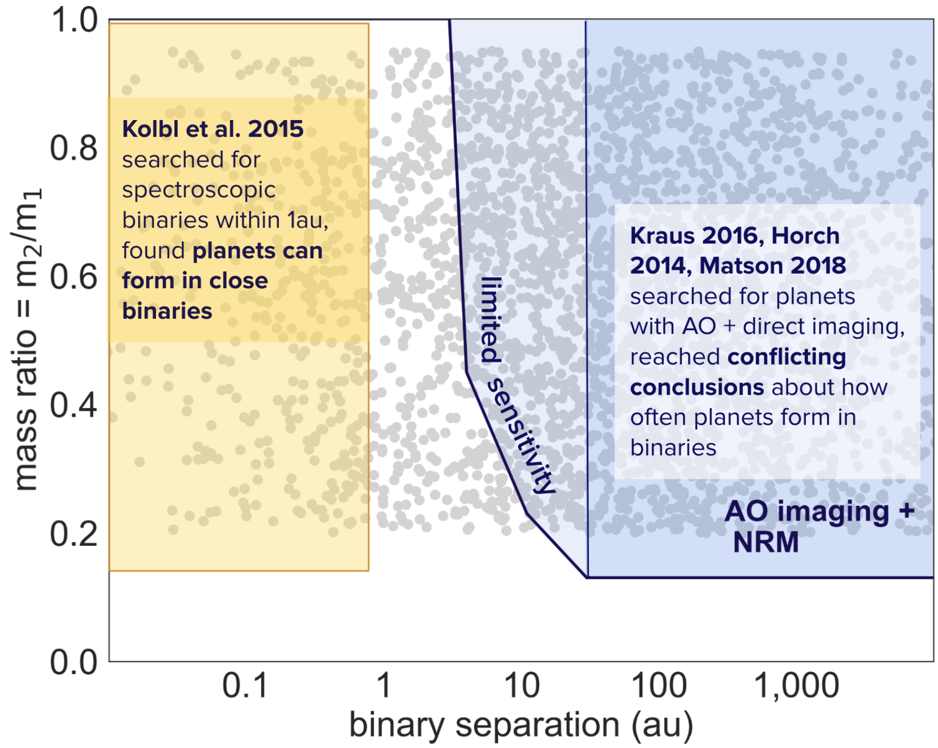

Figure 1.1: Points show simulated companions to field stars based on Raghavan et al. (2010). The separation distribution is log-normal and peaks at 40 au. Shaded regions show sensitives of previous imaging (blue) and spectroscopic (yellow) for binaries. There is broad consensus that the prevalence and properties of stellar companions to transiting planets is indistinguishable from that of field stars beyond 100 au. However, there are large uncertainties within 100 au, where dynamical effects are strongest. This dissertation aims to address major uncertainties in prevalence and properties of companions to planets hosts within 100 au by leveraging data-driven models to search for binaries at 1-200 au separations.

population (Figure 1.1). Binaries within this separation range are not typical spectroscopic binaries in that they do not possess two sets of spectral features with a detectable velocity shift. Instead, these binaries produce two sets of *overlapping* spectral features and can be identified via Bayesian model comparison of a single-star spectrum with a binary spectrum–that is, a linear combination of two single-star spectra (e.g., El-Badry et al., 2018b, etc.). They key to success for these methods is developing spectroscopic models that are accurate enough to resolve subtle deviations from single star spectra that overlapping spectra possess. Historically, spectroscopic models of stars have come from synthetic models and are limited by issues with computational efficiency, incomplete line lists, and radiative transfer calculations associated with physical stellar models. Fortunately, recent advances in data-driven spectroscopy offer increased accuracy in their predictions of stellar flux and may be a promising avenue for searches for planet-hosting binaries.

1.3 A Brief History of Data-Driven Spectroscopy

The origins of data-driven spectroscopy date as far back as Annie Jump Cannon, who worked as a Harvard computer to characterize thousands of stars by eye and create a renowned spectral classification system still used today (Cannon and Pickering, 1918, 1924). This classification was based on relative depths of spectral features that, at that time, could only be discerned by the well-trained eye. Today, it is understood that the spectral classes are in fact tracers of star properties like surface temperature, radius, and elemental abundances that imprint distinct markers into stellar spectra. In practice, this implies that spectra can be used to determine stellar properties, and vice versa.

More recently, data-driven models have sought out to accomplish with algorithms what Annie Jump Cannon achieved by eye in her catalogs– that is, to efficiently determine star properties based on observations of their spectra. In particular, *The Cannon*, named fondly after Annie Jump Cannon, is a supervised classification algorithm that learns the relationship between stellar labels (i.e., stellar properties like $T_{\rm eff}$, $\log g$, Fe/H, etc.) and per-pixel flx of stellar spectra (see Ness et al., 2015; Casey et al., 2016). Models trained by *The Cannon* are capable of deriving star properties from stellar spectra and predicting spectra of a star given its known labels. It is particularly successful at label transfer processes, whereby models are trained on samples of stars with pre-determined labels to infer properties of newly observed stars with high fidelity. *The Cannon*has been trained on a numerous datasets and has been used to derive accurate labels for 55,000 stars in APOGEE DR10 (Ness et al., 2015), determine masses and ages for 230,000 LAMOST Giants (Ho et al., 2017), and model spectra and abundances for 17,000 M Dwarfs in SDSS-V (Behmard et al., 2025), among many other applications.

There are other methodologies within spectroscopy that leverage data-driven and/or machine learning to model stellar spectra. These methodologies span a wide range of applications and degrees of complexity. For example, `SpecMatch-Emp` characterizes stars observed with Keck/HIRES by comparing their spectra to a library of well-characterized "touchstone" stars with known properties (Yee et al., 2017). Starfish (Czekala et al., 2015) emulates stellar spectra by computing a probability distribution over several probable spectra in a grid of high-resolution synthetic spectra. More recently, *The Payne* leveraged the computational efficiency and complexity of neural networks to train spectroscopic models on synthetic data from APOGEE (Ting et al., 2019). The Data-Driven Payne builds on these methods, and supplements a neural network trained on real spectra with differential (i.e., gradient) synthetic spectra from physical stellar atmospheric models (Ting et al., 2017; Xiang et al., 2019).

In Chapters 3 and 4 of this thesis, I outline two projects where I utilized The Cannon to train data-driven models on spectral from the Gaia Data Release 3 Radial Velocity Spectrometer (RVS) and Keck High Resolution Echelle Spectrometer (HIRES; Vogt et al., 1994). I found that these models are excellent at label transfer, however, more demanding applications like binary detection are limited by the models' ability to accurately emulate stellar spectra. I also describe my open-source model of Gaia DR3 spectra (`gaiaspec`) and present a novel wavelet-based method for removing non-astrophysical variations in stellar spectra. Finally, I comment on the applications and limitations of data-driven models more broadly, and offer insights into the prospects of data-driven searches for planet-hosting binaries.

Chapter 2

Kepler-1656b's Extreme Eccentricity: Signature of a Gentle Giant

A version of this chapter was previously published in *The Astronomical Journal* (Isabel Angelo et al., 2022, AJ, 163, 227).

Abstract

Highly eccentric orbits are one of the major surprises of exoplanets relative to the Solar System and indicate rich and tumultuous dynamical histories. One system of particular interest is Kepler-1656, which hosts a sub-Jovian planet with an eccentricity of 0.8. Sufficiently eccentric orbits will shrink in semi-major axis due to tidal dissipation of orbital energy during periastron passage. Here our goal was to assess whether Kepler-1656b is currently undergoing such high-eccentricity migration, and to further understand the system's origins and architecture. We confirm a second planet in the system with $M_{\mathrm{c}} = 0.40 \pm 0.09 M_{\mathrm{jup}}$ and $\mathrm{P}_{\mathrm{c}} = 1919 \pm 27$ days. We simulated the dynamical evolution of planet b in the presence of

planet c and find a variety of possible outcomes for the system, such as tidal migration and engulfment. The system is consistent with an in situ dynamical origin of planet b followed by subsequent Eccentric Kozai Lidov (EKL) perturbations that excite Kepler-1656b's eccentricity gently, i.e. without initiating tidal migration. Thus, despite its high eccentricity, we find no evidence that planet b is or has migrated through the high-eccentricity channel. Finally, we predict the outer orbit to be mutually inclined in a nearly perpendicular configuration with respect to the inner planet orbit based on the outcomes of our simulations, and make observable predictions for the inner planet's spin-orbit angle. Our methodology can be applied to other eccentric or tidally locked planets to constrain their origins, orbital configurations and properties of a potential companion.

2.1 Introduction

Despite their absence in our own solar system, highly eccentric planets have been discovered in a number of exoplanet systems. In contrast to the low-eccentricity orbits of planets in our own solar system, with mean eccentricity $\langle e \rangle = 0.06$, giant planets in radial velocity surveys have been found to span the full possible range of $0 \leq e < 1$ for bound orbits, with eccentricities as high as $e = 0.9$ in some cases (e.g., Masuda, 2017; Blunt et al., 2019; Schlecker et al., 2020). This range of eccentricities points to a diverse variety of formation scenarios (see Winn and Fabrycky, 2015, for a review).

Adding to this picture, a subset of these eccentric planets are giants orbiting much closer to their host star than the solar system Jovians. Roughly 1% of stars possess "Hot Jupiters" (giant planets with $P = 1 - 10$ days, see Udry and Santos, 2007; Cumming et al., 2008; Howard et al., 2010c). This number is even larger for for giant planets with $P < 200$ days (10.5%, see Wittenmyer et al., 2016). This class of close-orbiting giant planets must be corroborated by theories of planet formation, which originally predicted giant planets to form and accrete material further out beyond their host star's "ice line" (e.g., Boss, 1997; Rubie et al., 2015; Morbidelli et al., 2015).

Eccentric orbits are a relic of planet formation processes and/or the influence of another body— either a planet, brown-dwarf or star— in the system. Because giant planets possess significant envelopes, they must form in early disk phases when some gas is still present ($t <$ 10 Myr, see Wyatt, 2008). Since the presence of this gas tends to dampen eccentricity (e.g., Artymowicz et al., 1991; Ward, 1988; Lee and Chiang, 2016), some post-nebular eccentricity excitation mechanism is required to explain their eccentric orbits. For example, a planet can be perturbed into an eccentric orbit via disk migration, planet-planet scattering, or an outer perturber (e.g., Rasio and Ford, 1996; Fabrycky and Tremaine, 2007; Wu et al., 2007; Shara et al., 2016; Naoz et al., 2011, 2012; Dawson and Johnson, 2018). At this point, the planet may undergo high eccentricity migration, where it is subject to strong tidal forces at closest approach that dissipate energy and shrink the planet's orbit over time.

In this paper, we focus on Kepler-1656b (a.k.a KOI-367, KIC-4815520), a warm, highly eccentric sub-Saturn orbiting a Sun-like host star (Brady et al., 2018). Kepler-1656b's orbital configuration makes it a strong candidate for high eccentricity migration, and is likely a relic of aforementioned post-nebular eccentricity excitation[1]. Moreover, Sub-Saturns like Kepler-1656b lie in the sparsely populated regime of planets between the size of Neptune and Saturn, with densities spanning over an order of magnitude for a fixed size (Petigura et al., 2017c, 2018a). Planets of these sizes pose a challenge to standard core nucleated accretion theory, which predicts planets with similar-mass cores to have undergone runaway accretion and accumulate much larger gas envelopes than observed (Pollack et al., 1996). For example, a $10\,M_\oplus$ core at 5 au initiates runaway gas accretion in just 10^6 years. While there are a number

[1] While it's possible that the planet opened a gap during the disk phase and reduced the eccentricity damping, even then the eccentricity is not readily increased to ~ 0.8 values (Ward, 1988; Goldreich and Sari, 2003; Teyssandier and Ogilvie, 2016, 2017; Ragusa et al., 2018), and the expectation is that a more massive planet than Kepler-1656b is needed to open a gap (depending on the disk's viscosity, e.g., Armitage, 2007; Crida and Morbidelli, 2007; Edgar et al., 2007; Duffell and MacFadyen, 2013).

of theories that explain the envelope fractions seen in some sub-Saturns, (see for example Lee and Chiang, 2016; Millholland et al., 2020; Lee et al., 2014; Lee and Chiang, 2015), the detailed origin scenarios for these planets remains a mystery. Specifically, a tension exists between theories that sub-Saturns and giant planets more generally form in situ, or accrete material further out from their host star and migrate inwards later on (e.g, Batygin et al., 2016). This tension extends to Kepler-1656b, whose birth scenario (i.e., in situ versus inward migration) should corroborate its eccentric orbit and high reported core mass fraction (see Millholland et al., 2020).

Here we use Kepler-1656b as a case study of combining observational data and dynamical analysis to constrain a system's detailed architecture and dynamical history. Originally, Brady et al. (2018) predicted an upper limit of $\dot{\gamma}$ <1.4 $ms^{-1}yr^{-1}$on the long-term stellar radial velocity (RV) trend, with observations disfavoring— but not ruling out— a planetary or stellar companion in the system. In this study, more recent RV observations suggest underlying periodic RV behavior with period $P \gg 100$ days. We identify a giant planetary companion to Kepler-1656b and find that perturbations from said companion is driving Kepler-1656b into its eccentric orbit gently, i.e. without inducing high eccentricity migration. In doing so, we identify two classes of outer companions to sub-Saturns and giant planets based on their signature on the inner planet. The first class, "gentle companions", nudge the inner planet into an eccentric orbit in situ, without inducing migration. The second class, "strong companions" have a more dramatic effect on the inner planet, causing it to migrate inwards and settle in a close-in, tidally locked orbit. These classes of companions produce observable signatures on their inner companion's orbital configuration, and can potentially sculpt large-scale trends in exoplanet system architectures.

We begin by confirming the existence and planetary nature of the outer companion Kepler-1656c in Section 2.2. We then describe the physical setup and initial conditions we used to simulate the dynamical evolution of Kepler-1656b in the presence of this companion in

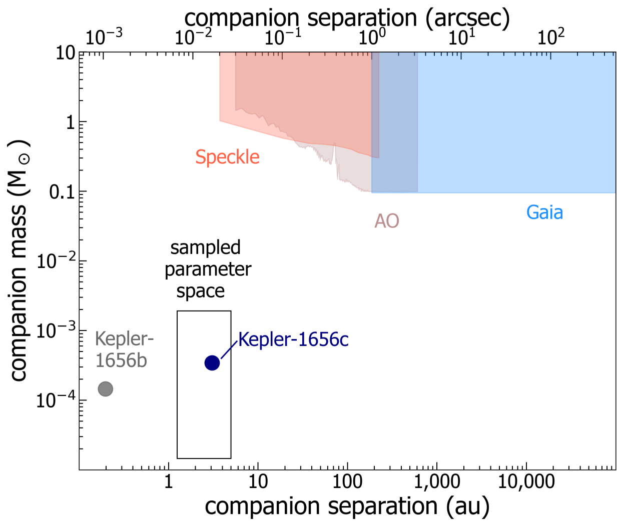

Figure 2.1: We consider regions of mass-separation parameter space in which a companion to Kepler-1656b can reside. Regions ruled out by Gaia (blue), AO (tan), and Speckle imaging (red) are shown. The explored region encapsulating Kepler-1656c is outlined, and both Kepler-1656b and Kepler-1656c are overplotted for reference.

Sections 2.3.1 and 2.3.2 respectively. In Section 2.4, we predict observable properties of the system's orbital configuration and discuss the dynamical origins of Kepler-1656b and eccentric and tidally locked planets more generally. Finally, in Section 2.5 we discuss how our proposed origin scenarios fit into our understanding of planet formation and the current exoplanet population at large.

2.2 Observational Signatures of a Planetary Companion

2.2.1 High-Resolution Imaging

We searched for companions to Kepler-1656b in imaging data and did not find evidence for a stellar companion in any of our images. To start, we used the Gaia database to search for nearby, co-moving companions to Kepler-1656. We found no targets within 10^5 au of the host star. The Gaia DR2 catalogue is complete to magnitudes as high as G$\sim 12-18$ (Gaia

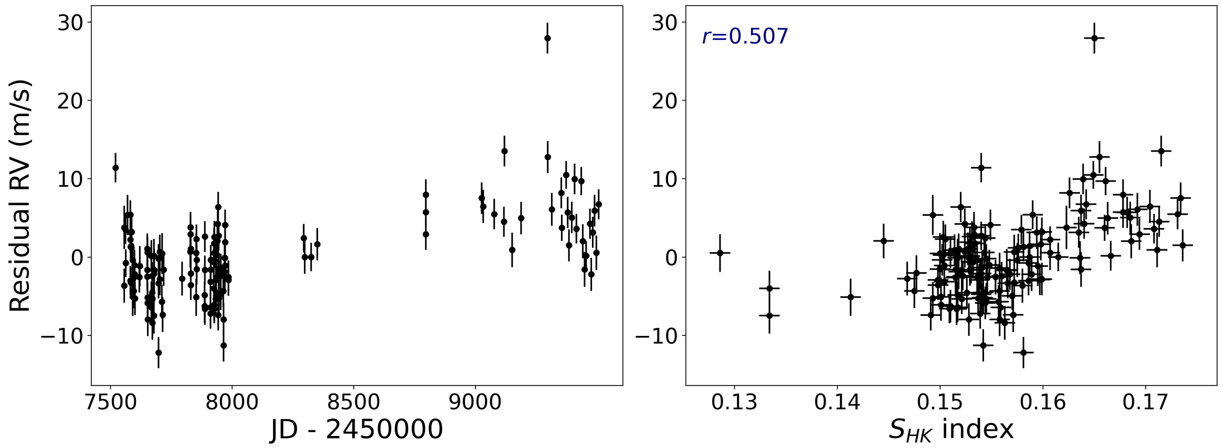

Figure 2.2: *Left*: Residual RV time series of Kepler-1656 shows a long-term trend with $P \gg 100$ days. *Right*: Residual RVs are plotted versus S_{HK}. A Pearson r statistic of 0.51 and median activity signal of S_{HK}=0.15 indicate that stellar activity is not sufficient at explaining the variability in our RV time series.

.

Collaboration et al., 2018), which conservatively rules out M> $0.1 M_{\odot}$ companions beyond Gaia's 1" detection limit. This is illustrated by the blue region in Figure 2.1.

We also searched for companions to Kepler-1656b using speckle and adaptive optics (AO) images of the host star. These images provide sensitivity to companion magnitudes as a function of radial distance from the host star. Following a similar process to that outlined in Furlan et al. (2017a), we converted to these sensitivities to corresponding companion masses using an interpolated magnitude-mass relationship from Pecaut and Mamajek (2013a). Speckle images taken by the Gemini-N telescope with a 692 nm-center wavelength filter effectively rule out companions residing in the red region of parameter space in Figure 2.1, and AO images from the Exoplanet Follow-up Observing Program (ExoFOP) archive [2] taken by the Keck-2 telescope *Ks* filter rule out the tan region. As can be seen in Figure 2.1, imaging observations of Kepler-1656 effectively rule out stellar-mass companions. Thus, the dynamical picture we consider for the remainder of our analysis is a 2-planet system in which Kepler-1656b is accompanied by a second planet in the system.

[2]https://exofop.ipac.caltech.edu/

2.2.2 Radial Velocities

Observations of Kepler-1656 were obtained through the California Planet Search (CPS; Howard et al., 2010a) using the High Resolution Echelle Spectrometer on the Keck-1 10m telescope (HIRES; Vogt et al., 1994). We collected 50 spectra between 07-11-2017 and 10-24-2020. These spectra were used along with 100 spectra reported in (Brady et al., 2018) to compute RV time series data for Kepler-1656 using standard CPS methodology. For all observations, spectra were observed through an iodine cell mounted in front of the spectrometer for wavelength calibration, with an SNR of 110 per pixel at 500nm. Our RVs are listed in Table 2.1.

Table 2.1: Radial Velocity and Activity Measurements

Time	**RV**	σ**(RV)**	S_{HK}
2457942.954	13.892	1.922	0.152
2457943.009	0.213	1.966	0.149
2457944.956	16.476	1.819	0.154
2457945.004	11.981	2.275	0.148
2457945.994	17.432	2.066	0.149

Table Comments: A portion of the dataset is shown here. S_{HK} errors are all set to 0.001, and RV uncertainties are photon-limited.

Our RV observations reveal a long-term trend in addition to the short-period signal produced by Kepler-1656b that is suggestive of underlying periodic RV behavior with period $P \gg 100$ days. This behavior can be seen in the residual RVs after the inner planet signal has been subtracted (Figure 2.2, left panel).

We first considered the possibility that this observed trend was due to stellar activity. Variability in magnetic field over the star's activity cycle can lead to apparent RV variability on multi-year timescales. In some cases, this variability can be mistaken for a planet signal (Robertson et al., 2014; Haywood et al., 2014). Using formalism outlined in Isaacson and Fischer (2010), we monitored stellar activity through Kepler-1656's S_{HK} index, which traces the CaII H&K emission lines as an indicator of chromospheric activity (e.g. Wilson, 1968; Duncan et al., 1991; Baliunas et al., 1995). Our results are shown in Table 2.1.

In the event that the observed trend is due to stellar activity, we should detect a strong correlation between the residual (i.e., inner planet subtracted) RV time series and the S_{HK} index, with $S_{\mathrm{HK}} \gtrsim 0.2$, consistent with significant activity for stars of similar color to Kepler-1656 (Isaacson and Fischer, 2010, Figure 9). We show the residual RV data as a function of S_{HK} in the right panel of Figure 2.2. We computed a Pearson r coefficient of 0.51 for this data set, suggesting a modest correlation; however, this is not uncommon for stars of this type and does not necessarily imply an activity-induced RV trend (Wright et al., 2008). Moreover, the data have a median S_{HK} of 0.15, similar to low-activity stars of similar color. S_{HK} values within this range are consistent with jitter $\sigma_{\mathrm{j}} \approx 2\mathrm{ms}^{-1}$, which cannot account for the full $\approx 7\mathrm{ms}^{-1}$ variability we see in our RVs. Thus, our observed RV trend is most likely due to an undetected companion–either a star, brown dwarf, or second planet– in the system.

2.2.3 Periodogram

We conducted a periodogram search for an outer planet in our RV data using *RVSearch*, an algorithm that iteratively searches for periodic signals in RV time series data (Rosenthal et al., 2021). We finely sampled a grid of outer planet periods (P_{c}) with $31 < P_{\mathrm{c}} < 10,000$ days. For each P_{c}, *RVSearch* compares the maximum likelihood 2-planet model to the single-planet model by evaluating the change in the BIC (see Schwarz, 1978). Peaks in the periodogram thus correspond to the most plausible periods for the outer planet. A

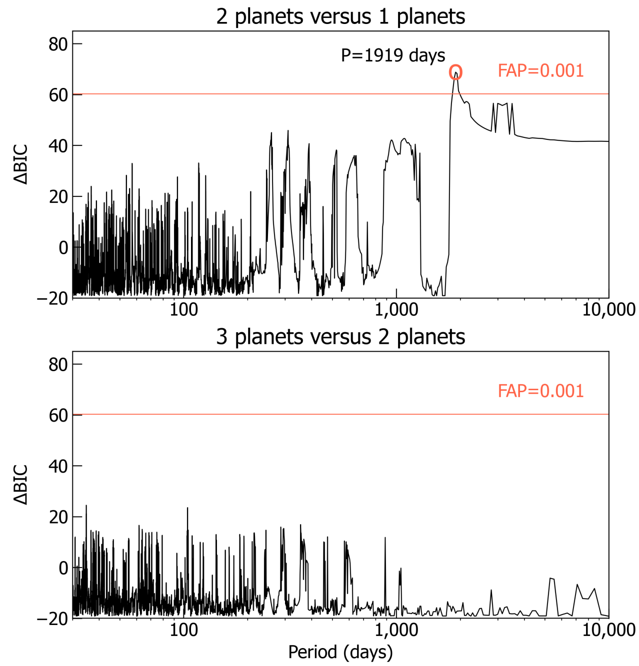

Figure 2.3: *Top:* Change in Bayesian Information Criterion (BIC) comparing 2-planet models and single-planet models. The trial period of the 2-planet model is shown on the horizontal axis. The BIC favors $P_\mathrm{c} \sim 1920\,$days, corresponding to a second planet in the system. The BIC associated with eFAP=0.001, above which peaks are considered true planets, is over-plotted in red. *Bottom:* The same as the top panel, but comparing 3-planet models to 2-planet models. No peaks lie above the eFAP=0.001 threshold, indicating no additional periodic signals are detected.

periodogram search for 2-planet signals is shown in the top panel of Figure 2.3. We also plot the ΔBIC threshold for false detections, corresponding to an empirical false alarm probability (eFAP) of 0.1%.

As can be see in the top panel of Figure 2.3, the BIC strongly favors the long-period candidate peak over the single-planet model with $\Delta\mathrm{BIC} = 79$. This model is also favored over other potential 2-planet models (i.e., other peaks in the periodogram) with $\Delta\mathrm{BIC} \gtrsim 30$, and lies above the false detection threshold. We find no additional periodic signals after subtracting the 2-planet signal from our RVs (i.e., considering the 3- versus 2-planet case), as shown in the bottom panel of Figure 2.3. Thus, we conclude that the long-term trend in our RVs is due to a second planet in the system, Kepler-1656c, whose orbital parameters correspond to the $P_c \sim 1920$ day peak in the periodogram.

Figure 2.4 shows the best-fit 2-planet Keplerian model of the system, including both Kepler-1656b and Kepler-1656c.[3] We summarize our adopted system parameters from this model in Table 2.2. We find that Kepler-1656c is a giant planet with $M_c = 0.43 \pm 0.10\, M_{\rm jup}$, semi-major axis $a_c = 3.057^{+0.059}_{-0.052}$ au, and eccentricity $e_c = 0.53^{+0.064}_{-0.08}$. For the remainder of our analysis, we explore a parameter space in the vicinity Kepler-1656c (see Figure 2.1) to to uncover the origins of Kepler-1656b's high eccentricity and to understand the dynamical evolution of sub-Saturns with outer companions more generally.

Table 2.2: Kepler-1656 System Parameters

Parameter	**Value**	**Notes**
Stellar		
$M_\star$ ($M_\odot$)	1.03 ± 0.04	A
$R_\star$ ($R_\odot$)	1.10 ± 0.13	A
age (Gyr)	$6.31^{+2.1}_{-2.9}$	A
$v\sin i$ (km s^{-1})	2.8 ± 1.0	B
Planet b		
P_b (days)	31.562 ± 0.011	C
$T_{c,b}$ (days)	2455011.47 ± 0.90	C
e_b	$0.838^{+0.045}_{-0.029}$	C
ω_b	$0.922^{+0.073}_{-0.095}$	C
K_b (m s^{-1})	$17.1^{+5.3}_{-1.6}$	C
a_b (au)	0.1974 ± 0.0026	D

[3] We note that there is an outlier in our RV data that do not fit the model. We examined the pipeline reduction and spectrum associated with this RV point, but do not find any anomalies. Thus, we include the outlier in our analysis.

Table 2.2 Continued

Parameter	Value	Notes
$M_{\rm b}\sin i$ ($M_\oplus$)	$47.8^{+6.2}_{-3.3}$	D
$M_{\rm b}\sin i$ ($M_{\rm jup}$)	$0.150^{+0.019}_{-0.010}$	D
$M_{\rm b}$ ($M_\oplus$)	$47.8^{+6.2}_{-3.3}$	D,E
$M_{\rm b}$ ($M_{\rm jup}$)	0.15 ± 0.02	D,E
Planet c		
$P_{\rm c}$ (days)	1919^{+27}_{-24}	C
$T_{\rm c,c}$ (days)	2459461^{+24}_{-26}	C
$e_{\rm c}$	$0.527^{+0.050}_{-0.054}$	C
$\omega_{\rm c}$	$1.53^{+0.24}_{-0.28}$	C
$K_{\rm c}$ (m s^{-1})	$6.35^{+0.56}_{-0.52}$	C
$a_{\rm c}$ (au)	3.053 ± 0.049	D
$M_{\rm c}\sin i$ ($M_\oplus$)	107.2 ± 10.2	D
$M_{\rm c}\sin i$ ($M_{\rm jup}$)	0.337 ± 0.032	D
$M_{\rm c}$ ($M_\oplus$)	126.4 ± 28.9	D,F
$M_{\rm c}$ ($M_{\rm jup}$)	0.40 ± 0.09	D,F

Table Comments: A: Johnson et al. (2017a), B: Petigura et al. (2017a), C: This work, adopted Keplerian model, D: This work, derived, E: Gaussian priors imposed on $M_{\rm b}\sin i$ based on our RV analysis and $i = 89.31 \pm 0.51$ reported in Brady et al. (2018), F: Gaussian prior imposed on $M_{\rm c}\sin i$ based on our RV analysis, and uniform prior imposed on $\cos(i)$.

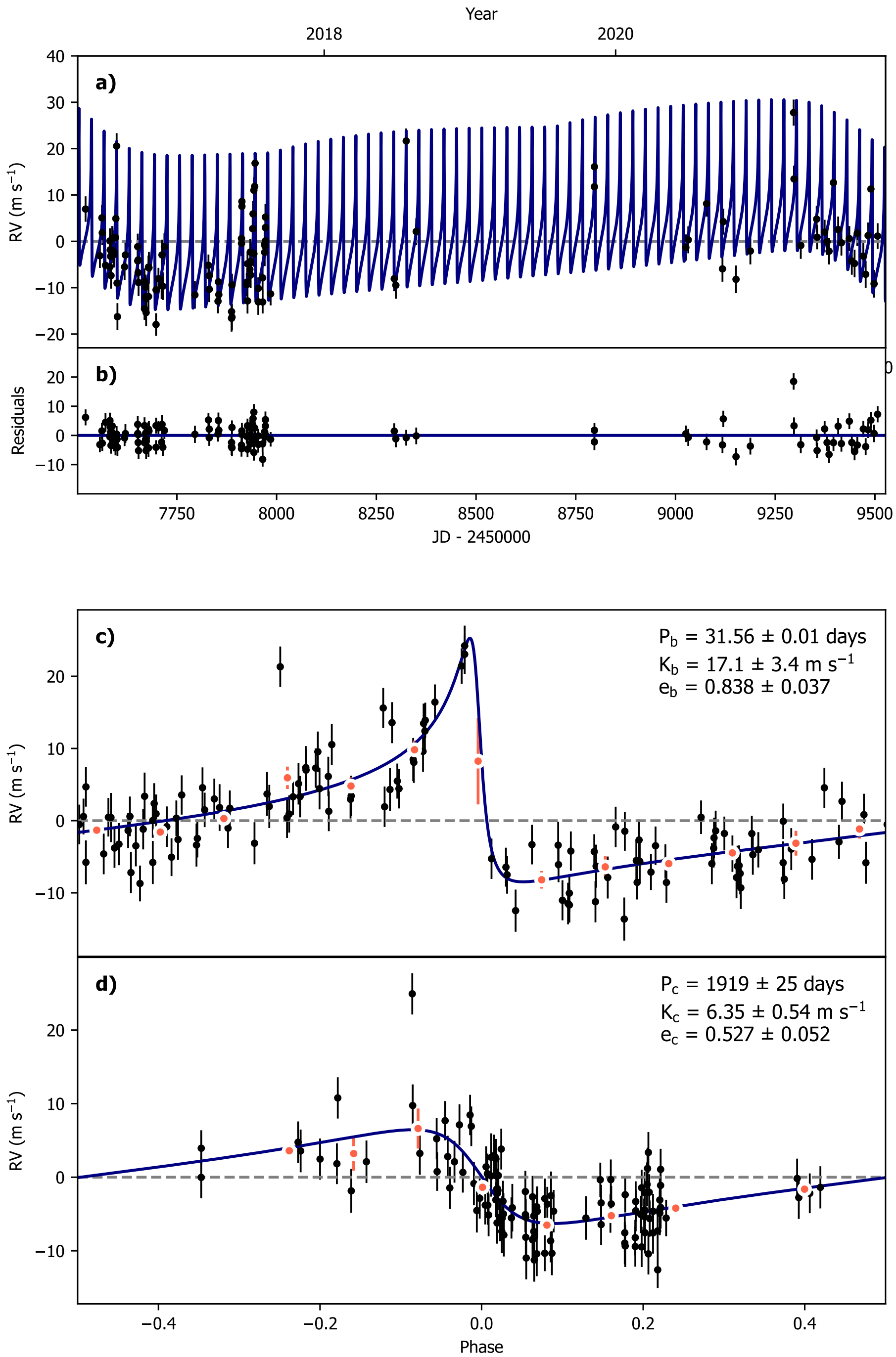

Figure 2.4: Adopted 2-planet Keplerian model for the Kepler-1656 system. **a)** Black points are RVs and the blue line shows our maximum a posteriori model. **b)** Model residuals are shown in black. **c)** Observations are phase-folded over the inner planet period and the phase-folded Keplerian model is over-plotted. RVs binned in 0.08 units of orbital phase are shown in red. **d)** Observations are phase-folded over the outer planet period, phase-folded Keplerian model is over-plotted.

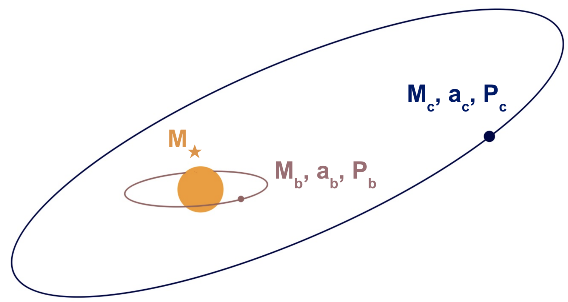

Figure 2.5: Schematic of a hierarchical triple system containing Kepler-1656b ($M_{\rm b}$) on a tight inner orbit and a companion ($M_{\rm c}$) on a distant, possibly inclined outer orbit around their host star ($M_\star$). Note that the sizes of the planets and their orbits are not drawn to scale.

2.3 Dynamical Model

Our observations in Section 2.2 revealed an outer planetary companion to Kepler-1656b. Gravitational perturbations from this companion should have a clear signature on the evolution and orbit of the inner planet. Thus, we can simulate the dynamical evolution of the Kepler-1656 system from an array of initial configurations to investigate formation scenarios that are consistent with the present-day properties.

An outer companion can excite a planet's eccentricity via dynamics that arise from a hierarchical three-body configuration- that is, a tight inner orbit of two bodies and a long-period outer orbit of a third body around the inner orbit's center of mass (see Figure 2.5 for visual depiction). In this case, the inner and outer orbits torque eachother and exchange angular momentum over timescales much longer than the planets' orbital periods. This exchange produces long-term eccentricity and inclination oscillations known as the the Eccentric Kozai-Lidov (EKL) mechanism (Kozai, 1962; Lidov, 1962; Naoz, 2016). In this regime, the secular (i.e. long-term, orbit-averaged) evolution of the system is chaotic and the planet can undergo extreme eccentricity oscillations (e.g., Li et al., 2014a). This evolution can lead to interesting dynamics, such as engulfment of the inner planet, retrograde orbits or high eccentricity migration (e.g., Naoz et al., 2011, 2012; Lai et al., 2018; Vick et al., 2019). For example, hot Jupiters on orbits that are misaligned with their host star's spin can be attributed to this mechanism (Albrecht et al., 2012; Naoz et al., 2011, 2012; Anderson et al., 2016; Storch et al., 2017; Stephan et al., 2018).

In this section, we outline the physical setup and relevant dynamical processes at play for the Kepler-1656 system - that is, a 2-planet system with a hierarchical orbital configuration subject to EKL behavior and other dynamical effects which we outline below. We then describe the underlying physics and initial conditions of our dynamical simulations.

2.3.1 Physical Setup & Relevant Timescales

Kepler-1656b's orbit together with the distant companion form a hierarchical triple system in which Kepler-1656b orbits on a tight inner orbit with respect to the companion. In this section, we examine the general dynamical landscape of hierarchical 2-planet systems like Kepler-1656 by comparing timescales associated with various physical processes that are at play. In particular, we consider hierarchical triple systems consisting of inner planets with outer companions in the vicinity of parameter space where Kepler-1656c resides (i.e., the outlined region of Figure 2.1) to capture a dynamical picture of Kepler-1656b and and similar sub-Saturns with distant companions.

For an inner companion of mass M_{b} and outer planet of mass M_{c} orbiting a host star of mass $M_\star$ the EKL-induced eccentricity oscillations occur on a characteristic timescale, hereafter

$t_{\rm EKL}$ (e.g., Antognini, 2015):

$$\begin{aligned} t_{\rm EKL} &= \frac{16}{30\pi}\frac{M_\star + M_{\rm b} + M_{\rm c}}{M_{\rm c}}\frac{P_{\rm c}^2}{P_{\rm b}(1-e_{\rm c}^2)^{3/2}} \\ &\sim 8.3\times10^4{\rm yr}\left(\frac{M_\star}{M_\odot}\right)\left(\frac{M_{\rm c}}{M_{\rm jup}}\right)^{-1} \\ &\times \left(\frac{P_{\rm b}}{30\,{\rm days}}\right)^{-1}\left(\frac{P_{\rm c}}{5{\rm yrs}}\right)^{2}\left[1-\left(\frac{e_{\rm c}}{0.5}\right)^2\right]^{-3/2}, \end{aligned} \tag{2.1}$$

where $P_{\rm b}$ and $P_{\rm c}$ are the inner and outer orbital periods and $e_{\rm c}$ is the eccentricity of the outer planet. Here we consider the lowest order level of secular approximation, namely the quadrupole-level, which describes the shortest timescale over which the eccentricity oscillates (e.g., Naoz et al., 2013a). We plot $t_{\rm EKL}$ versus inner planet semi-major axis ($a_{\rm b}$) for outer companions consistent with our observations as a blue shaded band in Figure 2.6. In our simulations, we also account for the fact that these oscillations may be modulated by the octupole-level order of the secular approximation, which can lead to more chaotic evolution and larger amplitudes (Naoz, 2016).

In addition to the EKL mechanism, there are several potentially relevant effects that may contribute to the evolution of the systems we consider. In particular, General Relativity (GR) precession, as well as tidal effects like circularization and shrinking, all contribute to the secular dynamical evolution of the Kepler-1656 system. While the EKL mechanism excites eccentricity, the aforementioned processes tend to suppress it (e.g., Fabrycky and Tremaine, 2007; Ford et al., 2000; Naoz et al., 2013b). For instance, the GR precession of the inner planet orbit occurs on a typical timescale of

$$\begin{aligned} t_{\rm GR} &= 2\pi\frac{a_{\rm b}^{5/2}c^2(1-e_{\rm b}^2)}{3G^{3/2}(M_\star+M_{\rm b})^{3/2}} \\ &\sim 2.2\times10^5{\rm yr}\left(\frac{M_\star}{M_\odot}\right)^{-3/2}\left(\frac{a_{\rm b}}{0.2\,{\rm au}}\right)^{5/2} \\ &\times \left[1-\left(\frac{e_{\rm b}}{0.8}\right)^2\right], \end{aligned} \tag{2.2}$$

where G is the gravitational constant and c is the speed of light. This precession suppresses EKL eccentricity excitations if $t_{\rm GR}$ is shorter than $t_{\rm EKL}$ (i.e., the upper left region of Figure 2.6).[4]

Given Kepler-1656b's high eccentricity it is also important to compare $t_{\rm EKL}$ to the shrinking and circularization timescales ($t_{\rm shrink}$, and $t_{\rm circ}$, respectively) of the inner orbit due to tides. We estimate these timescales for inner planet orbit due to equilibrium tides (e.g., Eggleton et al., 1998; Eggleton and Kiseleva-Eggleton, 2001), which for $e_{\rm b} \sim 0.8$ can be approximated as:

$$\begin{aligned} t_{\rm shrink} &\sim \frac{t_{\rm v,b} a_{\rm b}^8 M_{\rm b}^2 (1-e_{\rm b}^2)^{15/2} f_T(e_{\rm b})}{162(1+2k_{\rm s,b})^2 R_{\rm b}^8 M_\star (M_\star + M_{\rm b}) e_{\rm b}^2} \\ &\sim 4.1 \times 10^{10} {\rm yr} \left(\frac{M_\star}{M_\odot}\right)^{-2} \left(\frac{a_{\rm b}}{0.2\,{\rm au}}\right)^8 \left(\frac{M_{\rm b}}{50 M_\oplus}\right)^2 \\ &\times \left(\frac{R_{\rm b}}{5R_\oplus}\right)^{-8} \left[1 - \left(\frac{e_{\rm b}}{0.8}\right)^2\right]^{15/2} \left(\frac{e_{\rm b}}{0.8}\right)^{-2} \times f(e_{\rm b}) \end{aligned} \tag{2.3}$$

and

$$\begin{aligned} t_{\rm circ} &\sim \frac{t_{\rm v,b} a_{\rm b}^8 M_{\rm b}^2 (1-e_{\rm b}^2)^{13/2} f_T(e_{\rm b})}{81(1+2k_{\rm s,b})^2 R_{\rm b}^8 M_\star (M_\star + M_{\rm b})} \\ &\sim 1.5 \times 10^{11} {\rm yr} \left(\frac{M_\star}{M_\odot}\right)^{-2} \left(\frac{a_{\rm b}}{0.2\,{\rm au}}\right)^8 \left(\frac{M_{\rm b}}{50 M_\oplus}\right)^2 \\ &\times \left(\frac{R_{\rm b}}{5R_\oplus}\right)^{-8} \left[1 - \left(\frac{e_{\rm b}}{0.8}\right)^2\right]^{13/2} \times f(e_{\rm b}), \end{aligned} \tag{2.4}$$

where $R_{\rm b}$ is the radius of the inner planet, $t_{\rm v,b}$ is the viscous timescale of the inner planet, $k_{\rm s,b}$ is the classical apsidal motion constant, and we define $f(e) \equiv 1/(1 + 15/4e^2 + 15/8e^4 + 5/64e^6)$. Note that these are approximate scalings for the Kepler-1656 system in its cur-

[4]Although, in some cases, when the GR precession timescale is of the order of the quadrupole timescale, it can destabilize the resonance and potentially re-trigger high eccentricity excitations (e.g., Naoz et al., 2013b; Hansen and Naoz, 2020).

rent configuration–for our simulations, we use the full equations derived from Eggleton and Kiseleva-Eggleton (2001) which also depend on $\Omega_{s,\mathrm{b}}$, the spin rate of the inner planet (see Naoz, 2016, eq. 86-90). We adopt nominal values of $k_{\mathrm{s,b}} = 0.25$ and $t_{\mathrm{v,b}} = 1.5\,\mathrm{yr}$ in these equations. Our nominal $t_{\mathrm{v,b}}$ corresponds to a tidal quality factor of $Q_{\mathrm{b}} \approx 1.8 \times 10^6$ for the planet, where $t_{\mathrm{v,b}}$ is related to Q_{b} by following the equilibrium tide formalism from Naoz et al. (2016):

$$Q_{\mathrm{b}} = \frac{4}{3}\frac{k_{s,\mathrm{b}}}{(1+2k_{s,\mathrm{b}})^2}\frac{GM_{\mathrm{b}}}{R_{\mathrm{b}}^3}\frac{P_{\mathrm{b}}t_{v,b}}{2\pi}. \tag{2.5}$$

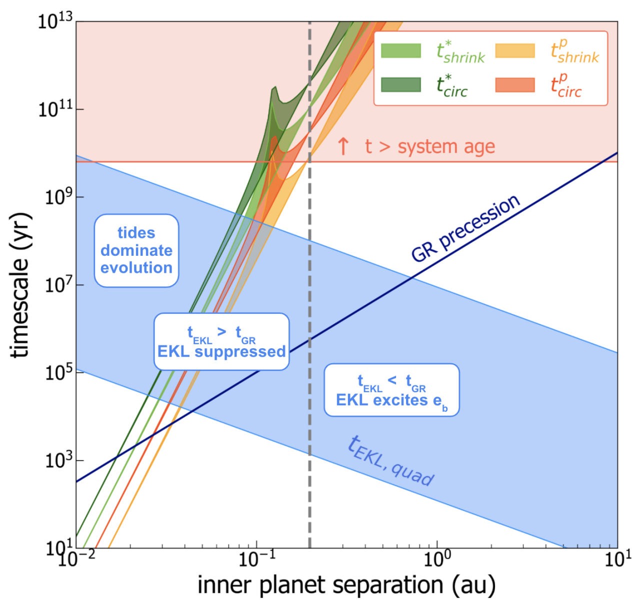

Figure 2.6: Relevant dynamical timescales for Kepler-1656b in the presence of companions similar to Kepler-1656c are plotted as a function of the inner planet orbital separation, a_{b}. The shaded blue band represents the range of eccentricity oscillation timescale t_{EKL} for outer companions in the parameter space surrounding Kepler-1656c, with $0.1M_{\mathrm{b}} < M_{\mathrm{c}} < 2M_{\mathrm{jup}}$, $P_2 = 600 - 4000\,\mathrm{days}$ ($a_{\mathrm{c}} = 1.4 - 5\,\mathrm{au}$), and $e_{\mathrm{c}} = 0 - 0.9$. For the tidal shrinking and circularization timescales, we consider a range of planet and star spin periods from 1-30 days. The upper boundary of the blue region is set by $\mathrm{a_c}$=5, M_{c}=$0.1M_b$, $\mathrm{e_c}$=0 and the lower boundary is set by a_{c}=1.4, M_{c}=$2M_{\mathrm{jup}}$, $\mathrm{e_c}$=0.9.

We show t_{shrink} and t_{circ} timescales for a range of a_{b} in Figure 2.6, where we consider both tides due to the planet and due to the star (see label). For the purpose of this figure we spin periods between $1-30$ days for both the planet and star (i.e., $\Omega_{s,\mathrm{b}}$=0.03-1 days^{-1}, shown as shaded bands). As can be seen, both the shrinking and circularization timescales are longer than the system's age for the planet's observed semi-major axis[5].

We see from Figure 2.6 that Kepler-1656b's current orbit (vertical dashed line) lies in the regime where EKL and GR precession timescales are shortest, indicating these timescales will compete to dominate the overall system behavior. Additionally, for some companions (i.e., in the left-hand region of the blue band in Figure 2.6), $t_{\rm shrink}$ and $t_{\rm circ}$ are similar in magnitude to $t_{\rm EKL}$. Accordingly, for initial configurations where Kepler-1656b orbits closer to the host star, tides may play a role in suppressing eccentricity excitations (note that these timescales are plotted for Kepler-1656b's current orbit with $e_{\rm b} \approx 0.8$, however, inner planets with higher eccentricities may have shorter tidal timescales). All this is to say that different companions will either induce or suppress eccentricity oscillations for Kepler-1656b.

We model the secular evolution of the system (up to the octupole-level of approximation) taking all of the aforementioned physical processes into account, i.e., EKL, GR, and tidal circularization and shrinking. We also include stellar evolution, which traces the host star's spin. Stellar evolution can play an important role in the dynamics of the planet. For example, an expanding star can engulf or circularize the inner planet, while a star undergoing mass loss can widen its orbit (e.g., Dobbs-Dixon et al., 2004; Stephan et al., 2017, 2018, 2020b,a). Our simulations compute the secular evolution of the inner and outer orbits over time from a suite of initial configurations which we describe in the next section.

2.3.2 Initial Conditions

In all of our simulations, we fix the host star and inner planet parameters to values reported in Table 2.2 and sample from the parameter space surrounding our best-fit values for planet c.

[5]We note that equilibrium tides tend to underestimate the efficiency of the tides compared to chaotic dynamical tides for sufficiently large eccentricity (Vick et al., 2019). Thus, for Kepler-1656b's eccentricity, the chaotic dynamical tides may result in more close-in, tidally locked systems than predicted by equilibrium tides, as we discuss in more detail in Section 2.4.1

We set the mass and radius for the host star and inner planet to $M_{\rm b} = 48.6 M_\oplus$, $R_{\rm b} = 5.02 R_\oplus$, $M_\star = 1.03 M_\odot$, and $R_\star = 1.10 R_\odot$. We considered candidate outer planets in the vicinity of Kepler-1656c with $P_{\rm c} = 600 - 4,000$ days, corresponding to a uniform distribution of $a_{\rm c} = 1.4 - 5$ au, and $M_{\rm c}$ from a log-uniform distribution within the planetary-mass regime of $0.1 M_{\rm b} < M_{\rm c} < 2 M_{\rm jup}$. We sample the mutual inclination from an isotropic distribution (i.e., uniform distribution for $\cos i_{\rm bc}$) and $e_{\rm c}$ from a uniform distribution of $(0, 1)$. These initial conditions are depicted in Table 2.3. To explore the origins of Kepler-1656b's eccentricity, we place Kepler-1656b at an near-circular orbit initially ($e = 0.01$).

We fix the viscous timescale of the star and planet to $t_{v,\star} = t_{v,b}$ =1.5 yr in our simulations. We note that while a more detailed modeling of tidal effects is valuable for this system, it is beyond the scope of this paper. We model the secular evolution of the Kepler-1656 system by numerically solving the octupole-level Hamiltonian for the hierarchical triple following Naoz et al. (2013a), folding in GR effects for the inner and outer orbit (Naoz et al., 2013b) and tidal effects (following Eggleton and Kiseleva-Eggleton, 2001). We account for the star's stellar evolution using *SSE* evolution code (Hurley et al., 2000), which is important particularly in accounting for the stellar spin's evolution (this combined code was tested and apply to various range of astrophysical systems, see for example Naoz et al., 2016; Stephan et al., 2016, 2018). To account for the mentioned uncertainty in the tidal timescale as well as the large uncertainty of the system's age, we focus on timestamps within ±20% of the system age, $t_{\rm sys} = 6.31$ Gyr. This window is conservative compared to the system errorbars ($\Delta t \approx 3$Gyr, see Johnson et al., 2017a).

We ran two sets simulations under the aforementioned initial conditions: one where we sample $\mathrm{a_b} = 0.1 - 10$ au to allow for the possibility of migration, and one in which begins in situ at $a_{\rm b}$=0.197 au (see Table 2.3). The first set of 800 simulations (hereafter Set 1) samples $a_{\rm b}$ from a log-uniform distribution between 0.1 au and a maximum value set by the following stability criterion:

$$a_{\rm b} \leq \epsilon_{\rm max} a_{\rm c} \frac{1-e_{\rm c}^2}{e_{\rm c}} \ , \tag{2.6}$$

where we fix $\epsilon_{\rm max} = 0.1$ to be the maximum relative amplitudes of the octupole and quadrupole terms in the Hamiltonian. This criterion verifies that the system is hierarchical and that the secular approximation is valid (Naoz, 2016)[6].

We ran two sets simulations under the aforementioned initial conditions: one where we sample $a_{\rm b} = 0.1 - 10$ au to allow for the possibility of migration, and one in which begins in situ at $a_{\rm b}$=0.197 au (see Table 2.3). The first set of 800 simulations (hereafter Set 1) samples $a_{\rm b}$ from a log-uniform distribution between 0.1 au and a maximum value set by eq. 2.6.

Following the results of Set 1, we find that *only the systems that began in situ* (i.e., with $a_{\rm b} = 0.197$ au initially) ended up in the regime that matches Kepler-1656b's observational constrains (see Section 2.4.1 for more details). This motivated us to run a second set of simulations that only considers systems beginning in situ with $a_{\rm b} = 0.197$ au, hereafter Set 2. For Set 2, we draw from the same distributions as Set 1 for all parameters except $a_{\rm b}$ and apply the aforementioned stability criterion by rejecting initial conditions for which $\epsilon > 0.1$ (see eq. 2.6). We run 1000 of these systems, as summarized in Table 2.3.

Table 2.3: Simulation Overview

Parameter	**Set 1**	**Set 2**	**Distribution**
$N_{\rm sim}$	800	1000	–
$a_{\rm b}$	0.1-10 au	0.197 au	log-uniform (1), fixed (2)
$a_{\rm c}$	1.4-5 au	”[a]	uniform
$M_{\rm b}$	48.6 $M_\oplus$	”	fixed

[6]To verify our stability condition, we calculated the timescale for the Kepler-1656 system to become unstable independently using stability metric from Zhang et al., *in prep.* We find that systems with $\epsilon > 0.1$ become unstable in timescales much smaller than the EKL timescale (eq. 17) by this metric, confirming that our stability criterion is reasonable.

Table 2.3 Continued

Parameter	Set 1	Set 2	Distribution
$M_{\rm c}$	0.1 $M_{\rm b}$-2$M_{\rm jup}$	"	log-uniform
$e_{\rm b}$	0.01^b	"	fixed
$e_{\rm c}$	0-1	"	uniform
$i_{\rm bc}$	0-180°	"	uniform in $\cos(i_{\rm bc})$

Table Comments:

^{a} denotes same values as in Set 1.

^{b} not fixed at $e_{\rm b} = 0$ exactly to avoid errors in our numerical calculations.

2.4 Dynamical analysis

2.4.1 Kepler-1656b Has Not Migrated

Results from Set 1 of our simulations (i.e., models in which Kepler-1656b was sampled at a range of initial semi-major axes) are plotted as a function of the planet's eccentricity and semi-major axis in the top panel of Figure 2.7. We overplot the region of eccentricity versus semi-major axes (hereafter "e-a") space that is consistent with our observations as a black rectangle. Here we define consistent as having a $a_{\rm b} = 0.16 - 0.24$ au, $e_{\rm b} = 0.67 - 1$ (note however that in some cases in the rectangle where $e_{\rm b} \approx 1$, the planet plunges into the host star and is thus inconsistent with our observations, as discussed further in the following section). For these simulations, only 2.0% ended up in the observed configuration at some point within 20% (1.2 Gyr) of the system's age. As seen in Figure 2.7, all of these models were in situ initially, providing early evidence that Kepler-1656b began its dynamical lifetime in situ.

The upper envelope of e-a space overplotted in Figure 2.7 shows the constant angular momentum track that planets undergoing high eccentricity migration follow. For these planets, scattering, EKL, or secular chaos (Rasio and Ford, 1996; Chatterjee et al., 2008; Naoz et al., 2011; Lithwick and Wu, 2014, see) can cause tidal effects to dominate over perturbations from the outer companions. These tidal effects cause the inner planet to circularize and migrate inward. The initial angular momentum for these eccentric systems, with initial eccentricity $e_{\rm b,i}$ and semi-major axis $a_{\rm b,i}$, can be written as:

$$J^2 \propto a_{\rm b,i}(1-e_{\rm b,i}^2) \approx 2a_{\rm b,i}(1-e_{\rm b,i}) \ , \tag{2.7}$$

since $e_{\rm b,i} \to 1$ for an initially eccentric planet. Similarly, the final angular momentum can be written as:

$$J^2 \propto a_{\rm b,f}(1-e_{\rm b,f}^2) \approx a_{\rm b,f} \ , \tag{2.8}$$

since $e_{\rm b,i} \to 0$ as the planet circularizes. The planet's orbit can continue to shrink without being tidally disrupted as long as its closest approach $r_{\rm peri} = a_{\rm b}(1-e_{\rm b})$ is larger than its Roche limit, so the lowest possible periastron the planet can shrink to is $a_{\rm b,f}(1-e_{\rm b,f}) \approx a_{\rm b,f} \approx R_{\rm Roche}$. Angular momentum is conserved in this process; thus, we can combine eq. 2.7 and 2.8 to get:

$$a_{\rm b,i}(1-e_{\rm b,i}^2) \approx a_{\rm b,f} = R_{\rm Roche}. \tag{2.9}$$

We show this envelope as solid black line in Figure 2.7 for $a_{\rm final} = 0.01\,$au. The shaded region around this line shows the same upper envelope for a range of $a_{\rm b,f}$ within a factor of 2. As we can see, Kepler-1656b does not migrate from its initial semi-major axis in any of our simulations prior to reaching this upper envelope.

We also consider a second e-a upper envelope for Kepler-1656b set by chaotic tides between the planet's orbital energy and fundamental modes (i.e., "f-modes", e.g., Wu, 2018; Vick et al., 2019). For close, eccentric orbits, a planet's orbital energy may be converted into internal fluid energy (a.k.a. "f-mode oscillations") via tidal stretching and compression at periastron passage, thereby causing the orbit to circularize over time. Wu (2018) describes the quantitative criterion for this chaotic tidal evolution to occur:

$$\begin{aligned} a_{\mathrm{b}}(1-e_{\mathrm{b}}) \quad \leq \quad & 0.02\,\mathrm{au}\left(\frac{R_{\mathrm{b}}}{1.1R_{\mathrm{jup}}}\right)^{2/21}\left(\frac{a_{\mathrm{b,i}}}{1\,au}\right)^{5/42} \\ & \times\left(\frac{Q'_{nl}}{0.5}\right)^{2/21}\left(\frac{P}{1.04\times10^{4}s}\right)^{11/21}, \end{aligned} \tag{2.10}$$

where $a_{\mathrm{b,i}}$ is Kepler-1656b's initial separation, (i.e., 0.197 for the in situ case), Q'_{nl} is a dimensionless form of the tidal integral (set to 0.5 for f-modes), and P is the period of the prograde mode ($P = 3.4 \times 10^4$ sec for a spin period 10 days).

We overplot this criterion as a dashed upper envelope to the e-a in Figure 2.7. We see here that our simulations do not reside in the regime where chaotic tides play a large role in the evolution of Kepler-1656b (i.e., the e-a region between the dashed and solid lines). Furthermore, we see from this figure that our simulations do not migrate prior to reaching the upper envelope set constant angular momentum. Thus, only systems that began in situ can reach the high eccentricity we observe. Motivated by these results, Set 2 requires Kepler-1656b to be in situ at the start of our simulations.

Set 2 produces results that are significantly more consistent with our observations, with 3.3% ending up in the observed configuration within 20% of the system's age (compared to 2.0% in Set 1). The absence of pre-envelope migration in both sets of simulations and the increased likelihood of matching our observations in for Set 2 both suggest that the planet began its dynamical lifetime in situ. We stress, however, that this does not require the planet to have originally formed in situ– there are a number of different paths for Kepler-1656b to reach this configuration prior to its in situ EKL onset, which we discuss in Section 2.5.

2.4.2 Fate of Sub-Saturns with Distant Giant Companions

Figure 2.7 shows the the evolution of all models in Set 1 (top row) and Set 2 (bottom row) of our simulations in e-a space, color-coded according to the final outcome for that system. In both rows, the left panel shows only the initial and final timestamps of each simulation, and the right panel shows all timestamps within 20% of the system's age. We overplot the constant angular momentum envelope in each panel, as well as the aforementioned diffusive tides criterion from Wu (2018) (see eq. 2.9 and 2.10, respectively).

In general, there are four possible outcomes for our simulations:

- **Inner planet crosses the star's Roche limit:** If the eccentricity of the inner planet is extremely high, as can be the case for the EKL mechanism (e.g., Li et al., 2014b), the orbit can fall within the star's Roche lobe at closest approach, causing the planet to plunge into its host star. We find that 0.8% (8.0%) of systems are plunge into their host in Set 2 (1) of our simulations. We mark these systems in red in Figure 2.7, and as is depicted, they are confined to the upper part of the angular momentum envelope.

  Crossing the host star's Roche limit is a typical outcome of the EKL mechanism (Naoz et al., 2012), and can have ramifications on the star's spin (e.g., Soker and Harpaz, 2000; Metzger et al., 2012; Qureshi et al., 2018a; Stephan et al., 2020b) and Lithium abundance (Aguilera-Gómez et al., 2016; Bharat Kumar et al., 2018). We see evidence for this in pre-engulfment hot Jupiters and ultra-short period planets (USPs) as well as post-engulfment white dwarfs (see Jura et al., 2009; Zuckerman et al., 2010; Maciejewski et al., 2016; Stephan et al., 2018).

  The orbits of the systems highlighted in Figure 2.7 are short-lived. They also tend to have closer and more eccentric companions ($e \approx 0.5$) and slightly misaligned mutual inclinations ($i_{\rm bc} \sim 40^o$), which are generally less stable than their retrograde counterparts (e.g., Innanen et al., 1997). This is evident in Figure 2.9 which shows the distribution

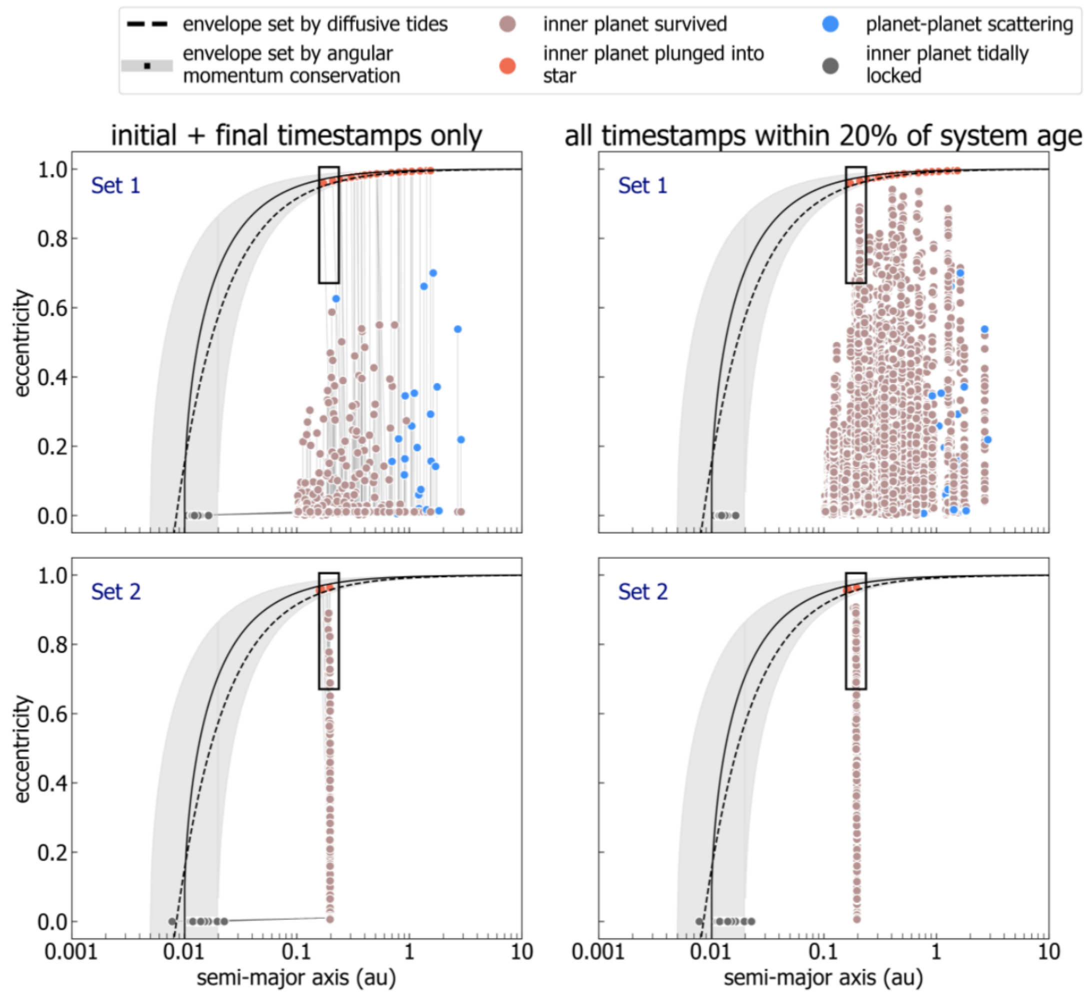

Figure 2.7: Eccentricity and semi-major axis of simulated planet b's with planet c parameters drawn according to the distributions described in Table 2.3. The top row shows selected timestamps from simulations Set 1, with just initial and final timestamps at $t = 0$ and $t = 7.5\,\mathrm{Gyr}$ connected by grey lines on the left, and all timestamps within 20% of the system's age (i.e. $t = 5.4 - 7.5\,\mathrm{Gyr}$) on the right. The bottom row shows the same for simulations Set 2. In all panels, the points are colored according to the simulation outcome. The upper envelope set by constant angular momentum (eq. 2.9) is plotted in black for a range of $a_{\mathrm{b,f}}$ within a factor of 2 of 0.01 au (shaded region), as well the upper envelope set by chaotic tides (see Wu, 2018, Equation 20). The region of the diagram that lies within 20% of the observed semi-major axis and eccentricity values reported in Brady et al. (2018) is outlined in black.

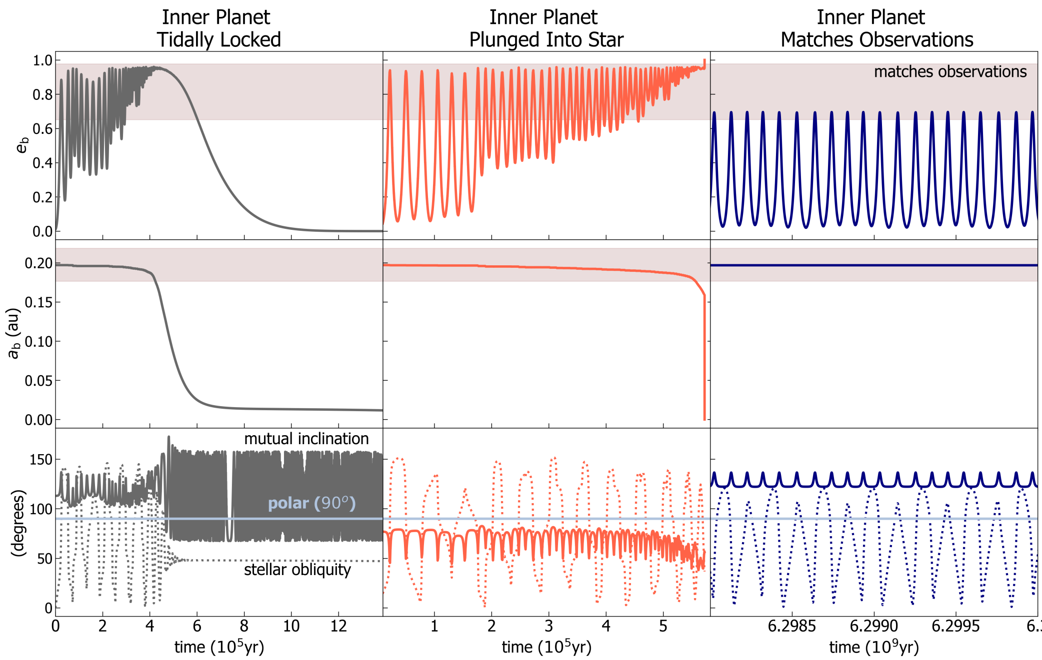

Figure 2.8: Evolution of three simulations with qualitatively different outcomes. *Left:* For a system with $M_\mathrm{c} = 1.52\,M_\mathrm{jup}$, $a_\mathrm{c} = 2.52\,\mathrm{au}$, $\mathrm{e_c} = 0.64$, $i_\mathrm{bc} = 112.96°$ initially, the inner planet becomes tidally locked after 1.2×10^6 years. *Middle:* For a system with M_c=$1.28\,M_\mathrm{jup}$, $a_\mathrm{c} = 1.88\,\mathrm{au}$, $\mathrm{e_c} = 0.51$, $i_\mathrm{bc} = 76.93°$ initially, the inner planet crosses the host star's Roche limit after 5×10^6 years. *Right:* For $M_\mathrm{c} = 1.09\,M_\mathrm{jup}$, $\mathrm{a_c} = 2.09\,\mathrm{au}$, $\mathrm{e_c} = 0.014$, $i_\mathrm{bc} = 121.65°$ initially, we reproduce our observations for a sustained, eccentric orbit for the inner planet past 6.3×10^9 years.

of companion parameters for the subset of simulations that plunge into their host star, as well as subsets for the other three simulations outcomes (as also highlighted in the density maps in Appendix B, Figure 13). Additionally, these systems are consistent with more massive outer companions (Figure 2.9, middle panel of row 2) which will tend to simultaneously perturb the inner planet and resist torque induced by it. An example of the evolution for such a system is shown in Figure 2.8. These outcomes suggest that sub-Saturns engulfed by their star's Roche lobe are more likely to arise from slightly misaligned ($i_{\rm bc} \sim 40^o$) 2-planet systems with eccentric, massive ($M \geq 1 M_{\rm jup}$) outer companions.

- **Inner planet is tidally locked:** When the eccentricity is excited to high values but the planet remains outside the star's Roche limit, tidal effects start to dominate and the planet's semi-major axis shrinks and circularizes, settling in a tidally locked configuration for the inner orbit. For our simulations, we found that 10.3% (15.3%) of systems in Set 2 (1) were tidally locked within the 7.5 Gyr of our simulations. They are marked in grey in Figure 2.7, and are connected to their initial e-a configurations in the left panel. We also show the evolution for one of these tidally locked systems in the left panel of Figure 2.8.

  The process of tidal locking has been discussed extensively in the literature (e.g., Fabrycky and Tremaine, 2007; Wu et al., 2007; Naoz et al., 2012; Petrovich, 2015). Along with stellar binary companions, planetary companions can also torque the inner planet and induce tidal migration and locking (Naoz et al., 2011; Teyssandier et al., 2013; Petrovich and Tremaine, 2016). It has been shown that mutual inclination between nominal Kozai angles of $\sim 40^\circ$ and $\sim 140^\circ$ are required to reach the high eccentricity needed for tides to shrink and circularized the orbit. We see here (top row, left panel) that initially large inclinations, along with moderate outer orbit eccentricity, are indeed needed to produce tidally locked orbits (see also Appendix B, Figure 13). These results imply that warm, tidally locked sub-Saturns in 2-planet systems are born out of configurations in which the outer companion is eccentric and mutually inclined with respect to the inner orbit. We classify outer companions that cause their inner planet to tidally lock as "strong companions", since they have drastic effects on the system's orbital architecture (see Section 2.5 for a detailed discussion).

- **Planet-planet scattering:** For configurations in which the inner and outer orbits are tightly packed, the system can become dynamically unstable or only marginally hierarchical (e.g., Bhaskar et al., 2021). For our simulations, this instability can arise from one of two possible outcomes: (1) the orbital configuration becomes non-hierarchical (i.e., $a_{\rm b,apo}/a_{\rm c,peri} \leq (M_{\rm b}/M_{\rm c})^{1/3}$), in which case the secular approximation breaks down and the inner planet is gravitationally captured by the outer planet (Naoz and Silk, 2014), or (2) the planets orbit within a factor of 5 of their mutual Hill radius, thus becoming dynamically unstable over timescales of $\sim 10^9$ years (Chatterjee et al., 2008). These scenarios cause the hierarchical approximation in our simulations to break down, thus we do not model the dynamical evolution beyond this point. Instead, we classify both of these events as planet-planet scattering outcomes, whereby the gravitational interactions between the two planets lead to dynamical instability and possible collisions or ejections (e.g., Rasio and Ford, 1996; Chatterjee et al., 2008). For our simulations, we find that 0.1% (6.0%) of systems in Set 2 (1) produce this outcome. As shown in Figure 2.7, simulations in which Kepler-1656b starts further out (i.e., simulations in Set 1) are more likely to scatter because the inner planet is closer, on average, to the companion from the start[7].

  Planet-planet scattering has been suggested to explain both individual planetary systems and population-level trends. For example, the misaligned orbits of KOI-13.01, HD 147506b, and HAT-P-7b, as well as multiple eccentric orbits in the upsilon-Andromedae system, are thought to be the outcome of planet-planet scattering events (Barnes et al.,

[7]We do not include these systems in Figure 2.9 because only 1 simulation from Set 2 has this outcome. This system had an outer companion with $M_{\rm c} = 0.04 M_{\rm jup}$, $a_{\rm c} = 2\,{\rm au}$, $e_{\rm c} = 0.42$, $i_{\rm bc} = 105^\circ$ initially.

2011; Winn et al., 2007; Ford et al., 2005; Narita et al., 2009). Beyond this, demographic trends such as the observed distribution of planet eccentricities, correlations between host star metallicity and orbital period, and orbital properties of hot Jupiters can be explained at least in part by planet-planet scattering (Ford and Rasio, 2008; Petigura et al., 2018c; Dawson and Johnson, 2018). We also speculate that perhaps a planet-planet scattering event resulted in the *in situ* configuration we propose for the onset of EKL effects in the Kepler-1656 system (see Section 2.5 for details). This speculation is consistent with the recent study of the HR 5183b system (Mustill et al., 2021).

- **Inner planet survives and matches our observations:** In some cases, the eccentricity of the inner planet is excited via the EKL mechanism to values that are just below the upper envelope of the e-a diagram. The result of this evolution is a significantly eccentric inner orbit ($e_{\mathrm{b}} \geq 0.67$, within 20% of the observed value for Kepler-1656b)[8], but still low enough to be stable over $\sim 10^9$-year timescales without being subject to the aforementioned tidal effects. Orbits like this are consistent with our observations of Kepler-1656b, and are the outcome for 3.3% (2.0%) of our simulations in Set 2 (1) (see Figure 2.8 for an example).

  Stable, eccentric orbits like Kepler-1656b's are a predicted outcome for systems subject to EKL-dominated evolution (e.g., Naoz et al., 2011, 2012; Li et al., 2014b; Petrovich, 2015; Petrovich and Tremaine, 2016). In these cases, torque exchanged by the inner and outer orbits causes the inner planet's eccentricity to oscillate with brief excursions to high values (see Figure 2.8 for detailed evolution). For simulations that remain below the upper e-a envelope, these oscillations are sustained over timescales up to and past $t = 6.3\,\mathrm{Gyr}$, thus reproducing our observations at the system's reported age.

  It is also important to note that even for systems that match our observations, Kepler-

[8] Note that this 20% is larger than the reported errorbars in Table 2.2. This relaxed constraint is to account for both the finite time sampling in our simulations and the choatic nature of the EKL mechanism (e.g., Naoz et al., 2016)

1656b spends a large fraction its lifetime in much less eccentric orbits due to the nature of its eccentricity oscillations, as can be illustrated in the upper right panel of Figure 2.8. As such, it is possible that several other known, less eccentric exoplanets are undergoing similar high-amplitude eccentricity oscillations induced by outer companions, but are observed during a period of lower eccentricity. We classify companions that induce these in situ eccentricity oscillations as "gentle companions", because they are relatively well-behaved compared to their more disruptive "strong companion" counterparts, as discussed further in Section 2.5. Kepler-1656cis an example of one such companion, and drives Kepler-1656b to oscillate between low and high values throughout it's lifetime.

- **Inner planet survives, but does not match our observations:** The majority of our simulations (68.7% and 85.5% for Set 1 and 2, respectively) do not fall in any of the categories described above, and instead simply retain less eccentric ($e_b < 0.67$) orbits for the duration of our simulations (see Figure 2.7). Still, these systems are valuable to our analysis because their initial conditions can be ruled out as possible origins scenarios to Kepler-1656b, since they are unable to reproduce our observations. Additionally, their low-to-moderate eccentricities and large distance from the upper envelope place them in a distinct region of *e*-*a* space compared to their more eccentric counterparts described above. Thus, planets in this region of parameter space may be signatures of distant companions that are exciting their eccentricities to values lower than that of Kepler-1656b. We discuss this possibility in more detail in Section 2.5.

The fact that a subset of our EKL simulations are consistent with our observations of Kepler-1656b indicates that the EKL mechanism is a plausible origin scenario for Kepler-1656b's eccentric orbit. Moreover, the range of possible outcomes in our simulations suggests a diverse set of dynamical histories for sub-Saturns with distant giant companions. We discuss these histories and their potential signatures on the *e*-*a* distribution of exoplanets in Section 2.5.

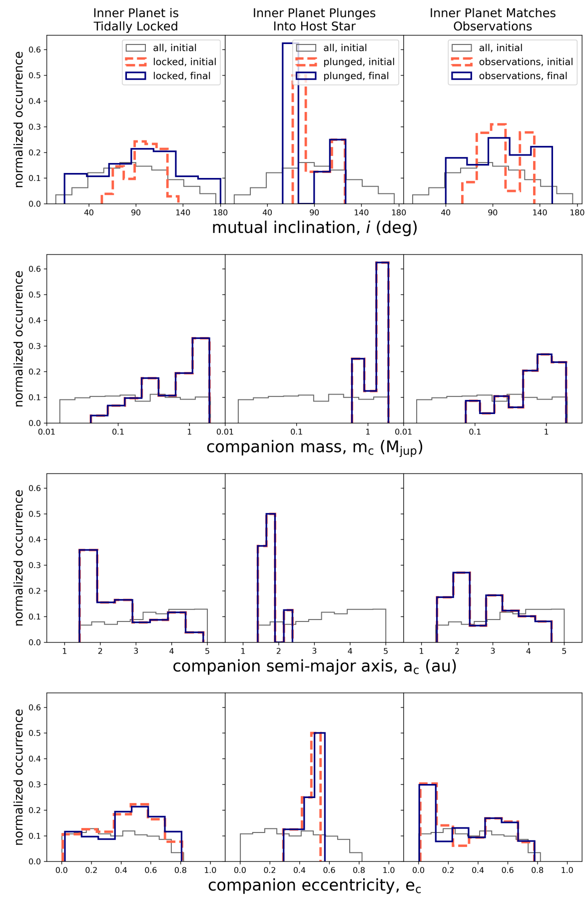

Figure 2.9: Normalized distributions are plotted for various orbital outer planet parameters in Set 2 our simulations. The models are divided into the following subsets: models in which the inner planet becomes tidally locked (left), models in which the inner planet plunges into its host star (middle), and models in which the inner planet is consistent with our observations (right). For all panels, the initial distribution of our entire set of models is shown in grey. Initial and final distributions of different subsets are over-plotted in red and blue, respectively.

2.4.3 Properties and Influence of the Outer Planet

Our simulations reveal that a subset of our sampled companions can reproduce the observed orbital configuration. Thus, we can search for commonalities in this subset to constrain both the evolution and observable properties of Kepler-1656c. Since Set 2 is more consistent with our observations, we focus on just this set for the remainder of our analysis of Kepler-1656 and return to our discussion of Set 1 in Section 2.5 to make generalizations about the sub-Saturn population.

The simulations that reproduce Kepler-1656b's orbit evolve from a unique set of orbital configurations (see Figure 2.9, right column), allowing us to uncover possible origin scenarios for the system. In particular, models that match our observations have identical initial and final distributions for the companion semi-major axis (a_c). This suggests that a_c does not change over time, and much like its inner companion, Kepler-1656c likely began its dynamical lifetime in situ.

Figure 2.9 also shows that although Kepler-1656b and Kepler-1656c can end up with any mutual inclination within $i_{bc} = 40 - 150°$, they are born out of a slightly more narrow distribution of $i_{bc} = 60 - 130°$. We also see that Kepler-1656c can be anywhere from $0 - 0.8$ initially. These initial conditions are consistent with predictions conditions required for EKL onset– that is, an eccentric outer orbit and significant mutual inclination between the inner and outer orbits (Naoz, 2016). In fact, our simulations show that Kepler-1656b's dynamical behavior is driven by quadrupole-level EKL oscillations (see Appendix A, eq. 17). These oscillations perturb the inner orbit just enough to excite it to its high eccentricity, but not enough to induce high eccentricity migration. We note that these are not test-particle quadrupole-level as the inner orbit's angular momentum is not conserved Naoz et al. (2011, 2013a). Further discussion on this dynamical behavior can be found in Appendix A.

Beyond this, commonalities in the simulations that reproduce our observations inform us about the companion's current orbital configuration. This is evident in Figure 2.10, which shows 2D distributions of the companion properties in our simulations, both initially (in tan) and today (i.e., when t is within 20% of t_{sys}, in blue). We also overlay Kepler-1656c for comparison based on orbital properties derived in Section 2.2. We see from the leftmost panel in this figure that Kepler-1656b and Kepler-1656c must orbit with a nearly polar mutual inclination ($i_{bc} \approx 90°$) in order to be consistent with our observed a_c =3.05 au and e_c =0.53. We also see that for all panels, Kepler-1656c's orbital properties are consistent with models that reproduce Kepler-1656b's high eccentricity. Thus, EKL oscillations induced by gravitational perturbations from Kepler-1656c is a promising origin scenario for Kepler-1656b.

2.4.4 Predictions for Inner Planet Obliquity

One of the parameters of planet b that is traced in our simulations is the orbital obliquity, λ_{b}, which defines the angle between the host star's spin axis and the planet's orbital angular momentum vector. In our analysis, we found that the subset of simulations that are consistent with our observations produce a wide range of values for Kepler-1656b's current spin-orbit angle, or obliquity λ_{b}, with values ranging from prograde to retrograde and a slight preference for polar orbits ($\lambda_{\mathrm{b}} \approx 90°$, see Figure 2.11). This, along with our predicted near-polar mutual inclination, is consistent with the previously suggested correlation between stellar obliquities and planetary mutual inclinations in systems with distant giant companions (Wang et al., 2021). Furthermore, a misaligned orbit for Kepler-1656b suggests that it is part of a growing sample of exoplanet systems hosting eccentric, misaligned, sub-Jovian planets with distant giant companions (see for example Yee et al., 2018; Dalal et al., 2019; Correia et al., 2020). While it has been suggested that a slowly decaying outer protoplanetary disk may produce sub-Neptunes in these architectures, whether or not this mechanism is relevant for sub-Saturns like Kepler-1656b remains to be seen (e.g., Petrovich et al., 2020). Nevertheless, the misaligned and often polar orbits favored in our models show that gravitational interactions from a distant, outer companion may be the instigator for these configurations beyond the case of Kepler-1656b.

Figure 2.11 also demonstrates that engulfment and tidal locking produce distinctive obliquity distributions for the inner planet. For instance, systems for which the inner planet plunges into the host star are overwhelmingly on prograde orbits right before they plunge in. This

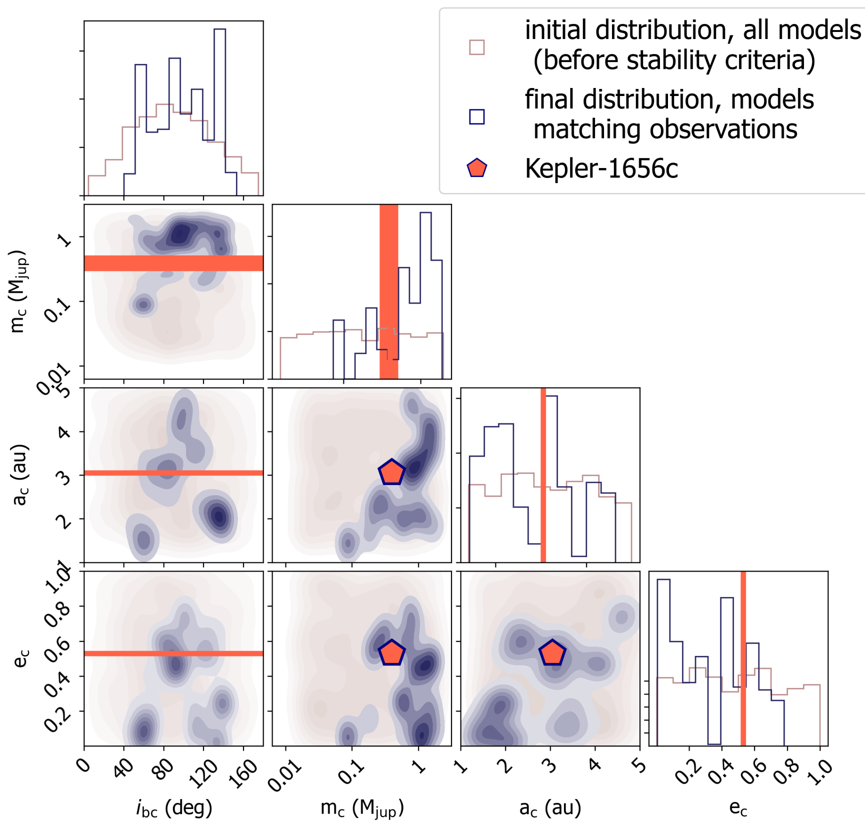

Figure 2.10: Smoothed 2D histograms of final conditions for all models (tan) and models that are consistent with our observations (blue) from Set 2 of our simulations are shown. We see that models consistent with our observations place constraints on the mutual inclination between Kepler-1656b and Kepler-1656c, which we refer to as $i_{\rm bc}$. We also overlay Kepler-1656c based on properties found in Section 2.2. The diagonal panels show 1D distributions of parameters, that have been normalized to sum to 1.

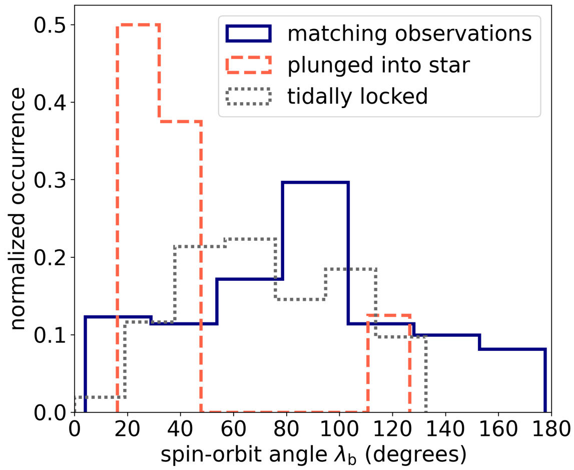

Figure 2.11: Final spin-orbit angle, or obliquity, of Kepler-1656b is shown for subsets of our simulations Set 2 for which inner planet matches our observations (blue line), plunges into its host star (grey dotted line), or becomes tidally locked (red dashed line).

trend should produce an observational signature on the population of post-engulfment host stars– specifically, plunging planets on prograde orbits will slow their host star's rotation (e.g., Qureshi et al., 2018b; Stephan et al., 2020b). On the other hand, for systems in which the inner planet becomes tidally locked, the final distribution of obliquities matches that of tidally locked hot Jupiters (e.g., Naoz et al., 2012). These predicted obliquity trends will be relevant to the next-generation spectrometers that can resolve obliquities of Jovian and sub-Jovian planets in multi-planet systems.

2.5 Discussion

Our analysis suggests an in situ beginning (here we use "beginning" to refer specifically to onset of EKL effects) with subsequent EKL-driven excitations is a plausible pathway for Kepler-1656b's orbital evolution to the close-in, highly eccentric orbit we see today. This proposed in situ configuration may point to in situ origins for Kepler-1656b, consistent with existing theories that dense planets with high core mass fractions can form in situ out of

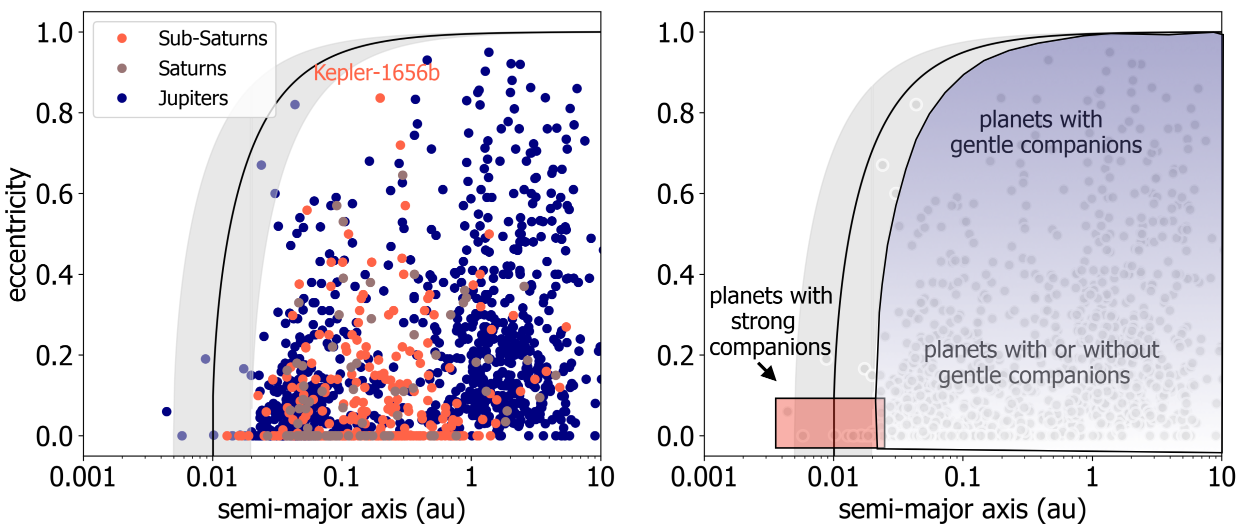

Figure 2.12: *Left:* The population of known giant planets is shown in e-a space. We plot sub-populations of Sub-Saturns ($M = 15 - 60M_{\oplus}$) in blue, Saturns ($M = 60 - 100M_{\oplus}$) in red, and Jupiters ($M > 100M_{\oplus}$) in tan. Kepler-1656b and Kepler-1656c is over-plotted and labeled accordingly. Data were obtained from the NASA Exoplanet Archive (Akeson et al., 2013). *Right:* The same population is plotted in grey. The shaded gradient represents an increasing need for gentle companions to explain orbital configurations towards the upper envelope, and the red region represents tidally locked configurations that may indicate the presence of strong companions. In both panels, the upper envelope set by angular momentum conservation is over-plotted.

particularly dusty disk environments (e.g., Lee et al., 2014; Lee and Chiang, 2015). It is also in line with more general theories of in situ giant formation (see for example Batygin et al., 2016). However, we cannot rule out the possibility that events such as flybys, scattering, or a merger occurred prior to EKL onset, in which case Kepler-1656b may not have formed in its present-day configuration. In fact, it is possible that a flyby or planet-planet scattering event pushed Kepler-1656b's orbit inwards from larger distances and/or into the required non-negligible mutual inclination between the two planets to trigger the onset of EKL (see Mustill et al., 2021, for an example of such evolution)[9]. In this case, it remains possible that a merger prior to EKL onset explains Kepler-1656b's below-average envelope mass fraction ($f_{\rm env} \approx 2.8\%$), as suggested by Millholland et al. (2020). Still, we stress that Kepler-1656b's eccentricity is not easily explained by flybys, scattering, or a merger alone– thus, perturbations from Kepler-1656c is our prevailing explanation for the inner orbital eccentricity [10]

Previously, Brady et al. (2018) suggested that Kepler-1656b might be undergoing high eccentricity migration given its proximity to the upper envelope of the e-a distribution. Here we find evidence for the contrary: our simulations indicate that Kepler-1656b's outer companion induces high-amplitude eccentricity oscillations gently, i.e. without triggering high eccentricity migration. From these results, we classify two types of outer planets: those that induce tidal migration of their inner companions ("strong companions"), and those that do not, while still exciting eccentricity ("gentle companions"). We note that for due to the nature of their eccentricity oscillations, systems with gentle companions spend more time in less eccentric orbits, and only a small fraction of time in orbits as eccentric as Kepler-1656b's (see Section 2.3.1 and Figure 2.8). On the other hand, strong companions cause their inner planet to tidally circularize. Thus, our simulations show that sub-Saturns with outer com-

[9]The analysis of the possible flybys and/or planet-planet scattering prior to the EKL evolution is beyond the scope of this study.

[10]We note that stellar flybys are an unlikely scenario for our system as they needed an unreasonable number of close interaction, see Batygin et al. (2020).

panions can populate wide regions of e-a parameter space, depending on effects from the companion.

Comparing Kepler-1656b to the general population of sub-Saturns ($M = 15 - 60M_\oplus$), Saturns ($M = 60 - 100M_\oplus$) and Jupiters ($M > 100M_\oplus$) suggests similar dynamical histories for these planets, as highlighted in Figure 2.12. While a more rigorous statistical analysis is required, already from comparing our simulation results to observations (see Figure 2.12), the similarities in the e-a distribution of the sub-Saturns, Saturns and Jupiters suggests that Saturn- and Jupiter-sized planets may be subject to similar perturbations from outer companions. Thus, a number of giant planets in the current exoplanet census may have orbital signatures of gentle and strong companions. In fact, a number of close-in, eccentric planets ranging from sub-Saturns to Jupiters have known companions that may be gently perturbing them into their eccentric configurations (e.g., CoRoT-20b, Kepler-419b, HD 156279b) or strongly driving them into a tidally locked orbits (see K2-43b, HAT-P-13b, Kepler-101b, HD 187123b). Beyond this, a larger subset of planets with eccentric or close, circular orbits do not have any confirmed companions (see for example GJ 2056 b, HD 20868b, HD 43197b, KOI-1257b, WASP-19b, TOI-1296b, TOI-1298b)[11], and would thus make a great laboratory to test our hypothesis of gentle and strong companions as the instigator for these configurations[12]. Long-term monitoring of these planet systems' RV and transit data, combined with dynamical modeling of potential 2-planet systems, can help us understand the origins of these systems and test our predicted positive correlation between eccentric or tidally sub-Saturns and outer companions.

[11] Note that Moutou et al. (2021) implied that TOI-1296b and TOI-1298b's orbits may connected to their high host star metallicity, implying that high metallicity produces more giant planet multiples (e.g., Dawson and Murray-Clay, 2013). Still, based on the non-negligible fraction of locked planets in our simulations (10.3%), we suggest that TOI-1296b and TOI-1298b, may have distant planetary companions.

[12] see the following references: CoRoT-20b–Deleuil et al. (2012), Kepler-419b–Dawson et al. (2014), HD 156279b–Díaz et al. (2012), K2-43b–Crossfield et al. (2016), HAT-P-13b–Bakos et al. (2009), Kepler-101b–Rowe et al. (2014), HD 187123b–Butler et al. (1998), GJ 2056b–Feng et al. (2020), HD 20868b– Moutou et al. (2009), HD 43179b–Naef et al. (2010), KOI-1257b–see Santerne et al. (2014), WASP-19b–Hebb et al. (2010), TOI-1296b & TOI-1298b–Moutou et al. (2021)

Additionally, the three populations plotted in Figure 2.12 do not appear to show distinctive trends, although more data are needed to confirm this. Still, the apparent similarities in these populations of Sub-Saturns, Saturns, and Jupiters in e-a space suggest that these planets are born from similar dynamical environments. This suggests that while these sub-populations of giant planets are often treated as distinct (e.g., Petigura et al., 2017c), they may be products of similar dynamical histories. Beyond this, it is thought that distant giant companions are responsible at least in part for the eccentricity distribution of small planets as well (Van Eylen et al., 2019). Future studies using a combination of observations and theoretical calculations to constrain outer companion properties will bring us closer to understanding the eccentricity distribution of giant planets and the exoplanet population as a whole, pulling the curtain back on the origins of these systems.

Our study demonstrates the advantages of combining observations and modeling to more rigidly constrain a planetary companion's origins and orbital properties (as highlighted in Figures 2.9 and 2.10). Here, we use observations to inform the types of outer planets in our simulations. From this, we uncover an in situ origin scenario for Kepler-1656b, and make inferences about the inclination of the inner and outer planets in the system. Thus, our combined observational and theoretical analysis allows us to paint a detailed picture of the Kepler-1656 system's origins. These methods can be applied to similar single- and 2−planet systems to constrain their architectures, histories, and properties of a potential companion.

2.6 Acknowledgements

IA is supported by the Dorothea Radcliffe Dea Fellowship and Dean's Scholarship at UCLA. We are grateful for the California Planet Search team and staff at W.M. Keck observatory for the remote observation efforts that made this study possible. IA would also like to thank Sanaea Rose, Judah Van Zandt, Dakotah Tyler, Jon Zink, Trevor David and Sarah Millholland for helpful insights on the analysis and discussion sections of this paper. IA and S.N. acknowledge partial support from the NSF through grant No. AST-1739160. S.N. thanks Howard and Astrid Preston for their generous support.

Software: *Numpy* (van der Walt et al., 2011), *Radvel* (Fulton et al., 2018), *RVSearch* (Rosenthal et al., 2021), *corner* (Foreman-Mackey, 2016), *astropy* (Astropy Collaboration et al., 2013a, 2018a)

Appendix 2A:

Kepler-1656b does not undergo high-eccentricity migration

Under the physical conditions outlined in Section 2.3.1, Kepler-1656b's eccentricity can behave in several ways. In the quadrupole-level approximation, Kepler-1656b's eccentricity will oscillate between fixed minima and maxima. Alternatively, for conditions in which tides or octupole-level behavior are dominant, high eccentricity migration or more chaotic eccentricity evolution can dominate. Consider the former case in which Kepler-1656b's eccentricity oscillates between a fixed minimum and maximum. In this scenario, Naoz et al. (2013a) showed that the total angular momentum is conserved, along with the eccentricity and angular momentum for the outer orbit. This implies that the initial and final mutual inclinations between the two planets (hereafter $i_{\mathrm{bc,i}}$ and $i_{\mathrm{bc,f}}$) obey

$$\cos(i_{\mathrm{bc,f}}) = \frac{A_f \cos(i_{\mathrm{bc,i}}) + L_1^2 e_1^2}{A_f \sqrt{1 - e_{\mathrm{b}}^2}}, \tag{2.11}$$

where $A_f = 2L_1 L_2 * \sqrt{1 - e_{\mathrm{c}}^2}$, and L_1 and L_2 are the respective conjugate angular momenta of the inner and outer orbit. Thus, if Kepler-1656b undergoes quadrupole-dominated ec-

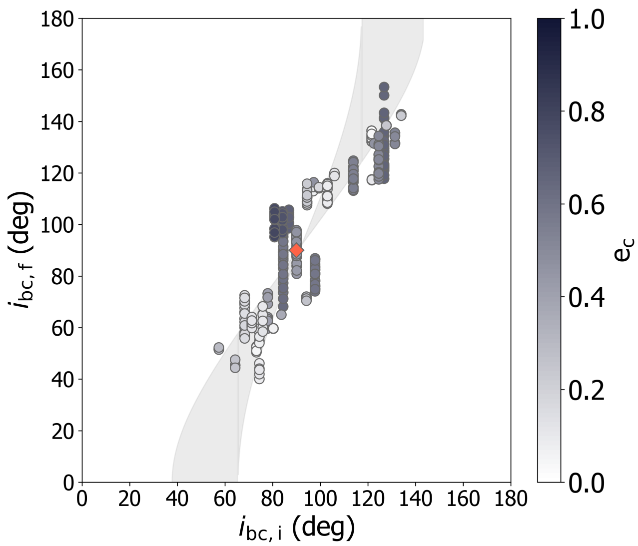

Figure 2.13: Initial and final mutual inclinations for all model timestamps consistent with observations of Kepler-1656b. The shaded region shows the initial and final inclinations predicted by the quadrupole-level approximation for inner planet eccentricities of $0.6 < e_{\rm b} < 0.9$.

centricity oscillations with no significant tides or octupole-level oscillations, it should obey the relationship described by eq. 17. We plot this relationship for a range of eccentricities for Kepler-1656b that are reasonably close to our observed value ($0.6 < e_p < 0.9$), and find that most simulations that match our observations lie in this regime. This indicates quadrupole-dominated behavior for Kepler-1656b, whereby the planet reached achieved its high eccentricity via sustained, high-amplitude eccentricity oscillations without tidal migration of chaotic behavior.

Appendix 2B:

Correlated Parameters for Plunging, Locked Models

We find in our simulations that scenarios in which the inner planet plunges into the host star or becomes tidally locked arise from a unique set of parameters for the outer compan-

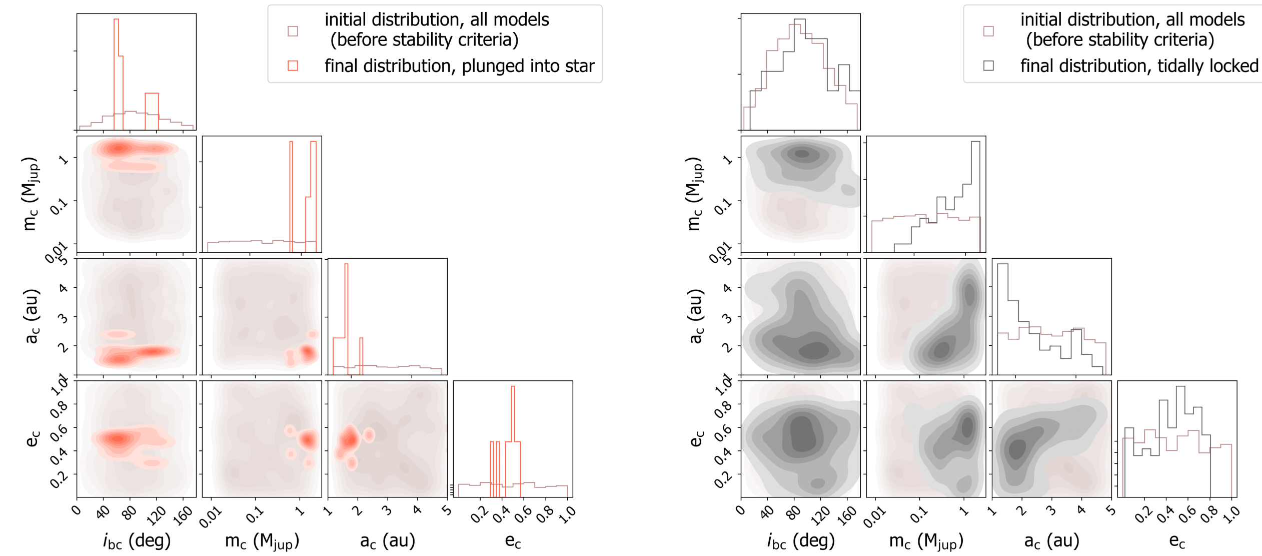

Figure 2.14: Smoothed 2D initial and final companion parameter distributions for models that plunged into their host star (left) and became tidally locked (right). The diagonal panels show 1D distributions of parameters that have been normalized to sum to 1.

ion. This is evident in Figure 13, which shows final distributions of companion parameters e_c, m_c, a_c, and $i_{\rm bc}$ for both cases. We see here that moderately eccentric ($e_c \approx 0.5$), aligned companions are more likely to cause the inner planet to plunge into its host star. On the other hand, companions on polar orbits with a wider range of e_c can cause the inner planet to become tidally locked. Thus, we predict that tidally locked sub-Saturns may be signatures of eccentric, polar companions.

Chapter 3

A Data-driven Spectral Model of Main Sequence Stars in Gaia DR3

A version of this chapter was previously published in *The Astrophysical Journal* (Isabel Angelo, Megan Bedell, & Erik A. Petigura, 2024, ApJ 974, 43).

Abstract

Precise spectroscopic classification of planet hosts is an important tool of exoplanet research at both the population and individual system level. In the era of large-scale surveys, data-driven methods offer an efficient approach to spectroscopic classification that leverages the fact that a subset of stars in any given survey has stellar properties that are known with high fidelity. Here, we use *The Cannon*, a data-driven framework for modeling stellar spectra, to train a generative model of spectra from the Gaia Data Release 3 Radial Velocity Spectrometer. Our model derives stellar labels with precisions of 72 K in T_{eff}, 0.09 dex in $\log g$, 0.06 dex in [Fe/H], 0.05 dex in [α/Fe] and 1.9 km/s in v_{broad} for main-sequence

stars observed by Gaia DR3 by transferring GALAH labels, and is publicly available at https://github.com/isabelangelo/gaiaspec [1]. We validate our model performance on planet hosts with available Gaia RVS spectra at SNR>50 by showing that our model is able to recover stellar parameters at ≥20% improved accuracy over the existing Gaia stellar parameter catalogs, measured by the agreement with high-fidelity labels from the Spectroscopic Observations of Cool Stars (SPOCS) survey. We also provide metrics to test for stellar activity, binarity, and reliability of our model outputs and provide instructions for interpreting these metrics. Finally, we publish updated stellar labels and metrics that flag suspected binaries and active stars for Kepler Input Catalog objects with published Gaia RVS spectra.

3.1 Introduction

Precise spectroscopic classification of stars has implications spanning all sub-fields of astronomy at both the population and individual system level. Surveys like APOGEE (Majewski et al., 2017) and LAMOST (Cui et al., 2012) have facilitated great strides in galactic archaeology and enabled discovery of binaries, transients and compact objects in the Milky Way. Spectroscopic observations of Kepler planet hosts illuminated features like the 'Radius Valley' and 'Hot Neptune Desert' that encode information planet formation (e.g., Fulton and Petigura, 2018; Petigura, 2020). Beyond this, spectroscopic observations of individual systems provide details the system age, composition and stellar properties (see for example Johnson et al., 2017b; Rice and Brewer, 2020).

Historically, spectroscopic classification has taken both data-driven and synthetic modeling approaches. The data-driven approach dates far back to the early 20th century, when Annie Jump Cannon classified thousands of stars by eye based on their optical spectra. In the following decades, spectroscopic classification shifted to rely on matching data with synthetic

[1] gaiaspec (2024)

spectral models. However, these models are not always calibrated to real stars and are based on incomplete physics. For example, many of these models assume local thermodynamic equilibrium (LTE) or employ incomplete atomic line lists, simplified convection treatment, or non-physical boundary conditions. More modern studies match data to real spectra from a library of well-characterized stars (e.g., Yee et al., 2017), and while these library spectra tend to resemble the data better than synthetic models, their performance decreases quickly for spectral resolutions below $R \sim 20,000$ (Behmard et al., 2019).

In recent years, data-driven spectral models have been used to infer stellar characteristics from spectra. These models are trained on large sets of spectroscopic data and pre-determined stellar labels like effective temperature, surface gravity, and elemental abundances using supervised learning algorithms to produce generative models of stellar spectra (see for example *The Cannon*; Ness et al. 2015 and the Data-Driven Payne; Xiang et al. 2019). These models have been employed for numerous science cases, from mapping elemental abundances (e.g., Ness et al., 2016; Buder et al., 2018) to identifying binaries (El-Badry et al., 2018b) and constraining properties of planet hosts (Behmard et al., 2019; Rice and Brewer, 2020). One of the key features of data-driven spectral models is that they leverage the fact that for a given survey, a subset of stars often have derived labels with higher fidelity from a previous survey. In training a model on these pre-determined labels, along with spectra from the survey of interest, precise and accurate spectral information encoded in the model is transferred across data sets. The model can then be used to derive high-fidelity labels for the rest of the stars in the survey of interest.

Nearly one million near-infrared spectra are included in Gaia Data Release 3 (DR3), and this number is expected to increase by ~ 2 orders of magnitude in future data releases (Katz et al., 2023). These spectra cover a wavelength range of 845–872 nm with a spectral resolution of $\lambda/\Delta\lambda \sim$ 11,500, with spectral information in the Ca II IR triplet at wavelengths of 849.8 nm, 854.2 nm, and 866.2 nm, as well as shallower lines from various other elements. While the large sample size lends itself well to large-scale classification and demographic analysis, the spectra have smaller bandwidth and lower signal-to-noise ratio (SNR) than a dedicated ground-based survey can typically achieve (e.g., GALAH, APOGEE, etc., see Majewski et al., 2017; De Silva et al., 2015). A data-driven spectral model trained on higher fidelity labels from one of these surveys can transfer pre-existing spectral information onto this new dataset and assess how far we can extrapolate high-fidelity labels to the growing sample of Gaia RVS spectra. For example, this method was shown to be effective for retrieving precise refractory elemental abundances of solar analogs in Gaia Rampalli et al., 2024, see.

In this paper, we use *The Cannon*'s data-driven framework to train a spectroscopic model of the Gaia RVS spectra using a set of well-characterized main sequence stars from GALAH. We outline the methods that *The Cannon* uses to train a model and use it to retrieve stellar labels in Section 3.2 and describe the data we train our model on in Section 3.3. In Section 3.4, we update our implementation of *The Cannon* and validate our model performance. In Section 3.5, we evaluate our model performance by measuring it's agreement with high-fidelity labels from the Spectroscopic Observations of Cool Stars (SPOCS) survey, and compare to performance of existing Gaia catalogs. We also publish updated stellar labels for stars in the Kepler Input Catalog (KIC) with published RVS spectra from Gaia DR3. We discuss applications for identifying unresolved binaries in the Gaia RVS spectra in Section 4.5 and outline metrics that our model reports for identifying active stars, evolved stars, and cases where our model performance is compromised in Section 3.6. We summarize our results in Section 3.7. Our model is available at `https://github.com/isabelangelo/gaiaspec` to enable characterization of user-specified stars with RVS spectra from DR3 or future data releases.

3.2 The Cannon Overview

We used *The Cannon 2*, an open-source implementation of *The Cannon* introduced in Casey et al. (2016), to train our data-driven model. In this section, we outline the basic methods of this implementation to train a Cannon model and fit a trained model to new spectra to derive stellar labels. A more detailed description of these methods can be found in Ness et al. (2015); Casey et al. (2016).

The Cannon interacts with spectroscopic data in two ways: it uses spectra during its “training step" to develop a spectral model, and then uses this model to fit to individual spectra during its “test step". To summarize, the training step determines the relationship between the observed flux and pre-determined “ground truth” labels at each individual pixel for all the stars in the training set. The test step fits a trained Cannon model to the observed flux of a star with labels to-be-determined by *The Cannon* model itself. These steps are described in detail below.

3.2.1 Training Step: Supervised Learning of “Ground Truth" Spectra

The Cannon is a generative linear regression model that predicts the flux of an object as a function of the labels it's trained on (in our case, $T_{\rm eff}$, $\log g$, [Fe/H], [α/Fe], $v_{\rm broad}$). The $v_{\rm broad}$ parameter here is a line broadening parameter that is determined by the combined effect of stellar rotation and macroturbulence, and is derived for both GALAH and Gaia RVS spectra (Buder et al., 2021; Frémat et al., 2023). A description of *The Cannon* can be found in Ness et al. (2015); Casey et al. (2016).

During the training step, *The Cannon* uses a supervised learning algorithm that takes stellar labels and flux as input. The goal during this step is for the model to learn how the flux at a particular wavelength interval, or “pixel" depends on a star’s labels. The training step learns from *The Cannon* model using the underlying assumptions that (1) stars with the same labels have the same per pixel flux, (2) the flux at a particular pixel is a polynomial function of the training labels, and (3) the flux at a particular pixel is independent of the flux at neighboring pixels. Note that the third assumption means that the pixel fits are performed independently of one another– the model does not assume any “smooth" behavior across pixels. The labels are treated as ground truth, and for each pixel the training step fits a polynomial to the relationship between the flux and the stellar labels.

For a particular pixel value j and stellar n, *The Cannon* assumes that the flux $f_{n,j}$ depends on the star labels l_n according to the following relation:

$$f_{n,j} = v(l_n) \cdot \theta_j + e_{n,j} \tag{3.1}$$

In this case, the “vectorizer" $v(l_n)$ is a function that describes how the flux is assumed to vary with the labels. For our model, we assume that the pixel flux varies linearly to second order with the 5 stellar labels so the vectorizer contains 21 terms (1 linear and 1 quadratic term for each label, with 10 cross-terms and 1 constant). The coefficients of the vectorizer polynomial are contained in θ_j, and e_{nj} are the flux errors, which for this paper are sampled from a Gaussian with zero mean and variance equal to the instrumental uncertainty and intrinsic pixel scatter of the model summed in quadrature, $\sigma_{nj}^2 + s_j^2$. We note that while *The Cannon 2* computes a model-specific s_j^2 during its training step, we derive an empirical s_j^2 that we use throughout this analysis (see Section 3.4.2).

In the training step, the flux (f_{nj}), flux errors from photon noise and instrumental effects (σ_{nj}) and the labels of the spectra (l_n) are known, but the θ_j coefficients are not. The θ_j coefficients govern the model behavior by predicting the flux at a given pixel according to

Equation 4.6. The log-likelihood of the model flux at a particular pixel is:

$$\ln p\,(f_{nj}|\theta_j, l_n, s_j^2) = -\frac{[f_{nj} - v(l_n)\cdot\theta_j]^2}{s_j^2 + \sigma_{nj}^2} - \ln\,(s_j^2 + \sigma_{nj}^2) \tag{3.2}$$

The training step determines the θ_j coefficients and s_j^2 associated with each pixel that maximize this likelihood function (or rather, minimizes the negative likelihood function) for all reference stars $N_{\rm ref}$ in the training set:

$$\theta_j, s_j^2 = \arg\min_{\theta_j, s_j^2} \sum_{n=1}^{N_{\rm stars}} -\ln p\,(y_{nj}|\theta_j, l_n, s_j^2), \tag{3.3}$$

thereby computing the θ_j coefficients that best describe the training flux. Once it has computed these best-fit coefficients, it can predict the flux at each pixel for a given set of stellar labels based on Equation 4.6 using the θ_j that is derived in the training step. This is described further in the following section.

We note that due to the data-driven nature of *The Cannon*, our model performance is limited by what we will refer to hereafter as the training domain. The training domain is a 5-dimensional space with each dimension corresponding to a single training label, populated by data points that each represent a single spectrum that the model was trained on (see Figure 3.3 for collapsed 2D visualizations of our model's training domain). In the event that *The Cannon* identifies best-fitting labels beyond this domain, the best-fit spectrum was achieved via extrapolation beyond the label range of the reference objects used to build the main-sequence model. In this case, the labels are not tied to any ground truth (see Section 3.4.3). Additionally, *The Cannon*'s label retrieval can only be as accurate and precise as the "ground truth" labels it's trained on.

3.2.2 Test Step: Model Fitting to Derive Stellar labels

Once *The Cannon* model is trained, we can use it to recover labels of stars outside the training set in the test step, which fits a trained Cannon model to spectroscopic observations of the star to be characterized. In the training step, the flux and labels of the training set are known, and the optimizer is fitting for θ_j and s_j^2. In the test step, the labels are unknown, but the values for θ_j and s_j^2 have been pre-determined by the test step. Thus, for a given spectrum, the optimal label vector $\hat{l}_n$ can be described as the labels associated with the maximum-likelihood Cannon model flux:

$$\hat{l}_n = \arg\min_{l_n} \sum_{j=1}^{N_{\text{pixels}}} -\ln p\left(f_{n_j} | \theta_j, l_n, s_j^2\right) \tag{3.4}$$

$$= \arg\min_{l_n} \sum_{j=1}^{N_{\text{pixels}}} \frac{[f_{nj} - v(l_n) \cdot \theta_j]^2}{s_j^2 + \sigma_{nj}^2} \tag{3.5}$$

where $-\ln p\left(f_{n_j} | \theta_j, l_n, s_j^2\right)$ (hereafter χ^2) represents the goodness-of-fit for a model with label vector l_n. Although *The Cannon 2* implementation that we used offers its own test step implementation[2], we wrote our own custom function that allows us to place further constraints on our model extrapolation during the test step, as described further in Section 3.4. We confirmed that our implementation performs just as well on ground truth label retrieval (see Figure 3.4) as *The Cannon 2*'s implementation. An illustration of the coefficient vectors that *The Cannon* computed during our model's training step can be found in Appendix .1.

3.3 Training Data

The Cannon requires a set of well-characterized stars with rest-frame, continuum normalized spectra on a uniform wavelength scale. It also requires the corresponding labels for all of

[2]https://github.com/andycasey/AnniesLasso

those stars. For our model, we train on labels T_{eff}, log$g$, [Fe/H], [α/H], v_{broad}. The stars and their labels should span the domain in which you want your model to perform well. Because *The Cannon* is measuring the flux dependence on the labels at the individual pixel level, spectra with high SNR and accurate, precise labels is best.

One million spectra of "well-behaved"[3] objects are available as part of the Gaia Data Release 3 (DR3) (Cropper et al., 2018). These spectra have spectral resolution $R = \Delta\lambda/\lambda = 11,500$ over infrared range $\lambda = 845 - 872\,\mathrm{nm}$. The spectra are rest-frame shifted, continuum-normalized and averaged over multiple visits over 34 months with an average of 8 transits per star per year (Katz et al., 2023). This, along with the large number of available spectra, makes them ideal for implementing a Cannon model. However, in order for our model to perform with higher precision and accuracy than the published Gaia stellar labels, we need to use higher fidelity reported labels for the stars in our training set.

For the labels corresponding to stars in our training set, we used labels computed by the GALAH+ survey's third data release (Buder et al., 2021). This survey published derived stellar labels for 588,571 nearby objects, including a large number of main sequence stars that were relevant as potential training objects for our model. Their labels were derived using a combination of GALAH spectra and Gaia DR2 and 2MASS photometry to improve the accuracy of the derived labels beyond the accuracy that the spectra alone can provide. Beyond this, GALAH spectra have higher resolution and more wavelength coverage than Gaia (R=$28,000$ over four $25\,\mathrm{nm}$ channels between 470 and $790\,\mathrm{nm}$), thereby providing higher fidelity labels that are ideal for label transfer. Their catalog also included a cross-match with Gaia DR3, which allowed us to map derived labels to the spectra in our training set.

Our original sample contained $588,571$ objects with GALAH-reported labels. From here,

[3] stars with SNR$\geq 15, 3100 < T_{\mathrm{eff}} < 14,500\,\mathrm{K}$, and no problems introduced by processing pipeline (Katz et al., 2023)

we applied a number of cuts to obtain a well-characterized sample of single stars that would optimize our model's performance. The cuts we made can be subdivided into three groups based on their broader purpose, as follows:

- *Quality of spectral data.* We required a SNR of at least 100/pixel for both the GALAH reference object (catalog label `snr_c3_iraf`) and Gaia RVS spectrum (catalog label `rvs_spec_sig_to_noise`) for each object in our training set. We also removed objects from GALAH that did not have a published Gaia RVS spectrum.[4]

- *Confidence in stellar labels.* We excluded objects that were flagged for poor label quality in the GALAH catalog. To enforce this, we required the stellar label, iron abundance, and α abundance quality flags (`flag_sp`, `flag_fe_h`, and `flag_alpha_fe`, respectively) to be zero. We also removed stars with any label whose uncertainty was more than twice the median uncertainty of the corresponding label in full GALAH catalog. Finally, we required finite GALAH-reported values for all 5 of our training labels.

- *Contamination from unresolved binaries.* In the event that the stars in our training sample are unresolved binaries, the GALAH-reported stellar labels will be biased and the model performance is compromised (El-Badry et al., 2018b, hereafter EB18). To ensure that the stars in our training set are all single stars, we removed GALAH stars that were identified as binaries in Traven et al. (2020). We also required stars to be well-fit by single star astrometric solutions. This information is encoded in the object's Renormalised Unit Weight Error (RUWE) value and Gaia `non_single_star` value (see Belokurov et al., 2020). For our training set, we require Gaia DR3 and Gaia DR2 RUWE < 1.4 and Gaia `non_single_star`$=0$. Finally, we remove potential binaries missed by the Gaia and GALAH detection methods with an iterative process

[4]The SNR>100 threshold was chosen because it was the lowest SNR threshold for which we were able to able to achieve output label precision comparable to the GALAH-reported label errors for stars in our training set.

described in Section 3.4.4.

We also limited our training set to main sequence stars by requiring $\log g > 4$. Of the 588,571 stars in the original GALAH sample, 28,530 remained after removing stars based on GALAH-reported SNR or quality flags, and 8,111 remained after removing stars with large label uncertainties, $\log g < 4$ and binaries identified by Traven et al. (2020). Only 632 of these had published RVS spectra with SNR$>$ 100, and this number was reduced to a final training set size of 563 after we iteratively removed potential binaries (see Section 3.4.4 for a description of this process). Figure 3.1 shows the $T_{\rm eff}$, $\log g$, [Fe/H] domain of our final training set.

Due to the fact that the RVS spectra are rest-frame shifted, the RVS spectra available in Gaia DR3 often have pixels with NaN values at either the beginning or the end of the spectrum. For our training set, $\sim 10\%$ of the spectra had NaN values in the first or last 20 pixels of the spectra. To address this, we clipped all of the spectra by 20 pixels on each side before training *The Cannon*. After clipping, only $\sim 1\%$ of training spectra had NaN pixels remaining in the spectra. To preserve the information in the 99% of spectra with finite values for these pixels, we interpolated the flux at these values and inflated their reported errors to infinity to effectively mask them during the training step.

3.4 Model Customizations

Our trained Cannon model predicts the flux at each RVS pixel as a function of the training labels over a wide range of parameter space. From there, we made a number of changes to the original Cannon test step described in Section 3.2.2 to improve our model performance and conducted tests to ensure that we can trust our model's output labels. The key customizations are described in the remainder of this section. We confirmed that our implementation

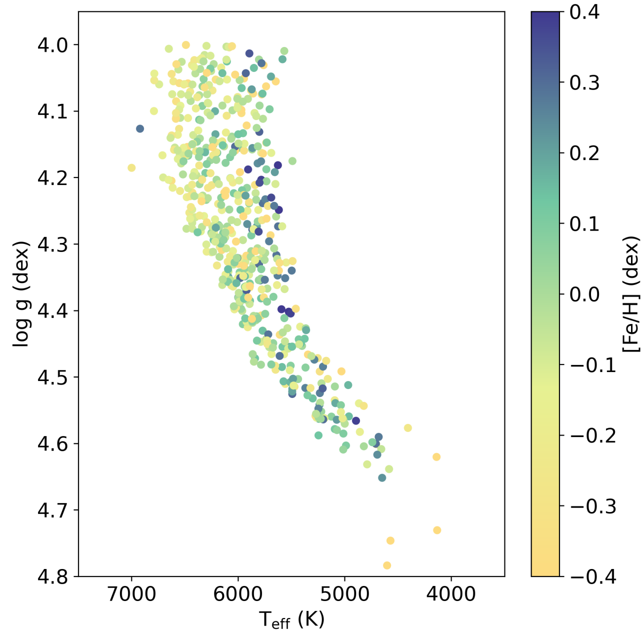

Figure 3.1: Distribution of T_{eff}, $\log g$, and [Fe/H] values for the stars in our model training set. The points represent individual stars in a 3D subset of our 5D training domain. Our model training set contains 563 well-characterized main sequence stars from GALAH with high-fidelity labels.

performs just as well on ground truth label retrieval (see Figure 3.4) as *The Cannon 2*'s implementation.

3.4.1 Masking the Calcium IR Triplet

The Gaia RVS spectral range includes 3 deep Calcium lines at 849.8 nm, 854.2 nm, and 866.2 nm, collectively known as the Ca-triplet. Certain astrophysical phenomena can manifest as superimposed emission and absorption contained to only these features in the RVS spectrum. Examples include interacting binaries, young accreting T Tauri stars, chromospheric activity in GKM main-sequence stars, emission linked to the accretion of matter in the stellar magnetosphere, etc (see Lanzafame et al., 2023, for a description). While we did not explicitly remove active stars from our training set, we assume that since our training set is composed of primarily of inactive, main sequence stars, the effects of activity would not be significant enough for *The Cannon* to learn for pixels outside the Ca-triplet. In other words, *The Cannon*'s assumption that flux depends solely on static labels breaks down at the Ca-triplet, where time-dependent surface features and activity contribute to the flux as well. This was demonstrated by Rampalli et al. (2021), who trained a polynomial regression model on RVS spectra and found that it over- or under-predicted flux at the Ca-triplet for 82% of the stars in their sample.

Given this assumption, while our best-fit Cannon model should produce a good fit to the spectrum at all the lines when the star is inactive, this should not be the case for active stars. Instead, for stars with activity-induced Calcium emission, *The Cannon* model should be a good fit everywhere *except* the Calcium features. In these cases, our Cannon model might fit to these features at the cost of providing a good fit to the rest of the spectrum. Thus, our model performance is compromised when we include these features in our Cannon test step. To avoid this, we mask out the Ca-triplet when we perform a model fit to a spectrum by setting their flux errors equal to be infinite. We note, however, that this mask is not driven by a need in the data; because stellar activity affects flux in the cores of the Ca-triplet lines and isn't included as a training label, we infer from this knowledge that the model performance is compromised for these features.

In the published model and validation steps in the subsequent section, we are fitting to the non-shaded parts of Figure 3.2. We stress that since *The Cannon* fits each wavelength pixel independently, including these masked pixels in the training set does not introduce any bias in the rest of the spectrum. Consequently, our trained Cannon model still predicts the flux for the Ca-triplet features as a function of stellar labels, though we caution that these predictions may not be reliable.

Figure 3.2 shows a test case where our Calcium mask is essential for recovering the correct labels for the spectrum. We show an active dwarf identified by Lanzafame et al. (2023) with a Cannon model fit that was computed with a mask on the Calcium features. We see

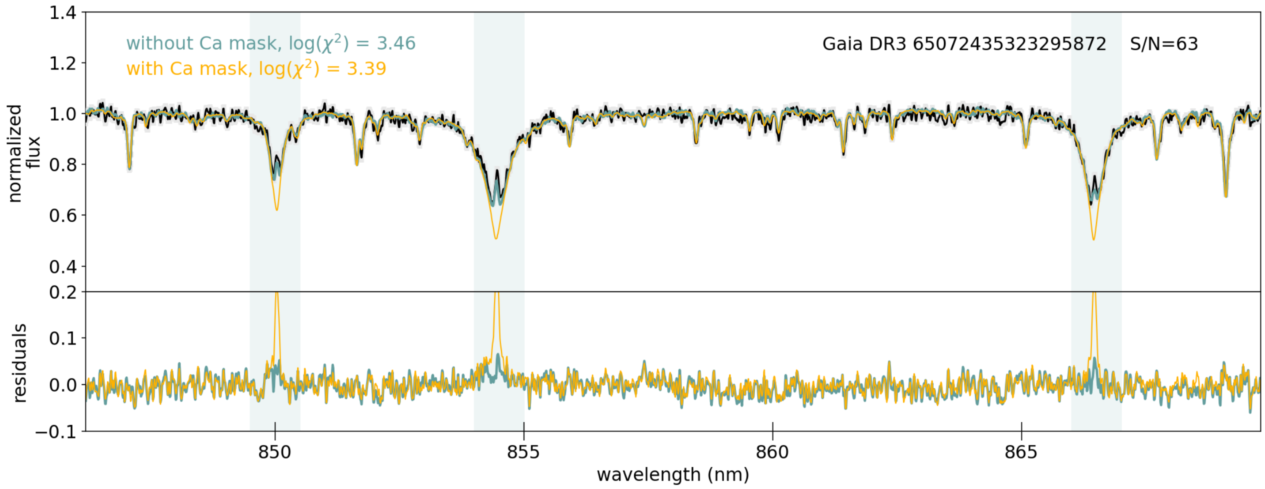

Figure 3.2: Demonstration of why a calcium mask is needed to correctly derive labels for RVS spectra. The spectrum shown in black is an active dwarf star with superimposed emission and absorption in the Ca-triplet lines. Without masking the shaded regions, *The Cannon* fits to the Ca-triplet at the expense of finding a good fit to the other features (shown in blue). When we include a mask, the best-fit model describes the spectrum well everywhere except the Calcium lines (shown in yellow).

that the model is a good fit to the entire spectrum, except the Calcium features which are shallower than the model predicts due to emission at those features that our model does not account for. Additionally, the best-fit model for a masked spectrum predicts what the flux at the Ca-triplet should be for that star in the absence of activity. Thus, we can identify active stars in cases where the true flux deviates significantly from the flux of the best-fit model computed with the Ca-triplet mask (see Section 3.6.1 for a detailed description of this process).

3.4.2 Empirical Model Scatter

As outlined in Section 3.2, our model uses a quantity called s^2 to compute the χ^2 goodness-of-fit associated with a particular set of labels for a given spectrum. This term describes the per-pixel intrinsic model scatter, or in other words, it encapsulates any pixel-level variation in the training set spectra that cannot be attributed to photon noise. When we first trained our model, the training step found s^2 to be $\sim 10^{-4}$ at the 3 pixels corresponding to the centers of the Calcium line cores and zero for all other pixels(see Casey et al., 2016, for a description of how the training step computes s^2). While this original s^2 is able to compute fits with reasonable χ^2 for SNR$\sim$50-100 where the deviation from the model is dominated by photon noise and instrumental effects, fitting our model to higher SNR targets from the SPOCS and CKS samples revealed clear correlation between SNR and χ^2 for stars with SNR$\gtrsim$ 200. This is because stars with significantly higher SNR than the training set possess intrinsic scatter that is lost to photon noise for the SNR$\sim$100 training set stars. In other words, the s^2 originally computed in our Cannon training step was underestimated, inflating the χ^2 for targets with high SNR.

To address this, we computed an empirical per pixel model scatter that we adopt for our model's s^2 term. We gathered all stars in Gaia with published SNR$\geq$500 RVS spectra and labels that reside reasonably within the training domain ($\log g >4$ dex, $4000 < T_{\rm eff} < 8000$ K). We removed obvious binaries by requiring RUWE$<$ 1.4, `non_single_star`=0 and downloaded the 94 remaining RVS spectra. We expect the primary deviation from our model to be due to intrinsic model scatter for these spectra since their photon noise is so low. Thus, for each model we computed the best-fit Cannon model and recorded the empirical scatter between the true flux and the best-fit model (i.e., $|\mathrm{data}-\mathrm{model}|^2$). We repeated this calculation for each of the spectra in our high SNR sample and computed the average across all stars for each pixel to arrive at an empirical s^2 prediction. We found that the best-fit Cannon model differs from the RVS spectra by $s^2 = 0.1 - 1\%$ for these objects, which is more than the $\leq 0.1\%$ variations predicted from their photon noise alone. We verified that including our empirical s^2 term successfully removed the observed correlation between high SNR and χ^2 that we saw for the SPOCS sample originally. Thus, we use the empirical s^2 term for our model throughout the entire analysis of this paper. A plot of our per pixel empirical s^2 can be found in Appendix .1.

We do not expect the training step to infer zero scatter for a training set with accurate flux errors. However, if the flux errors are overestimated, they may be sufficient to explain the model's deviation from the true fluxes without any intrinsic model scatter. Indeed, when we computed best-fit Cannon models for a sample of main sequence single stars in Section 3.6, we found that the distribution of χ^2 values peaks at $\sim$2150, which is slightly smaller than the expected peak near the total number of pixels in the spectrum (2361 for our training set). This indicates the Gaia RVS flux errors may be overestimated for the Gaia RVS spectra. Thus, we suspect that overestimated flux errors are the reason that our model's test step computes a zero scatter during its optimization.

3.4.3 Constraints on Extrapolating Beyond the Training Domain

During the test step outlined in Section 3.2, *The Cannon* is free to extrapolate to labels outside of the training domain. In this regime, however, the labels that best fit the model spectrum are not tied to any ground truth, and the labels output by the model are less reliable. To address this, we place constraints on the degree to which we allow *The Cannon* to extrapolate beyond the training domain during the test step. We measure the extrapolation using a 5-dimensional Gaussian kernel density estimation (KDE; Rosenblatt, 1956), which uses non-parametric kernel smoothing to approximate the probability density of stars in the training domain. We refer to this quantity hereafter as the training label density, $\rho(l_n)$. The training label density is a function of the 5 Cannon model training labels ($T_{\rm eff}$, log$g$, [Fe/H], [α/H], $v_{\rm broad}$), and the output represents the density of training set objects near those labels in the training domain. A Cannon fit that returns labels with a high $\rho(l_n)$ is the product of a reasonable interpolation within the training domain, thereby producing a reasonable model fit with accurate labels. In contrast, a Cannon model with a low $\rho(l_n)$ is extrapolating beyond the training domain to produce a model with labels that are not tied to ground truth.

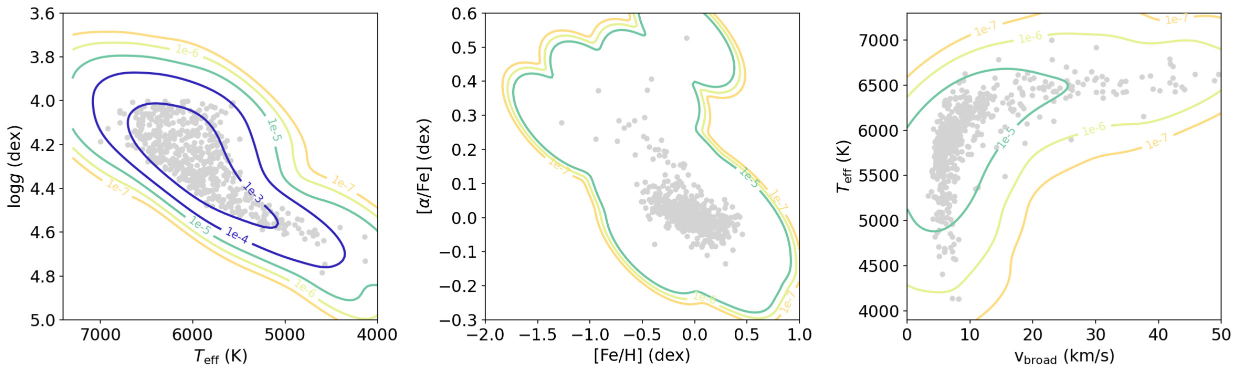

Figure 3.3: Illustration of our model's $\rho(l_n)$ parameter. We plot the distribution in 2D, where each point represents a model in our training density. Densely population regions of this space correspond to domains with high $\rho(l_n)$. We over plot lines of constant $\rho(l_n)$ in each panel. Our model penalizes labels that lie beyond the orange lines, where $\log \rho(l_n) > -7$.

To prevent *The Cannon* from extrapolating beyond the training domain when predicting labels, we determined a reasonable $\rho(l_n)$ range to allow *The Cannon* to model by inspecting the $\rho(l_n)$ distribution for all of the stars in our training set. Figure 3.3 shows all of the stars in the training set as a function of their labels, with contours to illustrate lines of constant KDE-estimated $\rho(l_n)$. We used a leave-one-out technique where for each star, the KDE used to determine $\rho(l_n)$ is calculated using the full training set with the individual star held out. This prevents the distribution from piling up around $\log \rho(l_n)$ =-4 due to the fact that the object we were evaluating went into estimating the KDE that was used to compute its $\rho(l_n)$. During our test step, we calculate $\rho(l_n)$ using a KDE computed with every object in the training set.

As shown in Figure 3.3, objects in the training set typically reside in densely populated regions of training domain, with $\rho(l_n) \sim 10^{-3}$. This number falls off quickly, and very few stars in the training set have $\rho(l_n) \sim 10^{-7}$. Thus, to prevent our model from extrapolating beyond the training domain, we updated the χ^2 value that is minimized in the test step

(Equation 4.10) to include an inflation factor that penalizes models with $\rho(l_n) < 10^{-7}$:

$$\hat{l}_n = \arg\min_{l_n} \sum_{j=1}^{N_{\rm pixels}} -\ln p \times P(l_n) \tag{3.6}$$

where $P(l_n)$ is a penalty factor that penalizes sets of labels l_n with $\rho(l_n) < 10^{-7}$:

$$P(l_n) = \left\{ \begin{array}{ll} 1, & \text{if } \rho(l_n) > 10^{-7} \\ 1 + \log\left[\frac{10^{-7}}{\rho(l_n)}\right], & \text{if } \rho(l_n) \leq 10^{-7} \end{array} \right\} \tag{3.7}$$

We do not include a penalty term for $\rho(l_n) > 10^{-7}$ to avoid favoring models simply for being in the most populated regions of training domain. The addition of 1 to the fractional penalty ensures that the likelihood function is continuous at $\rho(l_n) = 10^{-7}$. We also sample $\log \mathrm{v}_{\rm broad}$ instead of $\mathrm{v}_{\rm broad}$ in our optimization routine to avoid fitting with non-physical $\mathrm{v}_{\rm broad} < 0$ values.

3.4.4 Removing Spectral Binaries from the Training Set

While the binary cuts outlined in Section 3.3 remove binaries that were previously identified by Gaia, it is possible that our original training set is contaminated by unresolved spectroscopic binaries. In particular, binaries with separations wide enough to produce overlapping spectral features are often mistaken for single stars and are likely abundant in the RVS sample. In the event that our training set is contaminated with these “spectral" binaries, the training step will not accurately learn how flux depends on labels for single stars, and our Cannon model is not reliable.

To address this, we follow a similar procedure to that outlined in EB18 and iteratively remove these binaries from our training set. In brief, we train an original Cannon model on our initial training set and use this model to identify binaries (see Section 4.5 for a detailed explanation of this process). We then remove the identified binaries from the training set and train second model with the "cleaned" trained set, and repeat this process until model no longer identifies binaries in its own training set. We adopt a conservative criteria for binaries in this step and remove spectra with $\log \Delta\chi^2 \geq 2$ from the training set at each iteration, where $\Delta\chi^2 = \chi^2_{\text{single}} - \chi^2_{\text{binary}}$ describes how well the binary model fits the data compared to the single star model, as described further in Section 4.5.

3.5 Model Performance

3.5.1 Validation Against GALAH Labels

We investigated how well our model is able to predict stellar labels by running the test step on the spectra into the training set and compared *The Cannon*-inferred labels to known "ground truth" labels from GALAH. For this, we used leave-one-out cross-validation, where for each object in the training set we trained a new model with a new training set that is identical to the original, but with the object of interest held out. We then used each model to run the test step on the object that was held out of its training set. This allowed us to derive labels for each star in the training set without the bias introduced when the test step is run on an object that was in the model's original training set. Figure 3.4 shows the GALAH labels plotted versus *The Cannon* output labels found using *The Cannon* test step outlined in Section 3.2.2. The diagonal line represents the one-to-one line, where points should fall if our model is in perfect agreement with the ground truth labels.

Table 3.1: Average Cannon Model Disagreement with "Ground Truth" GALAH Labels for Training Set Stars

Label	rms (GALAH − Cannon)	**Training Set Range**
T_{eff}	72 K	4132-6999 K
$\log g$	0.09 dex	4.00-4.78 dex
[Fe/H]	0.06 dex	-1.27-0.45 dex
$[\alpha/Fe]$	0.05 dex	-0.13-0.53 dex
v_{broad}	1.9 km/s	3.34-50.23 dex

The root-mean-square (RMS) difference between the GALAH and *The Cannon* labels is listed in Table 3.1, along with the range for each label seen by our training set[5].

A Cannon model's performance is limited by the precision of its input training set labels, thus the model is performing well when it's agreement with the training labels is comparable to the label uncertainties. We overlay the average training set label uncertainties in Figure 3.4. As we can see from the figure, the model uncertainties in Table 3.1 are comparable, albeit slightly larger, than the average GALAH-reported uncertainties. The difference can likely be attributed to the difference in data quality for the two surveys – the GALAH spectra have higher resolution ($R \sim 28,000$, compared to 11,500 for Gaia) and a longer bandpass ($\sim$ 300 nm, compared to 25 nm for Gaia). Additionally, the RVS spectra of the objects in our training set are noisier than those in GALAH, with average SNR$\sim$ 139 and 168, respectively. The only exception to this is $\log g$, where our Cannon model agrees with the GALAH-reported logg to within 0.09 dex, which is less than the average uncertainty of 0.18 dex for our training set. This is likely due to the fact that *The Cannon* learns only the statistical uncertainty in the training set labels as opposed to any systematic uncertainties that may be present in the GALAH labels. Thus, if the GALAH labels have systematic

[5] While these ranges represent a useful 1D approximation of whether a given set of labels is the training set domain, $\rho(l_n)$ provides a better estimation.

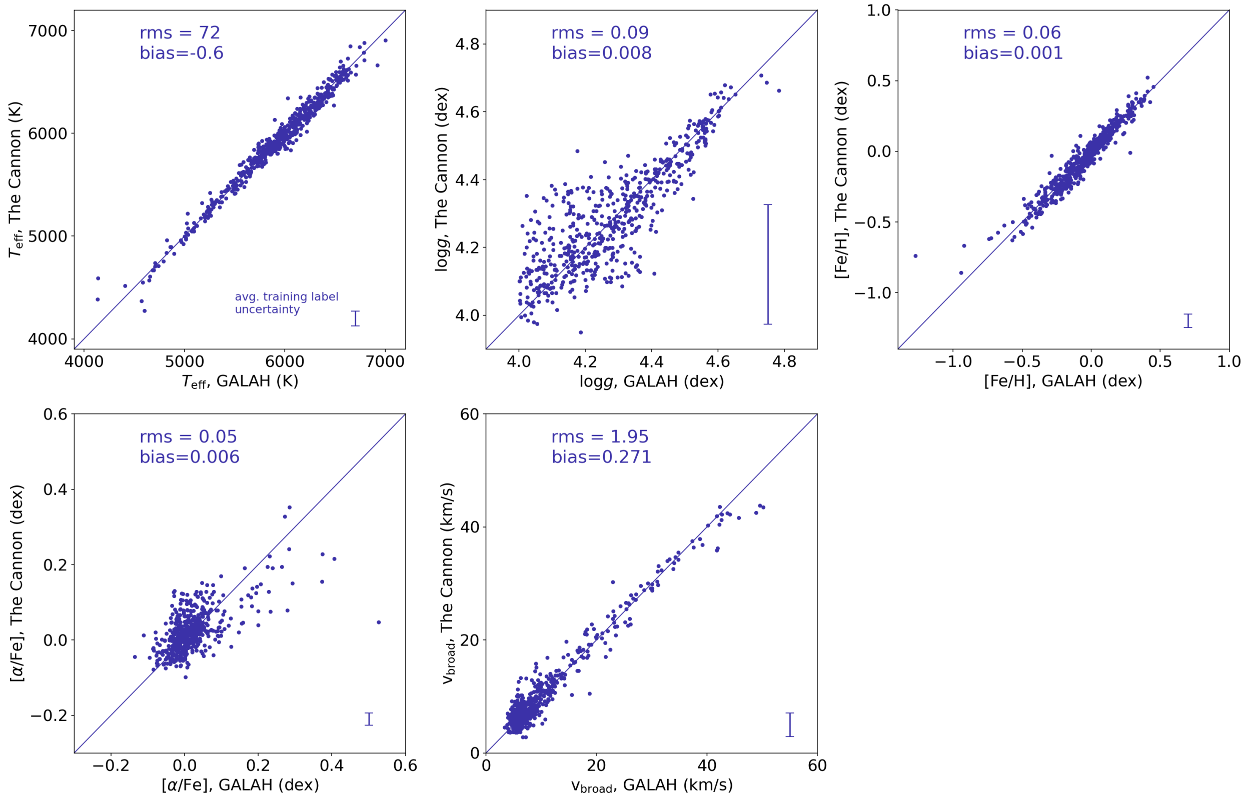

Figure 3.4: Cannon label outputs are plotted versus "ground truth" labels from the GALAH survey for each star in our training set. *The Cannon* output labels are computed using the leave-one-out cross-validation described in Section 3.4. The root-mean-square scatter and bias values are shown to quantify the average disagreement between *The Cannon* output and GALAH labels.

errors that effect multiple measurements of the same star in the same way, the GALAH errorbars will reflect this, but *The Cannon*/GALAH RMS scatter we report will not.

3.5.2 Performance on Planet Hosts

To investigate our model's ability to recover labels of planet hosts, we compared our Cannon output labels to labels reported by the Spectroscopic Properties of Cool Stars (SPOCS; Brewer et al., 2016). The SPOCS sample is a benchmark catalog of high fidelity stellar properties inferred from high-resolution ($R \sim 50,000$) high-SNR ($\gtrsim 100$/pixel) spectra of 1615 main sequence planet hosts using a bespoke line-by-line spectral analysis of high-resolution spectra from the Keck High Resolution Echelle Spectrometer (HIRES). Of the 1615 individual stars with SPOCS spectra and labels reported in Brewer et al. (2016), 107 are main sequence stars with published RVS spectra and reported labels spanning our model's training domain.

We used our model to compute best-fit models and corresponding stellar labels for each of these Gaia RVS spectra. The right panel of Figure 3.5 shows how well our Cannon-inferred labels agree with those reported by SPOCS for all 107 stars, and Figure 3.5 shows the same for the 80-star subset with RVS SNR>50. As we can see from these figures, the degree to which our model agrees with the high-fidelity SPOCS labels is largely limited by the SNR of the RVS spectrum that our fits to. For RVS spectra with SNR>50, we find that our model reproduces stellar labels that agree with SPOCS labels to within 85 K in $T_{\rm eff}$, 0.12 dex in $\log g$, 0.09 dex in [Fe/H], and 2.3 km/s in $v_{\rm broad}$. These label precisions indicate a 40% improvement over the Gaia `GspSpec` catalog's ability to reproduce the SPOCS labels, and 20% improvement over the `GspPhot` catalog.

The improved agreement with SPOCS labels over pre-existing Gaia catalogs has promising implications for constraining labels of exoplanet hosts without dedicated ground-based follow-up. In particular, for the majority of *Kepler* planet hosts, the most updated labels come from the Kepler Input Catalog (KIC), which published labels for stars in the `Kepler` field derived from stellar broadband photometry (Brown et al., 2011). These labels have large uncertainties (200 K in $T_{\rm eff}$ and 0.4 dex in $\log g$ and [Fe/H]), thus the characterization of these planets is limited by the degree to which their host properties are known. We used the Gaia-Kepler crossmatch database[6] to identify main sequence KIC stars with published RVS spectra with SNR>50 and $T_{\rm eff}$ within our training domain. We used our model to compute updated labels for these stars, along with their χ^2, $\rho(l_n)$, and $\chi^2_{\rm Ca}$ values (see Section 3.6 for a description of these metrics). Table 3.2 shows our reported labels for these stars.

[6]https://gaia-kepler.fun/

Table 3.2: Cannon Output Labels for Selection of KIC Stars

KIC ID	**Gaia DR3 ID**	$\mathbf{T}_{\mathrm{eff}}$	$\log g$	**[Fe/H]**	**[α/Fe]**	$\mathbf{v}_{\mathrm{broad}}$	**log** χ^2	**log** $\rho(l_n)$	**log** χ^2_{Ca}	**log** $\Delta\chi^2$
1430163	2051012738300892416	6410	4.1	-0.13	0.0	12.56	3.3	-2.68	2.95	0.46
1434277	2051704953890021504	5782	4.46	0.18	-0.02	6.39	3.37	-2.71	3.0	1.35
2010835	2051082179332118912	6061	4.35	-0.23	0.07	2.89	3.29	-3.55	2.83	1.08
2158850	2052529106573251584	5937	4.3	-0.12	0.09	7.78	3.3	-3.3	2.77	1.29
2285598	2099130979202197376	5575	4.54	0.07	0.13	8.51	3.76	-7.0	2.62	3.53
2297349	2051098465848130816	6149	4.2	-0.47	0.14	8.09	3.38	-4.12	2.79	2.21
2425346	2099128848898510336	6021	4.24	0.06	0.01	7.81	3.37	-2.45	2.72	1.59
2441864	2052545599247719808	5947	4.07	0.19	-0.02	7.09	3.33	-2.93	2.68	2.04
2447773	2051844591866815616	5930	4.11	0.15	-0.01	2.28	3.38	-3.21	2.69	1.56
2448382	2051841598266424448	5901	4.34	-0.42	0.12	4.88	3.34	-3.69	2.57	1.39

Note.

A machine-readable version of this table in its entirety is available in the accepted version of this paper, and a portion of the data is shown here. We report labels and metrics for all main sequence KIC stars with published RVS spectra in Gaia DR3 with SNR>50, but caution that labels may be unreliable for stars with $\log \rho(l_n) < -6$.

3.5.3 Comparison to Gaia-reported labels

Our Cannon model reports T_{eff}, log$g$, [Fe/H], [α/H], v_{broad} for any of the million stars with RVS spectra. Gaia reports these labels as well. Additionally, the Gaia `GspPhot` catalog reports stellar labels derived from the Bp/Rp spectra, which exist for ~ 100 million stars in Gaia DR3 (Andrae et al., 2023). Stellar labels derived from the RVS spectra are also reported in the `GspSpec` catalog using MatisseGauguin and Artificial Neural Net (ANN) algorithms (Recio-Blanco et al., 2023). To place our model in the context of the existing Gaia catalogs, we compared its output labels, the `GspPhot` labels, and the `GspSpec` labels to "ground truth" SPOCS labels for the same subsample of SPOCS stars from Section 3.5.3. We note that in the case of [Fe/H], SPOCS and Gaia `GspPhot` report [M/H] instead of [Fe/H], so the comparison may not be strictly one-to-one. However, since [Fe/H] and [M/H] are correlated and the model learns from every pixel and not just the iron lines, the [Fe/H] we predict should be an adequate predictor of M/H as well.

The labels from all three sources are plotted against the SPOCS labels in Figure 3.5, with the RMS scatter and bias listed for each panel. The subset of stars with Gaia RVS SNR>50 are outlined to show the improvement in our model performance for high SNR. Our Cannon model agrees more closely with the SPOCS labels than the `GspSpec` labels, and are on par with the `GspPhot` labels. This is true for both the full sample and the high-SNR subsample. Beyond this, our model is useful for predicting labels that are missing from either or both of the Gaia catalogs (plotted as vertical lines in Figure 3.5). As shown in the figures, Gaia is missing labels for a number of planet hosts in the sample. This is particularly the true for the RVS-derived metallicity and v_{broad} from `GspSpec` (middle panels of rows 3 and 4). Our Cannon model is able to fill these gaps, since it predicts all 5 labels for every star with an RVS spectrum. Beyond this, our model computes [α/Fe], which is missing entirely from both Gaia catalogs.

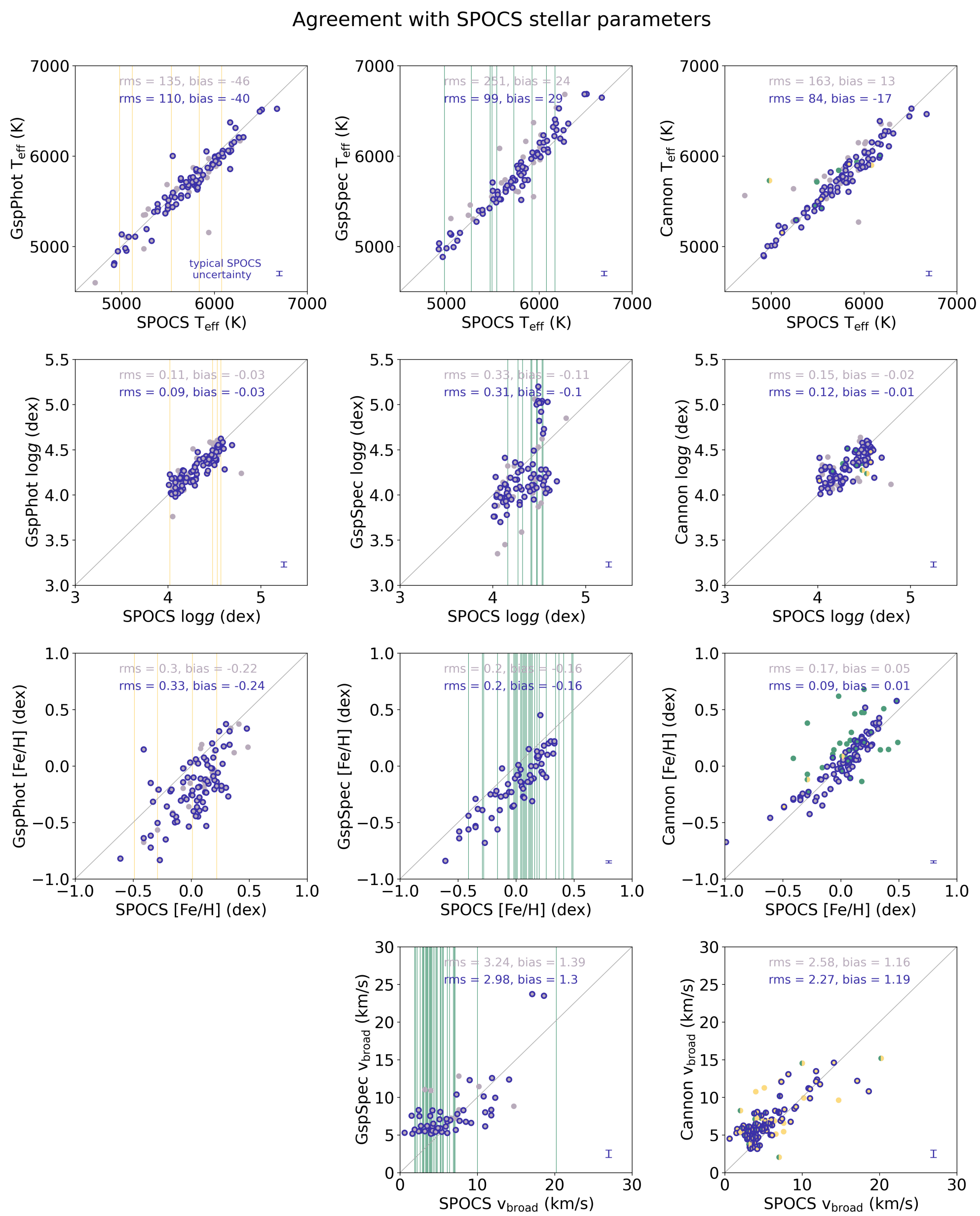

Figure 3.5: Stellar labels from Gaia `GspPhot` (left), Gaia `GspSpec` (middle), and our Cannon model (right) are plotted versus SPOCS labels for the 107 SPOCS stars with RVS spectra from DR3. The full sample is plotted in grey, and stars with RVS SNR>50 are outlined in blue with the rms scatter and bias in the corresponding colors. Labels missing from Gaia `GspPhot` are represented as yellow lines in the left panel and yellow circles in the right panel. Similarly, labels missing from Gaia `GspSpec` are represented as green lines in the middle panel and green circles in the right panel. Points that are half yellow and half green in the right panel are missing from both Gaia catalogs. We note that (1) We compare [Fe/H] values to [M/H] values for SPOCS and Gaia `GspPhot`, and compute [Fe/H] $=$`fem_gspspec`$+$`mh_gspspec` for Gaia `GspSpec`, since these catalogs do not report [Fe/H] directly. (2) while Gaia's `vbroad` is not technically part of the `GspSpec`, we include it because it's derived from the RVS spectrum (Frémat et al., 2023).

3.6 Applications for Identifying Anomalous Spectra

In addition to stellar label determination, our model can be used to identify spectra that deviate from the main sequence single stars that our model was trained on. In this section, we describe metrics that our model computes to flag these anomalous spectra and instructions on how to interpret them. We also provide specific examples of these cases.

Our model computes four metrics that can be used to identify anomalous spectra. The first two, $\chi^2_{\rm Ca}$ and $\Delta\chi^2$, are designed specifically to identify chromospherically active stars and binaries, respectively and are described in detail in the following sections. The third χ^2 metric measures how well *The Cannon* model fits the data. This value is minimized during the model's test step (see Section 3.2.2), and is large for any spectrum that is not well-fit by a main sequence single star model. The fourth metric is the $\rho(l_n)$ metric described in Section 3.4.3. A low $\rho(l_n)$ indicates that the model had to extrapolate to increase it's flexibility to describe the spectrum. In this case, the fit may or may not have a low χ^2, and the inferred labels are unreliable.

To determine reasonable ranges of our metrics that describe main sequence stars for which our model can compute accurate labels, we plot the distribution of these metrics in Figure 3.6 for a sample of main sequence single stars. This sample comes from Table E3 of EB18, and is composed of stars for which no unresolved spectroscopic binaries was detected in the paper. We removed stars with RUWE>1.4 or Gaia `non_single_star`=1 to further ensure the sample contained only single stars. We also required the stars to be main sequence ($\log g$ >4) and removed low SNR targets `rvs_spec_sig_to_noise`<50. The histograms in Figure 3.6 show the distributions of $\chi^2_{\rm Ca}$, $\Delta\chi^2$, χ^2 and $\rho(l_n)$ for our single star sample. The remainder of this section outlines specific test cases for which these metrics will be outside the nominal range for main sequence single stars.

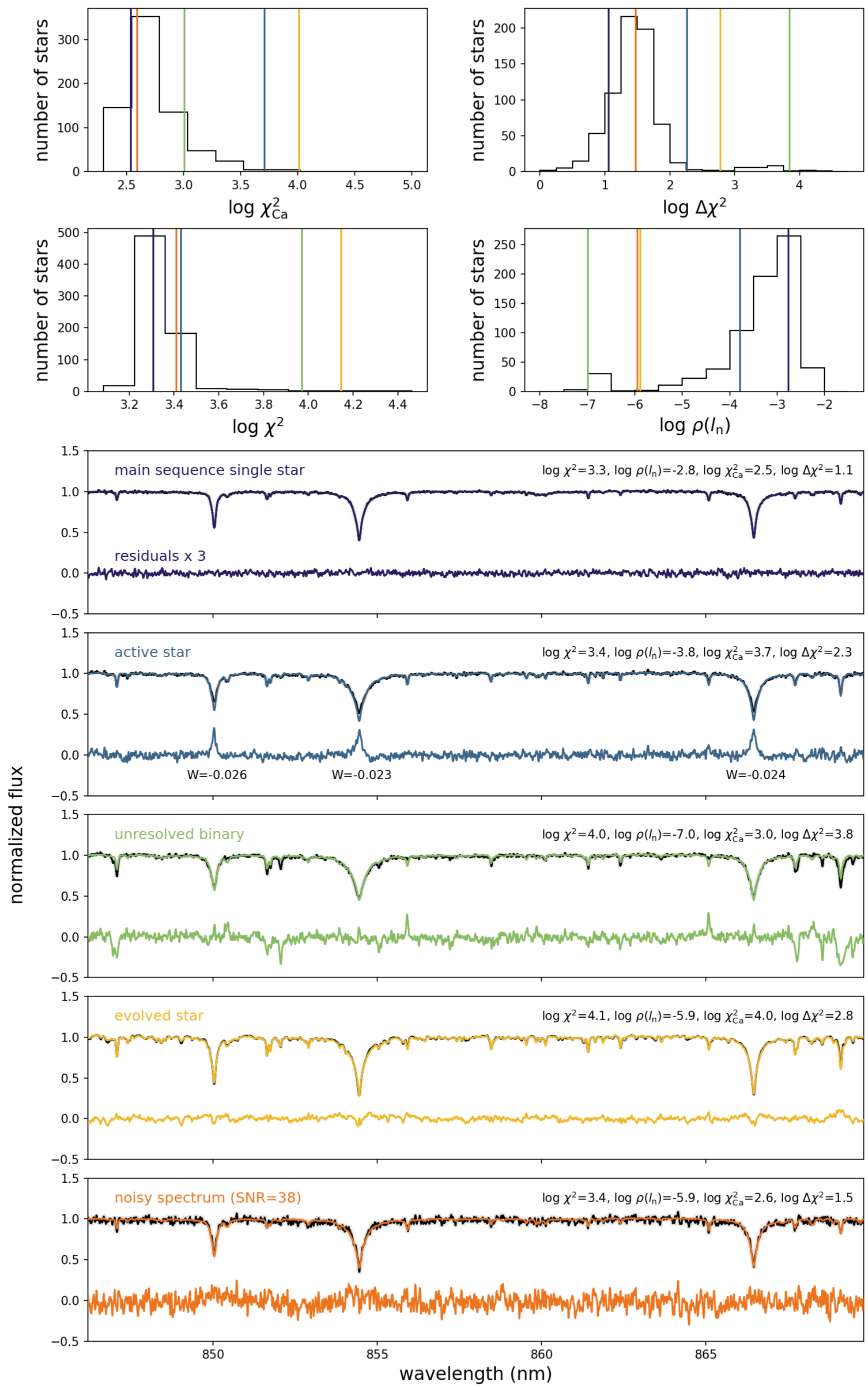

Figure 3.6: Top row: Distributions of abnormal star metrics for a vetted sample of main sequence single stars with SNR>50. Bottom panel: Example spectra with best-fit spectra and inferred parameters. Model residuals are plotted below and are inflated by a factor of three for visibility. From top to bottom: Gaia DR3 839728576668869760, a main sequence single star for comparison; Gaia DR3 4268620287278693120 (HD 176650), an evolved star with $\log g \sim 3$; Gaia DR3 3663127407180728704, a noisy (SNR=30) spectrum of a single star from EB18; Gaia DR3 3174658066485297152, an active dwarf with superimposed Calcium emission and absorption from Lanzafame et al. (2023); Gaia DR3 1535964555128078720, an unresolved binary identified in EB18.

3.6.1 Active Stars

While we mask out the Ca-triplet during our model's test step, these spectral features still contain important information about the star. In particular, stars that are chromospherically active, undergoing accretion or part of a pair of interacting binaries may exhibit Calcium emission that gets superimposed on the Ca-triplet absorption features (Lanzafame et al., 2023). Since our model is assumed to predict flux for *inactive* stars (see Section 3.4.1), activity should manifest as a poor model fit around the Calcium features where the model underestimates the flux. To provide a quantitative metric for this activity, we compute the star's χ^2_{Ca} evaluated only at the wavelengths containing the Ca-triplet– that is, the wavelengths that are masked during the model's test step.

We note that this metric is similar to the `activityindex_espcs` that Gaia reports, but with a few notable differences. First, we compute χ^2_{Ca} by comparing the spectrum to a best-fit Cannon model instead of a template spectrum, thereby measuring the extent to which the target object differs from other stars with otherwise similar labels. This allows us isolate the effect of chromospheric emission from natural variance of Ca-triplet flux for stars with different locations on the HR diagram. Second, our Cannon model allows us to inspect the nature of the star's behavior at the Ca-triplet through the visual inspection of the residuals and the lines' equivalent width, whose sign indicates whether the spectrum exhibits whether the spectrum exhibits Calcium emission or absorption relative to the best-fit model. Finally, Gaia doesn't report the `activityindex_espcs` for all stars. Of the $\sim$1 million stars with RVS spectra, about 1/3 have reported `activityindex_espcs`. Our model can compute reliable χ^2_{Ca} for any star with an RVS spectrum and reasonable (>50) SNR.

The distribution of χ^2_{Ca} values for our sample of main sequence single stars from EB18 is shown in the top right histogram in Figure 3.6. Objects with χ^2_{Ca} values that are much larger than the stars in this distribution have Ca triplet features that differ significantly than those in the training set. An example is shown in the second spectrum from the top for an active dwarf from Lanzafame et al. (2023) with an abnormal emission superimposed in the Calcium absorption features. In addition to computing χ^2_{Ca}, our model also enables visual inspection of the star's residuals at the Calcium III lines, along with their equivalent widths (which we denote as W). Equivalent width values in the single star sample from EB18 are typically within a range of 0 ± 0.01, so $W \approx -0.02$ for this target indicates significant activity-induced Calcium emission, along with a large χ^2_{Ca}. However, we caution that a large χ^2_{Ca} can also be the result of a poor quality-of-fit at *all* wavelengths. In this case, the large χ^2_{Ca} would instead be due to an unresolved binary, evolved star or noisy spectrum, as shown in the lower panels of the figure. Thus, χ^2_{Ca} is only a robust activity indicator in the absence of a large χ^2 and small $\rho(l_n)$ that suggest the model fit is unreliable, as described further in Section 3.6.4.

3.6.2 Unresolved Binaries

Binaries with < 1.8" orbital separations are not resolved by the Gaia RVS instrument and thus produce spectra that are inconsistent with the spectra of single stars (see Cropper et al., 2018, for a description of the RVS instrument design). Among these, binaries with $< 1\,\text{AU}$ physical separations appear as double-lined spectra ("SB2s") and are not processed by the Gaia's main spectroscopy pipeline (Gaia Collaboration et al., 2023). However, binaries with smaller relative velocities and separations close enough for both stars to fall within the RVS aperture may remain in the current RVS sample as mis-classified single stars. These spectral binaries produce two sets of *overlapping* spectral features and can be identified via Bayesian model comparison of a single-star spectrum with a binary spectrum–that is, a linear combination of two single-star spectra (e.g., EB18). Importantly, for a spectral binary to produce a continuum-normalized spectrum that is significantly different from a single star spectrum with the labels of the primary, the secondary must contribute significant flux without being too similar to the primary. This implies sensitivity to ∼1-200 AU binaries with mass ratios anywhere from $q \sim 0.4 - 0.9$[7].

We model a binary system as combination of a primary star and secondary star with respective labels l_1, l_2 and radial velocity (RV) offsets v_1, v_2 relative to the Gaia RVS instrument rest frame[8]. Our Cannon model predicts the rest-frame flux for each of the individual components as a function of stellar labels, $f_1(l_1)$ and $f_2(l_2)$. To account for the RV offsets, we compute wavelength shifts for each binary component relative to the rest frame $\Delta\lambda_{1,2} = \lambda_{1,2}v_{1,2}/\mathrm{c}$. We then interpolate *The Cannon*-predicted fluxes onto wavelength grids shifted by $\Delta\lambda_{1,2}$ to predict the individual spectra of the primary and secondary, $f_1(l_1, v_1)$ and $f_2(l_2, v_2)$. We account for the relative flux contributions of the primary and secondary by multiplying $f_1(l_1, v_1)$ and $f_2(l_2, v_2)$ by flux-weighted pre-factors W_1 and W_2:

$$f_{\mathrm{binary},j} = W_1 \times f_{1,j}(l_1) + W_2 \times f_{2,j}(l_2). \tag{3.8}$$

where W_1 and W_2 are calculated based on the flux ratio of the primary and secondary stars, $F_{\mathrm{rel}} = F_1/F_2 = 10^{-(m_1-m_2)/2.5}$. Here, m_1 and m_2 are the primary and secondary apparent magnitudes, which we compute using empirical magnitude-temperature relations for main sequence stars reported in Pecaut and Mamajek (2013b). W_1 and W_2 are then derived from from F_{rel} as follows:

$$W_1 = \frac{F_1}{F_1 + F_2} = \frac{F_{\mathrm{rel}}}{F_{\mathrm{rel}} + 1} \tag{3.9}$$

$$W_2 = 1 - W_1 \tag{3.10}$$

to assert that the binary flux is continuum-normalized to unity with realistic contributions from the primary and secondary star. The final product is a prediction of the flux at each pixel for a binary system as a function of the primary and secondary labels and RV offsets.

[7] The specific q sensitivity range varies slightly for different instruments, and sensitivity to binaries with $q > 0.9$ can be reached for binaries with larger velocity offsets (see El-Badry et al. 2018a, EB18)

[8] While we expect the relative RV to be small for overlapping spectra, we model non-zero v_1 and v_2 to account for sub-pixel RV shifts between the primary and secondary

To fit our binary model to an individual RVS spectrum, we repeat the test step described in Section 3.2.2 but evaluate the likelihood of our binary model instead of the single-star Cannon model. To ensure that our model predicts physically realistic binary systems, we assert that the primary and secondary stars have identical chemical compositions (i.e., [Fe/H] and [α/Fe] are the same for both stars). We also adopt the same training density threshold as our single star model (Section 3.4.3) to prevent fitting with unrealistic primary or secondary labels.

One notable difference between the Gaia RVS spectra and the APOGEE spectra used in EB18 is that the Gaia spectra are averaged over anywhere from $\sim 1-100$ epochs over a baseline of ∼three years. Consequently, for binary systems whose relative radial velocity changes by at least one resolution element ($26\,\mathrm{km/s}$)[9] the effect of the secondary is smeared out when the spectra are stacked according to the primary's spectral features and the spectrum does not represent a physical system. In this case, we expect that our model is unreliable, and an inferred RV offset $\geq 26\,\mathrm{km/s}$ should reflect this. For other binaries, the relative RV may not change significantly over the three-year baseline, and our model assumption of a constant relative RV between the primary and secondary stars is reasonable.

For a binary to be identified from its RVS spectrum, its spectrum should be significantly better fit by the binary model than the single star model. This information is contained in the value of $\Delta\chi^2 = \chi^2_{\mathrm{single}} - \chi^2_{\mathrm{binary}}$. It is important to note that due to increased flexibility of the binary model, an RVS spectrum should always be better fit by the binary model than the single star model. Thus, a robust binary detection requires $\Delta\chi^2$ to be large relative to a sample of single stars. To determine whether this is an effective metric for identifying binaries in the RVS sample, we computed $\Delta\chi^2$ for a sample of known single stars and binaries identified by EB18 (Table E3) with RVS SNR≥ 50. Our results are shown in Figure 3.7.

As we can see from this Figure, $\Delta\chi^2$ doesn't highlight a clear separation binaries and single

[9] $c/R = 26\,\mathrm{km/s}$ for the Gaia RVS instrument's R=11,500, see Cropper et al. (2018)

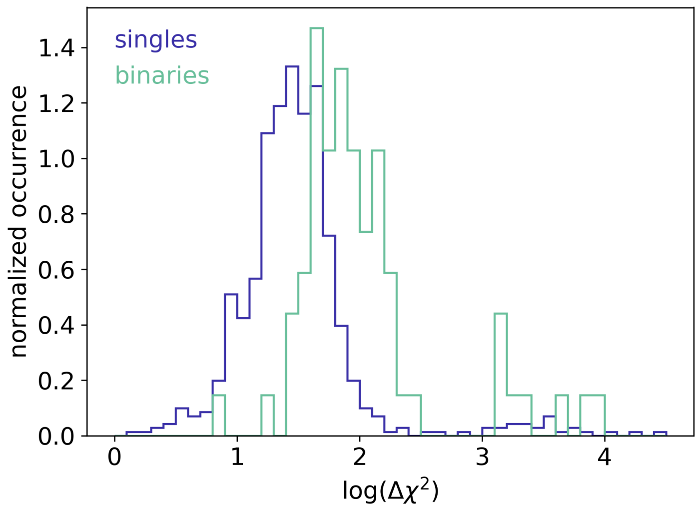

Figure 3.7: Distribution of $\Delta\chi^2$ for known single stars and binaries identified in EB18. Although most stars fall into the $\log\Delta\chi^2$ =1-3 region where our binary test is inconclusive, a $\log\Delta\chi^2 > 3$ is suggestive of binarity and a $\log\Delta\chi^2 < 1$ can rule out the possibility of a spectral binary.

stars in most cases- the $\log\Delta\chi^2$ =1-3 regime encapsulates both 85% of single stars and 87% of binaries. However, we find that a $\log\Delta\chi^2 > 3$ is suggestive of binarity. Indeed, while this region contains stars from both the single star and binary samples from EB18, the single star spectra in this region exhibit similar behavior to the binaries, in that the sign of the single star fit residuals is feature-dependent, and the magnitude of the residuals is significantly reduced for the binary fit (see Figure 1 of El-Badry et al., 2018a). Our validation results for abundance labels Section 3.5.1 indicate that this behavior can't be attributed to the model inferring incorrect abundances, thus we hypothesize that these stars are true binaries that were missed by EB18. We also find that the $\log\Delta\chi^2$ <1 region of this figure is largely populated by single stars, with only 1.5% contamination from known binaries. This suggests that our $\Delta\chi^2$ metric can be used in conjunction with other metrics that Gaia reports (RUWE, `non_single_star`) to rule out the possibility of binarity as opposed to definitively detecting them. Note that $\Delta\chi^2$ refers specifically to spectral binarity, since the system could be a wide binary with a resolved companion and appear as a single star in its spectrum. To summarize, while a our $\Delta\chi^2$ metric does not provide a definitive test for binarity for spectra with $\log\Delta\chi^2$ =1-3, a $\log\Delta\chi^2$ >3 and $\log\Delta\chi^2 < 1$ are suggestive of binary and single star scenarios, respectively. We include an example of a binary from EB18 with a $\log\Delta\chi^2$ >3 in Figure 3.6.

3.6.3 Evolved Stars

Post-main sequence stars ($\log g < 4$) will not be well-described by our model since it was only trained on stars with $\log g \geq 4$. In this case, the $\log\rho(l_n)$ will be low ($\lesssim -6$) since the model will need to extrapolate to explain behavior it was not trained on. However, since we limit the degree to which our model is allowed to extrapolate (Section 3.4.3), the extrapolation will not be enough for our model to provide a good fit and the χ^2 value should be large as well.

The fact that the model is unable to provide a good fit in these cases can also manifest in a large χ^2_{Ca} and $\Delta\chi^2$ for evolved stars, even if they are not active stars or binaries. We can distinguish evolved stars from active stars because active stars will be well-fit by our model everywhere except the Calcium lines, and should thus have a nominal χ^2. To this same end, we can distinguish between the evolved and binary cases based on their $\Delta\chi^2$ values and residuals as demonstrated in Figure 3.6. Binaries should posses larger $\Delta\chi^2$ values than evolved stars. Additionally the residuals of the single star fit to a binary possess characteristic positive and negative spikes in their residuals, while the evolved star shows residuals that are more smooth and typically negative.

3.6.4 Stars with Unreliable Labels

As discussed in Section 3.4.3, a low $\rho(l_n)$ is indicative of extrapolation beyond the training domain during the model's test step, in which case the inferred labels are unreliable. While we penalize models with $\log \rho(l_n) < -7$ (Equation 3.7), main sequence single stars typically have $\rho(l_n)$ well above this value. This is shown in the bottom right histogram in Figure 3.6, which shows the distribution of $\rho(l_n)$ values for main sequence single stars from EB18. As we can see, these stars typically have $\log \rho(l_n) \geq -5.5$. Thus, a $\log \rho(l_n) \lesssim -6$ is indicative of a star that deviates from a main sequence single star, and *The Cannon* output labels are unreliable in these cases.

There are a number of reasons that our model may predict unreliable labels:

- *Poor data quality.* If the spectrum is too noisy, the model may extrapolate to low $\rho(l_n)$ regions of the training domain to increase flexibility and over-fit to the data. This is typically the case for stars with Gaia RVS SNR<50 (see Section 3.5.2).

- *Unresolved Binaries.* As described in Section 4.5, the single star model that is used to infer stellar labels cannot adequately describe the spectrum of an unresolved spectral binary. As shown in El-Badry et al. (2018a), this can bias the inferred parameters of the primary. This can also manifest is a low $\rho(l_n)$, as is the case for the binary shown in Figure 3.6.

- *Evolved Stars.* Because our training set was composed solely of main sequence stars, accurately modeling an evolved star requires labels beyond the model's training domain. Since the training set contained zero information for these stars, the labels inferred from these spectra are not physical.

- *True labels are outside of the training domain.* Due to the finite size of our training set, there are bound to be main sequence single stars with labels that are beyond the training domain that do not satisfy any of the conditions listed above. For example, stars with $T_{\rm eff} < 4000\,$K, $T_{\rm eff} > 7000\,$K, or particularly large [Fe/H] or [α/Fe] abundances may not resemble the stars in our training set. By design, our model is not able to accurately predict labels for these stars.

In general, we caution that labels should not be trusted for any spectrum with a $\log\rho(l_n)$ < -6. For RVS spectra with SNR> 50, the metrics outlined in this section may also be unreliable since the model performance is compromised. However, for spectra with SNR< 50, a low $\rho(l_n)$ is correlated with metrics that can accurately flag binaries and evolved stars. In the binary case, the contamination of the spectral lines due to the secondary contribute to both a low $\rho(l_n)$ and a large $\Delta\chi^2$. In the evolved star case, the object should have a low $\rho(l_n)$ and large χ^2, and possibly a large χ^2 and $\chi^2_{\rm Ca}$ as well. Thus, while we can't trust the inferred labels in these cases, we can still trust the metrics and flag them as astrophysically anomalous. Note that this is not the case for active stars– $\chi^2_{\rm Ca}$ does not correlate with the $\rho(l_n)$ metric, since the labels are computed based on the fit everywhere except the lines where $\chi^2_{\rm Ca}$ is evaluated. As a result, our model can provide both accurate labels and metrics for active stars.

3.7 Conclusions

In this paper, we train a data-driven model on Gaia RVS spectra of main-sequence stars to predict stellar labels with precisions of 72 K in $T_{\rm eff}$, 0.09 dex in $\log g$, 0.06 dex in [Fe/H], 0.05 dex in [α/Fe] and 1.9 km/s in $v_{\rm broad}$, in agreement with SPOCS labels with roughly 20% and 40% improvement over the Gaia `GspPhot` and `GspSpec` catalogs, respectively for stars with SNR> 50. We also provide extended data tables with labels inferred by our model for stars from the Kepler Input Catalog (KIC) with SNR>50 with known Gaia DR3 source IDs identified using the Gaia-Kepler crossmatch database[10].

In addition to reporting stellar labels, our model also computes metrics to identify active stars, binaries, evolved stars, and cases where the model's inferred stellar parameters are unreliable. We compute these metrics for a sample of main sequence single stars from EB18

[10] https://gaia-kepler.fun/

to determine the nominal ranges of these metric values. Objects with metric values outside of these nominal ranges may be flagged as spectroscopic anomalies. We provide guidelines for interpreting the model metrics for these cases in Section 3.6.

We note that while our $\Delta\chi^2$ metric is able to recover some of the binaries identified in EB18, the majority of single stars and binaries in the sample lie in the $\log\Delta\chi^2 = 1 - 3$ regime where this metric is not conclusive. We interpret the difficulties identifying binaries in the RVS sample as a product of two factors. First, the low SNR for a large number of stars with published RVS spectra means that flux from the secondary star is likely lost to photon noise, even in cases where the SNR>50. Second, the fact that the RVS spectra are epoch-averaged over a three-year baseline presents a new obstacle for identifying binaries. In the event that the relative RV of the stars in the binary changes by $> 26\,\mathrm{km/s}$ over the observing baseline, the secondary features may be smeared out in the averaging step. This implies a complicated selection function that depends on the binary's mass ratio, orbital phase, and separation, and leaves fewer detectable binaries than a similar study with single-epoch spectra. However, it is likely that single-epoch spectra with SNR≥ 100 will be more amenable to identifying binaries via composite Cannon models like the one we develop here. The number of spectra that fit this criteria is large for planet hosts is expected to grow as Kepler and TESS planet hosts continue to be monitored with instruments like Keck-HIRES and the Keck Planet Finder.

Our model is publicly available at `https://github.com/isabelangelo/gaiaspec`. The model reports labels and metrics for stars with published Gaia RVS spectra, and provides accompanying plots to visualize and contextualize the best-fit spectra and their corresponding labels and metrics. Our model enables users to classify individual targets of interest, and is particularly useful for TESS planet hosts, many of which have published RVS spectra in Gaia DR3 with reasonable SNR. Beyond this, our code allows these methods to be readily applied to the ~10 million RVS spectra that are expected to be published in future Gaia data releases.

3.8 Acknowledgements

This work made use of the gaia-kepler.fun crossmatch database created by Megan Bedell. IA is supported by the Thacher Fellowship at UCLA. IA would also like to thank Adrian Price-Whelan, Dan Foreman-Mackey, Kareem El-Badry, Greg Gilbert, and Courtney Carter for their helpful insights on the analysis portion of this paper.

Software: astropy (Astropy Collaboration et al., 2013b, 2018b), numpy (Van Der Walt et al., 2011), matplotlib (Hunter, 2007), scipy (Jones et al., 2001), pandas (McKinney, 2010), thecannon (Casey et al., 2016)

3.9 Appendix 3: Model Coefficients

As described in Section 3.2, the trained model coefficients θ_j describe the model flux dependence on the stellar labels. The magnitude of these coefficient vectors is related to how strongly the model depends on a given label or combination of labels at a particular pixel. During the training step, *The Cannon* computes coefficient vectors for all linear and quadratic terms in our model, and we show the linear terms corresponding to each label here. We also show the model scatter, which was computed empirically as outlined in Section 3.4.2.

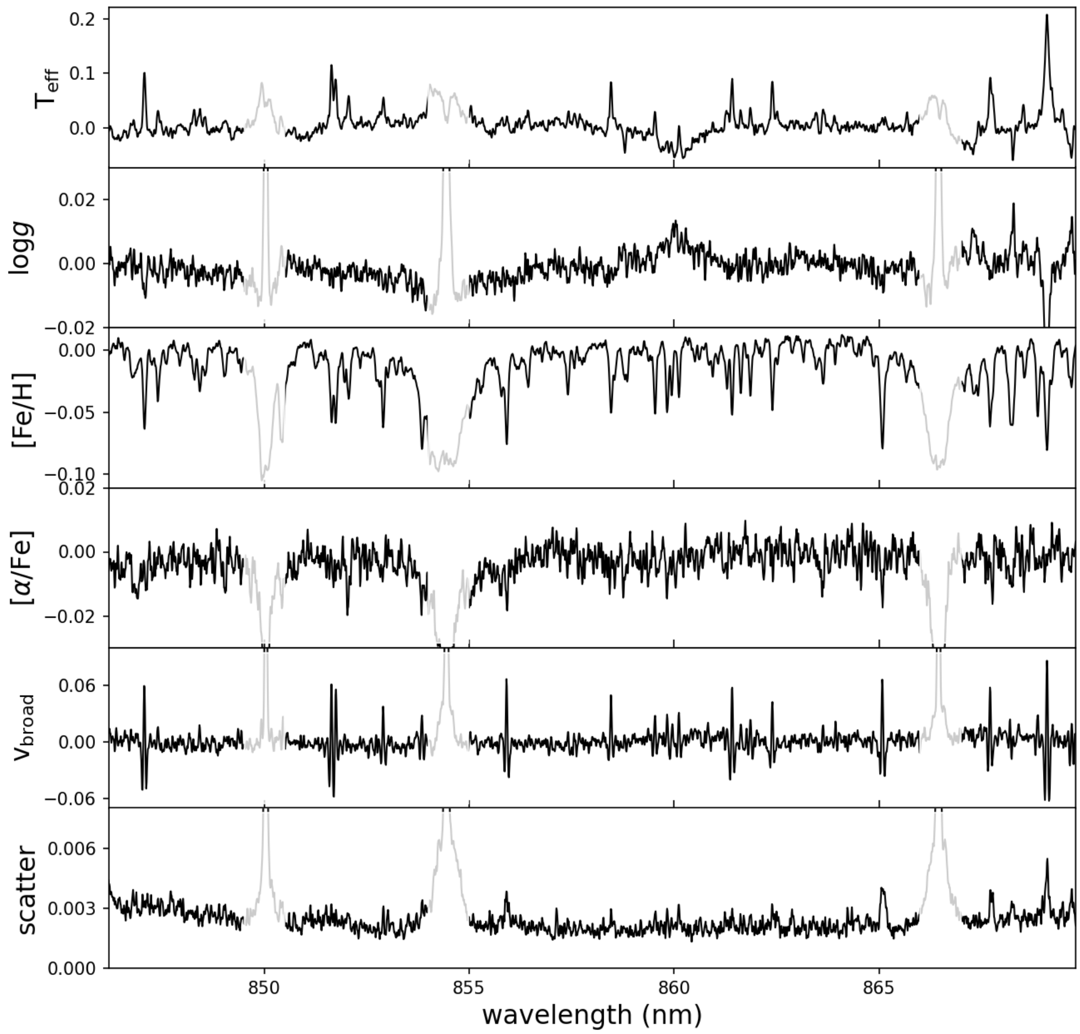

Figure 3.8: First-order coefficient vectors (i.e., spectral derivatives) of our trained model for each of our training labels (top 5 panels), and model scatter (bottom panel) as a function of the Gaia RVS wavelength scale. For each label, peaks in the coefficient vectors represent pixels in the spectrum that contain the most information for that particular label.

Chapter 4

Seeking Spectroscopic Binaries with Data-Driven Models

A version of this chapter was was submitted to *The Astrophysical Journal* on May 14, 2025.

Abstract

Data-driven stellar classification has a long and important history in astronomy, dating as far back as Annie Jump Cannon's "by eye" classifications of stars into spectral types still used today. In recent years, data-driven spectroscopy has proven to be an effective means of deriving stellar properties for large samples of stars, sidestepping issues with computational efficiency, incomplete line lists, and radiative transfer calculations associated with physical stellar models. A logical application of these algorithms is the detection of unresolved stellar binaries, which requires accurate spectroscopic models to resolve flux contributions from a fainter secondary star in the spectrum. Here we use *The Cannon* to train a data-driven model on spectra from the Keck High Resolution Echelle Spectrometer. We show that our model

is competitive with existing data-driven models in its ability to predict stellar properties $T_{\rm eff}$, $R_\star$, [Fe/H], $v \sin i$, and instrumental PSF, particularly when we apply a novel wavelet-based processing step to spectra before training. We find that even with accurate estimates of star properties, our model's ability to detect unresolved binaries is limited by its $\sim 3\%$ accuracy in per-pixel flux predictions, illuminating possible limitations of data-driven model applications.

4.1 Introduction

Across all subfields of astronomy, characterization of astronomical objects hinges on stellar spectroscopy. Spectroscopic observations of stars enable us to precisely determine their age, rotation, and chemical abundances, which play an important role in understanding the formation and evolution of objects as small as planets and as large as galaxies. In general, these properties are derived via comparison of stellar spectra to synthetic spectra derived from numerical approximations of stellar structure and radiative transfer (e.g., Kurucz (1979); Husser et al. (2013)). However, these models do not always resemble true observations of stellar spectra due to incomplete atomic line lists and numerous simplifications of the underlying physics needed to achieve sufficient computational efficiency.

More recently, data-driven models have successfully addressed many of these issues, particularly in the context of stellar label determination. Simple linear models of stellar flux as a function of stellar labels have shown to make consistent and accurate predictions of stellar properties like $T_{\rm eff}$, $\log g$, [Fe/H], and $v \sin i$ (e.g., *The Cannon*, Ness et al. (2015)). Additionally, more complex neural networks are also viable means of determining stellar properties from stellar spectra (e.g., *The Payne*, Ting et al. (2019)). These applications are particularly successful for large-scale surveys, where ample training data exist and high-fidelity labels of previous surveys can be leveraged to train the model. Indeed, data-driven models have been trained to predict stellar labels for spectra for a number of surveys, including APOGEE (Ness et al., 2015), LAMOST (Ho et al., 2017), and Gaia (Angelo et al., 2024).

Data-driven spectroscopic models also pose an exciting opportunity for science cases that require accurate predictions of the spectra themselves. For example, previous attempts to identify unresolved binaries in stellar spectra have historically been limited to close binaries (orbital separation $< 10\,\mathrm{au}$) with large radial velocity offsets that produce two distinct sets of spectral features. However, in recent years, accurate data-driven spectral models have been used to identify thousands of unresolved binaries with $> 10\,\mathrm{au}$ separations, which manifest as overlapping spectral features that produce subtle deviations from single star spectra (El-Badry et al., 2018a,b).

Applying these techniques to identify unresolved binaries in the planet host population is of particular interest. A large majority of the planets discovered to date come from space-based photometric missions like *Kepler*, *K2*, and *TESS*. However, these missions' large ≥4" pixels and considerable distances to most targets mean that many known planet hosts may have additional unresolved companions. Furthermore, follow-up studies to identify stellar companions to planet hosts have largely relied on ground-based adaptive optics imaging, which is only sensitive to companions beyond $> 50\,\mathrm{au}$ for most *Kepler* stars.

The California-Kepler Survey (CKS) provides an exciting opportunity to search for $< 50\,\mathrm{au}$ binaries in the planet census. The CKS survey is a *Kepler* follow-up program that obtained high-resolution spectra of over a thousand planet hosts using the High Resolution Echelle Spectrometer (HIRES; Vogt et al. 1994) instrument on Keck-I, providing a large sample of targets to search for binaries. Additionally, there is ample training data available of well-characterized stars and high-fidelity stellar labels. In fact, HIRES spectra have already been used to develop multiple data-driven spectroscopic models (Yee et al., 2017; Rice and Brewer, 2020) that accurately predict properties of stars with HIRES spectra. Finally, the high signal-to-noise of HIRES spectra offers the ability to detect subtle signatures binary that an unresolved secondary star imprints on the stellar spectrum.

In this paper, we explore the potential applications of data-driven models to identify unresolved binaries in the CKS sample. We perform simple experiments with synthetic HIRES spectra and show that unresolved binaries across a wide range of parameter space should produce significant markers of binarity in spectra with typical CKS measurement uncertainties. Next, we used *The Cannon*, a data-driven framework for modeling stellar spectra, to train a spectroscopic model of HIRES spectra. We also develop a novel wavelet-based filtering step to treat the stellar continuum and show that it significantly improves our model's ability to accurately predict stellar properties. Finally, we investigated our model's ability to identify binaries in a sample of previously identified binaries and single stars. We find that our data-driven model does not confidently resolve these two populations in our validation data, likely due to its limited accuracy in predicting the per-pixel flux of the single star spectra in our sample. We also comment on implications of our findings for future applications and limitations of data-driven spectroscopy.

4.2 Identifying Spectral Binaries in Simulated Data

We are particularly interested in identifying binaries with overlapping spectral features ("spectral binaries") that would have been missed by more traditional searches for SB2's with multiple sets of spectral features. We can identify these binaries as spectra that are significantly better fit by a composite spectrum–that is, a linear combination of two single star spectra– than by a single star spectrum (El-Badry et al., 2018a). In general, we expect to be sensitive to any spectral binary whose secondary star contributes more flux than the per-pixel flux uncertainties associated with the spectrum. However, a secondary star that is too similar in spectral type to its primary may not produce the unique markers of a non-single star at sufficiently high significance. To get a sense of our sensitivity to different types

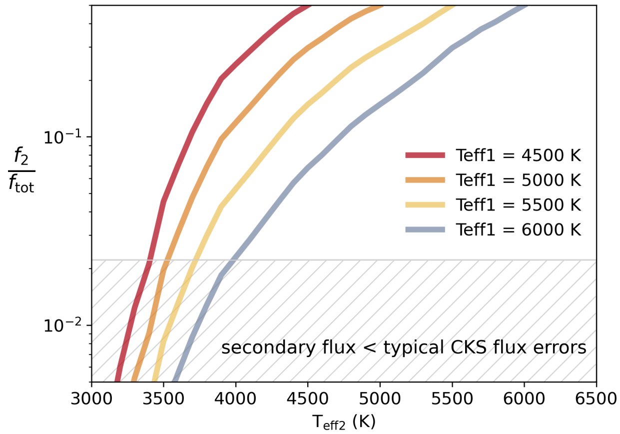

Figure 4.1: Flux contributions from the secondary star evaluated for binaries with a range of primary and secondary temperatures. For a HIRES spectrum with typical S/N=45 and 2.2% flux errors, secondary stars as cool as ∼3500 K should contribute to flux above the photon noise.

of binaries, we calculated the fractional flux contributions of the secondary star for a grid of binaries with varying primary and secondary T_{eff}:

$$\frac{f_2}{f_{tot}} = \frac{1}{1 + f_{rel}} \tag{4.1}$$

where f_{rel} is the flux ratio of the primary and secondary stars, $f_{rel} = f_1/f_2 = 10^{(m_2 - m_1)/2.5}$. Here, m_1 and m_2 are the V-band magnitudes of the primary and secondary star, respectively, which we estimate based on empirical T_{eff}-magnitude relations for main sequence stars published in Pecaut and Mamajek (2013b). Figure 4.1 shows f_2/f_{tot} as a function of secondary temperature T_{eff2} for a range of primary star temperatures T_{eff1}. We label $f_2/f_{tot} = 0.022$ for reference, since these are the flux errors associated with a typical CKS spectrum with S/N=45. We consider below this to be undetectable, since the secondary flux would be lost to photon noise. Based on this figure, we'd expect to be sensitive to secondary stars with $T_{eff2} \gtrsim 3500$ K for most a range of T_{eff1} in HIRES spectra with typical flux uncertainties.

We note that while this is a compelling proof-of-concept, this picture is complicated by two additional factors. First, binary spectra where the primary and secondary are spectrally similar (i.e., residing near the top of Figure 4.1 may not exhibit significant markers of binarity, compromising our sensitivity to more massive secondaries. Second, the relationship between the secondary and primary star might deviate from the simple temperature-flux relationship we assume for this calculation– in reality, the full spectrum contains pixels with different levels of information and photon noise that depend on both stellar properties and instrumental effects. To get a more accurate sense of the types of binaries we can detect, we developed a model that predicts the HIRES spectra of spectral binaries based synthetic, HIRES-like spectra. To do this, we used Starfish (Czekala et al., 2015) to simulate a grid of PHOENIX library spectra (Husser et al., 2013) for $\mathrm{T_{eff}}$ =3000-7000 K, convolved to the HIRES instrumental profile (FWHM=4.3 km/s) and re-sampled to the wavelengths of the HIRES middle detector CCD chip ($\lambda = 4985 - 6410$Å). We removed the continuum by dividing each individual spectrum order by a line fit to the continuum pixels (identified as pixels within 1% of the median flux) of that order.

To predict the HIRES flux of a single star, $\mathrm{f(T_{eff1})}$, we linearly interpolated between the two nearest grid spectra, weighted according to their distance from the $\mathrm{T_{eff}}$ we are modeling. We then modeled the HIRES flux of a spectral binary $\mathrm{f_{binary}(T_{eff1},\ T_{eff2})}$ as a linear combination of the primary and secondary flux:

$$f_{\rm binary} = W_1 \times f(T_{\rm eff1}) + W_2 \times f(T_{\rm eff2}). \tag{4.2}$$

where W1 and W2 are flux-weighted pre-factors calculated based on the flux ratio of the primary and secondary stars:

$$W_1 = \frac{f_1}{f_1 + f_2} = \frac{f_{\rm rel}}{f_{\rm rel} + 1} \tag{4.3}$$

$$W_2 = 1 - W_1, \tag{4.4}$$

We computed the best-fit single star and binary models for a simulated binary spectrum with `lmfit.minimize()` which uses linear least-squares to determine the $\mathrm{T_{eff}}$ and ($\mathrm{T_{eff1}}$, $\mathrm{T_{eff2}}$) that minimizes the χ^2 of the single star and binary model, respectively.

Since the binary model has more free parameters, we expect the χ^2 of the best-fit binary model to be consistently lower than that of the best-fit single star model. To account for this, we evaluated the extent to which a given simulated spectrum is better described by a binary model than a single star model based on the Bayesian information criterion (BIC) for the best-fit single star model and best-fit binary model:

$$\mathrm{BIC} = k \log N_{\mathrm{pixels}} - 2 \log \mathcal{L} \tag{4.5}$$

where k is the number of free parameters, N_{pixels} is the number of pixels in the spectrum, and $\ln\mathcal{L}$ is the Bayesian likelihood associated with the best-fit model (eq. 4.8). We then calculated the difference between the BIC of the best-fit single star binary models. A large ΔBIC indicates that the binary model is strongly favored over the single star model.

Figure 4.2 shows an example of a simulated binary spectrum with $\mathrm{T_{eff1}}$ =5900 K, $\mathrm{T_{eff2}}$ =4700 K and its best-fit single star and binary models. In the top panel we plot the individual primary and secondary star components, along with their composite spectral binary that we fit to. Below it, we plot the spectral binary degraded to S/N=500 along with the best-fit single star and binary models and their residuals to demonstrate our ability to distinguish between the single star and binary scenario in the limit of high S/N. The bottom panel shows the same thing, but the spectrum is degraded to S/N=45 to more closely resemble the CKS spectra. As we see from this figure, for a Sun-like primary and 4700 K secondary, the ΔBIC strongly favors the binary even in the lower S/N regime.

Next, we simulated $\mathrm{f_{binary}}(\mathrm{T_{eff1}}, \mathrm{T_{eff2}})$ for a range of $\mathrm{T_{eff1}}$ and $\mathrm{T_{eff2}}$ values and injected Gaussian noise into the spectra to degrade them to S/N=45. We then computed their best-fit

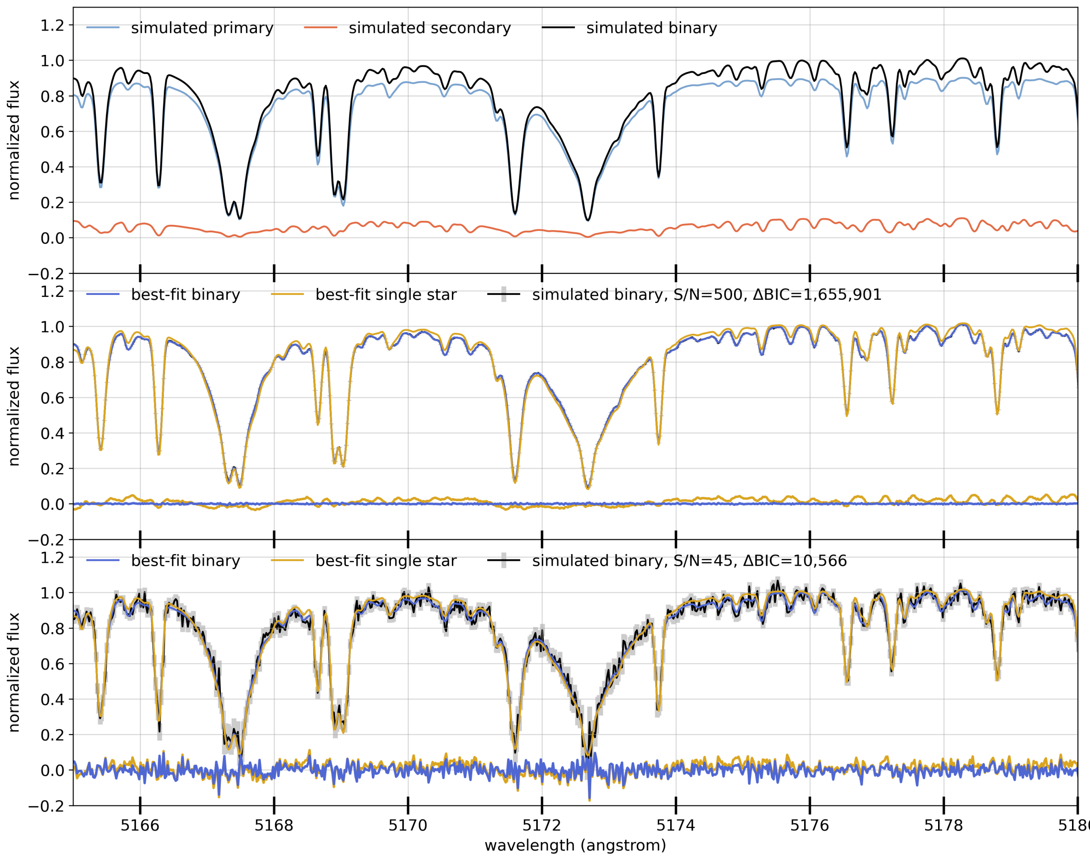

Figure 4.2: *Top:* Simulated spectrum of a 5900 K primary (blue) and 4700 K secondary (red) and their composite spectrum (black), weighted according to their relative flux contributions. *Middle:* Composite spectrum with added Gaussian noise and per-pixel S/N=500, with best-fit single star and binary models. The binary fit is strongly favored over the single star fit with $\Delta\mathrm{BIC} = 1.7 \times 10^6$. *Bottom:* Same as the middle panel, but with realistic HIRES S/N of 45. The binary fit is still favored over the single star, with $\Delta\mathrm{BIC} = 10^4$.

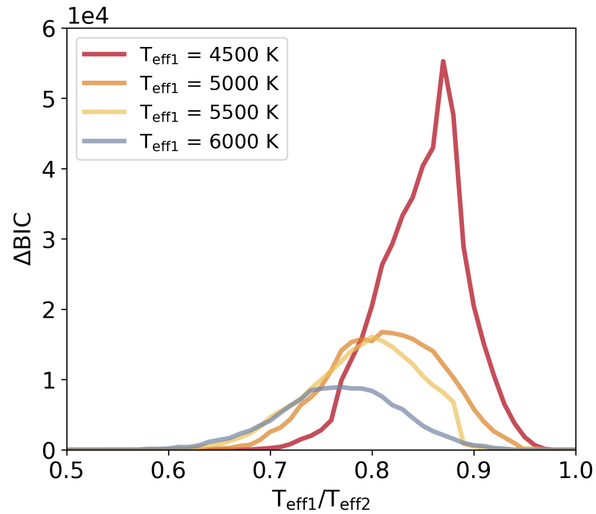

Figure 4.3: Difference in Bayesian Information Criterion (BIC) between the best-fit single star model and best-fit binary model is plotted versus secondary temperature for simulated binary spectra with a range of primary star temperatures. Simulated spectra include realistic CKS spectral resolution and S/N. The binary scenario is strongly favored over the single-star scenario for binaries with $T_{eff2}/T_{eff1} \approx$0.75-0.9, particularly for binaries with cooler primaries.

single star and binary models and their associated ΔBIC values. The ΔBIC values are plotted as a function of the simulated binary T_{eff2} for a number of different T_{eff1} values in Figure 4.3. Based on this figure, we expect to be sensitive to binaries with $T_{eff1}/T_{eff2} \sim$0.75-0.9, with increased sensitivity to binaries with cooler primaries where the secondary contributes a larger fraction of the total flux. This suggests that our binary detection methods will depend most sensitively on our model's ability to model cool stars with $T_{eff} \lesssim 4500\,K$. Our sensitivity drops off at low T_{eff1}/T_{eff2} where the secondary is too faint and high T_{eff1}/T_{eff2} where the normalized composite spectrum too closely resembles the spectrum of the primary alone.

4.3 A Data-driven Spectral Binary Emulator

Applying our simulated spectral binary model to real HIRES spectra poses a number of challenges. First, our model only varies T_{eff} of the primary and secondary stars while other important labels are fixed to log$g$=4.5 dex, [Fe/H] =0 dex, and $v \sin i$ =0 km/s. This is a simplification of the real spectra, which contain single stars and binaries spanning a wide range of stellar label space. Additionally, the simulation assumes that the model spectra used in the fits are a perfect representation of the real underlying spectra. In reality, the PHOENIX models do not agree with the per-pixel flux of real HIRES spectra of single stars to within our needed $\sim$ 2% accuracy. Thus, we developed a data-driven spectral binary emulator to use in the remainder of our analysis, as described in the the remainder of this section.

4.3.1 Training Data Selection

We trained our data-driven model on sample of high S/N HIRES spectra of well-characterized stars from the `SpecMatch-Emp` library. This library was published in Yee et al. (2017) and consists of 404 bright, well-characterized stars spanning a range of spectral types $\sim$ F1-M5. We selected this sample as our model training set due to its uniform coverage of stellar label space and high average per-pixel SNR$\sim$150. We then removed any reported unresolved binaries from our training set, since unresolved binaries will manifest as biased stellar parameters in our model training set (El-Badry et al., 2018a). To do this, we used the SIMBAD Astronomical Database to query the `SpecMatch-Emp` library and remove any unresolved binaries from our training set. In particular, we removed 22 stars flagged in SIMBAD as spectroscopic, eclipsing, or X-ray binaries, and 10 additional stars with reported stellar companions within the 0.8" HIRES slit width. We then divided our sample into hot and cool samples to train separate hot and cool star models, as described further in Section 4.4.

A number of `SpecMatch-Emp` library stars do not have reported $v \sin i$ values. We removed these stars from our hot star sample, but there were too many in the cool star sample to remove without compromising our model performance. In the cool star case, the $v \sin i$ values are typically not reported due to challenges in determining $v \sin i$ values for slow rotators ($\lesssim$2 km/s), thus we can assume the $v \sin i$ values for these stars $\leq$ 2 km/s. In order to preserve the information from these spectra and their associated labels, we included these spectra with assigned $v \sin i$ =0 km/s in our training set, along with "broadened" copies of each spectrum that we broadened to $v \sin i$ =3, 5 and 7 km/s with a rotational broadening kernel (Equation 17.12 in Gray 1992). Our final training set is composed of 135 unique hot stars, 140 unique cool stars, and 324 broadened cool stars. The distribution of training labels in our models is shown in Figure 4.4, and the spectroscopic properties of our training set are summarized in Table 4.1.

Table 4.1: HIRES spectra samples used in this analysis

Name	Source	Purpose	<S/N>	R	Number of Stars	Temperature Range
`SpecMatch-Emp`	Yee et al. (2017)	model training set	150	≥45,000	275	3056–6608 K
CKS	Petigura et al. (2022)	survey sample	≥20	≥60,000	1,332	3395–6538 K
Binary	Sullivan et al. (2024)	binary validation sample	≥20	≥60,000	97	3203–6838 K
Single Star	Raghavan et al. (2010)	single star validation sample	150	≥45,000	102	4879–6145 K

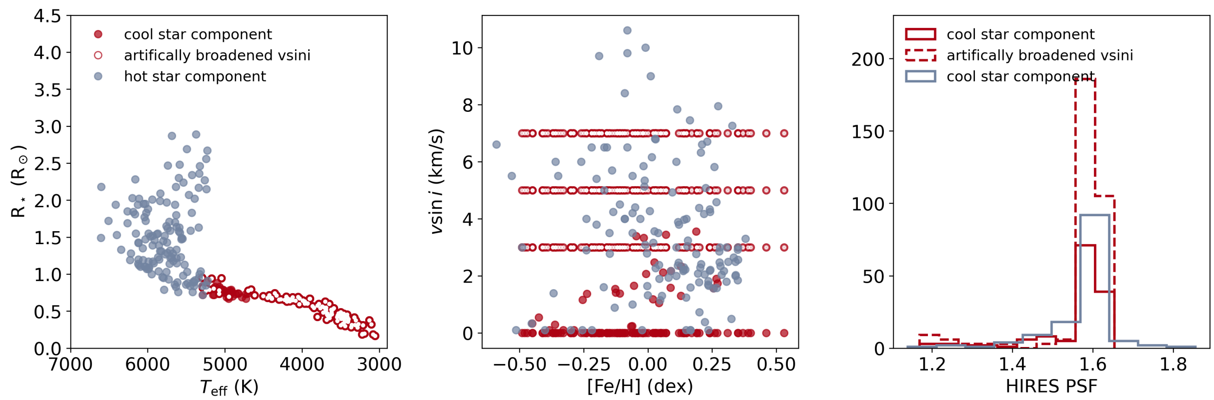

Figure 4.4: Distribution of `Specmatch-Emp` stellar labels $T_{\rm eff}$, $R_\star$, [Fe/H], $v \sin i$, and PSF values for both of our model training sets. Blue and red points indicate whether the star was part of the hot or cool components of our piecewise *Cannon* models, respectively. Stars that we included artificially broadened copies of in our cool star *Cannon* model are filled in with white in the left and middle panels and shown as a dashed histogram in the right panel. Stars with $R_\star < 2R_\odot$ and $5200 < T_{\rm eff} < 5300$ K are included in both training sets.

4.3.2 Data Preparation

The Cannon requires rest-frame, continuum normalized spectra as inputs for its training step. The HIRES spectra that are publicly available have undergone a preliminary step to remove their instrumental blaze function, as we describe below. We then used the `SpecMatch-Emp` code to shift the spectra into the rest-frame, and applied a second wavelet-based filtering treatment to remove residual low-frequency variations in the spectra, as detailed in the remainder of this section.

4.3.2.1 Blaze Function Removal, Part 1: Empirical Models

The HIRES spectra were processed according to standard practices of the California Planet Search (CPS; Howard et al. 2010b). This procedure removes large-frequency-scale flux variations due to the instrumental blaze function from each individual spectrum by dividing it by a composite spectrum of "spectrally flat" rapidly rotating B star spectra. However, we noticed that many of the HIRES spectra still contained low-frequency variations, especially for spectra with lower S/N. We suspect that this is due to either mismatched PSF or polarization between the target spectrum and the composite spectrum. In any case, this necessitated a second processing step to remove the residual blaze function, which we describe further in Section 4.3.2.3.

4.3.2.2 Spectral Registration

The spectra output by the CPS pipeline are shifted due to the stars' line-of-sight velocities, and need to be shifted into the rest-frame for the purposes of our analysis. We used the `SpecMatch-Emp` (Yee et al., 2017) to shift the spectra and register them back onto a common wavelength scale. Briefly, the spectra are shifted using a bootstrapping procedure that performs a cross-correlation of the target spectrum with a ladder of five rest-frame shifted reference spectra. The target spectrum is then shifted according to the reference spectrum that most closely resembles the target spectrum, identified as the spectrum with the largest peak in its cross-correlation function. Next, the spectra are registered onto a common wavelength scale, which we chose to be the standard CKS rest-frame wavelength solution (Petigura et al., 2017b). This produces a spectrum with 16 orders of 4021 pixels each. Finally, in order to produce a 1-D spectrum suitable for *The Cannon*, we clipped each order by 390 pixels on each side to ensure zero overlap between adjacent orders to produce a final spectrum of 16 individual orders of 3241 pixels.

4.3.2.3 Blaze Function Removal, Part 2: Wavelet Filtering

The processed HIRES observations are not strictly repeatable, and often contain non-astrophysical variations on different spectral scales that we wish to remove. To accomplish this, we used a wavelet-filtering process to decompose our signal into variations of different scales, allowing us to experiment with excluding variations that occur on larger scales than the typical spectral line (i.e., continuum-level variations) as well as variations comparable to or smaller than the typical spectra line. Our ultimate goal was to simultaneously minimize instrumental information and maximize astrophysical information in our spectra. This method has been used previously to remove systematics from transit signals (Stumpe et al., 2014; del Ser et al., 2018). In practice, it is similar in nature to a Fourier decomposition– the key difference is, instead of decomposing the spectrum into sine and cosine components, which describe oscillations that persist over all wavelengths, wavelet-filtering decompose the spectrum into wavelet components, which describe oscillations that are localized in wavelength space.

Figure 4.5 shows a visual representation of our wavelet filtering methods. To filter a single spectrum, we first perform a discrete, multi-level decomposition of each individual spectrum order using the `pywavelets` package. In the same way that a Fourier power spectrum describes the amplitude of sine wave with a specific frequency, wavelet transform coefficients describe the amplitude of a wavelet function with a specific frequency at a specific wavelength. Thus, the coefficients of a discrete wavelet transform correspond to the amplitude of a ladder of wavelets with discrete frequencies, evaluated as a function of wavelength. We then recombine the decomposed levels, leaving out the level associated with the lowest-frequency wavelet where we expect continuum information to be contained. We experimented with different wavelet modes and excluding different combinations of levels in our signal reconstruction. We chose to use a 'sym5' wavelet mode and leave out only the lowest-frequency "approximation" coefficients (variations on the order of $\sim 20\,\text{Å}$) because we found that this produced spectra with minimal differences across multiple exposures, thereby preserving only the star's time-invariant astrophysical signal. The output of our wavelet filtering process is a filtered signal that contains information from all but the lowest frequency modes of the original spectrum.

We emphasize that unlike traditional continuum treatment procedures, wavelet-filtering is not designed to produce flattened spectra normalized to unity. Instead, it produces signals

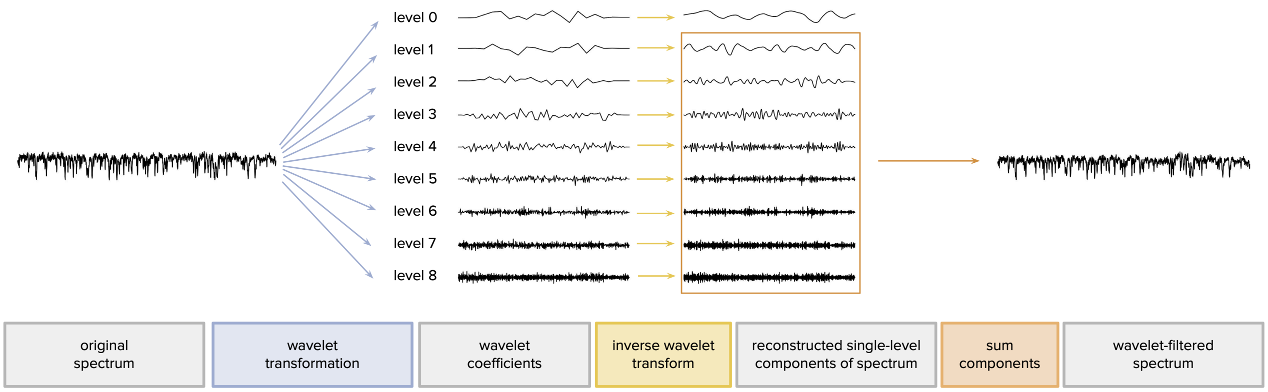

Figure 4.5: Schematic of our wavelet-filtering process to remove low-frequency, non-astrophysical variations in our HIRES spectra. The indiviual HIRES spectrum orders are decomposed into wavelet coefficients via discrete wavelet decomposition. We then reconstruct the signal via wavelet recombination (i.e., inverse wavelet decomposition) excluding the approximation coefficients associated with the lowest frequencies to produce a wavelet-filtered spectrum.

centered at zero, which do exhibit some low-frequency modulations. The key difference between the filtered and unfiltered spectra is that the modulations of the filtered spectra are homogeneous and do not differ from star-to-star, and thus do not provide any information about the star's physical properties or other systematics that might compromise the information our model learns during its training step. This is demonstrated in Figure 4.1, where we show two spectra of KOI-99 taken on different nights, where any differences should come from instrumental systematics. We plot the ratio of the spectra before and after wavelet-filtering, as well as the ratio of the filtered out wavelet components. As we can see from this figure, the ratio of the wavelet-filtered spectra is smaller than the ratio of the unfiltered spectra, indicating that the wavelet-filtered spectra vary less on a night-to-night bases compared to their unfiltered counterparts.

We proceed with our analysis using both spectra that are output by `SpecMatch-Emp` (hereafter "unfiltered" spectra), and spectra that gone through an extra step of wavelet-filtering and comparing our model performance for both cases.

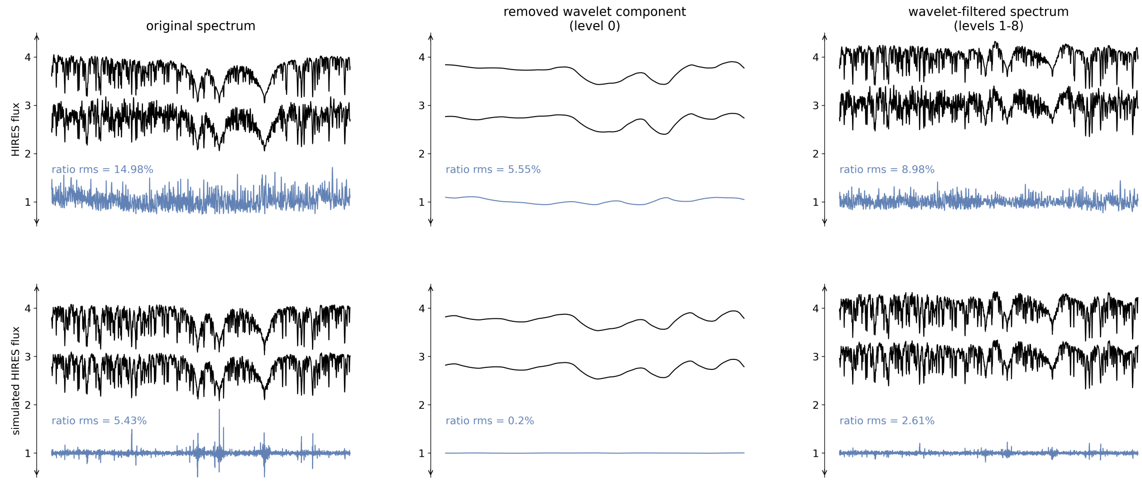

Figure 4.6: Demonstration of our wavelet-filtering process' ability to remove non-astrophysical night-to-night variations in HIRES spectra. The top panel shows snippets of two HIRES spectra of KOI-99 taken on different nights, where variations between the two spectra should be primarily due to instrumental effects. The bottom shows two simulated KOI-99-like spectra, where variations between the two spectra are due to injected noise only. While the wavelet-filtering does not produce flat spectra like traditional continuum treatments, it is effective at removing low-frequency night-to-night variations present in the HIRES spectra while simultaneously preserving astrophysical information in both real and simulated spectra.

4.3.2.4 Pixel Mask

The HIRES spectra are contaminated by telluric absorption features imprinted by molecules in the Earth's atmosphere. These features come from outside the target star reference frame; as a result, they are shifted out of the rest-frame when the rest of the spectrum is shifted into the rest frame and manifest at different pixels in the HIRES spectrum. We masked pixels prone to telluric contamination to both assure the quality of our training set spectra and avoid the model overfitting to telluric lines at the cost of producing a good fit elsewhere during its test step (see Section 4.3.3). We referenced a list of telluric lines [1], defined as the minimum and maximum wavelength of each telluric feature presented in the HIRES r chip bandpass. We added a 30 km/s padding to each wavelength range to account for shifting of telluric lines due to barycentric corrections, and another 100 km/s padding to account

[1]available at: https://www2.keck.hawaii.edu/inst/common/makeewww/Atmosphere/

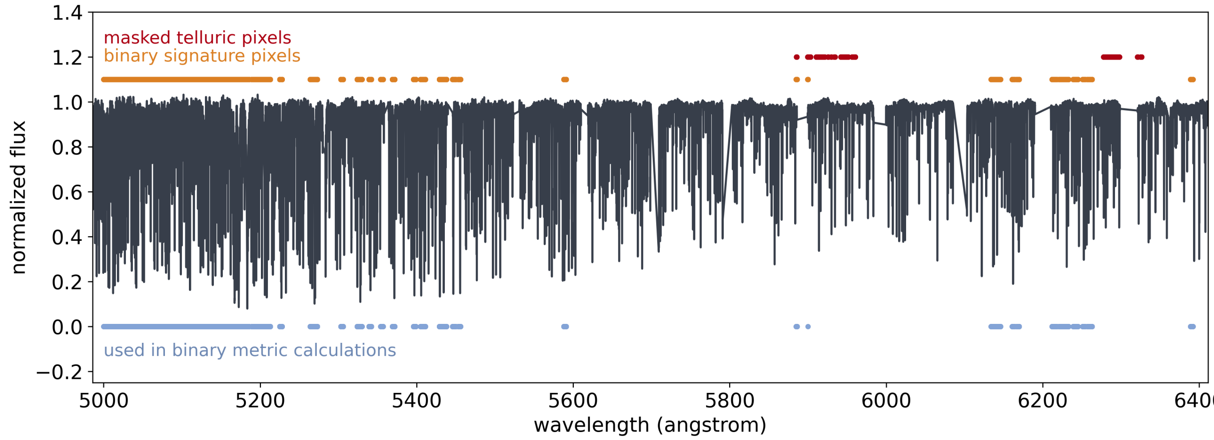

Figure 4.7: HIRES spectrum of sample star HD 106088. The pixels we mask due to telluric contamination are shown in red, and the binary signature pixels where we expect the strongest signatures of a secondary star are shown in orange. The binary signature pixels that are not contaminated by tellurics, shown in blue on the bottom, are the pixels we used to calculate the ΔBIC associated with binary detection (Section 4.5).

for shifting due to relative radial velocities of the stars before rest-frame shifting. Our final telluric mask is illustrated in Figure 4.7. Because *The Cannon* fits to each pixel individually in its training step, we don't mask telluric lines when training the model. Instead, we mask telluric lines during the test step by fitting to only the unmasked pixels in the spectrum, as described further in Section 4.3.3.

We expect our model to be more sensitive at specific features where the signature of the secondary is more prevalent. Along these lines, wavelengths where the single star and spectral binary spectrum are expected to be the same effectively add noise to our analysis when they are included in our ΔBIC calculation. Thus, we incorporate a second "binary signature" pixel mask using the methodology described below that includes only pixels expect to show the strongest signatures of the secondary star.

To determine which pixels to include in our binary signature pixel mask, we simulated a coarse grid of 20 binary spectra with T_{eff1} =4000-7000 K, T_{eff2} =3000-T_{eff1}. We calculated the difference between the binary spectrum and the primary spectrum for each binary to generate a grid of difference spectra. We then selected pixels in this grid whose average absolute difference or standard deviation were in the top 20th percentile of all pixels. These pixels (hereafter "binary signature pixels") represent the HIRES spectrum pixels that exhibit either the largest difference between the binary and single star spectra, or the largest scatter in difference between the binary and single star spectra. This binary signature pixel mask will be used in model comparison by computing the ΔBIC between best-fit single and binary star models using only the selected pixels.

4.3.3 Single Star Model

We used a data-driven algorithm called *The Cannon* (Ness et al., 2015) to train a generative model that predicts the per-pixel flux of single stars in HIRES as a function of four stellar labels l_n = ($T_{\rm eff}$, $R_\star$, [Fe/H], and $v \sin i$, PSF), where PSF is the instrumental point-spread function associated with each observation. While other models of HIRES spectra have included more training labels, it has been shown that $T_{\rm eff}$, $R_\star$, [Fe/H], and $v \sin i$ are sufficient for the purposes of binary detection (El-Badry et al., 2018a,b). We included the instrumental PSF as an additional training label to correct for differences in the PSF across different observations in our training set, which compromise our models' ability to accurately infer $v \sin i$ when left unaccounted for. Holding all other stellar labels constant, decreasing the resolution of a spectrum broadens the spectral lines, which is approximately similar to increasing $v \sin i$ for slowly rotating stars. Training our model on both $v \sin i$ and the observation-specific PSF allows it to learn how flux varies with both $v \sin i$ and PSF to produce a spectrum that is more accurately tethered to the star's physical $v \sin i$ value.

We used implementation of *The Cannon* developed by Casey et al. (2016)[2] to train our spectral model. A detailed description of these methods can be found in (Ness et al., 2015) and (Casey et al., 2016), but we provide a brief review here. In short, *The Cannon* assumes that the flux of star n at pixel j and the stellar labels l_n obey the following relationship:

[2]available at: https://github.com/andycasey/AnniesLasso

$$f_{nj} = \mathbf{v}(l_n) \cdot \theta_j + e_{nj} \tag{4.6}$$

where $\mathbf{v}(l_n)$ is a "vectorizing" function that describes how the flux is assumed to vary with the labels, and e_{nj} are the flux errors, which are sampled from a Gaussian with zero mean and variance equal to the instrumental uncertainty and intrinsic pixel scatter of the model summed in quadrature, $\sigma_{nj}^2 + s_j^2$. For our model, we assume the relationship between pixel flux and stellar labels is a second-order linear relationship with polynomial coefficient vector θ_j, where θ_j is a 21×1 matrix of label coefficients and their cross terms (T_{eff}, $R_\star$, $T_{\mathrm{eff}}{}^2$, $R_\star{}^2$, $T_{\mathrm{eff}} \cdot R_\star$, etc.). In other words, we model the flux as a polynomial function of the stellar labels, plus some deviation sampled from a Gaussian with zero mean and variance equal to the observation and model uncertainty summed in quadrature.

The Cannon uses training spectra during its "training step" to develop a spectral model and then uses this model to fit to individual spectra outside the training set during its "test step." The training step of *The Cannon* determines the polynomial coefficients and intrinsic model scatter (θ_j, s_j^2) that minimize the following function:

$$\theta_j, s_j^2 = \arg\min_{\theta_j, s_j^2} \sum_{n=1}^{N_{\mathrm{stars}}} -\ln p\,(y_{nj}|\theta_j, l_n, s_j^2), \tag{4.7}$$

where y_{nj} is the observed flux of training set star n at pixel j, and $\ln p$ describes the per pixel log-likelihood of the model flux:

$$\ln p\,(y_{nj}|\theta_j, l_n, s_j^2) = -\frac{[y_{nj} - \mathbf{v}(l_n) \cdot \theta_j]^2}{s_j^2 + \sigma_{nj}^2} - \ln\,(s_j^2 + \sigma_{nj}^2) \tag{4.8}$$

The final output of the training step is a set of polynomial coefficients θ_j and per-pixel intrinsic model scatter s_j^2 for each spectrum pixel that best describe how the flux of the training set stars depends on the stellar labels.

Once the model is trained and the coefficients are determined, we can use the model in the test step to estimate the labels of spectra outside the training step. This step typically varies only the stellar labels in l_n to determine the the best-fit *Cannon* model. However, we noticed during our analysis that many of the HIRES spectra in our samples show large residuals between the spectra and their best-fit single star models due to small errors in rest-frame shifting carried out by `SpecMatch-Emp` (see Section 4.3.2.2). Thus, we subsequently apply a relative radial velocity v_n to the *Cannon* model output to predict $g_{nj'}$, i.e. the flux corrected for imperfect rest-frame shifting during data processing:

$$g_{nj'}(l_{\text{single},n}) = \mathbf{D}(j', j, v_n) \cdot [v(l_n) \cdot \theta_j] \tag{4.9}$$

Here, $\mathbf{D}(j', j, v_n)$ is an operator that applies a Doppler shift by relative velocity v_n and interpolates *The Cannon* rest-frame flux $v(l_n) \cdot \theta_j$ from the j *Cannon* model grid to the j' HIRES pixel grid, and $l_{\text{single},n} = (l_n, v_n)$ describes the full set of six single model parameters that describe a particular star n. We note that while these shifting errors are present in the training data as well, we experimented with training a second model on velocity-corrected spectra and found that it does not substantially affect our model performance.

We estimate the best-fit model for a given spectrum by determining the model parameters in $l_{\text{single},n}$ that maximize the model likelihood function (or rather, minimize the negative likelihood function), evaluated at all non-telluric pixels (see Section 4.3.2.4 for a description of our telluric mask):

$$\hat{l}_{\text{single},n} = \arg \min_{l_{\text{single},n}} -\ln \mathcal{L}\left(y_n | l_{\text{single},n}\right) \tag{4.10}$$

where $\ln \mathcal{L}\left(y_{nj'} | \theta_{j'}, l_{\text{single},n}, s^2_{j'}\right)$ (hereafter $\ln\mathcal{L}$) represents the Bayesian likelihood associated with a particular set of parameters $l_{\text{single},n}$ for HIRES pixel j':[3]

$$-\ln\mathcal{L} = \frac{1}{2}\sum_{j'=1}^{N_{\text{pixels}}}\frac{[y_{nj'} - g_{nj'}(l_{\text{single},n})]^2}{s_{j'}^2 + \sigma_{nj'}^2} + \log 2\pi(s_{j'}^2 + \sigma_{nj'}^2) \tag{4.11}$$

To do this, we first conduct a brute search across the HR diagram, fixing all labels except $\mathrm{T_{eff}}$ and $\mathrm{R_\star}$ to determine the ballpark values of these labels for the star of interest. We then initialize an optimizer at these values and vary all six model parameters using the Nelder-Mead method in `scipy.optimize.minimize()`. We constrain the optimizer to only sample labels that are within the bounds of the training set (Figure 4.4), allowing v_n to vary from -20-20 km/s. We also re-parameterize $v\sin i$ to $\log(v\sin i)$ to avoid fitting $v\sin i < 0$.

The Cannon's quadratic approximation of f_{nj} has been shown to be sufficient for training sets spanning $\mathrm{T_{eff}}$ range of $\sim$ 2000 K, however, it has been suggested that broader applications may require increased model complexity (Ness et al., 2015). Since our training set spans a larger range of parameter space ($\mathrm{T_{eff}}$ =3000-7000 K), we split our training set into two separate training sets of hot ($\mathrm{T_{eff}}$ >5200 K) and cool ($\mathrm{T_{eff}}$ <5300 K) stars and trained two models that span smaller temperature ranges over which we'd expect the quadratic approximation to be more accurate. We allowed for 100 K overlap for the two components, but removed stars with $\mathrm{R_\star}$ >2 stars in the overlapping region from the cool star sample to avoid a training set with large gaps in the training set parameter space (see Figure 4.4). To carry out *The Cannon*'s test step, we fit each spectrum to both the cool and hot star models and select the fit with the larger $\ln\mathcal{L}$ (eq. 4.11), effectively fitting a piecewise *Cannon* model to each spectrum.

We note that another way to achieve increased model complexity is by training a neural network that assumes a more flexible relationship between the stellar flux and labels of the

[3]For our optimization we approximate $\theta_{nj'}$ and $s_{j'}^2$ as θ_{nj} and s_j^2. We recognize that this approximation is not technically correct, however, we find that this is a workable approximation in the low Doppler shift regime we are modeling.

training set (e.g., *The Payne*; Ting et al. 2019, and *The Data-Driven Payne*; Xiang et al. 2019). We experimented with using *The Payne* to train a neural network but found that for our training set, the models were prone to overfitting to training set spectra at the cost of providing a poor fit to spectra outside the training set. A more in-depth exploration of *The Payne* and its applications to HIRES spectra is beyond the scope of this paper.

4.3.4 Binary Model

We used our piecewise *Cannon* model to construct a binary model that predicts the flux of spectral binaries as a function of the labels associated with the primary and secondary stars. We model the spectral binary flux as a linear combination of the Doppler-corrected primary and secondary star flux, weighted according to their relative fluxes:

$$g_{\mathrm{binary},j'} = W_1 \times g_{1j'}(l_{\mathrm{single},1}) + W_2 \times g_{2j'}(l_{\mathrm{single},2}). \tag{4.12}$$

where W_1 and W_2 are estimated from the same empirical magnitude-temperature relations described in Section 4.2 and $l_{\mathrm{single},1}$ and $l_{\mathrm{single},2}$ are the respective label vectors of the primary and secondary star. While we are primarily interested in identifying spectral binaries with <10 km/s velocity offsets, we allow for the individual velocities of the primary and secondary (v_1 and v_2) to float to account for both imperfect rest-frame shifting of the spectrum and 1-10km s velocity offsets between the primary and secondary that would be missed by previous searches for SB2 signatures in CKS spectra (i.e., Kolbl et al. 2015).

We vary the following labels and velocities of the primary and secondary stars:

$$l_{\mathrm{binary}} = (\mathrm{T}_{\mathrm{eff}1}, \mathrm{R}_{\star 1}, [\mathrm{Fe/H}]_1, v\sin i_1, v_1, \mathrm{PSF}_1, \mathrm{T}_{\mathrm{eff}2}, \mathrm{R}_{\star 2}, v\sin i_2, v_2) \tag{4.13}$$

and assume that the metallicity [Fe/H] and the instrumental PSF are the same for both the primary and secondary star to derive $l_{\text{single},1}$ and $l_{\text{single},2}$ and compute our binary model prediction $g_{\text{binary},j'}$. We determine the parameters $l_{\text{binary},n}$ associated with the best-fit binary model for a particular star n using a similar optimization procedure to our single star optimization:

$$\hat{l}_{\text{binary},n} = \arg \min_{l_{\text{binary}}} -\ln \mathcal{L}_{\text{binary}} \left(y_n | l_{\text{binary},n} \right) \tag{4.14}$$

where $\ln \mathcal{L}_{\text{bin}}$ represents the Bayesian log-Likelihood associated with a particular set of binary model parameters:

$$-\ln \mathcal{L}_{\text{binary}} = \frac{1}{2} \sum_{j'=1}^{N_{\text{pixels}}} \frac{[y_{nj'} - g_{\text{binary,j'}}(l_{\text{binary},n})]^2}{s_{\text{binary},j'} + \sigma^2_{nj'}} + \log 2\pi (s_{\text{binary},j'} + \sigma^2_{nj'}) \tag{4.15}$$

and $s_{\text{binary},j'}$ is the intrinsic scatter of the binary model, which we estimate as a linear combination of the intrinsic scatter of the piecewise Cannon components used to model the primary and secondary stars ($s^2_{1j'}$ and $s^2_{2j'}$, respectively), weighted according to their relative flux contributions:

$$s_{\text{binary},j'} = W_1 \cdot s^2_{1j'} + W_2 \cdot s^2_{2j'} \tag{4.16}$$

We made two further modifications to our optimization procedure (Section 4.3.3) to adapt it for the binary scenario. First, we perform the optimization in three separate regimes of our piecewise Cannon *model*: (1) the primary and secondary are modeled by the hot piecewise component, (2) the primary and secondary are modeled by the cool piecewise component, and (3) the primary is modeled by the hot piecewise component and the secondary is modeled by the cool piecewise component. As with the single star optimization, we select the best-fit binary model as the model associated with the highest $\ln\mathcal{L}$ across all three regimes.

The second modification we make is to the initial brute optimization. To determine the best-fit binary in a particular regime of our piecewise *Cannon* model, we first conduct a brute search varying only $T_{\rm eff1}$ and $T_{\rm eff2}$ within the bounds of their respective piecewise component training sets. We then initialize an optimizer at these values and vary all 10 model parameters ($T_{\rm eff}$, log$g$, [Fe/H], [α/H], $v_{\rm broad}$ of the primary, $T_{\rm eff}$, $R_\star$, $v \sin i$ of the secondary, v_1, v_2) using the Nelder-Mead method in `scipy.optimize.minimize()`. Similarly to the single star optimization, the labels of the primary and secondary are constrained to within the bounds of their respective piecewise *Cannon* model components.

4.4 Validation of Label Transfer

Historically, label transfer has been a prominent application of data-driven spectroscopic models (e.g., Ness et al., 2015; Ho et al., 2017; Yee et al., 2017; Ting et al., 2019; Rice and Brewer, 2020). Briefly, label transfer leverages the fact that for a given survey, a subsample of stars typically have derived labels with higher fidelity from a previous survey. The process of label transfer involves first learning the relationship between flux and "ground truth" stellar labels for the subset of well-characterized stars (i.e., the model training set) and then using this model to infer stellar labels for a larger population.

To assess our model's ability to infer stellar labels, we performed a leave-20%-out validation of our piecewise *Cannon* model. For each 20% subset of the full training set, we trained a new model on the remaining 80% and used this model to compute labels of the held out subset to obtain Cannon-inferred labels for all of the stars in our training set. We do this for two models, one trained on unfiltered training spectra and one on wavelet-filtered training spectra (see Section 4.3.2.3). The left column of Figure 4.8 shows *The Cannon* output labels plotted versus `SpecMatch-Emp`-reported labels, which we treat as ground truth. The rms scatter around the one-to-one line represents our estimated 1σ label uncertainties for our model. As we can see from the Figure, wavelet filtering significantly improves our model's ability to predict labels that agree with the ground truth labels. The model trained on unfiltered spectra agree with ground truth labels to within 87 K in T_{eff}, 0.17 $R_\odot$ in $R_\star$, 0.11 dex in [Fe/H], 1.04 km/s in $v \sin i$, and 0.07 in instrumental PSF. On the other hand, the model trained on wavelet-filtered spectra show agreement of 71 K in T_{eff}, 0.1 $R_\odot$ in $R_\star$, 0.09 dex in [Fe/H], 0.91 km/s in $v \sin i$, and 0.05 in instrumental PSF with ground truth spectra. These precisions are comparable to model precisions of existing data-driven models of HIRES spectra (Yee et al., 2017; Rice and Brewer, 2020), and demonstrates the significant improvement in model performance when we apply wavelet-filtering.

The improvement from wavelet filtering is even more significant for spectra with more moderate signal-to-noise. We used our piecewise *Cannon* models to compute labels of the S/N∼45 stars in our CKS survey sample (described in further detail in Section 4.5.1). The right column of Figure 4.8 shows the unfiltered Cannon output labels inferred from unfiltered CKS spectra, plotted versus "ground truth" CKS labels reported in Petigura et al. (2022) in yellow. The wavelet-filtered Cannon output labels inferred from wavelet-filtered CKS spectra are plotted in blue.

We see from Figure 4.8 that the improved agreement with ground truth labels is even more stark for these spectra than for the training spectra. We also see that the unfiltered *Cannon* model produces unreliable [Fe/H] labels for a number of CKS spectra, whereas the filtered model doesn't have this issue. We conclude from these results that our wavelet-filtering process has the potential to drastically improve the quality of labels derived from data-driven models for spectra with S/N∼45. This is likely due to the shorter average exposure times for low S/N observations, which is correlated with larger blaze function variability that is more difficult to remove with preliminary removal methods detailed in Section 4.3.2.1.

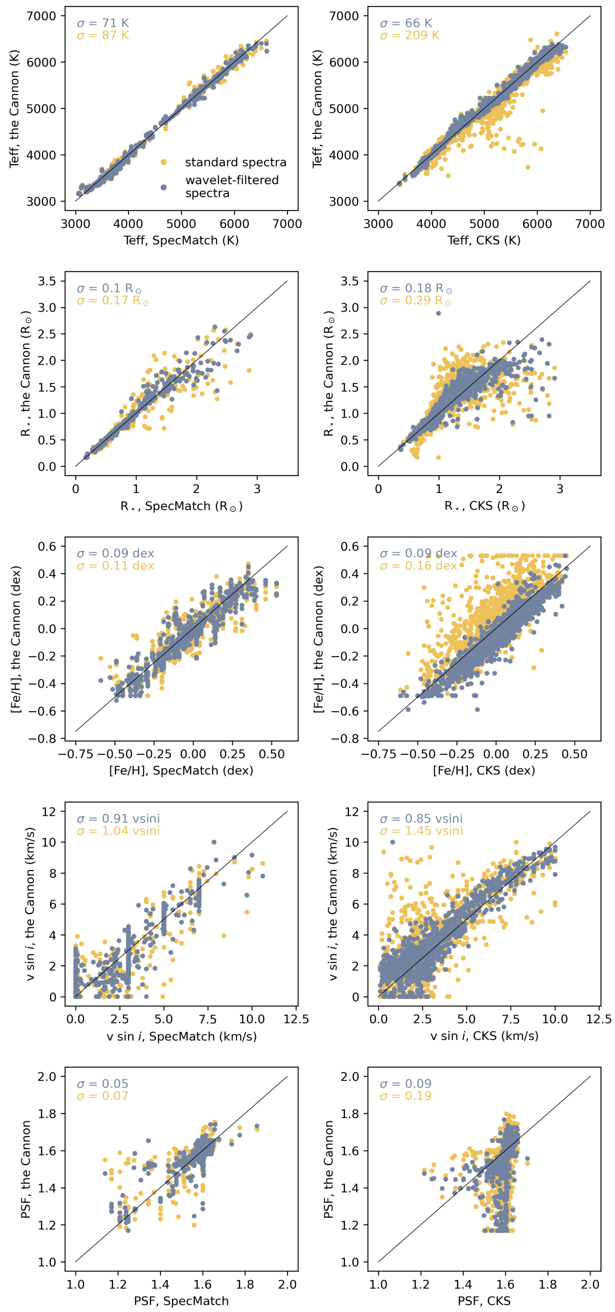

Figure 4.8: *Top:* Labels of our `SpecMatch-Emp` training set spectra (S/N≈150) were computed by our *Cannon* model using leave-20%-out validation for both original and wavelet-filtered models. *Bottom:* Demonstration of model label transfer performance on moderate signal-to-noise CKS spectra (S/N$\gtrsim$45). Labels of our CKS samples were computed by our *Cannon* models for CKS spectra using both unfiltered (blue) and wavelet-filtered (yellow) spectra. Wavelet-filtering significantly improves our *Cannon* model agreement with ground truth labels in both cases, but the improvement is more drastic at moderate S/N.

4.5 Sensitivity to Binaries

We have demonstrated that our model is effective determining stellar labels of stars based on their HIRES spectra. In this section, we explore our model's ability to detect signatures of unresolved binaries in these same spectra. As described below, this application of data-driven models is more demanding and requires more accurate per-pixel flux estimations of stellar spectra than tested and proved label transfer applications.

4.5.1 Survey Data

Our ultimate goal is to use our data-driven spectral binary emulator to distinguish between HIRES spectra of single stars and spectral binaries from the California-Kepler Survey (CKS) sample of known planet hosts. This sample includes HIRES spectra of 1305 planet hosts with surface temperatures of 4700-6500 K (CKS DR1, Petigura et al., 2017b), as well as 411 spectra of cooler planet hosts with $T_{\rm eff}$ =3500-5000 K (CKS DR2, Petigura et al., 2022) for a total of 1712 stars. These spectra have typical per-pixel S/N of 45 and spectral resolution of $R \geq$60,000, however, the exact S/N and spectral resolution vary for different objects.

Previous studies have identified unresolved binaries in this sample– that is, targets with companions with separations within the 0.8" HIRES slit width. In particular, 22 stars were flagged as spectroscopic binaries in Petigura et al. (2022). Another 132 were listed as binaries identified by *Kepler* follow-up imaging programs and further characterized in Sullivan et al. (2024), as detailed in the following section. Of the remaining targets, 59 were reported in a KOI binary catalog that combined results from Keck-II/NIRC2 imaging, UKIRT Hemisphere Survey photometry, and a patchwork of existing literature. This catalog will be described further in Kraus et al. (in preparation), and many are already published in Kraus et al. (2016). We removed these binaries from our survey sample, along with 143 stars with reported labels outside our model training set range, and reserved the remaining 1,332 stars our analysis, including our label transfer performance assessment in Section 4.4.

4.5.2 Validation Data

We reserved the aforementioned binaries in our CKS sample that appeared in the binary catalog published in Sullivan et al. (2024) for the purposes of validating our model's ability to identify spectral binaries. This catalog was assembled from a combination of KOI follow-up imaging programs from Furlan et al. 2017b, Ziegler et al. 2018 and supplemental AO imaging. The targets were then observed with the LRS2 instrument at McDonald to obtain revised star properties (T_{eff}, $R_\star$) for both the primary and secondary star. Of the 132 CKS targets that appeared in the complete binary catalog, 97 have companions that fall within the HIRES slit (angular separation≤0.8"), with 74 companions within 0.4" where we'd conservatively expect 100% of the secondary flux to fall within the slit. We compiled all binaries with ≤ 0.8" separations to to assemble our binary validation set.

We also composed a sample of single stars to validate our binary model. We selected this sample from a survey carried out in Raghavan et al. (2010). This study surveyed a sample of nearby stars with multiple complementary techniques that yielded near 100% completeness to stellar binaries at all separations. To construct our single star sample, we queried the 173 stars in this survey where no companion was detected to compile a subsample of 153 single stars with available HIRES spectra. We then removed stars with stellar labels outside the bounds of our training set model based on previously derived stellar properties from the Spectroscopic Properties of Cool Stars (SPOCS) catalog (Brewer et al., 2016). Our final validation sample consists of 102 single stars.

The Raghavan et al. (2010) survey targeted stars within 25 pc and is, by construction, much brighter than the *Kepler* field stars in our binary sample. Our single star and binary samples have large signal-to-noise discrepancies as a result, with average per-pixel S/N of 150 and 45 in each respective sample. Despite the ΔBIC 's incorporation of flux errors into its goodness-of-fit estimation, we found that S/N was the best predictor of a particular spectrum's ΔBIC when comparing our single star and binary validation samples. Thus, in order to obtain a validation sample that lends itself to a one-to-one comparison, we degraded each single star spectrum S/N by adding gaussian noise with standard deviation $\sigma_{\text{added}} = \sqrt{\sigma^2_{\text{binary}} - \sigma^2_{\text{single}}}$, where σ_{binary} are the flux errors of a randomly selected star from our binary validation sample. Our final validation sample consists of 102 single stars and 97 unresolved binaries with per-pixel S/N~ 45.

4.5.3 Performance on Validation Data

To assess our model's ability to distinguish between single stars and binaries, we calculated the ΔBIC for all of the stars in our single star and binary validation samples. We computed the best-fit single star and binary model for each spectrum masking only telluric pixels (Section 4.3.2.4). We report two ΔBIC values, one that incorporates all non-telluric pixels, and one that incorporates only pixels in our binary signature pixel mask. The ΔBIC distributions of our single star and binary validation samples for both cases are shown in Figure 4.9. As we can see from this Figure, the main advantage of using our binary signature pixel mask is that it reduces the spread of ΔBIC values of our single star spectra, both below zero and in the high ΔBIC tail of the distribution. This is in line with what we expect- the mask removes pixels with low information content, effectively removing noise in our analysis. Beyond this, we don't see much improvement in our model performance with or without this mask, and the ΔBIC doesn't successfully resolve the single star and binary samples in either case.

Figure 4.10 shows same ΔBIC of our binary validation sample plotted as a function of binary

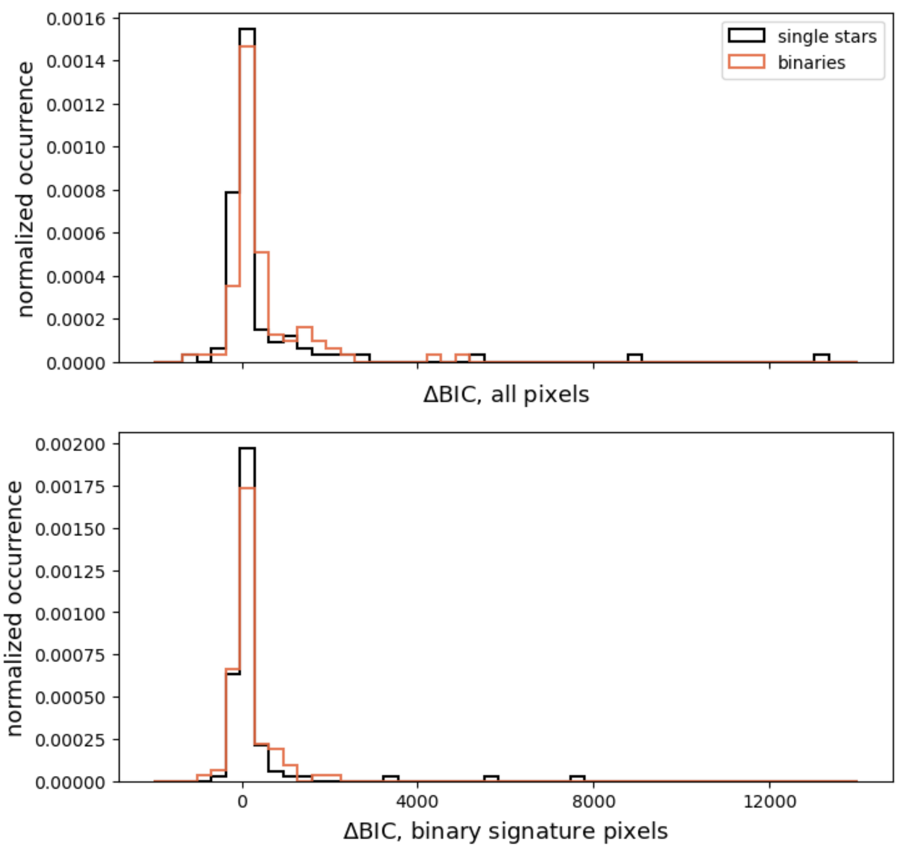

Figure 4.9: ΔBIC distributions of stars in our single star and binary validation samples. The top panel shows the ΔBIC values evaluated at all non-telluric HIRES specrum pixels, and the bottom panel shows the ΔBIC values evaluated only at our binary signature pixels where we'd expect the signature of the secondary star to be the strongest (see Section 4.2). In both cases, the ΔBIC doesn't meaningfully distinguish between the single star and binary samples.

separation and *Kepler* magnitude difference Δm_K as reported in Sullivan et al. (2024). While many of the binaries in our sample have comparable ΔBIC values to the single stars, binaries with the highest ΔBIC lie in regions where we'd expect to be most sensitive to binaries– that is, binaries with closer separations and intermediate Δm_K values. We expect binaries with closer ($\lesssim 0.4$") separations to show stronger signatures of binarity because (1) their larger RV offset which can alter the spectral line shape even when the features overlap to within one $\sim 10\,\mathrm{km/s}$ resolution element, and (2) the secondary flux is more likely to be entirely captured by the HIRES slit. Furthermore, binaries with intermediate Δm_K (i.e., intermediate T_{eff} ratios) will differ the most from similar single star spectra based on our simulated data analysis (Section 4.2).

Figure 4.10 also sheds light on our model's limited completeness. Several binaries with companions within 0.4" and intermediate Δm_K values have ΔBIC <300, indistinguishable from single stars. Furthermore, the fact that many of our single stars have comparable ΔBIC to the high ΔBIC binaries in our validation sample suggests that even single star spectra benefit significantly from increased complexity of our binary model. In short, our model's limited completeness and high false positive rate contribute to its inability to identify binaries in our validation sample.

Figure 4.11 shows the best-fit single star and binary models for two test cases– KOI-1002, a binary with a 5863 K primary and 4714 K secondary that we expect to detect based on ab initio models, and HD 78366, a single star with similar T_{eff} =5984 K to the primary star of KOI-1002. In both cases, the residuals for the single star and binary model fits are comparable with no significant improvement from the binary model. Additionally, the best-fit single star residuals have an amplitude of ≈2-5%, even for the single star where we'd expect the model to agree with the spectrum to within the measurement uncertainties. We interpret this as evidence that our ability to distinguish between single stars and binaries is limited by our model predictions of single star spectra. If our binary model is only as

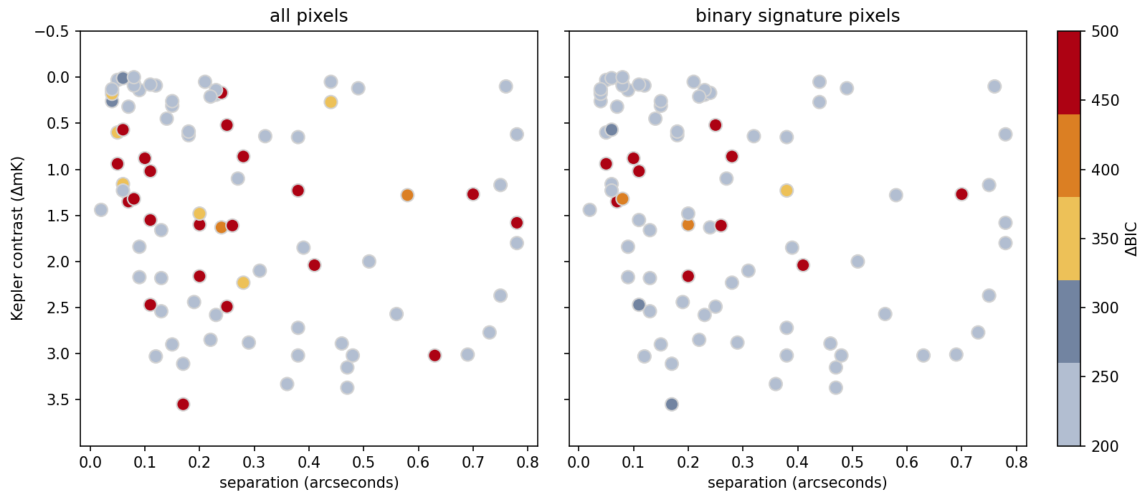

Figure 4.10: Magnitude difference Δm_K versus binary separation for all of the binaries in our validation sample, colored according to ΔBIC. While our model is unable to distinguish between the single stars and the binaries in most cases, the binaries with the highest ΔBIC values reside in regions of parameter space where we expect to be most sensitive to binaries.

accurate as its single star components, then we can only expect to model binary spectra accurately to within 2-5% of the true binary flux. This is larger than the accuracy we need to detect binaries in most cases, as seen in Section 4.2. Thus, we suspect that our models per-pixel flux predictions are not currently accurate enough to detect binaries.

4.6 Discussion

In this paper, we trained a data-driven model of HIRES spectra using the `Specmatch-Emp` stellar library as our training data. We demonstrate that our model is competitive with existing models (Yee et al., 2017; Rice and Brewer, 2020) at deriving stellar labels that agree with high-fidelity labels of previous surveys. We also develop a prescriptive, repeatable wavelet-filtering method for removing low-frequency, non-astrophysical variations in our spectra, which mitigates night-to-night variations in stellar spectra due to systematics. This processing step significantly improves the accuracy of our model's stellar label predictions,

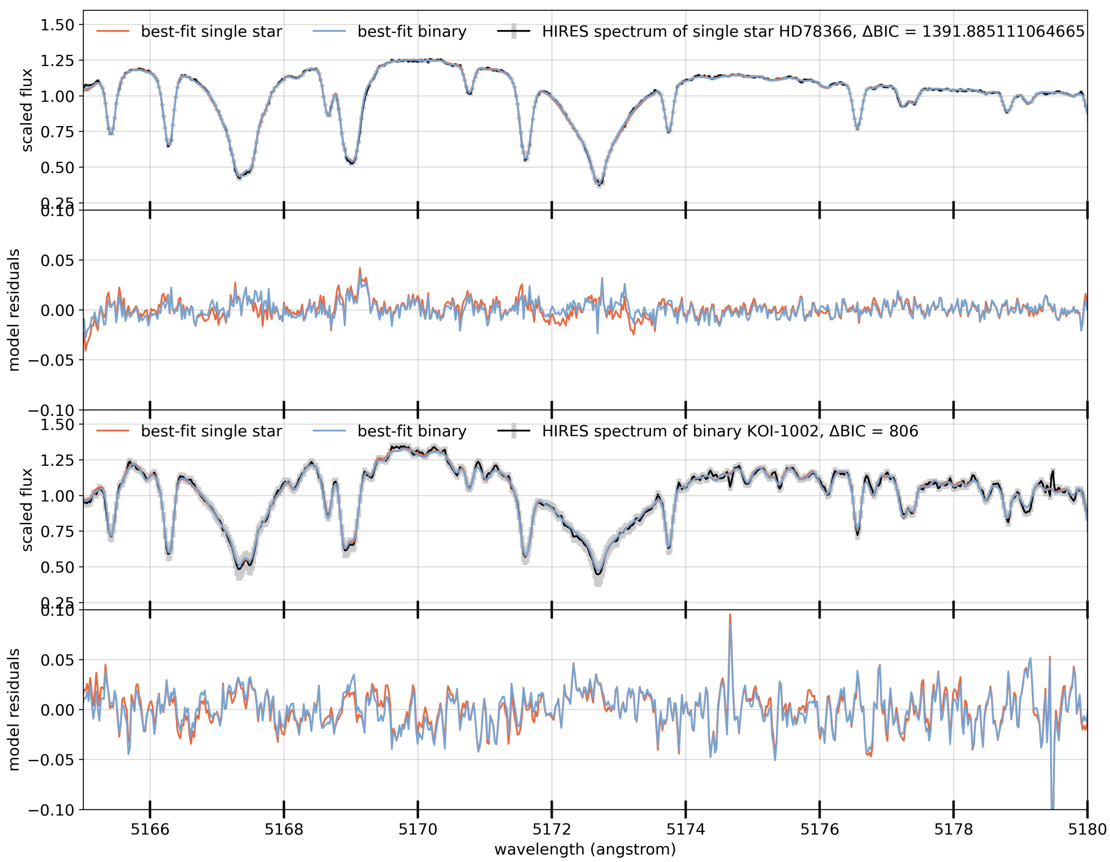

Figure 4.11: HIRES spectra of single star HD 78366 and binary KOI-1002, with their best-fit single star and binary spectra. We see that in both cases, both the single star and binary show ≈2-5% disagreement with the HIRES spectra, limiting our ability to detect signatures spectral binaries at sufficiently high significance.

particularly for low signal-to-noise spectra where traditional continuum treatments are less effective.

We used our data-driven model of single stars to develop a data-driven emulator of spectral binaries. Experiments with synthetic HIRES spectra in this analysis suggest that accurate models of HIRES spectra can discern between spectra of single stars and spectra with signatures of stars with unresolved stellar companions. We find that while our single star model is able to accurately determine stellar labels, its predictions of per-pixel HIRES flux are around $\sim 2\%$. This limits the degree to which our spectral binary emulator can model spectra of binaries in our validation sample, which, consequently, limits our model's ability to identify subtle ($\sim 2\%$) deviations from single star behavior that our binary spectra should exhibit.

One of our primary conclusions from our analysis is that accurate emulation of stellar spectra, even in the instance of accurate classification of stellar labels, is challenging. We see that our model is excellent at label transfer, but falls short of emulation. We suspect that this is because data-driven models may under- and over-estimate the true flux at different pixels; these discrepancies can balance one another out to provide reliable stellar labels, but will compound to produce a poor overall goodness-of-fit across the spectrum. Furthermore, the $\sim$2% accuracy in our model's flux predictions is comparable to intrinsic scatter of other data-driven models in the literature that are successful at label transfer (see for example Figure 5 of Ness et al. 2015 and Figure 6 of Rice and Brewer 2020). It's likely that these models may face similar challenges in applications that require accurate spectral emulation.

While El-Badry et al. (2018b) successfully identified spectral binaries in the APOGEE data, this analysis demonstrates these methods to different datasets is less straightforward. The HIRES spectra likely have higher latent dimensionality than APOGEE spectra– beyond the fact that these spectra are not observed with uniform spectral resolution, the fiber scrambler of HIRES has finite stability, and there is inherent scatter in resolution of spectra obtained under the same observing conditions (Spronck et al., 2015). Assembling a large, homogeneous training set is also a challenge for these data. While the `Specmatch-Emp` library provides an ideal training set in many regards, its labels are compiled from a patchwork of surveys with varying precision and underlying biases. This is compounded by the fact that our binary detection performance depends most sensitively on how well *The Cannon* works in the low-$T_{\rm eff}$ regime (see Figure 4.3), where `Specmatch-emp` $v \sin i$ information is incomplete. A training set composed of thousands of stars with uniformly derived labels spanning such a wide temperature range is not possible with the existing HIRES data.

It is worth exploring the benefits of utilizing data-driven models with higher complexity for the purposes of accurate spectral emulation. Here we demonstrate that a piecewise *Cannon* model is an effective means of adapting quadratic-in-labels models for training sets that span a wide range of stellar temperatures. However, higher complexity models may be a more suitable solution. Indeed, *The Payne* (Ting et al., 2019) uses a neural network to model much more complex flux-label relationships in its training data, and was shown to work for APOGEE data; however, we saw that in our case this model was prone to overfitting to training spectra. In the future, it is worth exploring *The Payne*'s applications to HIRES spectra further, along with other higher-complexity models such as higher-order polynomials or modeling a Gaussian process at every pixel, as proposed in Ness et al. (2015). It is also worth investigating the ways in which the intrinsic model scatter can be incorporated into model comparisons. In particular, assumptions encoded in the 68metric we used may break down in cases of large model uncertainty (i.e., large s_j^2).

Looking to the future, there remains opportunities for models with more detailed spectral emulation that enables detection of unresolved planet-hosting binaries. While we've seen that large surveys like APOGEE and LAMOST provide ideal training sets for data-driven models, it is likely that applications to search for binaries in the *Kepler* and *TESS* fields will require data-driven models of spectra from dedicated ground-based surveys to achieve the signal-to-noise needed for more distant planet hosts at $\sim$1 pc. Beyond this, processing pipelines of large surveys may produce spectra that are less ideal for spectral binary detection, as was shown to be the case for Gaia DR3 spectra (Angelo et al., 2024). More promising are high-resolution spectrometers like HARPS-N and KPF, whose stable PSF and legacy data may offer an opportunity to assemble large, homogeneous training sets with high-complexity models. These may bring us closer to detailed and accurate spectral emulation, an enormous achievement with far-reaching applications across all sub-fields of astronomy.

4.7 Acknowledgments

IA is supported by the Michael A. Jura Fellowship at UCLA. IA would also like to thank Sam Yee, Adam Kraus, Jack Lubin, Smadar Naoz, Kareem El-Badry, Greg Gilbert, Judah Van Zandt, David Ciardi, Brad Hansen and Tuan Do for their helpful insights on the analysis portion of this paper.

Software: astropy (Astropy Collaboration et al., 2013b, 2018b), numpy (Van Der Walt et al., 2011), matplotlib (Hunter, 2007), scipy (Jones et al., 2001), pandas (McKinney, 2010), thecannon (Casey et al., 2016)

Appendix 2A:

Kepler-1656b does not undergo high-eccentricity migration

Under the physical conditions outlined in Section 2.3.1, Kepler-1656b's eccentricity can behave in several ways. In the quadrupole-level approximation, Kepler-1656b's eccentricity will oscillate between fixed minima and maxima. Alternatively, for conditions in which tides or octupole-level behavior are dominant, high eccentricity migration or more chaotic eccentricity evolution can dominate. Consider the former case in which Kepler-1656b's eccentricity oscillates between a fixed minimum and maximum. In this scenario, Naoz et al. (2013a) showed that the total angular momentum is conserved, along with the eccentricity and angular momentum for the outer orbit. This implies that the initial and final mutual inclinations between the two planets (hereafter $i_{\mathrm{bc,i}}$ and $i_{\mathrm{bc,f}}$) obey

$$\cos(i_{\mathrm{bc,f}}) = \frac{A_f \cos(i_{\mathrm{bc,i}}) + L_1^2 e_1^2}{A_f \sqrt{1 - e_{\mathrm{b}}^2}}, \tag{17}$$

where $A_f = 2L_1 L_2 * \sqrt{1 - e_{\mathrm{c}}^2}$, and L_1 and L_2 are the respective conjugate angular momenta of the inner and outer orbit. Thus, if Kepler-1656b undergoes quadrupole-dominated eccentricity oscillations with no significant tides or octupole-level oscillations, it should obey the relationship described by eq. 17. We plot this relationship for a range of eccentricities for Kepler-1656b that are reasonably close to our observed value ($0.6 < e_p < 0.9$), and find that most simulations that match our observations lie in this regime. This indicates quadrupole-dominated behavior for Kepler-1656b, whereby the planet reached achieved its high eccentricity via sustained, high-amplitude eccentricity oscillations without tidal migration of chaotic behavior.

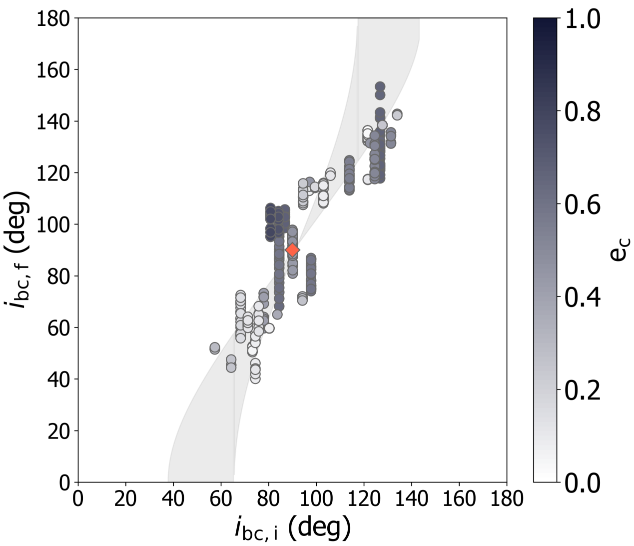

Figure 12: Initial and final mutual inclinations for all model timestamps consistent with observations of Kepler-1656b. The shaded region shows the initial and final inclinations predicted by the quadrupole-level approximation for inner planet eccentricities of $0.6 < e_{\mathrm{b}} < 0.9$.

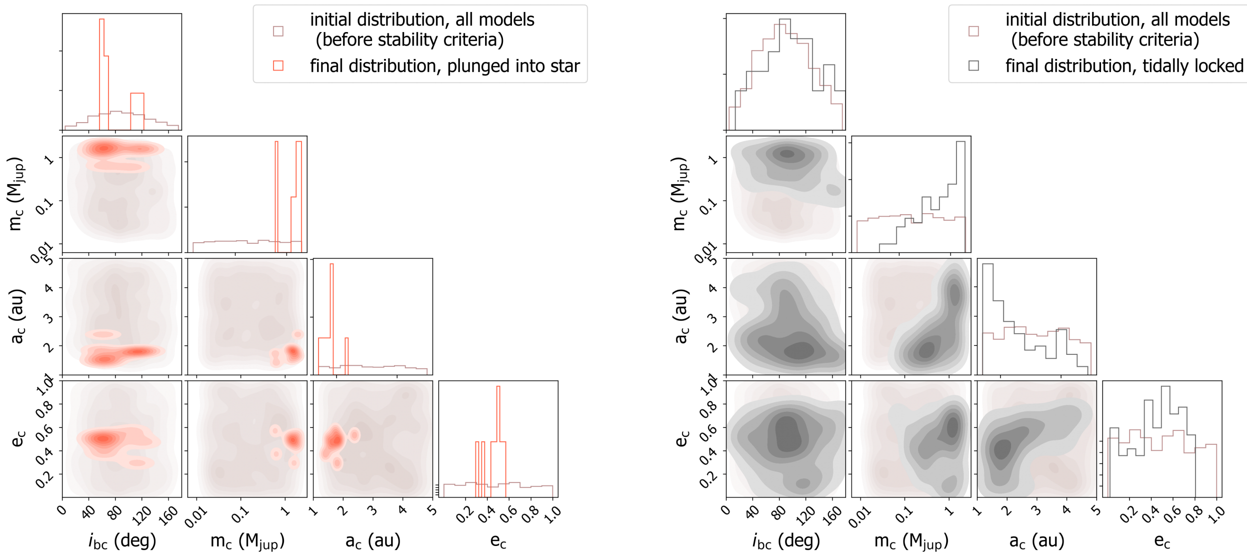

Figure 13: Smoothed 2D initial and final companion parameter distributions for models that plunged into their host star (left) and became tidally locked (right). The diagonal panels show 1D distributions of parameters that have been normalized to sum to 1.

Appendix 2B:

Correlated Parameters for Plunging, Locked Models

We find in our simulations that scenarios in which the inner planet plunges into the host star or becomes tidally locked arise from a unique set of parameters for the outer companion. This is evident in Figure 13, which shows final distributions of companion parameters e_c, m_c, a_c, and $i_{\rm bc}$ for both cases. We see here that moderately eccentric ($e_c \approx 0.5$), aligned companions are more likely to cause the inner planet to plunge into its host star. On the other hand, companions on polar orbits with a wider range of e_c can cause the inner planet to become tidally locked. Thus, we predict that tidally locked sub-Saturns may be signatures of eccentric, polar companions.

.1 Appendix 3: Model Coefficients

As described in Section 3.2, the trained model coefficients θ_j describe the model flux dependence on the stellar labels. The magnitude of these coefficient vectors is related to how strongly the model depends on a given label or combination of labels at a particular pixel. During the training step, *The Cannon* computes coefficient vectors for all linear and quadratic terms in our model, and we show the linear terms corresponding to each label here. We also show the model scatter, which was computed empirically as outlined in Section 3.4.2.

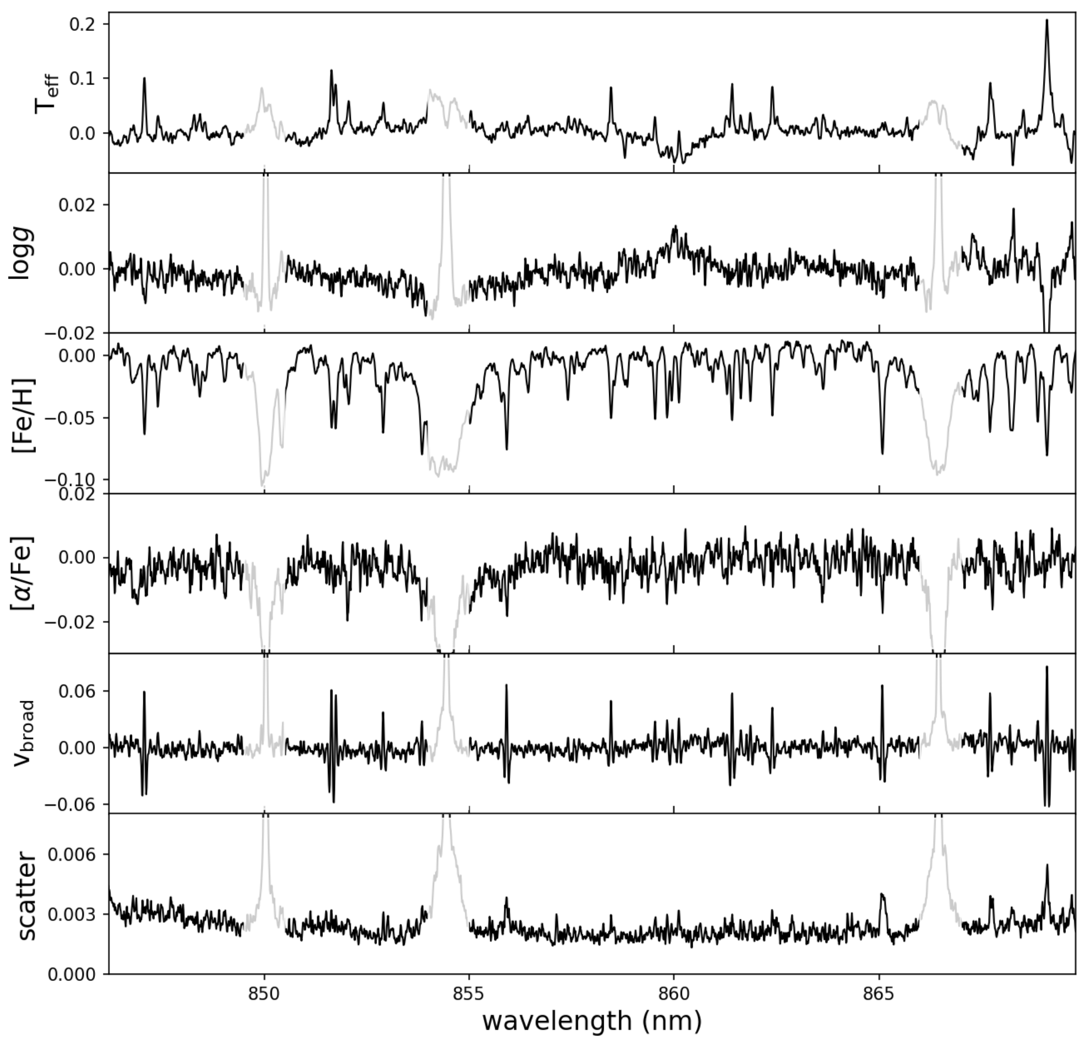

Figure 14: First-order coefficient vectors (i.e., spectral derivatives) of our trained model for each of our training labels (top 5 panels), and model scatter (bottom panel) as a function of the Gaia RVS wavelength scale. For each label, peaks in the coefficient vectors represent pixels in the spectrum that contain the most information for that particular label.

Bibliography

Aguilera-Gómez, C., Chanamé, J., Pinsonneault, M. H., and Carlberg, J. K. (2016). On Lithium-rich Red Giants. I. Engulfment of Substellar Companions. , 829(2):127.

Akeson, R. L., Chen, X., Ciardi, D., Crane, M., Good, J., Harbut, M., Jackson, E., Kane, S. R., Laity, A. C., Leifer, S., Lynn, M., McElroy, D. L., Papin, M., Plavchan, P., Ramírez, S. V., Rey, R., von Braun, K., Wittman, M., Abajian, M., Ali, B., Beichman, C., Beekley, A., Berriman, G. B., Berukoff, S., Bryden, G., Chan, B., Groom, S., Lau, C., Payne, A. N., Regelson, M., Saucedo, M., Schmitz, M., Stauffer, J., Wyatt, P., and Zhang, A. (2013). The NASA Exoplanet Archive: Data and Tools for Exoplanet Research. , 125(930):989.

Albrecht, S., Winn, J. N., Johnson, J. A., Howard, A. W., Marcy, G. W., Butler, R. P., Arriagada, P., Crane, J. D., Shectman, S. A., Thompson, I. B., Hirano, T., Bakos, G., and Hartman, J. D. (2012). Obliquities of Hot Jupiter Host Stars: Evidence for Tidal Interactions and Primordial Misalignments. , 757(1):18.

Anderson, K. R., Storch, N. I., and Lai, D. (2016). Formation and stellar spin-orbit misalignment of hot Jupiters from Lidov-Kozai oscillations in stellar binaries. , 456(4):3671–3701.

Andrae, R., Fouesneau, M., Sordo, R., Bailer-Jones, C. A. L., Dharmawardena, T. E., Rybizki, J., De Angeli, F., Lindstrøm, H. E. P., Marshall, D. J., Drimmel, R., Korn, A. J., Soubiran, C., Brouillet, N., Casamiquela, L., Rix, H. W., Abreu Aramburu, A., Álvarez, M. A., Bakker, J., Bellas-Velidis, I., Bijaoui, A., Brugaletta, E., Burlacu, A., Carballo, R., Chaoul, L., Chiavassa, A., Contursi, G., Cooper, W. J., Creevey, O. L., Dafonte, C., Dapergolas, A., de Laverny, P., Delchambre, L., Demouchy, C., Edvardsson, B., Frémat, Y., Garabato, D., García-Lario, P., García-Torres, M., Gavel, A., Gomez, A., González-Santamaría, I., Hatzidimitriou, D., Heiter, U., Jean-Antoine Piccolo, A., Kontizas, M., Kordopatis, G., Lanzafame, A. C., Lebreton, Y., Licata, E. L., Livanou, E., Lobel, A., Lorca, A., Magdaleno Romeo, A., Manteiga, M., Marocco, F., Mary, N., Nicolas, C., Ordenovic, C., Pailler, F., Palicio, P. A., Pallas-Quintela, L., Panem, C., Pichon, B., Poggio, E., Recio-Blanco, A., Riclet, F., Robin, C., Santoveña, R., Sarro, L. M., Schultheis, M. S., Segol, M., Silvelo, A., Slezak, I., Smart, R. L., Süveges, M., Thévenin, F., Torralba Elipe, G., Ulla, A., Utrilla, E., Vallenari, A., van Dillen, E., Zhao, H., and Zorec, J. (2023). Gaia Data Release 3. Analysis of the Gaia BP/RP spectra using the General Stellar Parameterizer from Photometry. , 674:A27.

Angelo, I., Bedell, M., Petigura, E., and Ness, M. (2024). A Data-driven Spectral Model of Main-sequence Stars in Gaia DR3. , 974(1):43.

Angelo, I., Naoz, S., Petigura, E., MacDougall, M., Stephan, A. P., Isaacson, H., and Howard, A. W. (2022). Kepler-1656b's Extreme Eccentricity: Signature of a Gentle Giant. , 163(5):227.

Antognini, J. M. O. (2015). Timescales of Kozai-Lidov oscillations at quadrupole and octupole order in the test particle limit. , 452(4):3610–3619.

Armitage, P. J. (2007). Massive Planet Migration: Theoretical Predictions and Comparison with Observations. , 665(2):1381–1390.

Artymowicz, P., Clarke, C. J., Lubow, S. H., and Pringle, J. E. (1991). The Effect of an External Disk on the Orbital Elements of a Central Binary. , 370:L35.

Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., Günther, H. M., Lim, P. L., Crawford, S. M., Conseil, S., Shupe, D. L., Craig, M. W., Dencheva, N., Ginsburg, A., VanderPlas, J. T., Bradley, L. D., Pérez-Suárez, D., de Val-Borro, M., Aldcroft, T. L., Cruz, K. L., Robitaille, T. P., Tollerud, E. J., Ardelean, C., Babej, T., Bach, Y. P., Bachetti, M., Bakanov, A. V., Bamford, S. P., Barentsen, G., Barmby, P., Baumbach, A., Berry, K. L., Biscani, F., Boquien, M., Bostroem, K. A., Bouma, L. G., Brammer, G. B., Bray, E. M., Breytenbach, H., Buddelmeijer, H., Burke, D. J., Calderone, G., Cano Rodríguez, J. L., Cara, M., Cardoso, J. V. M., Cheedella, S., Copin, Y., Corrales, L., Crichton, D., D'Avella, D., Deil, C., Depagne, É., Dietrich, J. P., Donath, A., Droettboom, M., Earl, N., Erben, T., Fabbro, S., Ferreira, L. A., Finethy, T., Fox, R. T., Garrison, L. H., Gibbons, S. L. J., Goldstein, D. A., Gommers, R., Greco, J. P., Greenfield, P., Groener, A. M., Grollier, F., Hagen, A., Hirst, P., Homeier, D., Horton, A. J., Hosseinzadeh, G., Hu, L., Hunkeler, J. S., Ivezić, Ž., Jain, A., Jenness, T., Kanarek, G., Kendrew, S., Kern, N. S., Kerzendorf, W. E., Khvalko, A., King, J., Kirkby, D., Kulkarni, A. M., Kumar, A., Lee, A., Lenz, D., Littlefair, S. P., Ma, Z., Macleod, D. M., Mastropietro, M., McCully, C., Montagnac, S., Morris, B. M., Mueller, M., Mumford, S. J., Muna, D., Murphy, N. A., Nelson, S., Nguyen, G. H., Ninan, J. P., Nöthe, M., Ogaz, S., Oh, S., Parejko, J. K., Parley, N., Pascual, S., Patil, R., Patil, A. A., Plunkett, A. L., Prochaska, J. X., Rastogi, T., Reddy Janga, V., Sabater, J., Sakurikar, P., Seifert, M., Sherbert, L. E., Sherwood-Taylor, H., Shih, A. Y., Sick, J., Silbiger, M. T., Singanamalla, S., Singer, L. P., Sladen, P. H., Sooley, K. A., Sornarajah, S., Streicher, O., Teuben, P., Thomas, S. W., Tremblay, G. R., Turner, J. E. H., Terrón, V., van Kerkwijk, M. H., de la Vega, A., Watkins, L. L., Weaver, B. A., Whitmore, J. B., Woillez, J., Zabalza, V., and Astropy Contributors (2018a). The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package. , 156(3):123.

Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., Günther, H. M., Lim, P. L., Crawford, S. M., Conseil, S., Shupe, D. L., Craig, M. W., Dencheva, N., Ginsburg, A., VanderPlas, J. T., Bradley, L. D., Pérez-Suárez, D., de Val-Borro, M., Aldcroft, T. L., Cruz, K. L., Robitaille, T. P., Tollerud, E. J., Ardelean, C., Babej, T., Bach, Y. P., Bachetti, M., Bakanov, A. V., Bamford, S. P., Barentsen, G., Barmby, P., Baumbach, A., Berry, K. L., Biscani, F., Boquien, M., Bostroem, K. A., Bouma, L. G., Brammer, G. B., Bray, E. M., Breytenbach, H., Buddelmeijer, H., Burke, D. J., Calderone, G., Cano Rodríguez, J. L., Cara, M., Cardoso, J. V. M., Cheedella, S., Copin, Y., Corrales, L., Crichton, D., D'Avella, D., Deil, C., Depagne, É., Dietrich, J. P., Donath, A., Droettboom, M., Earl, N., Erben, T., Fabbro, S., Ferreira, L. A., Finethy, T., Fox, R. T., Garrison, L. H., Gibbons, S. L. J., Goldstein, D. A., Gommers, R., Greco, J. P., Greenfield, P., Groener, A. M., Grollier, F., Hagen, A., Hirst, P., Homeier, D., Horton, A. J., Hosseinzadeh, G., Hu, L., Hunkeler, J. S., Ivezić, Ž., Jain, A., Jenness, T., Kanarek, G., Kendrew, S., Kern, N. S., Kerzendorf, W. E., Khvalko, A., King, J., Kirkby, D., Kulkarni, A. M., Kumar, A., Lee, A., Lenz, D., Littlefair, S. P., Ma, Z., Macleod, D. M., Mastropietro, M., McCully, C., Montagnac, S., Morris, B. M., Mueller, M., Mumford, S. J., Muna, D., Murphy, N. A., Nelson, S., Nguyen, G. H., Ninan, J. P., Nöthe, M., Ogaz, S., Oh, S., Parejko, J. K., Parley, N., Pascual, S., Patil, R., Patil, A. A., Plunkett, A. L., Prochaska, J. X., Rastogi, T., Reddy Janga, V., Sabater, J., Sakurikar, P., Seifert, M., Sherbert, L. E., Sherwood-Taylor, H., Shih, A. Y., Sick, J., Silbiger, M. T., Singanamalla, S., Singer, L. P., Sladen, P. H., Sooley, K. A., Sornarajah, S., Streicher, O., Teuben, P., Thomas, S. W., Tremblay, G. R., Turner, J. E. H., Terrón, V., van Kerkwijk, M. H., de la Vega, A., Watkins, L. L., Weaver, B. A., Whitmore, J. B., Woillez, J., Zabalza, V., and Astropy Contributors (2018b). The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package. , 156(3):123.

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., Greenfield, P., Droettboom, M., Bray, E., Aldcroft, T., Davis, M., Ginsburg, A., Price-Whelan, A. M., Kerzendorf, W. E., Conley, A., Crighton, N., Barbary, K., Muna, D., Ferguson, H., Grollier, F., Parikh, M. M., Nair, P. H., Unther, H. M., Deil, C., Woillez, J., Conseil, S., Kramer, R., Turner, J. E. H., Singer, L., Fox, R., Weaver, B. A., Zabalza, V., Edwards, Z. I., Azalee Bostroem, K., Burke, D. J., Casey, A. R., Crawford, S. M., Dencheva, N., Ely, J., Jenness, T., Labrie, K., Lim, P. L., Pierfederici, F., Pontzen, A., Ptak, A., Refsdal, B., Servillat, M., and Streicher, O. (2013a). Astropy: A community Python package for astronomy. , 558:A33.

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., Greenfield, P., Droettboom, M., Bray, E., Aldcroft, T., Davis, M., Ginsburg, A., Price-Whelan, A. M., Kerzendorf, W. E., Conley, A., Crighton, N., Barbary, K., Muna, D., Ferguson, H., Grollier, F., Parikh, M. M., Nair, P. H., Unther, H. M., Deil, C., Woillez, J., Conseil, S., Kramer, R., Turner, J. E. H., Singer, L., Fox, R., Weaver, B. A., Zabalza, V., Edwards, Z. I., Azalee Bostroem, K., Burke, D. J., Casey, A. R., Crawford, S. M., Dencheva, N., Ely, J., Jenness, T., Labrie, K., Lim, P. L., Pierfederici, F., Pontzen, A., Ptak, A., Refsdal, B., Servillat, M., and Streicher, O. (2013b). Astropy: A community Python package for astronomy. , 558:A33.

Bakos, G. Á., Howard, A. W., Noyes, R. W., Hartman, J., Torres, G., Kovács, G., Fischer, D. A., Latham, D. W., Johnson, J. A., Marcy, G. W., Sasselov, D. D., Stefanik, R. P., Sipőcz, B., Kovács, G., Esquerdo, G. A., Pál, A., Lázár, J., Papp, I., and Sári, P. (2009). HAT-P-13b,c: A Transiting Hot Jupiter with a Massive Outer Companion on an Eccentric Orbit. , 707(1):446–456.

Baliunas, S. L., Donahue, R. A., Soon, W. H., Horne, J. H., Frazer, J., Woodard-Eklund, L., Bradford, M., Rao, L. M., Wilson, O. C., Zhang, Q., Bennett, W., Briggs, J., Carroll, S. M., Duncan, D. K., Figueroa, D., Lanning, H. H., Misch, T., Mueller, J., Noyes, R. W., Poppe, D., Porter, A. C., Robinson, C. R., Russell, J., Shelton, J. C., Soyumer, T., Vaughan, A. H., and Whitney, J. H. (1995). Chromospheric Variations in Main-Sequence Stars. II. , 438:269.

Barnes, J. W., Linscott, E., and Shporer, A. (2011). Measurement of the Spin-Orbit Misalignment of KOI-13.01 from Its Gravity-darkened Kepler Transit Lightcurve. , 197(1):10.

Batygin, K., Adams, F. C., Batygin, Y. K., and Petigura, E. A. (2020). Dynamics of Planetary Systems within Star Clusters: Aspects of the Solar System's Early Evolution. , 159(3):101.

Batygin, K., Bodenheimer, P. H., and Laughlin, G. P. (2016). In Situ Formation and Dynamical Evolution of Hot Jupiter Systems. , 829(2):114.

Behmard, A., Ness, M. K., Casey, A. R., Angus, R., Cunha, K., Souto, D., Lu, Y., and Johnson, J. A. (2025). A Data-driven M Dwarf Model and Detailed Abundances for ∼17,000 M Dwarfs in SDSS-V. , 982(1):13.

Behmard, A., Petigura, E. A., and Howard, A. W. (2019). Data-driven Spectroscopy of Cool Stars at High Spectral Resolution. , 876(1):68.

Belokurov, V., Penoyre, Z., Oh, S., Iorio, G., Hodgkin, S., Evans, N. W., Everall, A., Koposov, S. E., Tout, C. A., Izzard, R., Clarke, C. J., and Brown, A. G. A. (2020). Unresolved stellar companions with Gaia DR2 astrometry. , 496(2):1922–1940.

Berger, T. A., Huber, D., Gaidos, E., and van Saders, J. L. (2018). Revised Radii of Kepler Stars and Planets Using Gaia Data Release 2. , 866(2):99.

Bharat Kumar, Y., Singh, R., Eswar Reddy, B., and Zhao, G. (2018). Two New Super Li-rich Core He-burning Giants: A New Twist to the Long Tale of Li Enhancement in K Giants. , 858(2):L22.

Bhaskar, H., Li, G., Hadden, S., Payne, M. J., and Holman, M. J. (2021). Mildly Hierarchical Triple Dynamics and Applications to the Outer Solar System. , 161(1):48.

Blunt, S., Endl, M., Weiss, L. M., Cochran, W. D., Howard, A. W., MacQueen, P. J., Fulton, B. J., Henry, G. W., Johnson, M. C., Kosiarek, M. R., Lawson, K. D., Macintosh, B., Mills, S. M., Nielsen, E. L., Petigura, E. A., Schneider, G., Vanderburg, A., Wisniewski, J. P., Wittenmyer, R. A., Brugamyer, E., Caldwell, C., Cochran, A. L., Hatzes, A. P., Hirsch, L. A., Isaacson, H., Robertson, P., Roy, A., and Shen, Z. (2019). Radial Velocity Discovery of an Eccentric Jovian World Orbiting at 18 au. , 158(5):181.

Borucki, W. J., Koch, D., Basri, G., Batalha, N., Brown, T., Caldwell, D., Caldwell, J., Christensen-Dalsgaard, J., Cochran, W. D., DeVore, E., Dunham, E. W., Dupree, A. K., Gautier, T. N., Geary, J. C., Gilliland, R., Gould, A., Howell, S. B., Jenkins, J. M., Kondo, Y., Latham, D. W., Marcy, G. W., Meibom, S., Kjeldsen, H., Lissauer, J. J., Monet, D. G., Morrison, D., Sasselov, D., Tarter, J., Boss, A., Brownlee, D., Owen, T., Buzasi, D., Charbonneau, D., Doyle, L., Fortney, J., Ford, E. B., Holman, M. J., Seager, S., Steffen, J. H., Welsh, W. F., Rowe, J., Anderson, H., Buchhave, L., Ciardi, D., Walkowicz, L., Sherry, W., Horch, E., Isaacson, H., Everett, M. E., Fischer, D., Torres, G., Johnson, J. A., Endl, M., MacQueen, P., Bryson, S. T., Dotson, J., Haas, M., Kolodziejczak, J., Van Cleve, J., Chandrasekaran, H., Twicken, J. D., Quintana, E. V., Clarke, B. D., Allen, C., Li, J., Wu, H., Tenenbaum, P., Verner, E., Bruhweiler, F., Barnes, J., and Prsa, A. (2010). Kepler Planet-Detection Mission: Introduction and First Results. *Science*, 327(5968):977.

Boss, A. P. (1997). Giant planet formation by gravitational instability. *Science*, 276:1836–1839.

Brady, M. T., Petigura, E. A., Knutson, H. A., Sinukoff, E., Isaacson, H., Hirsch, L. A., Fulton, B. J., Kosiarek, M. R., and Howard, A. W. (2018). Kepler-1656b: A Dense Sub-Saturn with an Extreme Eccentricity. , 156(4):147.

Brewer, J. M., Fischer, D. A., Valenti, J. A., and Piskunov, N. (2016). Spectral Properties of Cool Stars: Extended Abundance Analysis of 1,617 Planet-search Stars. , 225(2):32.

Brown, T. M., Latham, D. W., Everett, M. E., and Esquerdo, G. A. (2011). Kepler Input Catalog: Photometric Calibration and Stellar Classification. , 142(4):112.

Buder, S., Asplund, M., Duong, L., Kos, J., Lind, K., Ness, M. K., Sharma, S., Bland-Hawthorn, J., Casey, A. R., de Silva, G. M., D'Orazi, V., Freeman, K. C., Lewis, G. F., Lin, J., Martell, S. L., Schlesinger, K. J., Simpson, J. D., Zucker, D. B., Zwitter, T., Amarsi, A. M., Anguiano, B., Carollo, D., Casagrande, L., Čotar, K., Cottrell, P. L., da

Costa, G., Gao, X. D., Hayden, M. R., Horner, J., Ireland, M. J., Kafle, P. R., Munari, U., Nataf, D. M., Nordlander, T., Stello, D., Ting, Y.-S., Traven, G., Watson, F., Wittenmyer, R. A., Wyse, R. F. G., Yong, D., Zinn, J. C., Žerjal, M., and Galah Collaboration (2018). The GALAH Survey: second data release. , 478(4):4513–4552.

Buder, S., Sharma, S., Kos, J., Amarsi, A. M., Nordlander, T., Lind, K., Martell, S. L., Asplund, M., Bland-Hawthorn, J., Casey, A. R., de Silva, G. M., D'Orazi, V., Freeman, K. C., Hayden, M. R., Lewis, G. F., Lin, J., Schlesinger, K. J., Simpson, J. D., Stello, D., Zucker, D. B., Zwitter, T., Beeson, K. L., Buck, T., Casagrande, L., Clark, J. T., Čotar, K., da Costa, G. S., de Grijs, R., Feuillet, D., Horner, J., Kafle, P. R., Khanna, S., Kobayashi, C., Liu, F., Montet, B. T., Nandakumar, G., Nataf, D. M., Ness, M. K., Spina, L., Tepper-García, T., Ting, Y.-S., Traven, G., Vogrinčič, R., Wittenmyer, R. A., Wyse, R. F. G., Žerjal, M., and Galah Collaboration (2021). The GALAH+ survey: Third data release. , 506(1):150–201.

Butler, R. P., Marcy, G. W., Vogt, S. S., and Apps, K. (1998). A Planet with a 3.1 Day Period around a Solar Twin. , 110(754):1389–1393.

Cannon, A. J. and Pickering, E. C. (1918). The Henry Draper catalogue 0h, 1h, 2h, and 3h. *Annals of Harvard College Observatory*, 91:1–290.

Cannon, A. J. and Pickering, E. C. (1924). The Henry Draper catalogue : 21h, 22h, and 23h. *Annals of Harvard College Observatory*, 99:1.

Casey, A. R., Hogg, D. W., Ness, M., Rix, H.-W., Ho, A. Q. Y., and Gilmore, G. (2016). The Cannon 2: A data-driven model of stellar spectra for detailed chemical abundance analyses. *arXiv e-prints*, page arXiv:1603.03040.

Chatterjee, S., Ford, E. B., Matsumura, S., and Rasio, F. A. (2008). Dynamical Outcomes of Planet-Planet Scattering. , 686(1):580–602.

Ciardi, D. R., Beichman, C. A., Horch, E. P., and Howell, S. B. (2015). Understanding the Effects of Stellar Multiplicity on the Derived Planet Radii from Transit Surveys: Implications for Kepler, K2, and TESS. , 805(1):16.

Cieza, L. A., Padgett, D. L., Allen, L. E., McCabe, C. E., Brooke, T. Y., Carey, S. J., Chapman, N. L., Fukagawa, M., Huard, T. L., Noriga-Crespo, A., Peterson, D. E., and Rebull, L. M. (2009). Primordial Circumstellar Disks in Binary Systems: Evidence for Reduced Lifetimes. , 696(1):L84–L88.

Correia, A. C. M., Bourrier, V., and Delisle, J. B. (2020). Why do warm Neptunes present nonzero eccentricity? , 635:A37.

Crida, A. and Morbidelli, A. (2007). Cavity opening by a giant planet in a protoplanetary disc and effects on planetary migration. , 377(3):1324–1336.

Cropper, M., Katz, D., Sartoretti, P., Prusti, T., de Bruijne, J. H. J., Chassat, F., Charvet, P., Boyadjian, J., Perryman, M., Sarri, G., Gare, P., Erdmann, M., Munari, U., Zwitter, T., Wilkinson, M., Arenou, F., Vallenari, A., Gómez, A., Panuzzo, P., Seabroke, G., Allende Prieto, C., Benson, K., Marchal, O., Huckle, H., Smith, M., Dolding, C., Janßen, K., Viala, Y., Blomme, R., Baker, S., Boudreault, S., Crifo, F., Soubiran, C., Frémat, Y., Jasniewicz, G., Guerrier, A., Guy, L. P., Turon, C., Jean-Antoine-Piccolo, A., Thévenin, F., David, M., Gosset, E., and Damerdji, Y. (2018). Gaia Data Release 2. Gaia Radial Velocity Spectrometer. , 616:A5.

Crossfield, I. J. M., Ciardi, D. R., Petigura, E. A., Sinukoff, E., Schlieder, J. E., Howard, A. W., Beichman, C. A., Isaacson, H., Dressing, C. D., Christiansen, J. L., Fulton, B. J., Lépine, S., Weiss, L., Hirsch, L., Livingston, J., Baranec, C., Law, N. M., Riddle, R., Ziegler, C., Howell, S. B., Horch, E., Everett, M., Teske, J., Martinez, A. O., Obermeier, C., Benneke, B., Scott, N., Deacon, N., Aller, K. M., Hansen, B. M. S., Mancini, L., Ciceri, S., Brahm, R., Jordán, A., Knutson, H. A., Henning, T., Bonnefoy, M., Liu, M. C., Crepp, J. R., Lothringer, J., Hinz, P., Bailey, V., Skemer, A., and Defrere, D. (2016). 197 Candidates and 104 Validated Planets in K2's First Five Fields. , 226(1):7.

Cui, X.-Q., Zhao, Y.-H., Chu, Y.-Q., Li, G.-P., Li, Q., Zhang, L.-P., Su, H.-J., Yao, Z.-Q., Wang, Y.-N., Xing, X.-Z., Li, X.-N., Zhu, Y.-T., Wang, G., Gu, B.-Z., Luo, A. L., Xu, X.-Q., Zhang, Z.-C., Liu, G.-R., Zhang, H.-T., Yang, D.-H., Cao, S.-Y., Chen, H.-Y., Chen, J.-J., Chen, K.-X., Chen, Y., Chu, J.-R., Feng, L., Gong, X.-F., Hou, Y.-H., Hu, H.-Z., Hu, N.-S., Hu, Z.-W., Jia, L., Jiang, F.-H., Jiang, X., Jiang, Z.-B., Jin, G., Li, A.-H., Li, Y., Li, Y.-P., Liu, G.-Q., Liu, Z.-G., Lu, W.-Z., Mao, Y.-D., Men, L., Qi, Y.-J., Qi, Z.-X., Shi, H.-M., Tang, Z.-H., Tao, Q.-S., Wang, D.-Q., Wang, D., Wang, G.-M., Wang, H., Wang, J.-N., Wang, J., Wang, J.-L., Wang, J.-P., Wang, L., Wang, S.-Q., Wang, Y., Wang, Y.-F., Xu, L.-Z., Xu, Y., Yang, S.-H., Yu, Y., Yuan, H., Yuan, X.-Y., Zhai, C., Zhang, J., Zhang, Y.-X., Zhang, Y., Zhao, M., Zhou, F., Zhou, G.-H., Zhu, J., and Zou, S.-C. (2012). The Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST). *Research in Astronomy and Astrophysics*, 12(9):1197–1242.

Cumming, A., Butler, R. P., Marcy, G. W., Vogt, S. S., Wright, J. T., and Fischer, D. A. (2008). The Keck Planet Search: Detectability and the Minimum Mass and Orbital Period Distribution of Extrasolar Planets. , 120(867):531.

Czekala, I., Andrews, S. M., Mandel, K. S., Hogg, D. W., and Green, G. M. (2015). Constructing a Flexible Likelihood Function for Spectroscopic Inference. , 812(2):128.

Dalal, S., Hébrard, G., Lecavelier des Étangs, A., Petit, A. C., Bourrier, V., Laskar, J., König, P. C., and Correia, A. C. M. (2019). Nearly polar orbit of the sub-Neptune HD 3167 c. Constraints on the dynamical history of a multi-planet system. , 631:A28.

Dawson, R. I. and Johnson, J. A. (2018). Origins of Hot Jupiters. , 56:175–221.

Dawson, R. I., Johnson, J. A., Fabrycky, D. C., Foreman-Mackey, D., Murray-Clay, R. A., Buchhave, L. A., Cargile, P. A., Clubb, K. I., Fulton, B. J., Hebb, L., Howard, A. W., Huber, D., Shporer, A., and Valenti, J. A. (2014). Large Eccentricity, Low Mutual Inclination: The Three-dimensional Architecture of a Hierarchical System of Giant Planets. , 791(2):89.

Dawson, R. I. and Murray-Clay, R. A. (2013). Giant Planets Orbiting Metal-rich Stars Show Signatures of Planet-Planet Interactions. , 767(2):L24.

De Silva, G. M., Freeman, K. C., Bland-Hawthorn, J., Martell, S., de Boer, E. W., Asplund, M., Keller, S., Sharma, S., Zucker, D. B., Zwitter, T., Anguiano, B., Bacigalupo, C., Bayliss, D., Beavis, M. A., Bergemann, M., Campbell, S., Cannon, R., Carollo, D., Casagrande, L., Casey, A. R., Da Costa, G., D'Orazi, V., Dotter, A., Duong, L., Heger, A., Ireland, M. J., Kafle, P. R., Kos, J., Lattanzio, J., Lewis, G. F., Lin, J., Lind, K., Munari, U., Nataf, D. M., O'Toole, S., Parker, Q., Reid, W., Schlesinger, K. J., Sheinis, A., Simpson, J. D., Stello, D., Ting, Y. S., Traven, G., Watson, F., Wittenmyer, R., Yong, D., and Žerjal, M. (2015). The GALAH survey: scientific motivation. , 449(3):2604–2617.

del Ser, D., Fors, O., and Núñez, J. (2018). TFAW: Wavelet-based signal reconstruction to reduce photometric noise in time-domain surveys. , 619:A86.

Deleuil, M., Bonomo, A. S., Ferraz-Mello, S., Erikson, A., Bouchy, F., Havel, M., Aigrain, S., Almenara, J. M., Alonso, R., Auvergne, M., Baglin, A., Barge, P., Bordé, P., Bruntt, H., Cabrera, J., Carpano, S., Cavarroc, C., Csizmadia, S., Damiani, C., Deeg, H. J., Dvorak, R., Fridlund, M., Hébrard, G., Gandolfi, D., Gillon, M., Guenther, E., Guillot, T., Hatzes, A., Jorda, L., Léger, A., Lammer, H., Mazeh, T., Moutou, C., Ollivier, M., Ofir, A., Parviainen, H., Queloz, D., Rauer, H., Rodríguez, A., Rouan, D., Santerne, A., Schneider, J., Tal-Or, L., Tingley, B., Weingrill, J., and Wuchterl, G. (2012). Transiting exoplanets from the CoRoT space mission. XX. CoRoT-20b: A very high density, high eccentricity transiting giant planet. , 538:A145.

Díaz, R. F., Santerne, A., Sahlmann, J., Hébrard, G., Eggenberger, A., Santos, N. C., Moutou, C., Arnold, L., Boisse, I., Bonfils, X., Bouchy, F., Delfosse, X., Desort, M., Ehrenreich, D., Forveille, T., Lagrange, A. M., Lovis, C., Pepe, F., Perrier, C., Queloz, D., Ségransan, D., Udry, S., and Vidal-Madjar, A. (2012). The SOPHIE search for northern extrasolar planets. IV. Massive companions in the planet-brown dwarf boundary. , 538:A113.

Dobbs-Dixon, I., Lin, D. N. C., and Mardling, R. A. (2004). Spin-Orbit Evolution of Short-Period Planets. , 610(1):464–476.

Duffell, P. C. and MacFadyen, A. I. (2013). Gap Opening by Extremely Low-mass Planets in a Viscous Disk. , 769(1):41.

Duncan, D. K., Vaughan, A. H., Wilson, O. C., Preston, G. W., Frazer, J., Lanning, H., Misch, A., Mueller, J., Soyumer, D., Woodard, L., Baliunas, S. L., Noyes, R. W., Hartmann, L. W., Porter, A., Zwaan, C., Middelkoop, F., Rutten, R. G. M., and Mihalas, D. (1991). CA II H and K Measurements Made at Mount Wilson Observatory, 1966–1983. , 76:383.

Edgar, R. G., Quillen, A. C., and Park, J. (2007). The minimum gap-opening planet mass in an irradiated circumstellar accretion disc. , 381(3):1280–1286.

Eggleton, P. P., Kiseleva, L. G., and Hut, P. (1998). The Equilibrium Tide Model for Tidal Friction. , 499(2):853–870.

Eggleton, P. P. and Kiseleva-Eggleton, L. (2001). Orbital Evolution in Binary and Triple Stars, with an Application to SS Lacertae. , 562(2):1012–1030.

El-Badry, K., Rix, H.-W., Ting, Y.-S., Weisz, D. R., Bergemann, M., Cargile, P., Conroy, C., and Eilers, A.-C. (2018a). Signatures of unresolved binaries in stellar spectra: implications for spectral fitting. , 473(4):5043–5049.

El-Badry, K., Ting, Y.-S., Rix, H.-W., Quataert, E., Weisz, D. R., Cargile, P., Conroy, C., Hogg, D. W., Bergemann, M., and Liu, C. (2018b). Discovery and characterization of 3000+ main-sequence binaries from APOGEE spectra. , 476(1):528–553.

Fabrycky, D. and Tremaine, S. (2007). Shrinking Binary and Planetary Orbits by Kozai Cycles with Tidal Friction. , 669(2):1298–1315.

Feng, F., Shectman, S. A., Clement, M. S., Vogt, S. S., Tuomi, M., Teske, J. K., Burt, J., Crane, J. D., Holden, B., Wang, S. X., Thompson, I. B., Díaz, M. R., and Butler, R. P. (2020). Search for Nearby Earth Analogs .III. Detection of 10 New Planets, 3 Planet Candidates, and Confirmation of 3 Planets around 11 Nearby M Dwarfs. , 250(2):29.

Ford, E. B., Kozinsky, B., and Rasio, F. A. (2000). Secular Evolution of Hierarchical Triple Star Systems. , 535(1):385–401.

Ford, E. B., Lystad, V., and Rasio, F. A. (2005). Planet-planet scattering in the upsilon Andromedae system. , 434(7035):873–876.

Ford, E. B. and Rasio, F. A. (2008). Origins of Eccentric Extrasolar Planets: Testing the Planet-Planet Scattering Model. , 686(1):621–636.

Foreman-Mackey, D. (2016). corner.py: Scatterplot matrices in Python. *The Journal of Open Source Software*, 1:24.

Frémat, Y., Royer, F., Marchal, O., Blomme, R., Sartoretti, P., Guerrier, A., Panuzzo, P., Katz, D., Seabroke, G. M., Thévenin, F., Cropper, M., Benson, K., Damerdji, Y., Haigron, R., Lobel, A., Smith, M., Baker, S. G., Chemin, L., David, M., Dolding, C., Gosset, E., Janßen, K., Jasniewicz, G., Plum, G., Samaras, N., Snaith, O., Soubiran, C., Vanel, O., Zorec, J., Zwitter, T., Brouillet, N., Caffau, E., Crifo, F., Fabre, C., Fragkoudi, F., Huckle, H. E., Lasne, Y., Leclerc, N., Mastrobuono-Battisti, A., Jean-Antoine Piccolo, A., and Viala, Y. (2023). Gaia Data Release 3. Properties of the line-broadening parameter derived with the Radial Velocity Spectrometer (RVS). , 674:A8.

Fulton, B. J. and Petigura, E. A. (2018). The California-Kepler Survey. VII. Precise Planet Radii Leveraging Gaia DR2 Reveal the Stellar Mass Dependence of the Planet Radius Gap. , 156(6):264.

Fulton, B. J., Petigura, E. A., Blunt, S., and Sinukoff, E. (2018). RadVel: The Radial Velocity Modeling Toolkit. , 130(986):044504.

Furlan, E., Ciardi, D. R., Everett, M. E., Saylors, M., Teske, J. K., Horch, E. P., Howell, S. B., van Belle, G. T., Hirsch, L. A., Gautier, T. N., I., Adams, E. R., Barrado, D., Cartier, K. M. S., Dressing, C. D., Dupree, A. K., Gilliland, R. L., Lillo-Box, J., Lucas, P. W., and Wang, J. (2017a). The Kepler Follow-up Observation Program. I. A Catalog of Companions to Kepler Stars from High-Resolution Imaging. , 153(2):71.

Furlan, E., Ciardi, D. R., Everett, M. E., Saylors, M., Teske, J. K., Horch, E. P., Howell, S. B., van Belle, G. T., Hirsch, L. A., Gautier, III, T. N., Adams, E. R., Barrado, D., Cartier, K. M. S., Dressing, C. D., Dupree, A. K., Gilliland, R. L., Lillo-Box, J., Lucas, P. W., and Wang, J. (2017b). The Kepler Follow-up Observation Program. I. A Catalog of Companions to Kepler Stars from High-Resolution Imaging. , 153(2):71.

Gaia Collaboration, Brown, A. G. A., Vallenari, A., Prusti, T., de Bruijne, J. H. J., Babusiaux, C., Bailer-Jones, C. A. L., Biermann, M., Evans, D. W., Eyer, L., Jansen, F., Jordi, C., Klioner, S. A., Lammers, U., Lindegren, L., Luri, X., Mignard, F., Panem, C., Pourbaix, D., Randich, S., Sartoretti, P., Siddiqui, H. I., Soubiran, C., van Leeuwen, F., Walton, N. A., Arenou, F., Bastian, U., Cropper, M., Drimmel, R., Katz, D., Lattanzi, M. G., Bakker, J., Cacciari, C., Castañeda, J., Chaoul, L., Cheek, N., De Angeli, F., Fabricius, C., Guerra, R., Holl, B., Masana, E., Messineo, R., Mowlavi, N., Nienartowicz, K., Panuzzo, P., Portell, J., Riello, M., Seabroke, G. M., Tanga, P., Thévenin, F., Gracia-Abril, G., Comoretto, G., Garcia-Reinaldos, M., Teyssier, D., Altmann, M., Andrae, R., Audard, M., Bellas-Velidis, I., Benson, K., Berthier, J., Blomme, R., Burgess, P., Busso, G., Carry, B., Cellino, A., Clementini, G., Clotet, M., Creevey, O., Davidson, M., De Ridder, J., Delchambre, L., Dell'Oro, A., Ducourant, C., Fernández-Hernández, J., Fouesneau, M., Frémat, Y., Galluccio, L., García-Torres, M., González-Núñez, J., González-Vidal, J. J., Gosset, E., Guy, L. P., Halbwachs, J. L., Hambly, N. C., Harrison, D. L., Hernández, J., Hestroffer, D., Hodgkin, S. T., Hutton, A., Jasniewicz, G., Jean-Antoine-Piccolo, A., Jordan, S., Korn, A. J., Krone-Martins, A., Lanzafame, A. C., Lebzelter, T., Löffler, W., Manteiga, M., Marrese, P. M., Martín-Fleitas, J. M., Moitinho, A., Mora, A., Muinonen, K., Osinde, J., Pancino, E., Pauwels, T., Petit, J. M., Recio-Blanco, A., Richards, P. J., Rimoldini, L., Robin, A. C., Sarro, L. M., Siopis, C., Smith, M., Sozzetti, A., Süveges, M., Torra, J., van Reeven, W., Abbas, U., Abreu Aramburu, A., Accart, S., Aerts, C., Altavilla, G., Álvarez, M. A., Alvarez, R., Alves, J., Anderson, R. I., Andrei, A. H., Anglada Varela, E., Antiche, E., Antoja, T., Arcay, B., Astraatmadja, T. L., Bach, N., Baker, S. G., Balaguer-Núñez, L., Balm, P., Barache, C., Barata, C., Barbato, D., Barblan, F., Barklem, P. S., Barrado, D., Barros, M., Barstow, M. A., Bartholomé Muñoz, S., Bassilana, J. L., Becciani, U., Bellazzini, M., Berihuete, A., Bertone, S., Bianchi, L., Bienaymé, O., Blanco-Cuaresma, S., Boch, T., Boeche, C., Bombrun, A., Borrachero, R., Bossini, D., Bouquillon, S., Bourda, G., Bragaglia, A., Bramante, L., Breddels, M. A., Bressan, A., Brouillet, N., Brüsemeister, T., Brugaletta, E., Bucciarelli, B., Burlacu, A., Busonero, D., Butkevich, A. G., Buzzi, R., Caffau, E., Cancelliere, R., Cannizzaro, G., Cantat-Gaudin, T., Carballo, R., Carlucci, T., Carrasco, J. M., Casamiquela, L., Castellani, M., Castro-Ginard, A., Charlot, P., Chemin, L., Chiavassa, A., Cocozza, G., Costigan, G., Cowell, S., Crifo, F., Crosta, M., Crowley, C., Cuypers, J., Dafonte, C., Damerdji, Y., Dapergolas, A., David, P., David, M., de Laverny, P., De Luise, F., De March, R., de Martino, D., de Souza, R., de Torres, A., Debosscher, J., del Pozo, E., Delbo, M., Delgado, A., Delgado, H. E., Di Matteo, P., Diakite, S., Diener, C., Distefano, E., Dolding, C., Drazinos, P., Durán, J., Edvardsson, B., Enke, H., Eriksson, K., Esquej, P., Eynard Bontemps, G., Fabre, C., Fabrizio, M., Faigler, S., Falcão, A. J., Farràs Casas, M., Federici, L., Fedorets, G., Fernique, P., Figueras, F., Filippi, F., Findeisen, K., Fonti, A., Fraile, E., Fraser, M., Frézouls, B., Gai, M., Galleti, S., Garabato, D., García-Sedano, F., Garofalo, A., Garralda, N., Gavel, A., Gavras, P., Gerssen, J., Geyer, R., Giacobbe, P., Gilmore, G., Girona, S., Giuffrida, G., Glass, F., Gomes, M., Granvik, M., Gueguen, A., Guerrier, A., Guiraud, J., Gutiérrez-Sánchez, R., Haigron, R., Hatzidimitriou, D., Hauser, M., Haywood, M., Heiter, U., Helmi, A., Heu, J., Hilger, T., Hobbs, D., Hofmann, W., Holland, G., Huckle, H. E., Hypki, A., Icardi, V., Janßen, K., Jevardat de Fombelle, G., Jonker, P. G., Juhász, Á. L., Julbe, F., Karampelas, A., Kewley, A., Klar, J., Kochoska, A., Kohley, R., Kolenberg, K., Kontizas, M., Kontizas, E., Koposov, S. E., Kordopatis, G., Kostrzewa-Rutkowska, Z., Koubsky, P., Lambert, S., Lanza, A. F., Lasne, Y., Lavigne, J. B., Le Fustec, Y., Le Poncin-Lafitte, C., Lebreton, Y., Leccia, S., Leclerc, N., Lecoeur-Taibi, I., Lenhardt, H., Leroux, F., Liao, S., Licata, E., Lindstrøm, H. E. P., Lister, T. A., Livanou, E., Lobel, A., López, M., Managau, S., Mann, R. G., Mantelet, G., Marchal, O., Marchant, J. M., Marconi, M., Marinoni, S., Marschalkó, G., Marshall, D. J., Martino, M., Marton, G., Mary, N., Massari, D., Matijevič, G., Mazeh, T., McMillan, P. J., Messina, S., Michalik, D., Millar, N. R., Molina, D., Molinaro, R., Molnár, L., Montegriffo, P., Mor, R., Morbidelli, R., Morel, T., Morris, D., Mulone, A. F., Muraveva, T., Musella, I., Nelemans, G., Nicastro, L., Noval, L., O'Mullane, W., Ordénovic, C., Ordóñez-Blanco, D., Osborne, P., Pagani, C., Pagano, I., Pailler, F., Palacin, H., Palaversa, L., Panahi, A., Pawlak, M., Piersimoni, A. M., Pineau, F. X., Plachy, E., Plum, G., Poggio, E., Poujoulet, E., Prša, A., Pulone, L., Racero, E., Ragaini, S., Rambaux, N., Ramos-Lerate, M., Regibo, S., Reylé, C., Riclet, F., Ripepi, V., Riva, A., Rivard, A., Rixon, G., Roegiers, T., Roelens, M., Romero-Gómez, M., Rowell, N., Royer, F., Ruiz-Dern, L., Sadowski, G., Sagristà Sellés, T., Sahlmann, J., Salgado, J., Salguero, E., Sanna, N., Santana-Ros, T., Sarasso, M., Savietto, H., Schultheis, M., Sciacca, E., Segol, M., Segovia, J. C., Ségransan, D., Shih, I. C., Siltala, L., Silva, A. F., Smart, R. L., Smith, K. W., Solano, E., Solitro, F., Sordo, R., Soria Nieto, S., Souchay, J., Spagna, A., Spoto, F., Stampa, U., Steele, I. A., Steidelmüller, H., Stephenson, C. A., Stoev, H., Suess, F. F., Surdej, J., Szabados, L., Szegedi-Elek, E., Tapiador, D., Taris, F., Tauran, G., Taylor, M. B., Teixeira, R., Terrett, D., Teyssandier, P., Thuillot, W., Titarenko, A., Torra Clotet, F., Turon, C., Ulla, A., Utrilla, E., Uzzi, S., Vaillant, M., Valentini, G., Valette, V., van Elteren, A., Van Hemelryck, E., van Leeuwen, M., Vaschetto, M., Vecchiato, A., Veljanoski, J., Viala, Y., Vicente, D., Vogt, S., von Essen, C., Voss, H., Votruba, V., Voutsinas, S., Walmsley, G., Weiler, M., Wertz, O., Wevers, T., Wyrzykowski, Ł., Yoldas, A., Žerjal, M., Ziaeepour, H., Zorec, J., Zschocke, S., Zucker, S., Zurbach, C., and Zwitter, T. (2018). Gaia Data Release 2. Summary of the contents and survey properties. , 616:A1.

Gaia Collaboration, Vallenari, A., Brown, A. G. A., Prusti, T., de Bruijne, J. H. J., Arenou, F., Babusiaux, C., Biermann, M., Creevey, O. L., Ducourant, C., Evans, D. W., Eyer, L., Guerra, R., Hutton, A., Jordi, C., Klioner, S. A., Lammers, U. L., Lindegren, L., Luri, X., Mignard, F., Panem, C., Pourbaix, D., Randich, S., Sartoretti, P., Soubiran, C., Tanga, P., Walton, N. A., Bailer-Jones, C. A. L., Bastian, U., Drimmel, R., Jansen, F., Katz, D., Lattanzi, M. G., van Leeuwen, F., Bakker, J., Cacciari, C., Castañeda, J., De Angeli, F., Fabricius, C., Fouesneau, M., Frémat, Y., Galluccio, L., Guerrier, A., Heiter, U., Masana, E., Messineo, R., Mowlavi, N., Nicolas, C., Nienartowicz, K., Pailler, F., Panuzzo, P., Riclet, F., Roux, W., Seabroke, G. M., Sordo, R., Thévenin, F., Gracia-Abril, G., Portell, J., Teyssier, D., Altmann, M., Andrae, R., Audard, M., Bellas-Velidis, I., Benson, K., Berthier, J., Blomme, R., Burgess, P. W., Busonero, D., Busso, G., Cánovas, H., Carry, B., Cellino, A., Cheek, N., Clementini, G., Damerdji, Y., Davidson, M., de Teodoro, P., Nuñez Campos, M., Delchambre, L., Dell'Oro, A., Esquej, P., Fernández-Hernández, J., Fraile, E., Garabato, D., García-Lario, P., Gosset, E., Haigron, R., Halbwachs, J. L., Hambly, N. C., Harrison, D. L., Hernández, J., Hestroffer, D., Hodgkin, S. T., Holl, B., Janßen, K., Jevardat de Fombelle, G., Jordan, S., Krone-Martins, A., Lanzafame, A. C., Löffler, W., Marchal, O., Marrese, P. M., Moitinho, A., Muinonen, K., Osborne, P., Pancino, E., Pauwels, T., Recio-Blanco, A., Reylé, C., Riello, M., Rimoldini, L., Roegiers, T., Rybizki, J., Sarro, L. M., Siopis, C., Smith, M., Sozzetti, A., Utrilla, E., van Leeuwen, M., Abbas, U., Ábrahám, P., Abreu Aramburu, A., Aerts, C., Aguado, J. J., Ajaj, M., Aldea-Montero, F., Altavilla, G., Álvarez, M. A., Alves, J., Anders, F., Anderson, R. I., Anglada Varela, E., Antoja, T., Baines, D., Baker, S. G., Balaguer-Núñez, L., Balbinot, E., Balog, Z., Barache, C., Barbato, D., Barros, M., Barstow, M. A., Bartolomé, S., Bassilana, J. L., Bauchet, N., Becciani, U., Bellazzini, M., Berihuete, A., Bernet, M., Bertone, S., Bianchi, L., Binnenfeld, A., Blanco-Cuaresma, S., Blazere, A., Boch, T., Bombrun, A., Bossini, D., Bouquillon, S., Bragaglia, A., Bramante, L., Breedt, E., Bressan, A., Brouillet, N., Brugaletta, E., Bucciarelli, B., Burlacu, A., Butkevich, A. G., Buzzi, R., Caffau, E., Cancelliere, R., Cantat-Gaudin, T., Carballo, R., Carlucci, T., Carnerero, M. I., Carrasco, J. M., Casamiquela, L., Castellani, M., Castro-Ginard, A., Chaoul, L., Charlot, P., Chemin, L., Chiaramida, V., Chiavassa, A., Chornay, N., Comoretto, G., Contursi, G., Cooper, W. J., Cornez, T., Cowell, S., Crifo, F., Cropper, M., Crosta, M., Crowley, C., Dafonte, C., Dapergolas, A., David, M., David, P., de Laverny, P., De Luise, F., De March, R., De Ridder, J., de Souza, R., de Torres, A., del Peloso, E. F., del Pozo, E., Delbo, M., Delgado, A., Delisle, J. B., Demouchy, C., Dharmawardena, T. E., Di Matteo, P., Diakite, S., Diener, C., Distefano, E., Dolding, C., Edvardsson, B., Enke, H., Fabre, C., Fabrizio, M., Faigler, S., Fedorets, G., Fernique, P., Fienga, A., Figueras, F., Fournier, Y., Fouron, C., Fragkoudi, F., Gai, M., Garcia-Gutierrez, A., Garcia-Reinaldos, M., García-Torres, M., Garofalo, A., Gavel, A., Gavras, P., Gerlach, E., Geyer, R., Giacobbe, P., Gilmore, G., Girona, S., Giuffrida, G., Gomel, R., Gomez, A., González-Núñez, J., González-Santamaría, I., González-Vidal, J. J., Granvik, M., Guillout, P., Guiraud, J., Gutiérrez-Sánchez, R., Guy, L. P., Hatzidimitriou, D., Hauser, M., Haywood, M., Helmer, A., Helmi, A., Sarmiento, M. H., Hidalgo, S. L., Hilger, T., Hładczuk, N., Hobbs, D., Holland, G., Huckle, H. E., Jardine, K., Jasniewicz, G., Jean-Antoine Piccolo, A., Jiménez-Arranz, Ó., Jorissen, A., Juaristi Campillo, J., Julbe, F., Karbevska, L., Kervella, P., Khanna, S., Kontizas, M., Kordopatis, G., Korn, A. J., Kóspál, Á., Kostrzewa-Rutkowska, Z., Kruszyńska, K., Kun, M., Laizeau, P., Lambert, S., Lanza, A. F., Lasne, Y., Le Campion, J. F., Lebreton, Y., Lebzelter, T., Leccia, S., Leclerc, N., Lecoeur-Taibi, I., Liao, S., Licata, E. L., Lindstrøm, H. E. P., Lister, T. A., Livanou, E., Lobel, A., Lorca, A., Loup, C., Madrero Pardo, P., Magdaleno Romeo, A., Managau, S., Mann, R. G., Manteiga, M., Marchant, J. M., Marconi, M., Marcos, J., Marcos Santos, M. M. S., Marín Pina, D., Marinoni, S., Marocco, F., Marshall, D. J., Martin Polo, L., Martín-Fleitas, J. M., Marton, G., Mary, N., Masip, A., Massari, D., Mastrobuono-Battisti, A., Mazeh, T., McMillan, P. J., Messina, S., Michalik, D., Millar, N. R., Mints, A., Molina, D., Molinaro, R., Molnár, L., Monari, G., Monguió, M., Montegriffo, P., Montero, A., Mor, R., Mora, A., Morbidelli, R., Morel, T., Morris, D., Muraveva, T., Murphy, C. P., Musella, I., Nagy, Z., Noval, L., Ocaña, F., Ogden, A., Ordenovic, C., Osinde, J. O., Pagani, C., Pagano, I., Palaversa, L., Palicio, P. A., Pallas-Quintela, L., Panahi, A., Payne-Wardenaar, S., Peñalosa Esteller, X., Penttilä, A., Pichon, B., Piersimoni, A. M., Pineau, F. X., Plachy, E., Plum, G., Poggio, E., Prša, A., Pulone, L., Racero, E., Ragaini, S., Rainer, M., Raiteri, C. M., Rambaux, N., Ramos, P., Ramos-Lerate, M., Re Fiorentin, P., Regibo, S., Richards, P. J., Rios Diaz, C., Ripepi, V., Riva, A., Rix, H. W., Rixon, G., Robichon, N., Robin, A. C., Robin, C., Roelens, M., Rogues, H. R. O., Rohrbasser, L., Romero-Gómez, M., Rowell, N., Royer, F., Ruz Mieres, D., Rybicki, K. A., Sadowski, G., Sáez Núñez, A., Sagristà Sellés, A., Sahlmann, J., Salguero, E., Samaras, N., Sanchez Gimenez, V., Sanna, N., Santoveña, R., Sarasso, M., Schultheis, M., Sciacca, E., Segol, M., Segovia, J. C., Ségransan, D., Semeux, D., Shahaf, S., Siddiqui, H. I., Siebert, A., Siltala, L., Silvelo, A., Slezak, E., Slezak, I., Smart, R. L., Snaith, O. N., Solano, E., Solitro, F., Souami, D., Souchay, J., Spagna, A., Spina, L., Spoto, F., Steele, I. A., Steidelmüller, H., Stephenson, C. A., Süveges, M., Surdej, J., Szabados, L., Szegedi-Elek, E., Taris, F., Taylor, M. B., Teixeira, R., Tolomei, L., Tonello, N., Torra, F., Torra, J., Torralba Elipe, G., Trabucchi, M., Tsounis, A. T., Turon, C., Ulla, A., Unger, N., Vaillant, M. V., van Dillen, E., van Reeven, W., Vanel, O., Vecchiato, A., Viala, Y., Vicente, D., Voutsinas, S., Weiler, M., Wevers, T., Wyrzykowski, Ł., Yoldas, A., Yvard, P., Zhao, H., Zorec, J., Zucker, S., and Zwitter, T. (2023). Gaia Data Release 3. Summary of the content and survey properties. , 674:A1.

gaiaspec (2024). gaiaspec v1.0.0: Derive stellar properties for stars with gaia rvs spectra. Zenodo.

Goldreich, P. and Sari, R. (2003). Eccentricity Evolution for Planets in Gaseous Disks. , 585(2):1024–1037.

Hansen, B. M. S. and Naoz, S. (2020). The stationary points of the hierarchical three-body problem. , 499(2):1682–1700.

Haywood, R. D., Collier Cameron, A., Queloz, D., Barros, S. C. C., Deleuil, M., Fares, R., Gillon, M., Lanza, A. F., Lovis, C., Moutou, C., Pepe, F., Pollacco, D., Santerne, A., Ségransan, D., and Unruh, Y. C. (2014). Planets and stellar activity: hide and seek in the CoRoT-7 system. , 443(3):2517–2531.

Hebb, L., Collier-Cameron, A., Triaud, A. H. M. J., Lister, T. A., Smalley, B., Maxted, P. F. L., Hellier, C., Anderson, D. R., Pollacco, D., Gillon, M., Queloz, D., West, R. G., Bentley, S., Enoch, B., Haswell, C. A., Horne, K., Mayor, M., Pepe, F., Segransan, D., Skillen, I., Udry, S., and Wheatley, P. J. (2010). WASP-19b: The Shortest Period Transiting Exoplanet Yet Discovered. , 708(1):224–231.

Hirsch, L. A., Rosenthal, L., Fulton, B. J., Howard, A. W., Ciardi, D. R., Marcy, G. W., Nielsen, E., Petigura, E. A., de Rosa, R. J., Isaacson, H., Weiss, L. M., Sinukoff, E., and Macintosh, B. (2021). Understanding the Impacts of Stellar Companions on Planet Formation and Evolution: A Survey of Stellar and Planetary Companions within 25 pc. , 161(3):134.

Ho, A. Y. Q., Ness, M. K., Hogg, D. W., Rix, H.-W., Liu, C., Yang, F., Zhang, Y., Hou, Y., and Wang, Y. (2017). Label Transfer from APOGEE to LAMOST: Precise Stellar Parameters for 450,000 LAMOST Giants. , 836(1):5.

Horch, E. P., Howell, S. B., Everett, M. E., and Ciardi, D. R. (2014). Most Sub-arcsecond Companions of Kepler Exoplanet Candidate Host Stars are Gravitationally Bound. , 795(1):60.

Howard, A. W., Johnson, J. A., Marcy, G. W., Fischer, D. A., Wright, J. T., Bernat, D., Henry, G. W., Peek, K. M. G., Isaacson, H., Apps, K., Endl, M., Cochran, W. D., Valenti, J. A., Anderson, J., and Piskunov, N. E. (2010a). The California Planet Survey. I. Four New Giant Exoplanets. , 721(2):1467–1481.

Howard, A. W., Johnson, J. A., Marcy, G. W., Fischer, D. A., Wright, J. T., Bernat, D., Henry, G. W., Peek, K. M. G., Isaacson, H., Apps, K., Endl, M., Cochran, W. D., Valenti, J. A., Anderson, J., and Piskunov, N. E. (2010b). The California Planet Survey. I. Four New Giant Exoplanets. , 721(2):1467–1481.

Howard, A. W., Marcy, G. W., Johnson, J. A., Fischer, D. A., Wright, J. T., Isaacson, H., Valenti, J. A., Anderson, J., Lin, D. N. C., and Ida, S. (2010c). The Occurrence and Mass Distribution of Close-in Super-Earths, Neptunes, and Jupiters. *Science*, 330(6004):653.

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science and Engineering*, 9(3):90–95.

Hurley, J. R., Pols, O. R., and Tout, C. A. (2000). Comprehensive analytic formulae for stellar evolution as a function of mass and metallicity. , 315(3):543–569.

Husser, T. O., Wende-von Berg, S., Dreizler, S., Homeier, D., Reiners, A., Barman, T., and Hauschildt, P. H. (2013). A new extensive library of PHOENIX stellar atmospheres and synthetic spectra. , 553:A6.

Innanen, K. A., Zheng, J. Q., Mikkola, S., and Valtonen, M. J. (1997). The Kozai Mechanism and the Stability of Planetary Orbits in Binary Star Systems. , 113:1915.

Isaacson, H. and Fischer, D. (2010). Chromospheric Activity and Jitter Measurements for 2630 Stars on the California Planet Search. , 725(1):875–885.

Jang-Condell, H. (2015). On the Likelihood of Planet Formation in Close Binaries. , 799(2):147.

Johnson, J. A., Petigura, E. A., Fulton, B. J., Marcy, G. W., Howard, A. W., Isaacson, H., Hebb, L., Cargile, P. A., Morton, T. D., Weiss, L. M., Winn, J. N., Rogers, L. A., Sinukoff, E., and Hirsch, L. A. (2017a). The California-Kepler Survey. II. Precise Physical Properties of 2025 Kepler Planets and Their Host Stars. , 154(3):108.

Johnson, J. A., Petigura, E. A., Fulton, B. J., Marcy, G. W., Howard, A. W., Isaacson, H., Hebb, L., Cargile, P. A., Morton, T. D., Weiss, L. M., Winn, J. N., Rogers, L. A., Sinukoff, E., and Hirsch, L. A. (2017b). The California-Kepler Survey. II. Precise Physical Properties of 2025 Kepler Planets and Their Host Stars. , 154(3):108.

Jones, E., Oliphant, T., and Peterson, P. (2001). Scipy: Open source scientific tools for python. https://scipy.org/ [Accessed: Whenever].

Jura, M., Muno, M. P., Farihi, J., and Zuckerman, B. (2009). X-Ray and Infrared Observations of Two Externally Polluted White Dwarfs. , 699(2):1473–1479.

Katz, D., Sartoretti, P., Guerrier, A., Panuzzo, P., Seabroke, G. M., Thévenin, F., Cropper, M., Benson, K., Blomme, R., Haigron, R., Marchal, O., Smith, M., Baker, S., Chemin, L., Damerdji, Y., David, M., Dolding, C., Frémat, Y., Gosset, E., Janßen, K., Jasniewicz, G., Lobel, A., Plum, G., Samaras, N., Snaith, O., Soubiran, C., Vanel, O., Zwitter, T., Antoja, T., Arenou, F., Babusiaux, C., Brouillet, N., Caffau, E., Di Matteo, P., Fabre, C., Fabricius, C., Fragkoudi, F., Haywood, M., Huckle, H. E., Hottier, C., Lasne, Y., Leclerc, N., Mastrobuono-Battisti, A., Royer, F., Teyssier, D., Zorec, J., Crifo, F., Jean-Antoine Piccolo, A., Turon, C., and Viala, Y. (2023). Gaia Data Release 3. Properties and validation of the radial velocities. , 674:A5.

Kolbl, R., Marcy, G. W., Isaacson, H., and Howard, A. W. (2015). Detection of Stars Within ∼0.8 in of Kepler Objects of Interest. , 149(1):18.

Kozai, Y. (1962). Secular perturbations of asteroids with high inclination and eccentricity. , 67:591–598.

Kraus, A. L., Ireland, M. J., Huber, D., Mann, A. W., and Dupuy, T. J. (2016). The Impact of Stellar Multiplicity on Planetary Systems. I. The Ruinous Influence of Close Binary Companions. , 152(1):8.

Kurucz, R. L. (1979). Model atmospheres for G, F, A, B, and O stars. , 40:1–340.

Lai, D., Anderson, K. R., and Pu, B. (2018). How do external companions affect spin-orbit misalignment of hot Jupiters? , 475(4):5231–5236.

Lanzafame, A. C., Brugaletta, E., Frémat, Y., Sordo, R., Creevey, O. L., Andretta, V., Scandariato, G., Busà, I., Distefano, E., Korn, A. J., de Laverny, P., Recio-Blanco, A., Abreu Aramburu, A., Álvarez, M. A., Andrae, R., Bailer-Jones, C. A. L., Bakker, J., Bellas-Velidis, I., Bijaoui, A., Brouillet, N., Burlacu, A., Carballo, R., Casamiquela, L., Chaoul, L., Chiavassa, A., Contursi, G., Cooper, W. J., Dafonte, C., Dapergolas, A., Delchambre, L., Demouchy, C., Dharmawardena, T. E., Drimmel, R., Edvardsson, B., Fouesneau, M., Garabato, D., García-Lario, P., García-Torres, M., Gavel, A., Gomez, A., González-Santamaría, I., Hatzidimitriou, D., Heiter, U., Jean-Antoine Piccolo, A., Kontizas, M., Kordopatis, G., Lebreton, Y., Licata, E. L., Lindstrøm, H. E. P., Livanou, E., Lobel, A., Lorca, A., Magdaleno Romeo, A., Manteiga, M., Marocco, F., Marshall, D. J., Mary, N., Nicolas, C., Ordenovic, C., Pailler, F., Palicio, P. A., Pallas-Quintela, L., Panem, C., Pichon, B., Poggio, E., Riclet, F., Robin, C., Rybizki, J., Santoveña, R., Sarro, L. M., Schultheis, M. S., Segol, M., Silvelo, A., Slezak, I., Smart, R. L., Soubiran, C., Süveges, M., Thévenin, F., Torralba Elipe, G., Ulla, A., Utrilla, E., Vallenari, A., van Dillen, E., Zhao, H., and Zorec, J. (2023). Gaia Data Release 3. Stellar chromospheric activity and mass accretion from Ca II IRT observed by the Radial Velocity Spectrometer. , 674:A30.

Lee, E. J. and Chiang, E. (2015). To Cool is to Accrete: Analytic Scalings for Nebular Accretion of Planetary Atmospheres. , 811(1):41.

Lee, E. J. and Chiang, E. (2016). Breeding Super-Earths and Birthing Super-puffs in Transitional Disks. , 817(2):90.

Lee, E. J., Chiang, E., and Ormel, C. W. (2014). Make Super-Earths, Not Jupiters: Accreting Nebular Gas onto Solid Cores at 0.1 AU and Beyond. , 797(2):95.

Li, G., Naoz, S., Holman, M., and Loeb, A. (2014a). Chaos in the Test Particle Eccentric Kozai-Lidov Mechanism. , 791(2):86.

Li, G., Naoz, S., Kocsis, B., and Loeb, A. (2014b). Eccentricity Growth and Orbit Flip in Near-coplanar Hierarchical Three-body Systems. , 785(2):116.

Li, G., Naoz, S., Valsecchi, F., Johnson, J. A., and Rasio, F. A. (2014c). The Dynamics of the Multi-planet System Orbiting Kepler-56. , 794(2):131.

Lidov, M. L. (1962). The evolution of orbits of artificial satellites of planets under the action of gravitational perturbations of external bodies. , 9(10):719–759.

Lithwick, Y. and Wu, Y. (2014). Secular chaos and its application to Mercury, hot Jupiters, and the organization of planetary systems. *Proceedings of the National Academy of Science*, 111(35):12610–12615.

Maciejewski, G., Dimitrov, D., Fernández, M., Sota, A., Nowak, G., Ohlert, J., Nikolov, G., Bukowiecki, Ł., Hinse, T. C., Pallé, E., Tingley, B., Kjurkchieva, D., Lee, J. W., and Lee, C. U. (2016). Departure from the constant-period ephemeris for the transiting exoplanet WASP-12. , 588:L6.

Majewski, S. R., Schiavon, R. P., Frinchaboy, P. M., Allende Prieto, C., Barkhouser, R., Bizyaev, D., Blank, B., Brunner, S., Burton, A., Carrera, R., Chojnowski, S. D., Cunha, K., Epstein, C., Fitzgerald, G., García Pérez, A. E., Hearty, F. R., Henderson, C., Holtzman, J. A., Johnson, J. A., Lam, C. R., Lawler, J. E., Maseman, P., Mészáros, S., Nelson, M., Nguyen, D. C., Nidever, D. L., Pinsonneault, M., Shetrone, M., Smee, S., Smith, V. V., Stolberg, T., Skrutskie, M. F., Walker, E., Wilson, J. C., Zasowski, G., Anders, F., Basu, S., Beland, S., Blanton, M. R., Bovy, J., Brownstein, J. R., Carlberg, J., Chaplin, W., Chiappini, C., Eisenstein, D. J., Elsworth, Y., Feuillet, D., Fleming, S. W., Galbraith-Frew, J., García, R. A., García-Hernández, D. A., Gillespie, B. A., Girardi, L., Gunn, J. E., Hasselquist, S., Hayden, M. R., Hekker, S., Ivans, I., Kinemuchi, K., Klaene, M., Mahadevan, S., Mathur, S., Mosser, B., Muna, D., Munn, J. A., Nichol, R. C., O'Connell, R. W., Parejko, J. K., Robin, A. C., Rocha-Pinto, H., Schultheis, M., Serenelli, A. M., Shane, N., Silva Aguirre, V., Sobeck, J. S., Thompson, B., Troup, N. W., Weinberg, D. H., and Zamora, O. (2017). The Apache Point Observatory Galactic Evolution Experiment (APOGEE). , 154(3):94.

Masuda, K. (2017). Eccentric Companions to Kepler-448b and Kepler-693b: Clues to the Formation of Warm Jupiters. , 154(2):64.

Matson, R. A., Howell, S. B., Horch, E. P., and Everett, M. E. (2018). Stellar Companions of Exoplanet Host Stars in K2. , 156(1):31.

Mayor, M., Marmier, M., Lovis, C., Udry, S., Ségransan, D., Pepe, F., Benz, W., Bertaux, J. L., Bouchy, F., Dumusque, X., Lo Curto, G., Mordasini, C., Queloz, D., and Santos, N. C. (2011). The HARPS search for southern extra-solar planets XXXIV. Occurrence, mass distribution and orbital properties of super-Earths and Neptune-mass planets. *arXiv e-prints*, page arXiv:1109.2497.

Mayor, M. and Queloz, D. (1995). A Jupiter-mass companion to a solar-type star. , 378(6555):355–359.

McKinney, W. (2010). Data Structures for Statistical Computing in Python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 56 – 61.

Metzger, B. D., Giannios, D., and Spiegel, D. S. (2012). Optical and X-ray transients from planet-star mergers. , 425(4):2778–2798.

Millholland, S., Petigura, E., and Batygin, K. (2020). Tidal Inflation Reconciles Low-density Sub-Saturns with Core Accretion. , 897(1):7.

Morbidelli, A., Lambrechts, M., Jacobson, S., and Bitsch, B. (2015). The great dichotomy of the Solar System: Small terrestrial embryos and massive giant planet cores. , 258:418–429.

Moutou, C., Almenara, J. M., Hébrard, G., Santos, N. C., Stassun, K. G., Deheuvels, S., Barros, S., Benni, P., Bieryla, A., Boisse, I., Bonfils, X., Boyd, P. T., Collins, K. A., Baker, D., Cortés-Zuleta, P., Dalal, S., Debras, F., Deleuil, M., Delfosse, X., Demangeon, O., Essack, Z., Forveille, T., Girardin, E., Guerra, P., Heidari, N., Hesse, K., Hoyer, S., Jenkins, J. M., Kiefer, F., König, P. C., Laloum, D., Latham, D., Lopez, T., Martioli, E., Osborn, H. P., Ricker, G., Seager, S., Vanderspek, R., Vezie, M., Villaseñor, J., Winn, J., Wohler, B., and Ziegler, C. (2021). TOI-1296b and TOI-1298b observed with TESS and SOPHIE: two hot Saturn-mass exoplanets with different densities around metal-rich stars. , 653:A147.

Moutou, C., Mayor, M., Lo Curto, G., Udry, S., Bouchy, F., Benz, W., Lovis, C., Naef, D., Pepe, F., Queloz, D., and Santos, N. C. (2009). The HARPS search for southern extra-solar planets. XV. Six long-period giant planets around BD -17 0063, HD 20868, HD 73267, HD 131664, HD 145377, and HD 153950. , 496(2):513–519.

Mustill, A. J., Davies, M. B., Blunt, S., and Howard, A. (2021). Dynamical orbital evolution scenarios of the wide-orbit eccentric planet HR 5183b. *arXiv e-prints*, page arXiv:2102.06031.

Naef, D., Mayor, M., Lo Curto, G., Bouchy, F., Lovis, C., Moutou, C., Benz, W., Pepe, F., Queloz, D., Santos, N. C., Ségransan, D., Udry, S., Bonfils, X., Delfosse, X., Forveille, T., Hébrard, G., Mordasini, C., Perrier, C., Boisse, I., and Sosnowska, D. (2010). The HARPS search for southern extrasolar planets. XXIII. 8 planetary companions to low-activity solar-type stars. , 523:A15.

Naoz, S. (2016). The Eccentric Kozai-Lidov Effect and Its Applications. , 54:441–489.

Naoz, S., Farr, W. M., Lithwick, Y., Rasio, F. A., and Teyssandier, J. (2011). Hot Jupiters from secular planet-planet interactions. , 473(7346):187–189.

Naoz, S., Farr, W. M., Lithwick, Y., Rasio, F. A., and Teyssandier, J. (2013a). Secular dynamics in hierarchical three-body systems. , 431(3):2155–2171.

Naoz, S., Farr, W. M., and Rasio, F. A. (2012). On the Formation of Hot Jupiters in Stellar Binaries. , 754(2):L36.

Naoz, S., Fragos, T., Geller, A., Stephan, A. P., and Rasio, F. A. (2016). Formation of Black Hole Low-mass X-Ray Binaries in Hierarchical Triple Systems. , 822(2):L24.

Naoz, S., Kocsis, B., Loeb, A., and Yunes, N. (2013b). Resonant Post-Newtonian Eccentricity Excitation in Hierarchical Three-body Systems. , 773(2):187.

Naoz, S. and Silk, J. (2014). Formation of Dark Matter Tori around Supermassive Black Holes via the Eccentric Kozai-Lidov Mechanism. , 795(2):102.

Narita, N., Sato, B., Hirano, T., and Tamura, M. (2009). First Evidence of a Retrograde Orbit of a Transiting Exoplanet HAT-P-7b. , 61:L35–L40.

Ness, M., Hogg, D. W., Rix, H. W., Ho, A. Y. Q., and Zasowski, G. (2015). The Cannon: A data-driven approach to Stellar Label Determination. , 808(1):16.

Ness, M., Hogg, D. W., Rix, H. W., Martig, M., Pinsonneault, M. H., and Ho, A. Y. Q. (2016). Spectroscopic Determination of Masses (and Implied Ages) for Red Giants. , 823(2):114.

Ngo, H., Knutson, H. A., Hinkley, S., Bryan, M., Crepp, J. R., Batygin, K., Crossfield, I., Hansen, B., Howard, A. W., Johnson, J. A., Mawet, D., Morton, T. D., Muirhead, P. S., and Wang, J. (2016). Friends of Hot Jupiters. IV. Stellar Companions Beyond 50 au Might Facilitate Giant Planet Formation, but Most are Unlikely to Cause Kozai-Lidov Migration. , 827(1):8.

Ngo, H., Knutson, H. A., Hinkley, S., Crepp, J. R., Bechter, E. B., Batygin, K., Howard, A. W., Johnson, J. A., Morton, T. D., and Muirhead, P. S. (2015). Friends of Hot Jupiters. II. No Correspondence between Hot-jupiter Spin-Orbit Misalignment and the Incidence of Directly Imaged Stellar Companions. , 800(2):138.

Pecaut, M. J. and Mamajek, E. E. (2013a). Intrinsic Colors, Temperatures, and Bolometric Corrections of Pre-main-sequence Stars. , 208(1):9.

Pecaut, M. J. and Mamajek, E. E. (2013b). Intrinsic Colors, Temperatures, and Bolometric Corrections of Pre-main-sequence Stars. , 208(1):9.

Petigura, E. A. (2020). Two Views of the Radius Gap and the Role of Light Curve Fitting. , 160(2):89.

Petigura, E. A., Benneke, B., Batygin, K., Fulton, B. J., Werner, M., Krick, J. E., Gorjian, V., Sinukoff, E., Deck, K. M., Mills, S. M., and Deming, D. (2018a). Dynamics and Formation of the Near-resonant K2-24 System: Insights from Transit-timing Variations and Radial Velocities. , 156(3):89.

Petigura, E. A., Howard, A. W., Marcy, G. W., Johnson, J. A., Isaacson, H., Cargile, P. A., Hebb, L., Fulton, B. J., Weiss, L. M., Morton, T. D., Winn, J. N., Rogers, L. A., Sinukoff, E., Hirsch, L. A., and Crossfield, I. J. M. (2017a). The California-Kepler Survey. I. High-resolution Spectroscopy of 1305 Stars Hosting Kepler Transiting Planets. , 154(3):107.

Petigura, E. A., Howard, A. W., Marcy, G. W., Johnson, J. A., Isaacson, H., Cargile, P. A., Hebb, L., Fulton, B. J., Weiss, L. M., Morton, T. D., Winn, J. N., Rogers, L. A., Sinukoff, E., Hirsch, L. A., and Crossfield, I. J. M. (2017b). The California-Kepler Survey. I. High-resolution Spectroscopy of 1305 Stars Hosting Kepler Transiting Planets. , 154(3):107.

Petigura, E. A., Marcy, G. W., Winn, J. N., Weiss, L. M., Fulton, B. J., Howard, A. W., Sinukoff, E., Isaacson, H., Morton, T. D., and Johnson, J. A. (2018b). The California-Kepler Survey. IV. Metal-rich Stars Host a Greater Diversity of Planets. , 155(2):89.

Petigura, E. A., Marcy, G. W., Winn, J. N., Weiss, L. M., Fulton, B. J., Howard, A. W., Sinukoff, E., Isaacson, H., Morton, T. D., and Johnson, J. A. (2018c). The California-Kepler Survey. IV. Metal-rich Stars Host a Greater Diversity of Planets. , 155(2):89.

Petigura, E. A., Rogers, J. G., Isaacson, H., Owen, J. E., Kraus, A. L., Winn, J. N., MacDougall, M. G., Howard, A. W., Fulton, B., Kosiarek, M. R., Weiss, L. M., Behmard, A., and Blunt, S. (2022). The California-Kepler Survey. X. The Radius Gap as a Function of Stellar Mass, Metallicity, and Age. , 163(4):179.

Petigura, E. A., Sinukoff, E., Lopez, E. D., Crossfield, I. J. M., Howard, A. W., Brewer, J. M., Fulton, B. J., Isaacson, H. T., Ciardi, D. R., Howell, S. B., Everett, M. E., Horch, E. P., Hirsch, L. A., Weiss, L. M., and Schlieder, J. E. (2017c). Four Sub-Saturns with Dissimilar Densities: Windows into Planetary Cores and Envelopes. , 153(4):142.

Petrovich, C. (2015). Steady-state Planet Migration by the Kozai-Lidov Mechanism in Stellar Binaries. , 799(1):27.

Petrovich, C., Muñoz, D. J., Kratter, K. M., and Malhotra, R. (2020). A Disk-driven Resonance as the Origin of High Inclinations of Close-in Planets. , 902(1):L5.

Petrovich, C. and Tremaine, S. (2016). Warm Jupiters from Secular Planet-Planet Interactions. , 829(2):132.

Pollack, J. B., Hubickyj, O., Bodenheimer, P., Lissauer, J. J., Podolak, M., and Greenzweig, Y. (1996). Formation of the Giant Planets by Concurrent Accretion of Solids and Gas. , 124(1):62–85.

Quintana, E. V., Adams, F. C., Lissauer, J. J., and Chambers, J. E. (2007). Terrestrial Planet Formation around Individual Stars within Binary Star Systems. , 660(1):807–822.

Qureshi, A., Naoz, S., and Shkolnik, E. L. (2018a). Signature of Planetary Mergers on Stellar Spins. , 864(1):65.

Qureshi, A., Naoz, S., and Shkolnik, E. L. (2018b). Signature of Planetary Mergers on Stellar Spins. , 864(1):65.

Raghavan, D., McAlister, H. A., Henry, T. J., Latham, D. W., Marcy, G. W., Mason, B. D., Gies, D. R., White, R. J., and ten Brummelaar, T. A. (2010). A Survey of Stellar Families: Multiplicity of Solar-type Stars. , 190(1):1–42.

Ragusa, E., Rosotti, G., Teyssandier, J., Booth, R., Clarke, C. J., and Lodato, G. (2018). Eccentricity evolution during planet-disc interaction. , 474(4):4460–4476.

Rampalli, R., Ness, M., and Wylie, S. (2021). The Astrophysical Variance in Gaia-Radial Velocity Spectrometer Spectra. , 921(1):78.

Rampalli, R., Ness, M. K., Edwards, G. H., Newton, E. R., and Bedell, M. (2024). The Sun Remains Relatively Refractory Depleted: Elemental Abundances for 17,412 Gaia RVS Solar Analogs and 50 Planet Hosts. , 965(2):176.

Rasio, F. A. and Ford, E. B. (1996). Dynamical instabilities and the formation of extrasolar planetary systems. *Science*, 274:954–956.

Recio-Blanco, A., de Laverny, P., Palicio, P. A., Kordopatis, G., Álvarez, M. A., Schultheis, M., Contursi, G., Zhao, H., Torralba Elipe, G., Ordenovic, C., Manteiga, M., Dafonte, C., Oreshina-Slezak, I., Bijaoui, A., Frémat, Y., Seabroke, G., Pailler, F., Spitoni, E., Poggio, E., Creevey, O. L., Abreu Aramburu, A., Accart, S., Andrae, R., Bailer-Jones, C. A. L., Bellas-Velidis, I., Brouillet, N., Brugaletta, E., Burlacu, A., Carballo, R., Casamiquela, L., Chiavassa, A., Cooper, W. J., Dapergolas, A., Delchambre, L., Dharmawardena, T. E., Drimmel, R., Edvardsson, B., Fouesneau, M., Garabato, D., García-Lario, P., García-Torres, M., Gavel, A., Gomez, A., González-Santamaría, I., Hatzidimitriou, D., Heiter, U., Jean-Antoine Piccolo, A., Kontizas, M., Korn, A. J., Lanzafame, A. C., Lebreton, Y., Le Fustec, Y., Licata, E. L., Lindstrøm, H. E. P., Livanou, E., Lobel, A., Lorca, A., Magdaleno Romeo, A., Marocco, F., Marshall, D. J., Mary, N., Nicolas, C., Pallas-Quintela, L., Panem, C., Pichon, B., Riclet, F., Robin, C., Rybizki, J., Santoveña, R., Silvelo, A., Smart, R. L., Sarro, L. M., Sordo, R., Soubiran, C., Süveges, M., Ulla, A., Vallenari, A., Zorec, J., Utrilla, E., and Bakker, J. (2023). Gaia Data Release 3. Analysis of RVS spectra using the General Stellar Parametriser from spectroscopy. , 674:A29.

Rice, M. and Brewer, J. M. (2020). Stellar Characterization of Keck HIRES Spectra with The Cannon. , 898(2):119.

Robertson, P., Mahadevan, S., Endl, M., and Roy, A. (2014). Stellar activity masquerading as planets in the habitable zone of the M dwarf Gliese 581. *Science*, 345(6195):440–444.

Rosenblatt, M. (1956). Remarks on Some Nonparametric Estimates of a Density Function. *The Annals of Mathematical Statistics*, 27(3):832 – 837.

Rosenthal, L. J., Fulton, B. J., Hirsch, L. A., Isaacson, H. T., Howard, A. W., Dedrick, C. M., Sherstyuk, I. A., Blunt, S. C., Petigura, E. A., Knutson, H. A., Behmard, A., Chontos, A., Crepp, J. R., Crossfield, I. J. M., Dalba, P. A., Fischer, D. A., Henry, G. W., Kane, S. R., Kosiarek, M., Marcy, G. W., Rubenzahl, R. A., Weiss, L. M., and Wright,

J. T. (2021). The California Legacy Survey. I. A Catalog of 178 Planets from Precision Radial Velocity Monitoring of 719 Nearby Stars over Three Decades. , 255(1):8.

Rowe, J. F., Bryson, S. T., Marcy, G. W., Lissauer, J. J., Jontof-Hutter, D., Mullally, F., Gilliland, R. L., Issacson, H., Ford, E., Howell, S. B., Borucki, W. J., Haas, M., Huber, D., Steffen, J. H., Thompson, S. E., Quintana, E., Barclay, T., Still, M., Fortney, J., Gautier, T. N., I., Hunter, R., Caldwell, D. A., Ciardi, D. R., Devore, E., Cochran, W., Jenkins, J., Agol, E., Carter, J. A., and Geary, J. (2014). Validation of Kepler's Multiple Planet Candidates. III. Light Curve Analysis and Announcement of Hundreds of New Multi-planet Systems. , 784(1):45.

Rubie, D. C., Jacobson, S. A., Morbidelli, A., O'Brien, D. P., Young, E. D., de Vries, J., Nimmo, F., Palme, H., and Frost, D. J. (2015). Accretion and differentiation of the terrestrial planets with implications for the compositions of early-formed Solar System bodies and accretion of water. , 248:89–108.

Santerne, A., Hébrard, G., Deleuil, M., Havel, M., Correia, A. C. M., Almenara, J. M., Alonso, R., Arnold, L., Barros, S. C. C., Behrend, R., Bernasconi, L., Boisse, I., Bonomo, A. S., Bouchy, F., Bruno, G., Damiani, C., Díaz, R. F., Gravallon, D., Guillot, T., Labrevoir, O., Montagnier, G., Moutou, C., Rinner, C., Santos, N. C., Abe, L., Audejean, M., Bendjoya, P., Gillier, C., Gregorio, J., Martinez, P., Michelet, J., Montaigut, R., Poncy, R., Rivet, J. P., Rousseau, G., Roy, R., Suarez, O., Vanhuysse, M., and Verilhac, D. (2014). SOPHIE velocimetry of Kepler transit candidates. XII. KOI-1257 b: a highly eccentric three-month period transiting exoplanet. , 571:A37.

Schlecker, M., Kossakowski, D., Brahm, R., Espinoza, N., Henning, T., Carone, L., Molaverdikhani, K., Trifonov, T., Mollière, P., Hobson, M. J., Jordán, A., Rojas, F. I., Klahr, H., Sarkis, P., Bakos, G. Á., Bhatti, W., Osip, D., Suc, V., Ricker, G., Vanderspek, R., Latham, D. W., Seager, S., Winn, J. N., Jenkins, J. M., Vezie, M., Villaseñor, J. N., Rose, M. E., Rodriguez, D. R., Rodriguez, J. E., Quinn, S. N., and Shporer, A. (2020). A Highly Eccentric Warm Jupiter Orbiting TIC 237913194. , 160(6):275.

Schwarz, G. (1978). Estimating the Dimension of a Model. *Annals of Statistics*, 6(2):461–464.

Shara, M. M., Hurley, J. R., and Mardling, R. A. (2016). Dynamical Interactions Make Hot Jupiters in Open Star Clusters. , 816(2):59.

Soker, N. and Harpaz, A. (2000). Rotation, planets, and the 'second parameter' of the horizontal branch. , 317(4):861–866.

Spronck, J. F. P., Fischer, D. A., Kaplan, Z., Jurgenson, C. A., Valenti, J., Moriarty, J., and Szymkowiak, A. E. (2015). Fiber Scrambling for High-Resolution Spectrographs. II. A Double Fiber Scrambler for Keck Observatory. , 127(956):1027.

Stephan, A. P., Naoz, S., and Gaudi, B. S. (2018). A-type Stars, the Destroyers of Worlds: The Lives and Deaths of Jupiters in Evolving Stellar Binaries. , 156(3):128.

Stephan, A. P., Naoz, S., and Gaudi, B. S. (2020a). Giant Planets, Tiny Stars: Producing Short-Period Planets around White Dwarfs with the Eccentric Kozai-Lidov Mechanism. *arXiv e-prints*, page arXiv:2010.10534.

Stephan, A. P., Naoz, S., Gaudi, B. S., and Salas, J. M. (2020b). Eating Planets for Lunch and Dinner: Signatures of Planet Consumption by Evolving Stars. , 889(1):45.

Stephan, A. P., Naoz, S., Ghez, A. M., Morris, M. R., Ciurlo, A., Do, T., Breivik, K., Coughlin, S., and Rodriguez, C. L. (2019). The Fate of Binaries in the Galactic Center: The Mundane and the Exotic. , 878(1):58.

Stephan, A. P., Naoz, S., Ghez, A. M., Witzel, G., Sitarski, B. N., Do, T., and Kocsis, B. (2016). Merging binaries in the Galactic Center: the eccentric Kozai-Lidov mechanism with stellar evolution. , 460(4):3494–3504.

Stephan, A. P., Naoz, S., and Zuckerman, B. (2017). Throwing Icebergs at White Dwarfs. , 844(2):L16.

Storch, N. I., Anderson, K. R., and Lai, D. (2014). Chaotic dynamics of stellar spin in binaries and the production of misaligned hot Jupiters. *Science*, 345(6202):1317–1321.

Storch, N. I., Lai, D., and Anderson, K. R. (2017). Dynamics of stellar spin driven by planets undergoing Lidov-Kozai migration: paths to spin-orbit misalignment. , 465(4):3927–3942.

Stumpe, M. C., Smith, J. C., Catanzarite, J. H., Van Cleve, J. E., Jenkins, J. M., Twicken, J. D., and Girouard, F. R. (2014). Multiscale Systematic Error Correction via Wavelet-Based Bandsplitting in Kepler Data. , 126(935):100.

Sullivan, K., Kraus, A. L., Berger, T. A., Dupuy, T. J., Evans, E., Gaidos, E., Huber, D., Ireland, M. J., Mann, A. W., Petigura, E. A., Thao, P. C., Wood, M. L., and Zhang, J. (2024). Revising Properties of Planet–Host Binary Systems. IV. The Radius Distribution of Small Planets in Binary Star Systems Is Dependent on Stellar Separation. , 168(3):129.

Isabel Angelo, Bedell, M., Petigura, E., and Ness, M. (2024). A Data-driven Spectral Model of Main-sequence Stars in Gaia DR3. *The Astrophysical Journal*, 974(1):43.

Isabel Angelo, Duchene, G., Stapelfeldt, K., Telkamp, Z., Ménard, F., Padgett, D., Van der Plas, G., Villenave, M., Pinte, C., Wolff, S., Fischer, W. J., and Perrin, M. D. (2023). Demographics of Protoplanetary Disks: A Simulated Population of Edge-on Systems. *The Astrophysical Journal*, 945(2):130.

Isabel Angelo and Hu, R. (2017). A Case for an Atmosphere on Super-Earth 55 Cancri e. *The Astronimical Journal*, 154(6):232.

Isabel Angelo, Naoz, S., Petigura, E., MacDougall, M., Stephan, A. P., Isaacson, H., and Howard, A. W. (2022). Kepler-1656b's Extreme Eccentricity: Signature of a Gentle Giant. *The Astronomical Journal*, 163(5):227.

Isabel Angelo, Petigura, E., and Bedell, M. (2025). Seeking Spectral Binaries with Data-Driven Models. *submitted to The Astrophysical Journal.*

Isabel Angelo, Rowe, J. F., Howell, S. B., Quintana, E. V., Still, M., Mann, A. W., Burningham, B., Barclay, T., Ciardi, D. R., Huber, D., and Kane, S. R. (2017). Kepler-1649b: An Exo-Venus in the Solar Neighborhood. *The Astronomical Journal*, 153(4):162.

Teyssandier, J., Naoz, S., Lizarraga, I., and Rasio, F. A. (2013). Extreme Orbital Evolution from Hierarchical Secular Coupling of Two Giant Planets. , 779(2):166.

Teyssandier, J. and Ogilvie, G. I. (2016). Growth of eccentric modes in disc-planet interactions. , 458(3):3221–3247.

Teyssandier, J. and Ogilvie, G. I. (2017). Secular evolution of eccentricity in protoplanetary discs with gap-opening planets. , 467(4):4577–4590.

Thomas, C. A., Weiss, L. M., and He, M. Y. (2025). Biases from Missing a Small Planet in High Multiplicity Systems. , 984(1):L30.

Ting, Y.-S., Conroy, C., Rix, H.-W., and Cargile, P. (2019). The Payne: Self-consistent ab initio Fitting of Stellar Spectra. , 879(2):69.

Ting, Y.-S., Rix, H.-W., Conroy, C., Ho, A. Y. Q., and Lin, J. (2017). Measuring 14 Elemental Abundances with R = 1800 LAMOST Spectra. , 849(1):L9.

Traven, G., Feltzing, S., Merle, T., Van der Swaelmen, M., Čotar, K., Church, R., Zwitter, T., Ting, Y. S., Sahlholdt, C., Asplund, M., Bland-Hawthorn, J., De Silva, G., Freeman, K., Martell, S., Sharma, S., Zucker, D., Buder, S., Casey, A., D'Orazi, V., Kos, J., Lewis, G., Lin, J., Lind, K., Simpson, J., Stello, D., Munari, U., and Wittenmyer, R. A. (2020). The GALAH survey: multiple stars and our Galaxy. I. A comprehensive method for deriving properties of FGK binary stars. , 638:A145.

Udry, S. and Santos, N. C. (2007). Statistical Properties of Exoplanets. , 45(1):397–439.

van der Walt, S., Colbert, S. C., and Varoquaux, G. (2011). The NumPy Array: A Structure for Efficient Numerical Computation. *Computing in Science and Engineering*, 13(2):22–30.

Van Der Walt, S., Colbert, S. C., and Varoquaux, G. (2011). The NumPy Array: A Structure for Efficient Numerical Computation. *Computing in Science and Engineering*, 13(2):22–30.

Van Eylen, V., Albrecht, S., Huang, X., MacDonald, M. G., Dawson, R. I., Cai, M. X., Foreman-Mackey, D., Lundkvist, M. S., Silva Aguirre, V., Snellen, I., and Winn, J. N. (2019). The Orbital Eccentricity of Small Planet Systems. , 157(2):61.

Vick, M., Lai, D., and Anderson, K. R. (2019). Chaotic tides in migrating gas giants: forming hot and transient warm Jupiters via Lidov-Kozai migration. , 484(4):5645–5668.

Vogt, S. S., Allen, S. L., Bigelow, B. C., Bresee, L., Brown, B., Cantrall, T., Conrad, A., Couture, M., Delaney, C., Epps, H. W., Hilyard, D., Hilyard, D. F., Horn, E., Jern, N., Kanto, D., Keane, M. J., Kibrick, R. I., Lewis, J. W., Osborne, J., Pardeilhan, G. H., Pfister, T., Ricketts, T., Robinson, L. B., Stover, R. J., Tucker, D., Ward, J., and Wei, M. Z. (1994). HIRES: the high-resolution echelle spectrometer on the Keck 10-m Telescope. In Crawford, D. L. and Craine, E. R., editors, *Instrumentation in Astronomy VIII*, volume 2198 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 362.

Wang, J., Fischer, D. A., Xie, J.-W., and Ciardi, D. R. (2014). Influence of Stellar Multiplicity on Planet Formation. II. Planets are Less Common in Multiple-star Systems with Separations Smaller than 1500 AU. , 791(2):111.

Wang, X.-Y., Rice, M., Wang, S., Pu, B., Stefánsson, G., Mahadevan, S., Giacalone, S., Wu, Z.-Y., Esposito, T. M., Dalba, P. A., Avsar, A., Holden, B., Skiff, B., Polakis, T., Voeller, K., Logsdon, S. E., Klusmeyer, J., Schweiker, H., Wu, D.-H., Beard, C., Dai, F., Lubin, J., Weiss, L. M., Bender, C. F., Blake, C. H., Dressing, C. D., Halverson, S., Howard, A. W., Huber, D., Isaacson, H., Jackman, J. A. G., Llama, J., McElwain, M. W., Rajagopal, J., Roy, A., Robertson, P., Schwab, C., Shkolnik, E. L., Wright, J., and Laughlin, G. (2021). SOLES II: The Aligned Orbit of WASP-148b, the Only Known Hot Jupiter with a Nearby Warm Jupiter Companion, from NEID and HIRES. *arXiv e-prints*, page arXiv:2110.08832.

Ward, W. R. (1988). On disk-planet interactions and orbital eccentricities. , 73(2):330–348.

Wilson, O. C. (1968). Flux Measurements at the Centers of Stellar H- and K-Lines. , 153:221.

Winn, J. N. and Fabrycky, D. C. (2015). The Occurrence and Architecture of Exoplanetary Systems. , 53:409–447.

Winn, J. N., Johnson, J. A., Peek, K. M. G., Marcy, G. W., Bakos, G. Á., Enya, K., Narita, N., Suto, Y., Turner, E. L., and Vogt, S. S. (2007). Spin-Orbit Alignment for the Eccentric Exoplanet HD 147506b. , 665(2):L167–L170.

Wittenmyer, R. A., Butler, R. P., Tinney, C. G., Horner, J., Carter, B. D., Wright, D. J., Jones, H. R. A., Bailey, J., and O'Toole, S. J. (2016). The Anglo-Australian Planet Search XXIV: The Frequency of Jupiter Analogs. , 819(1):28.

Wright, J. T., Marcy, G. W., Butler, R. P., Vogt, S. S., Henry, G. W., Isaacson, H., and Howard, A. W. (2008). The Jupiter Twin HD 154345b. , 683(1):L63.

Wu, Y. (2018). Diffusive Tidal Evolution for Migrating Hot Jupiters. , 155(3):118.

Wu, Y., Murray, N. W., and Ramsahai, J. M. (2007). Hot Jupiters in Binary Star Systems. , 670(1):820–825.

Wyatt, M. C. (2008). Evolution of debris disks. , 46:339–383.

Xiang, M., Ting, Y.-S., Rix, H.-W., Sandford, N., Buder, S., Lind, K., Liu, X.-W., Shi, J.-R., and Zhang, H.-W. (2019). Abundance Estimates for 16 Elements in 6 Million Stars from LAMOST DR5 Low-Resolution Spectra. , 245(2):34.

Yee, S. W., Petigura, E. A., Fulton, B. J., Knutson, H. A., Batygin, K., Bakos, G. Á., Hartman, J. D., Hirsch, L. A., Howard, A. W., Isaacson, H., Kosiarek, M. R., Sinukoff, E., and Weiss, L. M. (2018). HAT-P-11: Discovery of a Second Planet and a Clue to Understanding Exoplanet Obliquities. , 155(6):255.

Yee, S. W., Petigura, E. A., and von Braun, K. (2017). Precision Stellar Characterization of FGKM Stars using an Empirical Spectral Library. , 836(1):77.

Ziegler, C., Law, N. M., Baranec, C., Morton, T., Riddle, R., De Lee, N., Huber, D., Mahadevan, S., and Pepper, J. (2018). Measuring the Recoverability of Close Binaries in Gaia DR2 with the Robo-AO Kepler Survey. , 156(6):259.

Zuckerman, B., Melis, C., Klein, B., Koester, D., and Jura, M. (2010). Ancient Planetary Systems are Orbiting a Large Fraction of White Dwarf Stars. , 722(1):725–736.