The Orbits of Young Extrasolar Planets as Formation Probes

Thesis by
Sarah Blunt

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2023
Defended May 10, 2023

Sarah Blunt
ORCID: 0000-0002-3199-2888

Over my career, I have benefited from the insights, expertise, and camaraderie of other women, particularly women of color, who have since left the field of academic physics because of the unacceptable way they were treated. Although they have gone on to lead brilliant careers on their own terms, this thesis is dedicated to them, with my most sincere thanks and an acknowledgement that I would not be here without them.

You can do it, Bub!
- *Devin Cody*

I mean, nobody *wants* to do statistics.
- *Devin Cody*

ABSTRACT

In this thesis, I have translated a few flavors of exoplanet timeseries measurements into constraints on exoplanet orbital parameters, and used these to make inferences about planet formation. I begin by introducing the two main observational techniques I used to perform these analyses: optical interferometry and stellar radial velocity monitoring. I then discuss some of the big open questions of exoplanet formation, particularly the mechanism for forming giant planets close to and far from their stars, where core accretion is thought to be too inefficient to form giant planet cores in time for them to accrete atmospheres.

High cadence radial velocity monitoring enables advances in our understanding of stellar activity, the fundamental stumbling block in the path to discovering and characterizing planets like the Earth. In my second thesis chapter, I present an argument that previously published RV-derived activity models of the PMS star V1298 Tau suffer from overfitting, casting doubt on published mass estimates of the young planets in the system which necessitated rapid contraction after formation, in tension with formation theory. I walk through several potential explanations for this overfitting, pointing out that the star has a strong differential rotation signal which is not included in published model fits, and encourage broader use of cross validation techniques in stellar activity model evaluation.

Optical interferometry, particularly using the VLTI/GRAVITY instrument, enables astrometry measurements that are orders of magnitude more precise than contemporary coronographic instruments, which translates to precise orbital parameters. In my third thesis chapter, I present and analyze two new VLTI/GRAVITY astrometric measurements of a young, widely separated planet and use them to make a preliminary argument that the planet's eccentricity is low or moderate. This sets an upper limit on the time (relative to disk dispersal) that the planet attained its current wide separation, and downweights the possibility of scattering after disk dispersal.

In the next two chapters, I showcase my contributions to two widely used open-source orbit-fitting software toolkits: `orbitize!` and `radvel`. I highlight two major new features of `orbitize!` that are available in the main code base as of the release of version 2: jointly fitting radial velocity measurements and jointly fitting absolute astrometry measurements. In the `radvel` chapter, I motivate and describe

an updated Gaussian Process regression model for stellar activity modeling that reduces the potential for overfitting.

PUBLISHED CONTENT AND CONTRIBUTIONS

Blunt, S., A. Carvalho, T. J. David, C. Beichman, J. K. Zink, E. Gaidos, A. Behmard, L. G. Bouma, D. Cody, F. Dai, D. Foreman-Mackey, S. Grunblatt, A. W. Howard, M. Kosiarek, H. A. Knutson, R. A. Rubenzahl, C. Beard, A. Chontos, S. Giacalone, T. Hirano, M. C. Johnson, J. Lubin, J. M. A. Murphy, E. A. Petigura, J. Van Zandt, and L. Weiss (2023a). "Overfitting Affects the Reliability of Radial Velocity Mass Estimates of the V1298 Tau Planets." In: *The Astronomical Journal (submitted)*.
S.B. conceived and carried out the analysis after being pointed to the star V1298 Tau and wrote the majority of the manuscript. Other authors contributed insights and ideas, and/or took HIRES observations.

Blunt, S., W. O. Balmer, J. J. Wang, S. Lacour, R. Abuter, A. Amorim, R. Asensio-Torres, M. Benisty, J.-P. Berger, H. Beust, A. Boccaletti, A. Bohn, M. Bonnefoy, H. Bonnet, G. Bourdarotm, W. Brandner, F. Cantalloube, P. Caselli, B. Charnay, G. Chauvin, A. Chavez, E. Choquet, V. Christiaens, Y. Clénet, V. Coudé du Foresto, A. Cridland, R. Dembet, A. Drescher, G. Duvert, A. Eckart, F. Eisenhauer, H. Feuchtgruber, P. Garcia, R. Garcia Lopez, T. Gardner, E. Gendron, R. Genzel, S. Gillessen, J. H. Girard, X. Haubois, G. Heißel, Th. Henning, S. Hinkley, S. Hippler, M. Horrobin, M. Houllé, Z. Hubert, L. Jocou, J. Kammerer, M. Keppler, P. Kervella, L. Kreidberg, A.-M. Lagrange, V. Lapeyrère, J.-B. Le Bouquin, P. Léna, D. Lutz, A.-L. Maire, F. Mang, G.-D. Marleau, A. Mérand, P. Mollière, J. D. Monnier, C. Mordasini, D. Mouillet, E. Nasedkin, M. Nowak, T. Ott, G. P. P. L. Otten, C. Paladini, T. Paumard, K. Perraut, G. Perrin, O. Pfuhl, E. Rickman, N. Pourré, L. Pueyo, J. Rameau, L. Rodet, Z. Rustamkulov, J. Shangguan, T. Shimizu, D. Sing, T. Stolker, C. Straubmeier, E. Sturm, L. J. Tacconi, E.F. van Dishoeck, A. Vigan, F. Vincent, K. Ward-Duong, F. Widmann, E. Wieprecht, E. Wiezorrek, T. Winterhalder, J. Woillez, S. Yazici, and A. Young (2023b). "VLTI/GRAVITY x HIP 65426 b: Preliminary Evidence for a Low or Moderate Orbital Eccentricity." In: *The Astronomical Journal (submitted)*.
S.B. set the direction of and carried out the majority of the analysis after being pointed to the planet HIP 65426 b. S.B. also wrote the majority of the manuscript. W.B. and S.L. wrote the section about new observations. W.B. performed an initial spectral analysis, and taught S.B. to use `species`. Other authors contributed insights and ideas, and/or contributed to the VLTI/GRAVITY observations.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

Chapter 1

INTRODUCTION

1.1 Introduction to the Exoplanet Zoo

The first exoplanets (Wolszczan and Frail, 1992) were discovered two years before I was born, and in my lifetime the number of known planets has grown exponentially to over 5000. While the first planets were discovered around pulsars, astronomers have since discovered planets around cool M dwarfs (Dressing and Charbonneau, 2013), white dwarfs (Vanderburg et al., 2020), evolved stars (Grunblatt et al., 2023), hot A stars (Lagrange et al., 2009), brown dwarfs (Chauvin et al., 2004), and solar analogues (Mayor and Queloz, 1995) (among other flavors of hosts). These planets span orders of magnitude of size and orbital period (Figure 1.1) and cover the gamut of eccentricities (e.g., O'Toole et al., 2009). Exoplanets have been detected using a variety of techniques (Figure 1.1), each of which has strengths and weaknesses. The three methods most relevant to this thesis are transit, radial velocity (RV), and direct imaging. The transit technique, by geometric probability, is most sensitive to planets closest to their host stars, while imaging is biased toward massive, young, and widely separated planets. The RV technique also preferentially detects close-in planets, but as observing baselines increase, it is becoming more sensitive to more widely-separated planets. All three techniques are discussed in detail in Section 1.3.

With such a wealth of data at our disposal, exoplanet scientists are beginning to focus on: 1) individual characterization of exoplanets, or systems of exoplanets, and 2) looking for population-level trends. The two papers that make up the bulk of my thesis are studies of individual stellar systems that dive into what these particular objects can tell us about planet formation. Even though we have detected thousands of exoplanets, the vast majority of them are billions of years old, whereas the planet formation process unfolds on a scale of millions of years. At these mature ages, many of physical "clues" imprinted during formation (see next section) have disappeared. The small number of known young planets makes it difficult to do large statistical studies to probe formation conditions. However, we can still learn a lot from detailed characterizations of individual young systems, as I aim to show in this thesis.

In this section, I have highlighted a few of the major observational results of the past three decades of exoplanet astrophysics, with the goal of communicating to

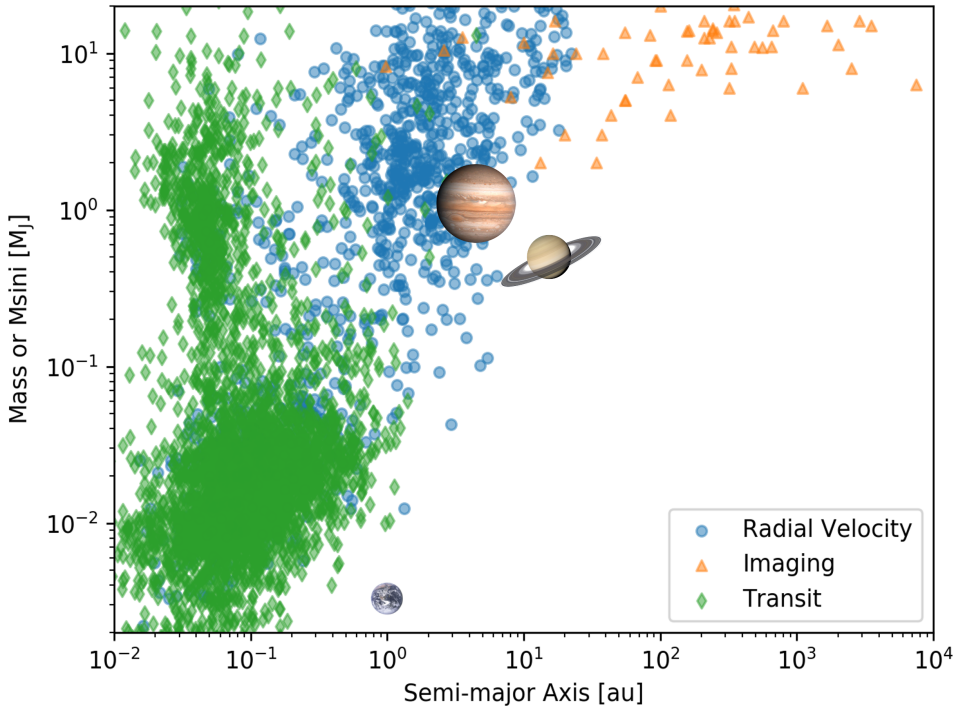

Figure 1.1: The exoplanet zoo as of May 10, 2023, using data from the NASA Exoplanet Archive. This figure only includes planets discovered via one of the three major methods discussed in this thesis: transit, radial velocity, and imaging. Due to selection effects, the imaged population is generally younger, more massive, and more widely separated than the transit and RV populations. However, some transiting and/or RV-detected planets are also young. The positions of Jupiter, Saturn, and Earth are shown for reference.

the reader some of the incredible diversity of exoplanet characteristics and habitats. The next section discusses the open questions in exoplanet astrophysics that I focus on in this document. In Section 1.3, I focus on the main observational techniques I will use in this thesis. Finally, Section1.4 includes an argument for the importance of high quality software in this field. I outline the rest of the thesis document in Section 1.5.

1.2 Open Questions in Planet Formation

How do Cold Jupiters Form?

The existence of cold Jupiters ($1 \lesssim M \lesssim 13\ M_J$, $a \gtrsim 10$ au) is unexpected from a formation theory perspective. These planets exist at disk locations where the classical

timescale to form a core sufficiently massive to undergo runaway accretion is significantly longer than the protoplanetary disk lifetime (Bowler, 2016). Therefore, for in situ formation to be viable, rocky planetary cores must have formed more quickly than previously expected in order for the cores to have enough time to enter the runaway accretion phase and obtain massive envelopes. Ex-situ formation via standard core accretion and subsequent outward migration is another possibility (e.g. Lega et al. 2021). Direct in-situ gravitational collapse is not expected to occur in the inner region of the protoplanetary disk: not only does the disk need to be Toomre-unstable, but also the local disk cooling time needs to be short enough to allow the formation of self-gravitating clumps (Rafikov, 2005). For cold Jupiters, however, in-situ gravitational collapse is a possibility, albeit complicated by the very fast inward migration timescales expected for any formed clumps (Nayakshin, 2017), as well as the uncertainty around how smaller giant planets (closer to Jupiter-mass) stop accreting gas after collapsing early in the lifetime of the disk (Drazkowska et al., 2022). Figure 1.2 beautifully illustrates these two alternate formation pathways for cold Jupiters.

Objects that formed rapidly (e.g., via gravitational collapse) are generally expected to retain more heat from their formation, resulting in a higher initial entropy, whereas slower formation processes (e.g., via core accretion) result in a lower formation entropy. Giant planets "forget" their initial entropy over time, and within a few Myr progress along the Kelvin-Hemholtz cooling track (Marley et al., 2007). This makes young giant planets important formation probes. Their observed masses and radii can be compared with expected cooling tracks calculated from different initial entropy assumptions, and therefore constrain the "efficiency" of the formation process. This motivates precise mass and radius measurements for young cold Jupiters, of which we know of precious few.

Mass and radius (which directly trace initial entropy) are not the only observational formation tracers for cold Jupiters. Another class of observational clues is dynamical constraints. In particular, eccentricity (e.g., Bowler, Blunt, and Nielsen, 2020a, Dupuy et al. 2019) and the various flavors of obliquity (e.g., the angle between the star's rotation axis and the planet's rotation axis, and the angle between the star's rotation axis and the planet's orbital axis; Bryan et al., 2016, Bowler et al., 2023) probe formation and evolution history, as discussed in much more detail in Chapter 3. Another important class of constraints is atmospheric metallicity (generally relative to the host star) and atmospheric C/O ratio (Öberg, Murray-Clay, and Bergin, 2011).

How do Small Highly Irradiated Planets Form?

The formation and evolution of small irradiated planets (orbital periods of $\lesssim$ 30 days) is also highly uncertain. Super-Earths and sub-Neptunes, which are planets in the radius range 1-4 $R_\oplus$ for which there are no analogues in our solar system, have been revealed to be very common: the super-Earth occurrence rate has been estimated at 30% (Zhu et al., 2018) based on Kepler results, as compared to less than 1% for cold Jupiters between 10 and 100 au (Nielsen et al., 2019). Fulton et al. (2017) famously revealed a gap in the radius number distribution of planets 1-4 $R_\oplus$ using Kepler data, which is interpreted as a dichotomy between bare rocky cores ("super-Earths") and rocky planets with voluminous H/He envelopes ("sub-Neptunes"). The formation pathways for these planets are unclear, and since the Kepler sample is (on average) several Gyr old (Berger et al., 2020), dedicated studies of young systems, particularly mass, radius, and age measurements, are necessary to gain insight.

As an example of how bulk physical properties constrain the formation processes of super-Earth and sub-Neptune progenitors, Owen (2020) walks through a procedure for estimating the entropy of formation for young rocky planets with inflated H/He envelopes (presumably $\sim$0.5 R_J, but see Chapter 2 for more discussion on this). taking into account several processes that would slow or reverse envelope accretion (in particular, two classes of mechanisms have been proposed: mechanisms that increase the entropy of the interior, and so slow accretion, and mechanisms that result in extra mass loss after accretion, called "boil-off"). Precise mass measurements (as well as well-known radii and ages) are needed to distinguish between these scenarios (20% for the young star V1298 Tau, studied in Chapter 2). This requirement motivates careful study of the stellar activity on young stars like V1298 Tau to enable such measurements and push forward our understanding of the formation of planets with rocky cores.

1.3 Observational Techniques

In this thesis, I study five previously discovered planets in two planetary systems. In this section, I provide some background on the techniques used to discover these planets, and the complementary techniques I used to study them.

Transits and Radial Velocities

The four planets orbiting V1298 Tau (Chapter 2) were initially discovered using the transit method (Winn, 2010), which works by observing the periodic dimming of a

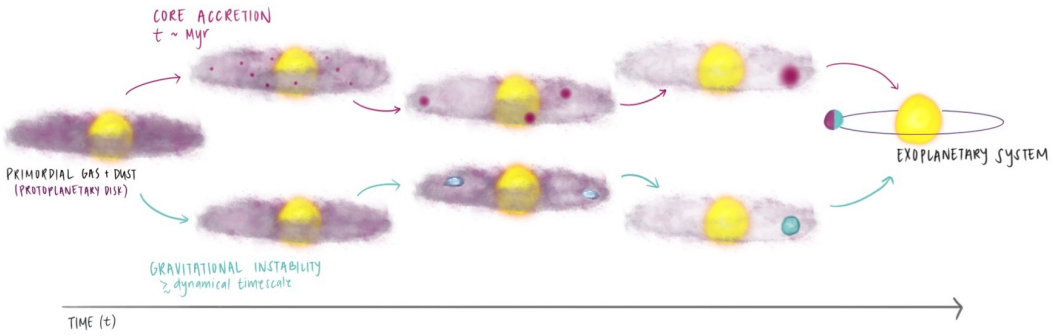

Figure 1.2: The two major possible pathways for the formation of cold Jupiters: core accretion, the slow build up of material over Myr, and gravitational collapse, the rapid collapse of material due to disk instabilities, which happens on approximately the dynamical timescale. This figure was designed by Suchitra Narayanan and Sarah Blunt, and drawn by Suchitra. Inspired by a figure by Tom Dunne (available at https://www.americanscientist.org/article/why-does-nature-form-exoplanets-easily.

star's brightness caused by a planet occulting the stellar surface. Planetary transits have a characteristic shape (Kreidberg, 2015), and occur periodically (modulo transit timing variations due to planetary interactions), which helps to identify them as bona fide planetary signals. The depth of the transit signal constrains R_p/R_*, which, together with an independent constraint on the stellar radius R_*, gives the planet radius. The measurement of accurate stellar radii is a topic unto itself (Fulton et al., 2017), but this is out of the scope of this thesis.

Transit-derived planet radii already give us insight into the planet as a "world" (e.g., the radii of small, presumably rocky planets are inconsistent with the presence of an extended H/He envelope, Zeng et al., 2019) but an understanding of a planet's bulk composition, atmosphere size and metallicity, and bulk metallicity (among other interesting derivable quantities) depends on independent measurements of mass. Masses for transiting planets can be derived from the magnitudes of transit timing variations (Agol and Fabrycky, 2018), but this technique only works for planets in multiple systems with detectable timing variations. The mass measurement technique most relevant to this thesis is the radial velocity detection method (Lovis and Fischer, 2010).

The radial velocity method is one of the oldest exoplanet detection techniques, having discovered the first planet around a main sequence (MS) star in 1995. It works by relating a star's velocity along the observer's line of right to the (non-relativistic, relative) Doppler shift measured from stellar spectral lines:

$$\mathbf{k} \cdot \mathbf{v}_* = c \frac{\lambda_B - \lambda_0}{\lambda_0}, \tag{1.1}$$

where $\mathbf{k} \cdot \mathbf{v}_*$ is the measured radial velocity, λ_B is the wavelength (of an absorption line center, for example) that would be measured in the International Celestial Reference System (IRCS), a reference frame with the solar system barycenter at rest at frame center, and λ_0 is the actual measured wavelength. The new RV data in this thesis were taken with Keck/HIRES (Vogt et al., 1994), a cross-dispersed echelle spectrograph. Stellar light is passed through a warm iodine gas cell, which imprints a forest of absorption lines on the spectrum. The HIRES Doppler pipeline (Butler et al., 1996a)[1] combines a high-resolution iodine spectrum taken in a lab, an iodine-free "template" observation of the star being observed, and a three Gaussian PSF model into a forward model of the stellar spectrum. A radial velocity is derived from a stellar spectrum by jointly fitting for the PSF parameters and a RV offset on 2 "chunks" of the spectrum. The average and statistical uncertainties of the resulting chunks are taken to be the measurement and observational uncertainty. The iodine cell technique on HIRES gives a characteristic RV uncertainty of 1 $\mathrm{m\,s^{-1}}$.

More modern methods of measuring radial velocities abound, from technological innovations like the use of wavelength calibrators more precise and stable than iodine cells (e.g., laser frequency combs, Steinmetz et al., 2008) to statistical methods like line-by-line radial velocities (Dumusque, 2018). As I write this document, the field is moving from the era of precise radial velocities (PRV; $\sim\mathrm{m\,s^{-1}}$ precision) into the era of extremely precise radial velocities (EPRV; $\sim\mathrm{cm\,s^{-1}}$ precision). While EPRV instruments still face technical challenges to obtaining ever-higher RV precision, most in the field agree that the limiting factor in EPRV analyses is not instrumental, but astrophysical.

In addition to planetary orbits, stellar surface processes can also induce apparent Doppler shifts. Asteroseismic oscillations, asymmetries in surface granulation patterns, rotationally modulated signals from magnetic surface features like spots, plage, and faculae, and multi-year activity cycles reflecting magnetic dynamo processes may all change the measured radial velocity signals. (See Section 5.2 for more detail.) Spot signals in particular are problematic, because they are (quasi)-periodic, and can occur on timescales similar to planetary orbits.

Young stars are particularly magnetically active, and vary up to $\mathrm{km\,s^{-1}}$ due to spots (e.g., Figure 2.1). For this reason, Doppler surveys looking to find and understand

[1] Runnable from the link: https://nexsci.caltech.edu/tools/prv_index.shtml

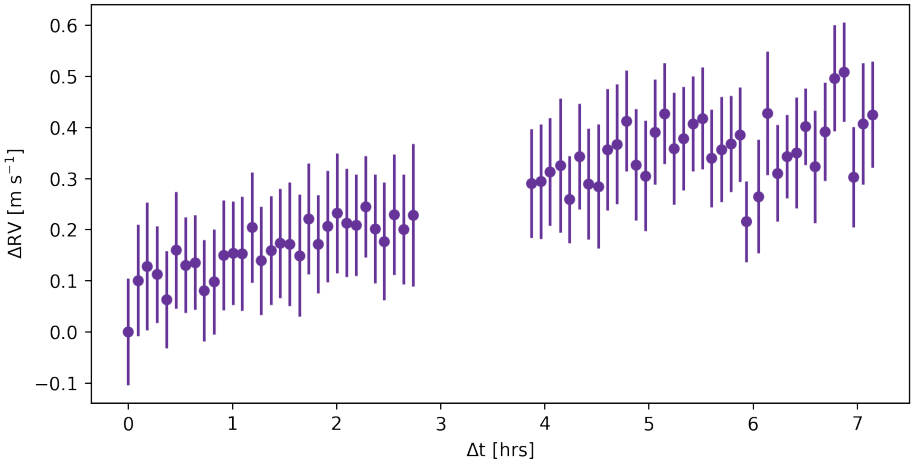

Figure 1.3: RVs of the active star V1298 Tau taken with the NEID spectrograph over 7 hours. While there is a clear increasing RV trend over the time period due to the spot activity, the RVs are stable at the tens of cm s^{-1} level over timescales shorter than an hour. This figure hightlights that active stars are not more active at **all timescales**, just on the rotation timescale.

planets have historically avoided these stars. However, the field is beginning to recognize young stars as a way to "isolate" spot activity; while large magnetic fields translate to a high spot filling fraction, the asteroseismic (Aerts, 2021) and granulation RV signals are not expected to be larger than those of their mature counterparts. To drive this point home, Figure 1.3 shows very high-cadence (several measurements per hour) RVs of V1298 Tau (Chapter 2) taken with the NEID spectrograph over a 7-hour timescale.[2] Although there is clear trend due to the activity, the star is coherent at the tens of cm s^{-1} level over timescales less than an hour.

The first chapter of my thesis studies V1298 Tau in detail, and in particular proposes a new methodology for evaluating activity models.

Direct Imaging

Exoplanet direct imaging was achieved successfully for the first time in 2004 (Chauvin et al., 2004), and in the intervening two decades has enabled detailed spectral and orbital characterization of individual objects. Imaging exoplanets requires a confluence of technology, both hardware and software. On the hardware side, imaging relies on adaptive optics (AO; Hampson et al., 2021), which itself is tied to

[2]These data are courtesy of PI Shreyas Vissapragada.

wavefront deformation sensing and real-time control algorithms, and coronagraphy (Ruane et al., 2018). On the software side, dedicated image post-processing algorithms such as KLIP (Soummer, Pueyo, and Larkin, 2012), a PCA-based algorithm for modeling the observed PSF, are necessary to image exoplanets at high contrast ratios.

Classical exoplanet imaging instruments have seen orders of magnitude of improvement in astrometric precision through software and hardware improvements (Bowler, 2016), but the resolution of a standard telescope is still fundamentally limited by the diffraction limit: $\theta = \frac{\lambda}{D}$. For the longest-period exoplanets typically amenable to imaging, with orbital periods of decades to tens of thousands of years, obtaining astrometric precision high enough to translate to meaningful orbital constraints is often impossible. The typical suggestion is to focus on the shortest-period imaged planets, which exhibit enough orbital motion to be detectable with classical imaging technology, or to wait years or decades until orbital curvature can be detected. However, the interferometer VLTI/GRAVITY (Gravity Collaboration et al., 2017) is changing the game.

Interferometers in general overcome the $\frac{\lambda}{D}$ limit (in the absence of impossibly large apertures) by using several separated apertures, rendering the fundamental resolution limit $\theta = \frac{\lambda}{b}$, where b is now the baseline between the apertures rather than the diameter of a single aperture. Light collected by each of these apertures has a slightly different path length, which translates to phase differences that vary as a function of position on the sky. The light entering the various apertures is then recombined and interfered, and the resulting fringes can be analyzed to give high spatial resolution measurements. Using only two apertures allows an observer to achieve high resolution only along one spatial dimension, so multiple baselines are typically used to achieve a high degree of spatial localization in two dimensions.

Radio astronomers have been using interferometers to achieve high spatial resolution for much longer than infrared (IR) astronomers, in part because of the λ dependence of the resolution equation; it is difficult to achieve high spatial resolution at long wavelengths without the assistance of interferometry. In addition, IR interferometry faces challenges that radio interferometry does not. First, the much smaller IR wavelengths necessitate a high degree of spatial coherence; a 0.1 μm shift is a negligible fraction of a 3 cm wavelength, but constitutes a 20% phase shift for a 0.5 μm photon. In addition, the distribution of spatial scales of atmospheric turbulence translates to a very short coherence time of IR radiation propagating through the

atmosphere ($\sim$ 1 ms), much shorter than that of radio wavelengths. This limits a naive IR interferometer to integration times shorter than the atmospheric coherence timescale, and therefore to detections of only the brightest objects.

The Very Large Telescope Interferometer (VLTI) is a set of four 8 m telescopes designed to overcome these limitations in order to obtain precise astrometry at IR wavelengths. Its new beam-combining instrument, VLTI/GRAVITY, saw first light in 2017 (Gravity Collaboration et al., 2017), and since then has been making $\sim \mu$as-level astrometric measurements of stars orbiting the galactic center, directly detecting exoplanets, and more.

The dual-field ability of VLTI/GRAVITY enables the detection of faint companions with optical interferometry, giving exquisite relative astrometric precision. Light from the bright primary star can be used for beam stabilization, while light from the fainter, secondary source can be directed to the science spectrometer, relying on simultaneous beam stabilization performed on light from the primary. As a point of comparison, Mt. Wilson's CHARA interferometer, a comparable facility, recently underwent an upgrade to add AO capabilities (Anugu et al., 2020), but lacks the dual-field capability that makes VLTI/GRAVITY so effective for exoplanet imaging.

Beam stabilization with VLTI/GRAVITY is achieved by a combination of adaptive optics and fringe tracking, which are performed fully independently. The AO system ensures that the sources are coupled efficiently to the single-mode fibers, while fringe tracking (i.e., using a control algorithm to stabilize the interference fringes) does the "heavy lifting" of computing and correcting for precise non-common path differences between light coming from each of the apertures introduced by atmospheric turbulence. The AO-corrected beams are also fed through single-mode fibers that spatially filter the light, removing all but the lowest spatial frequency mode. This filters high-frequency noise introduced by the turbulent atmosphere.

A next crucial ingredient of VLTI/GRAVITY is laser metrology. In order to make precise astrometric measurements using the dual-field capabilities, the optical path difference (OPD) between the science target and the phase reference must be known as exactly as possible. Because the paths of the two objects through the telescope and VLTI/GRAVITY itself are slightly different (due to e.g., nonuniformities in the optics which change over time with local temperature, pressure, and other factors), the VLTI/GRAVITY instrument propagates a suite of lasers through the instrument to continuously measure the OPD, and yet another control algorithm is deployed to correct for these OPDs.

These four major ingredients: AO, single mode fibers, laser metrology, and dual-field fringe tracking, allow for minute-long coherent integrations, phase-referenced astrometry of objects at fainter than 15 m_K, and a limiting magnitude of 17 m_K. This translates to relative astrometric measurements with μas precision of faint objects, and even high-contrast imaging capability.

The four UTs provide $\binom{4}{2} = 6$ unique baselines. The actual interfering of beams occurs on a silicon chip, which interferes the beams from these 6 baselines, which are then used to reconstruct the precise position of the object.

In addition to precise relative astrometry, VLTI/GRAVITY is equipped with a Wollaston prism to make polarization measurements, and two spectrometers to simultaneously obtain spectra of the science target and the fringe-tracking object. The fringe-tracking spectrometer is optimized for a high readout-rate, and has only 6 spectral pixels, while the science spectrometer is optimized for $\sim$ minute-long integrations, and allows the user to select from spectral resolutions of R=22, 500, and 4500 depending on the brightness of the target and user preference. Both spectrometers cover a spectral range of 1.95-2.45 μm.

GRAVITY astrometry is not just better in theory; it is better in practice (e.g., Nowak et al., 2020, (Hinkley et al., 2022)). Figure 1.4 shows uncertainties of actual measurements taken for HIP 65426 b, the imaged planet studied in Chapter 3. The interferometric measurements by GRAVITY are two orders of magnitude more precise than measurements taken by SPHERE and NaCo, advanced pieces of technology in their own rights. Following up exoplanets that have been discovered and astrometrically monitored by classical coronagraphic instruments with VLTI/-GRAVITY allows us to "speed up time": we no longer need to wait 10 years to observe orbital curvature of widely separated objects. Chapter 3 of this thesis dives deep into the orbital constraints we can obtain with a small number of epochs of GRAVITY astrometry, and relates these back to the big picture understanding of how HIP 65426 b, and exoplanets like it, formed.

1.4 The Importance of Software

Over the past decade, astronomers have begun to recognize the crucial role of software engineering in transformative science. Many of my thesis projects have centered around writing good code, so I want to get up on a high horse for a moment

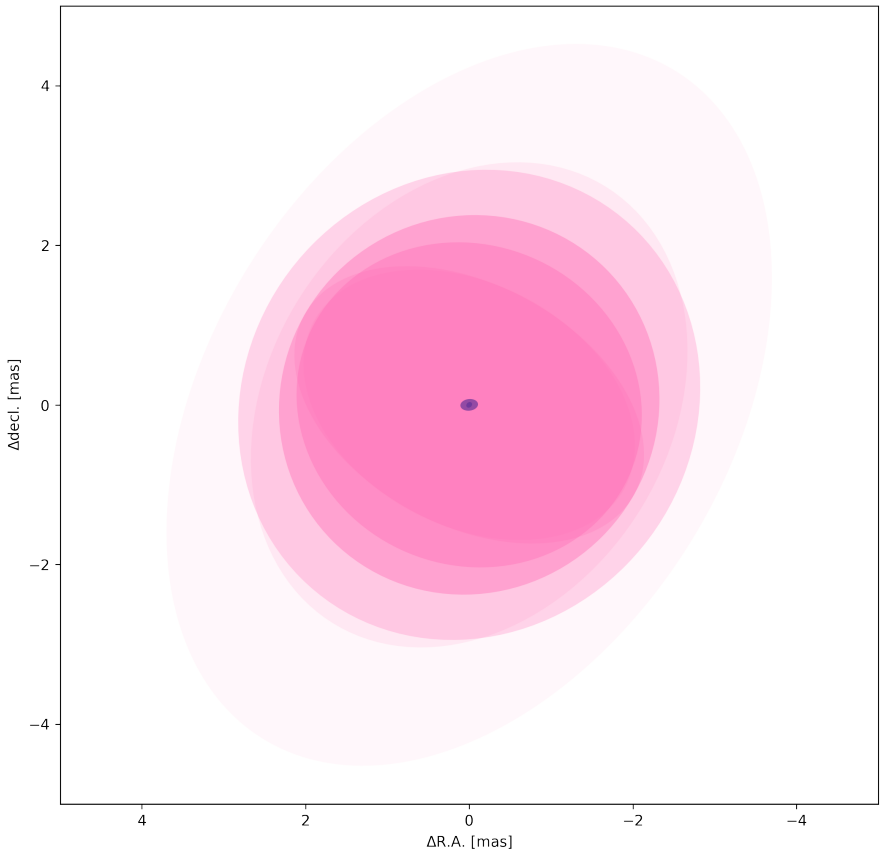

Figure 1.4: This figure highlights the improvement in uncertainties for the interferometric instrument VLTI/GRAVITY over contemporary coronographic instruments. Each pink ellipse (in various shades) shows the 1σ observational uncertainty of a relative astrometric measurement of HIP 65426 b taken with either SPHERE or NaCo. The two purple ellipses (small, at center) show the GRAVITY error bars to scale.

to emphasize *why it is important* for scientists to think about how to write code well.[3]

A central tenet of good science is reproducibility. New discoveries, particularly those that push the limits of current instruments and techniques, are not simply trusted; the measurements are repeated with other instruments, the relevant equations checked and rechecked, and the model assumptions questioned and varied. Easy-to-read, open-source software greatly speeds up this process of independent "checking of work," on the instrumentation, observation, analysis, and theory fronts. Open-sourcing software is critical; this practice enables early bug discovery and line by line comparison of independent codes using different approaches. Software that is easy to read is also crucially important for reproducibility. As an intern at Google, I was impressed how often best practices emphasized reducing the "cognitive load" of someone reading your software. How many pieces of information would a code reviewer have to hold in their mind simultaneously as they read through this nested `for` loop? Could I simplify the logic so that it would be more easily understandable? Writing readable code is not just good practice; it leads to more people understanding and questioning a code, which leads to robust, bug-free software. An impressive set of examples of the value of reproducible, open-source software is the results from early-release science (ERS) with JWST. As an example, Rustamkulov et al. (2023) used separate open-source reductions of JWST data to produce independent spectra, to remarkable agreement. The open-source nature of each of the codes used allowed the teams to "hit the ground running" as soon as the data were available, easily compute and compare results, and produce Nature-worthy results, specifically clear CO_2 and water transmission features, in a matter of weeks.

A next tenet of good science is testing. Not all science involves testing hypotheses, especially in exoplanet astrophysics, which also emphasizes exploration that is not necessarily seeking to test a particular idea. On the whole, however, science moves forward by developing and testing theories. The process of testing theories is creative and difficult (which is why observational/experimental astrophysics is a whole field), and relies heavily on good engineering. Designing tests that are plausible and informative is challenging! The same is true for software testing, which is so much more than checking that your code runs to completion. The codes we use for paradigm-shifting discoveries need to have a barrage of innovative tests thrown at them. Just as we do when developing theories, we need to think

[3] I (of course) still have a lot to learn about software development best practices!

up "antagonistic" tests that probe every conceivable edge case and actively try to make the code fail. I have much personal experience with the utility of writing good tests; tests that I have written for `orbitize!` have caught a wide variety of bugs before the code went to production (and after), from syntax errors introduced by a last-minute accidental keyboard slap to swaps in stellar mass and total mass variables that were unimportant for most of the objects we fit, but caused crucial inaccuracies for more massive secondaries.

To summarize, accurate, reproducible science relies on scientists and engineers investing time in their software.

1.5 Thesis Outline

The first two chapters (Chapter 2 and Chapter 3) of this thesis are studies of two individual stars using different techniques. Each seeks to ultimately answer the question of how these individual systems formed, and using those inferences to shed light on planetary formation as a whole. Chapter 2 is mostly methodological, using cross validation to argue that the published masses of the star V1298 Tau are unreliable, then brainstorming why. Chapter 3 focuses on what new information VLTI/GRAVITY astrometry lends to our understanding of the orbit of the young planet HIP 65426 b, and consequently what new information we can learn about the planet's formation. It too is methodological, seeking to provide a framework for how to interpret eccentricity posteriors of directly imaged planets.

Chapter 4 presents version 2 of `orbitize!`, a Python package for fitting the orbits of directly imaged planets (featured in Chapter 3). The chapter details the motivation for and implementation of each of the major new features in the new version of the code, giving examples throughout. Chapter 5 presents an updated Gaussian Process regression prescription for the radial velocity fitting software `radvel`, aimed at preventing overfitting (and featured in Chapter 2). This chapter motivates and explains Gaussian Process regression and its application to stellar activity, and lays out the changes in the new implementation. Both packages are open-source and available to use.

Finally, in Chapter 6, I summarize, discuss future directions, and conclude.

Chapter 2

OVERFITTING AFFECTS THE RELIABILITY OF RADIAL VELOCITY MASS ESTIMATES OF THE V1298 TAU PLANETS

Submitted to the Astronomical Journal

S. Blunt, A. Carvalho, T. J. David, C. Beichman, J. K. Zink, E. Gaidos, A. Behmard, L. G. Bouma, D. Cody, F. Dai, D. Foreman-Mackey, S. Grunblatt, A. W. Howard, M. Kosiarek, H. A. Knutson, R. A. Rubenzahl, C. Beard, A. Chontos, S. Giacalone, T. Hirano, M. C. Johnson, J. Lubin, J. M. Akana Murphy, E. A. Petigura, J. Van Zandt, & L. Weiss

2.1 Abstract

Mass, radius, and age measurements of young ($\lesssim$ 100 Myr) planets have the power to shape our understanding of planet formation. However, young stars tend to be extremely variable in both photometry and radial velocity, which makes constraining these properties challenging. The V1298 Tau system of four ~0.5 R_J planets transiting a pre-main sequence star presents an important, if stress-inducing, opportunity to directly observe and measure the properties of infant planets. Suárez Mascareño et al. (2021, hereafter SM21) published radial-velocity-derived masses for two of the V1298 Tau planets using a state-of-the-art Gaussian Process regression framework. The planetary densities computed from these masses were surprisingly high, implying extremely rapid contraction after formation in tension with most existing planet formation theories. In an effort to further constrain the masses of the V1298 Tau planets, we obtained 36 RVs using Keck/HIRES, and analyzed them in concert with published RVs and photometry. Through performing a suite of cross validation tests, we found evidence that the preferred model of SM21 suffers from overfitting, defined as the inability to predict unseen data, rendering the masses unreliable. We detail several potential causes of this overfitting, many of which may be important for other RV analyses of other active stars, and recommend that additional time and resources be allocated to understanding and mitigating activity in active young stars such as V1298 Tau.

Young Planets as Probes of Formation

Planet formation is an uncertain process. Giant planets are thought to form with large radii, inflated due to trapped heat, then cool and contract over the first few hundred Myr of their lives (Marley et al., 2007). However, the accretion efficiency of the formation process, which sets the planets' initial entropy and radii, spans orders of magnitude of uncertainty. The processes sculpting the post-formation masses and radii of smaller terrestrial exoplanets are also uncertain. Young, terrestrial planets also have uncertain initial entropies, and for highly irradiated planets, the unknown rate of photoevaporation (itself due to uncertainties in a planet's migration history, among other physical unknowns) during and after formation compounds this ambiguity (Lopez, Fortney, and Miller, 2012, Owen and Wu, 2013, Chen and Rogers, 2016, Owen, 2020).

Measuring the masses, radii, and ages of newly-formed planets presents a path forward (Owen, 2020). Young moving groups provide rigorous age constraints, and relatively model-independent methods of measuring planetary radii exist for both young directly-imaged and transiting planets (for transiting planets in particular, only stellar radius model dependencies impact the inferred planetary radius). However, in both situations, few model-independent mass measurements exist. For transiting planets, there are two complementary methods for measuring planetary masses: transit timing variations (TTVs), and stellar radial velocity (RV) timeseries.

Measuring RV masses of young planets is a difficult task, so some advocate to rely on transit timing variations (TTVs) alone to measure masses of young planets. However, not all planets transit, and only planets in multi-planet systems at or near mean motion resonance exhibit TTVs (Fabrycky et al., 2014). Even in systems that do, individual TTV mass posteriors are often covariant, since TTVs to first order constrain the planetary mass ratio (Lithwick, Xie, and Wu, 2012, Petigura et al., 2020). In an ideal scenario, both RVs and TTVs would be used to jointly constrain planetary masses in a given system, reducing posterior uncertainty and TTV degeneracies.

Stellar Activity & Overfitting

As the instrumental errors of extremely precise RV instruments approach 10 cm s^{-1}, and as the RV community begins to target more active stars, accurately modeling astrophysical noise is becoming more and more critical. Young stars present a particular challenge. These are highly magnetically active (Johns-Krull, 2007),

with starspots that occupy significant fractions of the stellar surface and induce RV variations on the order of ~km s^{-1} (Saar and Donahue, 1997). These RV variations are hundreds of times larger than the activity signals of older quiet stars typically targeted by RV surveys and complicate the detection of planet-induced Doppler shifts from even close-in Jupiter-mass planets (e.g., Huerta et al., 2008; Prato et al., 2008).

Other assumptions and/or information can be leveraged to model the activity signal, even if the signal isn't easily understandable from the RVs themselves. A widely used practice involves independently constraining the rotation period from a photometric timeseries, then using an informed prior on the rotation period to model the RVs (e.g., Grunblatt, Howard, and Haywood, 2015). Other related examples include specifying a quasi-periodic kernel for a Gaussian Process regression (GPR) model (i.e., assuming that the stellar activity has a quasi-periodic form), or modeling the RVs jointly with other datasets. The latter approach achieves better model constraints either by explicitly modeling the relationship between the datasets (e.g., Rajpaul et al., 2015) or by sharing hyperparameters between datasets (e.g., Grunblatt, Howard, and Haywood, 2015, López-Morales et al., 2016).

As is true for every model-fitting process, misspecifying the stellar activity model (i.e., fitting a model that is not representative of the process that generated the data) or allowing too many effective degrees of freedom can lead to overfitting.

Overfitting is a concept ubiquitous in machine learning, and in particular is often used to determine when a model has been optimally trained. One algorithm for determining whether a model is overfitting is as follows[1]: divide the data into a "training" set and an "evaluation" set (a common split is 80%/20%), and begin optimizing the model using just the training set. At each optimization step, calculate the goodness-of-fit metric for the model on the evaluation set, which is otherwise omitted from the training process altogether. This method of evaluating a model's ability to successfully predict new, or "out-of-sample," data is known as cross validation (CV).

The classic observed behavior is that the goodness-of-fit metrics for both the training and evaluation set improve as the model fits the training data better and better. At a certain point, the model begins to overfit to the training data, and the goodness-of-fit metric for the evaluation data worsens. This is because the model parameters have

[1] See also Cale et al. (2021), who define overfitting in terms of reduced χ^2 in the context of RV activity modeling

begun to reproduce the noise in the training set, at the expense of reproducing the signal common to both datasets. A model that is overfitting, then, can be defined as one that predicts the observations in a training set better than those in an evaluation set. An overfitting model fits aspects of the data that are not predictable or common to the entire data set, e.g., noise.

The optimally trained model is selected not by its performance relative to the *training* data, but by its performance relative to the *evaluation* data, which was omitted from the training process altogether. Making an analogy to Bayesian model comparison, we could imagine a similar process where the goodness-of-fit is evaluated for an "evaluation set" left out of the training process (i.e., posterior computation using MCMC, nested sampling, etc.) for a series of models. One benefit of this method over, e.g., formal Bayesian model comparison is that it also provides an easily-interpretable absolute metric for how well the model fits the data: if the evaluation set goodness-of-fit is significantly worse than that of the training set, we know the model is misspecified, even if it has (comparatively) the lowest Bayesian evidence.

In this study, we apply the CV technique as defined above to evaluate the predictiveness of one particular model fit to one particular star. This is intended as a case study, aiming to inspire further investigation into the extent of and causes of overfitting in RV modeling of young, active stars.

V1298 Tau

V1298 Tauri (hereafter V1298 Tau) is a young system of four $\gtrsim 0.5\ R_J$ planets transiting a K-type pre-main sequence (PMS) star (David et al., 2019a, David et al., 2019b). Very few transiting planets have been discovered around PMS stars (other notable systems being AU Mic, Plavchan et al., 2020, Cale et al., 2021, Zicher et al., 2022, Klein et al., 2022; K2-33, David et al., 2016; DS Tuc, Newton et al., 2019, Benatti et al., 2019; HIP 67522 Rizzuto et al., 2020; Kepler 1627A, Bouma et al., 2022; and TOI 1227, Mann et al., 2022). David et al. (2019a) reported the discovery of V1298 Tau b, a 0.9 R_J object with an orbital period of 24d. David et al. (2019b) discovered three additional planets in the system: V1298 Tau c at 8d, V1298 Tau d at 12d, and a single-transiting object, V1298 Tau e.

It is unclear from their radii ($0.5 - 1 R_J$) alone whether these planets are gas giants that contracted rapidly after forming, or young terrestrial or mini-Neptune planets, which will lose a large fraction of their atmospheres to photoevaporation (Owen and

Wu, 2013) and/or core-powered mass loss (Ginzburg, Schlichting, and Sari, 2018) as they age.

Suárez Mascareño et al. (2021), hereafter SM21, presented over 100 RVs of the system with four different instruments, and used a suite of Gaussian Process (GP) models in an effort to isolate the RV signals of the outer two transiting planets (b and e), finding masses of 0.64 and 1.16 M_J for each respectively. Combined with radii of 0.91 and 0.78 R_J, this implied high densities of 1.2 and 3.6 g cm^{-3}. Such high densities would require rapid contraction after within the 23 Myr lifetime of the system and place the outer V1298 Tau planets at the upper density boundary of even mature field exoplanets, where few theories predict planets to exist. Since V1298 Tau e only transited once while observed with K2, SM21 did not have a period constraint from transits, and derived a period of 40.2 ± 0.9 d from their RV measurements. After the publication of SM21, Feinstein et al. (2022) published updated ephemerides of all four planets using TESS photometry, including a second transit of planet e. They placed a strict lower limit of 42 d on V1298 Tau e's period, in tension with the value from SM21. In addition, Tejada Arevalo, Tamayo, and Cranmer (2022) performed a dynamical stability analysis using the mass measurements reported in SM21, finding that 97% of system configurations consistent with the SM21 posteriors are gravitationally unstable over the lifetime of the system.

SM21 performed a rigorous and state-of-the-art analysis, comparing several complex models with Bayesian methods. However, several independent lines of evidence appear to call the masses they report into question: the tension with formation theory discussed in SM21, the updated planet e orbital period of Feinstein et al. (2022), and the improbability of long-term stability derived by Tejada Arevalo, Tamayo, and Cranmer (2022).

This Paper

The purpose of this paper is two-fold: 1) to demonstrate the use of CV to show that the preferred model of SM21 is overfitting the RVs, and 2) to point out several potential causes of the overfitting, which are not unique to SM21 but common throughout the literature. We do not attempt to update the mass estimates for the V1298 Tau planets in this paper. We also do not attempt to prove that the mass estimates published in SM21 are incorrect, but instead seek to call into question their reliability. Our argument is that future joint models of the stellar activity and

planetary signals of V1298 Tau will need to prove their predictiveness in order to be trustworthy.

The structure of this paper is as follows: in Section 2.2, we review the literature data scrutinized in this paper and describe one additional contemporaneous RV dataset taken with Keck/HIRES. In section 2.3, we demonstrate that the preferred model of SM21 is overfitting. Section 2.4 discusses several potential causes of this overfitting, and advises modelers on how to detect and/or avoid these subtle pitfalls. In particular, Section 2.4 argues that differential rotation is an important effect for V1298 Tau, and must be modeled carefully. We conclude in Section 2.5. We also provide an appendix that provides a geometric interpretation of how GPR penalizes complexity.

2.2 Data

Throughout this paper, we reference several data sets: three photometric time series measured by different instruments and three RV timeseries derived from spectra measured by different instruments. Each dataset is detailed in the subsections below. All of the photometry is shown in Figure 2.9, and all of the RVs are shown in Figure 2.1.

K2 photometry

We downloaded EVEREST-processed (Luger et al., 2016, Luger et al., 2018) K2 lightcurves for V1298 Tau using the `lightkurve` package (Lightkurve Collaboration et al., 2018). We used built-in `lightkurve` functions to remove nans, remove outliers, and normalize the data.

LCO photometry

We obtained ground-based LCO photometry verbatim from SM21.

TESS photometry

We obtained TESS lightcurves from Feinstein et al. (2022), who combined timeseries photometry of V1298 Tau from TESS Sectors 43 and 44. Feinstein et al. (2022) used the 2-minute light curve created by the Science Processing Operations Center pipeline (SPOC; Jenkins et al., 2016), and binned those observations to 10 mins. We normalized the data for each TESS orbit separately, following Feinstein et al. (2022).

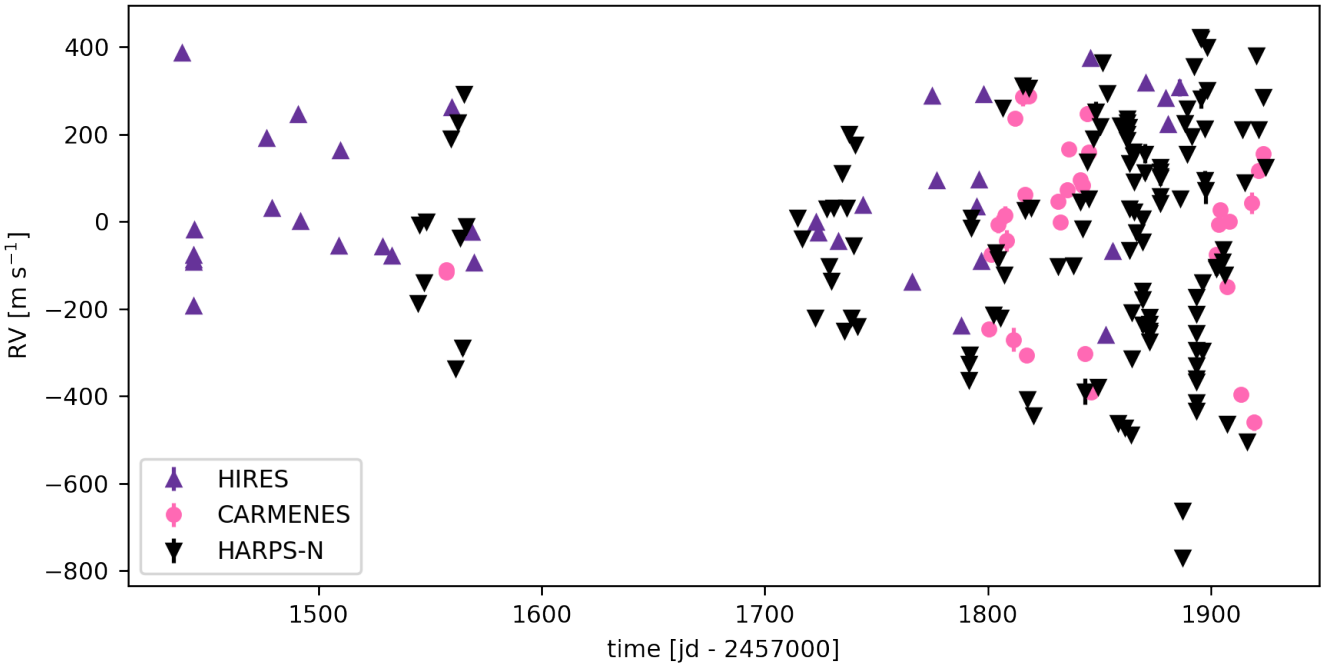

Figure 2.1: A tour of the RVs scrutinized in this study. The CARMENES and HARPS-N RVs are published in SM21, and the HIRES RVs are new in this study. **Takeaway:** the RV variability of V1298 Tau is hundreds of m/s, which is similar across all three instruments. The variability is significantly greater than the instrumental errors (which are included, but too small to see for the majority of points on this plot).

SM21 RVs

We obtained CARMENES and HARPS-N RVs directly from SM21. We note that SM21 excluded infrared-arm CARMENES RVs in its analysis, and we do the same here.

HARPS-N RVs are wavelength calibrated using a ThAr lamp, and the HARPS-N spectrograph covers 360-690 nm.

The visible arm of the CARMENES instrument covers the spectral range 520-960 nm, and spectra from this instrument are wavelength calibrated using a Fabry-Perot etalon, anchored using hollow cathode lamps.

Keck/HIRES RVs

Between November 16, 2018, and February 6, 2020, we obtained 36 RVs using the HIRES spectrograph on the Keck I telescope (Vogt et al., 1994). Wavelength calibration was performed by passing starlight through a warm iodine cell, and data reduction was performed using the California Planet Search pipeline described in Howard et al. (2010), which is adapted from Butler et al. (1996b). All HIRES RVs used in this study are given in Table 2.1. Some of these RVs were previously published in Johnson et al., 2022, and the processing is identical in that paper and this. The same stellar template, constructed from two stellar spectra taken on 24 Oct 2019 UT without the iodine in the light path, was used to derive RVs in both studies. In-transit RVs from that study have been excluded here. Spectra were typically taken using the C2 decker (14" x 0.861"), which enables sky subtraction, and is the CPS HIRES observer "decker of choice" for stars fainter than V∼10. However, a CPS HIRES observer "rule of thumb" is to use the shorter B5 decker (3.5" x 0.861") in poor seeing conditions, as the Doppler pipeline sky subtraction algorithm is unreliable when the stellar PSF fills the slit. Sky subtraction is not performed under such conditions. Accordingly, 7 RVs published here were calculated from spectra using the B5 decker. In both modes, HIRES has a resolving power of ∼60,000, and the iodine cell spectral grasp translates to contributions to the RV from wavelengths between 500 and 620 nm (Butler et al., 1996b).

2.3 Cross Validation Tests

Our intention in collecting additional RVs of V1298 Tau with HIRES was to jointly analyze these data together with literature data and update the masses published in SM21. However, early on in the analysis, we noticed clues that made us question our assumptions. In particular, the new data we had collected did not seem consistent with the models of SM21. In addition, many tested models converged on results that were physically unreasonable or clearly inconsistent with subsets of the data. We ultimately decided to test the predictive capability of the SM21 model that we were using as our starting point, as a check on our own assumptions. This section details the outcome of those experiments.

The main finding of this paper is that the median parameter estimate of the preferred model of SM21 (their $\mathbf{4p}_{PQP2}$) is overfitting. For convenience, we will refer to this model throughout the rest of this paper as the "SM21 preferred model." Showing that a point estimate is overfitting does not necessarily indicate that every model spanned by the posterior is overfitting. However, since the preferred model presented by

Table 2.1: New HIRES RV Data. Epochs are reported in units of BJD_{TDB}, with a 2457000 offset applied.

jd - 2457000	RV [m s^{-1}]	RV error [m s^{-1}]
1438.9463	387.47	7.885
1443.8205	-193.32	8.528
1443.9573	-87.96	7.941
1443.9682	-77.17	7.073
1443.9792	-92.67	8.609
1444.1566	-19.40	8.403
1476.8089	190.28	11.712
1479.0060	30.83	10.269
1490.7912	246.54	10.047
1491.7549	0.31	8.804
1508.9196	-56.33	9.530
1509.7900	163.24	12.032
1528.7841	-58.17	9.752
1532.7701	-78.17	10.465
1559.7316	261.44	9.531
1568.7290	-23.72	9.683
1569.7376	-94.99	11.022
1723.1354	-0.31	8.617
1724.0917	-26.27	9.648
1733.0739	-44.28	9.662
1743.9965	37.42	10.460
1765.9266	-138.49	11.108
1774.9154	287.77	9.431
1777.0163	94.72	9.400
1787.9817	-239.82	10.006
1794.8691	34.79	8.829
1795.9062	95.70	8.246
1796.9165	-90.09	8.827
1797.9532	292.26	8.131
1845.8966	375.28	9.472
1852.7705	-259.55	9.593
1855.8832	-68.08	16.101
1870.8459	318.54	8.901
1879.8314	282.03	9.115
1880.8314	222.22	8.493
1885.8611	306.95	19.622

SM21 (their figure 11) appears approximately Gaussian around the MAP estimates of the parameters relevant for us, (except for the kernel parameter C and the white noise jitter for CARMENES, which both peak at effectively 0), we assume that the MAP and median for this fit are close enough to make no difference, and that inferences made about the median fit hold true for other high-probability areas of parameter space.

Our goal was to test the predictiveness of the preferred SM21 model[2] using CV. In an ideal situation, we would do this by evaluating the model's performance on new HARPS-N data, unseen by the trained model. Lacking this, we constructed two ad hoc "validation sets:" a timeseries of Keck/HIRES data contemporaneous with the SM21 HARPS-N data, and the CARMENES data presented in SM21 (that the model was also trained on, but which were treated as independent from the HARPS-N data; see section 2.4). By chance, this results in a nearly perfect 80%/20% split for both validation sets (80.3%/19.7% for HARPS-N/CARMENES, and 78.9%/21.1% for HARPS-N/HIRES). In Figures 2.2 and 2.3, we show two visualizations of the results of performing CV on these two validation sets. Figure 2.2 shows the GP prediction of the SM21 preferred model, together with the HARPS-N data on which it was trained and conditioned. The contemporaneous HIRES data and CARMENES data and their residuals are overplotted. Figure 2.3 shows the residuals of this fit, given in terms of standard deviations from the mean GP prediction. In both figures, the residuals of the HIRES and CARMENES data have a much wider spread about 0 than the HARPS-N points. Because our intention was to evaluate the existing model, we did not re-train the GP hyperparameters in order to compute the prediction shown in Figure 2.2. Rather, we used the median parameters of the SM21 $\mathbf{4p}_{PQP2}$ model, conditioned on the HARPS-N data published in that study, to predict RV values at each of the CARMENES and HIRES epochs.

Our interpretation of the difference in residual distributions shown in these two figures is that the preferred SM21 model fits data included in its training set (i.e., the HARPS-N data) significantly better than *contemporaneous* data not included. In other words, the model is not predictive. This is a hallmark of overfitting, and indicates that the preferred SM21 model is not representative of the process generating the data.

An important counter-interpretation is that the V1298 Tau RVs measured by HARPS-N, HIRES, and CARMENES show different activity signals, and not that the pre-

[2] conditioned on the HARPS-N data of SM21; see Section 2.4.

ferred SM21 model is overfitting. In particular, starspots cooler than the stellar photosphere cause RVs collected in redder bands, where the contrast between spot and photosphere is lower, to show lower variability amplitudes (e.g., Carpenter, Hillenbrand, and Skrutskie, 2001; Prato et al., 2008; Mahmud et al., 2011). In addition, we expect different instruments to have different RV zero-point offsets. **Importantly, these two effects cannot explain the increased out-of-sample residual spread observed in Figures 2.2 and 2.3**[3]; the preferred SM21 model fitted the CARMENES zero-point offset, white noise jitter value, and activity amplitude, and those values have been applied to the CARMENES data. To account for the potential differences between the HIRES and HARPS-N RVs, we applied an RV zero-point offset and scale factor (0.76) that minimizes the residual spread (i.e., we applied a best-fit linear model to the HIRES data in order to minimize χ^2 with respect to the GP model prediction). See section 2.4 for further discussion of this point.

Another potential explanation for the phenomenon observed in Figures 2.2 and 2.3 is that the activity signals observed by HARPS-N, CARMENES, and HIRES are fundamentally different; i.e., the signal observed by one instrument is not a linear combination of the signal observed by another. This might occur because, for example, all three instruments have ~km/s instrumental systematics relative to one another, or because the shape of the activity signal changes significantly with wavelength. To rule out this explanation and provide more evidence that the effect we're seeing is actually overfitting, and not instrument-specific differences, we repeated the experiment above using only HARPS-N data. We randomly selected 80% of the HARPS-N data published in SM21, conditioned the preferred SM21 model on that subset, and computed the residuals for the random "held-out" 20%. The results are shown in Figures 2.4 and 2.5. Even though these held-out 20% were included in the training process (i.e., they informed the values of the hyperparameters), we observed substantially larger residuals than for the conditioned-on subset. This experiment provides additional evidence for overfitting, and not instrumental- or wavelength-dependent systematics.

It is worth noting that we distinguish between residual distributions (Figures 2.4 and 2.3) "by-eye" in this paper, but this technique will not generalize for more similar residual distributions. Residual diagnostic tests (see Caceres et al., 2019 for an example) will be helpful in generalizing this methodology.

[3] assuming that stellar activity signals observed by different instruments can be described as linear combinations; see Section 2.4.

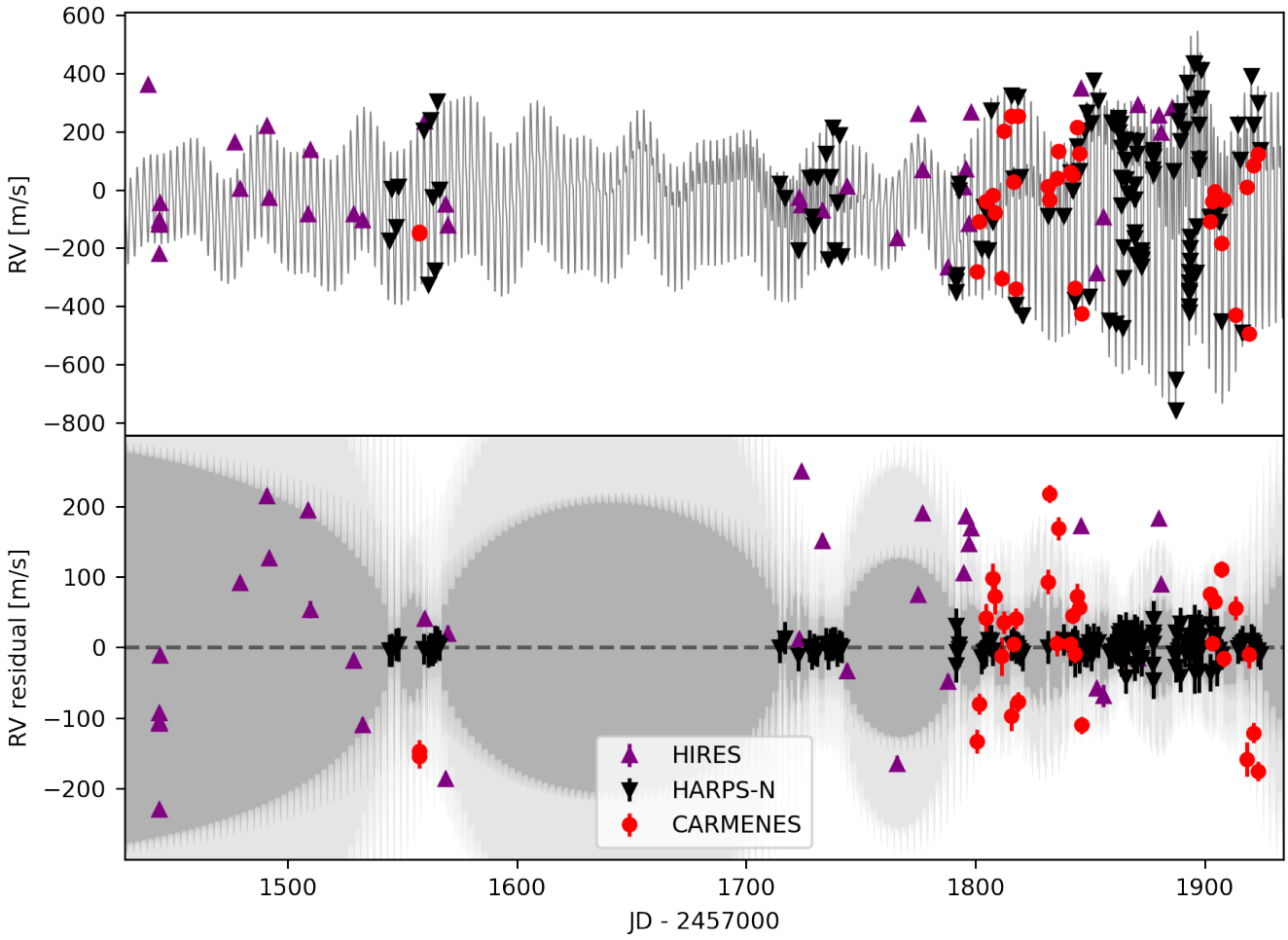

Figure 2.2: SM21 preferred model prediction and contemporaneous observed data. The HIRES data have been scaled and offset by linear parameters that minimize the residual spread with respect to the GP model, and the median $\mathbf{4p}_{PQP2}$ CARMENES data RV zero-point value was been applied in order to more easily compare both datasets with the model expectations. **Top:** mean model prediction (gray solid line), together with contemporaneous HARPS-N (black), CARMENES (red), and HIRES (purple) RVs overplotted. **Bottom:** model residuals, together with 1- and 2-σ GP uncertainty bands (shaded dark and light grey regions, respectively). **Takeaway:** The preferred SM21 model is overfitting to the HARPS-N data, which can be seen in the increased spread about the residual=0 line for both HIRES and CARMENES data during epochs with contemporaneous HARPS-N data.

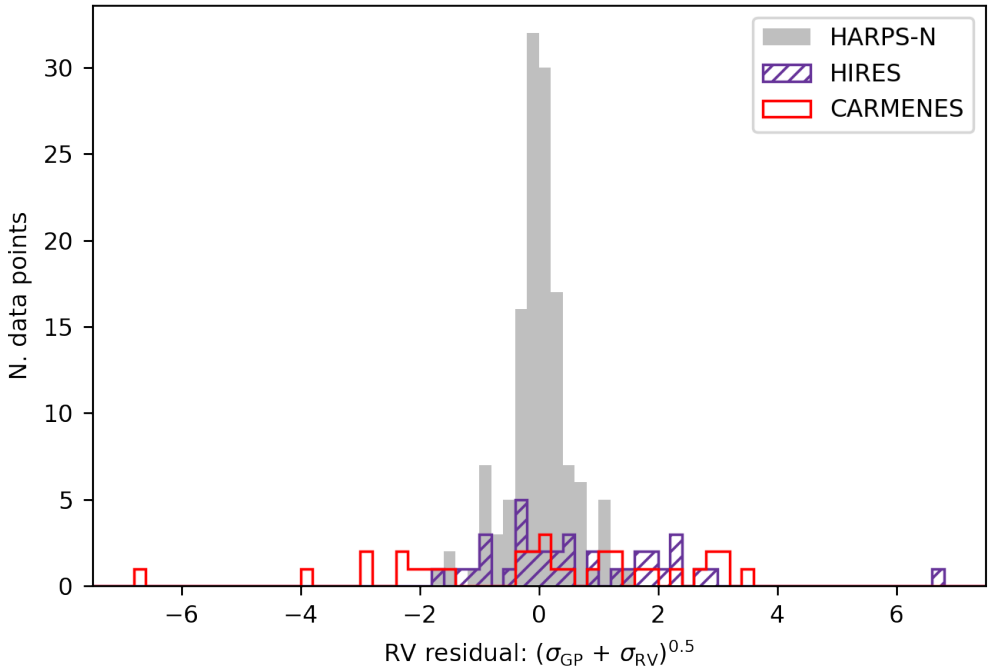

Figure 2.3: Another visualization of Figure 2.2. Histograms of the RV residuals, given in units of standard deviations from the mean prediction. **Takeaway:** The broader and more uniform distribution of HIRES and CARMENES residuals relative to the HARPS residuals is another hallmark of overfitting.

2.4 Potential Causes of Overfitting

This section points out several potential causes of the overfitting described in the previous section, and advises on how to detect and/or ameliorate these effects. We do not attempt to quantify the effect of each of these on the overfitting discussed in the previous section, but intend this as a qualitative discussion. Many of these effects are potentially relevant for stars other than V1298 Tau.

Importantly, this is not a list of "mistakes," but a list of assumptions we questioned throughout the process of trying to explain why the preferred SM21 fit was overfitting. We encourage future close investigation of each of these phenomena, both for V1298 Tau and other objects. This list is not exhaustive.

Correlated Datasets vs Datasets that Share Hyperparameters

The mathematical formalism in this section is essentially identical to that of Cale et al. (2021, see their section 3.2), but was developed independently. We encourage

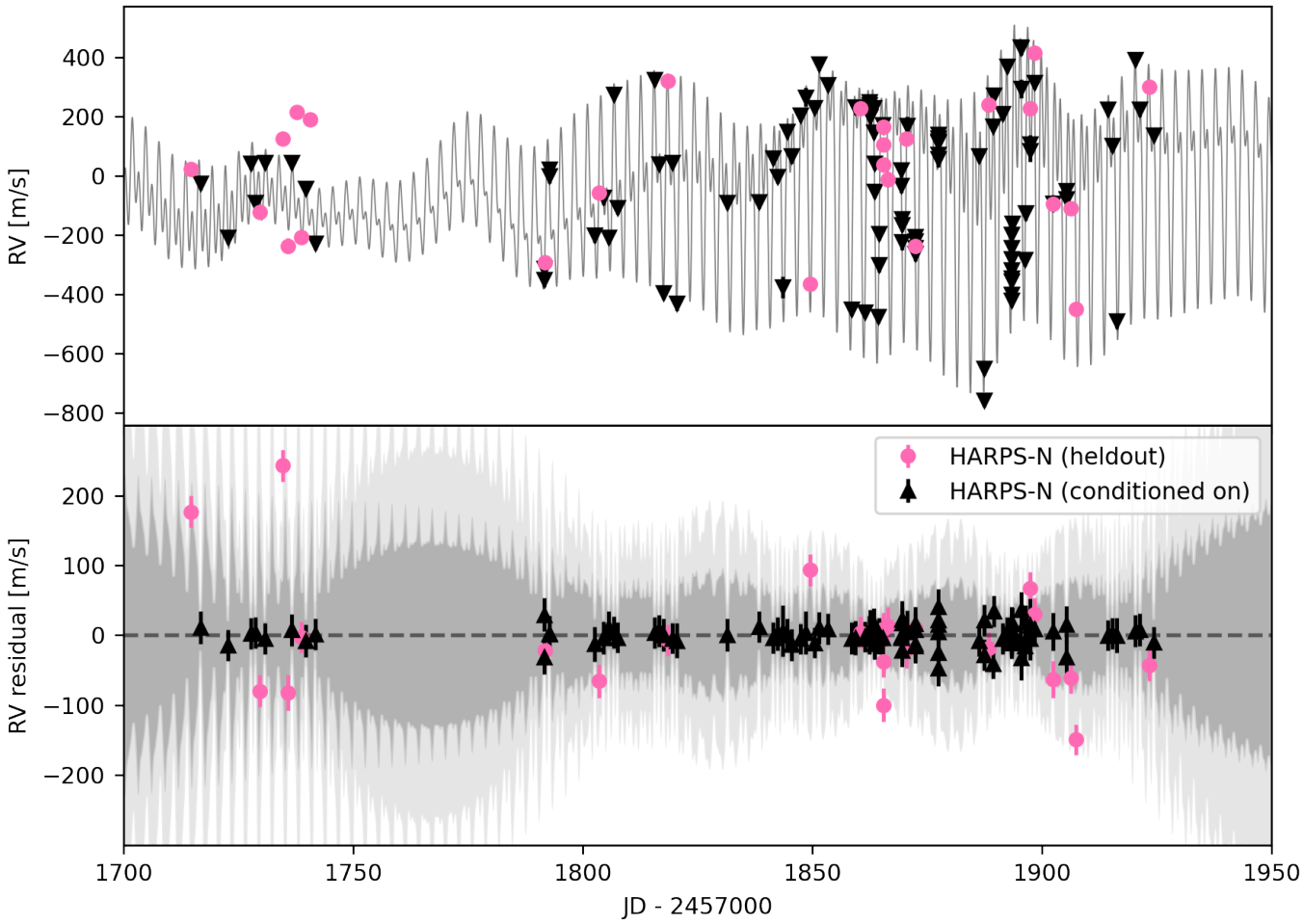

Figure 2.4: Same as Figure 2.2, except that the model prediction is computed by conditioning on a randomly-selected 80% subset of the HARPS-N data, as described in the text, as the residuals are computed for the 20% subset that was held-out. **Takeaway:** The effect seen in Figure 2.2 cannot be explained by instrument- or wavelength-dependent systematics, because the same larger residuals are seen within the data taken by only HARPS-N.

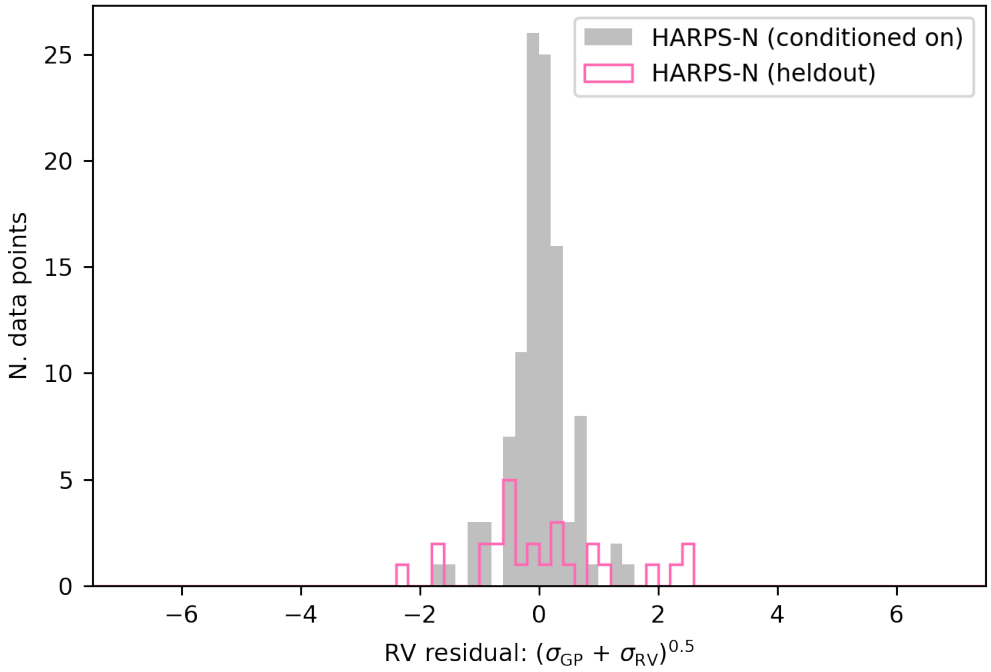

Figure 2.5: Another visualization of Figure 2.4. Same as Figure 2.3, except computed using the same method as for 2.4. **Takeaway:** the larger and more uniform spread of residuals for HARPS-N data on which the model was conditioned provides more evidence that the preferred SM21 model is overfitting.

readers to compare our explanations, and we ask readers to also cite Cale et al. (2021) whenever referencing Section 2.4 of this paper.

There is a difference between correlated measurements that are allowed to have different GP amplitudes and datasets that share GP hyperparameters but are themselves uncorrelated. We are motivated to stress this distinction by the need in RV timeseries fitting to write down the joint likelihood of a model applied to datasets taken from several different instruments. As a concrete example, let's consider three fictional RV data points, the first two from HIRES and the next one from CARMENES, to which we would like to fit a GP model. Because of the different bandpasses of HIRES and CARMENES, we might expect the same stellar activity signal to have a different amplitude when observed by these two instruments. However, we might expect the time-characteristics of the signals to be identical. **In other words, we expect the CARMENES activity signal to be a scalar multiple of the HIRES**

activity signal.[4] As discussed in Section 2.3, this assumption is borne out, at least to first order, in observations of other active stars at different wavelengths (see, e.g., Mahmud et al., 2011, who investigated the RV activity of the T-Tauri object Hubble I 4 with contemporaneous infrared and optical spectra taken with different instruments), but this point warrants further scrutiny. Comparing the variability of active stars with different instruments, as well as the variability of the sun with different solar instruments, is an important endeavor.

Another important caveat is the use of different techniques for computing RVs from stellar spectra (e.g., the iodine/forward-modeling technique of HIRES vs simultaneous reference/CCF technique of CARMENES and HARPS-N). Switching from one of these techniques to another is not expected to affect an astronomer's ability to recover common Keplerian signals, but spot activity is not a simple Doppler shift. More work is needed to understand and model spot activity at the spectral level. We proceed by assuming that modeling the same spectrum using an iodine/forward-model and with a simultaneous ThAr lamp reference (as an example) will only change the effective wavelength range of the spectrum that is used to compute RV, and therefore affect only the amplitude of spot-induced variations.

Assuming linearly-related GPs for different instruments, we can write down the joint covariance matrix for our three fictional data points, allowing unique amplitude terms $a_{\rm C}$ and $a_{\rm H}$ for each dataset, and assuming an arbitrary kernel function $\mathrm{k}_{i,j}$ describing the covariance between RVs at times t_i and t_j:

$$C_{\rm joint} = \begin{pmatrix} a_{\rm H}^2 k_{0,0} & a_{\rm H}^2 k_{0,1} & a_{\rm H} a_{\rm C} k_{0,2} \\ a_{\rm H}^2 k_{1,0} & a_{\rm H}^2 k_{1,1} & a_{\rm H} a_{\rm C} k_{1,2} \\ a_{\rm C} a_{\rm H} k_{2,0} & a_{\rm C} a_{\rm H} k_{2,1} & a_{\rm C}^2 k_{2,2} \end{pmatrix}. \tag{2.1}$$

Optimizing the hyperparameters of a fit that uses this covariance matrix to define the GP likelihood will give the desired result.

SM21, following many other fits in the literature, constructed an independent covariance matrix for each RV instrument in their dataset and summed the log(likelihoods) given by these together. This allows each RV dataset to be independent; i.e., a datapoint taken by HIRES is not correlated with a datapoint taken at exactly the same time by CARMENES. Figures 2.11 and 2.12 illustrate the difference between these two likelihood definitions for data for a different object (chosen because it is easier to see the effect using this dataset).

[4] With different RV offsets as well, so technically a linear combination, not just a scalar multiple.

This assumption of independent data for each instrument effectively adds additional free parameters to a model, and makes it more susceptible to overfitting. This is also why, in Figures 2.2 and 2.3, we could demonstrate that the preferred SM21 model was overfitting by comparing the model prediction conditioned on HARPS-N data to the CARMENES data; the CARMENES data influenced the final values of the hyperparameters, since they were shared between the two Gaussian processes, but otherwise the datasets were treated as independent.

To go along with this paper, we will release a new version of `radvel` (Fulton et al., 2016), built on `tinygp` (Foreman-Mackey et al., 2022), that treats the models for different instruments as correlated, but allows each instrument its own GP amplitude, white noise jitter term, and RV zero-point offset term.[5] The difference between the previous version of `radvel` and this version is also illustrated in Figures 2.11 and 2.12 in the Appendix.

Future work should continue to test this assumption by obtaining simultaneous (or near simultaneous) RVs for a variety of stellar types with different instruments, across a wide range of bandpasses.

P_{rot} and $P_{rot}/2$

Another practice that may have made the SM21 preferred fit susceptible to overfitting involves constructing a GP kernel with one term at the rotation period and another term at its first harmonic. In other words, the SM21 preferred model kernel has the following form:

$$C_{ij} = f_1(P_{\rm rot}) + f_2(P_{\rm rot}/2), \tag{2.2}$$

To understand the motivation for this, we first need to scrutinize the RV signal in Fourier space. Figure 2.6 shows the Lomb-Scargle periodogram of all RV data presented in SM21, zooming in on two important parts of period space. There are four extremely significant peaks in the RVs, which can all be explained with a single periodic signal at 2.91d, the rotation period identified by SM21. Along with a strong peak at 2.91d (hereafter P_{rot}), there is a signal at $P_{rot}/2$, which is often observed in RVs of stars showing starspot-induced variability (Nava et al., 2020). The other two strongly significant peaks can be explained as 1-day aliases of P_{rot} and $P_{rot}/2$. In other words, the dominant RV signal is periodic, but requires a two-component

[5] This is slightly different from the GP prescription in `juliet` (Espinoza, Kossakowski, and Brahm, 2018), which does not allow different amplitudes for individual RV instruments.

sinusoidal fit (i.e., it needs more terms in its Fourier expansion) in order for the fit to reproduce the shape of the curve. This is visualized in Figure 2.7, which shows the RVs phase-folded to P_{rot}. **In summary, the RV curve comprises a single periodic pattern, but that pattern is not a simple sinusoid.**

The preferred SM21 model kernel sums two approximately quasi-periodic terms, one at P_{rot} and one at $P_{rot}/2$, because the approximate quasi-periodic kernel used in SM21 (SM21 equation 1; derived in Foreman-Mackey et al., 2017) is less flexible than the standard quasi-periodic kernel (SM21 equation 3). In other words, the approximate kernel is less capable of fitting non-sinusoidal shapes. However, each term was modeled with its own independent exponential decay timescale. This adds an additional free parameter to the fit, which exacerbates the potential for overfitting.

The most straightforward way to address this is to construct a model with fewer unnecessary free parameters, for example by equating the parameters L_1 and L_2 in SM21 equation 1. A more complicated suggestion, which would be an excellent avenue for further study, is to leverage the correlation between the photometry and RVs, following, for example, Rajpaul et al. (2015). This requires assuming (or fitting for) a relationship between a photometric datapoint and an RV datapoint at the same time. Our preliminary investigations along these lines indicate that the FF' formalism, which models an RV signal as a function of a simultaneous photometric (F) dataset and the time derivative of the photometric dataset (F'; Aigrain, Pont, and Zucker, 2012), does not allow for a good phenomenological match between the LCO photometry and the contemporaneous RVs, but the derivative of the LCO photometry appears to fit better (i.e., the RV curve appears to be possible to model as a linear combination of the F' component only)[6]. Future work could write down a joint GP formalism that models RVs as the time derivative of the photometry (such a formalism would be very similar to that of Rajpaul et al., 2015).

Regardless, in order to be confident in the relationship between the photometry and the RVs, as well as to pick out the components of the RV that do not occur at P_{rot}, **we suggest a very high-cadence (several observations per night) RV follow-up campaign** with contemporaneous photometry[7] in order to develop a high-fidelity model of the stellar variability[8]. It is important to note that this campaign need

[6]This was also noted in SM21.

[7]As of 1-30-23, V1298 Tau will unfortunately not be reobserved with TESS through year 6. We used `tess-point` (Burke et al., 2020) to make this determination.

[8]It is worth pointing out that similar strategies have been successful before, e.g., to measure the mass of Kepler-78 b (Pepe et al., 2013, Howard et al., 2010)

not be performed by an RV instrument with 30 cm s^{-1} precision; Johnson et al. (2022) demonstrated 6-7 m s^{-1} RMS precision with HIRES over several hours, even though the stars moves by hundreds of m s^{-1} over even a single night. This level of instrumental RV error should be sufficient to understand the stellar activity, so long as the cadence is as high as possible.

Keplerian Parameters Enable Overfitting in the Presence of Un-modeled Noise
A Keplerian signal has five free parameters (semi-amplitude, eccentricity, argument of periastron, time of periastron, and period). A model with two Keplerian signals therefore has 10 additional free parameters than a model without. To first order, more free parameters means more model flexibility. This problem can be addressed using model comparison, which penalizes complexity. However, if there is un-modeled noise in the data, including additional Keplerian signals in the model can lead to overfitting; for example, high eccentricity Keplerian models have similar properties to delta functions, which have relatively "flat" RV curves, except for a spike in RV near periastron. With insufficient sampling, outlier data points can be overfit with eccentric Keplerian signals.

A common worry in the RV modeling community is that using GPR to model stellar activity will "soak up" Keplerian signals, leading to underestimates of Keplerian RV semi-amplitudes (discussed in Aigrain and Foreman-Mackey, 2022), even when modeled jointly. However, we find evidence for the opposite effect in the SM21 preferred fit: that the Keplerian signals function as extra parameters that make the model susceptible to overfitting, and the GP is forced to compensate. Examining Figure 2.8, which shows the contributions to the mean model prediction from the Keplerians and the activity-only portion of the mean GP model[9], we find that the activity model interferes with the Keplerian model where RV data exists. This is seen most readily when smoothing the activity model over several rotation periods (effectively low-pass filtering the activity model).

We can explain this behavior by imagining that there is some un-modeled noise source in the data that is inconsistent with Keplerian motion or quasi-periodic variability (see next section). If some non-physical combination of parameters fits the data better at an epoch with many data points that is affected by this noise source, this may outweigh the negative Bayesian evidence contributions from 1) the added complexity and 2) the worse fit at epochs with fewer data points. We would then

[9]The activity-only portion is isolated following SM21, subtracting the Keplerian mean model from the total mean GP prediction.

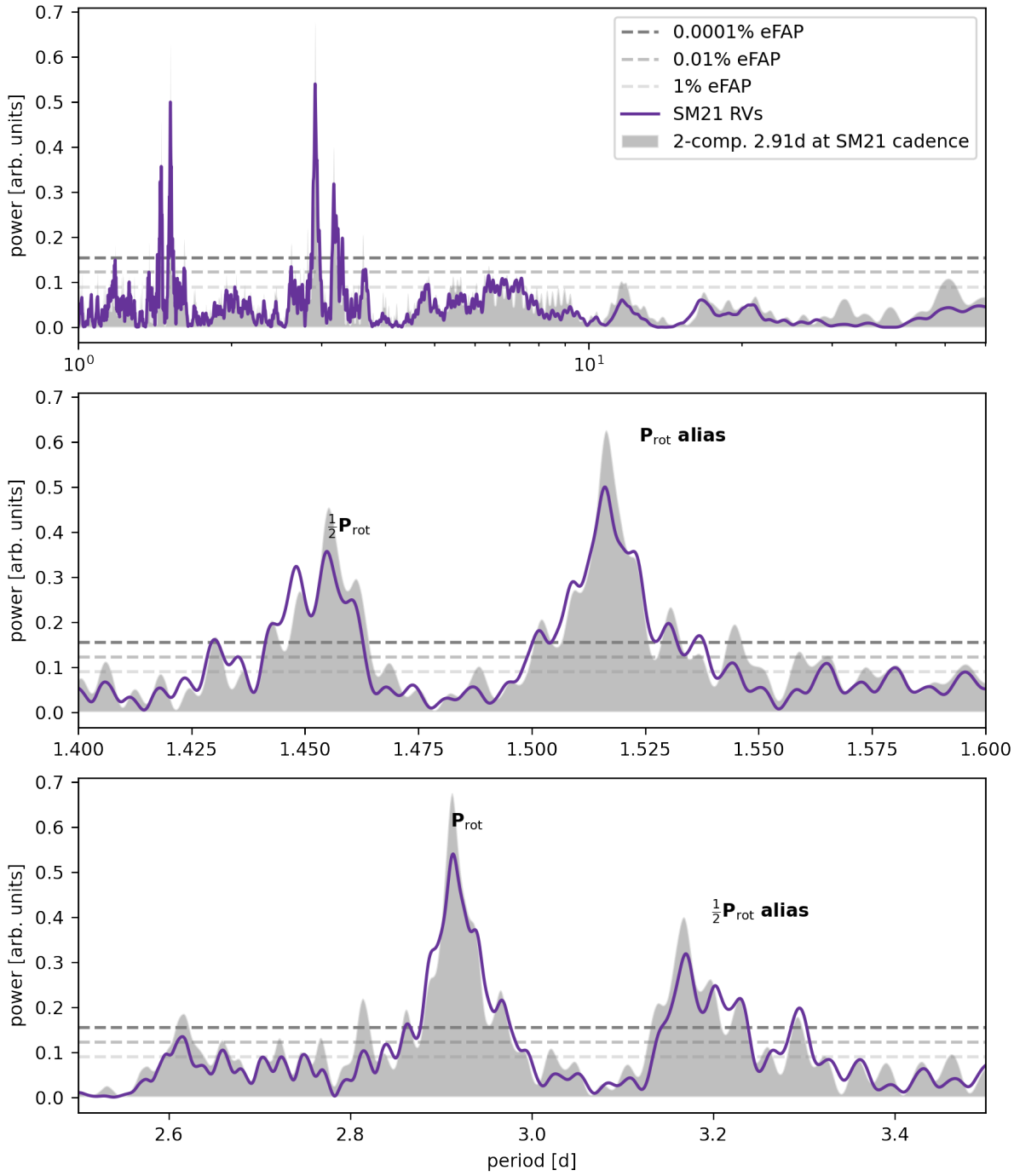

Figure 2.6: Lomb-scargle periodogram of all RV data presented in SM21, and 2-component sinusoidal fit passed through the same window function. **Top:** Periodogram of all RVs (solid purple line) and a 2-component sinusoidal fit to the data (filled grey). **Middle/bottom:** same, but zoomed in. The rotation period, its harmonic, and its 1d aliases are labeled. **Takeaway:** the dominant Lomb-Scargle periodogram structure can be explained as harmonics and aliases of a single period at 2.91d.

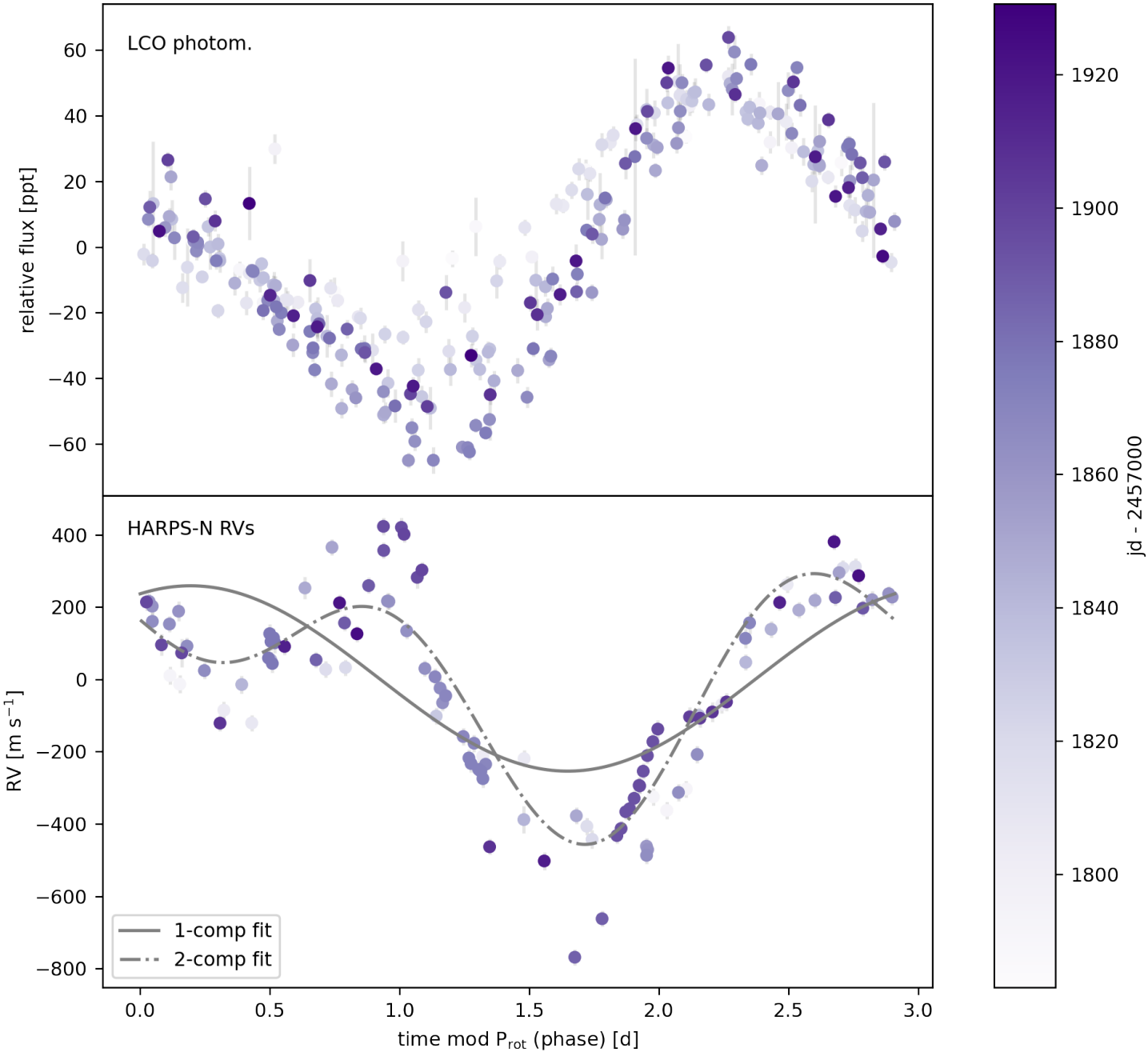

Figure 2.7: HARPS-N RVs and contemporaneous LCO photometry from SM21, phase-folded to the rotation period and colored by observation time. **Top:** LCO photometry. **Bottom:** HARPS-N RVs, with fitted jitter values from the preferred SM21 fit added to the error bars. 1- and 2-component sinusoidal fits are also shown. **Takeaway:** the presence of a strong periodogram peak at $P_{rot}/2$ results from the higher-order shape of the RV rotation pattern. This pattern is not present in the LCO photometry, which is approximately sinusoidal over the rotation period.

expect the Keplerian model to *oversubtract* at epochs with fewer data points (e.g., around jd = 1725 in Figure 2.8).

This effect suggests that the Keplerian signals in the SM21 preferred fit are not a viable description of the RV variability at timescales greater than the rotation period. More effort certainly needs to be spent understanding this phenomenon, but in the meantime we suggest performing CV tests in order to detect overfitting of this nature.

Differential Rotation

The previous subsections all argue that the preferred SM21 fit had too many free parameters (or effective free parameters) that allowed the model to overfit. In other words, we have argued that a simpler model (one for which the GP predictions for each instrument are scalar multiples of each other, a single period is present in the kernel, and no Keplerian signals are present in the model) would be more predicitive, albiet perhaps with larger uncertainties. In this section, we suggest that this much simpler proposed model is still insufficient, because the host star has multiple, differentially rotating, active regions.

Differential rotation may not be the un-modeled noise source that we propose is affecting the SM21 preferred fit. The conclusions of this paper do not change if this is true. We discuss it here because it is potentially widely relevant, especially for young stars. We call for more work on modeling and understanding differential rotation in RVs.

Evidence for a Strong Differential Rotation Signal from Photometry

In the K2 and TESS photometry of V1298 Tau (Figure 2.9), two periodic signals of different amplitudes are visible by eye. These peaks are coherent in phase towards the end of both baselines, producing a larger overall photometric variability amplitude. Although each baseline covers only a portion of the beat periods implied by these different periods coming into and out of phase, the beating "envelope" is still easily distinguished. To guide the eye, we over-plotted the shape of the beating envelope formed by the three dominant periods in the Lomb-Scargle periodogram of the K2 data.

Multiple closely-related periodicities are also apparent in the periodograms of the K2 and TESS data (and the LCO data, albeit at lower significance, potentially due to the lower cadence of that dataset; Figure 2.10). In particular, over both the K2

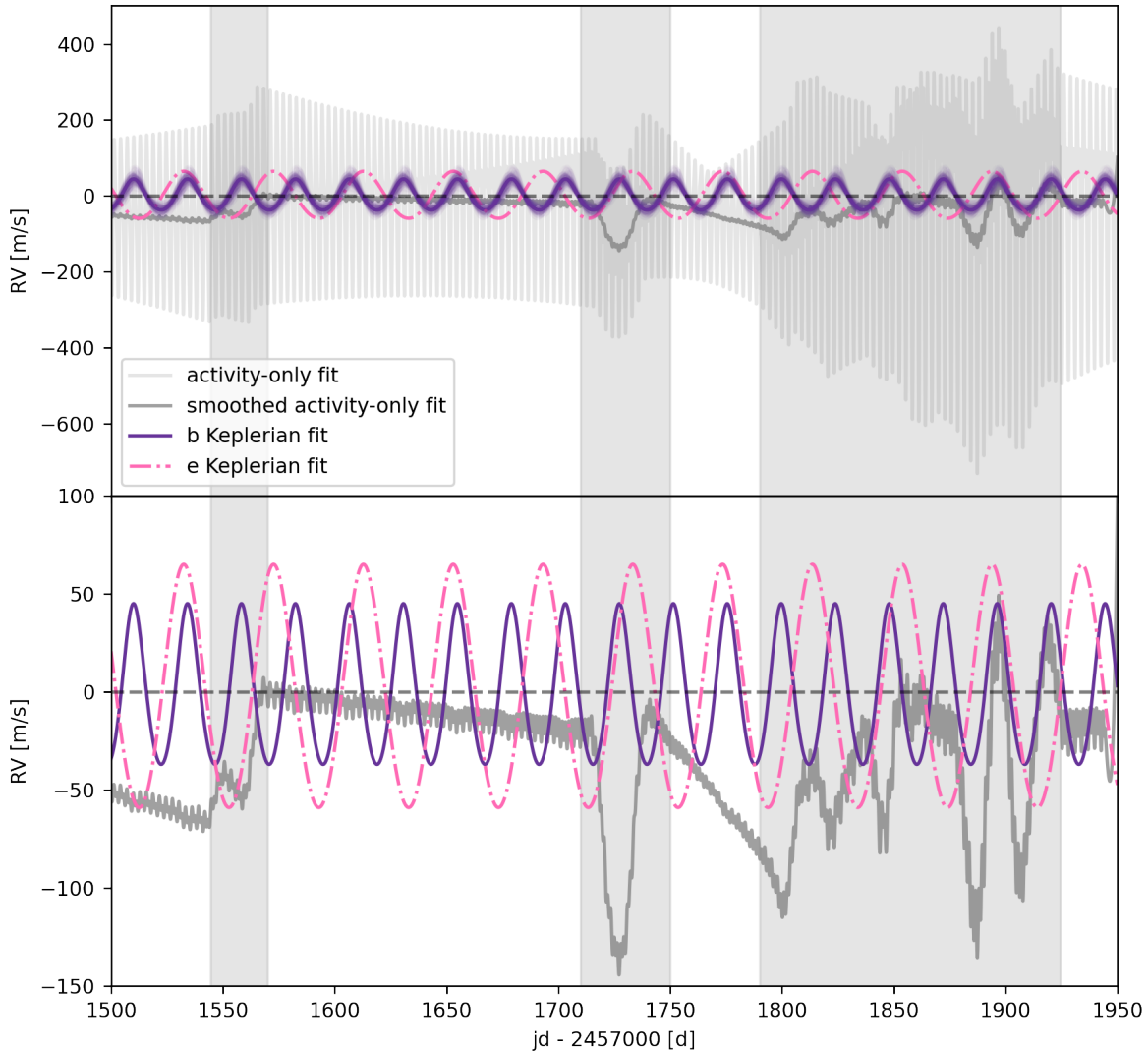

Figure 2.8: Smoothed activity-only component of the preferred model of SM21, together with the Keplerian model components. **Top:** 100 random draws from the posterior describing the planet b Keplerian are also shown, to illustrate that this effect holds true across the posterior, and not simply for one point estimate. The light gray solid line shows the full activity-only model component, and the darker grey shows this model averaged over a (randomly chosen) 11.2 d timescale. (Note that the same pattern holds when choosing a slightly different smoothing timescale; i.e., this is not a result of aliasing.) Shaded grey regions indicate where there are observations. **Bottom:** same as top, but with a zoomed-in y axis. **Takeaways:** the activity-only component changes suddenly in windows of time where there are observations. When the activity-only component is averaged over shorter-timescale variations, the GP contributes to the fit on timescales similar to the Keplerians, even interfering destructively at some times. This casts doubt on the reality of the Keplerian signals reported in SM21, indicating that they may be favored because of overfitting.

and TESS baselines, a dominant periodicity at 2.85 and 2.92 d, respectively, and two less prominent periodicities (one at a larger period, and one at a smaller period) are present. The multiple periodicities in the light curve, visible both in the shape of the beating envelope and in Fourier space, have often been interpreted as a smoking gun of differential rotation (see, e.g., Lanza, Rodono, and Zappala, 1994, Frasca et al., 2011). It is important to note, however, that short spot lifetimes may also produce the observed photometric pattern, and have been shown in simulations to be easily confused with differential rotation (see, e.g., Basri and Shah, 2020). Longer photometric time baselines than are available in the photometric data presented in this paper are needed to distinguish between the two. The conclusion of this section (that there is a noise source visible in photometry that is un-modeled in the SM21 preferred model) would remain unchanged in this case, but this interpretation has important implications for future modeling efforts. That the signals arise from a close binary is ruled out by the multiple nearby periods in the light curve (rotation of two tidally extended binary stars can produce a similar pattern, but with a single period), while astroseismic pulsations are ruled out by the amplitude and period of the variability; V1298 Tau is a PMS $1.2 M_\odot$ star with $\log g$=4.48 (SM21), which we would expect to be oscillating on the scales of minutes and $\lesssim$1 ppt, not days and 20 ppt (Chaplin and Miglio, 2013; see their Figure 3).

Effect on RVs

Assuming that V1298 Tau is differentially rotating, it is possible that the combination of a multiply periodic structure with insufficient cadence is leading the GP to prefer a more complex model. In other words, the data is not consistent with a quasi-periodic structure, so a simple quasi-periodic model will not be preferred over a more complex model (e.g., one with Keplerians at longer periods), even if neither is predictive. Even a secondary active region with 5% the RV amplitude of the primary structure (reasonable given the photometric amplitude ratios) would incur an RV variability of 20 m/s, significantly greater than the instrumental floor of HARPS-N, CARMENES, and HIRES.

An important clarification is that this conclusion is consistent with the discussion in Section 2.4. Although there is a clear periodic 2.91 d signal visible in Figure 2.7, there is also $\sim$200 m s^{-1} of scatter around this signal. It is possible that this scatter may contain coherent signals at other periods that are unresolvable with the current RV cadence.

Complicating this already complicated story is the fact that the dominant periodicity appears to change over time (Figure 2.10). This provides further motivation for our major recommendation, first given in Section 2.4: **V1298 Tau appears to be a multiply-periodic star with evolving periodicity. A high-cadence (several data points per night) RV campaign is necessary to construct a high-fidelity activity model.** The high cadence is necessary to resolve the close periodicities due to apparent differential rotation. Care should be taken to ensure that the periods do not evolve significantly over the observing baseline, or that this effect is sufficiently modeled.

2.5 Summary & Discussion

In this study, we have presented evidence that the preferred model of SM21 is overfitting using two ad hoc "validation" data sets: one set of contemporaneous HIRES and CARMENES data, and one set of artificially held-out HARPS-N data. The effects that we have proposed may be responsible for the non-predictiveness of the preferred SM21 model are:

- The RV datasets from different instruments are treated as uncorrelated, allowing the model more freedom. We will release a new version of `radvel` removing this assumption.
- The SM21 preferred model includes two summed quasi-periodic terms at P_{rot} and $P_{rot}/2$ in their kernel, each with its own free exponential decay parameter. This additional free parameter grants the model unnecessary flexibility.
- The SM21 model also includes parameters describing eccentric Keplerian signals, which grant even more degrees of freedom.
- We find evidence from multiple independent photometric datasets that this star has a strong differential rotation signal, indicating that a singly (quasi)-periodic activity model is insufficient. This explains why more complex models were favored over simpler models in SM21, even though the preferred model fell victim to overfitting.

The first point, in particular, warrants further scrutiny for stars across a range of ages and spectral types. We argued in Section 2.4 that RV datasets taken by instruments with different bandpasses and calculated using different RV extraction techniques should be linear combinations of each other, recapitulating the observation made

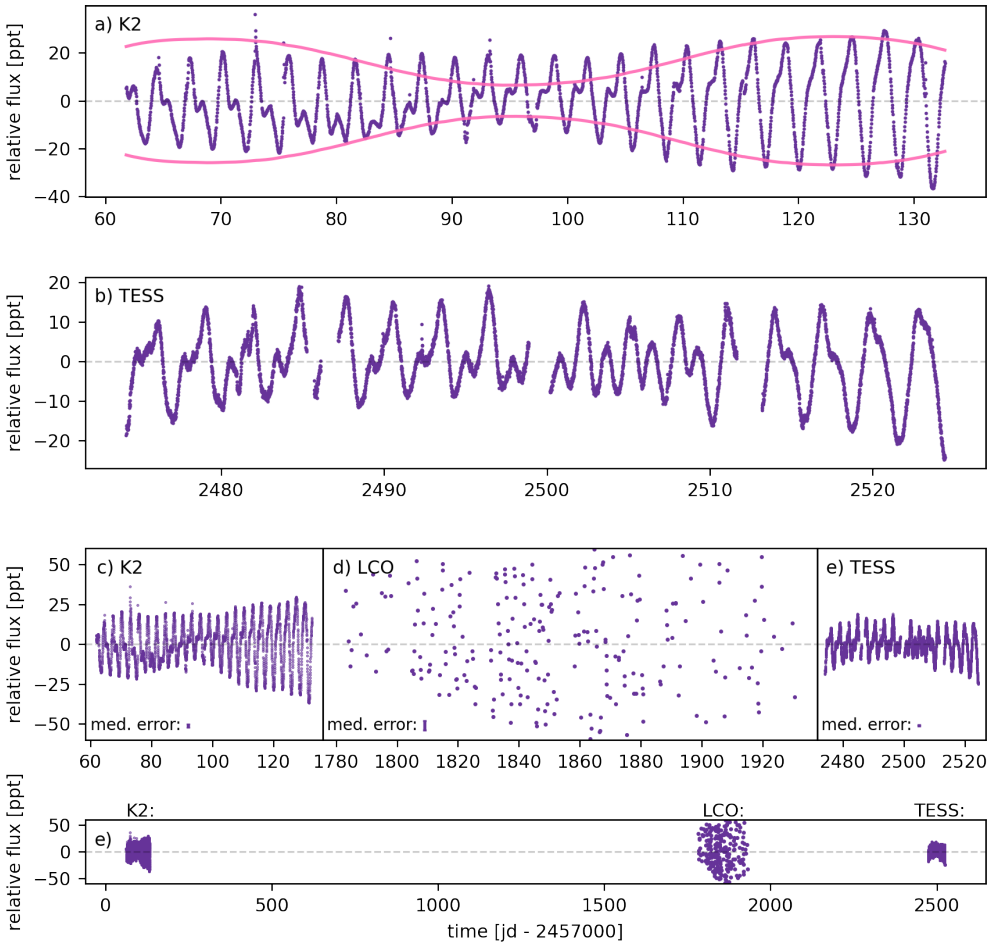

Figure 2.9: A tour of the relevant photometry of the star V1298 Tau. **Panel a:** detailed view of the K2 photometry (purple points), with a beating envelope overplotted in solid pink. The beating envelope is drawn to illustrate the effect of spot beating on overall variability amplitude, not to precisely fit the data. The envelope drawn is constructed from the beating of three sinusoids at 2.70, 2.85, and 3.00 d. Signatures of beating can be seen by eye: two peaks of different amplitudes phase up toward the end of the K2 baseline, producing a single-peaked variability pattern and a larger overall variability amplitude. **Panel b:** detailed view of the TESS photometry (purple points). Beating characteristics are also visible, although the baseline is shorter than that of K2. **Panels c, d, and e:** relative views of K2, LCO, and TESS photometry, emphasizing relative time baseline and variability amplitude. A typical error bar for each dataset is also shown in the bottom left corner of each panel. The differences in wavelength coverage and flux dilution between the K2, LCO, and TESS photometry largely account for the overall differences in amplitude of the signals. Both the K2 and TESS data cover less than one complete beat period of the two largest-amplitude periodic signals, but the LCO photometry (which is contemporaneous with the RVs of SM21) covers a longer time baseline. **Panel e:** All photometry, plotted on the same panel to emphasize relative time elapsed between each dataset. **Takeaway:** differential rotation effects are visible by eye in both the K2 and TESS datasets.

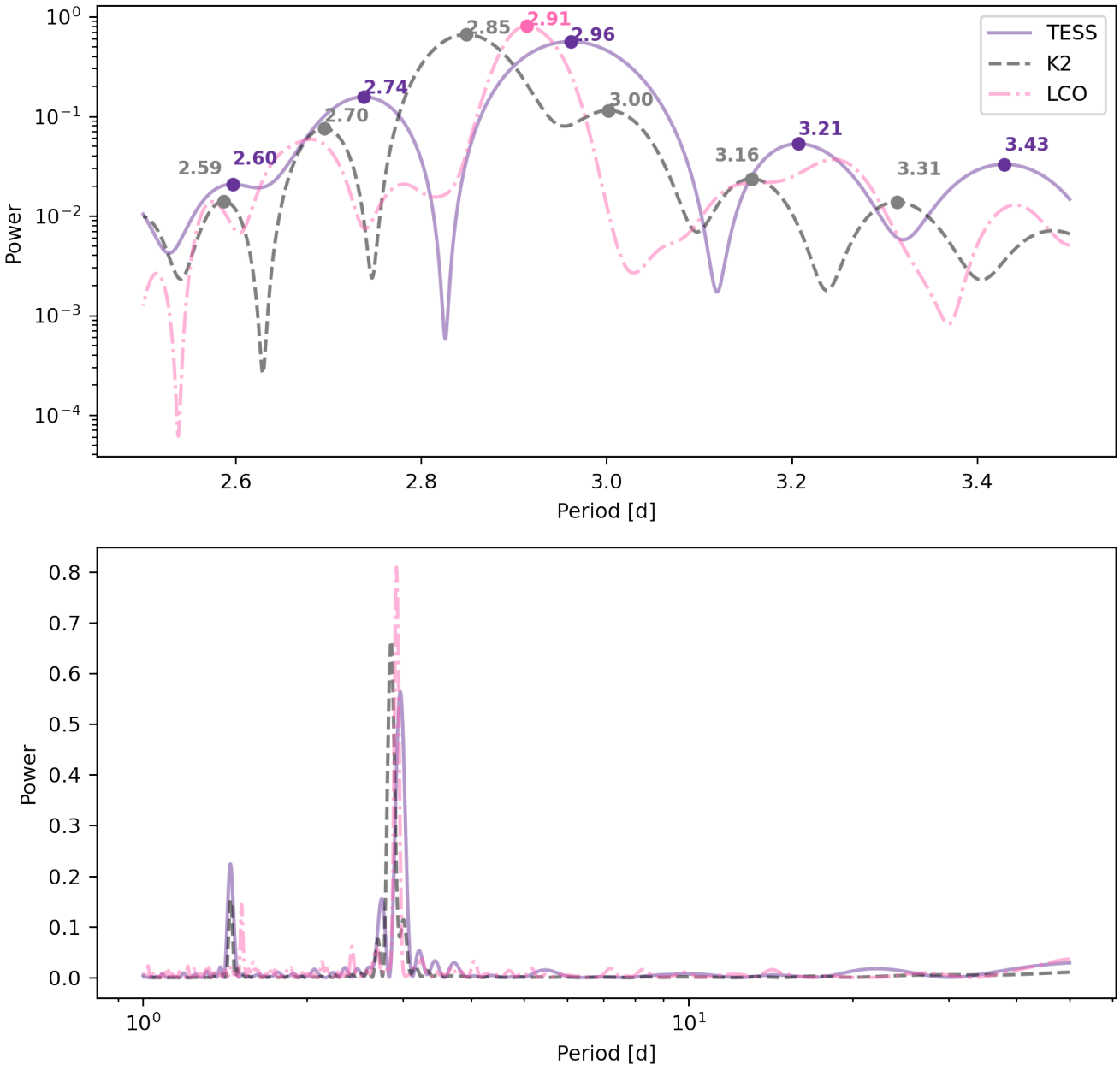

Figure 2.10: Lomb-Scargle periodograms of the photometric data shown in Figure 2.9. **Top:** Zoom-in on the presumed rotation period, showing several nearby peaks in all three datasets. **Bottom:** Same as top over a wider period range. **Takeaway:** multiple closely-related periodicities are visible in Fourier space for all three photometric datasets, more evidence for differentially rotating active regions.

in Cale et al. (2021), but this assumption may not be true. Contemporaneous RV datasets made by different instruments will help test this assumption.

These authors have devoted significant person- and computer-power to producing a fit to the data presented here that take into account all of these effects. However, we have found that jointly fitting all the data using only a single rotation period forces all of the instrumental GP amplitudes to 0. We interpret this as evidence that a singly (quasi)-periodic GP model is incapable of fitting the data (i.e., a more complex model is needed), and differential rotation provides a ready (but not sole) explanation. However, the differential rotation effects are very complicated to disentangle with the current dataset.[10] Again, we suggest a high-cadence RV campaign to resolve the multiple, nearby periodicities in the RVs and construct a high-fidelity model.

One important detail to note is that the GP kernel which best-fits a highly active, rapid-rotator like V1298 Tau may be wholly inappropriate to fit the activity signal of an older, quieter, Sun-like star. In young, rapid-rotators, the activity signal is relatively long-lived, often stable across several observation epochs (e.g., Yu et al., 2019; Carvalho et al., 2021).

On the other hand, Sun-like stars have much shorter-lived spots, sometimes evolving over the course of one or two week observing campaigns (Giles, Collier Cameron, and Haywood, 2017; Namekata et al., 2019). A GP kernel describing the activity of Sun-like stars should be more flexible, allowing for more rapidly changing and decaying signals. While a single kernel may be capable of spanning these regimes of period evolution, the attempt to construct one should be made with caution. For the time being, the best approach may be to treat the two regimes of activity with unique kernels.

This analysis is imperfect and incomplete. Many of the effects we have discussed are subtle, and we encourage others to study them further. This analysis has also evolved (quite a lot) over the preparation of this study.

There are many exciting follow-up avenues for the V1298 Tau system. First, an independent determination of the planet masses with TTVs would be enormously helpful in providing a "check" for RV modelers. Second, we believe it is worthwhile to explore modeling frameworks for V1298 Tau that explicitly model the relationship between contemporaneous photometry, activity indices, and multiple RV datasets.

[10] Although we highly encourage others to try!

These frameworks (such as that of Rajpaul et al., 2015 and Cale et al., 2021) move beyond sharing hyperparameters between contemporaneous photometric and RV datasets and allow a function of one dataset to be directly correlated with the other, decreasing the overfitting potential. In the longer term, comparing or jointly modeling these data with Doppler tomographic information and spectrum-level measurements, as in Yu et al. (2019), Finociety et al. (2021), and Klein et al. (2022) will provide even stronger constraints.

In addition to working toward an optimal physical model of all available data, it is worth investigating alternative statistical modeling pathways to GPR, especially low computational cost techniques like autoregressive moving average (ARMA) models (Feigelson, Babu, and Caceres, 2018, Durbin and Koopman, 2001). ARMA models treat the *i*th datapoint as a linear combination of past data points and model residuals, and "training" involves optimizing the linear coefficients. Directly comparing models constructed with ARMA and GPR would be a worthwhile exercise in general for datasets containing stellar activity, and in particular for young, active stars.

We believe that understanding the RV variability of young stars is an endeavor that will pay dividends in the near future. The relative long-term stability of activity on young stars allows for detailed study of a given spot geometry and its impact on both photometric and spectroscopic observations across multiple bands. As we work to understand how to best fit activity with GPs, young stars, particularly WTTSs, provide good laboratories on which to test our techniques.

Just as we validate the performance of a new instrument on stars with large, well-studied Keplerian signals, we must, as a field, validate the performance of our activity-modeling techniques on stars with large, well-studied activity signals before we can trust activity-models applied to Sun-like stars at 30 $\mathrm{cm\,s^{-1}}$ precision[11]. This starts by allocating resources to the construction of high-cadence RV datasets of young stars, and continues by studying a) the relationship between RVs and auxiliary data, such as photometry and activity indices, b) the best phenomenological models (kernels, etc) for the data, c) the best methods for validating a given model's accuracy, and d) the cadence needed to resolve periodic signals (and combinations of signals). We believe that these studies, on young stars, will pave the way for stellar activity models with 30 $\mathrm{cm\,s^{-1}}$ predictive capability, on which the characterization of Earth 2.0 depends.

[11] In fact, the activity-to-Keplerian ratio of 1000 $\mathrm{m\,s^{-1}}$: 50 $\mathrm{m\,s^{-1}}$ for warm giant planets around a young star like V1298 Tau is reminiscent of the 1 $\mathrm{m\,s^{-1}}$: 10 $\mathrm{cm\,s^{-1}}$ ratio for an Earth around a Sun-like star.

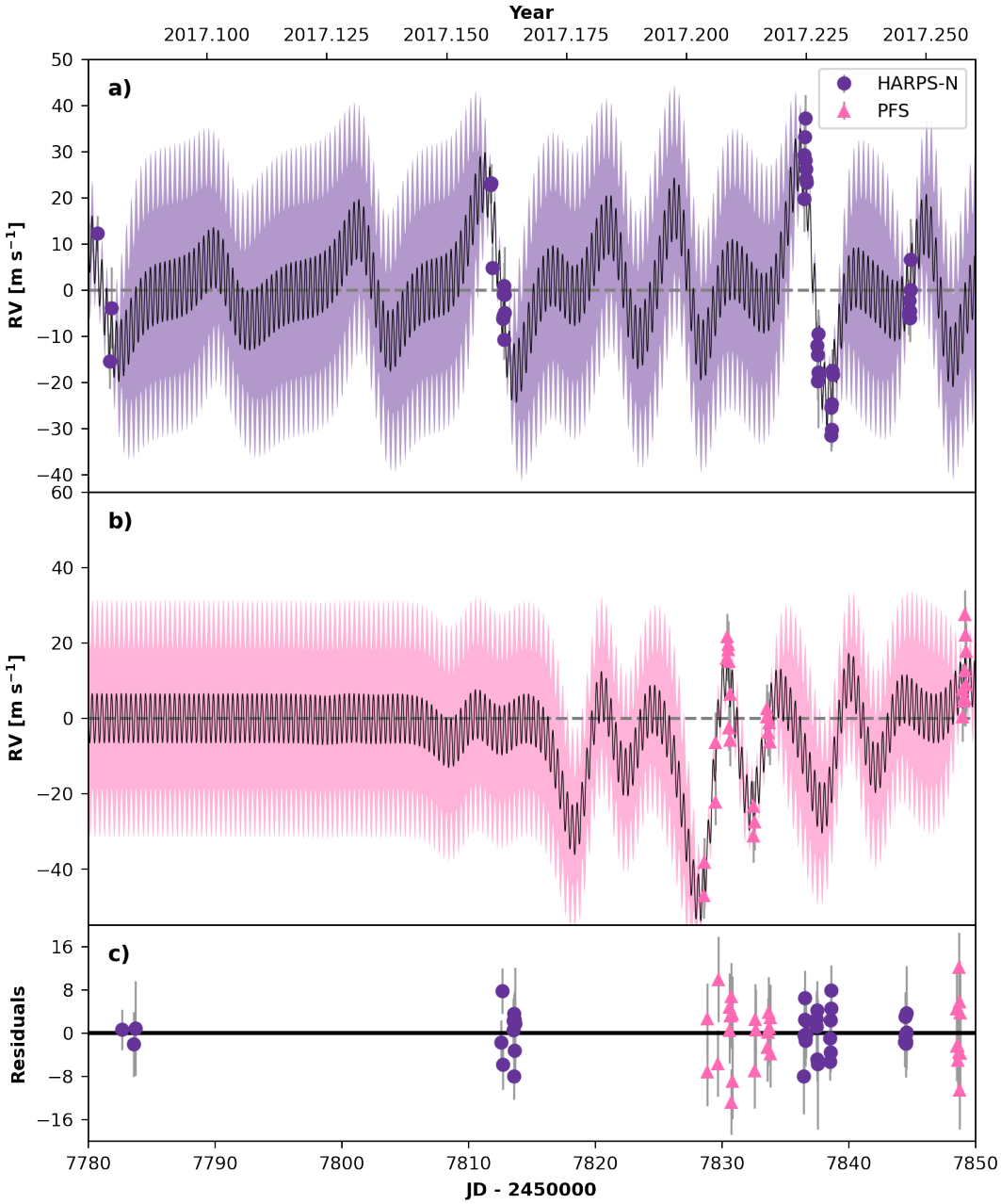

Figure 2.11: Demonstration of the impact of constructing separate covariance matrices and adding the log(likelihoods). Compare with Figure 2.12. The data and best-fit parameters are for K2-131, published in Dai et al. (2017), for demonstration purposes only. **Top:** GP mean prediction (black solid line) and 1-σ uncertainties (purple filled), together with the HARPS-N data points on which the GP is conditioned (purple points). **Middle:** Same as top, but for PFS data. **Bottom:** Residuals with respect to the GP mean prediction. **Takeaway:** When separate covariance matrices for each RV instrument are used, contemporaneous data are uncorrelated in the model, allowing additional degrees of freedom.

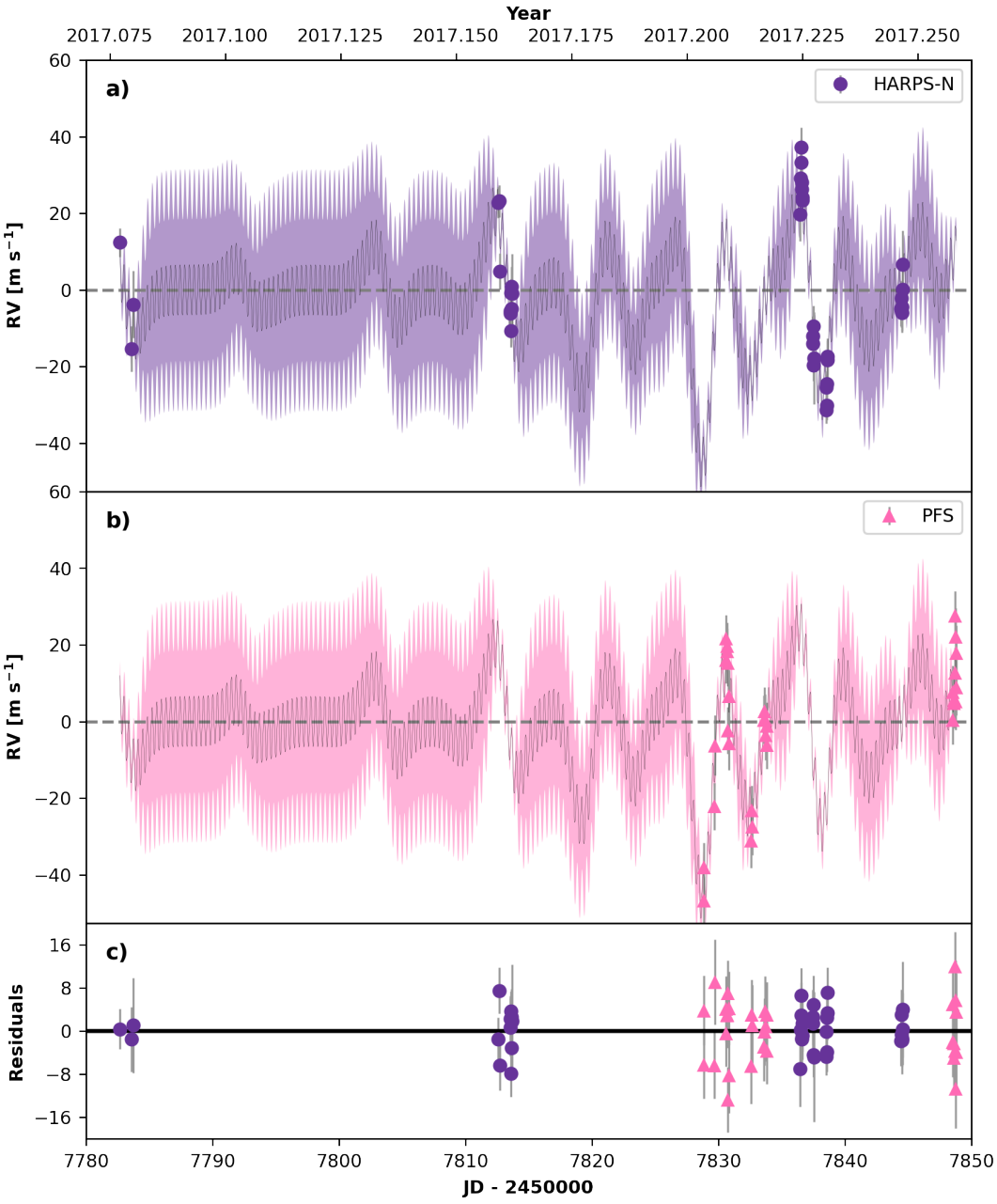

Figure 2.12: Same as Figure 2.11 (in particular, using the exact same data and GP hyperparameters), but here a single covariance matrix is constructed, following the suggestion in Section 2.4. **Takeaway:** Constructing a single covariance matrix requires that GP predictions for separate instruments are scalar multiples of one another, which is more consistent with physical expectations and results in a more constrained model than one with a separate covariance matrix for each instrument.

2.6 Appendix: Gaussian Processes and Occam's Razor

Many introductions to GPR (e.g., Aigrain and Foreman-Mackey, 2022) mention that the GP likelihood has an "Occam's razor" term built in that penalizes complexity. This section briefly reviews GPR, then outlines a geometrical interpretation of the complexity penalty in order to further readers' understanding.

A Gaussian process regression model parameterizes the covariance between data points using a kernel function. A statistician may pick an arbitrary function (subject to certain mathematical requirements, see Rasmussen and Williams, 2006 for the gory details) to be the kernel, which can then be used to calculate the covariance between any two data points. As an example, let's consider the periodic kernel:

$$C_{ij} = \eta_1^2 \exp\left[-\frac{\sin^2(\frac{\pi|t_i - t_j|}{P_{\rm rot}})}{\eta_3^2}\right], \tag{2.3}$$

where η_1 is the amplitude, $P_{\rm rot}$ is the variability period (often the star's rotation period), and η_3 is the harmonic complexity, or degree of "wiggly-ness" of the repeating signal. Given this model for the covariance of our data, and some data, we can make a prediction, which is the conditional probability distribution over expected values at new measurement times. This is referred to as *conditioning* a GP on a set of data.

Importantly, Gaussian process regression does not inherently involve training (i.e., parameter tuning, generally via an optimization and/or MCMC step). Gaussian process regression is just the process of using a parametrization of your covariance matrix to predict the values and uncertainties of new data points given existing data points.

The "training" part comes in when you are optimizing the hyperparameters of your kernel (optionally jointly with parameters of a mean function, which could be a function of Keplerian orbital parameters). Now, it becomes important to compute a statistic describing how well your GP model fits your data, so that you can optimize the (hyper)parameters to obtain your result. This is where the Gaussian process likelihood comes in:

$$\log p(d|m) = -\frac{N}{2}\log 2\pi - \frac{1}{2}\mathbf{r}^T\mathbf{C}^{-1}\mathbf{r} - \frac{1}{2}\log|\mathbf{C}|, \tag{2.4}$$

where $\mathbf{C}$ is the covariance matrix computed for the times at which you have data, N is the number of measurements, and $\mathbf{r}$ is the vector of residuals (data - mean

model). The first term is a constant, and does not change as a function of the kernel hyperparameters, and the second term is analogous to χ^2 (in fact, it reduces to χ^2 in the limit of no off-diagonal covariance). The second term describes how well your mean model and correlated noise description matches your data. The third term is the "Occam's razor" term that penalizes complexity.

To understand how the third term penalizes complexity, recall that the determinant of a matrix can be understood as the hypervolume between vectors defined by the columns of the matrix. To make this concrete, consider the 3x3 identity matrix:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{2.5}$$

The vectors defined by the columns of this matrix are (1,0,0), (0,1,0), and (0,0,1). The volume of the 3D shape defined by these vectors (the unit cube) is 1, the same as the matrix determinant!

The i-th column vector of a covariance matrix can be interpreted as the vector of covariances between a data point taken at t_i and every other data point in the dataset. The determinant of this matrix, then, is the hypervolume defined by these covariance vectors. A perfectly covariant matrix, in which all data points are perfectly correlated, will consist of all 1s[12], and the covariance vectors will all "point" in the same direction. This results in a third-term contribution of:

$$\begin{aligned} -\frac{1}{2}\log|\mathbf{C}| &= -\frac{1}{2}\log 0 \\ &= -(-\infty) \\ &= \infty. \end{aligned} \tag{2.6}$$

A matrix of perfectly *independent* data points, on the other hand, is (a scalar multiple of) the identity matrix. The covariance vectors all "point" in orthogonal directions. This matrix results in a third-term contribution of:

$$\begin{aligned} -\frac{1}{2}\log|\mathbf{C}| &= -\frac{1}{2}\log 1 \\ &= 0. \end{aligned} \tag{2.7}$$

[12] Or a scalar multiple of the matrix of all 1s.

This exercise demonstrates that the determinant of the covariance matrix quantifies how "clustered" the covariance vectors corresponding to each data point are in hyperspace. More clustered covariance vectors get a big likelihood boost, while less clustered/more independent covariance vectors get a smaller boost. Figure 5.3 in Rasmussen and Williams, 2006 decomposes the likelihood contributions of the second and third terms in Equation 2.4, illustrating how they combine to produce a local likelihood maximum in parameter space for a toy model.

2.7 Acknowledgements

S.B. wishes to thank first and foremost Alejandro Suárez-Mascareño for constructive and helpful thoughts throughout the preparation of this study. S.B. also wishes to thank the small army of people who shaped this analysis through conversation: Jason Wang and his research group, Heather Knutson's research group, the folks at the Flatiron CCA, the University of Michigan Exoplanet Journal Club, the UC Riverside Astrobiology Seminar group, Johanna Teske and the astronomers of the Carnegie Earth & Planets Laboratory, Jéa Adams, Kim Paragas, Shreyas Vissapragada, Ward Howard, and Roberto Tejada Arevalo.

J.M.A.M. is supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1842400. J.M.A.M. acknowledges the LSSTC Data Science Fellowship Program, which is funded by LSSTC, NSF Cybertraining Grant No. 1829740, the Brinson Foundation, and the Moore Foundation; his participation in the program has benefited this work. T.H. is supported by JSPS KAKENHI Grant Numbers JP19K14783 and JP21H00035.

This research was enabled by the following software: `numpy` (Harris et al., 2020), `Lightkurve`, a Python package for Kepler and TESS data analysis (Lightkurve Collaboration et al., 2018), `pandas` (McKinney, 2010), `matplotlib` (Hunter, 2007), `scipy` (Virtanen et al., 2020), `astropy` (Astropy Collaboration et al., 2013, Astropy Collaboration et al., 2018, Astropy Collaboration et al., 2022), `jax` (Bradbury et al., 2018), `george` (Ambikasaran et al., 2014), `celerite` (Foreman-Mackey et al., 2017), `tinygp` (github.com/dfm/tinygp), and `radvel` (Fulton et al., 2018).

S.B. wishes to acknowledge her status as a settler on the ancestral lands of the Gabrielino/Tongva people, and to recognize that the astronomical observations described in this paper were only possible because of the dispossession of Maunakea from Kanāka Maoli. We seek to work toward a scientific practice guided by pono and a future in which we all honor the land.

Chapter 3

FIRST VLTI/GRAVITY OBSERVATIONS OF HIP 65426 B: EVIDENCE FOR A LOW OR MODERATE ORBITAL ECCENTRICITY

Submitted to the Astronomical Journal

S. Blunt, W.-O. Balmer, J. J. Wang, S. Lacour, G. Bourdarot, J. Kammerer, N. Pourré, E. Rickman, J. Shangguan, T. Winterhalder, R. Abuter, A. Amorim, R. Asensio-Torres, M. Benisty, J.-P. Berger, H. Beust, A. Boccaletti, A. Bohn, M. Bonnefoy, H. Bonnet, W. Brandner, F. Cantalloube, P. Caselli, B. Charnay, G. Chauvin, A. Chavez, E. Choquet, V. Christiaens, Y. Clénet, V. Coudé du Foresto, A. Cridland, R. Dembet, A. Drescher, G. Duvert, A. Eckart, F. Eisenhauer, H. Feuchtgruber, P. Garcia, R. Garcia Lopez, E. Gendron, R. Genzel, S. Gillessen, J. H. Girard, X. Haubois, G. Heißel, Th. Henning, S. Hinkley, S. Hippler, M. Horrobin, M. Houllé, Z. Hubert, L. Jocou, M. Keppler, P. Kervella, L. Kreidberg, A.-M. Lagrange, V. Lapeyrère, J.-B. Le Bouquin, P. Léna, D. Lutz, A.-L. Maire, F. Mang, G.-D. Marleau, A. Mérand, P. Mollière, J. D. Monnier, C. Mordasini, D. Mouillet, E. Nasedkin, M. Nowak, T. Ott, G. P. P. L. Otten, C. Paladini, T. Paumard, K. Perraut, G. Perrin, O. Pfuhl, L. Pueyo, J. Rameau, L. Rodet, Z. Rustamkulov, T. Shimizu, D. Sing, T. Stolker, C. Straubmeier, E. Sturm, L. J. Tacconi, E.F. van Dishoeck, A. Vigan, F. Vincent, K. Ward-Duong, F. Widmann, E. Wieprecht, E. Wiezorrek, J. Woillez, S. Yazici, and A. Young

3.1 Abstract

Giant exoplanets have been directly imaged over orders of magnitude of orbital separations, prompting theoretical and observational investigations of their formation pathways. In this paper, we present new VLTI/GRAVITY astrometric data of HIP 65426 b, a cold, giant exoplanet at a projected separation of 92 au from its primary. Leveraging GRAVITY's astrometric precision, we present an updated eccentricity posterior that disfavors large eccentricities. The eccentricity posterior is still prior-dependent, and we extensively interpret and discuss the limits of the posterior constraints presented here. We also perform updated spectral comparisons with self-consistent forward-modeled spectra, finding a best fit `ExoREM` model with solar metallicity and C/O=0.6. An important caveat is that it is difficult to estimate

robust errors on these values, which are subject to interpolation errors as well as potentially missing model physics. Taken together, the orbital and atmospheric constraints paint a preliminary picture of formation inconsistent with scattering after disk dispersal. Further work is needed to validate this interpretation.

3.2 Introduction

The existence of cold Jupiters (CJs) at large separations (10s of au) is a formation puzzle. Direct imaging surveys have revealed that these planets are intrinsically rare, with an occurrence rate of about 1% (Bowler and Nielsen, 2018), but given that the nominal core formation timescale at typical CJ separations is much longer than the disk dispersal timescale (Armitage, 2020), we would not expect CJs to exist at all, assuming in-situ core accretion. Direct gravitational collapse within a disk is an alternative to core accretion, but with its own issues, particularly very fast migration after formation (Nayakshin, 2017, Vorobyov and Elbakyan, 2018). The CJ population is quite diverse, with masses spanning an order of magnitude and separations spanning several (Bowler, 2016), so it is also possible that multiple formation mechanisms are at play.

The eccentricities of CJs are a useful tracer of formation history, both at the population level (Bowler, Blunt, and Nielsen, 2020a), and for individual systems. Marleau et al. (2019) conducted a detailed investigation of possible formation scenarios via core accretion for the ~14 Myr CJ HIP 65426 b (Table 3.1). At a projected separation of 92 au (Chauvin et al., 2017), this object is significantly farther from its star than, for example, the famous HR 8799 planets (at 16, 26, 41, and 72 au, Wang et al., 2018), and therefore even more of a challenge for formation via core accretion. Marleau et al. (2019) derived distributions over the initial mass and luminosity of HIP 65426 b under several assumptions of post-formation entropy, then coupled these initial conditions with N-body simulations to investigate which models could match the present-day conditions of the planet. They varied the initial number of planets in the system, and included prescriptions for type I and II migration in a protoplanetary disk. Through a suite of such simulations, they found two families of explanations for HIP 65426 b's current separation and luminosity. The first is core formation at close separations, followed by outward scattering and subsequent runaway gas accretion at the present-day location. The slower timescale of type II migration and subsequent disk dispersal allows the planet to remain in place after scattering and damps the post-scattering eccentricity. The second scenario is similar, except runaway accretion and subsequent disk dispersal occur *before* scattering. Under this

scenario, most simulations resulted in a high (>0.5) present-day eccentricity for HIP 65426 b. A third scenario the authors mentioned but did not investigate in detail is the prospect of in-situ formation, invoking a more rapid core formation process than typically assumed, such as pebble accretion (Lambrechts and Johansen, 2014, Rosenthal and Murray-Clay, 2018). This scenario would presumably result in a circular orbit.

In summary, Marleau et al. (2019) propose three statistically distinct formation pathways via core accretion for HIP 65426 b: 1) in-situ formation, resulting in a circular orbit, 2) scattering before disk dispersal, resulting in a low-to-moderate eccentricity orbit, and 3) scattering after disk dispersal, resulting in a high eccentricity orbit. Both scenarios 2 and 3 are often accompanied by an inner giant planet. An important caveat is that these are not hard-and-fast distinctions; an eccentricity of 0, even with no model uncertainty, would not unequivocally rule out two of the three scenarios. However, such a measurement will allow us to assign statistical probabilities to each formation scenario by comparing with the population synthesis outputs of studies like Marleau et al. (2019). Precise measurements of eccentricity will also better constrain the population-level eccentricity distribution of CJs, which is currently still prior-dependent (Nagpal et al., 2023), allowing us to model the formation mechanisms responsible for the population as a whole.

HIP 65426 b has been astrometrically monitored since its discovery by Chauvin et al. (2017), and previous papers report an essentially unconstrained eccentricity posterior that reproduces the prior (Chauvin et al., 2017, Cheetham et al., 2019, Carter et al., 2022). In this paper, we update the orbit model of HIP 65426 b using new astrometry from the optical interferometer VLTI/GRAVITY (~50x more precise than previous astrometry) as part of the ExoGRAVITY project (Lacour et al., 2020).

Spectral and photometric measurements of HIP 65426 b have also been used to infer the planet's atmospheric properties. Petrus et al. (2021) recovered a bimodal posterior by comparing the HIP 65426 b measurements with `BT-Settl CIFIST` models, with one small radius ($\sim 1 R_J$) peak and one larger radius peak ($\sim 1.2 R_J$). Their `Exo-REM` comparisons yielded broader posteriors encompassing both of these possibilities, as `Exo-REM` predictions were only available for the K-band at the time, and they could not compare with all available data. These yielded T_{eff}=$1560 \pm 100 K$, $\log g < 4.40$, a slightly super-solar metallicity of $0.05^{+0.24}_{-0.22}$, and an upper limit on the C/O ratio (≤ 0.55). Carter et al. (2022) subsequently published 7 long-wavelength

Table 3.1: Relevant physical properties of HIP 65426 A and b

quantity	value	source
Stellar spectral type	A2V	Carter et al. (2022)
Stellar mass	1.96 ± 0.04	Chauvin et al., 2017
Moving group membership	Lower Centaurus-Crux, 14 ± 4 Myr	Gagné et al. (2018)
Stellar parallax	9.30 ± 0.03 mas	Gaia DR3 (Gaia Collaboration et al., 2016, Gaia Collaboration et al., 2022)
Hot-start planetary mass estimate	7.1 ± 1.2 M_J	Carter et al. (2022)
Hot-start planetary radius estimate	1.44 ± 0.03 R_J	Carter et al. (2022)
Hot-start planetary T_{eff}	1283^{+25}_{-31} K	Carter et al. (2022)

Table 3.2: Observing log. NEXP, NDIT, and DIT denote the number of exposures, the number of detector integrations per exposure, and the detector integration time, respectively. τ_0 is the atmospheric coherence time during each exposure. The fiber pointing is the placement of the science fiber relative to the fringe tracking fiber (which is placed on the central star). HD 91881 and HD 73900 are two binary systems used for phase referencing.

Date	UT time		Target	NEXP/NDIT/DIT	Airmass	τ_0	Seeing	Fiber pointing
	Start	End						ΔRA/ΔDEC [mas]
2021-01-07	06:22:05	06:48:54	HD 91881	8 / 64 / 1 s	1.05–1.10	6.6–8.0 ms	0.51–0.65″	-1108,710
2021-01-07	06:59:10	07:57:00	HIP 65426 b	4 / 8 / 100 s	1.38–1.64	5.0–6.6 ms	0.61–1.19″	416,-704
2021-01-07	08:03:17	08:06:18	HIP 65426 A	2 / 64 / 1 s	1.35–1.36	4.6–5.4 ms	1.05–1.23″	0,0
2022-01-23	05:16:56	05:30:01	HD 73900	6 / 64 / 1 s	1.05–1.10	13.2–20.4 ms	0.41–0.57″	-825,-455
2022-01-23	05:42:05	05:55:37	HD 91881	6 / 64 / 1 s	1.38–1.64	6.4–13.8 ms	0.30–0.67″	-1108,710
2022-01-23	06:08:31	06:51:52	HIP 65426 b	6 / 4 / 100 s	1.35–1.36	5.4–13.9 ms	0.54–0.72″	418,-699

photometric datapoints of HIP 65426 b using the NIRCAM and MIRI instruments, and used them to update the empirical bolometric luminosity and of the planet to $\log L/L_\odot$ = (-4.31, -4.14). Together with hot-start cooling models, they used this luminosity estimate to infer a planetary mass, radius, and effective temperature (given in Table 3.1). They found that an independent comparison to the `BT-Settl CIFIST` grid yielded a good fit, but resulted in an unphysically small planetary radius of 1.06 ± 0.05 R_J (consistent with the smaller radius mode of Petrus et al., 2021).

This paper is structured as follows: in Section 2.2, we present our new GRAVITY data, including two new astrometric epochs and a new K-band spectrum. In Section 3.4, we present our updated orbital solution and discuss the significance of our eccentricity measurement in detail. In Section 3.5, we compare all existing photometry and spectra measurements of HIP 65426 with self-consistent model spectra to update our understanding of the planet's atmospheric properties, finding results that are consistent with but more precise than previous work, keeping in mind that systematic uncertainties in these fits are likely underestimated. Finally, in Section 2.5 we discuss the implications of our orbital and atmospheric inferences and call for additional observational and theoretical work.

3.3 Data

We observed HIP 65426 b on the 7th of January, 2021 and the 23rd of January, 2022 as part of the ExoGRAVITY Large Program (ESO Program ID 1104.C-0651, Lacour et al., 2020). We used the European Southern Observatory (ESO) Very Large Telescope Interferometer (VLTI)'s four 8.2m Unit Telescopes (UTs) and the GRAVITY instrument (Gravity Collaboration et al., 2017). The observations primarily used the "off-axis dual field" mode, in which the roof mirror is used to split the telescope field into two. The star light goes to the fringe tracker to correct for atmospheric perturbations (Lacour et al., 2019a). The exoplanetary light goes to the spectrograph configured with the medium resolution grism (R=500). Phase referencing of the metrology is obtained by swapping on a binary just before the observation. On 2021-01-07 we used the binary system HD 91881, and on 2022-01-23 we used two binary systems: HD 73900 and HD 91881. During the night of the 2021-01-07, we also performed a sequence of "on-axis single-field" observations of the host star to calibrate the spectrum of the planet for that night. The second night of observations does not have an on-axis calibrator, and so we do not consider the spectrum of the planet from those observations. The observing log, presented in Table 3.2, records the length of the observations, the number of files recorded and the atmospheric conditions. The placement of the science fiber was based on preliminary orbit predictions fit to the available relative astrometry at the time using the `whereistheplanet`[1] software (Wang, Kulikauskas, and Blunt, 2021), and resulted in an efficient coupling of the planet flux into the science fiber.

We calculated the complex visibilites of the host and the companion, which were phase-referenced with the metrology system using observations of the binary calibrator HD 91881, using the Public Release 1.5.0 (1 July 2021[2]) of the ESO GRAVITY pipeline (Lapeyrere et al., 2014). We decontaminated the flux of the planet due to the host using a custom python pipeline (see Appendix A, Gravity Collaboration et al., 2020), which treats the contamination as a polynomial dependent on time and baseline: a polynomial of fourth order was used for stellar light suppression.

We obtained the astrometric position of the planet relative to the star at each epoch by analysing the phase of the ratio of coherent fluxes, computing a periodogram power map over the fiber's field-of-view (Figure 3.1). The mean astrometric position is taken to be the minimum χ^2 value of this map. We estimated the uncertainty on

[1] `http://www.whereistheplanet.com`

[2] `https://www.eso.org/sci/software/pipelines/gravity/`

Table 3.3: New relative astrometry of HIP 65426 b presented in this paper. $\sigma_{\Delta\text{R.A.}}$ and $\sigma_{\Delta\text{Dec.}}$ denote the uncertainties in astrometric position, and $\rho_{\Delta\text{R.A.},\Delta\text{Decl.}}$ denotes the correlation between the $\sigma_{\Delta\text{R.A.}}$ and $\sigma_{\Delta\text{Dec.}}$ measurements.

Date [jd − 2400000.5]	ΔR.A. [mas]	ΔDecl. [mas]	$\sigma_{\Delta\text{R.A.}}$ [mas]	$\sigma_{\Delta\text{Decl.}}$ [mas]	$\rho_{\Delta\text{R.A.},\Delta\text{Decl.}}$
59221.312	415.613	0.107	-708.133	0.073	-0.095
59602.271	416.269	0.035	-705.051	0.035	-0.224

Table 3.4: Astrometric measurements from the literature used in the orbit fits presented in this paper. Here, RV_{pl} indicates a measurement of the planet's radial velocity *relative to the primary*. Making this measurement involved separately measuring the absolute radial velocity of the planet and the absolute radial velocity of the star, subtracting these quantities, and propagating the uncertainty.

Date [jd − 2400000.5]	separation (ρ) [mas]	σ_ρ [deg]	position angle (P.A.) [mas]	$\sigma_{\text{P.A.}}$ [deg]	RV_{pl} [km s^{-1}]	$\sigma_{\text{RV}_{\text{pl}}}$ [km s^{-1}]	reference
57538.4	830.4	4.9	150.28	0.22	–	–	Chauvin et al. (2017)
57565.5	830.1	3.2	150.14	0.17	–	–	Chauvin et al. (2017)
57791.0	827.6	1.5	150.11	0.15	–	–	Chauvin et al. (2017)
57793.1	828.8	1.5	150.05	0.16	–	–	Chauvin et al. (2017)
57891.0	832	3	149.52	0.19	–	–	Cheetham et al. (2019)
57892.0	850	20	148.5	1.6	–	–	Cheetham et al. (2019)
58250.0	822.9	2.0	149.85	0.15	–	–	Cheetham et al. (2019)
58250.0	826.4	2.4	149.89	0.16	–	–	Cheetham et al. (2019)
58263.5	–	–	–	–	14	15	Petrus et al. (2021)

each astrometric measurement using the scatter of mean astrometric values between individual exposures. These new astrometric datapoints are provided in Table 3.3.

We then extracted the ratio of the coherent flux between the two sources at the location of the companion, generating a contrast spectrum from 2 to 2.5 microns. We converted this contrast spectrum into a flux calibrated spectrum by multiplying it by a BT-NextGen (Allard, Homeier, and Freytag, 2012) spectrum fit to archive photometry of the host. We noted a tension between the absolute flux of the resultant spectrum and published SPHERE K-band photometry (Chauvin et al., 2017). For our 2021-01-07 observation, $F_{K2} = 1.40 \times 10^{-16}$ W/m$^2/\mu$m vs the SPHERE value $F_{K2} = 7.21 \times 10^{-17}$ W/m$^2/\mu$m. We opted to integrate the GRAVITY spectrum over the Paranal/IRDIS_D_K12_2 filter and scale the GRAVITY spectrum by $\times 0.52$ to match the SPHERE K2 photometry. The existing SPHERE photometry matches the existing SINFONI spectrum (described more and compared with our new GRAVITY spectrum in Section 3.5) without any scaling factor applied.

The fluxes, uncertainties, and inter-channel flux covariances are available as a machine-readable table published along with this paper.

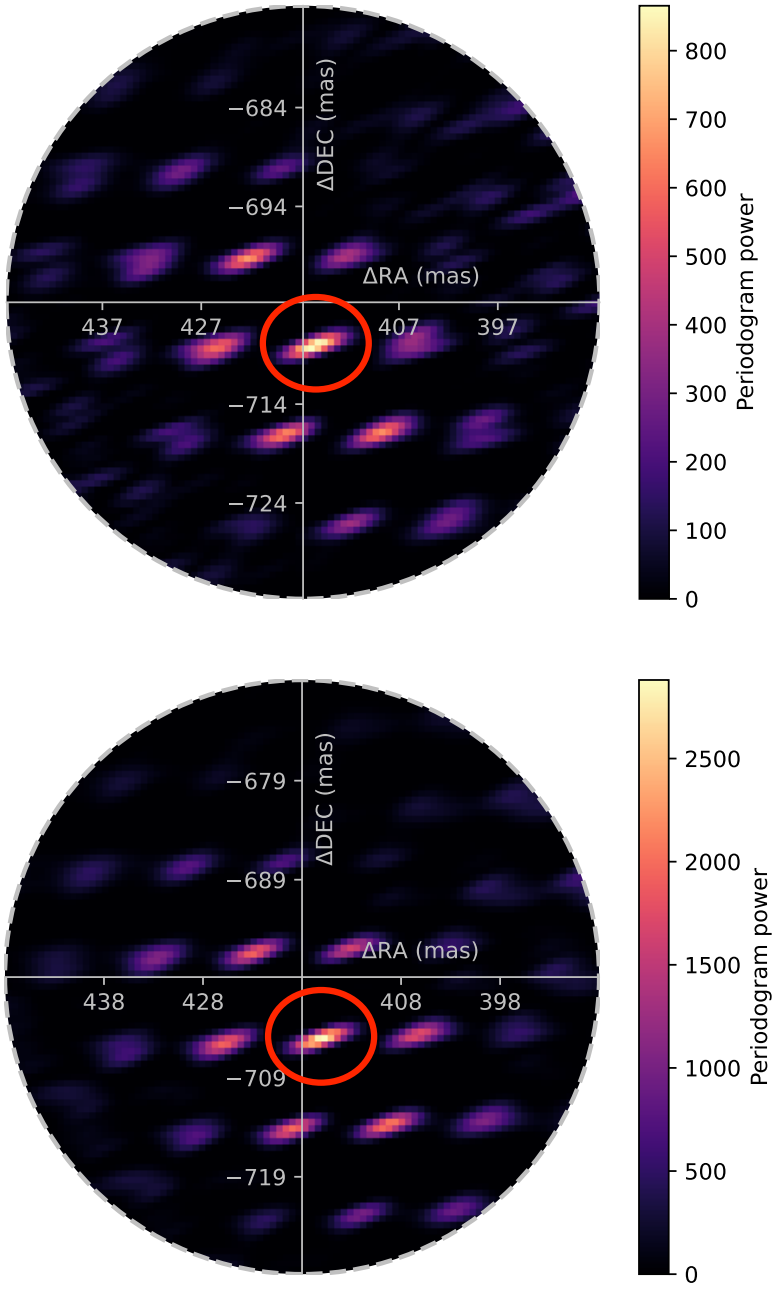

Figure 3.1: Detections of HIP 65426 b with VLTI/GRAVITY. Both periodogram power maps visualizes the χ^2 fit to the interferometric observables assuming a point source, after removing the contribution of the star using a 4th order polynomial. The outer dashed grey circle indicates the effective fiber field of view (60mas in diameter), and the red circles indicate the most probable planet position at each epoch. The planet is detected at high confidence in both epochs (periodogram power > 500).

Literature Data

Literature spectral and photometric data used in this study (all of which are plotted in Figure 3.14) come from several sources. The VLT/SPHERE IFS spectra at Y-H (0.96 – 1.64 μm) bands, four photometric VLT/SPHERE measurements at H2 (λ_0 = 1.593 μm and $\Delta\lambda$ = 0.11 μm) and H3 (λ_0 = 1.667 μm and $\Delta\lambda$ = 0.12 μm), at K1 (λ_0 = 2.1025 μm and $\Delta\lambda$ = 0.204 μm) and K2 (λ_0 = 2.2550 μm and $\Delta\lambda$ = 0.218 μm) bands, and two NACO measurements at Lp (λ_0 = 3.80 μm, $\Delta\lambda$ = 0.62 μm) and Mp (λ_0 = 4.78 μm, $\Delta\lambda$ = 0.59 μm) bands were originally published in Cheetham et al. (2019). An additional NACO point obtained using the NB405 filter was originally published in Stolker et al. (2020). The medium-resolution SINFONI spectrum at K band come from Petrus et al. (2021). All spectral and photometric data (with the exception of the SINFONI data, which were kindly provided by S. Petrus) were accessed from `species`[3](version 0.5.5).

3.4 Orbit Analysis

In order to understand how our new GRAVITY data constrains the orbit of HIP 65426 b, we performed seven different orbit-fits, varying the data subset used and the priors applied. These fits use literature data from Table 3.4[4] and new VLTI/GRAVITY data from Table 3.3. The results are summarized in Table 3.6 and described in more detail below. We used version 2.1.3 of `orbitize!` (Blunt et al., 2020) for all fits, taking advantage of the ability to fit companion RVs introduced in version 2.0.0. `orbitize!` is a Bayesian tool for computing the posteriors over orbital elements of directly-imaged exoplanets, and is intended to meet the orbit-fitting needs of the high-contrast imaging community. All orbit fits used the following orbital basis: semimajor axis (a), eccentricity (ecc), inclination (inc, with inc=0 defined to be a face-on orbit), argument of periastron (ω_p), position angle of nodes (Ω), and epoch of periastron, defined as fraction of the orbital period past a specified reference date (see Blunt et al. (2020) section 2.1 for a bit more detail). We use mjd=58859 as the reference date in this paper. Following the suggestion in Householder and Weiss (2022), we also explicitly specify here that we assume that ω_p is defined relative to the longitude of ascending node, which in turn is defined as the intersection point between the orbit track and the line of nodes when the object is moving *away from the observer*. Our coordinate system is illustrated in Figure 1 of Blunt et al., 2020. An interactive tutorial is also available to help users build an intuitive understanding

[3]https://github.com/tomasstolker/species/blob/main/species/data/companion_data.json

[4]There are additional astrometric datapoints given in Stolker et al. (2020) and Carter et al. (2022) which were excluded because of their large error bars.

of the coordinate system[5]. We used the same priors on all parameters as given in Blunt et al. (2020), unless otherwise specified.

The orbit fits presented in this paper include two types of data: relative astrometry, and companion radial velocities. The ability to fit companion radial velocities is new in `orbitize!` since the publication of Blunt et al. (2020), so we discuss it in a bit of detail here. Following standard practice (e.g. Chapter 1 of Seager, 2010), `orbitize!` by default defines orbital parameters (a, e, etc.) as those of the *relative* orbit, meaning that astrometric and radial velocity measurements of the secondary are both assumed to be relative to the primary.[6] In order to use the RV measurement of Petrus et al. (2021) in our orbit fits, we therefore subtracted the absolute measurement of the planet's RV (derived from the planetary spectrum) and an independent measurement of the star's RV derived from 78 HARPS spectra of the primary (see Section 4.2 of Petrus et al., 2021 for details), and propagated the uncertainties in both measurements.[7] Relative companion RV measurements like this do not allow us to measure a dynamical mass for the companion, but have the potential to reduce posterior uncertainties of the relative orbit. However, the uncertainty on the available companion RV measurement is too large to meaningfully constrain the orbital parameters beyond the constraints from the relative astrometry (see Figure 3.2). It does, however, reduce the 180° degeneracy between Ω and ω, which usually shows up as double-peaked posteriors on those parameters for relative-astrometry-only orbits. A more precise planetary RV would uniquely orient the planet in 3D space (to within the orbital parameter uncertainties). Figure 3.2 shows the RV predictions for each orbit in the posterior of our accepted fit, together with the measurement from Petrus et al., 2021. The current uncertainties of the companion and stellar RVs of HIP 65426 A and b limit the ability of the relative RV to constrain the orbit fit.

We performed the following orbit fits (summarized in Table 3.6). All fits included the companion RV as described in the above paragraph.

[5]https://github.com/sblunt/orbitize/blob/main/docs/tutorials/show-me-the-orbit.ipynb

[6]When radial velocity measurements are available for both the star and the planet, `orbitize!` instead assumes that planetary and stellar RV measurements are relative to the system barycenter.

[7]It is important to note that systematic offsets in the wavelength solutions used for the HARPS and SINFONI spectra (not to mention RV offsets due to astrophysical variability, such as pulsations) would affect the RV value derived here. However, the RV value derived from the SINFONI data is not sufficiently precise to impact our measurements of eccentricity and inclination, the main focus of this paper. Therefore we do not investigate any potential systematics in detail.

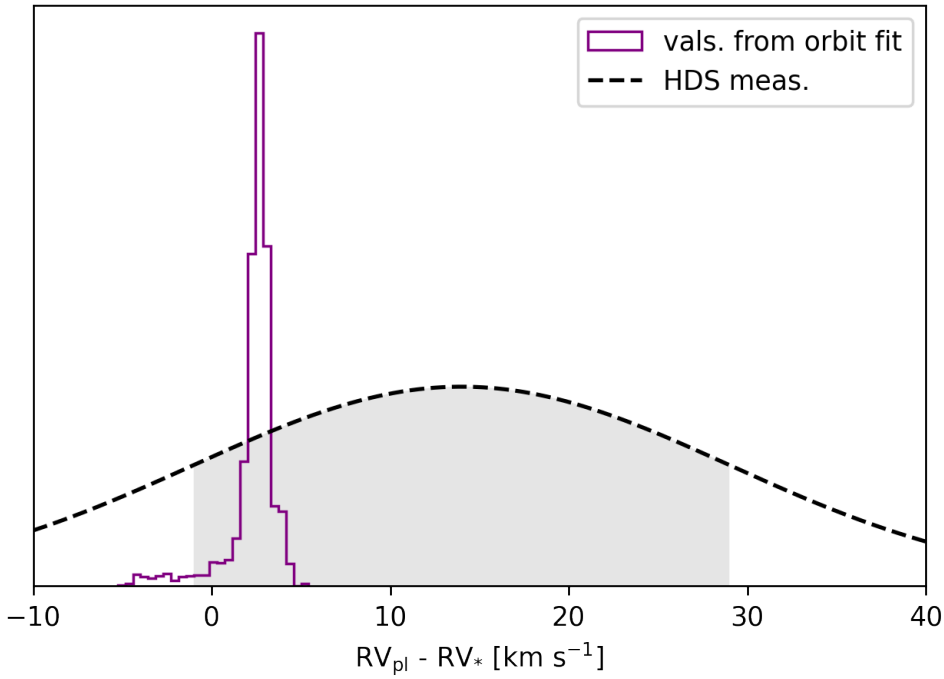

Figure 3.2: Planetary radial velocity predictions from the accepted fit (purple histogram), together with the HDS planetary RV measurement from Petrus et al. (2021) (dashed black line, with 1-σ range shaded grey). **Takeaway:** The planetary RV measurement does not constrain the orbital parameters, beyond breaking the 180° degeneracy for Ω and ω. A planetary RV of -3 km/s is also allowed, given the astrometry alone, but is disfavored because of the relative RV measurement.

1. Only including literature data from SPHERE and NACO (i.e. no GRAVITY data)
2. Literature data plus the earlier epoch of GRAVITY astrometry
3. Literature data plus the later epoch of GRAVITY astrometry
4. Literature data plus both epochs of GRAVITY astrometry
5. Both epochs of GRAVITY data alone (i.e. no literature data)
6. Fit 4, except fixing eccentricity to 0[8]
7. Fit 4, except applying a decreasing prior on eccentricity.

[8]Note that ω_p and τ_{58849} are undefined for a circular orbit.

For the final fit, we used the following prior distribution, following Nielsen et al. (2008)[9]:

$$p(e) \propto -2.18e. \tag{3.1}$$

Fits 1–5 were performed in order to assess the outlier sensitivity of our fits, as well as to understand the relative constraining power of the GRAVITY astrometry and the less precise literature measurements. The final two fits were performed to understand the prior dependence of the inferred eccentricity, as well as the impact of the eccentricity-inclination degeneracy (see, e.g., Ferrer-Chávez, Wang, and Blunt, 2021).

For all fits, we used the `ptemcee` implementation of `emcee` (Vousden, Farr, and Mandel, 2016; Foreman-Mackey et al., 2013), a parallel-tempered affine-invariant MCMC sampling algorithm. All runs used 20 temperatures and 1000 walkers. After an initial burn-in period of 100,000 steps, each walker was run for 100,000 steps. Every 100th step was saved.[10] The chains were examined visually to determine whether they had converged (see Figures 3.3 and 3.4), and the 100,000 steps of each chain post-burn-in were kept as the posterior estimate.

We applied Gaussian priors on parallax and total mass using the values given in Table 3.1. Both parameters do not significantly correlate with any other fit parameters, and the marginalized posteriors on these parameters reproduce their priors (visible, for example, in Figure 3.6).

The Impact of GRAVITY

Figures 3.5 and 3.6 visualize the posterior of fit #4, using all astrometric data and assuming a uniform prior on eccentricity, and Figure 3.7 compares each of the orbit fits that use varying data subsets. These plots show the transformative impact of GRAVITY data on the orbital uncertainties. The semimajor axis, eccentricity, and inclination posteriors all tighten significantly after including just two GRAVITY astrometric epochs. Figure 3.7 also shows that the results of a preference for moderate eccentricity and near edge-on inclination are robust to outliers, by showing nearly identical posterior distributions regardless of which GRAVITY epoch is used in the fit. It is also apparent from Figure 3.7 that the majority of the orbital parameter information is coming from the two GRAVITY epochs, as evident by the similarity

[9] The linear coefficient was shared via private communication by E. Nielsen.

[10] In `orbitize!`, this configuration corresponds to the following variable definitions: `num_temps = 20; num_walkers = 1000; num_steps = 1_000_000_000; burn_steps = 100_000; thin = 100`.

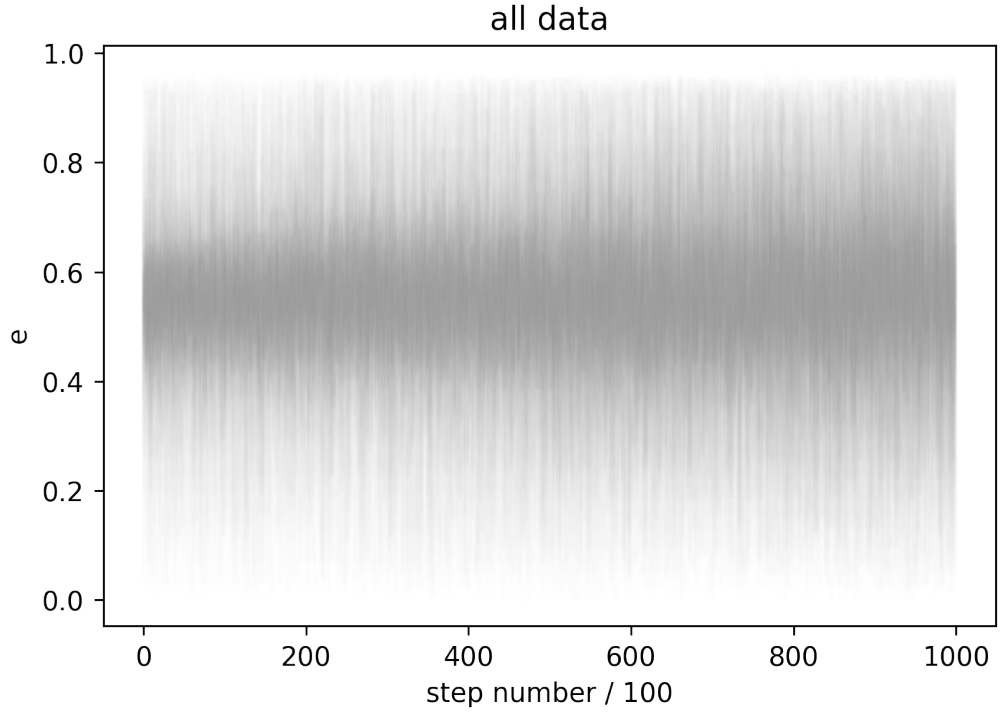

Figure 3.3: An example converged MCMC run that was used to determine that the chains were sufficiently burned-in. A random subset of 100 chains are plotted as a function of step number. The similar distribution of walkers at each step is an indication that the number of burn-in and total steps is adequate, and that the resulting posterior estimates are trustworthy.

between the accepted fit and the GRAVITY-only fit. This test highlights the power of GRAVITY precision astrometry, but also warrants a warning: any unquantified systematics in the GRAVITY data could significantly change the recovered orbital parameters. The orbital constraints reported here rely heavily on the detailed work that the exoGRAVITY team has undertaken to extract accurate astrometry (e.g. Lacour et al., 2019b, Gravity Collaboration et al., 2020). HIP 65426 b should continue to be astrometrically monitored by GRAVITY in order to redistribute the impact of the two GRAVITY epochs reported here.

A Moderate Eccentricity?

All of the orbit fits that include GRAVITY data show a preference for moderate eccentricities ($e \sim 0.5$). In order to understand the significance of this preference, we performed an additional fit using a linearly descending eccentricity prior (Equation 3.1). Figure 3.8 shows that the eccentricity posterior is prior-dependent, even though

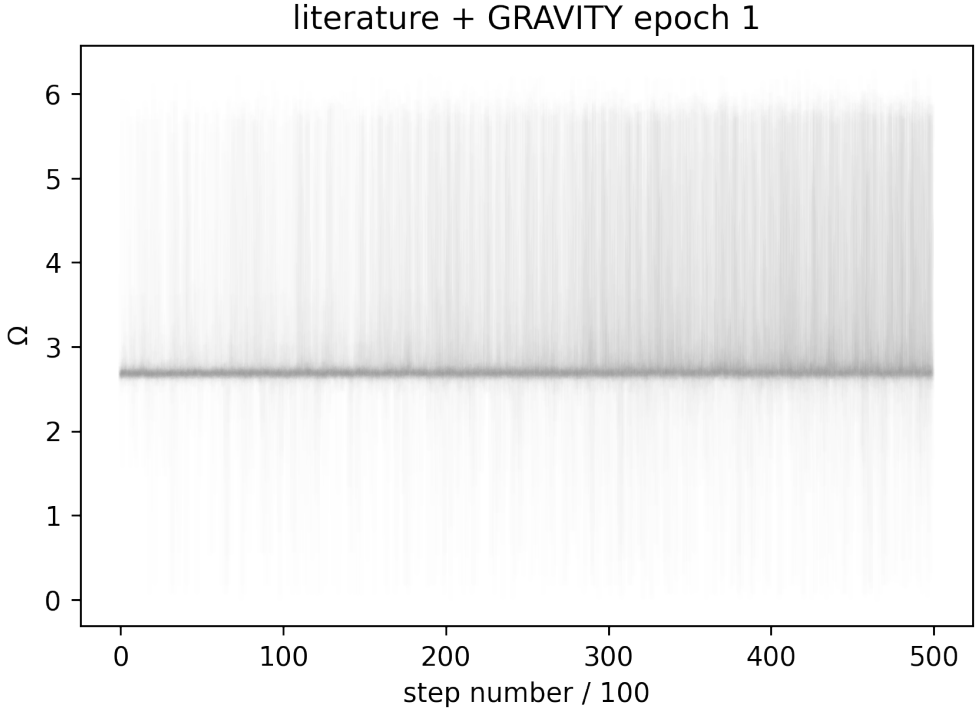

Figure 3.4: An example unconverged MCMC run that was used to determine that more burn-in steps were needed in order to rule out non-convergence. A random subset of 100 chains are plotted as a function of step number. The color gradient from left to right (getting darker toward the right) indicates that the chains are not fully converged. As an aside, because they are bimodal for astrometry-only orbits, Ω and ω are good parameters to use for assessing convergence.

both a linearly decreasing and a uniform eccentricity prior result in preferences for non-zero eccentricities. We next performed a series of maximum-likelihood fits, fixing eccentricity to a specific value for each, in order to examine how the changing maximum likelihood value was influencing the posterior shape. The results of this test are shown in Figure 3.9. This figure shows that the maximum likelihoods achieved by low ($e < 0.2$) and moderate ($e = 0.5$) orbits are comparable, but that the highest achieveable maximum likelihood occurs at low eccentricities. This plot, along with Figures 3.6 and 3.8 allow us to construct the following explanation of the shape of the eccentricity posterior: at low ($e < 0.2$) eccentricities, likelihood is high, but prior volume is low (i.e. there are "fewer" circular orbits that fit the data well, even though those that do fit tend to fit very well). At moderate eccentricities, prior volume is high, and likelihood is still very good, leading to a posterior peak. At high eccentricities ($e>0.6$), prior volume is very high, but likelihood is low, leading

to a decreasing posterior probability as a function of eccentricity. The "uptick" at very high eccentricities (e>0.8) is caused by a degeneracy between eccentricity and inclination (see Ferrer-Chávez, Wang, and Blunt, 2021 for a detailed discussion), together with a nonlinear relationship between eccentricity and sky projection. As the eccentricity asymptotically approaches 1, the corresponding inclination must increase more and more rapidly to reproduce the astrometric data (see the covariance between eccentricity and inclination in Figure 3.6). This causes a characteristic "banana-shaped" covariance, typical of incomplete orbits (see Blunt et al., 2019 for another example in the context of an incomplete radial velocity orbit). We can explain the maximum a posteriori (MAP) shift to lower eccentricities when using a linearly decreasing prior, therefore, as occurring because the prior volume at moderate and high eccentricities decreases when we switch to a linearly decreasing eccentricity prior.

Model Selection

There exist many model selection metrics a statistician can use to pick an "optimal" model. In this section, we briefly explain two such metrics, drawing heavily from Gelman, Hwang, and Vehtari (2014), and use them to support the explanation in the previous section.

We can choose to define a model's "goodness" by its ability to predict unseen data points. A "good" model will predict new, unseen data better than a "bad" model. However, a posterior comprises many models, each with its own probability based on existing data. In addition, we do not know a priori what we expect unseen data to look like; this is the purpose of model fitting. The logic Gelman, Hwang, and Vehtari (2014) review is that one can define the "expected (log) predictive density" (elpd) for a new data point as the expected posterior probability of the new datapoint, weighted by the probability of the new datapoint itself (typically unknown):

$$\text{elpd} = \int \log p_{\text{post}}(y_i) f(y_i) dy \tag{3.2}$$

where p_{post} is the posterior probability, y_i is an unseen data point, and $f(y)$ is the unknown physical function generating new data. Taking it one step further, we can aim to compute the expected log *pointwise* predictive density (elppd), which sums the elpd for each new datapoint over an arbitrary-sized unseen dataset. The problem is then determining an estimator of this expected predictive density. (See Gelman, Hwang, and Vehtari, 2014 section 2.3 for more detail).

Both the AIC and the WAIC take the approach of defining the elppd of a new datapoint as the sum of two terms: the first term represents the model's ability to predict *existing* datapoints, and the second term is an overfitting penalty. They differ in their definitions of each of these terms. Both of these estimators converge to the actual lppd under certain conditions.

The AIC computes the first term as the probability of the existing data *given the maximum likelihood model*. The second term is simply the number of free parameters in the model. This definition has the pleasing property of being an unbiased estimator of the lppd for Gaussian posteriors which were computed using a flat prior. It may already be clear that we will choose to argue that the AIC is insufficient for the orbit described in this paper, as the orbital posterior (see Figure 3.6) is quite non-Gaussian. In addition, effects of prior volume are important, and the maximum likelihood estimate (which is the only relevant quantity for the AIC) is different from the MAP estimate for both eccentricity priors discussed in the previous section. Because the maximum likelihood estimate is approximately equal for the free-eccentricity and fixed-eccentricity models, while the number of free parameters differs, we can understand why the AIC metric favors the ecc=0 model (Table 3.5).

The WAIC uses the whole posterior, not just the maximum likelihood estimate, to compute both the first and second terms. Because of this, Gelman, Hwang, and Vehtari, 2014 call it a "a more fully Bayesian approach for estimating the out-of-sample expectation." The first term is exactly the average predictive density for all existing datapoints (Gelman, Hwang, and Vehtari, 2014 Eq 5), and the second term has a few definitions in the literature. In this article, we define the WAIC using Gelman, Hwang, and Vehtari (2014) Eq 12, which sets the overfitting term to be the sum of the variances of the lppd values of individual existing datapoints. Under this definition, a model which is overfitting will have (on average) smaller variances in predicted posterior probabilities of existing datapoints. Because the WAIC includes information from the whole posterior, implicitly taking into account effects of prior volume, we might expect that the WAIC is a better predictor for our thoroughly non-Gaussian posterior. Indeed, the WAIC shows a much less clear distinction between the various eccentricity models than the AIC (Table 3.5). The circular model is still preferred, but only by ΔWAIC=1, which many authors argue is "essentially indistinguishable." This motivates us to argue that, although the AIC prefers the fixed ecc=0 model, the three models are indistinguishable in terms

Table 3.5: Model comparison metrics for orbit-fits including all available astrometric data and varying the eccentricity prior. Note: for the AIC computation, M_{tot} and parallax were not included in the number of free parameters, since they were both highly constrained by their respective priors.

	uniform e prior	linear e prior	e=0
AIC	46.0	46.0	40.0
WAIC	42.4	42.1	41.6

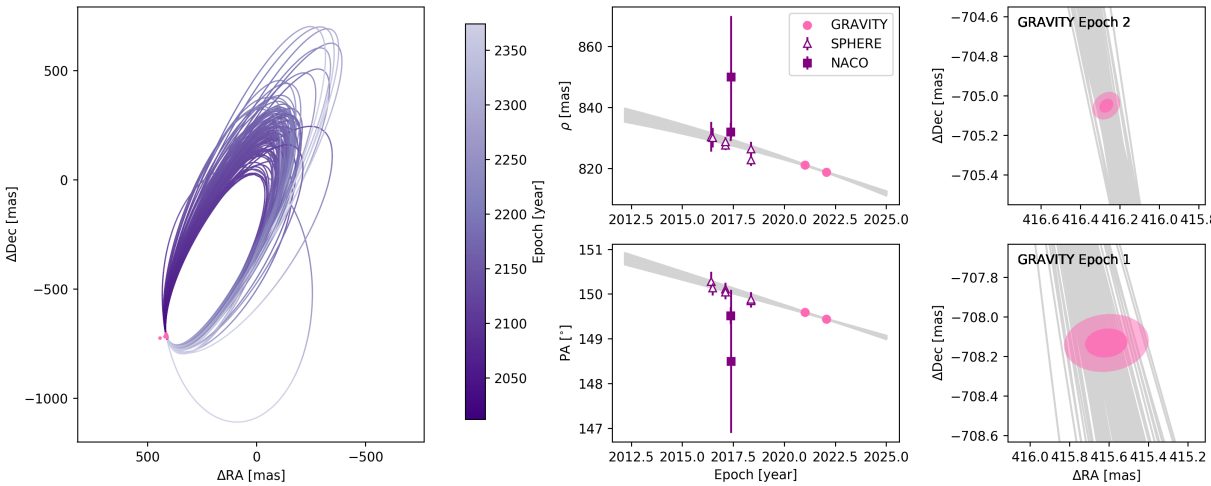

Figure 3.5: Sky-projected visualization of the posterior of the orbit fit #4 described in the text. **Left:** 100 orbit tracks projected onto the plane of the sky, colored by elapsed time. The astrometric data are visible as pink points in the bottom left corner of the panel. **Middle column:** the same 100 posterior orbits (grey) in separation (top) and position angle (bottom) vs time, together with the astrometric data used for orbit-fitting. **Right column:** the same 100 posterior orbits (grey), together with earlier (bottom) and later (top) astrometric measurements taken with VLTI/GRAVITY. 1- and 2-σ error ellipses are shaded in dark and light pink, respectively. **Takeaway:** The two VLTI/GRAVITY epochs are ~50x more precise than existing astrometric measurements of HIP 65426, and reduce the posterior uncertainty.

of expected predictive ability. In other words, **we do not rule out a moderate eccentricity**.

3.5 Spectral Analysis

The GRAVITY K-band spectrum (R=500) published in this work overlaps in wavelength coverage almost completely with the higher-resolution SINFONI spectrum (R~5600), and therefore does not add significant spectral information to the HIP

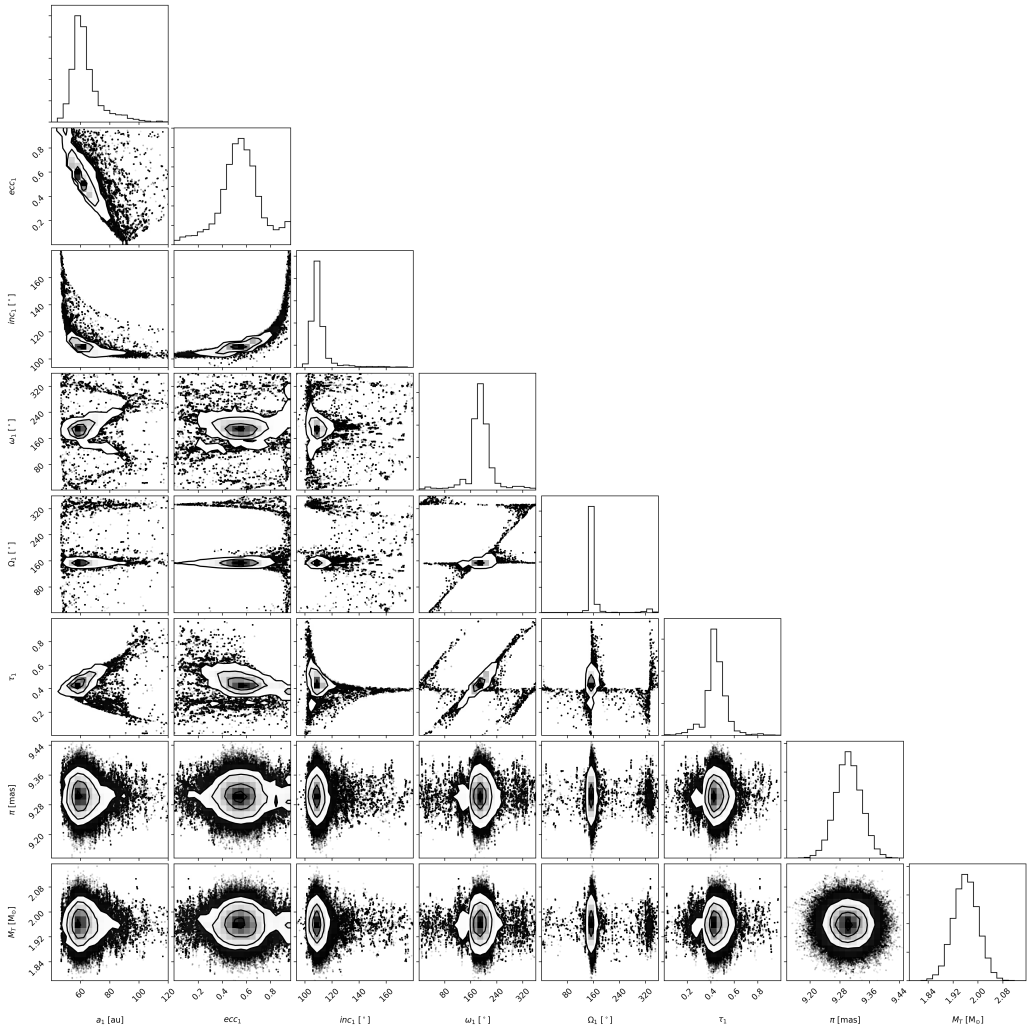

Figure 3.6: Corner plot of the posterior of the accepted orbit fit, using all data and assuming a uniform eccentricity prior. Diagonal panels show marginalized 1D histograms of posterior elements, and off-diagonals show 2D covariances between posterior elements. 1, 2, and 3-σ contours are outlined in the covariance panels, and individual posterior samples outside of the 3-σ boundaries are plotted directly as black dots. **Takeaway:** the 1D marginalized posterior distributions of semimajor axis and inclination are well constrained. Strong covariances are apparent, in particular between eccentricity, inclination, and semimajor axis.

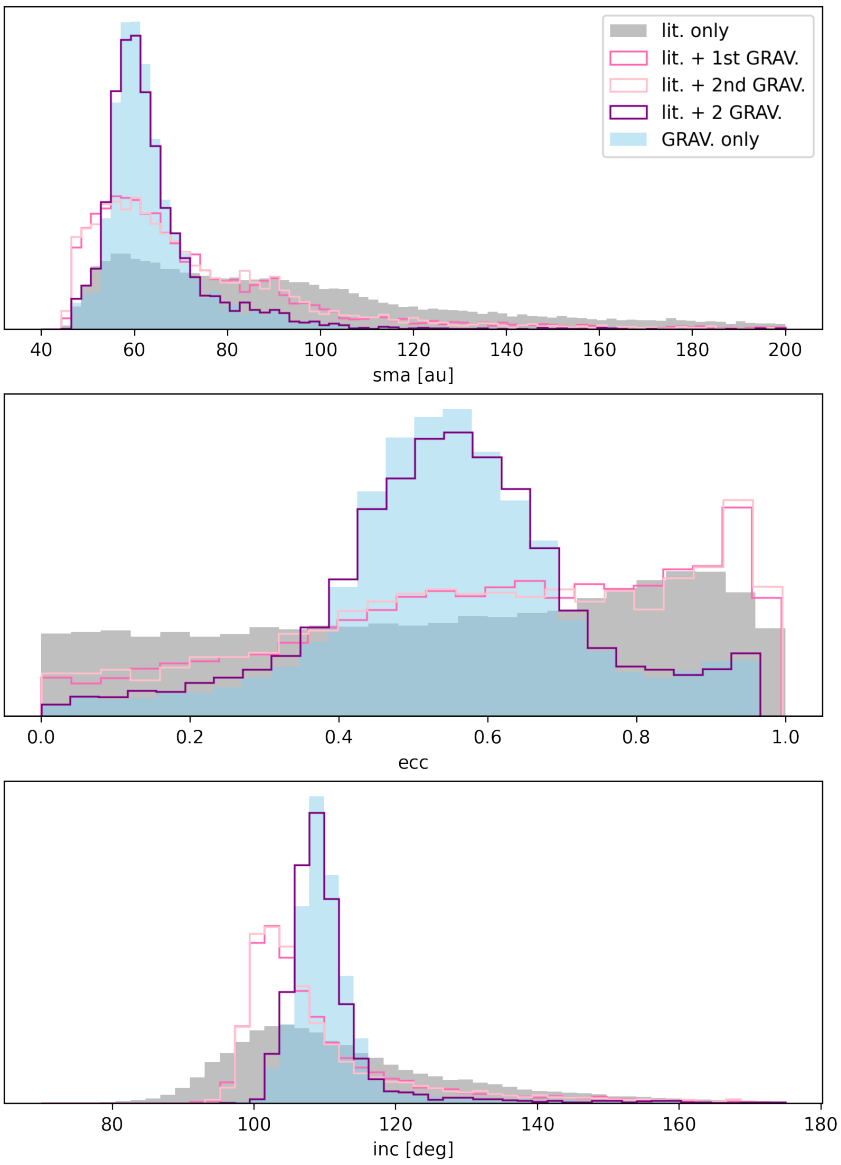

Figure 3.7: Relative constraining power of the astrometric data for semimajor axis (top), eccentricity (middle), and inclination (bottom). The results of the following fits are shown and compared: (1) only literature astrometry (i.e. no GRAVITY data; grey), (2) literature astrometry and the first epoch of GRAVITY data (dark pink outline), (3) literature astrometry and the second epoch of GRAVITY data (light pink outline), (4) only GRAVITY astrometry (i.e. no literature data), and (5) all astrometric data (i.e. the accepted fit; purple outline). **Takeaway:** most of the constraining power of the fit comes from the GRAVITY data, evident by the similarity between the GRAVITY-only fit and the accepted fit. In addition, neither GRAVITY point alone drives the fit, as evidenced by the similarity between fits (2) and (3). In other words, the posterior preference for moderate eccentricities is robust to the possibility that one of the two GRAVITY epochs is an outlier.

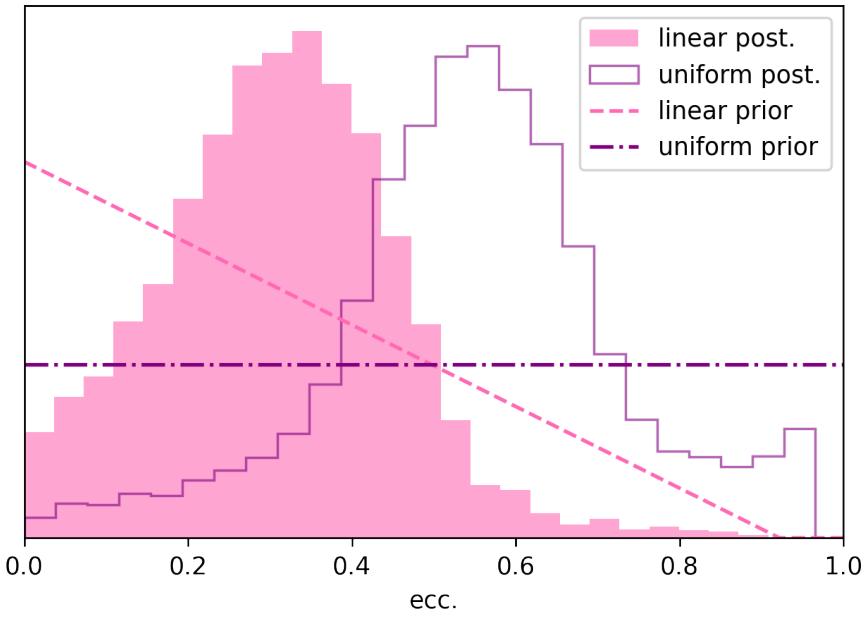

Figure 3.8: 1D marginalized eccentricity posteriors for fits with uniform (purple) and linearly decreasing (pink) priors on eccentricity. The priors themselves are also plotted as lines of the same colors. **Takeaway:** the eccentricity posterior depends on the choice of prior. However, both the linearly decreasing prior and the uniform prior result in posterior peaks at moderate eccentricity values.

Table 3.6: 68% marginalized posterior credible intervals for each free parameter/orbital elements in each of the orbit fits described in section 3.4.

it name	a [au]	ecc	inc [deg]	ω_p [deg]	Ω [deg]	τ_{58849}	π [mas]	M_{tot} [$M_\odot$]
literature data only	$92.5^{+88.7}_{-32.0}$	$0.58^{+0.29}_{-0.38}$	$109.5^{+20.5}_{-11.2}$	$181.6^{+107.6}_{-128.9}$	$171.9^{+163.9}_{-45.3}$	$0.36^{+0.30}_{-0.17}$	9.30 ± 0.03	1.96 ± 0.04
literature + GRAVITY epoch 1	$66.1^{+28.3}_{-12.7}$	$0.64^{+0.26}_{-0.31}$	$105.6^{+13.9}_{-5.0}$	$195.6^{+74.0}_{-95.9}$	$156.9^{+171.0}_{-8.1}$	0.44 ± 0.15	9.30 ± 0.03	1.96 ± 0.04
literature + GRAVITY epoch 2	$66.3^{+27.2}_{-13.0}$	0.63 ± 0.27	$105.4^{+14.9}_{-4.8}$	$194.3^{+70.5}_{-101.6}$	$156.5^{+172.2}_{-7.6}$	0.43 ± 0.15	9.30 ± 0.03	1.96 ± 0.04
all data	$61.1^{+10.4}_{-5.5}$	0.55 ± 0.15	$109.0^{+4.4}_{-3.1}$	$189.4^{+25.0}_{-19.0}$	$154.9^{+5.8}_{-3.4}$	0.44 ± 0.08	9.30 ± 0.03	1.96 ± 0.04
GRAVITY data only	$61.4^{+9.7}_{-5.1}$	0.55 ± 0.14	$109.6^{+4.4}_{-2.9}$	$190.3^{+25.8}_{-19.3}$	$155.4^{+7.2}_{-3.6}$	0.44 ± 0.07	9.30 ± 0.03	1.96 ± 0.04
all data, ecc. fixed to 0	91.4 ± 0.5	=0	102.9 ± 0.6	–	153.1 ± 0.3	–	9.30 ± 0.03	1.96 ± 0.04
all data, decreasing ecc. prior	$72.5^{+12.9}_{-7.5}$	0.31 ± 0.13	$105.3^{+2.0}_{-1.8}$	$187.0^{+35.8}_{-30.3}$	$153.8^{+2.9}_{-2.1}$	0.45 ± 0.11	9.30 ± 0.03	1.96 ± 0.04

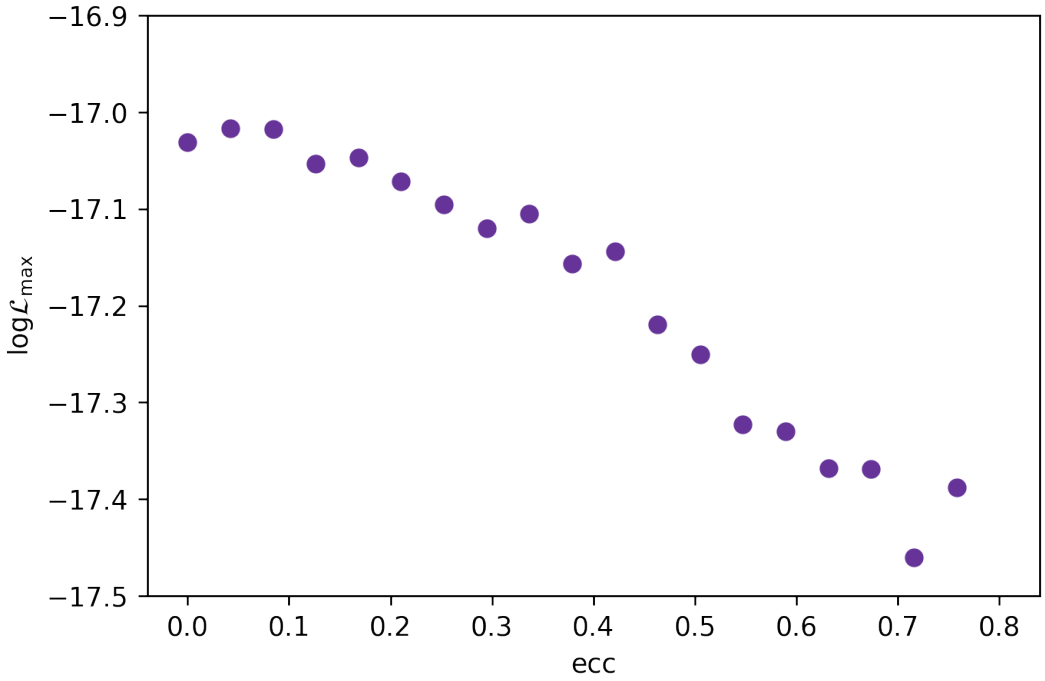

Figure 3.9: Maximum log(likelihood) as a function of eccentricity. Although the maximum a posteriori eccentricity is moderate ($\sim$ 0.5), the maximum likelihood occurs at lower eccentricities. This allows us to understand the shape of the eccentricity posterior (Figure 3.8); the likelihood is slightly higher at lower eccentricities, but the prior volume here is lower. The posterior "drop-off" at higher eccentricities is caused by a real decrease in likelihood. More eccentric orbits are less consistent with the data.

Table 3.7: 68% credible intervals of posterior fits to self-consistent model spectra grids. G is short for GRAVITY (i.e. the GRAVITY K-band spectrum was included in the fit), and Si is short for SINFONI (i.e. the SINFONI K-band spectrum was included in the fit). "yes GP" means that a Gaussian Process regression was enabled for the SPHERE IFS spectrum, the hyperparameter constraints for which are reported in the table.

fit name	$T_{\rm eff}$ [K]	log g	π [mas]	log $l_{\rm SPHERE}$	$A_{\rm SPHERE}$	[Fe/H]	C/O	R [R_J]	$RV_{\rm SINFONI}$
`BT-Settl` (G only, yes GP)	1637 ± 8	3.85 ± 0.03	$9.3^{+0.03}_{-0.04}$	$0.2^{+0.2}_{-0.1}$	$0.4^{+0.3}_{-0.2}$	–	–	1.0 ± 0.01	–
	1477^{+8}_{-7}	3.93 ± 0.07	9.3 ± 0.03	$0.3^{+0.1}_{-0.0}$	0.6 ± 0.1	–	–	1.19 ± 0.03	–
`BT-Settl` (Si only, yes GP)	1624^{+8}_{-7}	3.89 ± 0.04	9.3 ± 0.03	0.4 ± 0.1	0.5 ± 0.1	–	–	0.99 ± 0.01	75395^{+1}_{-2}
	1469^{+5}_{-4}	$3.71^{+0.07}_{-0.06}$	9.3 ± 0.03	0.4 ± 0.1	0.6 ± 0.1	–	–	1.24 ± 0.02	75396^{+1}_{-2}
`BT-Settl` (G only, no GP)	1639 ± 7	3.85 ± 0.03	9.3 ± 0.03	–	–	–	–	1.0 ± 0.01	–
	1487^{+7}_{-6}	$3.96^{+0.05}_{-0.06}$	9.3 ± 0.03	–	–	–	–	$1.17^{+0.03}_{-0.02}$	–
`BT-Settl` (Si only, no GP)	1630 ± 7	3.92 ± 0.03	9.31 ± 0.03	–	–	–	–	0.97 ± 0.01	75389^{+1}_{-2}
	1475^{+5}_{-4}	$3.81^{+0.07}_{-0.06}$	9.3 ± 0.03	–	–	–	–	$1.21^{+0.02}_{-0.03}$	75378 ± 1
`Exo-REM` (Si+G, yes GP)	1337 ± 9	$3.52^{+0.03}_{-0.02}$	9.3 ± 0.03	0.4 ± 0.1	0.5 ± 0.1	$0.15^{+0.08}_{-0.1}$	$0.595^{+0.008}_{-0.009}$	$1.51^{+0.03}_{-0.02}$	18 ± 8

65426 b SED. However, it is an independent constraint on the K-band spectrum. In this section we first compare the GRAVITY and SINFONI spectra, finding good agreement, then globally compare the HIP 65426 b spectral data with self-consistent atmosphere models to update our understanding of this object's atmosphere.

In Figure 3.10, we plot the GRAVITY and SINFONI spectra. The agreement is excellent. Both spectra agree in magnitude[11] and shape.

Following Petrus et al. (2021), we derived atmospheric properties for HIP 65426 b by comparing with two sets of self-consistent atmospheric model grids: `BT-SETTL CIFIST 2011c`[12] (Allard et al., 2003, Allard et al., 2007, Allard, Homeier, and Freytag, 2011) and `Exo-REM` (Allard, Homeier, and Freytag, 2012, Charnay et al., 2018), excellent summaries of which are given in Section 3.1 of Petrus et al. (2021). In the temperature range relevant for HIP 65426 b, the major advantages of `Exo-REM` include 1) ability to explore non-solar metallicities, and 2) the success of the `Exo-REM` cloud prescription at reproducing observations of dusty planets (like HIP 65426 b) near the L-T transition (Charnay et al., 2018). Ultimately, however, we are interested in comparing constraints from multiple independent models.

We used `species` (Stolker et al., 2020) to perform comparisons to both sets of models. In all cases, we computed posteriors using the `pyMultinest` (Buchner et al., 2014) Python interface to `multinest` (Feroz and Hobson, 2008, Feroz, Hobson, and Bridges, 2009, Feroz et al., 2019) with 1000 live points. The results of all atmosphere model fits are shown in Table 3.7.

`BT-SETTL CIFIST`

We performed four variations of `BT-SETTL CIFIST` comparisons to our full spectral dataset, by 1) using either the GRAVITY or SINFONI spectrum[13] in order to compare their relative constraining power, and 2) fitting for correlated noise in the SPHERE IFS data, following Wang et al. (2020) (see their Equation 4). All fits performed in this and the next section are summarized in Table 3.7. The results of the two fits allowing correlated noise in the SPHERE data are shown in Figure 3.11. The fit only including GRAVITY data is shown in purple, and the fit only including SINFONI data is shown in pink. The two fits are consistent overall, as we would

[11] Bearing in mind that the raw extracted GRAVITY spectrum was scaled to match the SPHERE K-band photometry (which overlap in wavelength)

[12] https://phoenix.ens-lyon.fr/Grids/BT-Settl/CIFIST2011c/

[13] Although we downsampled the resolution of the SINFONI spectrum in order to compare with the GRAVITY spectrum in Figure 3.10, we fit the SINFONI spectrum at its native resolution.

expect given the consistency of the spectra themselves. Like Petrus et al. (2021) (but unlike Carter et al. (2022)), we also recover a bimodal posterior in surface gravity and effective temperature, regardless of which of the two K-band spectra we use.

Our initial `BT-SETTL CIFIST` fits allowed a free RV offset parameter for the SINFONI data, but this value consistently converged to unphysically large values, perhaps indicating an underlying problem with the model grid. We therefore opted to fix the RV of the SINFONI data to 0 for the fits presented here. Toggling on and off a GP for the SPHERE data does not significantly impact the physical parameters inferred, even though there are clear correlated residuals in the SPHERE data (Figure 3.12). The SPHERE residuals are visualized in Figure 3.12, where both modes are plotted along with the SPHERE IFS data. Toggling on a GP for this dataset allows the model to treat the deviations from both models as correlated noise, perhaps due to imperfect speckle subtraction at the planet location, photometric calibration errors due to telluric absorption, and/or model imperfections.

`Exo-REM`

Because the agreement between the `BT-SETTL CIFIST` model fits is good regardless of whether we use the SINFONI or GRAVITY spectrum, we only perform `Exo-REM` fits using *both* datasets.

The available `Exo-REM` grid is computed for a grid of metallicities, allowing us estimate the atmospheric metallicity and C/O ratio. However, the grid we used only computes predictions out to 5 μm, so we excluded the two longer-wavelength JWST/MIRI photometry for these fits. The results are shown in Figures 3.13 and 3.14, and summarized in Table 3.7. A solar metallicity with a C/O ratio of 0.6 is preferred. We find GP parameters for the SPHERE dataset that are consistent with the results from the `BT-Settl CIFIST` fits (previous section), a good sanity check that the model is similarly treating correlated noise in both cases. We also recover a planetary radial velocity value consistent with that reported in Petrus et al., 2021, which is an excellent sanity check given that their planetary RV was computed by cross-correlation with the SINFONI spectrum alone, and not through SED fitting. Unsurprisingly, our lower-resolution GRAVITY spectrum does not permit a direct RV measurement. The effective temperature recovered by this fit is about 150K lower than the low-temperature `BT-SETTL CIFIST` mode, increasing our overall confidence in this low-temperature interpretation. However, the surface

gravity value pushes against the low-gravity end of the grid, which casts doubt on the results of this fit.

The constraints we derived in this analysis are significantly more precise than those reported in Petrus et al. (2021), because at the time of that paper's publication the `Exo-REM` grid predictions were only available for K-band. We were able to use the updated `Exo-REM` grid to compare with all available spectral information. However, the uncertainties are likely underestimated, in particular because these fits do not account for interpolation errors, an important point that we discuss in more detail in the next section.

From scrutinizing the residuals shown in Figure 3.14, it is also clear that the NACO and JWST photometry beyond 3 μm is systematically higher than the `Exo-REM` models. Therefore, in addition to potentially underestimated errors due to unmodeled physics in the atmosphere grid, interpolation errors, and correlated noise in the SPHERE IFS spectrum, the best fit is still not perfect, again pointing to potential inaccuracies in our comparison. Future work could attempt to fit offsets to the three NACO points in order to reduce the discrepancy, and/or inflate the errors in these or other spectral datapoints. Altogether, the values and errors derived from grid comparison should be treated with caution.

3.6 Discussion & Conclusion

Interpreting the Eccentricity Constraints

Because of the strong degeneracy between eccentricity and inclination, it is most straightforward to report constraints on these parameters in two dimensions.[14] For fit # 4, including all available astrometry and applying a uniform eccentricity prior, the 1-σ upper limit on both parameters is e=0.7/inc=110°, and the 2-σ upper limit is e=0.8/inc=120°. Applying a non-uniform, linearly decreasing prior on eccentricity, which previous work shows is appropriate for the cold Jupiter population (Bowler, Blunt, and Nielsen, 2020a, Nagpal et al., 2023), tightens these upper limits even more. These are still tenuous constraints, but they are driven by the likelihood, not just by the prior. This is the first time we are obtaining eccentricity posteriors on this object that do not simply reproduce the prior. It is worth emphasizing the difficulty of measuring the eccentricity of an object almost 100 au from its star; previous studies of this object report that it would take 5-10 years of orbit monitoring before resolving orbital curvature and constraining HIP 65426 b's eccentricity. With the

[14] The direction of orbital motion on the plane of the sky constrains the inclination to > 90°.

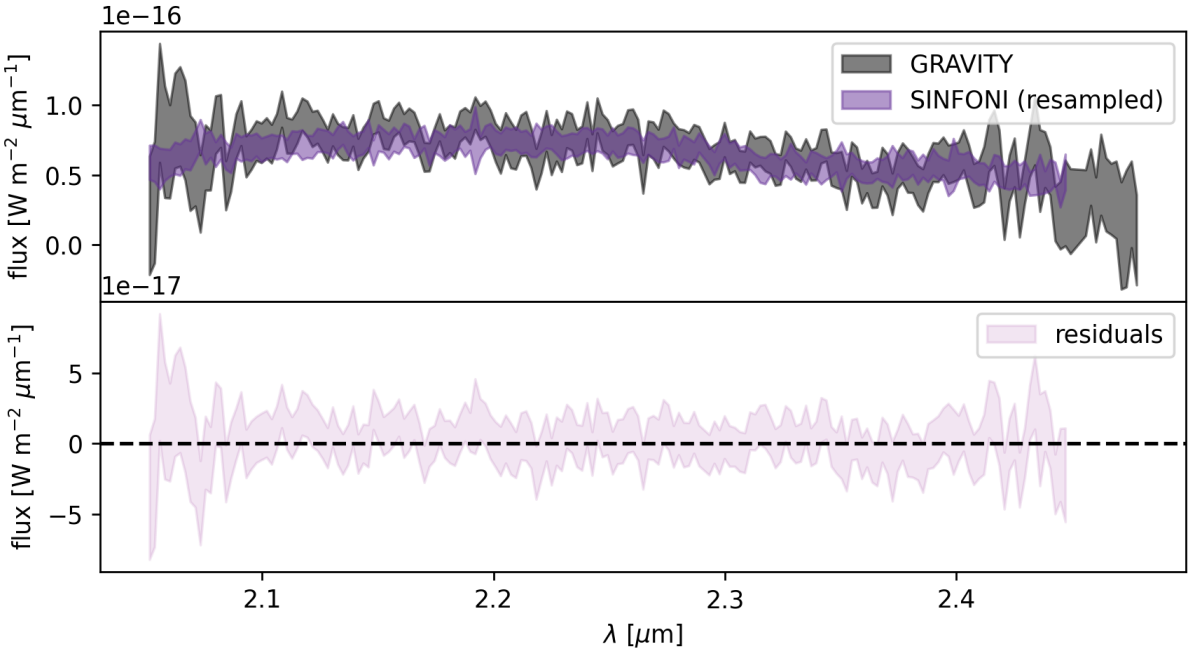

Figure 3.10: GRAVITY and SINFONI K-band spectra comparison. **Top:** GRAVITY (grey) and SINFONI (purple) 1σ flux confidence intervals are shown as filled bands. The SINFONI spectrum was resampled onto the GRAVITY wavelength grid using `spectres` (Carnall, 2017). **Bottom:** The residuals, with propagated uncertainties, are shown relative to the flux=0 line. **Takeaway:** the agreement between these two independent datasets is excellent.

precision of VLTI/GRAVITY, we are able to "speed up time" and obtain eccentricity constraints sooner, with fewer measurements. Other directly imaged exoplanets with well-constrained eccentricities are generally much closer to their stars (e.g. β Pic b and c, at 3 and 10 au).

A potentially useful outcome of this paper is a generalize-able prescription for interpreting eccentricity posterior for incomplete orbits. To have confidence in an eccentricity measurement from a posterior, we suggest showing that the eccentricity posterior is:

- prior-independent, that is, driven by the likelihood and not phenomenologically different under different prior assumptions,
- inconsistent with a circular orbit, ideally beyond 3-σ for varied prior assumptions, and
- not dependent on a single data point.

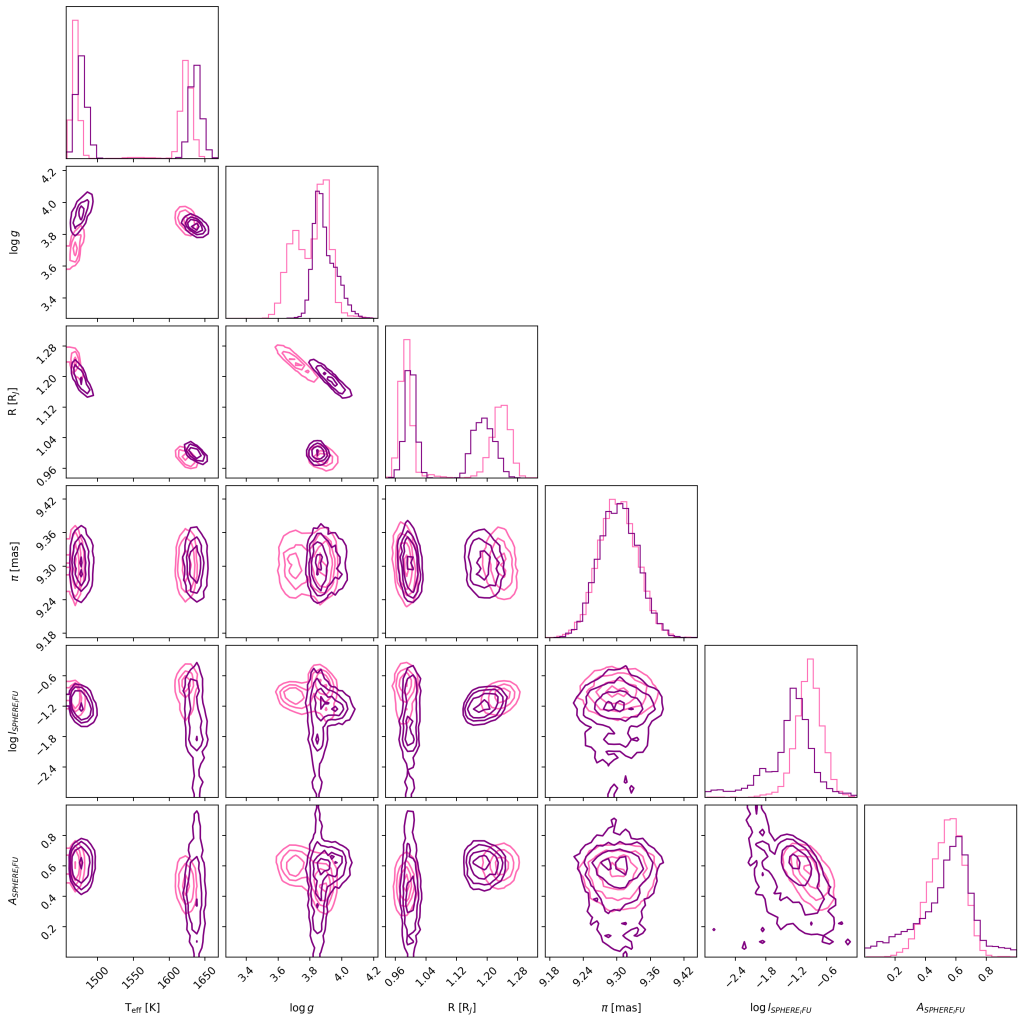

Figure 3.11: Results of forward-modeling the photometric and spectral data of HIP 65426 b by comparing with the `BT-SETTL CIFIST` model grid. Posteriors over the free parameters in the fit, as well radius, a derived parameter, are shown. Fits performed using GRAVITY K-band spectra are shown in purple, and fits performed using SIFNONI K-band spectra are shown in pink. The GP hyperparameters (defined as in Wang et al., 2020 Equation 4) to the SPHERE IFS spectral data (length scale and amplitude) are shown as well. **Takeaways:** as expected, log g correlates strongly with radius and $T_{\rm eff}$. Two families of solutions are apparent at high (1.3 R_J) and low (0.9 R_J) radii.

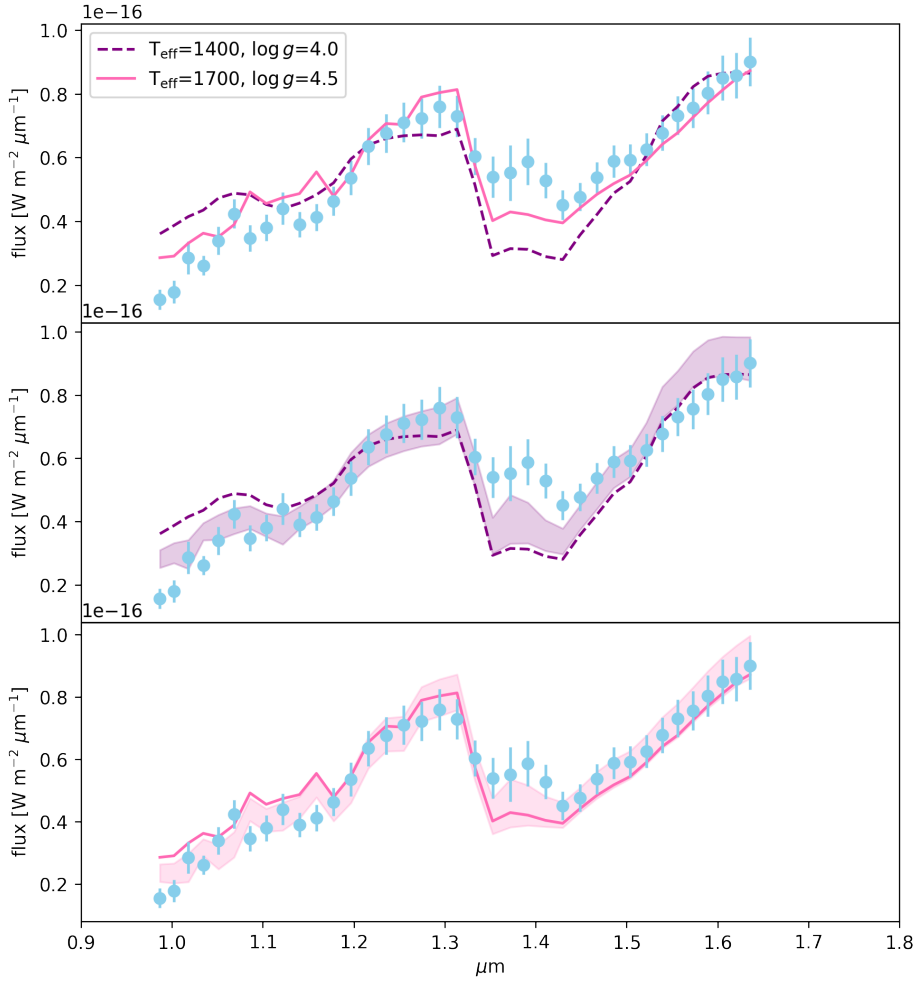

Figure 3.12: `BT-SETTL CIFIST` models representing the two posterior peaks shown in Figure 3.11, together with the SPHERE IFS data. **Top:** both models, resampled onto the SPHERE IFS wavelength grid using `spectres` (Carnall, 2017), and multiplied by a scalar chosen to minimize the sum of squared residuals for the SPHERE IFS data alone. The SPHERE IFS data are shown as blue points, with error bars representing their reported statistical uncertainties. **Middle:** the low-$T_{\rm eff}$ model (dashed purple line), SPHERE IFS data, and Gaussian Process $1-\sigma$ uncertainties (solid purple band; computed using the MAP GP parameters). **Bottom:** same as middle, but the high-$T_{\rm eff}$ model is shown as a solid pink line, and GP uncertainties as a pink band. **Takeaway:** there are correlated residuals in the SPHERE passband for both of the $T_{\rm eff}$ modes recovered from comparisons to the `BT-Settl CIFIST` grid, which we model with a squared-exponential Gaussian process.

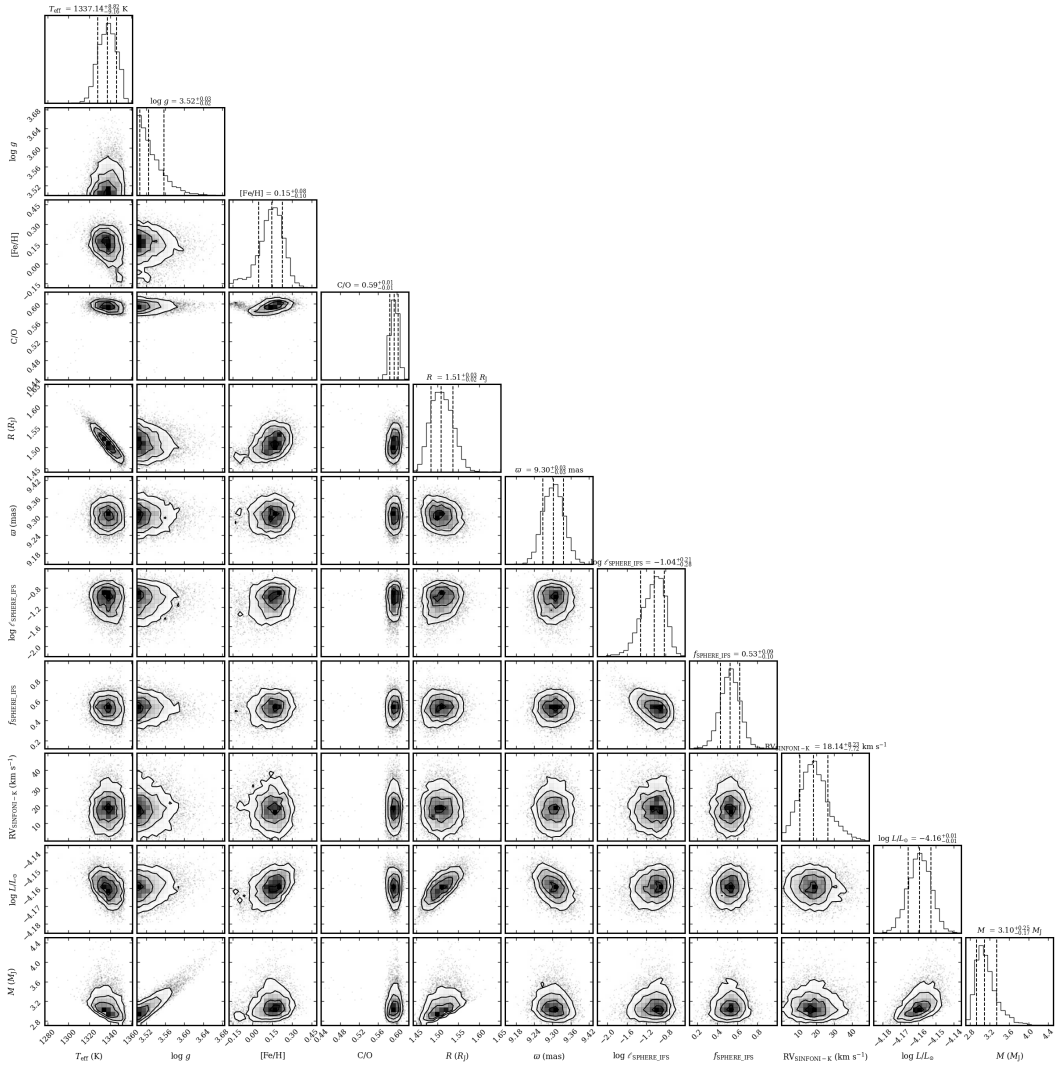

Figure 3.13: `Exo-REM` posterior fits to all data, showing 2D covariances and 1D marginalized posteriors over fitted model parameters and derived parameters (radius, luminosity, and mass). Also see Figure 3.14. Note: parallax is denoted ϖ here, and π elsewhere in the text. `Takeaways:` the `Exo-REM` derived atmosphere parameters are about 150K lower than the low-$T_{\rm eff}$ `BT-SETTL CIFIST` parameters (Figure 3.11). The surface gravity hits the edge of the available grid. A slightly super-solar-metallicity atmosphere with a C/O ratio of 0.6 is favored, although there are likely systematic errors unaccounted for in this fit.

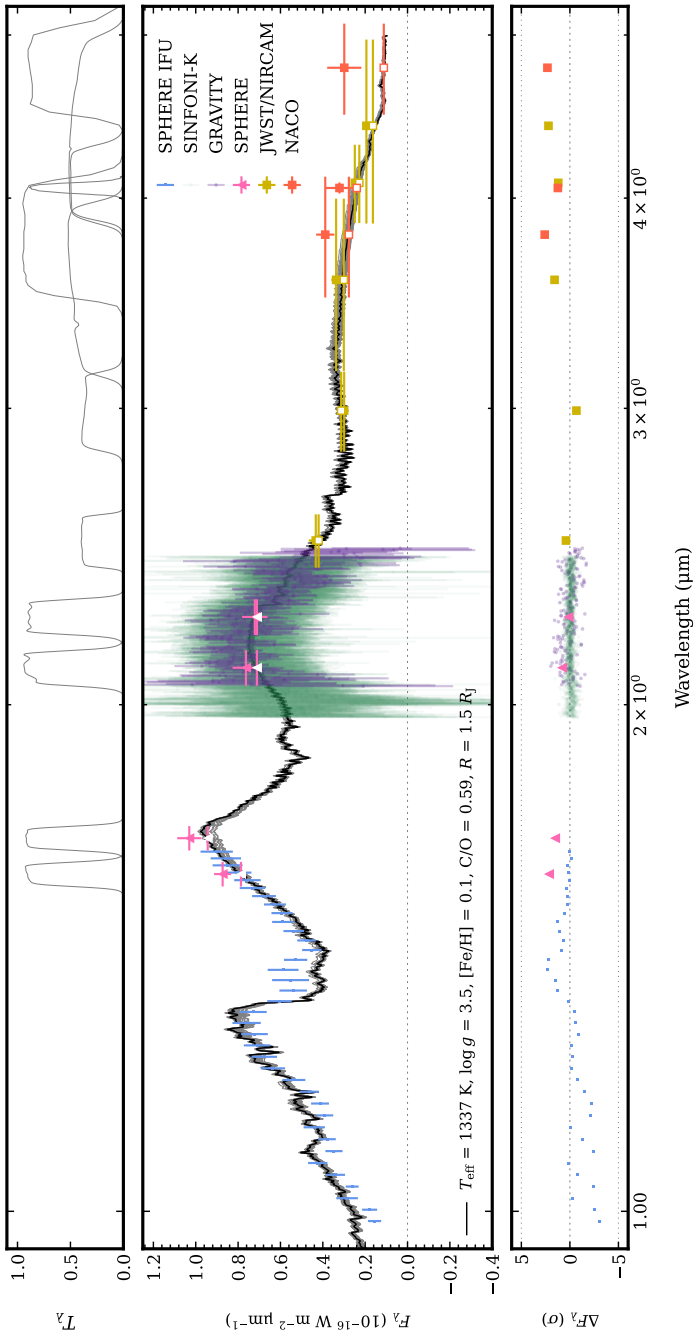

Figure 3.14: Full best-fit model SED from comparison with the Exo-REM grid, together with all fitted spectral data. **Top**: transmission functions of photometric bands (each plotted as a data point with errors immediately beneath). **Middle:** MAP model spectrum, along with 100 random draws from the posterior (in grey; difficult to distinguish from MAP spectrum). The spectral data are plotted as points with error bars. The horizontal bars of the photometry points indicate their spectral bandpass. The corresponding band-integrated model predictions are overplotted as empty symbols. **Bottom:** MAP model residuals. **Takeaways:** Overall, the model spectrum fits well. Correlated noise is visible in the residuals of the SPHERE dataset. The NACO and JWST/NIRCAM points beyond 3 μm are underestimated by the model.

HIP 65426 b does not yet satisfy the first two criteria, so continued orbital monitoring with VLTI/GRAVITY will be important to robustly measure its eccentricity. Planetary RVs with sub-km s^{-1} precision (Figure 3.2), which could be obtained with high spectral resolution instruments like CRIRES, would also further constrain the eccentricity (e.g., Snellen et al., 2014, Schwarz et al., 2016).

With these caveats in mind, the posteriors presented in this paper favor a low or moderate eccentricity. Given the results of Marleau et al., 2019, this is a preliminary hint that HIP 65426 b did not attain its current position via scattering after disk dispersal.

Interpreting the Atmosphere Constraints

First, it is important to understand the limits of the spectral interpretation approach we have taken in this paper. Self-consistent grid modeling involves interpolating spectra in multiple dimensions during the spectral inversion, while the variations of the synthetic spectra along the grid are not linear in these dimensions (e.g. Czekala et al., 2015, Petrus et al., 2021). Missing or incorrect physics in the model grids, therefore, is not the only source of error. Performing atmospheric retrievals to evaluate the grid comparison results of this study is an important next step.

Keeping this limitation in mind, in this work we repeated the analysis presented in Petrus et al. (2021), which compared the available spectral and photometric data of HIP 65426 b with the self-consistent `BT-Settl` and `Exo-REM` grids. This work benefited from additional data (in particular, the medium-resolution K-band GRAVITY spectrum, an additional NACO photometric point presented in Stolker et al., 2020, and the JWST photometery presented in Carter et al., 2022), and the expanded capability of `Exo-REM` to handle data outside of K-band. Like Petrus et al. (2021), we recovered two modes in our `BT-Settl` posterior fit, regardless of which K-band spectrum we use and whether we included a correlated noise model for the SPHERE IFS data: one at a higher radius of 1.2 R_J, and one at a lower radius of 1.0 R_J. Both modes are significantly below the hot-start radius of 1.4 R_J derived in Carter et al., 2022, reinforcing the tension that paper originally pointed out.

Like Petrus et al. (2021), we only recover a single posterior mode when comparing with the `Exo-REM` grid, which is about 150K cooler than the coolest MAP `BT-Settl` fit. The Gaussian process hyperparameters applicable to the SPHERE datasets are consistent across both grids, which we interpret as correlated observational noise, but which could also be a systematic problem common to both grids. The `Exo-REM`

fit posterior favors a slightly super-solar metallicity and a C/O ratio of 0.6. However, the log g posterior hits the edge of the grid, so we strongly encourage skepticism of these derived values. The C/O ratio, metallicity, and other atmosphere properties we present here are consistent but more precise than those presented in Petrus et al. (2021), keeping in mind that the systematic errors are likely underestimated. A planetary metallicity *relative* to its host star's is generally expected to reflect its formation condition (Öberg, Murray-Clay, and Bergin 2011, Madhusudhan, Amin, and Kennedy 2014); in particular, formation via gravitational instability is generally held to produce an object with the same metallicity as its primary. In order to interpret HIP 65426 b's metallicity, it is important to understand the primary's metallicity. Because HIP 65426 A is a fast rotator with few spectral lines, a direct metallicity measurement is difficult, but if we take the metallicity of other members of its moving group as representative, it should be approximately sub-solar, modulo scatter among individual stars. (see Section 4.3 of Petrus et al., 2021).

Taken with a big grain of salt, the super-solar metallicity of HIP 65426 b, expected ~sub-solar metallicity of HIP 65426 A, and the planetary C/O ratio of 0.6, as constrained by comparison with the `Exo-REM` grid, are consistent with formation via core accretion past the CO snowline (Öberg, Murray-Clay, and Bergin 2011). The exact location of this snowline for HD 163296, which has a similar mass to HIP 65426 A, was observed to be 75 au (Qi et al. 2015, Petrus et al. 2021). This (very tentative, with many caveats, see, e.g., Mollière et al. 2022) picture provides a parallel constraint on the picture of how HIP 65426 b attained its current separation by setting an outer limit on the initial formation location before any scattering occurred.

Future Directions

Continued orbital monitoring, both with VLTI/GRAVITY and spectrographs capable of measuring additional planetary RVs will refine the eccentricity measurement of HIP 65426 b over the next few years, allowing us more insight into this specific planet's formation and further constraining the population-level eccentricity distribution of cold Jupiters. Uncertainties of $\sim \mathrm{km\,s^{-1}}$ or below are needed for relative RV measurements of HIP 65426 b to constrain orbital parameters like a and ecc, motivating observation with high resolution spectrographs like CRIRES. Other tracers of formation condition that will be measurable in the near future include the planet's obliquity (Sepulveda et al., 2023), dynamical mass (hopefully measurable upon the release of Gaia timeseries data), and spin (from high resolution spectra). Further

atmospheric characterization work, particularly retrievals that jointly model bulk atmosphere properties and trace chemical species fingerprints (see, e.g. Xuan et al., 2022) represent an important parallel path toward assessing the trustworthiness of the metallicity and C/O ratio, which will be helpful for pinpointing the formation location of HIP 65426 b. Xuan et al. (2022) showed that high resolution spectra are sensitive to a broader range of atmospheric pressures than lower-resolution spectra, leading to more robust abundance measurements, motivating high resolution measurements of HIP 65426 b in particular.

It is worth mentioning that the rapid rotation and early spectral type[15] of the primary star, HIP 65426 A, precludes precise RV measurements, so a dynamical mass measurement will rely entirely on the combination of relative (i.e. from high contrast imaging) and absolute (i.e. from Gaia) astrometry. However, radial velocity monitoring of the planet, together with continued relative astrometric monitoring, may allow for the detection of unseen inner companions, through Keplerian effects on the host star (Lacour et al., 2021) and/or planet-planet interactions (Covarrubias, Blunt, and Wang, 2022).

More theoretical work is also needed in order to interpret these results. In particular, population-level studies à la Marleau et al., 2019 could be conducted for alternate plausible formation pathways, particularly more rapid core formation via pebble accretion and gravitational instability in the protoplanetary disk. It would also be interesting to compare the existing measurements of other formation tracers, for example orbital inclination, metallicity, and C/O ratio, with their corresponding predictions in existing models.

VLTI/GRAVITY is a powerful instrument. This study has mostly focused on its ability to refine orbital eccentricity measurements in order to make dynamical inferences useful for commenting on planet formation. However, the ExoGRAVITY program is actively investigating a number of other scientific questions and observational constraints, particularly precise dynamical mass measurements and bolometric luminiosities (e.g., Hinkley et al., 2023).

Planet formation is a complex process, and we will need a diverse set of observational and theoretical tools to unravel its secrets. This paper represents a small step toward better constraints and deeper understanding.

[15] Hot, early-type stars have few spectral lines, and rapid rotators have broad spectral lines, which both decrease RV precision.

3.7 Acknowledgements

Thanks to Simon Petrus for sharing the SINFONI spectrum used in this work, and for very helpful guidance and discussion. Thanks to Aldo Sepulveda and Dan Huber for collaborative and informative discussions. S.B. and J.J.W. are supported by NASA Grant 80NSSC23K0280. G.-D.M. acknowledge the support of the DFG priority program SPP 1992 "Exploring the Diversity of Extrasolar Planets" (MA 9185/1) and from the Swiss National Science Foundation under grant 200021_204847 "PlanetsInTime." Parts of this work have been carried out within the framework of the NCCR PlanetS supported by the Swiss National Science Foundation. This work has made use of data from the European Space Agency (ESA) mission *Gaia* (`https://www.cosmos.esa.int/gaia`), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, `https://www.cosmos.esa.int/web/gaia/dpac/consortium`). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement.

S.B. wishes to acknowledge her status as a settler on the ancestral lands of the Gabrielino/Tongva people.

This research was enabled by the following software: `tinygp` (github.com/dfm/tinygp), `jax` (Bradbury et al., 2018), `numpy` (Harris et al., 2020), `pandas` (McKinney, 2010), `matplotlib` (Hunter, 2007), `spectres` (Carnall, 2017), `species` Stolker et al. (2020), `corner` (Foreman-Mackey, 2016), `scipy` (Virtanen et al., 2020), `astropy` (Astropy Collaboration et al., 2013, Astropy Collaboration et al., 2018, Astropy Collaboration et al., 2022), and `orbitize!` (Blunt et al., 2020).

Chapter 4

`orbitize!` VERSION 2

This chapter describes work that I have led myself, collaborated on, and/or reviewed, with a focus on lines of code that I have personally written.[1] *However,* `orbitize!` *is a collaborative project, and it is difficult to separate out features that are solely attributable to me or to others. I try to be as inclusive as possible in the following sections by explicitly naming major contributors for each module, but since the focus is on what I personally did, I do not go into a huge amount of detail about who did what. Parts of this chapter will be submitted to the* Journal of Open Source Software, *and on that paper, all* `orbitize!` *v2 contributors (including those who contributed legacy code) will be appropriately credited.*

4.1 Origin Story & Scope of this Chapter

As an undergrad researcher, I led a paper describing the Orbits for the Impatient algorithm (Blunt et al., 2017), a method for fitting the orbits of directly imaged planets. The original OFTI code was implemented in IDL, and was not publicly available. Over my post-bacc year at Caltech (2017-2018), Jason Wang and I collaborated with a small group of others to port the OFTI algorithm to Python, add an additional MCMC backend computation option that Jason had previously implemented, and make the merged code open-source. The resulting software package was `orbitize!`, version 1 of which we officially released in late 2018 and published in *PASP* in the following year (Blunt et al., 2020).

Some exciting uses of `orbitize!` v1 include: determination of the population-level eccentricity distributions of directly imaged planets and brown dwarfs, pointing to distinct formation mechanisms for the two populations (Bowler, Blunt, and Nielsen, 2020a), the first obliquity measurement (constraints on the planetary spin, stellar spin, and orbital plane angular momentum vectors) of a planetary mass companion (Bryan et al., 2020), and orbital constraints for the first exoplanet directly imaged with JWST (Carter et al., 2022).

[1] GitHub keeps track of who contributed each line of code, so if you want to, you can check! This metric is not perfect, however; for example, if I port some IDL code to Python and add it to GitHub, GitHub will credit only me, not the original author of the IDL code. It also, of course, does not take into account conversations or pair programming.

Since the publication of Blunt et al. (2020), myself and the `orbitize!` team have been hard at work adding new features and expanding the code's functionality. This chapter describes a subset of the new additions to the code in version 2, released in 2021, and since.

4.2 How `orbitize!` Works

`orbitize!` turns data into orbits. More specifically, relative kinematic measurements of a secondary (which I will refer to as a planet throughout this section, but which could just as easily be a white dwarf, MS star, or other massive companion) and its primary are converted to posteriors over orbital parameters through Bayesian analysis.

`orbitize!` hinges on the two-body problem, which describes the paths of two bodies gravitationally bound to each other. The solution of the two-body problem (see Seager, 2010 Chapter 2) describes the motion of each body as a function of time, given parameters determining the position and velocity of both objects at a particular epoch. There are many basis sets (orbital bases) that can be used to describe an orbit, which can then be solved using Kepler's equation. It is important, then, to be explicit about coordinate systems. I coded up an interactive visualization to define and help users understand our coordinate system, which you can check out here[2]. Blunt et al. (2020) also hosts a video of myself walking through a recording where I use and explain the coordinate system. Several still frames from that video are reproduced in Figure 4.1.

In its "standard" mode, `orbitize!` assumes that the user only has relative astrometric data to fit. To obtain these measurements, an astronomer takes an image containing two point sources[3] and measures the position of the planet relative to the star in angular coordinates. In the `orbitize!` coordinate system, relative R.A. and decl. can be expressed as the following functions of orbital parameters (Green, 1985):

[2]https://github.com/sblunt/orbitize/blob/main/docs/tutorials/show-me-the-orbit.ipynb

[3]This is extremely simplified and not exactly accurate for high-contrast imaging, but I do not think additional detail would be beneficial here.

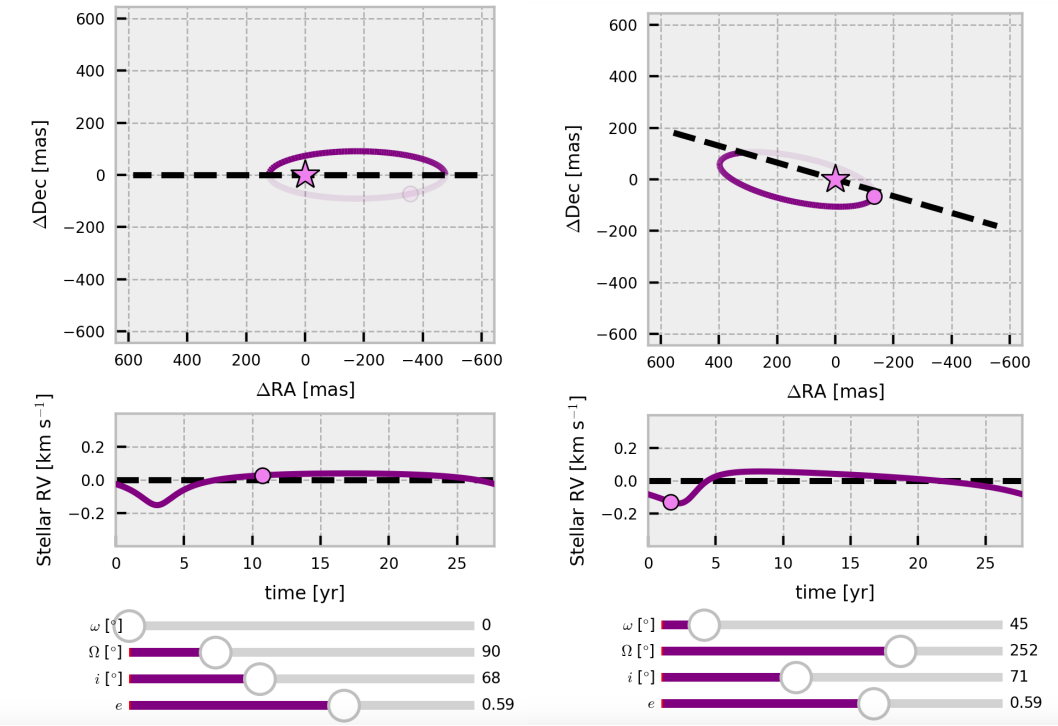

Figure 4.1: Stills from the show-me-the-orbit visualization of the `orbitize!` coordinate system. Positive radial velocity is described as moving away from the observer, and Ω=ω=0 describes an orbit with periastron pointed towards the North celestial pole.

$$\begin{aligned}\Delta\text{R.A.} &= \pi a(1 - e\cos\text{E})\left[\cos^2\frac{i}{2}\sin\left(f + \omega_\text{p} + \Omega\right) - \sin^2\frac{i}{2}\sin\left(f + \omega_\text{p} - \Omega\right)\right] \\ \Delta\text{decl.} &= \pi a(1 - e\cos\text{E})\left[\cos^2\frac{i}{2}\cos\left(f + \omega_\text{p} + \Omega\right) + \sin^2\frac{i}{2}\cos\left(f + \omega_\text{p} - \Omega\right)\right]\end{aligned} \tag{4.1}$$

where a, e, ω_p, Ω, and i are orbital parameters, and π is the system parallax. f is the true anomaly, and E is the eccentric anomaly, which are related to elapsed time through Kepler's equation and Kepler's third law:

$$
\begin{aligned}
M &= 2\pi(\frac{t}{P} - (\tau - \tau_{\rm ref})) \\
(\frac{P}{yr})^2 &= (\frac{a}{au})^3(\frac{M_\odot}{M_{tot}}) \\
M &= \mathrm{E} - e\sin\mathrm{E} \\
f &= 2\tan^{-1}\left[\sqrt{\frac{1+e}{1-e}}\tan\frac{E}{2})\right]
\end{aligned}
\tag{4.2}
$$

Kepler's equation is transcendental and nonlinear, and must in general be solved numerically. `orbitize!` employs two Kepler solvers to convert between mean and eccentric anomaly: one that is efficient for the highest eccentricities (Mikkola, 1987), and Newton's method in other cases, which is more efficient for the average orbit needing to be solved. See Blunt et al. (2020) for more detail.

From scrutinizing the set of equations 4.1 and 4.2, it is possible to understand how relative astrometry measurements constrain orbital parameters, and to understand a few important inherent degeneracies. First, notice that the individual component masses do not show up anywhere in this equation set. Thus, it is impossible to measure dynamical masses for either the primary or the secondary using just relative astrometry. That said, if the mass of the planet can be safely assumed to be negligible compared to the mass of the star, then the total mass derived from Keplerian analysis can be treated as a constraint on the dynamical mass of the primary. In practice, the reverse logic is often employed: an independent constraint on the mass of the primary (from e.g., spectroscopic analysis) is used as a prior on the total mass when the planet mass is small and can be ignored.

A second important degeneracy is between semimajor axis a, total mass $M_{\rm tot}$, and parallax π. If we just had relative astrometric measurements and no external knowledge of the system parallax, we would not be able to distinguish between a system that has larger distance and larger semimajor axis (and therefore larger total mass, assuming a fixed period) from a system that has smaller distance, smaller semimajor axis, and smaller total mass. Luckily, we live in an era where parallax measurements are excellent overall thanks to the Gaia mission (Gaia Collaboration et al., 2016), and strict priors can often be applied to parallax, breaking the degeneracy and enabling dynamical mass measurements of stars (when planet mass is negligible). However, this degeneracy is important to understand when considering the impact of potential biases in parallax or stellar mass measurements.

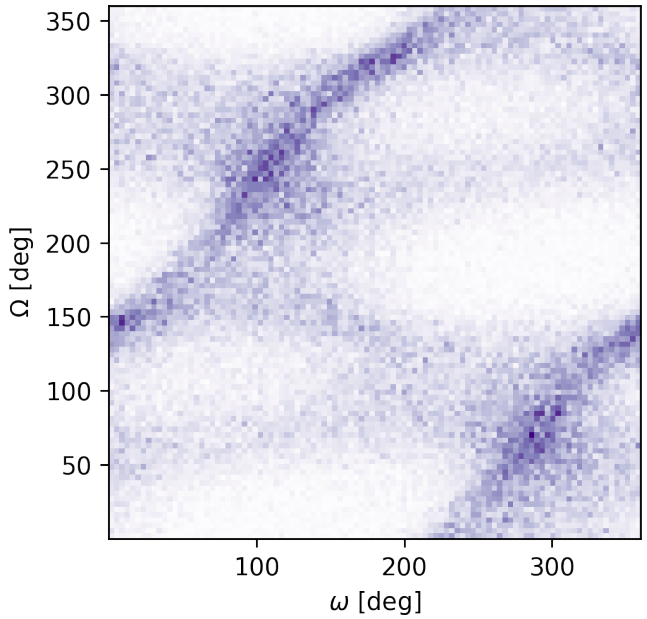

Figure 4.2: Posterior distributions over Ω and ω_p for GJ 504 b, computed via the OFTI algorithm using data from Kuzuhara et al. (2013). The 180° symmetry in both posteriors, discussed in the text, is apparent.

A final degeneracy I would like to point out concerns the argument of periastron ω_P and the position angle of nodes Ω. Equation 4.1 is invariant to the transformation:

$$\begin{aligned} \omega_p^{'} &= \omega_p + \pi \\ \Omega^{'} &= \Omega - \pi \end{aligned} \tag{4.3}$$

which creates a 180° degeneracy between particular values of ω_p and Ω, and a characteristic "double-peaked" structure in marginalized 1D posteriors of these parameters (see Figure 4.2 for an example). Physically, this degeneracy comes about because relative astrometry alone only constrains motion in the plane of the sky; an orbit tilted toward the observer, with the planet moving away from the observer has the same projection on the plane of the sky as an orbit tilted away from the observer, with the planet moving toward the observer. In practice, this degeneracy is handy, because if the ω_p/Ω posteriors do not appear identical before and after 180°, it is generally an indication that the MCMC chains are unconverged.

Although the degeneracies between individual component masses, parallax and semimajor axis, and "the omegas" are baked into the equations describing relative

astrometric measurements, the small orbital fractions over which directly imaged planets are observed introduces additional uncertainties and degeneracies. Because direct imaging today is sensitive to only planets at large separations from their stars ($\gtrsim 200mas$), directly imaged planets have orbital periods of in the best cases years, and in the worst cases thousands or tens of thousands of years (Bowler, 2016). The short orbital fractions over which we can observe these objects translates to often broad orbital posteriors which are similar to the priors. A common recurring degeneracy for incomplete orbits is between eccentricity and inclination (Blunt et al., 2019; Ferrer-Chávez, Wang, and Blunt, 2021; Chapter 3 of this thesis); in the absence of complete information about the orbital acceleration over the entire orbit, the observed data can be equally well explained by a more inclined, circular orbit, and by a more face-on, eccentric orbit. Directly imaged planets often have astrometric measurements that are change approximately linearly in time, which translates to: 1) broad posteriors, and 2) strong eccentricity-inclination degeneracies.

Because the shapes of the posteriors of directly imaged planet orbits are often multimodal, highly covariant, and non-Gaussian, many standard posterior computation tools, such as the widely used Affine-invariant MCMC sampler `emcee` (Foreman-Mackey et al., 2013) struggle to converge. This motivated the development of the OFTI algorithm, and the incorporation of other specialized sampling tools, such as MCMC with parallel tempering (Vousden, Farr, and Mandel, 2016), into `orbitize!`.

For more details about how `orbitize!` sets up likelihood, defines priors, and computes posteriors, I refer readers to Blunt et al. (2020) and Blunt et al. (2017).

4.3 Jointly Fitting Radial Velocities

Including RVs in the Model

In the `orbitize!` coordinate system, and relative to the system barycenter, the radial velocity of the *planet* due to the gravitational influence of the star is:

$$rv_p(f) = \sqrt{\frac{G}{(1-e^2)}} M_* \sin i \, (M_{\rm tot})^{-1/2} \, a^{-1/2} (\cos(\omega_p + f) + e\cos\omega_p), \quad (4.4)$$

and the radial velocity of the *star* due to the gravitational influence of the planet is:

$$rv_*(f) = \sqrt{\frac{G}{(1-e^2)}} M_p \sin i \, (M_{\rm tot})^{-1/2} \, a^{-1/2} (\cos(\omega_* + f) + e\cos\omega_*), \quad (4.5)$$

where ω_* is the argument of periastron of the star's orbit, which is equal to ω_p + 180°.

In these equations, the individual component masses m_p and m_* enter. This means radial velocity measurements break the total mass degeneracy and enable measurements of individual component masses ("dynamical" masses). Crucially, however, radial velocities of a planet do not enable dynamical mass measurements of the planet itself, but of the star[4]. RV measurements of high-contrast companions are increasingly common (e.g., Snellen et al., 2014, Petrus et al., 2021) and increasingly precise thanks to high-resolution instruments like CRIRES (Kaeufl et al., 2004) and KPIC (Pezzato et al., 2019). The RV measurements of high-contrast companions are typically made relative to the star, not the barycenter, which in fact translates to a constraint not on the star's dynamical mass, but on the total mass.

Radial velocity measurements also break the Ω/ω degeneracy discussed in the previous section, uniquely orienting the orbit in 3D space.

How it Works

`orbitize!` can perform joint fits of RV and astrometric data in two different ways, which have complementary applications. The first method was primarily implemented by Roberto Tejada Arevalo and Vighnesh Nagpal, with discussion and guidance primarily contributed by Lea Hirsch, myself, and Jason Wang. I implemented and tested the calculation of secondary RVs, and reorganized the API for computing radial velocities. This method is automatically triggered when an `orbitize!` user inputs radial velocity data. `orbitize!` automatically parses the data sets up an appropriate model, then runs the user's Bayesian computation algorithm of choice to jointly constrain all free parameters in the fit. `orbitize!` can handle both primary and secondary RVs, and fits for the appropriate dynamical masses when RVs are present; when primary RVs are included, `orbitize!` fits for the dynamical masses of secondary objects, and vice versa. Instrumental nuisance parameters (RV zeropoint offset, γ, and white noise jitter, σ) for each RV instrument are also included as additional free parameters in the fit if the user specifies different instrument names in the data file.

The second method of jointly fitting RV and astrometric data in `orbitize!` is a feature I designed and primarily implemented. Jorge Llop-Sayson became an expert

[4]This is very often misunderstood; people see "RV" and ask why we can not get a planetary dynamical mass.

in the use of this feature, wrote a tutorial (linked in the previous section) on how to use it, and added new features to the implementation. This method separates out the fitting of radial velocities and astrometry, enabling a user to fit "one at a time," and combine the results in a Bayesian framework. First, a user performs a fit to just the radial velocity data using, for example, `radvel` (but can be any radial velocity orbit-fitting code)[5]. The user then feeds the numerical posterior samples into `orbitize!` through the `orbitize.priors.KDEPrior` object. This prior creates a representation of the prior using kernel density estimation[6] (Scott, 2015), which can then be used to generate random prior samples or compute the prior probability of a sample orbit. Importantly, this prior preserves covariances between input parameters, allowing `orbitize!` to use an accurate representation of the RV posterior to constrain the fit. I often refer to this method as the "posteriors as priors" method, since posteriors output from a RV fitting code are, through KDE sampling, being applied as priors in `orbitize!`.

In implementing and testing this method, I also wrote a separate `orbitize!` module[7] which a) converts arbitrary `radvel` outputs to `orbitize!`-readable prior samples, and computes predicted separation as a function of input epoch. This method is particularly useful for proposals; the posterior positions of planets known through only RV monitoring can be predicted and plotted using only a `radvel` posterior chain.

How to Use It

Detailed tutorials for both joint fitting methods are publicly available on the `orbitize!` documentation website[8].

Validation

`orbitize!` maintains a suite of unit tests, many of which were specifically written to make sure that the features described in this section are working correctly. Specifically, interested readers are directed to: test_secondary_rvs.py, test_radvel_utils.py, test_mcmc_rv.py, test_kepler_solver.py, and test_kde_and_ndinterpolator.py.

[5] There are several advantages to fitting RVs separately using a dedicated RV orbit-fitting package, which are discussed more in Section 4.3.

[6] A truly awesome visualization of which is shown at https://mathisonian.github.io/kde/

[7] https://github.com/sblunt/orbitize/blob/main/orbitize/radvel_utils/compute_sep.py

[8] Joint fitting: https://orbitize.readthedocs.io/en/latest/tutorials/RV_MCMC_Tutorial.html; trained fitting: https://orbitize.readthedocs.io/en/latest/tutorials/Using_nonOrbitize_Posteriors_as_Priors.html

For the purposes of this document, I set up an experiment to demonstrate and validate both of the RV fitting methods in `orbitize!`, which should produce identical results.

Figure 4.3 shows a simulated dataset of relative astrometry and stellar radial velocity measurements. I generated 100 epochs of simultaneous relative astrometry and RV measurements, uniformly spaced between MJD=51550 and 52650, using the following orbital parameters: sma=50 au, ecc=0.5, inc=$\pi/4$, total mass=1.2 $M_\odot$, plx=60 mas, M_p = 0.5 $M_\odot$, $\omega = \Omega = 0$, $\tau = 0.8$. I added Gaussian noise with amplitude 2 mas and 2 km s^{-1} to each dataset, respectively. I first ran `orbitize!` using the joint constraint method, not fitting for instrumental nuisance parameters (σ and γ) for simplicity. I chose a well-constrained orbit to test the code's ability to correctly recover the input orbital parameters. Figure 4.4 shows a subset of the fitted parameter posteriors, together with the underlying true values from which the data were generated, shown as solid pink lines. To generate these results, I used the parallel tempered MCMC implementation in `orbitize!`, with 1000 walkers and 20 temperatures. I generated 1e6 total samples, discarding the first 1e3 samples as burn-in. These chains are likely not completely converged, but even this minimal run resulted in posteriors consistent with the underlying true values.

I next performed the same test using the "trained" `orbitize!` KDE prior method. I first performed an independent `radvel` fit to only the RV data, the results of which are shown in Figure 4.5. I applied uniform priors to all parameters[9] and fixed σ and γ to 0, as in the joint `orbitize!` fit. The next step is creating a KDE representation of the RV posterior. `orbitize!` does this automatically, using `scipy.stats.gaussian_kde`, but it is important to choose the bandwidth of the KDE kernel correctly. Figures 4.6 and 4.7 show the samples from the RV posterior together with samples from the KDE representation using two different bandwidths, highlighting the need to check that the chosen bandwidth allows for an accurate KDE representation of the posterior. Having determined an appropriate KDE bandwidth, I then ran `orbitize!` on just the astrometric simulated data, and recovered the posteriors shown in Figure 4.8. I used the same MCMC parameters as for the joint test: 1000 walkers and 20 temperatures. I again generated 1e6 total samples, discarding the first 1e3 samples as burn-in. Again, these chains are likely not totally converged, but the minimal run gives totally consistent results.

[9]When using this method of jointly fitting RVs + astrometry, it is important to keep track of all priors that are applied, as one could accidentally end up "doubly" applying priors. For example, if I set a log-uniform prior on K when running the `radvel` fit, then also set a log-unifom prior on the

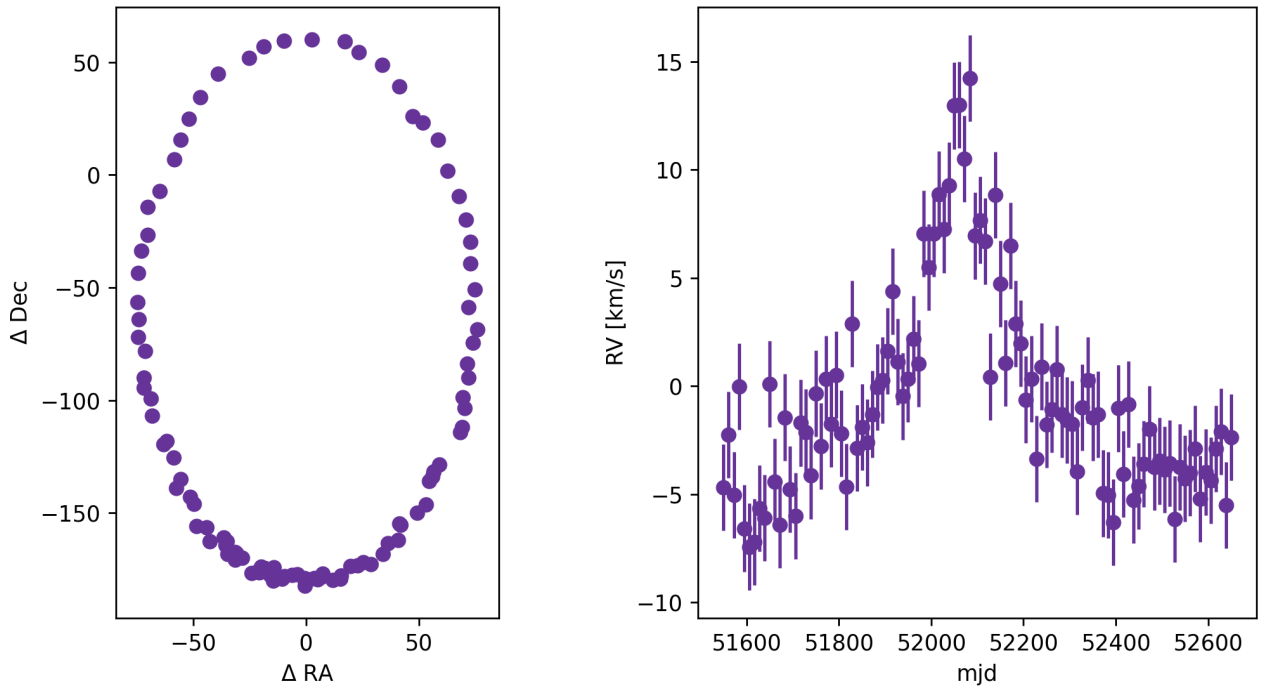

Figure 4.3: Simulated data used to test the RV+astrometry fitting capabilities of `orbitize!`. The left panel shows the simulated astrometry, and the right panel shows simulated stellar radial velocities. Data were generated assuming masses of 0.7 and 0.5 $M_{\odot}$, parallax of 60 mas, a=2au, e=0.5, inc=45 degrees, ω=Ω=0, and τ=0.8 (assuming the standard $\tau_{\rm ref}$=58849, see Section 3.4. Random Gaussian noise with σ of 2mas (for astrometry) and 2 km s^{-1} (for RVs) was (somewhat arbitrarily) chosen and applied to all data.

Applications

There are many applications of jointly fitting RVs and astrometry, and I will highlight a few here, citing relevant papers along the way. I will also discuss situations in which users might choose to use the joint fit or trained fit methods in `orbitize!`.

Jointly Fitting Planetary RVs

As outlined in this Section, jointly fitting planetary RVs does not allow for a dynamical mass measurement of the planet. However, these types of measurements can:

secondary mass when running the subsequent `orbitize!` fit, I have effectively applied two priors on secondary mass.

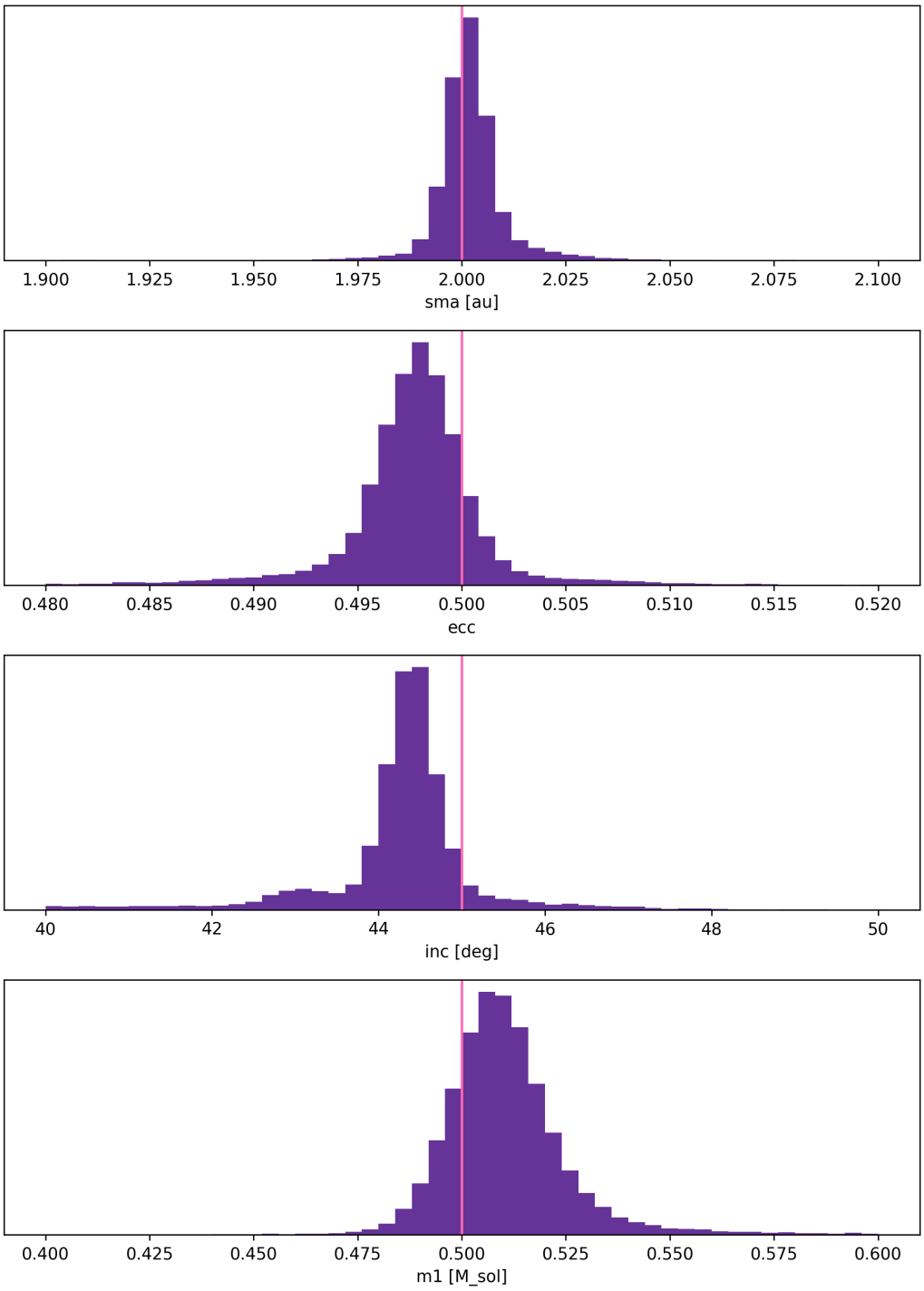

Figure 4.4: Selected marginalized posterior elements for the joint RV + astrometry `orbitize!` fit to data shown in Figure 4.3 described in the text. Purple histograms show posterior samples, and pink solid lines denote underlying truth values from which the data were generated.

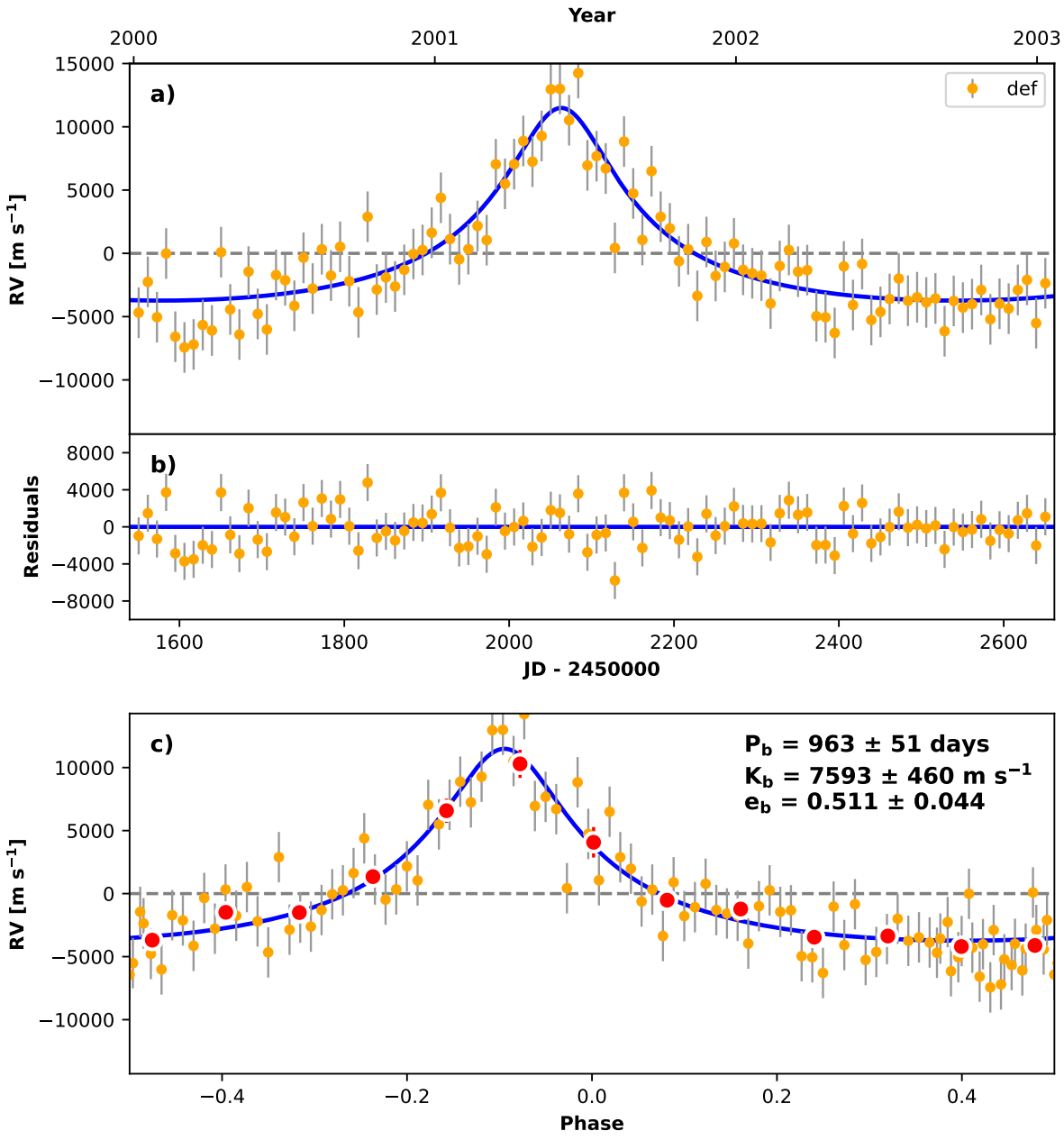

Figure 4.5: `radvel` fit to the simulated dataset shown in Figure 4.3. Top: all simulated RV data, together with the maximum a posteriori RV orbit model. Middle: residuals to the MAP fit. Bottom: phase-folded RVs to the MAP recovered period. Uncertainties in recovered orbital parameters are shown in the top right corner of this panel.

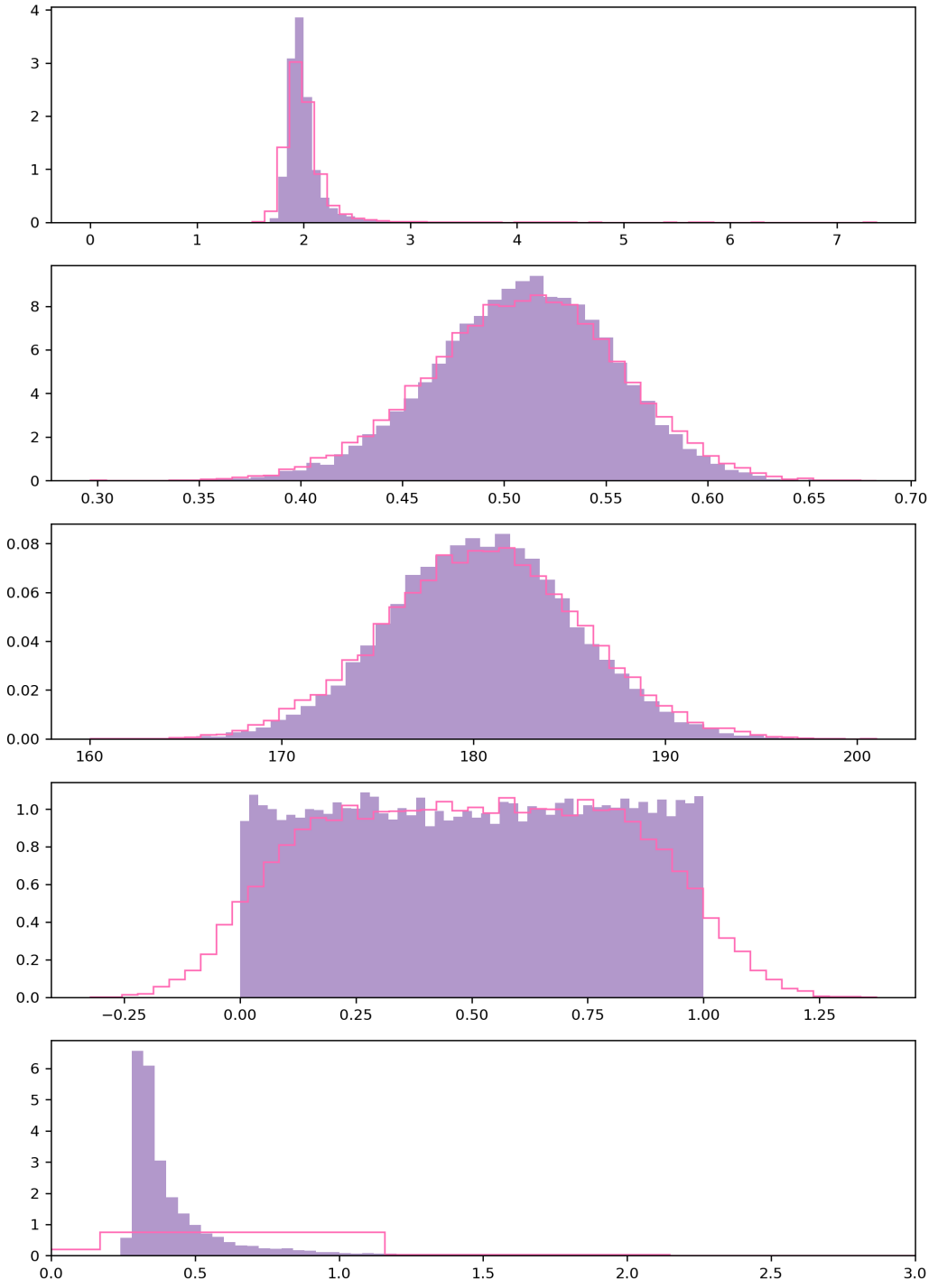

Figure 4.6: Posterior samples for the fit shown in Figure 4.5 (purple), and samples drawn from its KDE representation using the default bandwidth (pink). The default bandwidth is too wide, resulting in a KDE representation that does not capture sharp boundaries. Compare with Figure 4.7.

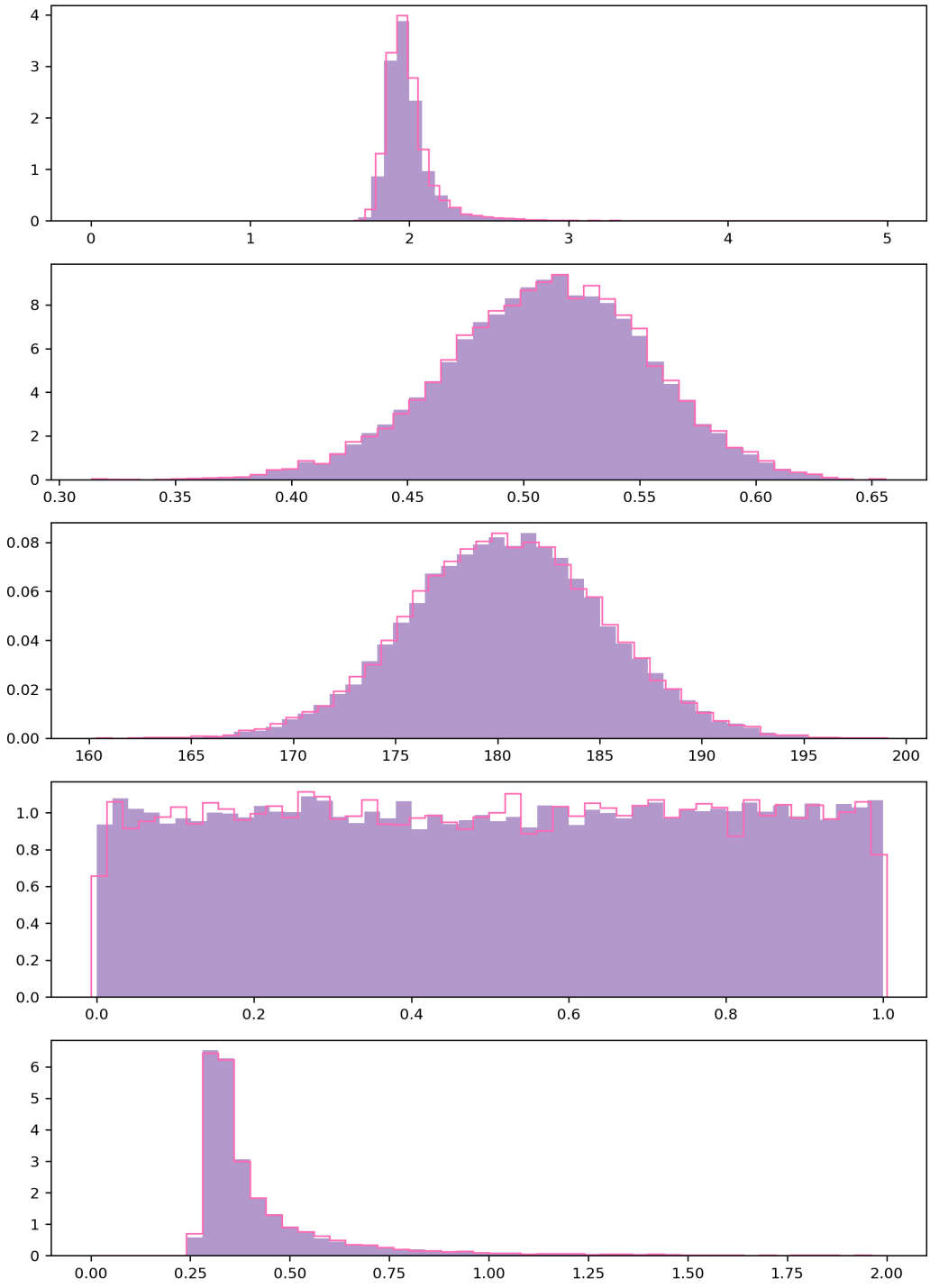

Figure 4.7: Posterior samples for the fit shown in Figure 4.5 (purple), and samples drawn from its KDE representation using a bandwidth of 0.1 (pink). This smaller bandwidth results in a better KDE representation than that shown in Figure 4.7.

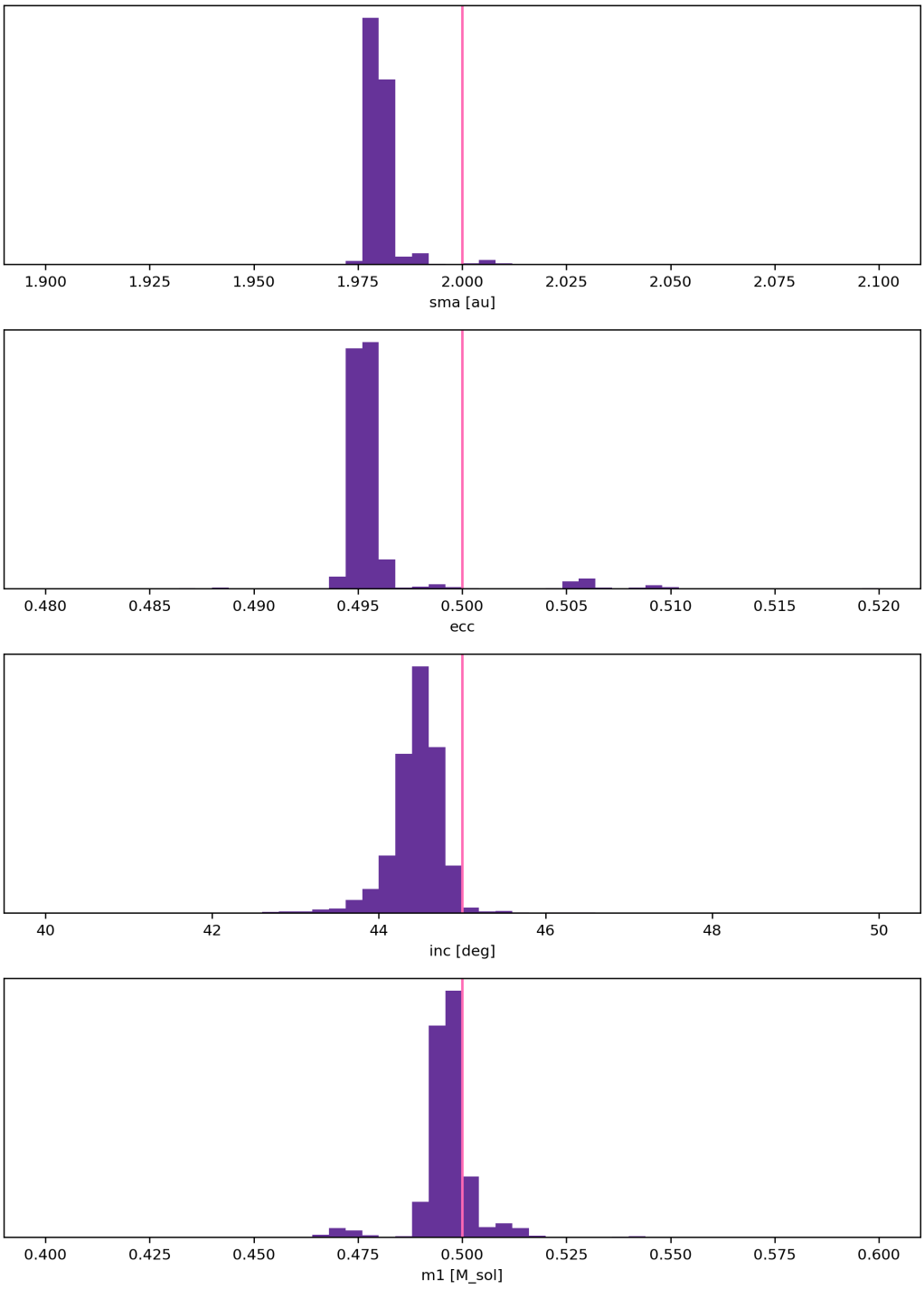

Figure 4.8: Selected marginalized posterior elements for the trained RV + astrometry `orbitize!` fit to data shown in Figure 4.3 described in the text. Purple histograms show posterior samples, and pink solid lines denote underlying truth values from which the data were generated. Compare with Figure 4.4.

a) break the Ω/ω degeneracy, uniquely orienting the orbit in 3D space[10], b) confirm that a companion is gravitationally bound to its primary, a canonical problem for companions discovered via high-contrast imaging, and c) reduce orbital parameter posterior uncertainty. In particular, planetary RV measurements have been shown to break the inclination/eccentricity degeneracy (Schwarz et al., 2016). It is also worth pointing out that planetary RV measurements are a feature of high spectral resolution studies of directly imaged companions, typically carried out to study the companions' atmospheres.

The first application is useful for obliquity constraints, which themselves are a tracer of dynamical formation and evolution (Bowler et al., 2023). Most obliquity measurements that have been made to date are not true 3D obliquities, but projected obliquities on the plane of the sky (see Bowler et al., 2023 for the direct imaging case and Rubenzahl et al., 2021 for the transit case). The limiting factor is the orientation of the stellar rotation vector in 3D space. Stellar inclination can be determined through astroseismology (Zhang et al., 2021) or a combination of spectroscopic $v\sin i$ determination and photometric monitoring, but determining the position angle of nodes (Ω) of the star's angular momentum vector has not been possible until recently. Kraus et al. (2020) presented the first fully 3D obliquity measurement of an exoplanet (i.e., orienting both the orbital angular momentum vector and the stellar rotation angular momentum vector in 3D space) by interferometrically measuring the photocenter shift of a star across the Brγ absorption line. There is interest in the community in performing these types of measurements for other stars and obtaining more 3D stellar angular momentum vector measurements, but joint planetary RV+astrometric fitting capabilities are needed to turn these stellar measurements into obliquity constraints.

The second application is more of a sanity check than an interesting scientific constraint, as generally companions being studied at high spectral resolution will already have passed several rounds of checks testing that the companion is not an unbound background star. However, it is still worth mentioning, as labeling candidate companions as background stars is difficult, and multiple rounds of checks are needed (see e.g., Nielsen et al., 2017, which showed that the imaged companion HD 131399 Ab, published in Wagner et al., 2016, was in fact a background star).

[10] Stellar RVs also have this ability, but not every primary star is amenable to RV measurement. For example, hot and rapidly rotating stars, common directly imaged planet hosts, make RV measurements difficult.

Petrus et al. (2021) used their RV measurement of the planet HIP 65426 b was consistent with orbital motion around the primary HIP 65426 A.

The third application is straightforward to understand: more data gives better constraints. However, there is some subtlety to how precise planetary RVs need to be in order to achieve improved orbital constraints. Because of the low SNR of planetary companions (both because of their intrinsic faintness and because of contaminating stellar light), it is challenging to obtain high enough spectral resolution measurements of high-contrast companions in order to derive precise RV measurements. Typical uncertainties are on the order of a few $\mathrm{km\,s^{-1}}$. However, higher spectral resolution measurements are possible, and have been made for a few planets (e.g., β Pic b, Snellen et al., 2014). I recently dPI'ed a CRIRES proposal (PI: Simon Petrus) to obtain high resolution spectra of the exoplanet HIP 65426 b, both to study the exoplanet's atmosphere at high spectral resolution, and to obtain a higher precision planetary RV measurement. Chapter 3 discusses the eccentricity of HIP 65426 b in detail, but the constraints from existing data are still inconclusive. However, the expected RVs of HIP 65426 b are significantly different assuming a zero or moderate eccentricity (see Figure 4.9) because of the eccentricity-inclination degeneracy; the same (small) amount of orbital motion in the plane of the sky can be explained as an orbit that is either less eccentric and closer to edge-on (with more orbital motion in the radial direction) or more eccentric and closer to face-on (with less orbital motion in the radial direction). RV measurements of HIP 65426 b will help us better constrain the planet's eccentricity, and therefore its formation history.

Jointly Fitting Stellar RVs

The major application of jointly fitting stellar radial velocities and astrometry is measuring dynamical masses. The joint fitting capabilities of `orbitize!` are most useful when both the RV and astrometric datasets are (individually) not very constraining. A common scenario is: a star has been RV monitored for years, and has a measured RV trend (i.e., the star's radial velocity increases linearly over time), indicative of a long-period companion whose orbit has not been resolved. Follow-up imaging is performed, and a companion is discovered, and a few epochs of relative astrometric measurements are added to the dataset. The reverse also occurs for companions discovered with imaging that are then followed up with radial velocities. This series of events most commonly plays out for massive, super-planetary mass companions (e.g., brown dwarfs and MS companions, such as HD 104304, Schnupp

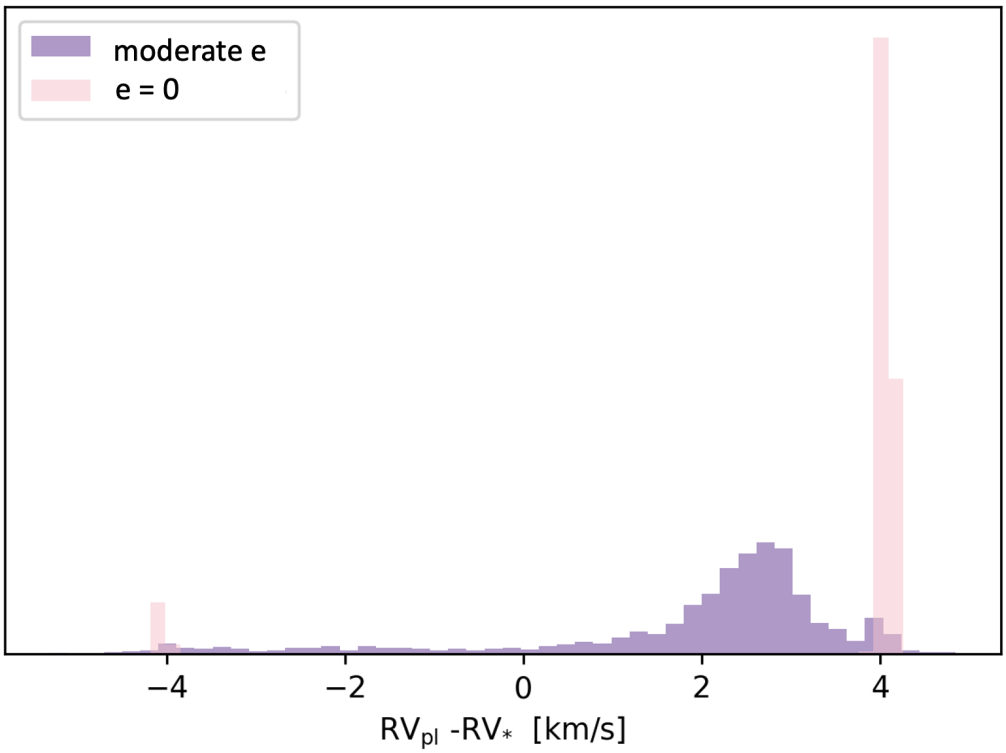

Figure 4.9: The predicted radial velocity of HIP 65426 b for each posterior sample assuming (in purple) a uniform prior on eccentricity, and (in pink) a fixed circular eccentricity. With sub-km s^{-1} RV precision, RV measurements of HIP 65426 b will enable improved eccentricity constraints.

et al., 2010, Howard and Fulton, 2016; and even compact objects, e.g., HD 159062, Hirsch et al., 2019), but also happens for planets (e.g., the discovery of β Pictoris c, Lagrange et al., 2019, Nowak et al., 2020).

The value of dynamical mass measurements of young planets has been discussed in several places in this thesis (see the Introduction and Chapter 2), but it is worth highlighting the value of precise dynamical masses for other objects as well. Many M dwarfs, in particular, have been shown to be anomalously inflated relative to model expectations (Boyajian et al., 2012, Kesseli et al., 2018, Pegues et al., 2021). Theories explaining this phenomenon range from magnetic spot activity to measurement systematics to binary effects. Stassun et al. (2011) and (Pegues et al., 2021) specifically call for more dynamical mass measurements of widely separated M dwarf companions, particularly those near the fully convective boundary, in order to make progress on this issue. This motivates my ongiong work with Vighnesh

Nagpal on the HD 104304 G-M binary, which we have been working to characterize with RVs and astrometry (Nagpal et al., 2021).

Joint fitting of the RV and astrometric datasets simultaneously is generally more efficient when fitting either dataset alone would lead to MCMC chains struggling to converge. In other words, if the orbit fraction is low in both datasets, fitting them jointly will often speed up convergence time by reducing or breaking the multi-modal parameter degeneracies (e.g. between Ω/ω) that would exist if fitting each dataset independently.

Training on Stellar RVs

Using the trained fitting method of `orbitize!` is preferable when the RV fit is complicated in its own right. For example, an RV dataset with significant stellar activity may warrant a full separate orbit fit (see the next chapter), the relevant orbital parameter posteriors of which can then be applied as priors in the astrometry fit. To illustrate the utility of this functionality, I direct readers to Llop-Sayson et al. (2021). I performed the RV fit published in this paper, and Jorge used the joint fitting capabilities of `orbitize!` to produce constraints on the orbit of ϵ Eridani. I reproduce the portion of this paper that I wrote below.

> Mawet et al. (2019) performed a thorough series of tests to evaluate the possibility that ϵ Eridani b is an artifact of stellar activity, finding that the ~7 yr orbital period is distinct from periods and harmonics of the periodicities in the S_{HK} activity indicator timeseries. Our aim is not to recapitulate their analysis, but to update their orbital solution using the additional data obtained since the paper's publication, which spans approximately half of one orbital period of ϵ Eridani b.
>
> Mawet et al. (2019) identified three peaks in a Lomb-Scargle periodogram of the RVs that rose above the 1% eFAP: one at the putative planet period of 7.3 yr, one at 2.9 yr, and one at 11d. Applying `RVSearch` (Rosenthal et al., 2021a) to the full dataset, we recover this structure of peaks. Like Mawet et al. (2019), we also identify two major periods in the S_{HK} timeseries for both HIRES and APF, which coincide with the 11 d and 2.9 yr periods in the RVs. We interpret these as the signatures of rotationally-modulated stellar activity and a long-term activity cycle, respectively. To investigate the effects of these signals

on the physical parameters of planet b, we performed two separate RV orbit-fits using `RadVel` (Fulton et al., 2018) to try different priors on the Gaussian Process (GP) timescale for modeling the stellar activity. In each of these fits, we assume a one-planet orbital solution, parameterized as $\sqrt{e_b}\cos\omega_b$, $\sqrt{e_b}\sin\omega_b$, $T_{conj,b}$, P_b, K_b. We also included RV offset (γ) and white noise (σ) parameters for each instrument in the fit, treating the four Lick velocity datasets independently to account for instrumental upgrades as in Mawet et al. (2019). Finally, we allowed an RV trend $\dot{\gamma}$.[11]

In the first fit, we included a GP noise model to account for the impact of rotationally-modulated magnetic activity on the RVs (Rajpaul et al., 2015). We used a quasi-periodic kernel, following (Mawet et al., 2019). This kernel has hyperparameters η_2, the exponential decay timescale (analagous to the lifetime of active regions on the stellar disk), η_3, the stellar rotation period, and η_4, which controls the number of local maxima in the RVs per rotation period, and η_1, the amplitude of the GP mean function, which we treated as independent for each instrument dataset. Following López-Morales et al. (2016), we fixed η_4 to 0.5, which allows approximately two local maxima per rotation period. In this first fit, we allowed η_2 and η_3 to vary in the range (0, 100d). We calculated a Markov chain representation of the posterior using `emcee` (Foreman-Mackey et al., 2013), We visually inspected the chains to ensure appropriate burn-in and production periods. In total, the chain contained 450080 samples. The resulting orbital and nuisance parameters are given in Table 4.1. Both the orbital parameters and the GP hyperparameters are well constrained; in particular, the marginalized posterior over the rotation period η_3 is Gaussian about the expected period of 11d. The data allow a trend, although the value is consistent with no trend at the 1σ level, which allows us to conclude that there is no evidence in the current data for a RV trend. One non-intuitive feature of this fit is a preference for extremely small values (10^{-6} m/s) of white noise jitter for the second Lick dataset. Even when we ran a fit requiring that all jitter values be at least 0.5 m/s, the posterior peaked at this lower bound. This may be evidence that much of the noise in this particular dataset is correlated, and therefore well-modeled by the

[11] The reference epoch for γ,$\dot{\gamma}$,$\ddot{\gamma}$ was 2457454.642028.

GP noise model. It could also indicate that the reported observational uncertainties are overestimated for this dataset. Whatever the reason, there are fewer than 10 measurements that are affected by the value of this jitter parameter, and neither the orbital parameters nor the GP timescale parameters are affected by its particular value.

To investigate the impact of the long-term activity cycle on the marginalized orbital parameter posteriors, we performed another fit identical to the one described above, except we required that η_2 and η_3 vary between 1yr and the $\sim$ 30 yr observation baseline. The marginalized posteriors for η_2 and η_3 were broad, with power across the entire allowed space, although the period parameter η_3 showed local maxima at both $\sim$ 1100 d and $\sim$ 2000 d. The marginalized 1 d posteriors for both of these parameters did not vary with those of any of the orbital parameters, allowing us to conclude that the long-term activity cycle, while somewhat present in the RVs, did not significantly affect the derived values of the orbital parameters. We therefore adopt the rotation-only GP fit described above.

Our derived orbital parameters, account for the effect of rotationally-modulated stellar noise, are very similar to those of Mawet et al. (2019) (see their Table 3). We derive an orbital period of 2671^{+17}_{-23} days, a slightly reduced median semiamplitude of 10.3 m s^{-1}, and a low median eccentricity of 0.067. We show the series of RV measurements, the residuals to the fit and the phase folded RV curve in Fig. 4.10.

The remainder of this paper uses the RV constraints derived in the section quoted above as priors for an `orbitize!` fit using upper limits on the planet location as well as Hipparcos/Gaia relative astrometry (see next section). The ability to separately treat the stellar activity greatly simplified the workflow and utility of the `orbitize!` fit.

4.4 Jointly Fitting Absolute Astrometric Data

Up to this point, I have only discussed *relative* astrometry, which is a measurement of a secondary's position with respect to the primary. However, it is also possible to measure *absolute* astrometry of individual stars by precisely tracking their sky positions relative to "fixed" background stars at large distances. Timeseries of absolute astrometry measurements enable orbit constraints. In the `orbitize!`

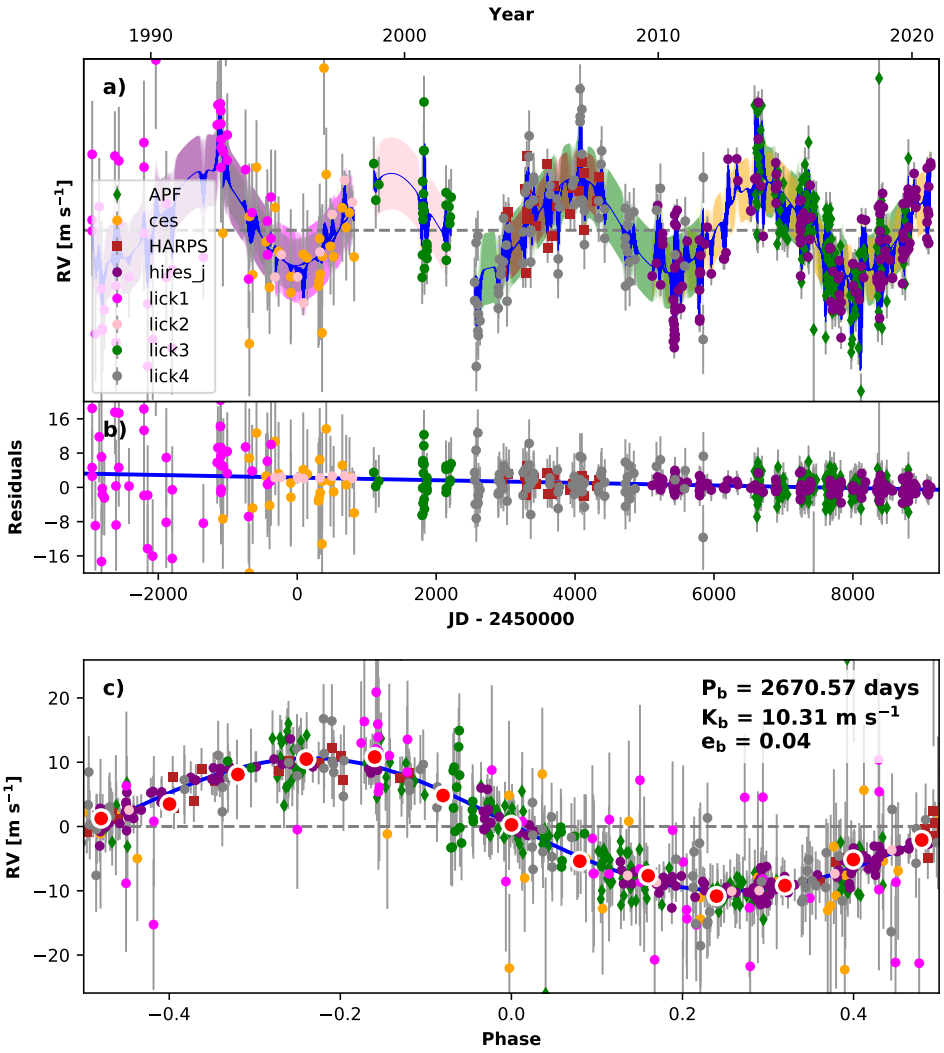

Figure 4.10: Multi-panel plot showing the data, maximum a posteriori model fit and residuals. (a) Time series of radial velocities from all data sets, (b) residuals to the RV fit, (c) phase-folded RV curve. The maximum probability one-planet model is overplotted (*blue*), as well as the binned data (*red dots*). Reproduced from Llop-Sayson et al. (2021).

Table 4.1: RV Fit MCMC Posteriors for ϵ Eridani.

Parameter	Credible Interval	Maximum Likelihood	Units
Modified MCMC Step Parameters			
P_b	2671^{+17}_{-23}	2661	days
$T\mathrm{conj}_b$	2460017^{+76}_{-32}	2460023	JD
$T\mathrm{peri}_b$	2460054^{+680}_{-690}	2460235	JD
e_b	$0.055^{+0.067}_{-0.039}$	0.046	
ω_b	$57.3^{+80.2}_{-154.7}$	2.1	$^\circ$
K_b	$10.34^{+0.95}_{-0.93}$	10.33	m s^{-1}
Other Parameters			
γ_{lick4}	$-1.0^{+2.7}_{-2.6}$	-0.9	m s^{-1}
γ_{lick3}	10.0 ± 4.9	9.9	m s^{-1}
γ_{lick2}	$7.5^{+5.6}_{-5.7}$	7.6	m s^{-1}
γ_{lick1}	$4.4^{+6.2}_{-6.1}$	4.5	m s^{-1}
$\gamma_{\mathrm{hires_j}}$	2 ± 1	2	m s^{-1}
γ_{harps}	$16442.5^{+3.2}_{-3.1}$	16442.4	m s^{-1}
γ_{ces}	$16446.6^{+5.7}_{-5.5}$	16446.6	m s^{-1}
γ_{apf}	-0.9 ± 1.3	-0.9	m s^{-1}
$\dot{\gamma}$	$-0.00026^{+0.00063}_{-0.0006}$	-0.00026	m s^{-1} d^{-1}
$\ddot{\gamma}$	$\equiv 0.0$	$\equiv 0.0$	m s^{-1} d^{-2}
σ_{lick4}	$5.17^{+1.1}_{-0.95}$	5	m s^{-1}
σ_{lick3}	$5.6^{+2.1}_{-2.3}$	5.3	m s^{-1}
σ_{lick2}	$2.3^{+3.6}_{-1.4}$	0.5	m s^{-1}
σ_{lick1}	$7.2^{+3.5}_{-3.7}$	7.4	m s^{-1}
$\sigma_{\mathrm{hires_j}}$	$2.36^{+0.46}_{-0.41}$	2.26	m s^{-1}
σ_{harps}	$4.8^{+2.2}_{-2.7}$	4.3	m s^{-1}
σ_{ces}	$8.1^{+3.8}_{-4.5}$	7.3	m s^{-1}
σ_{apf}	$3.64^{+0.62}_{-0.55}$	3.51	m s^{-1}
$\eta_{1,\mathrm{apf}}$	$7.82^{+0.97}_{-0.92}$	7.71	m s^{-1}
$\eta_{1,\mathrm{hires_j}}$	$6.92^{+0.64}_{-0.59}$	6.78	m s^{-1}
$\eta_{1,\mathrm{ces}}$	$7.5^{+4.1}_{-4.7}$	7.0	m s^{-1}
$\eta_{1,\mathrm{harps}}$	$5.7^{+2.3}_{-3.0}$	5.3	m s^{-1}
$\eta_{1,\mathrm{lick,fischer}}$	$8.7^{+1.3}_{-1.4}$	8.0	m s^{-1}
η_2	$37.6^{+6.4}_{-5.4}$	36.4	days
η_3	$11.68^{+0.14}_{-0.13}$	11.66	days
η_4	$\equiv 0.5$	$\equiv 0.5$	

coordinate system discussed in the previous subsections of this chapter, the absolute offset in stellar position due to orbital motion of a planet is:

$$\Delta\text{R.A.}_* = -\frac{M_p}{M_{\text{tot}}}\pi a(1 - e\cos\text{E})\left[\cos^2\frac{i}{2}\sin\left(f + \omega_\text{p} + \Omega\right) - \sin^2\frac{i}{2}\sin\left(f + \omega_\text{p} - \Omega\right)\right]$$
$$\Delta\text{decl.}_* = -\frac{M_p}{M_{\text{tot}}}\pi a(1 - e\cos\text{E})\left[\cos^2\frac{i}{2}\cos\left(f + \omega_\text{p} + \Omega\right) + \sin^2\frac{i}{2}\cos\left(f + \omega_\text{p} - \Omega\right)\right] \tag{4.6}$$

which is equal to $-\frac{M_p}{M_{\text{tot}}}$ multiplied by the relative astrometry measurements. As with the introduction of RV measurements, adding absolute astrometry enables dynamical mass measurements of the planet when combined with relative astrometry measurements, motivating joint fits. (Alone, absolute astrometry constrains the quantity $M_p {M_{\text{tot}}}^{2/3}$.)

Fitting absolute astrometry is complicated for many reasons. A (relatively) straightforward complication is that absolute stellar position timeseries fits must also fit for the star's parallactic ellipse and proper motion. In practice, these parameters can be treated as free parameters in an orbit fit, and jointly constrained along with an orbit model.

A messier complication is handling the available absolute astrometry data. In the remainder of this section, I briefly recap the method of Nielsen et al. (2019), describe its implementation in `orbitize!` (code that I designed, wrote, and tested), and point to a second method whose implementation was led by Jason Wang.

The goal of the Nielsen et al. method is to use absolute astrometric measurements from Hipparcos (van Leeuwen, 2007), a space satellite from the 1990s, along with a single astrometric measurement from Gaia, to jointly constrain an orbit model. The Hipparcos + Gaia combination of absolute astrometry data has been used to great success. In general, Hipparcos and Gaia can both be used separately to measure proper motion, and the baseline between them gives a change in proper motion over time, or acceleration (presumably due to orbital motion). The 20 yr baseline between the two missions makes the combination especially sensitive to longer-period companions (e.g., Xuan and Wyatt, 2020, Lubin et al., 2022). Methods like this, using Gaia data in any form, are "stop-gaps" until individual astrometric measurements from Gaia are made public; the reported Gaia astrometric quantities available today are fitted quantities, but we do not have access to the underlying data

or model parameters. This is not true for Hipparcos, for which we have individual astrometric measurements.

The first issue Nielsen et al. (and Snellan et al., before them) had to contend with was reconstructing the timeseries Hipparcos measurements. The individual star Hipparcos intermediate astrometric data (IAD), computed in an impressive full-survey reduction by van Leeuwen (2007), are not publicly available as individual measurements, but as residuals to astrometric model fits. Equations 1-4 in Nielsen et al. (2019) use these residuals and the reported model parameters to reconstruct the underlying data points. Some of the residuals published in van Leeuwen (2007) were corrupted (Brandt et al., 2021), which Nielsen et al. addressed by proposing that a user "re-fit" the IAD using the best-fit model parameters reported by van Leeuwen (2007) (a sanity check I have taken to calling "the Nielsen test"). Results of the Nielsen test (performed with `orbitize!`[12]) for β Pictoris are shown in Figure 4.11. A nominal Nielsen test fit has five parameters: parallax, proper motion (in RA and decl), and RA/decl offsets from the Hipparcos reported positions. Given a file for a particular star containing Hipparcos residuals and model fit parameters, `orbitize!` can automatically perform a Nielsen test to establish that the IAD are appropriate to use before being incorporated into an orbit fit.

The math in Nielsen et al. (2019) is clearly laid out, so I will not rehash it here except to point out that Equation 10 contains a typo (which I've also made Eric aware of). The corrected equation is:

$$f = \left(G\sqrt{\frac{2}{9D}} + 1 - \left(\frac{2}{9D}\right) \right)^{3/2} . \tag{4.7}$$

Having reconstructed the Hipparcos IAD, we can proceed with orbit fitting. If directed by a user to incorporate Hipparcos IAD, `orbitize!` automatically sets up and performs an orbit fit jointly constraining the astrometric and orbital parameters, using `astropy.coordinates` to forward-model the parallactic ellipse. Code for the `orbitize!` Hipparcos IAD fitting implementation is available at https://github.com/sblunt/orbitize/blob/main/orbitize/hipparcos.py[13] Unit tests relevant to this module include: test_hipparcos.py and test_abs_astrometry.py. A pedagogical tutorial is hosted at https://orbitize.readthedocs.io/en/latest/tutorials/Hipparcos_IAD.html.

[12] The source code is available at https://orbitize.readthedocs.io/en/latest/tutorials/Hipparcos_IAD.html

[13] I encourage you to look at this code! I'm proud of the way I wrote this module.

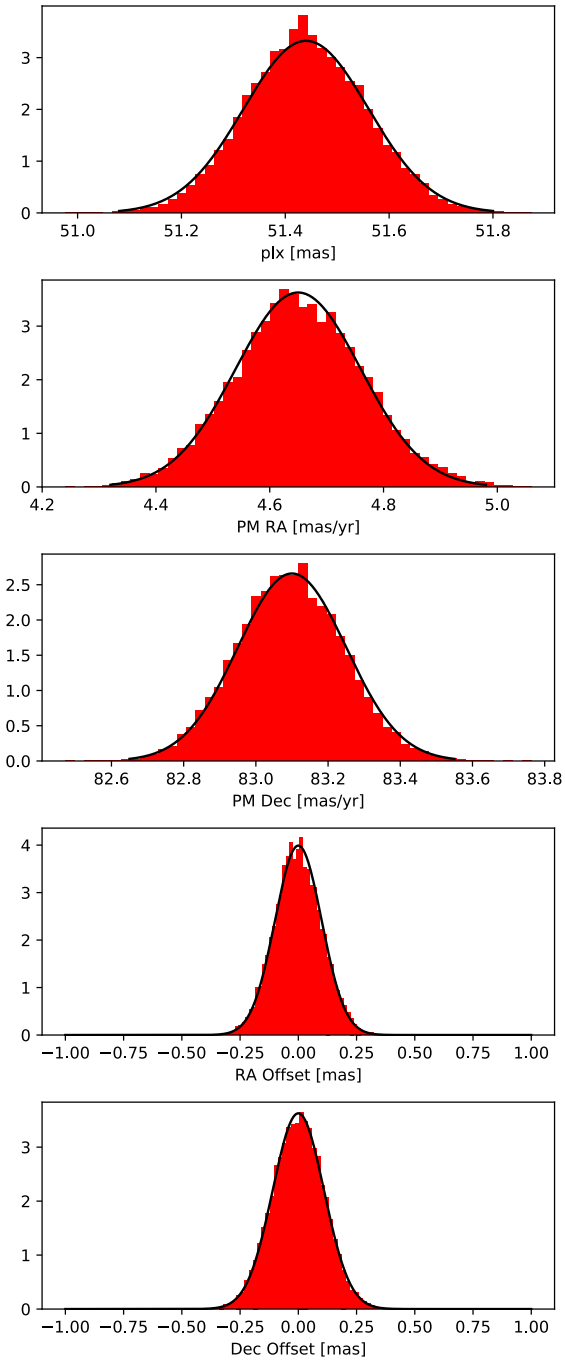

Figure 4.11: Results of the Nielsen test for β Pic. Using the Hipparcos IAD measurements, reconstructed from the best-fit model parameters and residuals reported in van Leeuwen (2007), posteriors over the 5-parameter astrometric model are re-fit and reported in red. The expected outcome, based on the van Leeuwen reported parameters and uncertainties, is shown in black. The excellent agreement indicates that the reconstructed IAD are suitable for orbit-fitting.

In order to validate my implementation, I set up and repeated "case 3" in Nielsen et al. (2019), a single-planet β Pictoris b fit including relative astrometry, Hipparcos and Gaia absolute astrometry, and a planetary radial velocity measurement (it goes without saying, hopefully, that the fit in Eric's paper was performed using independent, closed-source software). This test is also included in the main `orbitize!` code base[14]. The result is shown in Figure 4.12. Even though our fit uses the eDR3 Gaia position, and the Nielsen et al. test uses the DR2 Gaia position, the results are consistent well within 1-σ: for example we recover a planetary dynamical mass of $13.7^{+6.1}_{-3.8}$ M_J, whereas Nielsen et al. recover $12.8^{+5.5}_{-3.2}$ M_J.

A parallel treatment of fitting Hipparcos and Gaia data is developed in Brandt (2018), and has been implemented in `orbitize!` by Jason Wang. The major differences between this method and the Nielsen et al. method described above are: 1) the Brandt et al. method does not use the individual IAD Hipparcos measurements, but instead uses three proper motion vectors computed from: 1) the reported Gaia proper motion, 2) the change in reported Gaia and Hipparcos astrometric positions, divided by the time baseline, and 3) the proper motion over the Hipparcos baseline, which itself is computed from a linear combination of two separate Hipparcos reductions with inflated errors. An important caveat to this method is that it does not fit for the parallax or proper motion, but assumes these values from the reported Gaia measurements. It therefore relies on the accuracy of these measurements.

Published work using `orbitize!` has used both methods to constrain exoplanet and brown dwarf dynamical masses (e.g., Brandt method: Hinkley et al., 2023; Nielsen method: Nowak et al., 2020).

[14] orbitize/tests/end-to-end-tests/betaPic_hipIAD.py

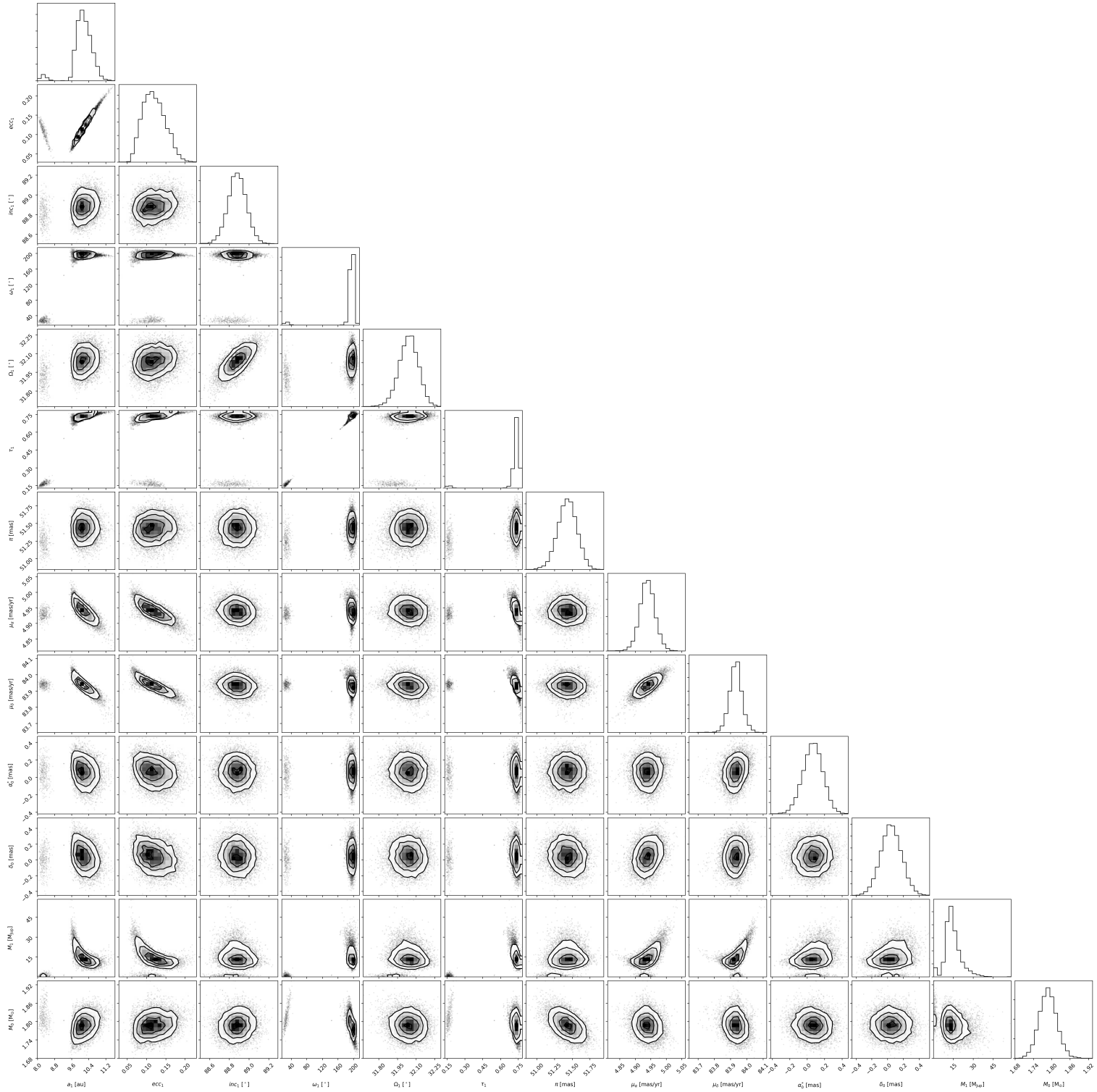

Figure 4.12: Fit results for "case 3" of Nielsen et al. (2019), a single-planet β Pictoris b fit including relative astrometry, Hipparcos + Gaia absolute astrometry, and a planetary radial velocity measurement. Compare with Nielsen et al. (2019) Figure 7. The results appear identical, validating the implementation.

Chapter 5

A NEW-AND-IMPROVED GAUSSIAN PROCESS REGRESSION MODULE FOR RADVEL

Parts of this chapter will be submitted to the Journal of Open-Source Software. *The ideas about wavelength dependence of stellar activity signals were developed in conversation with Jared Siegel. A new version of* `radvel` *will be released soon that implements the model described here.*

5.1 `radvel`

`radvel` is open-source software written in Python and C, primarily developed by BJ Fulton, Erik Petigura, and myself. Like `orbitize!`, it is an orbit-fitting code, although its intended audience is the radial velocity planet detection community, and therefore its methodology is different than that of `orbitize!`. `radvel` is quite flexible; it has an object-oriented model fitting interface (Figure 5.1) that allows users to customize the orbital basis used to fit a set of RVs, the priors applied (including the ability to fix or free individual model parameters when performing a fit), and the number of Keplerian signals that make up the model. `radvel` users primarily deal with RV timeseries with many potentially low SNR planet signals whose orbital periods have been independently determined by, for example, some flavor of periodic signal analysis (Hara et al., 2017). `radvel` posteriors are typically Gaussian, with low covariance between individual parameters, so convergence speed is not a priority, and most posterior-computation techniques work just fine (a fundamentally different situation than the typical `orbitize!` use case of a high-SNR, low orbit-fraction, single-planet signal).

As of Version 1.1.0[1], `radvel` has added Gaussian Process capabilities for jointly modeling correlated stellar noise and planetary Doppler signals. This capability, and a more recent update, are described in detail in the rest of this chapter.

5.2 Motivation & Gaussian Process Model Description

Separating out the impacts of stellar activity and planetary signals on radial velocity (RV) measurements is widely accepted to be the most important problem in radial velocity exoplanet science today (e.g., Collier Cameron, 2018), and one that will

[1] https://github.com/California-Planet-Search/radvel/releases/tag/v1.1.0

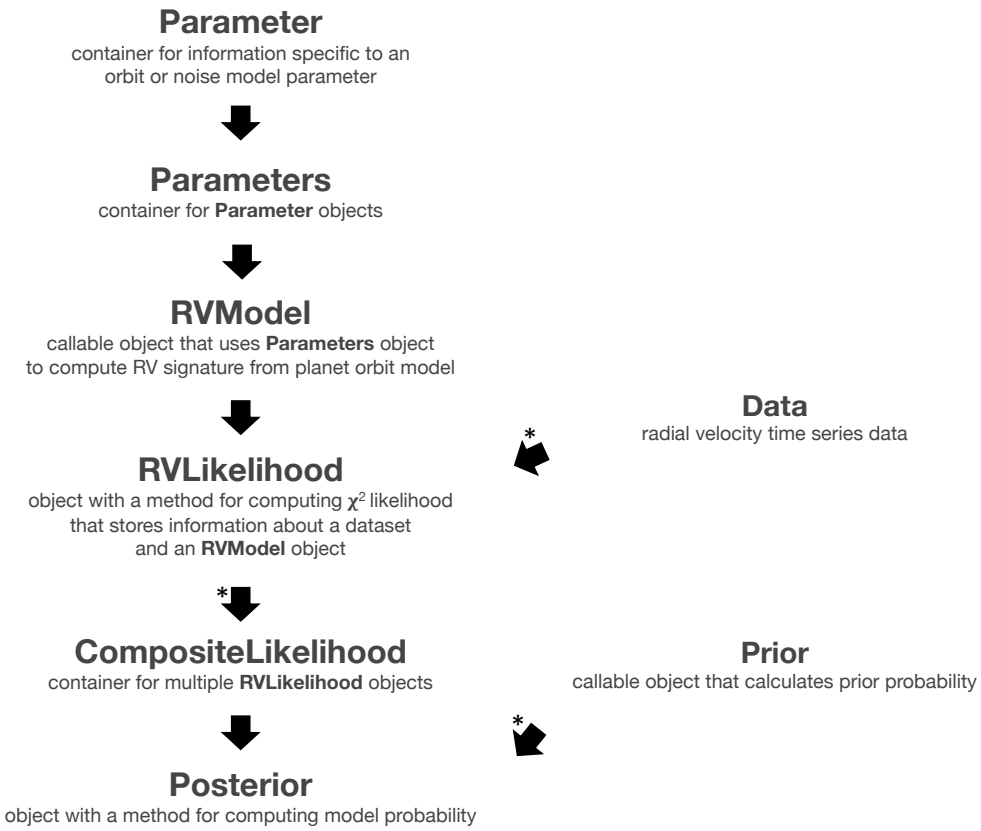

Figure 5.1: Class diagram for the RadVel package showing the relationships between the various objects contained within the RadVel package. Arrows point from attributes to their container objects (i.e. a radvel.Parameter object is an attribute of a radvel.Parameters object). Pertinent characteristics are summarized beneath each object. An asterisk next to an arrow indicates that the container object typically contains multiple attribute objects of the indicated type. This figure was constructed by myself, and was originally published in Fulton et al. (2017).

be critical to solve in order to discover and characterize Earth-like planets around Sun-like stars. Deriving an exoplanet-induced Doppler shift from a star's spectrum relies on our ability to either simultaneously model or remove Doppler shifts induced by stellar surface processes.

Stellar activity processes can be orders of magnitude greater amplitude than planetary signals (e.g., on the Sun, activity is at the level of 1 $\mathrm{m\,s^{-1}}$, while the RV shift induced by the Earth is 10 $\mathrm{cm\,s^{-1}}$). In addition, they occur across timescales spanning orders of magnitude. On timescales of seconds to minutes, stars jiggle due to stochastically excited sound waves propagating through their interiors (p-mode asteroseismic oscillations). The asymmetric bulk motion of the stellar surface toward

or away from the observer introduces Doppler shifts. On timescales of minutes, stellar surface granulation due to asymmetries in photosphere convection affects the RV measurement. Both of these effects are typically mitigated by lengthening exposure times (or taking a series of exposures, for bright stars), effectively averaging over p-mode oscillations and granulation noise (Dumusque et al., 2011). This strategy will likely need to be revised as we enter the era of extremely precise radial velocity (EPRV) instruments like NEID with instrumental precision of tens of $\mathrm{cm\,s^{-1}}$, but this discussion is beyond the scope of this chapter. On timescales of the stellar rotation period ($\sim$ 1 day, for the fastest rotators, to hundreds of days for the slowest), active regions dominate the activity signal. On the sun, we observe clusters of dark magnetized sunspots covering generally a few percent of the solar surface, which contribute to the RV signal by blocking first blueshifted then redshifted photosphere as they rotate across the solar surface. We also observe brighter magnetized regions (faculae and plage) that cover a larger fraction of the solar surface and primarily contribute to RV signal by suppressing convection through magnetic pressure. The quiet sun has a net blueshift due to convection (upwelling plasma is hotter, and therefore brighter, than sinking plasma), so convection suppression results in a net redshift. Because the pattern of faculae and plage across the solar surface is asymmetric, this convective blueshift suppression is also variable on the timescale of stellar rotation (Aigrain, Pont, and Zucker, 2012).

When RV astronomers say "stellar activity" we are typically referring to the impacts of rotationally-modulated active regions (both bright faculae/plage and dark spots) on RV signals. These active regions are not static; sunspots evolve over timescales of one to a few rotation periods, and the overall *occurrence* of spots and plage is known to vary on an 11-year "solar cycle" on the sun, which is understood as magnetic energy oscillating between storage in poloidal and toroidal components (Muñoz-Jaramillo, Nandy, and Martens, 2009).[2]. RV baselines are typically too short to observe these long-period cycles on other stars, but surveys with the longest baselines are beginning to be able to probe magnetic activity cycles (Rosenthal et al., 2021b).

Modeling rotationally-modulated stellar activity is difficult. Stellar surfaces are complex, with brightness patterns that vary not only as a function of 2D location on the stellar surface but also as a function of wavelength of observation. Compounding the problem of this large parameter space is the difficulty of obtaining RV data with

[2] A truly fantastic visualization of this dynamo can be found at https://svs.gsfc.nasa.gov/3521.

sufficient cadence to understand the activity. Even a single stellar spectrum is expensive, and weather, time allocation committees, and cost can interfere with the construction of a high quality and high cadence dataset.

Gaussian process regression (GPR) was originally introduced as a "stop-gap" measure for jointly modeling planetary signals and stellar activity while more physically-motivated stellar activity models were developed, but is widely used today to jointly model stellar activity and planetary signals. Personally, I believe that GPR will be an important tool in the hunt for Earth 2.0, and at the same time that the current models are quite flexible, and require more physical constraints. The following description deepens and builds on the material in Section 2.6 in order to give the reader some background in this important statistical technique, then describes the GPR model that I constructed and implemented as part of my thesis. The background GPR material in this section draws heavily from Rasmussen and Williams (2006), Foreman-Mackey (2018), and Aigrain and Foreman-Mackey (2022).

A staple model-fitting prescription in astrophysics is Bayesian χ^2 regression. Given a model prediction, $\mathbf{m}$ which uses model parameters θ, data $\mathbf{y}$, and observational uncertainties σ, the model log-likelihood is:

$$\mathcal{L}(\boldsymbol{\theta}) \equiv p(\mathbf{y} \mid \boldsymbol{\theta}) = \prod_{i=1}^{N} \mathcal{L}_i = \prod_{i=1}^{N} \left\{ \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{(y_i - m_i)^2}{2\sigma_i^2} \right] \right\} \tag{5.1}$$

$$= \frac{1}{\sqrt{2\pi|\mathbf{K}|}} \exp\left(-\frac{1}{2} (\mathbf{y} - \mathbf{m})^{\mathrm{T}} \mathbf{K}^{-1} (\mathbf{y} - \mathbf{m}) \right) \tag{5.2}$$

where the second line is given in an equivalent matrix formulation. $\mathbf{K}$ is called the covariance matrix, and in the standard χ^2 model fitting formulation, all of its off-diagonal elements are zero, while the diagonal elements are σ_i^2:

$$K_{ij} = \delta_{ij} \sigma_i^2. \tag{5.3}$$

In words, the χ^2 likelihood is the probability of drawing the N existing data points from an N-dimensional Gaussian model centered on the model prediction vector, where the width of the Gaussian in the ith dimension corresponds to the observational uncertainty on the ith datapoint.

So far, I have restricted the discussion to χ^2 likelihood, but having recast this more familiar model-fitting problem into the equivalent problem of drawing from an N-dimensional Gaussian, the jump to Gaussian Process regression (GPR) is not so

extreme. The first tenet of GPR is removing the assumption that the off-diagonal elements in the covariance matrix **K** must be zero. This allows datapoints to have nonzero correlations in the likelihood model, and therefore for the likelihood to describe "correlated noise." Removal of this assumption is not unique to GPR; for example, a dataset might have empirically computed covariances between datapoints that can be incorporated into the model using the same equation above. The special sauce of Gaussian Process regression is modeling this covariance between datapoints functionally. This is often called "nonparametric regression," but it might be more accurately referred to as "semi-parametric" regression. Rather than forward-modeling **m**, one forward-models the covariance structure by specifying its functional form.

GPR is used widely in the machine learning community to model datasets with correlated noise. Even if the modeler does not know the functional form of the process generating the correlated noise, they can use GPR to model it in a data-driven way. Assumptions or prior knowledge about the correlated noise can also be incorporated into the model in a Bayesian framework, making it attractive to astronomers; for example, the timescale over which the noise is coherent can be assumed, and fixed to a certain value, or included as a free model parameter, in which case a Bayesian prior can be applied. These properties make GPR a powerful and flexible tool for pattern-fitting.

In order to model the rotationally-modulated activity signal of an active star, we can construct the following GPR model, using Equation 5.2 to compute likelihood. We set **m**=0, effectively assuming that all of the stellar signal can be treated as correlated noise. To model planetary signals, **m** becomes the Keplerian prediction.

$$\begin{aligned} \mathbf{m} &= 0 \\ K_{ij} &= k\left(t_i, t_j, \phi\right) + \delta_{ij}\sigma_i^2 \\ k\left(t_i, t_j, \phi\right) &= \eta_1^2 \cdot \exp\left[-\frac{(t-t')^2}{\eta_2^2} - \frac{\sin^2\left(\frac{\pi(t-t')}{\eta_4}\right)}{\eta_3^2}\right] \end{aligned} \tag{5.4}$$

ϕ represents the Gaussian process hyperparameters (parameters of the kernel function used to construct the covariance matrix). For this kernel, the "quasi-periodic" kernel, ϕ is a vector consisting of η_1, η_2, η_3, and η_4. η_1 is the amplitude of the Gaussian process, analogous to the amplitude of the correlated deviation from RV=0

due to stellar activity. η_2 is the characteristic "lifetime" of an active region, or in other words how long a given RV signal due to stellar activity will be coherent for. η_4 is the rotation period, and η_3, which is sometimes called the harmonic complexity (Nicholson and Aigrain, 2022) or periodic lengthscale, and which I like to call the shape parameter, requires a bit more explanation. Figure 5.2 shows the impact of this parameter on the best-fit GPR models of a noisy dataset. Importantly, for this experiment, the active region lifetime, η_2, is set to infinity; in other words, this is a strictly periodic kernel, and the model is unchanging over time. The impact of η_3 is to allow more or fewer "wiggles" in the correlated noise model, which must then repeat exactly after each period. Small values of η_3 are the most wiggly, and larger values less so. For very large values of η_3 (e.g., η_3=10, bottom panel of the figure), the mean model prediction is flattened out relative to the true amplitude of the generating function. Figures 5.3 and 5.4 show comparable experiments illustrating the impacts of η_2 and η_4.

From scrutinizing the quasi-periodic kernel function (Eq 5.4), we can observe where physical constraints can be levied. η_4, the rotation period, is most straightforward to understand. A common strategy is to monitor a star one wishes to model with photometry, determine the rotation period from photometric variations (there is much discussion about the best way to do this– autocorrelation, e.g., Holcomb et al., 2022; periodogram analysis; GPR, etc), then set an informed prior on this parameter when modeling RVs. Another widely used method is to jointly model RVs and photometry or simultaneous activity indices, allowing all of these data, which in principle share a single rotation period, to jointly constrain this parameter (e.g., Suárez Mascareño et al., 2021).

A less common strategy for introducing physical constraints is to set a prior on η_2, the characteristic active region lifetime, based on either observations of the specific star of interest and/or the theoretically expected or observed distribution from a population of similar objects (Basri and Shah, 2020; Giles, Collier Cameron, and Haywood, 2017). Approaches vary on what to do about η_3, and here I will contribute only a warning: the shape parameter of the same activity signal need not be the same in photometry and RVs. This is apparent from looking at, e.g., Figure 2.7, which shows contemporaneous RVs and photometry of V1298 Tau phase-folded to the period of the largest-amplitude rotation signal. The photometric and RV data have different shapes. More work is needed in general to sufficiently model the relationship between photometry and RV activity signals, and for now I will content

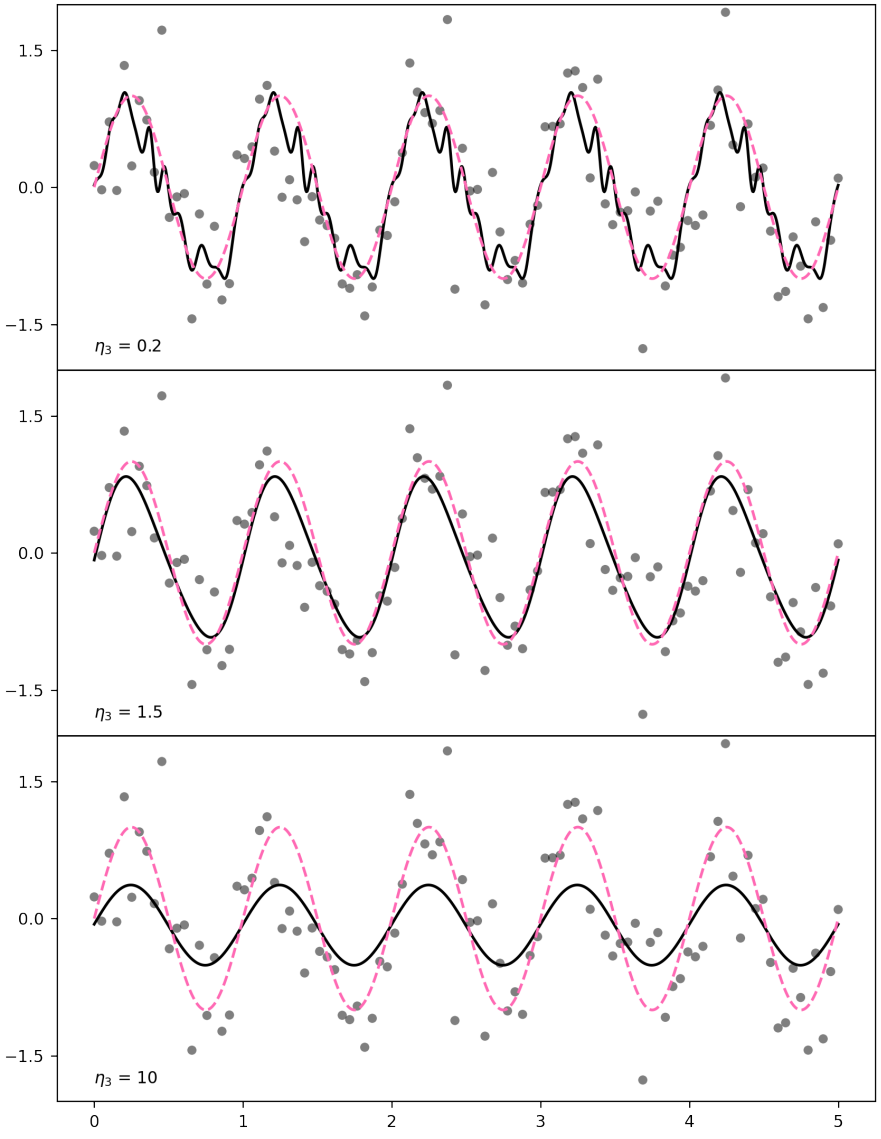

Figure 5.2: The impact of η_3 on the best-fit model for a noisy dataset drawn from a sinusoid. Each panel shows the same noisy dataset (grey points) drawn from a true underlying model (pink dashed curve). The black curve shows the best-fit GPR model prediction, given a fixed value of the shape parameter η_3 (the value of which is given in the bottom left corner of each panel).

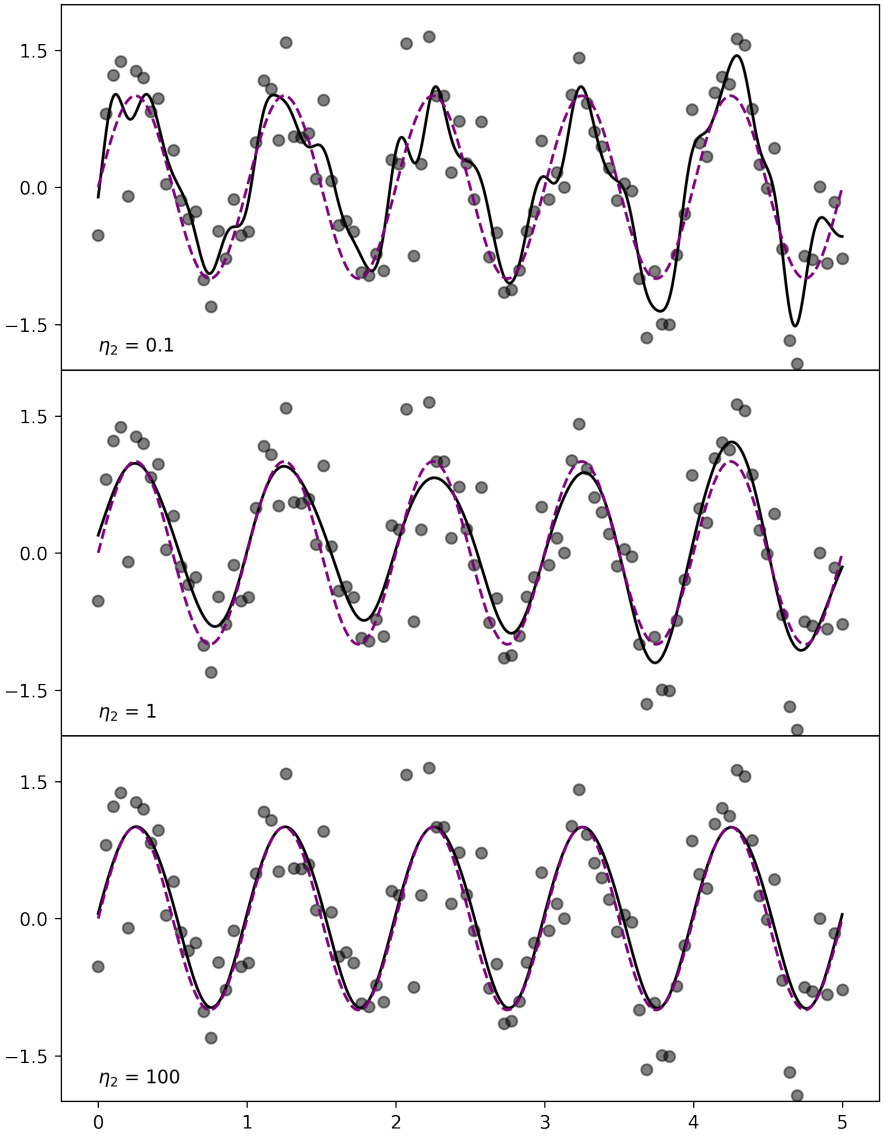

Figure 5.3: The impact of η_2 on the best-fit model for a noisy dataset drawn from a sinusoid. Compare with Figure 5.2; the setup is the same, except in this figure, the shape parameter η_3 is held constant, while the exponential decay timescale η_2 is varied. I am sure it is apparent why these parameters are often a) confusing and confused with one another, and b) degenerate in model fits. η_2 controls the characteristic amount of time over which the periodic pattern evolves.

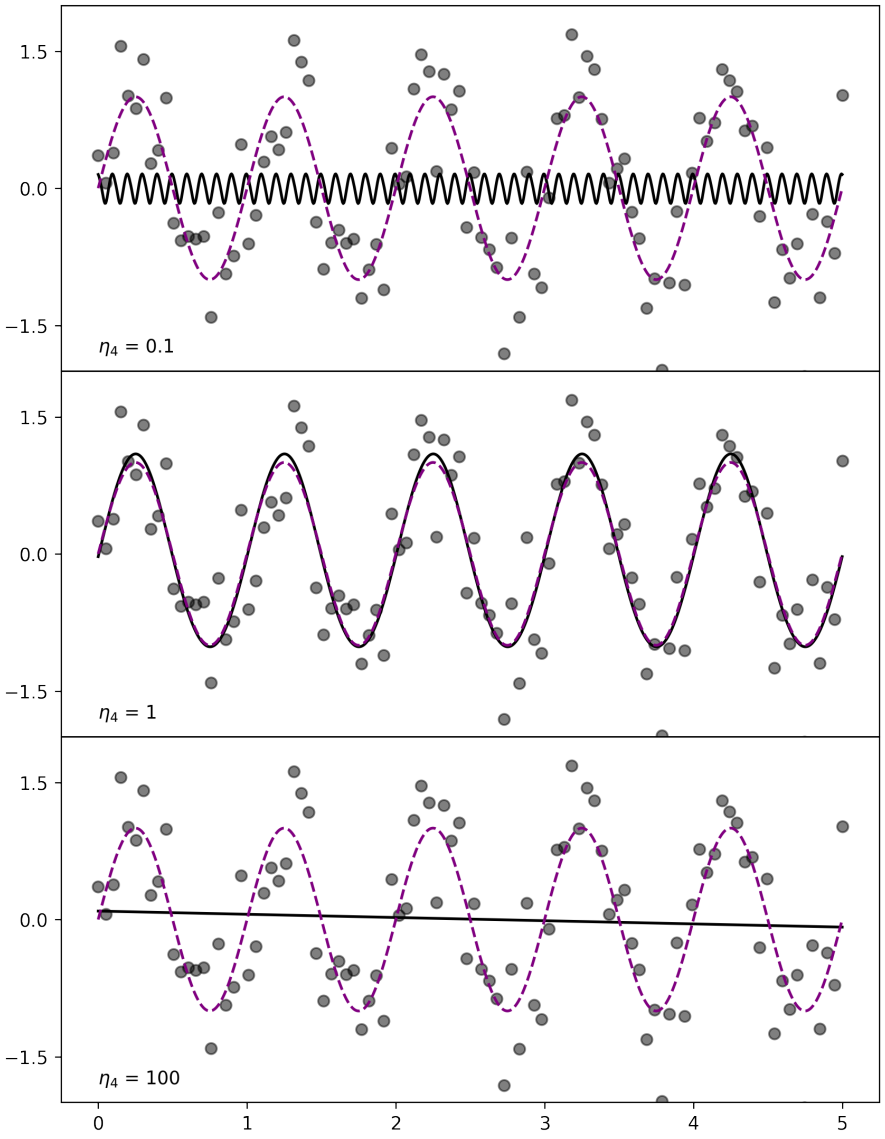

Figure 5.4: The impact of η_4, the period, on the best-fit model for a noisy dataset drawn from a sinusoid. Compare with Figures 5.2 and 5.3. To make this figure, η_2, the exponential decay timescale, has been set to ∞ (i.e., same setup for Figure 5.2). η_4 controls the periodic timescale over which the GP pattern must repeat exactly.

myself with just the warning that the assumption of the same η_3 for both data types may not be valid.

Having explained in general how GPR is applied to model rotationally-modulated stellar activity signals in RVs, I will now explain an observation and subsequent model modification I made during my PhD. This model modification was proposed in parallel by Cale et al. (2021). This description builds on the description in Section 2.4.

The insight that motivated the modification I made relates to the way RV datasets from different instruments are treated in the "classical" GPR framework.[3] When applying a parametric model to a dataset, it is sufficient to compute the likelihood for each dataset separately, then multiply them together. To make this concrete, we can imagine fitting a sine curve to some data taken by two independent instruments. We might generate model predictions for a proposed parameter vector (period, say), compute a likelihood for each individual datapoint, then multiply them together to obtain the overall likelihood using Equation 5.1. In practice, astronomers will define a function that computes log(likelihood) given a set of data, then add these values together in loop applied to several sets of data to compute a total likelihood. This choice is often not specified; for example Suárez Mascareño et al. (2021) performs GPR this way, but I only realized this after explicitly asking the first author.

My insight was that adding log(likelihood) for a GPR model applied to two independent datasets leads to the (potentially) unintended consequence that data from different instruments ends up being treated as uncorrelated in the model. This is visualized in Figures 2.11 and 2.12. Constructing a single covariance matrix for the entire RV dataset, as discussed in Section 2.4, reduces effective degrees of freedom in the model, making the model less susceptible to overfitting. This tweaked model was also constructed to be more consistent with physical expectations, which I discuss more in Section 5.4 below. However, the relationship between radial velocities taken by instruments with different spectral efficiencies needs to be studied further.

5.3 How to Use it

The version of `radvel` that incorporates this model is working and validated, but currently requires some optimization before being released publicly. Nevertheless,

[3] "Classical" is perhaps too strong of a word here. It is often unclear how authors are choosing to model multiple independent datasets. Part of the motivation for my V1298 Tau paper was to point out that this modeling choice is not standardized and often unclear in the literature, but likely impacts reported constraints.

the code is available on this GitHub branch: https://github.com/California-Planet-Search/radvel/tree/tinygp. In writing this implementation, I gutted and re-wrote the `radvel` Gaussian Process back-end (the original implementation, which has been used since 2018, I also wrote). These are breaking changes, although existing `radvel` setup files and scripts using the API only require small modifications in order to use the new code. These changes are outlined for a sample script using the API in the code snippet below. While the user interface and API are very similar, the backend is completely different. The code is now built on `jax` by way of `tinygp`. `jax` is Google software that provides "replacements" for `numpy` functions to enable just-in-time compilation and execution on GPUs (and/or TPUs). The matrix inversion and determinant computation steps of computing a Gaussian Process likelihood are computationally intensive (GP likelihood calculation scales as $O(n^3)$), but parallelizable, so being able to run code on GPUs will speed up runtime significantly and enable application to larger datasets (e.g., Foreman-Mackey et al., 2017).

I include a tutorial of how to use the new code below, based on the data and model published in Dai et al. (2017). I originally wrote this tutorial in 2018 when the first version of the Gaussian Process code that I wrote in `radvel` was released, and I updated it to use the new underlying model in 2022, highlighting changes one would have to make in order to use the new `tinygp` implementation. This version of the tutorial can be compared to the version publicly hosted on the radvel documentation website[4].

```
import numpy as np
import pandas as pd
import os
import radvel
import radvel.likelihood
from radvel.plot import orbit_plots, mcmc_plots
from scipy import optimize

# read in data from Dai et al 2017
data = pd.read_csv(os.path.join(radvel.DATADIR,'k2-131.txt'), sep=' ')

t = np.array(data.time)
vel = np.array(data.mnvel)
errvel = np.array(data.errvel)
tel = np.array(data.tel)

telgrps = data.groupby('tel').groups
```

[4] https://radvel.readthedocs.io/en/latest/tutorials/GaussianProcess-tutorial.html

```
instnames = telgrps.keys()

# initialize Keplerian model
nplanets=1
params = radvel.Parameters(nplanets,basis='per tc secosw sesinw k')
params['per1'] = radvel.Parameter(value=0.3693038)
params['tc1'] = radvel.Parameter(value=2457582.9360)
params['sesinw1'] = radvel.Parameter(value=0.,vary=False) # fix eccentricity = 0
params['secosw1'] = radvel.Parameter(value=0.,vary=False)
params['k1'] = radvel.Parameter(value=6.55)
params['dvdt'] = radvel.Parameter(value=0.,vary=False)
params['curv'] = radvel.Parameter(value=0.,vary=False)

# initialize GP parameters at Dai+ published values
# hyperparameters are shared between instruments except activity amplitudes.
params['gp_amp_harps-n'] = radvel.Parameter(value=25.0) # eta1
params['gp_amp_pfs'] = radvel.Parameter(value=25.0)
params['gp_explength'] = radvel.Parameter(value=13.4) # eta2
params['gp_perlength'] = radvel.Parameter(value=0.54) # eta3
params['gp_per'] = radvel.Parameter(value=9.64)  # eta4

gpmodel = radvel.model.RVModel(params)

jit_guesses = {'harps-n':0.5, 'pfs':5.0}

likes = []
def initialize(tel_suffix):

    # instantiate a separate likelihood object for each instrument.

    indices = telgrps[tel_suffix]

    # <<<<<<<<< changed on tinygp branch <<<<<<<<
    # <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
    like = radvel.likelihood.RVLikelihood(
        gpmodel, t[indices], vel[indices], errvel[indices],
        suffix='_'+tel_suffix
    )
    # >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>

    # add in instrumental nuisance parameters
    like.params['gamma_'+tel_suffix] = radvel.Parameter(
        value=np.mean(vel[indices]), vary=False, linear=True
    )
    like.params['jit_'+tel_suffix] = radvel.Parameter(
        value=jit_guesses[tel_suffix], vary=True
    )
    likes.append(like)


for tel in instnames:
    initialize(tel)
```

```
70
71 # define the likelihood and posterior
72
73 # <<<<<<<<< changed on tinygp branch <<<<<<<<
74 # <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
75 gplike = radvel.likelihood.GPLikelihood(likes)
76 # >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
77
78 gppost = radvel.posterior.Posterior(gplike)
79
80 # apply same priors as Dai+
81 gppost.priors += [radvel.prior.Gaussian('per1', 0.3693038, 0.0000091)]
82 gppost.priors += [radvel.prior.Gaussian('tc1', 2457582.9360, 0.0011)]
83 gppost.priors += [radvel.prior.Jeffreys('k1', 0.01, 10.)]
84 gppost.priors += [radvel.prior.Jeffreys('gp_amp_pfs', 0.01, 100.)]
85 gppost.priors += [radvel.prior.Jeffreys('gp_amp_harps-n', 0.01, 100.)]
86 gppost.priors += [radvel.prior.Jeffreys('jit_pfs', 0.01, 10.)]
87 gppost.priors += [radvel.prior.Jeffreys('jit_harps-n', 0.01,10.)]
88 gppost.priors += [
89     radvel.prior.Gaussian('gp_explength', 13.4, sqrt(2))
90 ]
91 gppost.priors += [
92     radvel.prior.Gaussian('gp_per', 9.64, 0.12  )
93 ]
94 gppost.priors += [
95     radvel.prior.Gaussian('gp_perlength', 0.54, 0.10)
96 ]
97
98 # run MCMC
99 chains = radvel.mcmc(gppost,nrun=100,ensembles=3,savename='rawchains.h5')
100
101 # make a plot visualizing the RV data and model
102 GPPlot = orbit_plots.GPMultipanelPlot(
103     gppost,
104     subtract_gp_mean_model=False,
105     subtract_orbit_model=False
106 )
107 fig, axes = GPPlot.plot_multipanel()
108
109 # make a corner plot visualizing the posterior
110 Corner = mcmc_plots.CornerPlot(gppost, chains) # posterior distributions
111 Corner.plot()
```

The first plot output by this tutorial is Figure 2.12, and the second is reproduced below as Figure 5.5.

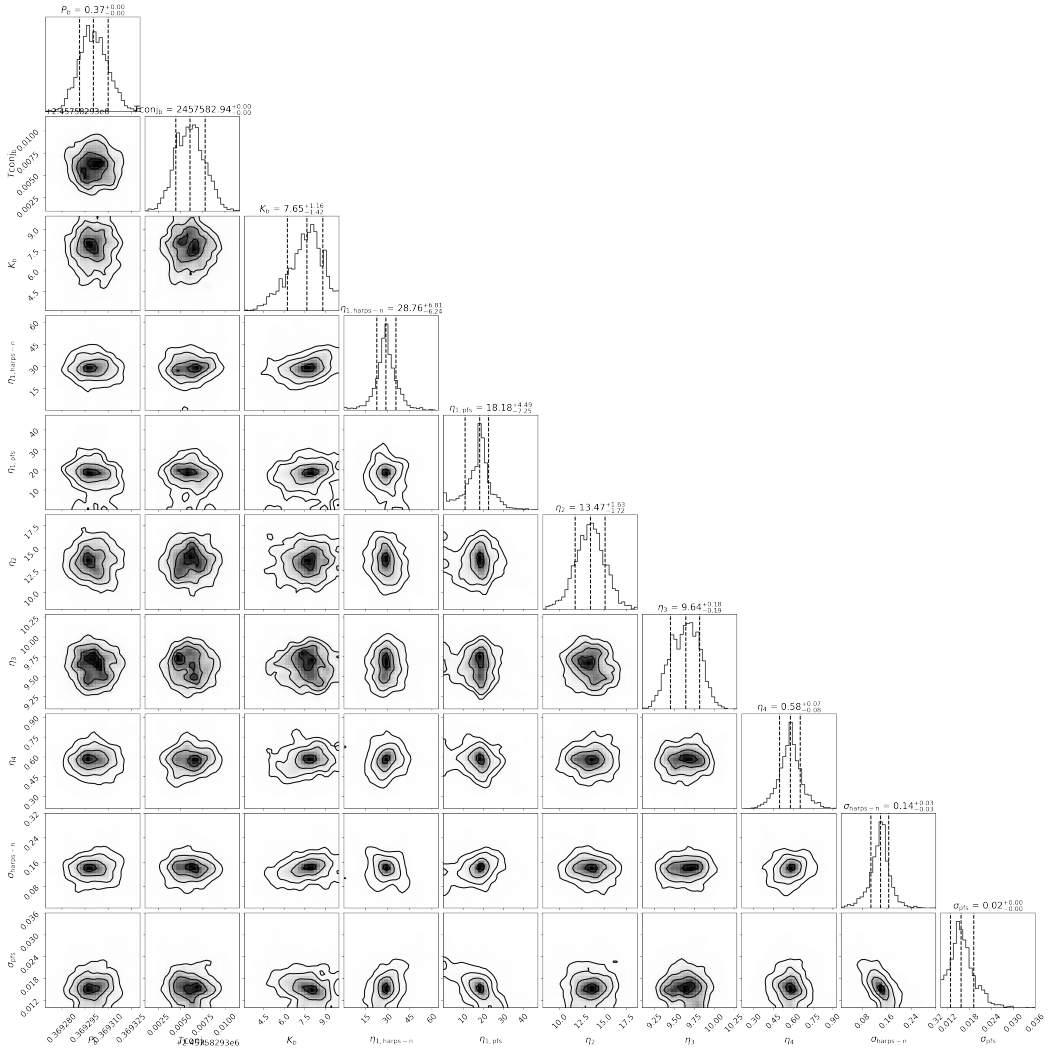

Figure 5.5: Posterior samples returned by the code snippet computing a GP model for K2-131 in section 5.3. Keplerian elements like Keplerian semi-amplitude are constrained jointly with Gaussian process hyperparameters like η_2. The Gaussian process amplitude parameters are unique to each instrument, in accordance with the model described in this chapter, but other hyperparameters that are more fundamentally tied to stellar processes (i.e., the rotation period, η_3 are shared across and jointly constrained by all RV data. From visual inspection, these chains are not completely converged, but this is unimportant since this plot is meant to help visualize the model-fitting process, rather than be interpreted as new scientific constraints.

5.4 Validating the Model

Validating the GPR model I have outlined in the previous sections involves two steps: first validating that the model is implemented correctly, and second validating that the model assumptions are correct.

The first test I performed seeking to provide evidence that the model was implemented correctly was to code up a visualization of the GP prediction for a model with two different datasets. This visualization became Figure 2.12. As expected, the GP predictions appeared to be linear combinations, with constraints coming from both datasets. I then ran a full MCMC fit to the K2-131 dataset shown in this figure (the results of which are shown in Figure 5.5), and compared it to the outputs of the same test using the current version of `radvel`, which assumes uncorrelated GPs for each instrument (the results of the same test using the old `radvel` GP implementation are shown in Figure 5.6 for comparison, and the relevant parameters are directly overplotted in Figure 5.7). Here are a few "sanity checks" I performed:

- Overall, the posteriors for the two tests are similar. The rotation period, η_4, planet semi-amplitude, and active region lifetime are within 1σ. In both cases, the posteriors are similar to those of Dai et al. (2017), who modeled all data as if it came from a single instrument using independent software.
- When the GPs are treated as correlated (i.e., when I use the new model), the GP shape parameter posterior, η_3, is on average larger than when the GPs are treated as uncorrelated. This means that the GP is not allowed to be as "wiggly." This is consistent with the story that when effective free parameters are removed, noise is less able to be treated as signal.
- The average of the Gaussian Process amplitude posteriors of the two instruments in the correlated GP case ($\eta_{1,\mathrm{pfs}}$ and $\eta_{1,\mathrm{harps-n}}$ is consistent with the single Gaussian process amplitude (η_1) posterior in the uncorrelated GP case.

I also repeated this test after 1) adding a large arbitrary offset to all of the HARPS-N RVs, and 2) multiplying all of the HARPS-N RVs by 0.01 to test that only the recovered γ and η_1 values were affected.

The next step in validating the model is checking that the underlying model itself, outside of the implementation details, is valid. This is an open problem, as there is a lot of physics in stellar activity that is not understood. I hope to spend time over the

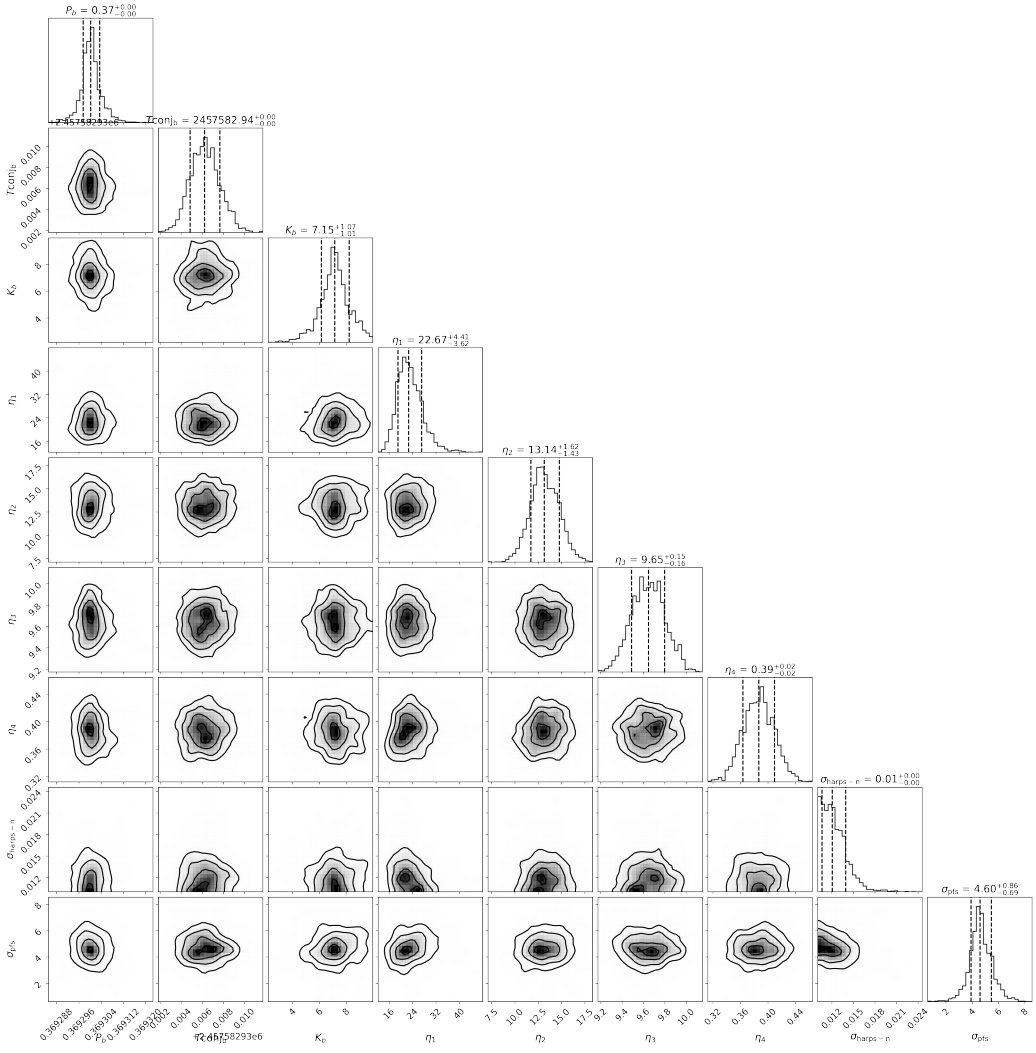

Figure 5.6: Posterior samples computed for a fit to the same K2-131 RVs analyzed in the code snippet in this section. Compare with Figure 5.5. The only difference between this result and the result shown in Figure 5.5 is that this result uses the current version of `radvel`, which assumes uncorrelated Gaussian Processes for each instrument.

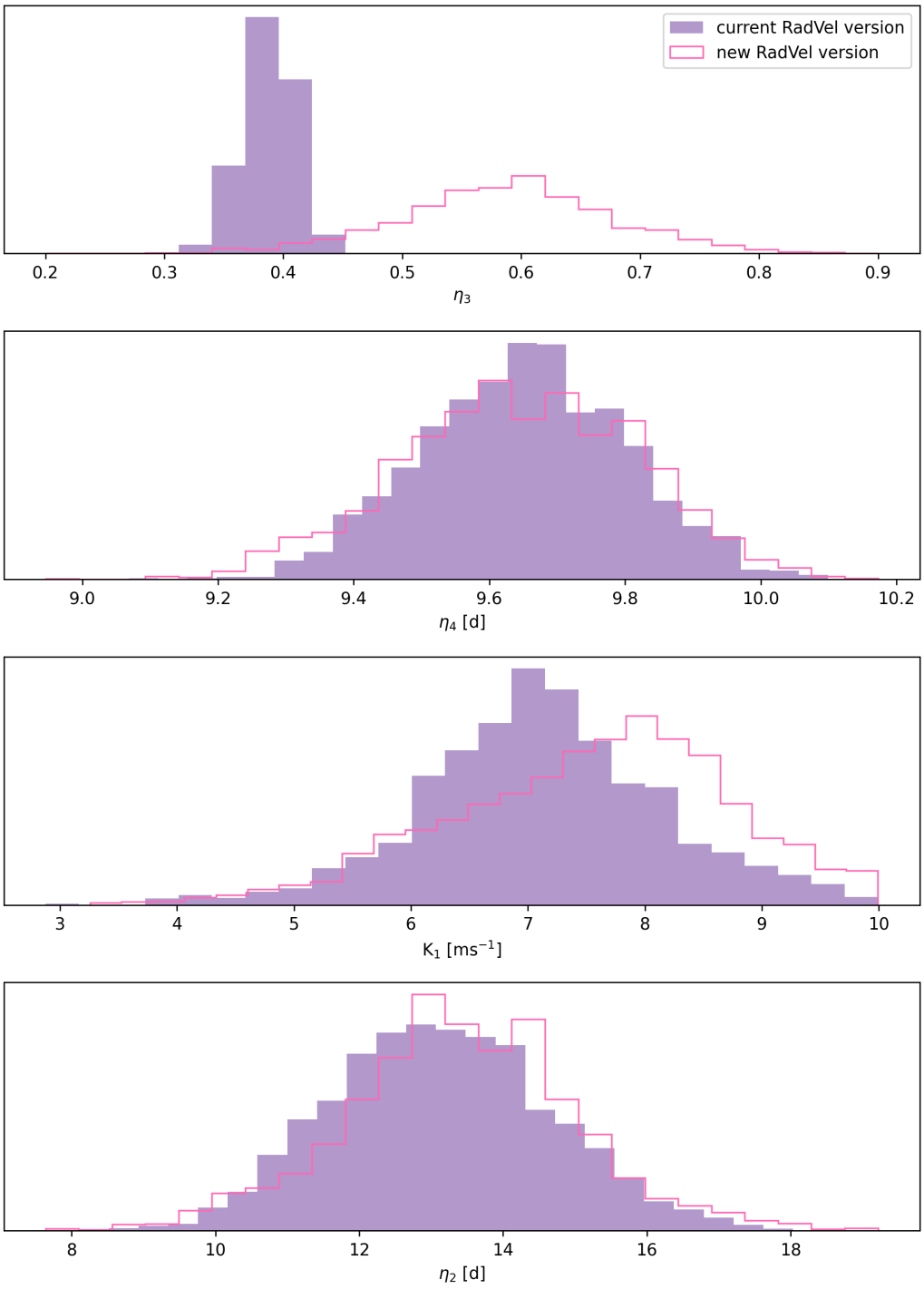

Figure 5.7: Direct comparison of relevant parameters shown in Figure 5.6 and 5.5.

next few years researching this further. However, I have done several preliminary checks and thought experiments that I will detail below.

The first piece of evidence that it is valid to treat stellar activity signals as linear combinations comes from actual data. For observations of some active pre-MS T-Tauri stars, such as Hubble I 4, the radial velocity signals, to first order, appear by-eye to be proportional (Figure 5.8). The phased cadence of these observations is still quite low, however, and it is difficult to tell whether the RV curves have different shapes as a function of wavelength. A similar phenomenological trend has been found in preliminary comparisons of simultaneous observations of the Sun-as-a-star with high resolution spectrographs: solar RVs taken with different instruments appear to be proportional to one another (Lily Zhao, private communication). This is the very beginning of these types of investigations, however, and more work (and higher quality contemporaneous data across a range of bandpasses) is needed to test this assumption.

We can also attempt to theoretically model the wavelength-dependence of stellar activity in order to justify or refute the assumption of stellar activity proportionality. As discussed previously, active regions are thought to affect RV measurements in two ways: blocking light from the photosphere coming from either the red- or blue-shifted side of the star, and suppressing the net convective blueshift on the stellar surface. Aigrain, Pont, and Zucker (2012) constructed the following model of a single point-like spot's effect on the stellar RV signature:

$$\begin{aligned} F(t) &= f\mathrm{MAX}\{\cos\beta(t); 0\} \\ \Delta\mathrm{RV}_{\mathrm{rot}}(t) &= -F(t)V_{\mathrm{eq}}\cos\delta\sin\phi(t)\sin i \\ \Delta\mathrm{RV}_{\mathrm{c}}(t) &= F(t)\delta V_{\mathrm{c}}\kappa\cos\beta(t), \end{aligned} \tag{5.5}$$

where $F(t)$ is the stellar flux, $f = 2(1-c)(1-\cos\alpha)$ is the flux decrease when the spot is at the disk center, c is the spot contrast, β is the angle between the spot area vector and the line of sight, and α is the angular radius of the spot on the stellar surface (in this model assumed to be small).

The resulting RV perturbations from this point spot for the rotation and convective suppression components are given below the flux variation equation. Additional variables defined here are as follows: V_{eq} is the equatorial velocity, $\phi(t)$ is the rotation phase of the spot, δ is the spot latitude, δV_c is the change in convective velocity due to the spot, and κ is the ratio of the magnetized area to the spot area.

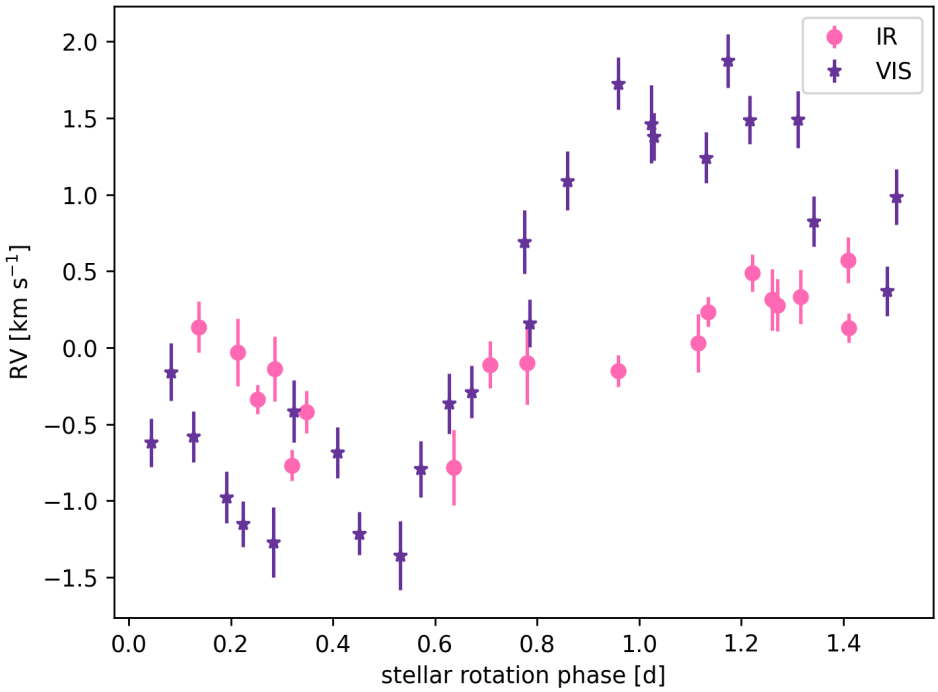

Figure 5.8: Radial velocities of Hubble I 4, from Mahmud et al. (2011). Optical RVs are from the Tull spectrograph, and NIR RVs are in the K-band, and taken with IRTF. The median of the NIR and visible RV datasets has been subtracted off, and the data have been phased to the 1.5459d stellar rotation period identified in Mahmud et al. (2011). **Takeaway:** to first order, the proportionality of stellar activity signals appears consistent with the data for this star, but higher cadence data are needed.

Scrutinizing $\Delta\mathrm{RV}_{\mathrm{rot}}(t)$, I see only a single wavelength-dependent parameter: $F(t)$.[5] Because both the spot brightness and the blocked photosphere brightness change as a function of wavelength, $F(t)$ will too. Therefore, we would expect RV variations due to blocking of rotationally-induced Doppler shifts to be scalar multiples at different wavelengths.

It is worth reiterating at this point that this model makes several potentially important simplifications, in particular that spots are small relative to the stellar surface (a simplification that is not present in the original Dorren (1987) model upon which

[5] V_{eq} also changes as a function of depth in the atmosphere (Christensen-Dalsgaard and Thompson, 2007), but for the Sun at least this is a small enough change relative to the location of the photosphere so as to make no difference. It is perhaps worth checking for other stars with stronger differential rotation signals.

this formalism is based) and that their contrast is independent of viewing angle. This model also ignores the flux contributions of bright magnetic features like faculae and plage, the contrasts of which do vary strongly from star center to limb (Hirayama and Moriyama, 1979). These and other assumptions will need to be tested against high-cadence data.

On the Sun, the suppression of convective blueshift is thought to dominate the activity signal over rotation timescales (Meunier, Desort, and Lagrange, 2010). Unlike in $\Delta\mathrm{RV}_{\mathrm{rot}}(t)$, I see several potentially wavelength-dependent components in the $\Delta\mathrm{RV}_c(t)$ term. The magnetized area may change as a function of depth in the stellar atmosphere (analogous to wavelength of observation), and the convective velocity also changes as a function of atmospheric depth (e.g., Cretignier et al., 2020). In fact, Bauer et al. (2018) found a strong relationship between convective zone depth and the R'HK activity metric, alluding to the importance of understanding convective blueshift suppression in modeling activity signals. However, all of these proportionalities should be constant in time, assuming that they do not change significantly with viewing angle.

This analysis suggests that the assumption of proportionality may only be valid when *either* the rotational component *or* the convective component dominate. When they are of roughly equal magnitude, however, the differences between convective blueshift velocity and magnetized area as a function of photosphere temperature mean that both components do not scale the same way with wavelength, making for more complicated relationship between the activity signals. On the Sun, this checks out because the convective component does dominate the RV signal. I think this is an intruiging direction for further thought and research, particularly using new techniques by Siegel et al (in prep) to separate out the convective and rotational components of radial velocity measurements by computing radial velocities using lines forming at different formation heights.

A shortcoming of this model worth mentioning that is worth mentioning is that it does not account for instrumental systematics, which must be treated separately.

5.5 Applications

This feature is brand-new and not yet available on pypi. Its capabilities were highlighted in Chapter 2, and, modulo small changes to existing scripts (pointed out in the code snippet in this chapter), it can be used as a "drop-in" replacement for "classical" GPR in the future.

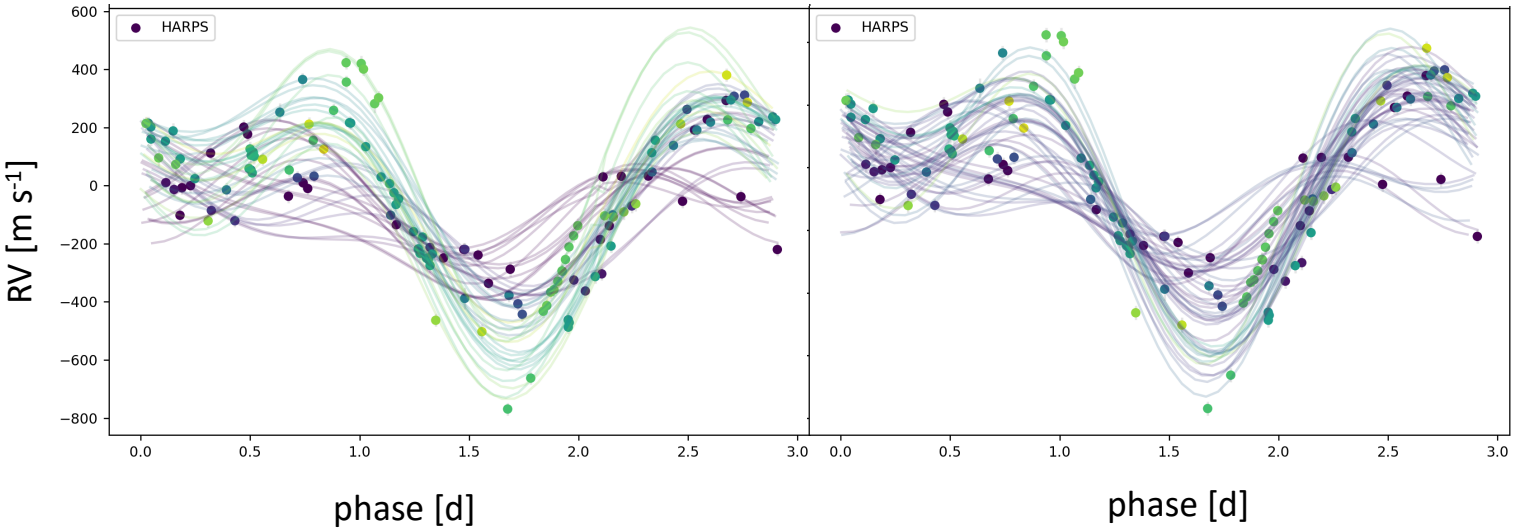

Figure 5.9: HARPS-N data of V1298 Tau, from Suárez Mascareño et al., 2021, phase-folded to the major rotation period of the star at 2.9 days. Points are colored by elapsed time since the first observation (darker points were taken earlier). In the left panel, only the stellar activity portion of the preferred SM21 model (see Chapter 2) is shown, and in the right panel, the full model fit, incorporating stellar activity and Keplerian parameters, is shown. The models are extremely similar, highlighting the outsized impact of the stellar activity contribution.

The two panels shown in Figure 5.9 illustrate the importance of preventing overfitting in models that jointly fit stellar activity signals and small planetary Doppler signals.

Chapter 6

FUTURE DIRECTIONS

6.1 Radial Velocity Frontiers

In (E)PRV science, the clear next frontier is understanding stellar activity. Operational RV instruments are consistently achieving Doppler precisions of tens of $\mathrm{cm\,s^{-1}}$, which means that we are no longer limited by our ability to measure RV shifts precisely, but by our understanding of stellar activity. The importance of high-quality datasets cannot be overstated here. Dedicated solar feeds for EPRV spectrographs (e.g., SoCal; Rubenzahl et al., in prep) are currently obtaining extremely high cadence (i.e., exposures every day, as often as readout time and weather allow) datasets of the Sun. These datasets will be invaluable stellar activity probes, particularly as more IR EPRV instruments come online. However, solar datasets alone will not be sufficient. We know that the Sun is not a "typical" star in many respects, and we will need datasets of comparable quality across diverse stellar activity archetypes. Older stars with faculae-dominated activity signals, more similar to the Sun, will also need to be monitored intensively, lest we start fitting "noise" in the solar activity signal to the detriment of our understanding of stellar activity more generally. A white paper on this topic, one of the outcomes of the recent EPRV conference, is in preparation.

In addition to improving our understanding of old stars like our Sun, high-cadence RV datasets of young, spot-dominated stars like V1298 Tau will be critically important to help us develop a physical understanding of how starspots impact radial velocity measurements. For V1298 Tau in particular, constructing a contemporaneous RV dataset that spans optical and IR bandpasses will not only be a helpful resource for this individual star, potentially enabling revised planet mass measurements, but for researchers seeking to understand stellar activity as a whole. Because young stars "isolate" spot activity from other sources of astrophysical and instrumental noise (by nature of the amplitude of this activity being orders of magnitude larger than for stars typically targeted by RV surveys), obtaining this dataset for V1298 Tau should be high priority as we enter the EPRV era.

Aside from more and better data, there are several modeling avenues I believe are worth pursuing for V1298 Tau, which will be broadly valuable for understanding

stellar activity. First, I think more time spent modeling the photometry will be fruitful. Much of the current literature on photometric modeling either uses very high-quality data for stars that have continuous photometric monitoring over the observational baseline (i.e., Kepler targets, Frasca et al., 2011), or does not treat differential rotation (e.g., Luger et al., 2018, which does contain a differential rotation prescription but runs into technical issues; individual spots are "stretched" out as they rotate differentially, which is perhaps nonphysical). Developing a viable photometric model will, I believe, go a long way for V1298 Tau. Independent constraints on the spot sizes and evolution timescales, stellar activity cycle timescale, and periods of the dominant active latitudes at each RV epoch will provide much needed physical priors on RV activity models.

Another important modeling avenue involves deriving relationships between all available datasets, and using them to maximize the constraint of each epoch of data. As an example, I mentioned in Chapter 2 an empirical correlation between the photometry derivative and RV when the two are phased to the dominant spot rotation period of 2.91d. This relationship could be explored and validated, then used to reduce effective free parameters in a model that jointly fits the RVs and photometry.

A final direction I believe is worth pursuing is going back to the fundamental methodology of deriving RVs from spectra, a current hot topic (for good reason, in my opinion) in RV exoplanet science. There are, in my view, two parallel approaches to mitigating stellar activity signals: methods that seek to empirically reduce scatter, for example by only using activity-insensitive lines to compute an RV signal, and methods that seek to understand and model activity signals. One popular scatter-reducing idea is line-by-line (e.g., Dumusque, 2018) analysis, which involves separately computing RV measurements for individual spectral lines, then performing outlier-rejection, or otherwise using only consistent lines to compute the final RV measurement. Another scatter-reducing idea is using machine learning, particularly algorithms like convolutional neural networks originally developed for image-recognition, to compute RVs directly from a time series of raw spectra (e.g., de Beurs et al. 2022). The current "state of the art" for these empirical scatter-reducing methods is ~20-50% Doppler error reduction, several orders of magnitude off from the necessary reduction needed to detect 8 $\mathrm{cm\,s^{-1}}$ rocky-planet signals around G-dwarfs with 1 $\mathrm{m\,s^{-1}}$ activity levels. A newer idea in the second category of better understanding activity, which I think holds much promise for young stars,

is performing formation-temperature dependent RV calculations (Al Moulla et al., 2022). This idea has the potential to separate out convective and rotational RV activity components (Siegel et al., in prep), which could lead to deep insights for young stars, and V1298 Tau in particular. This idea is still in its infancy.

Next Steps for `radvel`

`radvel` is a tool intended for state-of-the-art RV modeling, and it should continue to evolve as best practices for stellar activity mitigation change. The next immediate step for `radvel` is optimizing and publicly releasing the new Gaussian Process prescription described in Chapter 2, which should be a short and straightforward process. In the slightly longer-term, I think it will be important to study the wavelength dependence of stellar activity timeseries in order to evaluate the model I presented in Chapter 5. As I mentioned above, I am excited about the potential of deriving independent RV datasets as a function of formation temperature, and think that this will be a useful methodology for separating the convective and rotational components of stellar activity, and testing the model I presented.

6.2 Frontiers in Exoplanet Imaging

On the optical interferometry front, there are many exciting future avenues for understanding young planets. Refining our dynamical understanding of individual planetary systems is top priority here, as better individual constraints will also lead to better population-level constraints. The ExoGRAVITY team is closely monitoring ~25 directly-imaged planets and brown dwarfs, with the goal to refine individual orbital measurements, particularly eccentricities and obliquities, over the next few years. With refined orbital parameters (and a slightly larger sample, taking into account the ~10 more recent discoveries), the team will be able to revisit the population-level eccentricity (Bowler, Blunt, and Nielsen, 2020b) and obliquity (Bowler et al., 2023) distributions of these populations. I see this exercise as analogous to the revision of stellar radii in the era of Kepler that eventually led to the discovery of the super-Earth/sub-Neptune radius gap; the directly-imaged planet and brown dwarf populations are likely composed of sub-populations that formed differently. These differences may imprint features in the eccentricity and obliquity distributions that are "washed out" with the current level of orbital precision. GRAVITY measurements will, hopefully, bring these distributions into sharper focus, revealing features shaped by different underlying physics.

In order to better constrain the orbit of HIP 65426 b (Chapter 3 specifically, I believe that the most helpful next step is more measurements, both additional interferometric epochs and/or high spectral resolution RV measurements of the planet itself. There are also several physical quantities that I believe would be helpful to have better theoretical predictions for: in particular, the differing expectations for the dynamical and atmospheric properties of the planet as a function of core formation timescale (in particular, how the predictions change when pebble accretion or another more rapid core formation process is invoked) would help us make better inferences about how the planet's present-day properties relate to its formation conditions. A gravitational instability parameter study would also be extremely insightful. Obliquity predictions, in particular, would be valuable, since the obliquity of the planetary orbit (relative to the stellar spin axis) is measurable with current data and (as far as I am aware) has not been explicitly predicted from existing models.

A parallel effort in direct imaging studies over the next few years will be further developing our understanding of cold Jupiter atmospheres. For HIP 65426 b, performing full Bayesian retrievals is the next logical step, both for evaluating the radius, metallicity, and C/O constraints derived in this thesis using forward-modeling of self-consistent atmosphere models, which are limited in the parameter space they explore, and for evaluating the models themselves. For the population of cold Jupiters as a whole, more retrieval studies seeking to specifically evaluate the trustworthiness of parameters derived from these models (such as C/O; Xuan et al., 2022) will increase our confidence in physically interpreting the values in spectral fitting corner plots.

Another exciting general avenue in exoplanet orbit modeling is looking for non-Keplerian motion in astrometric residuals due to planet-planet interactions (e.g. Covarrubias, Blunt, and Wang, 2022, who showed that VLTI/GRAVITY will be sensitive to planet-planet effects in HR 8799 within 5 years) or moons (Ruffio et al., 2023) is an exciting and surprisingly realistic goal. The number of directly-imaged multi-planet systems is small, but growing (e.g., β Pictoris, Nowak et al., 2020; HD 209458 Hinkley et al., 2023), and multi-planet effects are already evident in newly-revealed multi-planet systems like β Pictoris (Lacour et al., 2021). Such measurements give independent dynamical mass measurements. Non-Keplerian motion as an independent mass-measurement technique for directly-imaged objects is important in and of itself; as stressed in Section 1.2, masses of cold Jupiters are crucial probes of formation, but most directly imaged planets we know of do not

have dynamical mass measurements, either because their host stars are too active (or too early type, or too quickly rotating, etc.) for radial velocity measurements, and/or because their absolute astrometric or RV signals are too small to be measurable with available data. However, a very widely separated planet with very little orbital motion around its primary, and therefore small astrometric and RV signals, could still have a nearby moon that enables a direct dynamical mass measurement. Exo-moon searches are also interesting from a formation theory standpoint in their own right, but this is beyond the scope of this document.

Next Steps for `orbitize!`

Like `radvel`, `orbitize!` is a toolkit for the direct-imaging community, and it should continue to grow in functionality and accessibility. An important immediate next step is the publication of a dedicated v2 documentation paper.

In the mid-term, a priority for the collaboration is benchmarking the two absolute astrometry implementations against one another. I also think that more effort should be spent visualizing fits to Hipparcos and Gaia measurements. Another consistent thread of effort among our collaboration is improving the efficiency of calculation, both by optimizing existing code and by developing and/or implementing additional backend algorithms. Porting existing rejection sampling algorithms like OFTI and TheJoker (Price-Whelan et al., 2017) to GPUs and other accelerated computing environments will help with this, as will adapting these algorithms to work on datasets of mixed type (e.g., both RVs and astrometry). Implementing the α-deep probabilistic inference formalism published in Sun et al. (2022) is another important priority.

The `orbitize!` collaboration maintains a list of open issues[1], and readers should feel free to submit their own issues with additional ideas.

Over the next decade, the `orbitize!` team will focus on 1) preparing for and (once available) developing methods for incorporating Gaia timeseries data products into orbit-fits, and 2) preparing for the launch of the Nancy Grace Roman Space Telescope. The sensitivity of Gaia is highly uncertain, and depends on the ultimate astrometric precision Gaia is able to obtain. This is difficult to determine from current data, since the community currently only has access to derived data products (e.g., parallax), and not raw intermediate data. Nevertheless, Gaia data will likely push the absolute astrometric detection limits towards shorter orbital

[1] https://github.com/sblunt/orbitize/issues

periods (comparable to and shorter than the Gaia mission lifetime, rather than the Hipparcos-Gaia baseline), and (hopefully) lower masses. Figure 3 of Brandt et al. (2021) shows an example simulated improvement for Gaia timeseries data relative to Gaia+Hipparcos accelerations for β Pictoris. Absolute astrometry from Gaia can be combined with relative astrometry and radial velocity measurements to refine orbits and dynamical masses of cold Jupiters, as well as to reveal (or not) previously undetected inner planets to cold Jupiters, which will be important for continued progress on the question of whether very widely separated cold Jupiters like HIP 65426 form via scattering (Bryan et al., 2016).

Preparing for the launch of Roman will bring its own set of challenges. Roman will image planets much closer to their stars than the current generation of exoplanet imagers, and most, if not all, of these will have prior radial velocity measurements. Although Roman is primarily a technology demonstration, it is expected to achieve visible-light contrast of 10^{-8} at 200 mas, which means that it is expected to be sensitive to giant planets in reflected light. This will mark a fundamental shift in the direct-imaging community; from young, self-luminous planets that are most visible in the infrared, to older, reflecting planets first discovered with RV[2]. Tools like RV "deconfusers" (Pogorelyuk et al., 2022) that assign radial velocity signals to each of the detected points of light in a Roman image of a multi-planet system will need to be developed, tested, and integrated into the `orbitize!` ecosystem.

6.3 The Next 50 Years

The Astro2020 Decadal Survey (National Academies of Sciences and Medicine, 2021) established that one of the three major goals of astrophysics over the next decade is to identify and characterize potentially habitable planets. There are two complementary pathways that the report proposes will be used for this landmark achievement: 1) radial velocity detection around Sun-like stars, and subsequent imaging at high spectral resolution, and 2) transit detection around M-dwarfs, and subsequent characterization via transmission spectroscopy. Both avenues have pros and cons; a clear advantage of the M-dwarf opportunity is the relative signal sizes of Earth-like planets in the habitable zone. Both radial velocity and transit signals are significantly larger, and Earth-like signals, especially transits, are detectable with current technology (e.g., the Tierras Observatory; Garcia-Mejia et al., 2020). Although Sun-like stars are more massive, and have more widely-separated habitable zones, which translate to smaller transit and RV signals, they are two major

[2] 47 Uma is a canonical Roman target; Smith et al. (2022).

advantages over M-dwarfs: first, that the only known habitable planet (ours) orbits a solar-type star, and second, that the high-energy flux environments of M-dwarfs may make circum-M dwarf planets inhospitable for life (Shields, Ballard, and Johnson, 2016). I am personally more interested in the first pathway, although both will (and should) be rigorously pursued over the timeline of my career. Studying Earth analogues is a goal I want to spend my career building up to, and I believe that building expertise in imaging and radial velocity detection is a step in this direction.

In order to discover Earth-like planets around Sun-like stars, we will need to understand stellar activity. RV instrumentation improvements (both hardware and software) will help, but we are already at the order of magnitude of Doppler precision that we will need. Stellar activity is by far the confounding factor here. On the imaging side, the Astro2020 report set up a clear (if still extremely challenging) technological development path towards the 10^{-11} intensity contrast we will need to be able to achieve to image habitable-zone Earths. The Roman Space telescope, which will achieve contrasts of 10^{-8}, a significant improvement over today's state-of-the-art of 10^{-4}-10^{-5}, is a stepping stone toward the next great observatory ("LuvEX"), intended to image Earths.

In 50 years, I hope that we will be able to approach the imaging and characterization of other Earths with refined observational and analytical techniques, a thorough and well-tested theory of planet formation and evolution, and a sense of adventure.

BIBLIOGRAPHY

Aerts, C. (Jan. 2021). "Probing the Interior Physics of Stars through Asteroseismology." In: *Reviews of Modern Physics* 93.1, 015001, p. 015001. DOI: 10.1103/RevModPhys.93.015001. arXiv: 1912.12300 [astro-ph.SR].

Agol, E. and D. C. Fabrycky (2018). "Transit-Timing and Duration Variations for the Discovery and Characterization of Exoplanets." In: *Handbook of Exoplanets*. Ed. by H. J. Deeg and J. A. Belmonte, 7, p. 7. DOI: 10.1007/978-3-319-55333-7_7.

Aigrain, S., F. Pont, and S. Zucker (Feb. 2012). "A Simple Method to Estimate Radial Velocity Variations due to Stellar Activity using Photometry." In: *Monthly Notices of the Royal Astronomical Society* 419.4, pp. 3147–3158. DOI: 10.1111/j.1365-2966.2011.19960.x. arXiv: 1110.1034 [astro-ph.SR].

Aigrain, S. and D. Foreman-Mackey (Sept. 2022). "Gaussian Process Regression for Astronomical Time-series." In: *arXiv e-prints*, arXiv:2209.08940, arXiv:2209.08940. arXiv: 2209.08940 [astro-ph.IM].

Al Moulla, K. et al. (Aug. 2022). "Measuring Precise Radial Velocities on Individual Spectral Lines. III. Dependence of Stellar Activity Signal on Line Formation Temperature." In: *Astronomy and Astrophysics* 664, A34, A34. DOI: 10.1051/0004-6361/202243276. arXiv: 2205.07047 [astro-ph.SR].

Allard, F., D. Homeier, and B. Freytag (Dec. 2011). "Model Atmospheres From Very Low Mass Stars to Brown Dwarfs." In: *16th Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun*. Ed. by C. Johns-Krull, M. K. Browning, and A. A. West. Vol. 448. Astronomical Society of the Pacific Conference Series, p. 91. DOI: 10.48550/arXiv.1011.5405. arXiv: 1011.5405 [astro-ph.SR].

Allard, F., D. Homeier, and B. Freytag (June 2012). "Models ofVvery-low-mass Stars, Brown Dwarfs and Exoplanets." In: *Philosophical Transactions of the Royal Society of London Series A* 370.1968, pp. 2765–2777. DOI: 10.1098/rsta.2011.0269. arXiv: 1112.3591 [astro-ph.SR].

Allard, F. et al. (Nov. 2007). "K-H_2 Quasi-molecular Absorption Detected in the T-dwarf ε Indi Ba." In: *Astronomy and Astrophysics* 474.2, pp. L21–L24. DOI: 10.1051/0004-6361:20078362. arXiv: 0709.1192 [astro-ph].

Allard, F. et al. (June 2003). "Model Atmospheres and Spectra: The Role of Dust." In: *Brown Dwarfs*. Ed. by E. Martín. Vol. 211, p. 325.

Ambikasaran, S. et al. (Mar. 2014). "Fast Direct Methods for Gaussian Processes and the Analysis of NASA Kepler Mission Data." In: URL: http://arxiv.org/abs/1403.6015.

Anugu, N. et al. (Dec. 2020). "CHARA Array Adaptive Optics: Complex Operational Software and Performance." In: *Optical and Infrared Interferometry and Imaging VII*. Ed. by P. G. Tuthill, A. Mérand, and S. Sallum. Vol. 11446. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 1144622, p. 1144622. DOI: 10.1117/12.2561560. arXiv: 2012.11667 [astro-ph.IM].

Armitage, P. J. (2020). *Astrophysics of Planet Formation, Second Edition*.

Astropy Collaboration et al. (Oct. 2013). "Astropy: A Community Python Package for Astronomy." In: *Astronomy and Astrophysics* 558, A33, A33. DOI: 10.1051/0004-6361/201322068. arXiv: 1307.6212 [astro-ph.IM].

Astropy Collaboration et al. (Sept. 2018). "The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package." In: *The Astronomical Journal* 156.3, 123, p. 123. DOI: 10.3847/1538-3881/aabc4f. arXiv: 1801.02634 [astro-ph.IM].

Astropy Collaboration et al. (Aug. 2022). "The Astropy Project: Sustaining and Growing a Community-oriented Open-source Project and the Latest Major Release (v5.0) of the Core Package." In: *apj* 935.2, 167, p. 167. DOI: 10.3847/1538-4357/ac7c74. arXiv: 2206.14220 [astro-ph.IM].

Basri, G. and R. Shah (Sept. 2020). "The Information Content in Analytic Spot Models of Broadband Precision Light Curves. II. Spot Distributions and Lifetimes and Global and Differential Rotation." In: *The Astrophysical Journal* 901.1, 14, p. 14. DOI: 10.3847/1538-4357/abae5d. arXiv: 2008.04969 [astro-ph.SR].

Bauer, F. F. et al. (Feb. 2018). "The Influence of Convective Blueshift on Radial Velocities of F, G, and K stars." In: *Astronomy and Astrophysics* 610, A52, A52. DOI: 10.1051/0004-6361/201731227.

Benatti, S. et al. (Oct. 2019). "A Possibly Inflated Planet around the Bright Young Star DS Tucanae A." In: *Astronomy and Astrophysics* 630, A81, A81. DOI: 10.1051/0004-6361/201935598. arXiv: 1904.01591 [astro-ph.SR].

Berger, T. A. et al. (June 2020). "The Gaia-Kepler Stellar Properties Catalog. I. Homogeneous Fundamental Properties for 186,301 Kepler Stars." In: *The Astronomical Journal* 159.6, 280, p. 280. DOI: 10.3847/1538-3881/159/6/280. arXiv: 2001.07737 [astro-ph.SR].

Blunt, S. et al. (May 2017). "Orbits for the Impatient: A Bayesian Rejection-sampling Method for Quickly Fitting the Orbits of Long-period Exoplanets." In: *The Astronomical Journal* 153.5, 229, p. 229. DOI: 10.3847/1538-3881/aa6930. arXiv: 1703.10653 [astro-ph.EP].

Blunt, S. et al. (Nov. 2019). "Radial Velocity Discovery of an Eccentric Jovian World Orbiting at 18 au." In: *The Astronomical Journal* 158.5, 181, p. 181. DOI: 10.3847/1538-3881/ab3e63. arXiv: 1908.09925 [astro-ph.EP].

Blunt, S. et al. (Mar. 2020). "orbitize!: A Comprehensive Orbit-fitting Software Package for the High-contrast Imaging Community." In: *The Astronomical Journal* 159.3, 89, p. 89. DOI: 10.3847/1538-3881/ab6663. arXiv: 1910.01756 [astro-ph.EP].

Bouma, L. G. et al. (Mar. 2022). "A 38 Million Year Old Neptune-sized Planet in the Kepler Field." In: *The Astronomical Journal* 163.3, 121, p. 121. DOI: 10.3847/1538-3881/ac4966. arXiv: 2112.14776 [astro-ph.EP].

Bowler, B. P. (Oct. 2016). "Imaging Extrasolar Giant Planets." In: *Publications of the Astronomical Society of the Pacific* 128.968, p. 102001. DOI: 10.1088/1538-3873/128/968/102001. arXiv: 1605.02731 [astro-ph.EP].

Bowler, B. P., S. C. Blunt, and E. L. Nielsen (Feb. 2020a). "Population-level Eccentricity Distributions of Imaged Exoplanets and Brown Dwarf Companions: Dynamical Evidence for Distinct Formation Channels." In: *The Astronomical Journal* 159.2, 63, p. 63. DOI: 10.3847/1538-3881/ab5b11. arXiv: 1911.10569 [astro-ph.EP].

Bowler, B. P., S. C. Blunt, and E. L. Nielsen (Feb. 2020b). "Population-level Eccentricity Distributions of Imaged Exoplanets and Brown Dwarf Companions: Dynamical Evidence for Distinct Formation Channels." In: *AJ* 159.2, 63, p. 63. DOI: 10.3847/1538-3881/ab5b11. arXiv: 1911.10569 [astro-ph.EP].

Bowler, B. P. and E. L. Nielsen (2018). "Occurrence Rates from Direct Imaging Surveys." In: *Handbook of Exoplanets*. Ed. by H. J. Deeg and J. A. Belmonte, 155, p. 155. DOI: 10.1007/978-3-319-55333-7_155.

Bowler, B. P. et al. (Apr. 2023). "Rotation Periods, Inclinations, and Obliquities of Cool Stars Hosting Directly Imaged Substellar Companions: Spin-Orbit Misalignments Are Common." In: *The Astronomical Journal* 165.4, 164, p. 164. DOI: 10.3847/1538-3881/acbd34. arXiv: 2301.04692 [astro-ph.EP].

Boyajian, T. S. et al. (Oct. 2012). "Stellar Diameters and Temperatures. II. Main-sequence K- and M-stars." In: *The Astrophysical Journal* 757.2, 112, p. 112. DOI: 10.1088/0004-637X/757/2/112. arXiv: 1208.2431 [astro-ph.SR].

Bradbury, J. et al. (2018). *JAX: composable transformations of Python+NumPy programs*. Version 0.3.13. URL: http://github.com/google/jax.

Brandt, G. M. et al. (Dec. 2021). "htof: A New Open-source Tool for Analyzing Hipparcos, Gaia, and Future Astrometric Missions." In: *The Astronomical Journal* 162.6, 230, p. 230. DOI: 10.3847/1538-3881/ac12d0. arXiv: 2109.06761 [astro-ph.IM].

Brandt, T. D. (Dec. 2018). "The Hipparcos-Gaia Catalog of Accelerations." In: *The Astrophysical Journal Supplement Series* 239.2, 31, p. 31. DOI: 10.3847/1538-4365/aaec06. arXiv: 1811.07283 [astro-ph.SR].

Bryan, M. L. et al. (Apr. 2016). "Statistics of Long Period Gas Giant Planets in Known Planetary Systems." In: *The Astrophysical Journal* 821.2, 89, p. 89. DOI: 10.3847/0004-637X/821/2/89. arXiv: 1601.07595 [astro-ph.EP].

Bryan, M. L. et al. (Apr. 2020). "Obliquity Constraints on an Extrasolar Planetary-mass Companion." In: *The Astronomical Journal* 159.4, 181, p. 181. DOI: 10.3847/1538-3881/ab76c6. arXiv: 2002.11131 [astro-ph.EP].

Buchner, J. et al. (Apr. 2014). "X-ray Spectral Modelling of the AGN Obscuring Region in the CDFS: Bayesian Model Selection and Catalogue." In: *Astronomy and Astrophysics* 564, A125, A125. DOI: 10.1051/0004-6361/201322971. arXiv: 1402.0004 [astro-ph.HE].

Burke, C. J. et al. (Mar. 2020). *TESS-Point: High Precision TESS Pointing Tool*. Astrophysics Source Code Library, record ascl:2003.001. ascl: 2003.001.

Butler, R. P. et al. (June 1996a). "Attaining Doppler Precision of 3 M s-1." In: *Publications of the Astronomical Society of the Pacific* 108, p. 500. DOI: 10.1086/133755.

Butler, R. P. et al. (June 1996b). "Attaining Doppler Precision of 3 M s-1." In: *Publications of the Astronomical Society of the Pacific* 108, p. 500. DOI: 10.1086/133755.

Caceres, G. A. et al. (Aug. 2019). "Autoregressive Planet Search: Application to the Kepler Mission." In: *The Astronomical Journal* 158.2, 58, p. 58. DOI: 10.3847/1538-3881/ab26ba. arXiv: 1905.09852 [astro-ph.EP].

Cale, B. L. et al. (Dec. 2021). "Diving Beneath the Sea of Stellar Activity: Chromatic Radial Velocities of the Young AU Mic Planetary System." In: *The Astronomical Journal* 162.6, 295, p. 295. DOI: 10.3847/1538-3881/ac2c80. arXiv: 2109.13996 [astro-ph.EP].

Carnall, A. C. (May 2017). "SpectRes: A Fast Spectral Resampling Tool in Python." In: *arXiv e-prints*, arXiv:1705.05165, arXiv:1705.05165. DOI: 10.48550/arXiv.1705.05165. arXiv: 1705.05165 [astro-ph.IM].

Carpenter, J. M., L. A. Hillenbrand, and M. F. Skrutskie (June 2001). "Near-Infrared Photometric Variability of Stars toward the Orion A Molecular Cloud." In: *The Astronomical Journal* 121.6, pp. 3160–3190. DOI: 10.1086/32108610.48550/arXiv.astro-ph/0102446. arXiv: astro-ph/0102446 [astro-ph].

Carter, A. L. et al. (Aug. 2022). "The JWST Early Release Science Program for Direct Observations of Exoplanetary Systems I: High Contrast Imaging of the Exoplanet HIP 65426 b from 2-16 μm." In: *arXiv e-prints*, arXiv:2208.14990, arXiv:2208.14990. DOI: 10.48550/arXiv.2208.14990. arXiv: 2208.14990 [astro-ph.EP].

Carvalho, A. et al. (Mar. 2021). "Radial Velocity Monitoring of the Young Star Hubble 4: Disentangling Star-spot Lifetimes from Orbital Motion." In: *The Astrophysical Journal* 910.1, 33, p. 33. DOI: 10.3847/1538-4357/abe237. arXiv: 2102.06257 [astro-ph.SR].

Chaplin, W. J. and A. Miglio (Aug. 2013). "Asteroseismology of Solar-Type and Red-Giant Stars." In: *Annual Reviews of Astronomy and Astrophysics* 51.1, pp. 353–392. DOI: 10.1146/annurev-astro-082812-140938. arXiv: 1303.1957 [astro-ph.SR].

Charnay, B. et al. (Feb. 2018). "A Self-consistent Cloud Model for Brown Dwarfs and Young Giant Exoplanets: Comparison with Photometric and Spectroscopic Observations." In: *The Astrophysical Journal* 854.2, 172, p. 172. DOI: 10.3847/1538-4357/aaac7d. arXiv: 1711.11483 [astro-ph.EP].

Chauvin, G. et al. (Oct. 2004). "A Giant Planet Candidate Near a Young Brown Dwarf. Direct VLT/NACO Observations using IR Wavefront Sensing." In: *Astronomy and Astrophysics* 425, pp. L29–L32. DOI: 10.1051/0004-6361:200400056. arXiv: astro-ph/0409323 [astro-ph].

Chauvin, G. et al. (Sept. 2017). "Discovery of a Warm, Dusty Giant Planet around HIP 65426." In: *Astronomy and Astrophysics* 605, L9, p. L9. DOI: 10.1051/0004-6361/201731152. arXiv: 1707.01413 [astro-ph.EP].

Cheetham, A. C. et al. (Feb. 2019). "Spectral and Orbital Characterisation of the Directly Imaged Giant Planet HIP 65426 b." In: *Astronomy and Astrophysics* 622, A80, A80. DOI: 10.1051/0004-6361/201834112. arXiv: 1812.07198 [astro-ph.EP].

Chen, H. and L. A. Rogers (Nov. 2016). "Evolutionary Analysis of Gaseous Sub-Neptune-mass Planets with MESA." In: *The Astrophysical Journal* 831.2, 180, p. 180. DOI: 10.3847/0004-637X/831/2/180. arXiv: 1603.06596 [astro-ph.EP].

Christensen-Dalsgaard, J. and M. J. Thompson (Jan. 2007). "Observational Results and Issues Concerning the Tachocline." In: *The Solar Tachocline*. Ed. by D. W. Hughes, R. Rosner, and N. O. Weiss, p. 53.

Collier Cameron, A. (2018). "The Impact of Stellar Activity on the Detection and Characterization of Exoplanets." In: *Handbook of Exoplanets*. Ed. by H. J. Deeg and J. A. Belmonte, 23, p. 23. DOI: 10.1007/978-3-319-55333-7_23.

Covarrubias, S., S. Blunt, and J. J. Wang (Mar. 2022). "N-body Interactions will be Detectable in the HR-8799 System within 5 yr with VLTI-GRAVITY." In: *Research Notes of the American Astronomical Society* 6.3, 66, p. 66. DOI: 10.3847/2515-5172/ac61d8. arXiv: 2204.03679 [astro-ph.EP].

Cretignier, M. et al. (Jan. 2020). "Measuring Precise Radial Velocities on Individual Spectral Lines. II. Dependance of Stellar Activity Signal on Line Depth." In: *Astronomy and Astrophysics* 633, A76, A76. DOI: 10.1051/0004-6361/201936548. arXiv: 1912.05192 [astro-ph.EP].

Czekala, I. et al. (Oct. 2015). "Constructing a Flexible Likelihood Function for Spectroscopic Inference." In: *The Astrophysical Journal* 812.2, 128, p. 128. DOI: 10.1088/0004-637X/812/2/128. arXiv: 1412.5177 [astro-ph.SR].

Dai, F. et al. (Dec. 2017). "The Discovery and Mass Measurement of a New Ultra-short-period Planet: K2-131b." In: *The Astronomical Journal* 154.6, 226, p. 226. DOI: 10.3847/1538-3881/aa9065. arXiv: 1710.00076 [astro-ph.EP].

David, T. J. et al. (June 2016). "A Neptune-sized Transiting Planet Closely Orbiting a 5-10-Million-Year-Old Star." In: *Nature* 534.7609, pp. 658–661. DOI: 10.1038/nature18293. arXiv: 1606.06729 [astro-ph.EP].

David, T. J. et al. (Aug. 2019a). "A Warm Jupiter-sized Planet Transiting the Pre-main-sequence Star V1298 Tau." In: *The Astronomical Journal* 158.2, 79, p. 79. DOI: 10.3847/1538-3881/ab290f. arXiv: 1902.09670 [astro-ph.EP].

David, T. J. et al. (Nov. 2019b). "Four Newborn Planets Transiting the Young Solar Analog V1298 Tau." In: *The Astrophysical Journal Letters* 885.1, L12, p. L12. DOI: 10.3847/2041-8213/ab4c99. arXiv: 1910.04563 [astro-ph.EP].

de Beurs, Z. L. et al. (Aug. 2022). "Identifying Exoplanets with Deep Learning. IV. Removing Stellar Activity Signals from Radial Velocity Measurements Using Neural Networks." In: *The Astronomical Journal* 164.2, 49, p. 49. DOI: 10.3847/1538-3881/ac738e. arXiv: 2011.00003 [astro-ph.EP].

Dorren, J. D. (Sept. 1987). "A New Formulation of the Starspot Model, and the Consequences of Starspot Structure." In: *The Astrophysical Journal* 320, p. 756. DOI: 10.1086/165593.

Drazkowska, J. et al. (Mar. 2022). "Planet Formation Theory in the Era of ALMA and Kepler: from Pebbles to Exoplanets." In: *arXiv e-prints*, arXiv:2203.09759, arXiv:2203.09759. DOI: 10.48550/arXiv.2203.09759. arXiv: 2203.09759 [astro-ph.EP].

Dressing, C. D. and D. Charbonneau (Apr. 2013). "The Occurrence Rate of Small Planets around Small Stars." In: *The Astrophysical Journal* 767.1, 95, p. 95. DOI: 10.1088/0004-637X/767/1/95. arXiv: 1302.1647 [astro-ph.EP].

Dumusque, X. (Nov. 2018). "Measuring Precise Radial Velocities on Individual Spectral Lines. I. Validation of the Method and Application to Mitigate Stellar Activity." In: *Astronomy and Astrophysics* 620, A47, A47. DOI: 10.1051/0004-6361/201833795. arXiv: 1809.01548 [astro-ph.SR].

Dumusque, X. et al. (Jan. 2011). "Planetary Detection Limits Taking into Account Stellar Noise. I. Observational Strategies to Reduce Stellar Oscillation and Granulation Effects." In: *Astronomy and Astrophysics* 525, A140, A140. DOI: 10.1051/0004-6361/201014097. arXiv: 1010.2616 [astro-ph.EP].

Dupuy, T. J. et al. (Jan. 2019). "A Model-independent Mass and Moderate Eccentricity for β Pic b." In: *The Astrophysical Journal Letters* 871.1, L4, p. L4. DOI: 10.3847/2041-8213/aafb31. arXiv: 1812.11530 [astro-ph.EP].

Durbin, J. and S. J. Koopman (2001). *Time Series Analysis by State Space Methods*.

Espinoza, N., D. Kossakowski, and R. Brahm (Dec. 2018). "Juliet: a Versatile Modelling Tool for Transiting and Non-transiting Exoplanetary Systems." In: *arXiv e-prints*, arXiv:1812.08549, arXiv:1812.08549. arXiv: 1812.08549 [astro-ph.EP].

Fabrycky, D. C. et al. (Aug. 2014). "Architecture of Kepler's Multi-transiting Systems. II. New Investigations with Twice as Many Candidates." In: *The Astrophysical Journal* 790.2, 146, p. 146. DOI: 10.1088/0004-637X/790/2/14610.48550/arXiv.1202.6328. arXiv: 1202.6328 [astro-ph.EP].

Feigelson, E. D., G. J. Babu, and G. A. Caceres (Aug. 2018). "Autoregressive Times Series Methods for Time Domain Astronomy." In: *Frontiers in Physics* 6, 80, p. 80. DOI: 10.3389/fphy.2018.00080. arXiv: 1901.08003 [astro-ph.IM].

Feinstein, A. D. et al. (Jan. 2022). "V1298 Tau with TESS: Updated Ephemerides, Radii, and Period Constraints from a Second Transit of V1298 Tau e." In: *The Astrophysical Journal Letters* 925.1, L2, p. L2. DOI: 10.3847/2041-8213/ac4745. arXiv: 2111.08660 [astro-ph.EP].

Feroz, F. and M. P. Hobson (Feb. 2008). "Multimodal Nested Sampling: an Efficient and Robust Alternative to Markov Chain Monte Carlo Methods for Astronomical Data Analyses." In: *Monthly Notices of the Royal Astronomical Society* 384.2, pp. 449–463. DOI: 10.1111/j.1365-2966.2007.12353.x. arXiv: 0704.3704 [astro-ph].

Feroz, F., M. P. Hobson, and M. Bridges (Oct. 2009). "MULTINEST: an Efficient and Robust Bayesian Inference Tool for Cosmology and Particle Physics." In: *Monthly Notices of the Royal Astronomical Society* 398.4, pp. 1601–1614. DOI: 10.1111/j.1365-2966.2009.14548.x. arXiv: 0809.3437 [astro-ph].

Feroz, F. et al. (Nov. 2019). "Importance Nested Sampling and the MultiNest Algorithm." In: *The Open Journal of Astrophysics* 2.1, 10, p. 10. DOI: 10.21105/astro.1306.2144. arXiv: 1306.2144 [astro-ph.IM].

Ferrer-Chávez, R., J. J. Wang, and S. Blunt (May 2021). "Biases in Orbital Fitting of Directly Imaged Exoplanets with Small Orbital Coverage." In: *The Astronomical Journal* 161.5, 241, p. 241. DOI: 10.3847/1538-3881/abf0a8. arXiv: 2103.12877 [astro-ph.EP].

Finociety, B. et al. (Dec. 2021). "The T Tauri star V410 Tau in the Eyes of SPIRou and TESS." In: *Monthly Notices of the Royal Astronomical Society* 508.3, pp. 3427–3445. DOI: 10.1093/mnras/stab2778. arXiv: 2109.11755 [astro-ph.SR].

Foreman-Mackey, D. et al. (2017). "Fast and Scalable Gaussian Process Modeling with Applications to Astronomical Time Series." In: *ArXiv*. URL: https://arxiv.org/abs/1703.09710.

Foreman-Mackey, D. et al. (Oct. 2022). *dfm/tinygp: v0.2.3*. Version v0.2.3. DOI: 10.5281/zenodo.7269074. URL: https://doi.org/10.5281/zenodo.7269074.

Foreman-Mackey, D. (June 2016). "corner.py: Scatterplot Matrices in Python." In: *The Journal of Open Source Software* 1.2, p. 24. DOI: 10.21105/joss.00024. URL: https://doi.org/10.21105/joss.00024.

Foreman-Mackey, D. (Jan. 2018). "An Astronomer's Introduction to Gaussian Processes." In: *American Astronomical Society Meeting Abstracts #231*. Vol. 231. American Astronomical Society Meeting Abstracts, 225.01, p. 225.01.

Foreman-Mackey, D. et al. (Mar. 2013). "emcee: The MCMC Hammer." In: *Publications of the Astronomical Society of the Pacific* 125.925, p. 306. DOI: 10.1086/670067. arXiv: 1202.3665 [astro-ph.IM].

Frasca, A. et al. (Aug. 2011). "Magnetic Activity and Differential Rotation in the Very Young Star KIC 8429280." In: *Astronomy and Astrophysics* 532, A81, A81. DOI: 10.1051/0004-6361/201116980. arXiv: 1106.4928 [astro-ph.SR].

Fulton, B. J. et al. (Oct. 2016). "Three Temperate Neptunes Orbiting Nearby Stars." In: *The Astrophysical Journal* 830.1, 46, p. 46. DOI: 10.3847/0004-637X/830/1/46. arXiv: 1607.00007 [astro-ph.EP].

Fulton, B. J. et al. (Sept. 2017). "The California-Kepler Survey. III. A Gap in the Radius Distribution of Small Planets." In: *The Astronomical Journal* 154.3, 109, p. 109. DOI: 10.3847/1538-3881/aa80eb. arXiv: 1703.10375 [astro-ph.EP].

Fulton, B. J. et al. (Apr. 2018). "RadVel: The Radial Velocity Modeling Toolkit." In: *Publications of the Astronomical Society of the Pacific* 130.986, p. 044504. DOI: 10.1088/1538-3873/aaaaa8. arXiv: 1801.01947 [astro-ph.IM].

Gagné, J. et al. (Mar. 2018). "BANYAN. XI. The BANYAN Σ Multivariate Bayesian Algorithm to Identify Members of Young Associations with 150 pc." In: *The Astrophysical Journal* 856.1, 23, p. 23. DOI: 10.3847/1538-4357/aaae09. arXiv: 1801.09051 [astro-ph.SR].

Gaia Collaboration et al. (Nov. 2016). "The Gaia Mission." In: *Astronomy and Astrophysics* 595, A1, A1. DOI: 10.1051/0004-6361/201629272. arXiv: 1609.04153 [astro-ph.IM].

Gaia Collaboration et al. (July 2022). "Gaia Data Release 3: Summary of the Content and Survey roperties." In: *arXiv e-prints*, arXiv:2208.00211, arXiv:2208.00211. DOI: 10.48550/arXiv.2208.00211. arXiv: 2208.00211 [astro-ph.GA].

Garcia-Mejia, J. et al. (Dec. 2020). "The Tierras Observatory: An ultra-precise photometer to characterize nearby terrestrial exoplanets." In: *Ground-based and Airborne Telescopes VIII*. Ed. by H. K. Marshall, J. Spyromilio, and T. Usuda. Vol. 11445. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 114457R, 114457R. DOI: 10.1117/12.2561467. arXiv: 2012.09744 [astro-ph.IM].

Gelman, A., J. Hwang, and A. Vehtari (2014). "Understanding Predictive Information Criteria for Bayesian Models." In: *Statistics and Computing* 24.6, pp. 997–1016. DOI: 10.1007/s11222-013-9416-2. URL: https://doi.org/10.1007/s11222-013-9416-2.

Giles, H. A. C., A. Collier Cameron, and R. D. Haywood (Dec. 2017). "A Kepler Study of Starspot Lifetimes with respect to Light-curve Amplitude and Spectral Type." In: *Monthly Notices of the Royal Astronomical Society* 472.2, pp. 1618–1627. DOI: 10.1093/mnras/stx1931. arXiv: 1707.08583 [astro-ph.SR].

Ginzburg, S., H. E. Schlichting, and R. Sari (May 2018). "Core-Powered Mass-Loss and the Radius Distribution of Small Exoplanets." In: *Monthly Notices of the Royal Astronomical Society* 476.1, pp. 759–765. DOI: 10.1093/mnras/sty290. arXiv: 1708.01621 [astro-ph.EP].

Gravity Collaboration et al. (June 2017). "First Light for GRAVITY: Phase Referencing Optical Interferometry for the Very Large Telescope Interferometer." In: *Astronomy and Astrophysics* 602, A94, A94. DOI: 10.1051/0004-6361/201730838. arXiv: 1705.02345 [astro-ph.IM].

Gravity Collaboration et al. (Jan. 2020). "Peering into the Formation History of β Pictoris b with VLTI/GRAVITY Long-baseline Interferometry." In: *Astronomy and Astrophysics* 633, A110, A110. DOI: 10.1051/0004-6361/201936898. arXiv: 1912.04651 [astro-ph.EP].

Green, R. M. (1985). *Spherical Astronomy*.

Grunblatt, S. et al. (Mar. 2023). "An Unlikely Survivor: a Low-density Hot Neptune Orbiting a Red Giant Star." In: *arXiv e-prints*, arXiv:2303.06728, arXiv:2303.06728. DOI: 10.48550/arXiv.2303.06728. arXiv: 2303.06728 [astro-ph.EP].

Grunblatt, S. K., A. W. Howard, and R. D. Haywood (Aug. 2015). "Determining the Mass of Kepler-78b with Nonparametric Gaussian Process Estimation." In: *The Astrophysical Journal* 808.2, 127, p. 127. DOI: 10.1088/0004-637X/808/2/127. arXiv: 1501.00369 [astro-ph.EP].

Hampson, K. M. et al. (2021). "Adaptive Optics for High-Resolution Imaging." In: *Nature Reviews Methods Primers* 1.1, p. 68. DOI: 10.1038/s43586-021-00066-7. URL: https://doi.org/10.1038/s43586-021-00066-7.

Hara, N. C. et al. (Jan. 2017). "Radial Velocity Data Analysis with Compressed Sensing Techniques." In: *Monthly Notices of the Royal Astronomical Society* 464.1, pp. 1220–1246. DOI: 10.1093/mnras/stw2261. arXiv: 1609.01519 [astro-ph.IM].

Harris, C. R. et al. (Sept. 2020). "Array Programming with NumPy." In: *Nature* 585.7825, pp. 357–362. DOI: 10.1038/s41586-020-2649-2. URL: https://doi.org/10.1038/s41586-020-2649-2.

Hinkley, S. et al. (Aug. 2022). "Direct Discovery of the Inner Exoplanet in the HD206893 System. Evidence for Deuterium Burning in a Planetary-mass Companion." In: *arXiv e-prints*, arXiv:2208.04867, arXiv:2208.04867. DOI: 10.48550/arXiv.2208.04867. arXiv: 2208.04867 [astro-ph.EP].

Hinkley, S. et al. (Mar. 2023). "Direct Discovery of the Inner Exoplanet in the HD 206893 System. Evidence for Deuterium Burning in a Planetary-mass Companion." In: *Astronomy and Astrophysics* 671, L5, p. L5. DOI: 10.1051/0004-6361/202244727.

Hirayama, T. and F. Moriyama (Sept. 1979). "Center to Limb Variation of the Intensity of the Photospheric Faculae." In: *Solar Physics* 63.2, pp. 251–255. DOI: 10.1007/BF00174531.

Hirsch, L. A. et al. (June 2019). "Discovery of a White Dwarf Companion to HD 159062." In: *The Astrophysical Journal* 878.1, 50, p. 50. DOI: 10.3847/1538-4357/ab1b11. arXiv: 1905.06440 [astro-ph.SR].

Holcomb, R. J. et al. (Sept. 2022). "SpinSpotter : An Automated Algorithm for Identifying Stellar Rotation Periods with Autocorrelation Analysis." In: *The Astrophysical Journal* 936.2, 138, p. 138. DOI: 10.3847/1538-4357/ac8990. arXiv: 2206.10629 [astro-ph.SR].

Householder, A. and L. Weiss (Dec. 2022). "The Standard RV Equation uses ω_p, not ω_*." In: *arXiv e-prints*, arXiv:2212.06966, arXiv:2212.06966. DOI: 10.48550/arXiv.2212.06966. arXiv: 2212.06966 [astro-ph.EP].

Howard, A. W. and B. J. Fulton (Nov. 2016). "Limits on Planetary Companions from Doppler Surveys of Nearby Stars." In: *Publications of the Astronomical Society of the Pacific* 128.969, p. 114401. DOI: 10.1088/1538-3873/128/969/114401. arXiv: 1606.03134 [astro-ph.EP].

Howard, A. W. et al. (Oct. 2010). "The Occurrence and Mass Distribution of Close-in Super-Earths, Neptunes, and Jupiters." In: *Science* 330.6004, p. 653. DOI: 10.1126/science.1194854. arXiv: 1011.0143 [astro-ph.EP].

Huerta, M. et al. (May 2008). "Starspot-Induced Radial Velocity Variability in LkCa 19." en. In: *The Astrophysical Journal* 678.1. Publisher: IOP Publishing, p. 472. ISSN: 0004-637X. DOI: 10.1086/526415. URL: https://iopscience.iop.org/article/10.1086/526415/meta (visited on 04/24/2020).

Hunter, J. D. (2007). "Matplotlib: A 2D Graphics Environment." In: *Computing in Science & Engineering* 9.3, pp. 90–95. DOI: 10.1109/MCSE.2007.55.

Jenkins, J. M. et al. (Aug. 2016). "The TESS Science Processing Operations Center." In: *Software and Cyberinfrastructure for Astronomy IV*. Ed. by G. Chiozzi and J. C. Guzman. Vol. 9913. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 99133E, 99133E. DOI: 10.1117/12.2233418.

Johns-Krull, C. M. (Aug. 2007). "The Magnetic Fields of Classical T Tauri Stars." In: *The Astrophysical Journal* 664.2, pp. 975–985. DOI: 10.1086/519017. arXiv: 0704.2923 [astro-ph].

Johnson, M. C. et al. (June 2022). "An Aligned Orbit for the Young Planet V1298 Tau b." In: *The Astronomical Journal* 163.6, 247, p. 247. DOI: 10.3847/1538-3881/ac6271. arXiv: 2110.10707 [astro-ph.EP].

Kaeufl, H.-U. et al. (Sept. 2004). "CRIRES: a High-resolution Infrared Spectrograph for ESO's VLT." In: *Ground-based Instrumentation for Astronomy*. Ed. by A. F. M. Moorwood and M. Iye. Vol. 5492. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, pp. 1218–1227. DOI: 10.1117/12.551480.

Kesseli, A. Y. et al. (June 2018). "Magnetic Inflation and Stellar Mass. II. On the Radii of Single, Rapidly Rotating, Fully Convective M-Dwarf Stars." In: *The Astronomical Journal* 155.6, 225, p. 225. DOI: 10.3847/1538-3881/aabccb. arXiv: 1804.04133 [astro-ph.SR].

Klein, B. et al. (June 2022). "One Year of AU Mic with HARPS - II. Stellar Activity and Star-Planet Interaction." In: *Monthly Notices of the Royal Astronomical Society* 512.4, pp. 5067–5084. DOI: 10.1093/mnras/stac761. arXiv: 2203.08190 [astro-ph.EP].

Kraus, S. et al. (July 2020). "Spin-Orbit Alignment of the β Pictoris Planetary System." In: *The Astrophysical Journal Letters* 897.1, L8, p. L8. DOI: 10.3847/2041-8213/ab9d27. arXiv: 2006.10784 [astro-ph.EP].

Kreidberg, L. (Nov. 2015). "batman: BAsic Transit Model cAlculatioN in Python." In: *Publications of the Astronomical Society of the Pacific* 127.957, p. 1161. DOI: 10.1086/683602. arXiv: 1507.08285 [astro-ph.EP].

Kuzuhara, M. et al. (Sept. 2013). "Direct Imaging of a Cold Jovian Exoplanet in Orbit around the Sun-like Star GJ 504." In: *The Astrophysical Journal* 774.1, 11, p. 11. DOI: 10.1088/0004-637X/774/1/11. arXiv: 1307.2886 [astro-ph.EP].

Lacour, S. et al. (Apr. 2019a). "The GRAVITY Fringe Tracker." In: *Astronomy and Astrophysics* 624, A99, A99. DOI: 10.1051/0004-6361/201834981. arXiv: 1901.03202 [astro-ph.IM].

Lacour, S. et al. (Dec. 2020). "The ExoGRAVITY Project: Using Single Mode Interferometry to Characterize Exoplanets." In: *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*. Vol. 11446. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 114460O, 114460O. DOI: 10.1117/12.2561667. arXiv: 2101.07098 [astro-ph.EP].

Lacour, S. et al. (Oct. 2021). "The Mass of β Pictoris c from β Pictoris b Orbital Motion." In: *Astronomy and Astrophysics* 654, L2, p. L2. DOI: 10.1051/0004-6361/202141889. arXiv: 2109.10671 [astro-ph.EP].

Lacour, S. et al. (2019b). "First Direct Detection of an Exoplanet by Optical Interferometry." In: *Astronomy and Astrophysics*.

Lagrange, A. .-. et al. (Jan. 2009). "A Probable Giant Planet Imaged in the β Pictoris Disk. VLT/NaCo Deep L'-band Imaging." In: *Astronomy and Astrophysics* 493.2, pp. L21–L25. DOI: 10.1051/0004-6361:200811325. arXiv: 0811.3583 [astro-ph].

Lagrange, A. .-. et al. (Aug. 2019). "Evidence for an Additional Planet in the β Pictoris System." In: *Nature Astronomy* 3, pp. 1135–1142. DOI: 10.1038/s41550-019-0857-1.

Lambrechts, M. and A. Johansen (Dec. 2014). "Forming the Cores of Giant Planets from the Radial Pebble Flux in Protoplanetary Discs." In: *Astronomy and Astrophysics* 572, A107, A107. DOI: 10.1051/0004-6361/201424343. arXiv: 1408.6094 [astro-ph.EP].

Lanza, A. F., M. Rodono, and R. A. Zappala (Oct. 1994). "Fourier Analysis of Spotted Star Light Curves as a Tool to Detect Stellar Differential Rotation. II. Spots' Evolution and Binarity." In: *Astronomy and Astrophysics* 290, pp. 861–867.

Lapeyrere, V. et al. (July 2014). "GRAVITY Data Reduction Software." In: *Optical and Infrared Interferometry IV*. Ed. by J. K. Rajagopal, M. J. Creech-Eakman, and F. Malbet. Vol. 9146. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 91462D, p. 91462D. DOI: 10.1117/12.2056850.

Lega, E. et al. (Feb. 2021). "Migration of Jupiter-mass Planets in Low-viscosity Discs." In: *Astronomy and Astrophysics* 646, A166, A166. DOI: 10.1051/0004-6361/202039520. arXiv: 2012.12930 [astro-ph.EP].

Lightkurve Collaboration et al. (Dec. 2018). *Lightkurve: Kepler and TESS Time Series Analysis in Python*. Astrophysics Source Code Library. ascl: 1812.013.

Lithwick, Y., J. Xie, and Y. Wu (Dec. 2012). "Extracting Planet Mass and Eccentricity from TTV Data." In: *The Astrophysical Journal* 761.2, 122, p. 122. DOI: 10.1088/0004-637X/761/2/122. arXiv: 1207.4192 [astro-ph.EP].

Llop-Sayson, J. et al. (Nov. 2021). "Constraining the Orbit and Mass of epsilon Eridani b with Radial Velocities, Hipparcos IAD-Gaia DR2 Astrometry, and Multiepoch Vortex Coronagraphy Upper Limits." In: *The Astronomical Journal* 162.5, 181, p. 181. DOI: 10.3847/1538-3881/ac134a. arXiv: 2108.02305 [astro-ph.EP].

Lopez, E. D., J. J. Fortney, and N. Miller (Dec. 2012). "How Thermal Evolution and Mass-loss Sculpt Populations of Super-Earths and Sub-Neptunes: Application to the Kepler-11 System and Beyond." In: *The Astrophysical Journal* 761.1, 59, p. 59. DOI: 10.1088/0004-637X/761/1/59. arXiv: 1205.0010 [astro-ph.EP].

López-Morales, M. et al. (Dec. 2016). "Kepler-21b: A Rocky Planet Around a V = 8.25 Magnitude Star." In: *The Astronomical Journal* 152.6, 204, p. 204. DOI: 10.3847/0004-6256/152/6/204. arXiv: 1609.07617 [astro-ph.EP].

Lovis, C. and D. Fischer (2010). "Radial Velocity Techniques for Exoplanets." In: *Exoplanets*. Ed. by S. Seager, pp. 27–53.

Lubin, J. et al. (Feb. 2022). "TESS-Keck Survey. IX. Masses of Three Sub-Neptunes Orbiting HD 191939 and the Discovery of a Warm Jovian plus a Distant Substellar Companion." In: *The Astronomical Journal* 163.2, 101, p. 101. DOI: 10.3847/1538-3881/ac3d38. arXiv: 2108.02208 [astro-ph.EP].

Luger, R. et al. (Oct. 2016). "EVEREST: Pixel Level Decorrelation of K2 Light Curves." In: *The Astronomical Journal* 152.4, 100, p. 100. DOI: 10.3847/0004-6256/152/4/10010.48550/arXiv.1607.00524. arXiv: 1607.00524 [astro-ph.EP].

Luger, R. et al. (Sept. 2018). "An Update to the EVEREST K2 Pipeline: Short Cadence, Saturated Stars, and Kepler-like Photometry Down to Kp = 15." In: *The Astronomical Journal* 156.3, 99, p. 99. DOI: 10.3847/1538-3881/aad23010.48550/arXiv.1702.05488. arXiv: 1702.05488 [astro-ph.IM].

Madhusudhan, N., M. A. Amin, and G. M. Kennedy (Oct. 2014). "Toward Chemical Constraints on Hot Jupiter Migration." In: *The Astrophysical Journal Letters* 794.1, L12, p. L12. DOI: 10.1088/2041-8205/794/1/L12. arXiv: 1408.3668 [astro-ph.EP].

Mahmud, N. I. et al. (July 2011). "Starspot-Induces Optical and Infrared Radial Velocity Variability in T Tauri Star Hubble I 4." en. In: *The Astrophysical Journal* 736.2. Publisher: IOP Publishing, p. 123. ISSN: 0004-637X. DOI: 10.1088/0004-637X/736/2/123. URL: https://doi.org/10.1088%5C%2F0004-637x%5C%2F736%5C%2F2%5C%2F123 (visited on 04/24/2020).

Mann, A. W. et al. (Apr. 2022). "TESS Hunt for Young and Maturing Exoplanets (THYME). VI. An 11 Myr Giant Planet Transiting a Very-low-mass Star in Lower Centaurus Crux." In: *The Astronomical Journal* 163.4, 156, p. 156. DOI: 10.3847/1538-3881/ac511d. arXiv: 2110.09531 [astro-ph.EP].

Marleau, G.-D. et al. (Apr. 2019). "Exploring the Formation by Core Accretion and the Luminosity Evolution of Directly Imaged Planets. The Case of HIP 65426 b." In: *Astronomy and Astrophysics* 624, A20, A20. DOI: 10.1051/0004-6361/201833597. arXiv: 1902.01869 [astro-ph.EP].

Marley, M. S. et al. (Jan. 2007). "On the Luminosity of Young Jupiters." In: *The Astrophysical Journal* 655.1, pp. 541–549. DOI: 10.1086/509759. arXiv: astro-ph/0609739 [astro-ph].

Mawet, D. et al. (Jan. 2019). "Deep Exploration of $\in$ Eridani with Keck Ms-band Vortex Coronagraphy and Radial Velocities: Mass and Orbital Parameters of the Giant Exoplanet." In: *The Astronomical Journal* 157.1, 33, p. 33. DOI: 10.3847/1538-3881/aaef8a. arXiv: 1810.03794 [astro-ph.EP].

Mayor, M. and D. Queloz (Nov. 1995). "A Jupiter-mass Companion to a Solar-type Star." In: *Nature* 378.6555, pp. 355–359. DOI: 10.1038/378355a0.

McKinney, W. (2010). "Data Structures for Statistical Computing in Python." In: *Proceedings of the 9th Python in Science Conference*. Ed. by S. van der Walt and J. Millman, pp. 56–61. DOI: 10.25080/Majora-92bf1922-00a.

Meunier, N., M. Desort, and A. .-. Lagrange (Mar. 2010). "Using the Sun to Estimate Earth-like Planets Detection Capabilities . II. Impact of Plages." In: *Astronomy and Astrophysics* 512, A39, A39. DOI: 10.1051/0004-6361/200913551. arXiv: 1001.1638 [astro-ph.EP].

Mikkola, S. (Sept. 1987). "A Cubic Approximation for Kepler's Equation." In: *Celestial Mechanics* 40.3-4, pp. 329–334. DOI: 10.1007/BF01235850.

Mollière, P. et al. (July 2022). "Interpreting the Atmospheric Composition of Exoplanets: Sensitivity to Planet Formation Assumptions." In: *The Astrophysical Journal* 934.1, 74, p. 74. DOI: 10.3847/1538-4357/ac6a56. arXiv: 2204.13714 [astro-ph.EP].

Muñoz-Jaramillo, A., D. Nandy, and P. C. H. Martens (June 2009). "Helioseismic Data Inclusion in Solar Dynamo Models." In: *The Astrophysical Journal* 698.1, pp. 461–478. DOI: 10.1088/0004-637X/698/1/461. arXiv: 0811.3441 [astro-ph].

Nagpal, V. et al. (Jan. 2021). "Using Long Baseline Radial Velocities and Direct Imaging to Make 13-Sigma Dynamical Mass Measurements for the Components of the HD 104304 Stellar Binary." In: *American Astronomical Society Meeting Abstracts*. Vol. 53. American Astronomical Society Meeting Abstracts, 530.03, p. 530.03.

Nagpal, V. et al. (Feb. 2023). "The Impact of Bayesian Hyperpriors on the Population-level Eccentricity Distribution of Imaged Planets." In: *The Astronomical Journal* 165.2, 32, p. 32. DOI: 10.3847/1538-3881/ac9fd2. arXiv: 2211.02121 [astro-ph.EP].

Namekata, K. et al. (Feb. 2019). "Lifetimes and Emergence/Decay Rates of Star Spots on Solar-type Stars Estimated by Kepler Data in Comparison with Those of Sunspots." In: *The Astrophysical Journal* 871.2, 187, p. 187. DOI: 10.3847/1538-4357/aaf471. arXiv: 1811.10782 [astro-ph.SR].

National Academies of Sciences, E. and Medicine (2021). *Pathways to Discovery in Astronomy and Astrophysics for the 2020s*. DOI: 10.17226/26141.

Nava, C. et al. (Jan. 2020). "Exoplanet Imitators: A Test of Stellar Activity Behavior in Radial Velocity Signals." In: *The Astronomical Journal* 159.1, 23, p. 23. DOI: 10.3847/1538-3881/ab53ec. arXiv: 1911.04106 [astro-ph.EP].

Nayakshin, S. (Sept. 2017). "A Desert of Gas Giant Planets beyond Tens of au: from Feast to Famine." In: *Monthly Notices of the Royal Astronomical Society* 470.2, pp. 2387–2409. DOI: 10.1093/mnras/stx1351.

Newton, E. R. et al. (July 2019). "TESS Hunt for Young and Maturing Exoplanets (THYME): A Planet in the 45 Myr Tucana-Horologium Association." In: *The Astrophysical Journal Letters* 880.1, L17, p. L17. DOI: 10.3847/2041-8213/ab2988. arXiv: 1906.10703 [astro-ph.EP].

Nicholson, B. A. and S. Aigrain (Oct. 2022). "Quasi-Periodic Gaussian Processes for Stellar Activity: From Physical to Kernel Parameters." In: *Monthly Notices of the Royal Astronomical Society* 515.4, pp. 5251–5266. DOI: 10.1093/mnras/stac2097. arXiv: 2207.12164 [astro-ph.SR].

Nielsen, E. L. et al. (Feb. 2008). "Constraints on Extrasolar Planet Populations from VLT NACO/SDI and MMT SDI and Direct Adaptive Optics Imaging Surveys: Giant Planets are Rare at Large Separations." In: *The Astrophysical Journal* 674.1, pp. 466–481. DOI: 10.1086/524344. arXiv: 0706.4331 [astro-ph].

Nielsen, E. L. et al. (Dec. 2017). "Evidence That the Directly Imaged Planet HD 131399 Ab Is a Background Star." In: *The Astronomical Journal* 154.6, 218, p. 218. DOI: 10.3847/1538-3881/aa8a69. arXiv: 1705.06851 [astro-ph.EP].

Nielsen, E. L. et al. (July 2019). "The Gemini Planet Imager Exoplanet Survey: Giant Planet and Brown Dwarf Demographics from 10 to 100 au." In: *The Astronomical Journal* 158.1, 13, p. 13. DOI: 10.3847/1538-3881/ab16e9. arXiv: 1904.05358 [astro-ph.EP].

Nowak, M. et al. (Oct. 2020). "Direct Confirmation of the Radial-velocity Planet β Pictoris c." In: *Astronomy and Astrophysics* 642, L2, p. L2. DOI: 10.1051/0004-6361/202039039. arXiv: 2010.04442 [astro-ph.EP].

O'Toole, S. J. et al. (Jan. 2009). "Selection Functions in Doppler Planet Searches." In: *Monthly Notices of the Royal Astronomical Society* 392.2, pp. 641–654. DOI: 10.1111/j.1365-2966.2008.14051.x. arXiv: 0810.1589 [astro-ph].

Öberg, K. I., R. Murray-Clay, and E. A. Bergin (Dec. 2011). "The Effects of Snowlines on C/O in Planetary Atmospheres." In: *The Astrophysical Journal Letters* 743.1, L16, p. L16. DOI: 10.1088/2041-8205/743/1/L16. arXiv: 1110.5567 [astro-ph.GA].

Owen, J. E. (Nov. 2020). "Constraining the Entropy of Formation from Young Transiting Planets." In: *Monthly Notices of the Royal Astronomical Society* 498.4, pp. 5030–5040. DOI: 10.1093/mnras/staa2784. arXiv: 2009.03919 [astro-ph.EP].

Owen, J. E. and Y. Wu (Oct. 2013). "Kepler Planets: A Tale of Evaporation." In: *The Astrophysical Journal* 775.2, 105, p. 105. DOI: 10.1088/0004-637X/775/2/105. arXiv: 1303.3899 [astro-ph.EP].

Pegues, J. et al. (Feb. 2021). "Dynamical Masses and Stellar Evolutionary Model Predictions of M Stars." In: *The Astrophysical Journal* 908.1, 42, p. 42. DOI: 10.3847/1538-4357/abd4eb. arXiv: 2101.05838 [astro-ph.SR].

Pepe, F. et al. (Nov. 2013). "An Earth-Sized Planet with an Earth-Like Density." In: *Nature* 503.7476, pp. 377–380. DOI: 10.1038/nature12768. arXiv: 1310.7987 [astro-ph.EP].

Petigura, E. A. et al. (Jan. 2020). "K2-19b and c are in a 3:2 Commensurability but out of Resonance: A Challenge to Planet Assembly by Convergent Migration." In: *The Astronomical Journal* 159.1, 2, p. 2. DOI: 10.3847/1538-3881/ab5220. arXiv: 1910.12899 [astro-ph.EP].

Petrus, S. et al. (Apr. 2021). "Medium-resolution Spectrum of the Exoplanet HIP 65426 b." In: *Astronomy and Astrophysics* 648, A59, A59. DOI: 10.1051/0004-6361/202038914. arXiv: 2012.02798 [astro-ph.EP].

Pezzato, J. et al. (Sept. 2019). "Status of the Keck Planet Imager and Characterizer Phase II Development." In: *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*. Vol. 11117. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 111170U, 111170U. DOI: 10.1117/12.2529685. arXiv: 1909.06487 [astro-ph.IM].

Plavchan, P. et al. (June 2020). "A Planet within the Debris Disk around the Pre-Main-Sequence Star AU Microscopii." In: *Nature* 582.7813, pp. 497–500. DOI: 10.1038/s41586-020-2400-z. arXiv: 2006.13248 [astro-ph.EP].

Pogorelyuk, L. et al. (Oct. 2022). "Deconfusing Detections in Directly Imaged Multiplanet Systems." In: *The Astrophysical Journal* 937.2, 66, p. 66. DOI: 10.3847/1538-4357/ac8d56.

Prato, L. et al. (Nov. 2008). "A Young-Planet Search in Visible and Infrared Light: DN Tauri, V836 Tauri, and V827 Tauri." In: *The Astrophysical Journal Letters* 687.2, p. L103. DOI: 10.1086/59320110.48550/arXiv.0809.3599. arXiv: 0809.3599 [astro-ph].

Price-Whelan, A. M. et al. (Mar. 2017). "The Joker: A Custom Monte Carlo Sampler for Binary-star and Exoplanet Radial Velocity Data." In: *The Astrophysical Journal* 837.1, 20, p. 20. DOI: 10.3847/1538-4357/aa5e50. arXiv: 1610.07602 [astro-ph.SR].

Qi, C. et al. (Nov. 2015). "Chemical Imaging of the CO Snow Line in the HD 163296 Disk." In: *The Astrophysical Journal* 813.2, 128, p. 128. DOI: 10.1088/0004-637X/813/2/128. arXiv: 1510.00968 [astro-ph.SR].

Rafikov, R. R. (Mar. 2005). "Can Giant Planets Form by Direct Gravitational Instability?" In: *The Astrophysical Journal Letters* 621.1, pp. L69–L72. DOI: 10.1086/428899. arXiv: astro-ph/0406469 [astro-ph].

Rajpaul, V. et al. (Sept. 2015). "A Gaussian Process Framework for Modelling Stellar Activity Signals in Radial Velocity Data." In: *Monthly Notices of the Royal Astronomical Society* 452.3, pp. 2269–2291. DOI: 10.1093/mnras/stv1428. arXiv: 1506.07304 [astro-ph.EP].

Rasmussen, C. E. and C. K. I. Williams (2006). *Gaussian processes for machine learning*. Adaptive computation and machine learning. MIT Press, pp. I–XVIII, 1–248. ISBN: 026218253X.

Rizzuto, A. C. et al. (July 2020). "TESS Hunt for Young and Maturing Exoplanets (THYME). II. A 17 Myr Old Transiting Hot Jupiter in the Sco-Cen Association." In: *The Astronomical Journal* 160.1, 33, p. 33. DOI: 10.3847/1538-3881/ab94b7. arXiv: 2005.00013 [astro-ph.EP].

Rosenthal, L. J. et al. (July 2021a). "The California Legacy Survey. I. A Catalog of 178 Planets from Precision Radial Velocity Monitoring of 719 Nearby Stars over Three Decades." In: *The Astrophysical Journal Supplement Series* 255.1, 8, p. 8. DOI: 10.3847/1538-4365/abe23c. arXiv: 2105.11583 [astro-ph.EP].

Rosenthal, L. J. et al. (July 2021b). "The California Legacy Survey. I. A Catalog of 178 Planets from Precision Radial Velocity Monitoring of 719 Nearby Stars over Three Decades." In: *The Astrophysical Journal Supplement Series* 255.1, 8, p. 8. DOI: 10.3847/1538-4365/abe23c. arXiv: 2105.11583 [astro-ph.EP].

Rosenthal, M. M. and R. A. Murray-Clay (Sept. 2018). "Restrictions on the Growth of Gas Giant Cores via Pebble Accretion." In: *The Astrophysical Journal* 864.1, 66, p. 66. DOI: 10.3847/1538-4357/aad560. arXiv: 1808.05947 [astro-ph.EP].

Ruane, G. et al. (Aug. 2018). "Review of High-Contrast Imaging Systems for Current and Future Ground- and Space-Based Telescopes I: Coronagraph Design Methods and Optical Performance Metrics." In: *Space Telescopes and Instrumentation 2018: Optical, Infrared, and Millimeter Wave*. Ed. by M. Lystrup et al. Vol. 10698. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 106982S, 106982S. DOI: 10.1117/12.2312948. arXiv: 1807.07042 [astro-ph.IM].

Rubenzahl, R. A. et al. (Mar. 2021). "The TESS-Keck Survey. IV. A Retrograde, Polar Orbit for the Ultra-low-density, Hot Super-Neptune WASP-107b." In: *The Astronomical Journal* 161.3, 119, p. 119. DOI: 10.3847/1538-3881/abd177. arXiv: 2101.09371 [astro-ph.EP].

Ruffio, J.-B. et al. (Mar. 2023). "Detecting Exomoons from Radial Velocity Measurements of Self-luminous Planets: Application to Observations of HR 7672 B and Future Prospects." In: *The Astronomical Journal* 165.3, 113, p. 113. DOI: 10.3847/1538-3881/acb34a. arXiv: 2301.04206 [astro-ph.EP].

Rustamkulov, Z. et al. (Feb. 2023). "Early Release Science of the exoplanet WASP-39b with JWST NIRSpec PRISM." In: *Nature* 614.7949, pp. 659–663. DOI: 10.1038/s41586-022-05677-y. arXiv: 2211.10487 [astro-ph.EP].

Saar, S. H. and R. A. Donahue (Aug. 1997). "Activity-Related Radial Velocity Variation in Cool Stars." In: *The Astrophysical Journal* 485.1, pp. 319–327. DOI: 10.1086/304392.

Schnupp, C. et al. (June 2010). "Discovery of a Stellar Companion to the Nearby Solar-Analogue HD 104304." In: *AAP* 516, A21, A21. DOI: 10.1051/0004-6361/201014740. arXiv: 1005.0620 [astro-ph.SR].

Schwarz, H. et al. (Sept. 2016). "The Slow Spin of the Young Substellar Companion GQ Lupi b and its Orbital Configuration." In: *Astronomy and Astrophysics* 593, A74, A74. DOI: 10.1051/0004-6361/201628908. arXiv: 1607.00012 [astro-ph.EP].

Scott, D. W. (2015). *Multivariate Density Estimation: Theory, Practice, and Visualization*.

Seager, S. (2010). *Exoplanets*.

Sepulveda, A. et al. (Jan. 2023). "TESS Analysis of The Directly Imaged Exoplanet Host Star HIP 65426." In: *American Astronomical Society Meeting Abstracts*. Vol. 55. American Astronomical Society Meeting Abstracts, 164.02, p. 164.02.

Shields, A. L., S. Ballard, and J. A. Johnson (Dec. 2016). "The habitability of planets orbiting M-dwarf stars." In: *Physics Reports* 663, p. 1. DOI: 10.1016/j.physrep.2016.10.003. arXiv: 1610.05765 [astro-ph.EP].

Smith, A. J. et al. (June 2022). "47 Ursae Majoris c: Exploring Roman CGI Observation Strategies." In: *Bulletin of the American Astronomical Society*. Vol. 54, 102.63, p. 102.63.

Snellen, I. et al. (Apr. 2014). "The Fast Spin-rotation of a Young Extrasolar Planet." In: *arXiv e-prints*, arXiv:1404.7506, arXiv:1404.7506. DOI: 10.48550/arXiv.1404.7506. arXiv: 1404.7506 [astro-ph.EP].

Soummer, R., L. Pueyo, and J. Larkin (Aug. 2012). "Detection and Characterization of Exoplanets and Disks Using Projections on Karhunen-Loève Eigenimages." In: *The Astrophysical Journal Letters* 755.2, L28, p. L28. DOI: 10.1088/2041-8205/755/2/L28. arXiv: 1207.4197 [astro-ph.IM].

Stassun, K. G. et al. (Dec. 2011). "The M4 Transition: Toward a Comprehensive Understanding of the Transition into the Fully Convective Regime." In: *16th Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun*. Ed. by C. Johns-Krull, M. K. Browning, and A. A. West. Vol. 448. Astronomical Society of the Pacific Conference Series, p. 505. arXiv: 1012.2580 [astro-ph.SR].

Steinmetz, T. et al. (Sept. 2008). "Laser Frequency Combs for Astronomical Observations." In: *Science* 321.5894, p. 1335. DOI: 10.1126/science.1161030. arXiv: 0809.1663 [astro-ph].

Stolker, T. et al. (Mar. 2020). "MIRACLES: Atmospheric Characterization of Directly Imaged Planets and Substellar Companions at 4-5 μm. I. Photometric Analysis of β Pic b, HIP 65426 b, PZ Tel B, and HD 206893 B." In: *Astronomy and Astrophysics* 635, A182, A182. DOI: 10.1051/0004-6361/201937159. arXiv: 1912.13316 [astro-ph.EP].

Suárez Mascareño, A. et al. (Dec. 2021). "Rapid Contraction of Giant Planets Orbiting the 20-Million-Year-Old Star V1298 Tau." In: *Nature Astronomy* 6, pp. 232–240. DOI: 10.1038/s41550-021-01533-7. arXiv: 2111.09193 [astro-ph.EP].

Sun, H. et al. (June 2022). "α-deep Probabilistic Inference (α-DPI): Efficient Uncertainty Quantification from Exoplanet Astrometry to Black Hole Feature Extraction." In: *The Astrophysical Journal* 932.2, 99, p. 99. DOI: 10.3847/1538-4357/ac6be9. arXiv: 2201.08506 [astro-ph.IM].

Tejada Arevalo, R., D. Tamayo, and M. Cranmer (June 2022). "Stability Constrained Characterization of the 23 Myr Old V1298 Tau System: Do Young Planets Form in Mean Motion Resonance Chains?" In: *The Astrophysical Journal Letters* 932.1, L12, p. L12. DOI: 10.3847/2041-8213/ac70e0. arXiv: 2203.02805 [astro-ph.EP].

van Leeuwen, F. (Nov. 2007). "Validation of the New Hipparcos Reduction." In: *Astronomy and Astrophysics* 474.2, pp. 653–664. DOI: 10.1051/0004-6361:20078357. arXiv: 0708.1752 [astro-ph].

Vanderburg, A. et al. (Sept. 2020). "A Giant Planet Candidate Transiting a White Dwarf." In: *Nature* 585.7825, pp. 363–367. DOI: 10.1038/s41586-020-2713-y. arXiv: 2009.07282 [astro-ph.EP].

Virtanen, P. et al. (2020). "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python." In: *Nature Methods* 17, pp. 261–272. DOI: 10.1038/s41592-019-0686-2.

Vogt, S. S. et al. (June 1994). "HIRES: the High-resolution Echelle Spectrometer on the Keck 10-m Telescope." In: *Proceedings of the SPIE*. Ed. by D. L. Crawford and E. R. Craine. Vol. 2198. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, p. 362. DOI: 10.1117/12.176725.

Vorobyov, E. I. and V. G. Elbakyan (Oct. 2018). "Gravitational Fragmentation and Formation of Giant Protoplanets on Orbits of Tens of au." In: *Astronomy and Astrophysics* 618, A7, A7. DOI: 10.1051/0004-6361/201833226. arXiv: 1806.07675 [astro-ph.SR].

Vousden, W. D., W. M. Farr, and I. Mandel (Jan. 2016). "Dynamic Temperature Selection for Parallel Tempering in Markov Chain Monte Carlo Simulations." In: *Monthly Notices of the Royal Astronomical Society* 455.2, pp. 1919–1937. DOI: 10.1093/mnras/stv2422. arXiv: 1501.05823 [astro-ph.IM].

Wagner, K. et al. (Aug. 2016). "Direct Imaging Discovery of a Jovian Exoplanet within a Triple-Star System." In: *Science* 353.6300, pp. 673–678. DOI: 10.1126/science.aaf9671. arXiv: 1607.02525 [astro-ph.EP].

Wang, J. J., M. Kulikauskas, and S. Blunt (Jan. 2021). *whereistheplanet: Predicting Positions of Directly Imaged Companions*. Astrophysics Source Code Library, record ascl:2101.003. ascl: 2101.003.

Wang, J. J. et al. (Nov. 2018). "Dynamical Constraints on the HR 8799 Planets with GPI." In: *The Astronomical Journal* 156.5, 192, p. 192. DOI: 10.3847/1538-3881/aae150. arXiv: 1809.04107 [astro-ph.EP].

Wang, J. J. et al. (June 2020). "Keck/NIRC2 L'-band Imaging of Jovian-mass Accreting Protoplanets around PDS 70." In: *The Astronomical Journal* 159.6, 263, p. 263. DOI: 10.3847/1538-3881/ab8aef. arXiv: 2004.09597 [astro-ph.EP].

Winn, J. N. (Jan. 2010). "Transits and Occultations." In: *arXiv e-prints*, arXiv:1001.2010, arXiv:1001.2010. DOI: 10.48550/arXiv.1001.2010. arXiv: 1001.2010 [astro-ph.EP].

Wolszczan, A. and D. A. Frail (Jan. 1992). "A Planetary System around the Millisecond Pulsar PSR1257 + 12." In: *Nature* 355.6356, pp. 145–147. DOI: 10.1038/355145a0.

Xuan, J. W. and M. C. Wyatt (Sept. 2020). "Evidence for a High Mutual Inclination Between the Cold Jupiter and Transiting Super Earth Orbiting π Men." In: *Monthly Notices of the Royal Astronomical Society* 497.2, pp. 2096–2118. DOI: 10.1093/mnras/staa2033. arXiv: 2007.01871 [astro-ph.EP].

Xuan, J. W. et al. (Oct. 2022). "A Clear View of a Cloudy Brown Dwarf Companion from High-resolution Spectroscopy." In: *The Astrophysical Journal* 937.2, 54, p. 54. DOI: 10.3847/1538-4357/ac8673. arXiv: 2208.01657 [astro-ph.EP].

Yu, L. et al. (Nov. 2019). "Magnetic Field, Activity, and Companions of V410 Tau." In: *Monthly Notices of the Royal Astronomical Society* 489.4, pp. 5556–5572. DOI: 10.1093/mnras/stz2481. arXiv: 1909.01199 [astro-ph.SR].

Zeng, L. et al. (May 2019). "Growth Model Interpretation of Planet Size Distribution." In: *Proceedings of the National Academy of Science* 116.20, pp. 9723–9728. DOI: 10.1073/pnas.1812905116. arXiv: 1906.04253 [astro-ph.EP].

Zhang, J. et al. (Sept. 2021). "Long-period Jovian Tilts the Orbits of Two sub-Neptunes Relative to Stellar Spin Axis in Kepler-129." In: *The Astronomical Journal* 162.3, 89, p. 89. DOI: 10.3847/1538-3881/ac0634. arXiv: 2105.03446 [astro-ph.EP].

Zhu, W. et al. (June 2018). "About 30% of Sun-like Stars Have Kepler-like Planetary Systems: A Study of Their Intrinsic Architecture." In: *The Astrophysical Journal* 860.2, 101, p. 101. DOI: 10.3847/1538-4357/aac6d5. arXiv: 1802.09526 [astro-ph.EP].

Zicher, N. et al. (May 2022). "One Year of AU Mic with HARPS - I. Measuring the Masses of the Two Transiting Planets." In: *Monthly Notices of the Royal Astronomical Society* 512.2, pp. 3060–3078. DOI: 10.1093/mnras/stac614. arXiv: 2203.01750 [astro-ph.EP].