UNLV University Libraries
University of Nevada, Las Vegas

UNLV Theses, Dissertations, Professional Papers, and Capstones

December 2018

Bayesian Model Comparisons in Planetary Science

John Henry Boisvert
boisver3@unlv.nevada.edu

Follow this and additional works at: https://digitalscholarship.unlv.edu/thesesdissertations

Part of the Astrophysics and Astronomy Commons, and the Other Physics Commons

Repository Citation

Boisvert, John Henry, "Bayesian Model Comparisons in Planetary Science" (2018). *UNLV Theses, Dissertations, Professional Papers, and Capstones*. 3404.
https://digitalscholarship.unlv.edu/thesesdissertations/3404

This Dissertation is protected by copyright and/or related rights. It has been brought to you by Digital Scholarship@UNLV with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you need to obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/or on the work itself.

This Dissertation has been accepted for inclusion in UNLV Theses, Dissertations, Professional Papers, and Capstones by an authorized administrator of Digital Scholarship@UNLV. For more information, please contact digitalscholarship@unlv.edu.

BAYESIAN MODEL COMPARISONS IN PLANETARY SCIENCE

by

John Henry Boisvert

Bachelor of Science — Physics
University of Nevada, Las Vegas
2010

Master of Science — Physics
University of Nevada, Las Vegas
2013

A dissertation submitted in partial fulfillment
of the requirements for the

Doctor of Philosophy — Astronomy

Department of Physics and Astronomy
College of Sciences
The Graduate College

University of Nevada, Las Vegas
December 2018

Copyright 2019 by John Henry Boisvert
All Rights Reserved

UNLV | GRADUATE COLLEGE

Dissertation Approval

The Graduate College
The University of Nevada, Las Vegas

November 15, 2018

This dissertation prepared by

John Henry Boisvert

entitled

Bayesian Model Comparison in Planetary Science

is approved in partial fulfillment of the requirements for the degree of

Doctor of Philosophy – Astronomy
Department of Physics and Astronomy

Jason Steffen, Ph.D.
Examination Committee Chair

Kathryn Hausbeck Korgan, Ph.D.
Graduate College Interim Dean

Daniel Proga, Ph.D.
Examination Committee Member

Rebecca Martin, Ph.D.
Examination Committee Member

George Rhee, Ph.D.
Examination Committee Member

Rodney Metcalf, Ph.D.
Graduate College Faculty Representative

Abstract

Model comparison in the modern era allows us to use statistical methods that were previously difficult with older machines. I present a state-of-the-art model comparison code that uses modern Bayesian statistics to measure the Bayes factor between two competing models. The Bayes factor is the ratio of the probability of the data given one model to the probability of the data given another model. My code was used to compare models in five problems in planetary science. The first three pertain to radial velocity exoplanet data. There is a degeneracy in the radial velocity exoplanet signal between a single planet on an eccentric orbit and a two-planet system with a period ratio of 2:1. This degeneracy could lead to misunderstandings of the dynamical histories of planetary systems as well as measurements of planetary abundances if the correct architecture is not established. We constrain the rate of mischaracterization by analyzing a sample of 60 non-transiting, radial velocity systems orbiting main sequence stars from the NASA Exoplanet Archive (NASA Archive) using my model comparison pipeline. We find that 15 systems (25% of our sample) show compelling evidence for the two-planet case with a confidence level of 95%. The *Automated Planet Finder* obtained additional data for seven of the best candidates. My pipeline finds that six of them continue to show strong evidence for the two-planet case. Observational strategies to break the 2:1 degeneracy are explored using two thousand synthetically generated single planets with eccentric orbits and two planets with circular orbits. We find that focusing on taking observations where the degeneracy is the weakest decreases the ambiguity between the models more than taking observations at random phases. The final two problems are model comparison of high-pressure/high-temperature experimental data. My code is able to identify two phase transitions in pressure-temperature water-ice data taken at Argonne National Lab in two separate datasets. One of the phase transitions found—cubic ice-VII to tetragonal ice-$\mathrm{VII_t}$—is previously unreported until now. My code sees these phase transitions in X-ray diffraction data, which uses Bragg's law to peer into the crystal lattice of water, and in pressure–volume equation of state fits.

Acknowledgements

"If I have seen further it is by standing on the shoulders of Giants."

— Sir Issac Newton, 1675

I must start by thanking Dr. Jason Steffen for providing me with the opportunity to work with you after already starting my Ph.D. studies. Thank you for showing me the world of exoplanets, research, and problem solving. I am grateful that you sent me to the 2016 Sagan summer workshop, the 2017 Aspen winter conference, and the winter AAS meeting in 2018. The experiences from working together have given me what's necessary to start my career. It's hard to put into words how much I've learned from you. Thank you so much.

Thank you to my love and fiancée Mariela Preciado for keeping me sane during the rough times. You're overwhelming support gave me the foundation necessary to stay focused through the end. I also express my thanks and love to my mother and father, Barbara and Louis Boisvert. Without your unconditional love and support, I would not have made it anywhere near where I stand today. Thank you so much for everything.

Thank you to Gail Michel-Parsons, Natasha Korceba, and Cindy Lee for all of your help with paper work, unlocking doors, and everything else you all do to keep the lights on in the department. You all are not thanked enough. I would also like to thank my committee members, Dr. George Rhee, Dr. Daniel Proga, Dr. Rebecca Martin, and Dr. Rodney Metcalf for their support and encouragement. Thanks for sticking with the process until the end. Thank you to Keala Kiko for being a great bridge between the graduate students and the graduate college. And I cannot forget all of the graduate students and postdoctoral researchers who I've had great intellectual conversations with over the past couple years. Talking with you all about the latest topics in the field has been a rewarding experience and brings everything into perspective.

Finally, I am forever in debt to Dr. Lon Spight and Dr. Jim Selser for showing me the wonders of physics and for being positive forces of encouragement before and during my graduate career. You both have changed my life in numerous ways and will be missed greatly.

Table of Contents

Abstract iii

Acknowledgements iv

List of Tables vii

List of Figures xi

List of Abbreviations xii

1 Introduction 1
- 1.1 Exoplanet Historical Overview . . . 2
  - 1.1.1 Architectures of Exoplanets . . . 2
    - 1.1.1.1 Radial Velocity . . . 3
  - 1.1.2 Planetary Density Estimates . . . 3
- 1.2 Dissertation Overview . . . 3

2 The Pipeline 4
- 2.1 The Pipeline Algorithm . . . 5
- 2.2 Measuring the Fully Marginalized Likelihood . . . 6

3 Radial Velocity Exoplanets 8
- 3.1 The Models . . . 8
  - 3.1.1 Single Eccentric . . . 8
  - 3.1.2 Circular Double . . . 9
  - 3.1.3 Degeneracy at the 2:1 Mean-Motion Resonance . . . 9
- 3.2 Preliminary Sample . . . 9
  - 3.2.1 Models Tested . . . 11
  - 3.2.2 Pipeline Set-up . . . 11
  - 3.2.3 Pipeline Characterization . . . 11
    - 3.2.3.1 Synthetic Results . . . 12
- 3.3 Real System Analysis and Results . . . 12
  - 3.3.1 Real Results . . . 40
- 3.4 Discussion . . . 42
  - 3.4.1 Follow-up Observations with the Automated Planet Finder . . . 42

4 Modeling Stellar Noise **75**
4.1 Stellar Activity . . . 75
4.2 Modeling Stellar Noise: Gaussian Processes . . . 75
4.2.1 Kernel Selection and New Likelihood Function . . . 75
4.3 New Synthetic Sample . . . 76
4.3.1 The Synthetic Recipe . . . 76
4.4 Preliminary Results . . . 78

5 Observation Strategy: Optimum Phases or More Data? **83**
5.1 Breaking the Degeneracy: Optimum Phases . . . 83
5.1.1 Optimum Phase Parameter-Space . . . 84
5.2 Set-up . . . 91
5.3 Results . . . 91
5.4 Conclusions . . . 105

6 The Shape of Water: High–Pressure/High–Temperature Data Modeling **106**
6.1 Bragg's Law . . . 106
6.1.1 Pipeline Set-Up . . . 107
6.1.2 Results . . . 107
6.2 Equation of State . . . 107
6.2.1 Single-Phase Birch-Murnaghan . . . 110
6.2.1.1 Pipeline Set-Up . . . 110
6.2.2 Single-Phase Rose-Vinet . . . 110
6.2.2.1 Pipeline Set-Up . . . 110
6.2.3 Two-Phase Birch-Murnaghan . . . 111
6.2.3.1 Pipeline Set-Up . . . 111
6.2.4 Two-Phase Rose-Vinet . . . 111
6.2.4.1 Pipeline Set-Up . . . 112
6.2.5 Triple-Phase Birch-Murnaghan . . . 112
6.2.5.1 Pipeline Set-Up . . . 112
6.2.6 Triple-Phase Rose-Vinet . . . 113
6.2.6.1 Pipeline Set-Up . . . 113
6.2.7 Results . . . 113
6.3 Conclusions . . . 114

7 Discussion and Conclusions **126**

References **128**

Curriculum Vitae **134**

List of Tables

3.1 The preferred fixed circular double models of the extended sample that has Bayes factor larger than the 95th percentile of the synthetic systems. 39
3.2 *Automated Planet Finder* target list. 44
3.3 Results with new data from *Automated Planet Finder*. 45

6.1 Table of Bayes factors from the model comparison between the cubic and tetragonal unit cell. 109
6.2 Fit parameters for the single-phase Burch-Murnaghan Equation of State. . . 115
6.3 Fit parameters for the single-phase Rose-Vinet Equation of State. 116
6.4 Fit parameters for the two-phase Burch-Murnaghan Equation of State. . . . 117
6.5 Fit parameters for the two-phase Rose-Vinet Equation of State. 118
6.6 Fit parameters for the three-phase Burch Murnaghan Equation of State, fixing B'_o to 4. 119
6.7 Fit parameters for the three-phase Rose-Vinet Equation of State, fixing B'_o to 4.121
6.8 Fit parameters for the three-phase Burch Murnaghan Equation of State. . . 123
6.9 Fit parameters for the three-phase Rose-Vinet Equation of State. 124
6.10 Bayes factor square between all EOS models fit to the pressure/volume data 125

List of Figures

3.1 Stellar effective temperature vs. stellar surface gravity plot for the RV multi-planet systems from the NASA Archive, the *Kepler* multi-planet systems, and the sample of 95 stars. . . . 10
3.2 Property distributions for the sample of 95 real systems from the NASA Exoplanet Archive and the 1,000 synthetic time series. . . . 13
3.3 The log Bayes factor distribution for the 1,000 synthetic single eccentric time series and 60 real systems hosted by main sequence stars. . . . 14
3.4 HD 240237 time series with the circular double at 2.17:1 model. . . . 17
3.5 HD 240237 phase plots for the outer and inner planet of the circular double model fixed at 2.17:1. . . . 18
3.6 HD 240237 time series with the single eccentric model. . . . 19
3.7 HD 240237 phase plot the single eccentric model. . . . 20
3.8 HD 2952 time series with the circular double at 2.17:1 model. . . . 21
3.9 HD 2952 phase plots for the outer and inner planet of the circular double model fixed at 2.17:1. . . . 22
3.10 HD 2952 time series with the single eccentric model. . . . 23
3.11 HD 2952 phase plot the single eccentric model. . . . 24
3.12 α Arietis time series with the circular double at 2.17:1 model. . . . 25
3.13 α Arietis phase plots for the outer and inner planet of the circular double model fixed at 2.17:1. . . . 26
3.14 α Arietis time series with the single eccentric model. . . . 27
3.15 α Arietis phase plot the single eccentric model. . . . 28
3.16 HD 96127 time series with the circular double at 2.17:1 model. . . . 29
3.17 HD 96127 phase plots for the outer and inner planet of the circular double model fixed at 2.17:1. . . . 30
3.18 HD 96127 time series with the single eccentric model. . . . 31
3.19 HD 96127 phase plot the single eccentric model. . . . 32
3.20 HD 95089 time series with the circular double at 2.17:1 model. . . . 33
3.21 HD 95089 phase plots for the outer and inner planet of the circular double model fixed at 2.17:1. . . . 34
3.22 HD 95089 time series with the single eccentric model. . . . 35
3.23 HD 95089 phase plot the single eccentric model. . . . 36
3.24 Orbital period vs. planetary mass for all RV planets. . . . 37
3.25 Mass ratio distribution for all RV adjacent planet pairs. . . . 38

3.26 The posterior distributions for the period ratio when considering the inner planet period as a free parameter. 41
3.27 HD 2952 time series with new data from the *Automated Planet Finder* and the best fit circular double model. 47
3.28 HD 2952 phase plots with new data from the *Automated Planet Finder* and the best fit circular double model for the outer and inner planet. 48
3.29 HD 2952 time series with new data from the *Automated Planet Finder* and the best fit single eccentric model. 49
3.30 HD 2952 phase plot with new data from the *Automated Planet Finder* and the best fit single eccentric model. 50
3.31 α Arietis time series with new data from the *Automated Planet Finder* and the best fit circular double model. 51
3.32 α Arietis phase plots with new data from the *Automated Planet Finder* and the best fit circular double model for the outer and inner planet. 52
3.33 α Arietis time series with new data from the *Automated Planet Finder* and the best fit single eccentric model. 53
3.34 α Arietis phase plot with new data from the *Automated Planet Finder* and the best fit single eccentric model. 54
3.35 HD 32518 time series with new data from the *Automated Planet Finder* and the best fit circular double model. 55
3.36 HD 32518 phase plots with new data from the *Automated Planet Finder* and the best fit circular double model for the outer and inner planet. 56
3.37 HD 32518 time series with new data from the *Automated Planet Finder* and the best fit single eccentric model. 57
3.38 HD 32518 phase plot with new data from the *Automated Planet Finder* and the best fit single eccentric model. 58
3.39 HD 81688 time series with new data from the *Automated Planet Finder* and the best fit circular double model. 59
3.40 HD 81688 phase plots with new data from the *Automated Planet Finder* and the best fit circular double model for the outer and inner planet. 60
3.41 HD 81688 time series with new data from the *Automated Planet Finder* and the best fit single eccentric model. 61
3.42 HD 81688 phase plot with new data from the *Automated Planet Finder* and the best fit single eccentric model. 62
3.43 14 Andromedae time series with new data from the *Automated Planet Finder* and the best fit circular double model. 63
3.44 14 Andromedae phase plots with new data from the *Automated Planet Finder* and the best fit circular double model for the outer and inner planet. 64
3.45 14 Andromedae time series with new data from the *Automated Planet Finder* and the best fit single eccentric model. 65
3.46 14 Andromedae phase plot with new data from the *Automated Planet Finder* and the best fit single eccentric model. 66
3.47 γ_1 Leonis time series with new data from the *Automated Planet Finder* and the best fit circular double model. 67

3.48 γ_1 Leonis phase plots with new data from the *Automated Planet Finder* and the best fit circular double model for the outer and inner planet. 68
3.49 γ_1 Leonis time series with new data from the *Automated Planet Finder* and the best fit single eccentric model. 69
3.50 γ_1 Leonis phase plot with new data from the *Automated Planet Finder* and the best fit single eccentric model. 70
3.51 HD 86081 time series with new data and the best fit circular double model. . 71
3.52 HD 86081 phase plots with new data and the best fit circular double model for the outer and inner planet. 72
3.53 HD 86081 time series with new data and the best fit single eccentric model. . 73
3.54 HD 86081 phase plot with new data and the best fit single eccentric model. . 74

4.1 The log Bayes factor distribution for the 2,000, ten-year synthetic time series, considering single eccentric and fixed circular double models. 81
4.2 The log Bayes factor distribution for the 2,000, ten-year synthetic time series, considering all models. 82

5.1 Single eccentric model with $K = 10$ m s^{-1}, $e = 0.15$, $\omega = 1.4$ rad and its equivalent 2:1 circular double model 85
5.2 Single eccentric model with $K = 10$ m s^{-1}, $e = 0.6$, $\omega = 1.4$ rad and its equivalent 2:1 circular double model. 86
5.3 The solutions to Equation 5.1.0.2. 87
5.4 The solutions to Equation 5.1.0.2. 88
5.5 The solutions to Equation 5.1.0.2. 89
5.6 The solutions to Equation 5.1.0.2. 90
5.7 The log Bayes factor distribution for the sixteen-year synthetic single eccentric time series, considering the fixed circular double models and the single eccentric model, and doubling the data. 93
5.8 The log Bayes factor distribution for the sixteen-year synthetic single eccentric time series, considering all models and doubling the data. 94
5.9 The log Bayes factor distribution for the sixteen-year synthetic circular double time series, considering the fixed circular double models and the single eccentric model, and doubling the data. 95
5.10 The log Bayes factor distribution for the sixteen-year synthetic circular double time series, considering all models and doubling the data. 96
5.11 The log Bayes factor distribution for the sixteen-year synthetic single eccentric time series, considering the fixed circular double models and the single eccentric model, and bringing the number of data to twice the maximum. . . 97
5.12 The log Bayes factor distribution for the sixteen-year synthetic single eccentric time series, considering all models and bringing the number of data to twice the maximum. 98
5.13 The log Bayes factor distribution for the sixteen-year synthetic circular double time series, considering the fixed circular double models and the single eccentric model, and bringing the number of data to twice the maximum. . . 99

5.14 The log Bayes factor distribution for the sixteen-year synthetic circular double time series, considering all models and bringing the number of data to twice the maximum. 100
5.15 The log Bayes factor distribution for the sixteen-year synthetic single eccentric time series, considering the fixed circular double models and the single eccentric model, and bringing the number of data to match the maximum. . 101
5.16 The log Bayes factor distribution for the sixteen-year synthetic single eccentric time series, considering all models and bringing the number of data to match the maximum. 102
5.17 The log Bayes factor distribution for the sixteen-year synthetic circular double time series, considering the fixed circular double models and the single eccentric model, and bringing the number of data to match the maximum. . 103
5.18 The log Bayes factor distribution for the sixteen-year synthetic circular double time series, considering all models and bringing the number of data to match the maximum. 104

6.1 Plot of the Bayes factors from the model comparison between the cubic and tetragonal unit cell. 108
6.2 Single-phase Burch-Murnaghan Equation of State Fit. 115
6.3 Single-phase Rose-Vinet Equation of State Fit. 116
6.4 Two-phase Burch-Murnaghan Equation of State Fit. 117
6.5 Two-phase Rose-Vinet Equation of State Fit. 118
6.6 Three-phase Burch-Murnaghan Equation of State Fit, fixing B'_o to 4. 120
6.7 Three-phase Rose-Vinet Equation of State Fit, fixing B'_o to 4. 122
6.8 Three-phase Burch-Murnaghan Equation of State Fit. 123
6.9 Three-phase Rose-Vinet Equation of State Fit, fixing B'_o to 4. 124

List of Abbreviations

APF	*Automated Planet Finder*
BF	Bayes Factor
EOS	Equation of State
FML	Fully Marginalized Likelihood
GP	Gaussian Process
LBF	Logarithm of the Bayes Factor
MCMC	Markov Chain-Monte Carlo
MLE	Maximum Likelihood Estimation
MMR	Mean-Motion Resonances
NASA	National Aeronautics and Space Administration
NASA Archive	NASA Exoplanet Archive
RV	Radial Velocity

Chapter 1

Introduction

Model comparison is the most basic scientific exercise. If there are two models that fit a set of data reasonably well, then the goal of the scientific exercise becomes determining the model fits the data better. But, what does "better" mean? One approach is to measure a goodness-of-fit for each model, and the comparison of the calculated statistics determine which model more likely fits the data. The classic goodness-of-fit that is taught at the undergraduate physics level is Pearson's χ^2 statistic, first introduced in 1900 [1]:

$$\chi^2(\mathrm{d}(x), \mathcal{M}(x, \vartheta)) = \sum_{i=1}^{\mathrm{N}} \frac{[\mathrm{d}(x_i) - \mathcal{M}(x_i, \vartheta)]^2}{\sigma(x_i)}, \tag{1.0.0.1}$$

where $\mathrm{d}(x)$ is the data, $\sigma(x)$ is the variance, $\mathcal{M}(x, \vartheta)$ is the model as a function of its parameters, N is the number of data points, and the independent variable is x.

However, this statistic has some caveats. For instance, there lacks a way to penalize datasets of varying size. The reduced χ^2 is used to get around this, which takes into account the number of degrees of freedom in the fit:

$$\chi^2_{\mathrm{red}}(\mathrm{d}(x), \mathcal{M}(\vartheta)) = \frac{\chi^2(\mathrm{d}(x), \mathcal{M}(\vartheta))}{\nu}, \tag{1.0.0.2}$$

where ν is the degrees of freedom and is calculated as the difference between the number of data points and the number of model parameters. Another caveat to consider is that this statistic assumes the errors in the dataset—i.e. $\sigma_i(x)$—are all independent to each other and Gaussian distributed. This likely not the case in every problem because noise is often correlated. In addition, it assumes the correlations between the model parameters are Gaussian. This might not be the case for some pairs of model parameters.

In 1763, Reverend Thomas Bayes formulated an equation that describes the probability of a set of model parameters given the data and the model. His equation also quantifies prior knowledge about the system and includes it when computing the probability [2]. Bayes was not able to utilize his theorem during his lifetime. However, with today's advancements in computational technology, his theorem became the basis of Bayesian statistics. Using Bayes statistics in combination with the Markov-Chain Monte Carlo method, we can measure a goodness-of-fit that does take into account correlations between model parameters, the dependence of the errors in the dataset, and—if desired—incorporates prior knowledge about the problem.

1.1 Exoplanet Historical Overview

While it may seem like a part of the natural order of the universe for planets to be found around stars, the first confirmed planet outside the solar system was not discovered until relatively recently. Wolszczan and Frail [3] found a planet orbiting around the millisecond pulsar PSR1257+12 in 1992 using the radial velocity (RV) technique. The radial velocity of a star is computed by measuring the shifts in the stellar absorption features as it orbits around the barycenter of the system [4]. These shifts are due to the presence of a planet(s) gravitationally tugging at the star. This technique can measure the orbital period, eccentricity, and sky-projected mass ($M \sin i$) of the planets in the system. This same method is used to measure spectroscopic stellar binary systems [4]. The first planet detected around a main sequence star was in 1995 by Mayor and Queloz [5] using the same technique. The first multi-planet system was found around υ AndromedæA, the primary of a binary system, by Butler et al. in 1999 with a detection of three planets [6]. The inner-most planet was discovered in 1997 [7]. In 2010, Curiel et al. detected a fourth planet [8]. All of these detections were using the radial velocity technique.

There are other techniques for detecting exoplanets. The transit technique measures the dimming and re-brightening of starlight received as a planet passes in front of it [9]. This technique can measure the orbital period and the planet radius. The first transiting exoplanet is HD 209458b, seen in 2000 [10, 11], however this planet was originally detected via the radial velocity technique. Gravitational micro-lensing is another detection technique. In general, when a massive foreground object passes in front of a background object, the distortion of space-time caused by the foreground object makes it act like a lens [12]. The image of background object becomes elongated and magnified based on the shape of the lens. If the foreground star hosts a planet, then the micro-lensed image will become distorted. Information about the planet can be measured from these deviations [13]. In 2006, Beaulieu et al. discovered OGLE-2005-BLG-390Lb, the first planet discovered via micro-lensing [14]. Direct imaging of exoplanets is the final detection technique covered here [15]. Observers block the starlight with chronographs in order to view the planets. Fomalhaut b was the first directly imaged planet taken by Kalas et al. [16] in 2008.

1.1.1 Architectures of Exoplanets

The architectures of planetary systems give insight into their formation and dynamical histories. For example, interactions with the protoplanetary disk tend to drive adjacent planets into first-order, mean-motion resonances (MMRs, such as the 2:1), while simultaneously damping their eccentricities to values that are difficult to measure [17, 18]. On the other hand, planet-planet scattering [19, 20] or Kozai-Lidov oscillations [21, 22, 23] can produce single planets with eccentric orbits. While not all planetary systems must pass through these phases of disk migration or eccentricity growth, the system architectures that they produce rarely occur from in-situ formation. Thus, reliable estimates of the frequencies of different architectures will reveal the relative importance of these processes in planet formation and evolution in general.

1.1.1.1 Radial Velocity

For RV observations in particular, the challenge in identifying the true system architecture is a degeneracy between two models—one with a single planet with eccentric orbits (single eccentrics) and one with two planets with circular orbits at the 2:1 (circular doubles) [24, 25]. Historically, the single-planet model has been favored on the grounds of Occam's razor [26], since a system with a single planet is simpler than a system with two. However, the circular double model has the same number of model parameters as single eccentrics (it is just as simple) and it is a consequence of dynamical processes known to occur. These facts motivate careful scrutiny of existing discoveries in order to properly characterize the systems. If circular doubles are more common than currently suggested, then disk-migration may be more important than previously thought [18].

This degeneracy is widely known though rarely addressed. Nevertheless, there is precedent for reconsidering certain systems. For example, Kürster et al. [26] reanalyzed RV data for HD 27894 and found that a circular double model was a better fit than the reported single eccentric model. Also, Angelada et al. and Wittenmyer et al.[24, 25] found similar results for several RV systems. At the same time, new measurements from the *Kepler* mission show that planet pairs near 2:1 are quite common. For example, using the method of Steffen et al. [27] on the *Kepler* DR25 catalog [28], we estimate that 20% of *Kepler*'s transiting adjacent planet pairs with period ratios between 1 and 6 are within ten percent of 2—including the most prominent peak of the period ratio distribution at 2.17 [27]. Motivated by these new facts and the results of previous studies, Chapter 3 begins with reanalyzing a sample of 60 single eccentric planetary systems using a new Bayesian analysis pipeline introduced in Chapter 2.

1.1.2 Planetary Density Estimates

Planets that have both RV and transit data are able to have their densities estimated [10, 11]. Planetary composition models are used to model the different depths as realistic as possible [29]. However, accurate measurements of the equation of state (EOS) for the included compounds are required. In this dissertation, we focus on the results of measuring water-ice data from a high-pressure/high-temperature X-ray diffraction experiment to measure its EOS.

1.2 Dissertation Overview

The chapters of my dissertation cover the following topics: Chapter 2 introduces the model comparison pipeline, the likelihood, priors, and the fully marginalized likelihood. Chapter 3 focuses on my work from Boisvert et al. [30] and follow-up observations using the *Automated Planet Finder* (*APF*). Chapter 4 introduces adding realistic stellar noise to synthetic systems and the results of analyzing them with a common observational technique. Chapter 5 investigates different observational campaigns in an effort to find the most efficient way to break the 2:1 degeneracy. Chapter 6 describes my model comparison pipeline analyzing two high-temperature/high-pressure datasets relating to a newly discovered phase of water-ice at high pressures. Finally, Chapter 7 is a short conclusion.

Chapter 2

The Pipeline

The model comparison pipeline utilizes Bayes' Theorem to determine the probability of some data given a model and other prior knowledge about the problem. Bayes' Theorem is as follows:

$$\mathcal{P}(\vartheta|\mathrm{d},\mathcal{M}) = \frac{\mathcal{P}(\vartheta|\mathcal{M}) \times \mathcal{P}(\mathrm{d}|\vartheta,\mathcal{M})}{\mathcal{P}(\mathrm{d}|\mathcal{M}),} \tag{2.0.0.1}$$

where $\mathcal{P}$ is the probability, d is the data, $\mathcal{M}$ is the model, and ϑ are the parameters of $\mathcal{M}$. $\mathcal{P}(\mathrm{d}|\vartheta,\mathcal{M})$ is the likelihood, $\mathcal{P}(\vartheta|\mathcal{M})$ is the prior, $\mathcal{P}(\vartheta|\mathrm{d},\mathcal{M})$ is the posterior probability distribution of the parameters, and $\mathcal{P}(\mathrm{d}|\mathcal{M})$ is the fully marginalized likelihood (FML, a.k.a. the Bayesian evidence). Before going through the pipeline algorithm, let's define the likelihood function and the various priors that the algorithm is currently built to use.

The likelihood function—$\mathcal{P}(\mathrm{d}|\vartheta,\mathcal{M})$—is the probability of a dataset given the model and the model parameters. The likelihood function we use is based on an assumption that the errors are Gaussian distributed [31]. With normalization, it is:

$$\ln \mathcal{P}(\mathrm{d}|\vartheta,\mathcal{M}) = -\frac{1}{2}\sum_n \left[\frac{(\mathrm{d}_n - \mathcal{M}(\vartheta,x_n))^2}{\sigma_{\mathrm{d},n}^2} + \ln 2\pi\,\sigma_{\mathrm{d},n}^2\right], \tag{2.0.0.2}$$

where σ_d are the measurement errors. The likelihood is equivalent to negative χ^2—Equation 1.0.0.1. The likelihood is weighted by the individual errors in each data point.

The prior—$\mathcal{P}(\vartheta|\mathcal{M})$—represents all prior knowledge known about the system. The purpose of the prior is to adjust likelihood distribution from Gaussian if there is any reason for it to deviate. A prior that does this is an informative prior. An uninformed prior allows the posterior probability distribution to look similar to the likelihood distribution. The four types of priors used in the model comparison code are uniform, modified Jeffreys, Gaussian, and Rayleigh.

The uniform prior is the most basic prior. It gives the parameter an equal probability of being between a range of values. This prior is normalizable and objective. It is given by the following:

$$\ln \mathcal{P}(\vartheta|\mathcal{M}) = -\ln\left(\vartheta_\mathrm{max} - \vartheta_\mathrm{min}\right), \tag{2.0.0.3}$$

where ϑ_max and ϑ_min are the maximum and minimum limits on the parameter value.

The Jeffreys prior, first introduced by Sir Harold Jeffreys in 1946 [32], is used when the scale of the parameter is not known. It is given by the following:

$$\ln \mathcal{P}(\vartheta|\mathcal{M}) = -\ln\left[\vartheta \times \ln\left(\frac{\vartheta_{\rm max}}{\vartheta_{\rm min}}\right)\right], \tag{2.0.0.4}$$

where $\vartheta_{\rm max}$ and $\vartheta_{\rm min}$ are the maximum and minimum limits on the parameter value. However, this prior is not normalizable because it can not take zero for a minimum value. In 1996, Berger and Pericchi [33] formed the modified Jeffreys prior, a Jeffreys prior that turns into a uniform prior at small value—ϑ_0. Thus, it is used when a scalable parameter could have zero for a minimum value. It is also normalizable and objective. It is given by the following:

$$\ln \mathcal{P}(\vartheta|\mathcal{M}) = -\ln\left[(\vartheta_0 + \vartheta) \times \ln\left(1 + \frac{\vartheta_{\rm max}}{\vartheta_0}\right)\right], \tag{2.0.0.5}$$

where $\vartheta_{\rm max}$ is the maximum limit on the parameter value and ϑ_0 is where the uniform prior takes over the Jeffreys prior.

The Gaussian prior is used when the value of the parameter and the error are known in advance. For example, the parameter could be constrained from a different dataset. This prior is normalizable but not objective. The prior probability is penalized the further the parameter is from the mean value of the parameter—$\vartheta_{\rm mean}$. The error in the parameter is used to inform the width of the Gaussian—$\vartheta_{\rm width}$. The prior is given by the following:

$$\ln \mathcal{P}(\vartheta|\mathcal{M}) = -\ln\left[\sqrt{2\pi}\,\vartheta_{\rm width}\right] - \frac{1}{2}\left(\frac{\vartheta - \vartheta_{\rm mean}}{\vartheta_{\rm width}}\right). \tag{2.0.0.6}$$

The Rayleigh prior is used in many areas of science. A couple examples are in radar imaging to measure sea winds [34] and in speech algorithms to decrease the signal-to-noise [35]. This prior is normalizable and objective. The Rayleigh prior is described by the Rayleigh parameter—$\vartheta_{\rm R}$— that alters the shape and range of the distribution. The value of chosen based on the range of parameter-values. The prior is given by the following:

$$\ln \mathcal{P}(\vartheta|\mathcal{M}) = -\ln\left[\frac{\vartheta}{\vartheta_{\rm R}^2} \times \exp\left(-\frac{\vartheta^2}{\vartheta_{\rm R}^2}\right)\right]. \tag{2.0.0.7}$$

In principle, any function can be used to build a prior.

2.1 The Pipeline Algorithm

The model fitting procedure is a three step process. First, we determine the starting seed for our Markov Chain-Monte Carlo (MCMC) by doing a maximum likelihood estimation (MLE). Since the likelihood is equivalent to χ^2, this step is equivalent to minimizing χ^2.

The second step in our pipeline estimates the posterior distributions of the model parameters using an ensemble sampler MCMC, *emcee* [36]. Each run has a number of Markov chains equal to five times the number of model parameters (plus one if the result if odd), thins the chains every hundred steps, and ignores the first 20% of the chain as burn-in. The

chains evolve until they yield a set of at least ten thousand independent samples per model per dataset. We measure the autocorrelation length after each run to determine the number of independent samples. If the number of independent samples falls short of ten thousand, then the autocorrelation length is used to determine how many additional steps are needed to yield ten thousand independent samples and the MCMC is rerun with the new number of steps. The different chains were initialized using the parameter values from the MLE, with each parameter scattered by a sufficiently small amount to allow the ensemble sampler to fill the posterior mode. The resulting posterior distributions yield accurate, correlated errors on the model parameters. Finally, our procedure quantitatively compares models by calculating the Bayes factor (BF)—the ratio of the probabilities of the data given two competing models:

$$\mathrm{BF} = \frac{\mathcal{P}(\mathrm{d}|\mathcal{M}_a)}{\mathcal{P}(\mathrm{d}|\mathcal{M}_b)}, \tag{2.1.0.1}$$

where $\mathcal{P}(\mathrm{d}|\mathcal{M}_i)$ is the FML for model i. The Bayes factor also accounts for the number of model parameters, that means it can be used for comparing any two models—even those of varying complexity. The FML is approximated using an importance sampling algorithm where the sampling distribution is informed by a set of posterior samples taken from the aforementioned MCMC [37, 30].

2.2 Measuring the Fully Marginalized Likelihood

In this context, importance sampling is essentially a general form of Monte Carlo integration to estimate the fully marginalized likelihood, $\mathcal{P}(\mathrm{d}|\mathcal{M})$ in Equation 2.0.0.1. The value of $\mathcal{P}(\mathrm{d}|\mathcal{M})$ is the integral over the prior probability distribution $\mathcal{P}(\vartheta|\mathcal{M})$ times the likelihood function $\mathcal{P}(\mathrm{d}|\vartheta, \mathcal{M})$, i.e.,

$$\mathcal{P}(\mathrm{d}|\mathcal{M}) = \int \mathcal{P}(\vartheta|\mathcal{M})\mathcal{P}(\mathrm{d}|\vartheta, \mathcal{M})d\vartheta \tag{2.2.0.1}$$

We multiply the numerator and denominator of the integrand by $g(\vartheta)$, a distribution over the model parameters with a known normalization.

$$\mathcal{P}(\mathrm{d}|\mathcal{M}) = \int \frac{\mathcal{P}(\mathrm{d}|\vartheta, \mathcal{M})\mathcal{P}(\vartheta|\mathcal{M})}{g(\vartheta)} g(\vartheta)d\vartheta. \tag{2.2.0.2}$$

Equation 2.2.0.2 is in a form such that $\mathcal{P}(\mathrm{d}|\mathcal{M})$ can be estimated numerically by drawing N samples from $g(\vec{\vartheta})$,

$$\widehat{\mathcal{P}(\mathrm{d}|\mathcal{M})} \approx \frac{1}{N} \sum_{\vartheta_i \sim g(\vartheta)} \frac{\mathcal{P}(\mathrm{d}|\vartheta_i, \mathcal{M})\mathcal{P}(\vartheta_i|\mathcal{M})}{g(\vartheta_i)}. \tag{2.2.0.3}$$

The key to an accurate and efficient estimate of $\widehat{\mathcal{P}(\mathrm{d}|\mathcal{M})}$ lies in choosing an appropriate $g(\vartheta)$. Assuming our parameter space contains one dominant posterior mode, we choose a multivariate normal $\mathcal{N}(\vec{\mu}_g, \vec{\Sigma}_g)$, where $\vec{\mu}_g$ and $\vec{\Sigma}_g$ describe the mean vector and covariance matrix of the model parameters respectively. After we perform an MCMC on a particular model/dataset, we can estimate $\vec{\mu}_g$ and $\vec{\Sigma}_g$ using a set of posterior samples. That information

is fed into our importance sampling algorithm to estimate $\widehat{\mathcal{P}(\mathrm{d}|\mathcal{M})}$ for that model. Nelson et al. [37], Guo et al. [38], and Weinberg et al. [39] provide more detailed prescriptions and investigations of this method [30].

Chapter 3

Radial Velocity Exoplanets

The first project for my model comparison pipeline is fitting models near the 2:1 degeneracy—introduced in Section 1.1.1—to archival RV exoplanet data, and then perform a Bayes factor analysis to determine the model that is favored by the data. The families of RV models the pipeline fit are introduced below.

3.1 The Models

3.1.1 Single Eccentric

One planet on an eccentric orbit is modeled with the following equations:

$$RV_{\rm single}(t) = K[\cos(\omega + f(t)) + e\,\cos(\omega)] + A \times (t - t_{\rm middle}) + C \quad (3.1.1.1)$$

$$M(t) = \left(2\pi\,\frac{(t-t_o)}{P} + M_o\right) \bmod 2\pi \quad (3.1.1.2)$$

$$M(t) = E(t) - e\,\sin E(t) \quad (3.1.1.3)$$

$$\cos f(t) = \frac{\cos E(t) - e}{1 - e\,\cos E(t)} \quad (3.1.1.4)$$

$$\sin f(t) = \frac{\sqrt{1-e^2}\,\sin E(t)}{1 - e\,\cos E(t)}, \quad (3.1.1.5)$$

where t_o is the time of the earliest RV measurement, $t_{\rm middle}$ is the time in the middle of the RV measurements, $f(t)$ is the true anomaly, $E(t)$ is the eccentric anomaly, $M(t)$ is the mean anomaly, and the model parameters are: orbital period (P), velocity semi-amplitude (K), eccentricity (e), longitude of periastron (ω), mean anomaly at t_o ($M_{\rm o} \equiv M(t_{\rm o})$), linear trend ($A$) and a velocity offset (C). The algorithm for solving for $RV_{\rm single}(t)$ is as follows:

- Solve Equation 3.1.1.2 for $M(t)$
- Numerically solve Kepler's equation—Equation 3.1.1.3—for $E(t)$
- Solve and substitute Equations 3.1.1.4 – 3.1.1.5 into Equation 3.1.1.1

3.1.2 Circular Double

Two planets on circular orbits are straightforward to model because $e = 0$ and ω is undefined. It is modeled with the following equations:

$$RV_{\rm double}(t) = K_{\rm out}\cos f_{\rm out}(t) + K_{\rm in}\cos f_{\rm in}(t) + A(t - t_{\rm middle}) + C \quad (3.1.2.1)$$

$$M(t) = \left(2\pi \frac{(t - t_o)}{P} + M_o\right) \bmod 2\pi \quad (3.1.2.2)$$

$$M(t) = E(t) = f(t), \quad (3.1.2.3)$$

where the model parameters are: outer/inner planet period ($P_{\rm out}/P_{\rm in}$), the velocity semi-amplitude for the outer/inner planets ($K_{\rm out}/K_{\rm in}$), and mean anomaly at the time of the earliest RV measurement for the outer/inner planets ($M_{\rm o,out}/M_{\rm o,in}$). Each model also has a linear trend (A) and a velocity offset (C). When the orbit is circular the true anomaly is equal to the mean anomaly. This makes the algorithm for solving $RV_{\rm double}(t)$ easier than for the single planet case: Equation 3.1.2.2 needs to be substituted into Equation 3.1.2.1 for the outer and inner planet.

3.1.3 Degeneracy at the 2:1 Mean-Motion Resonance

The source of the model degeneracy is in the first-order expansion of the RV signal of a single eccentric planet [40]:

$$RV_{\rm single} \approx K\cos(M + \omega) + Ke\cos(2M + \omega) + \mathcal{O}(e^2), \quad (3.1.3.1)$$

where $RV_{\rm single}$ is the observed radial velocity, K is the velocity semi-amplitude, e is the eccentricity, ω is the longitude of periastron, and M is the mean anomaly, which is a function of time. By comparison, the signal of a circular double is

$$RV_{\rm double} = K_{\rm out}\cos(M_{\rm out}) + K_{\rm in}\cos(M_{\rm in}), \quad (3.1.3.2)$$

where $RV_{\rm double}$ is the observed radial velocity, $K_{\rm out}$ and $K_{\rm in}$ are the velocity semi-amplitudes, and $M_{\rm out}$ and $M_{\rm in}$ are the mean anomalies. At the 2:1 MMR, $M_{\rm in} = 2M_{\rm out}$ and the inner planet signal ($K_{\rm in}$) masquerades—to first-order—as the eccentricity signal (Ke) from the single planet model.

3.2 Preliminary Sample

The sample introduced in Section 1.1.1.1 contains 60 systems and comprises every non-transiting RV system from the NASA Exoplanet Archive (NASA Archive), as of November 2016 [41], that is listed as having only a single planet, orbiting a main sequence star, and whose system properties were derived from a single data set. We did not limit our sample by eccentricity. Our main results will focus on the main sequence stars, but we will also report on an extended sample which ignores stellar type. The extended sample contains 95 systems, which is nearly a quarter of all RV-discovered single-planet systems. Figure 3.1 shows how we determine stellar types based on their reported surface gravity and how we select the main sequence sample.

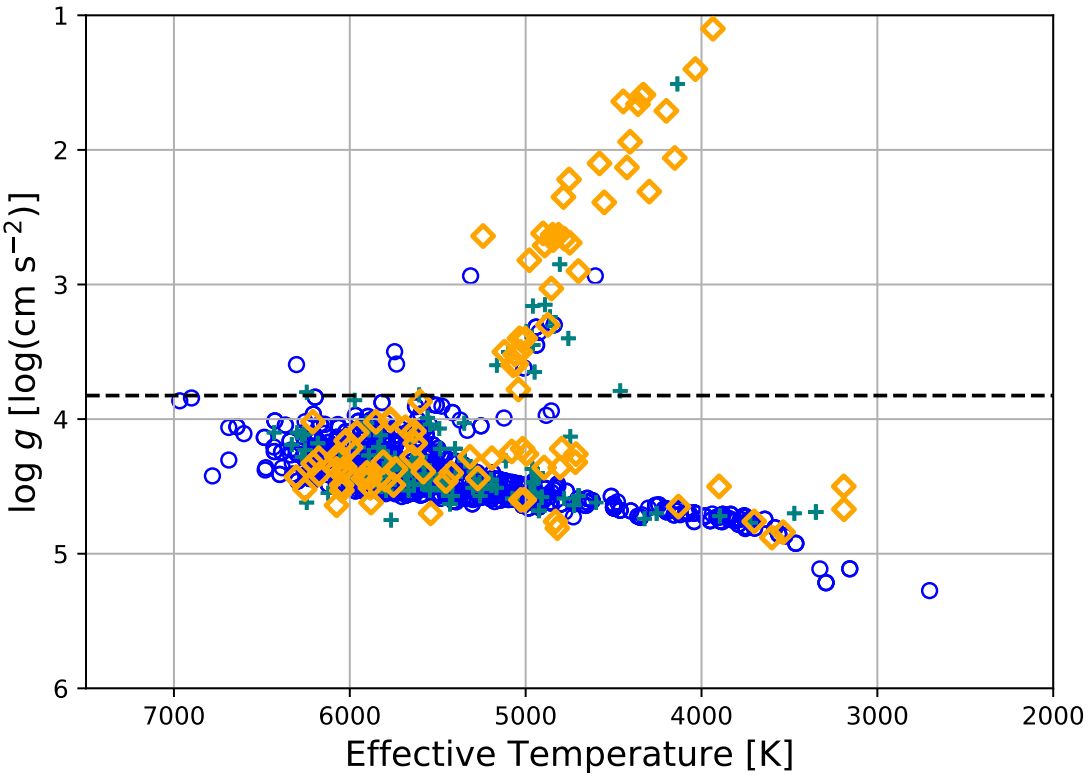

Figure 3.1: Stellar effective temperature vs. stellar surface gravity plot for the RV multi-planet systems from the NASA Archive as grey crosses and the *Kepler* multi-planet systems as blue circles. Our sample of 95 stars, ignoring stellar type, are orange diamonds. For our sample of main sequence stars we select those with $\log g \geq 3.825$, there are 60 main sequence stars in the main sample and 95 stars in the entire sample.

3.2.1 Models Tested

For each system we test four planetary system models: a single eccentric; two circular doubles (one with a period ratio fixed at 2 and the other fixed at 2.17—where there are two large peaks in the period ratio distribution from *Kepler* Steffen et al. [27]); and a "floating" circular double with no period ratio constraint. This last model has an additional model parameter, but the Bayes factor calculation can account for different numbers of model parameters. Our primary results work with the two fixed models given the compelling theoretical and observational reasons to consider them, the fact that the number of model parameters are identical (and thus more directly comparable), and because a narrow-band signal at a fixed period ratio is less susceptible to a false positive detection from stellar RV jitter or statistical noise. For each system we only consider the larger of the Bayes factor for the two fixed circular double models.

3.2.2 Pipeline Set-up

We draw our set of initial conditions for the MLE from the NASA Archive. The time of periastron passage is used to determine the initial $M_{\rm o}$. We first fit for C, fixing the other parameters at their nominal values and setting $A = 0$. We next fit for A and C simultaneously. Some systems did not have K, ω, and/or the time of periastron passage reported on the NASA Archive. In those cases, an MLE was done with the missing quantities as the only free parameters. We initialized the fixed circular double models to their first-order, single eccentric equivalent values using equations (3.1.3.1) and (3.1.3.2). For the floating circular double, the inner planet orbital period is initialized to either the 2:1 or the 2.17:1, depending on which fixed model produced a larger Bayes factor.

We impose a modified Jeffery's prior for the orbital period and velocity semi-amplitude with bounds between 0–10,000 days and 0–2,000 m s^{-1} respectively and ϑ_0 equal to 1 day and 1 m s^{-1} respectively. We use this prior because it is normalizable, objective, and intended for scalable parameters that could have zero as a value. We use uniform priors for the remaining parameters, (e, ω, M_0), because they are also normalizable and objective. We sample the parameters for the single eccentric model in $\{P,\, K,\, \sqrt{e}\sin(\omega),\, \sqrt{e}\cos(\omega),\, \omega + M_0\}$-space in order to maintain uniform priors [42].

The prior bounds for K, $K_{\rm out}$, and $K_{\rm in}$ are between 0 and 2,000 m s^{-1}. The prior bounds for P, $P_{\rm out}$, and $P_{\rm in}$ are between 0 and 10,000 days. The prior bounds for $\sqrt{e}\,\sin(\omega)$ and $\sqrt{e}\cos(\omega)$ are such that $0 < (\sqrt{e}\sin(\omega))^2 + (\sqrt{e}\cos(\omega))^2 < 1$, i.e. $0 < e < 1$. The prior bounds for $(M_0 + \omega)$, $M_{0,\rm out}$, and $M_{0,\rm in}$ are between -2π and 4π. These limits allow the Markov chains to cross the 0 and 2π coordinate singularities while remaining well-behaved. Furthermore, these values are modded by 2π before doing any calculations. The prior bounds for C are between -100,000 m s^{-1} and 100,000 m s^{-1} to accommodate the wide range in offset values in the real sample.

3.2.3 Pipeline Characterization

We characterized the pipeline efficiency with an ensemble of 1,000 synthetic RV time series whose system and data properties match the real systems. We use the Bayes factors of

these synthetic systems to characterize our model comparison pipeline. This Monte Carlo simulation was initialized as follows.

The start time (t_0) is a uniform random draw between 1 and 1,000 days. The number of observations are drawn from the real systems with a normally-distributed adjustment with a standard deviation 10% of the nominal value rounded to the nearest whole number. The observation time series is produced by selecting a set of observation differences ($t_i - t_{i-1}$) from the real distribution of observation differences with a similar, normally-distributed 10% variation added to each difference. The number of orbits is the number of orbits of a randomly chosen real system with a normally distributed 10% variation.

We determine the orbital period (P) using the selected number of orbits and the observation time series. The velocity semi-amplitude (K) and the eccentricity (e) are separate random draws from the real systems. The mean anomaly of the start time (M_0) and argument of periastron (ω) are randomly drawn between 0 and 2π. The linear trend (A) is a 10% variation to a random draw from the real systems.

We assume that the RV errors are normally distributed with a standard deviation that is the quadrature sum of stellar jitter and instrumental and photon noise. The instrumental and photon noise ($\sigma_{\rm RV}$) are drawn randomly from the RV errors of the real systems and our error bars are assigned to this value. Steller jitter is selected from a log uniform distribution between 0.5 and 5 m s^{-1}. The observation errors are added to the synthetic RV measurement—not to the error in the RV measurement. Figure 3.2 shows the parameter distributions for the 1,000 synthetic time series and the real systems as reported in the NASA Archive.

3.2.3.1 Synthetic Results

The resulting Bayes factors from this characterization are shown as the blue distribution in Figure 3.3. The vertical lines denote the 95th and 90th percentiles of the distribution. The shape of the distribution is not symmetric, and the vast majority of our synthetic datasets favor the single-planet case—as expected since the synthetic systems were constructed to be single eccentrics. Real systems with Bayes factors larger than those thresholds may be circular double systems mischaracterized as single eccentrics.

Our approach to model comparison to our sample is different from earlier studies. For example, Wittenmyer et al. [25] used the reduced χ^2 to determine the preferred model and refined their results with stability tests using the N-body integrator *Mercury* [43]. Angelada et al. [24] randomized individual sets of data to calculate the false positive rate per system. Their model selection was also based on the reduced χ^2 of least-squares fitting. In this work, we use a fully marginalized likelihood to calculate the Bayes factor for the model comparison and we estimate our false positive rate by analyzing a large simulated dataset with our pipeline.

3.3 Real System Analysis and Results

After analyzing our synthetic systems, we ran our sample of 60 real systems (95 systems for the extended sample) through the same pipeline. Figure 3.3 shows that the Bayes factor

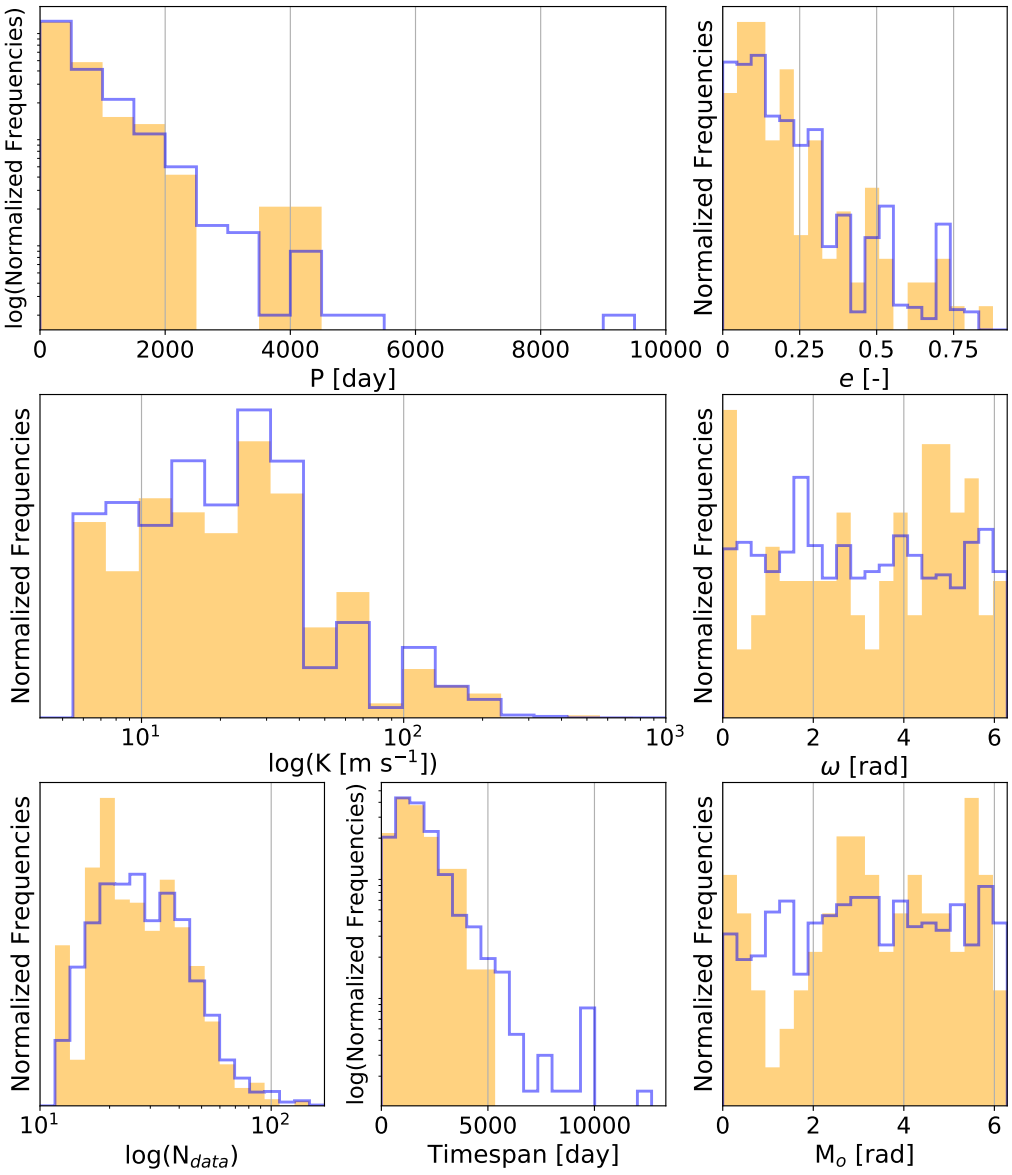

Figure 3.2: Property distributions for our sample of 95 real systems from the NASA Exoplanet Archive in orange and the 1,000 synthetic time series in blue.

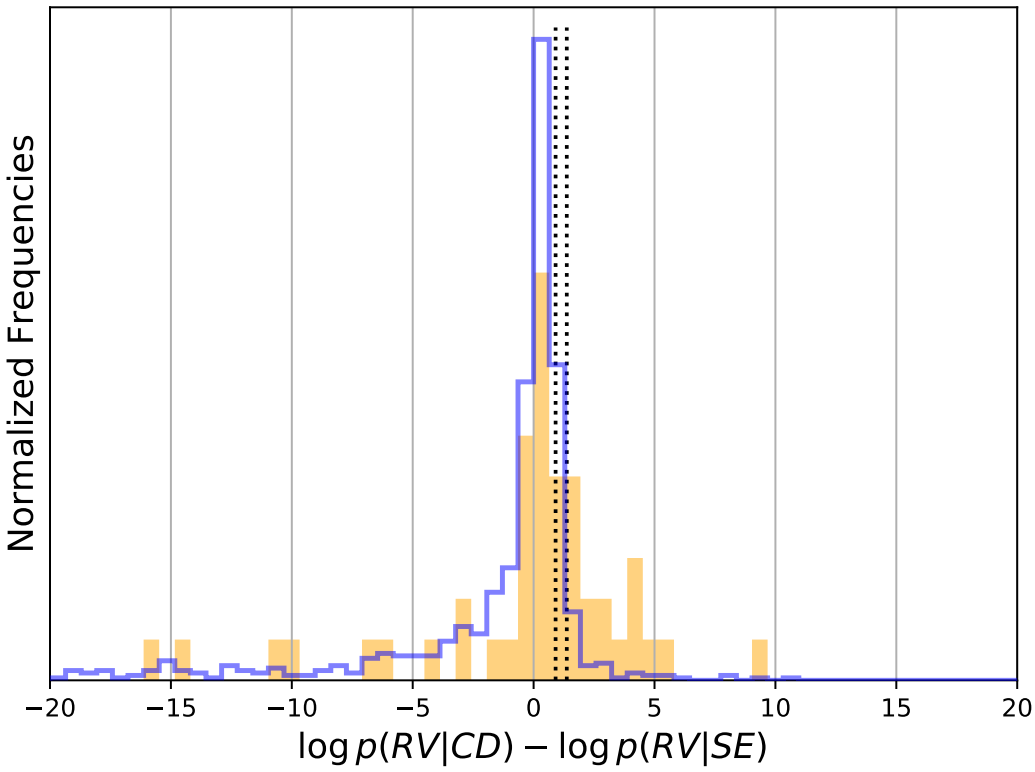

Figure 3.3: The log Bayes factor distribution for the 1,000 synthetic single eccentric time series in blue and 60 real systems hosted by main sequence stars in orange. Here, we compare only the single eccentric model to the fixed, circular double model with the largest Bayes factor. The 95th and 90th percentiles are indicated with the dotted lines near a Bayes factor of 24 and 8, respectively.

distributions for the synthetic and real systems (in orange) are not similar. We find that 15 (25%) of the systems have Bayes factors larger than the 95th percentile of the synthetic systems. (For the extended sample of 95 systems the numbers are 30 and 31% of the entire sample respectively.) 9 of these systems prefer the 2.17:1 model (22 of the extended sample) while the remaining 6 (8 from the extended sample) prefer the 2:1 model. Assuming a false positive rate of 5% from our 95% confidence level, our estimate of the number of false positives is 0.75 ± 0.87 (1.5 ± 1.2 for the extended sample). The systems from the extended sample that prefer the fixed circular double model, the model parameters, and Bayes factors are shown in Table 3.1. The timeseries and phase plots for the best fixed circular double model and single eccentric model for HD 240237, HD 2952, α Arietis, HD 96127, and HD 95089 are shown in Figures 3.4–3.23. A CSV file containing the model parameters with errors for all four models, Bayes factors between the circular double models and the single eccentric model, and percentile of the best fixed model for each system in the extended sample are available in Boisvert et al. as Table 2 [30].

We examine the consequences of these potential discoveries on several distributions of planet properties. Figure 3.24 shows the planet mass vs. orbital period for known RV planets along with the new planets favored by our analysis orbiting main sequence stars. These potential new systems lie well within the range of values measured in known systems. We point out, however, that some systems may yet be false positives. For instance, there are a few candidate circular double systems that would be hot Jupiters (planet with $P \lesssim 10$ days) with interior companions. Presently, there is only a single known system (WASP-47 [44]) where a hot Jupiter has a known interior companion. And the period ratio in this case is over 5:1—far from the degeneracy we consider here. However, the hot Jupiter has an outer companion with a period ratio near 2.17. Figure 3.25 shows how the predicted mass ratios for the main sequence systems that favor the two-planet model compare with the mass ratios for RV systems on the NASA Archive.

(This page is intentionally left black so that Figure 3.4 is on the back-side of the page.)

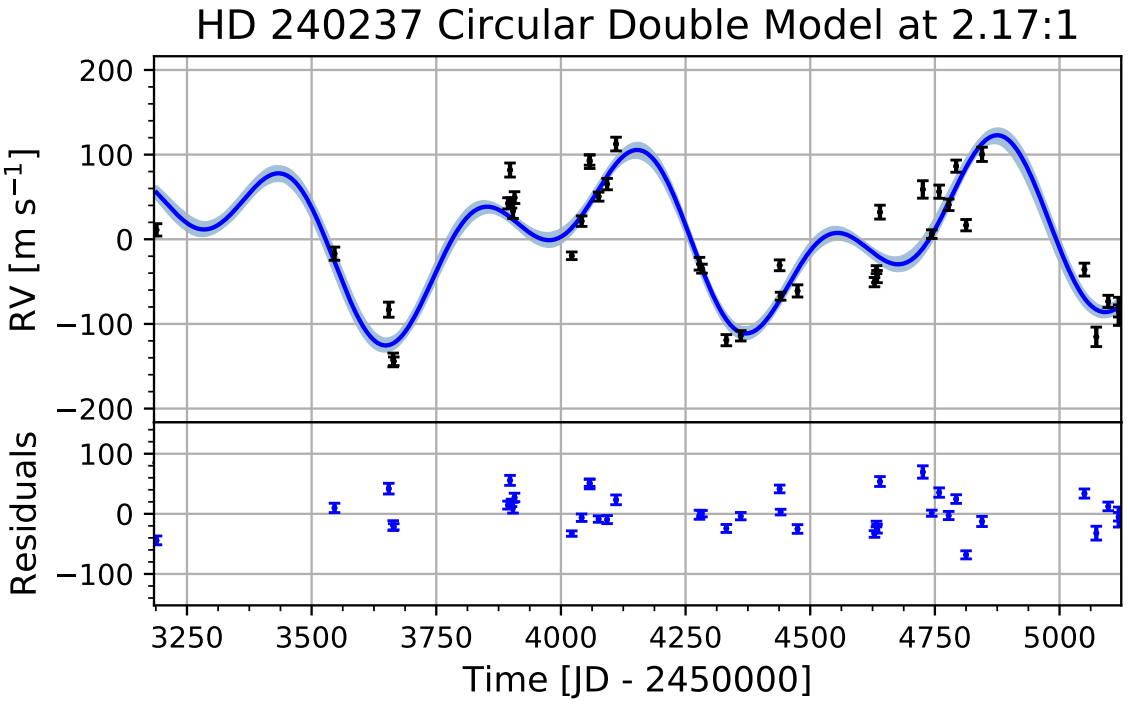

Figure 3.4: Time series for HD 240237. The offset data are in black, the circular double planet model fixed at 2.17:1 in blue and the 68.2% error from the posterior distribution is in light blue. This is the best fit out of the single eccentric, fixed circular double at 2:1, and fixed circular double at 2.17:1 models.

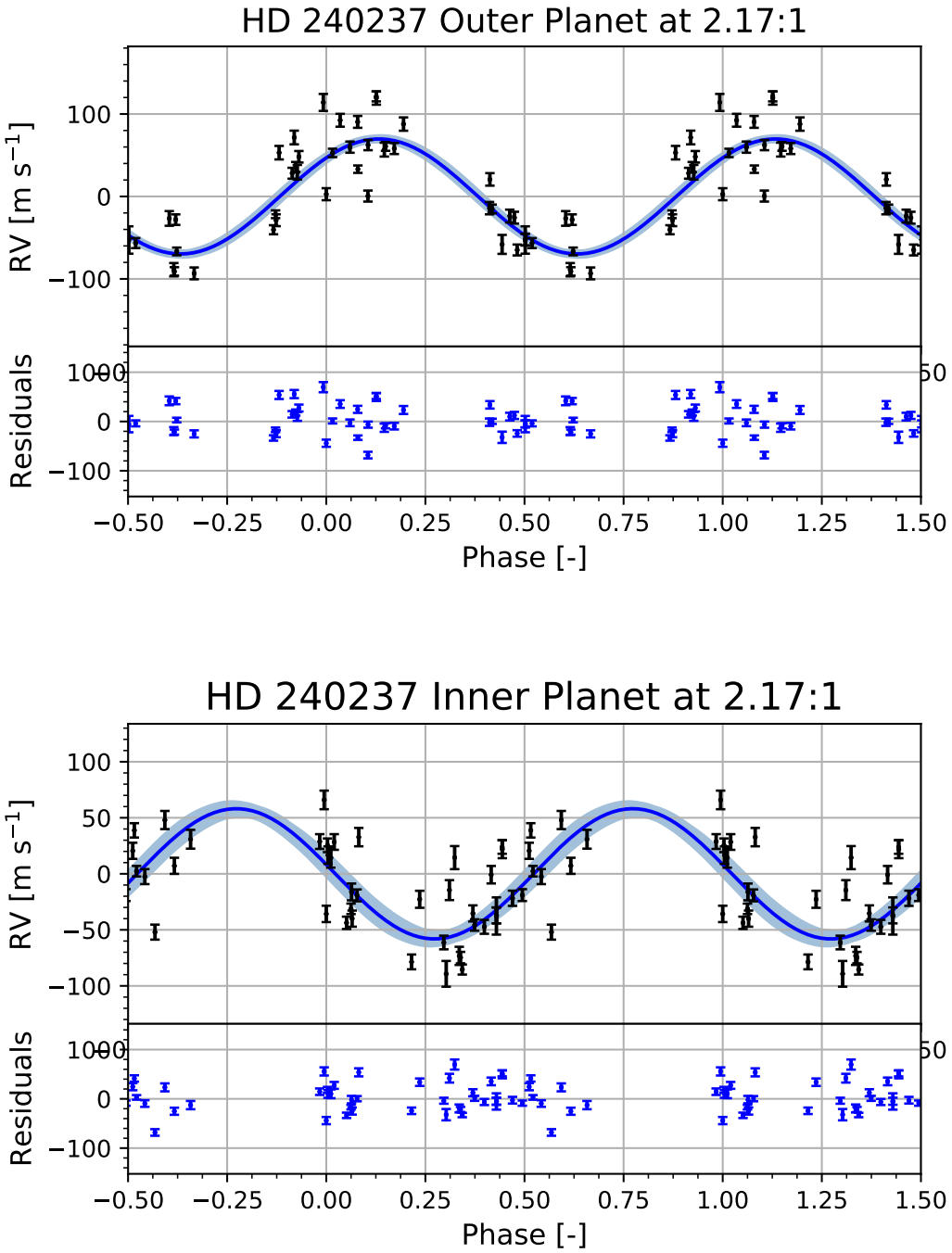

Figure 3.5: Phase plot for HD 240237. The offset data are in black, the circular double planet model fixed at 2.17:1 in blue and the 68.2% error from the posterior distribution is in light blue. This is the best fit out of the single eccentric, fixed circular double at 2:1, and fixed circular double at 2.17:1 models. The outer planet is on top and the inner planet is on the bottom.

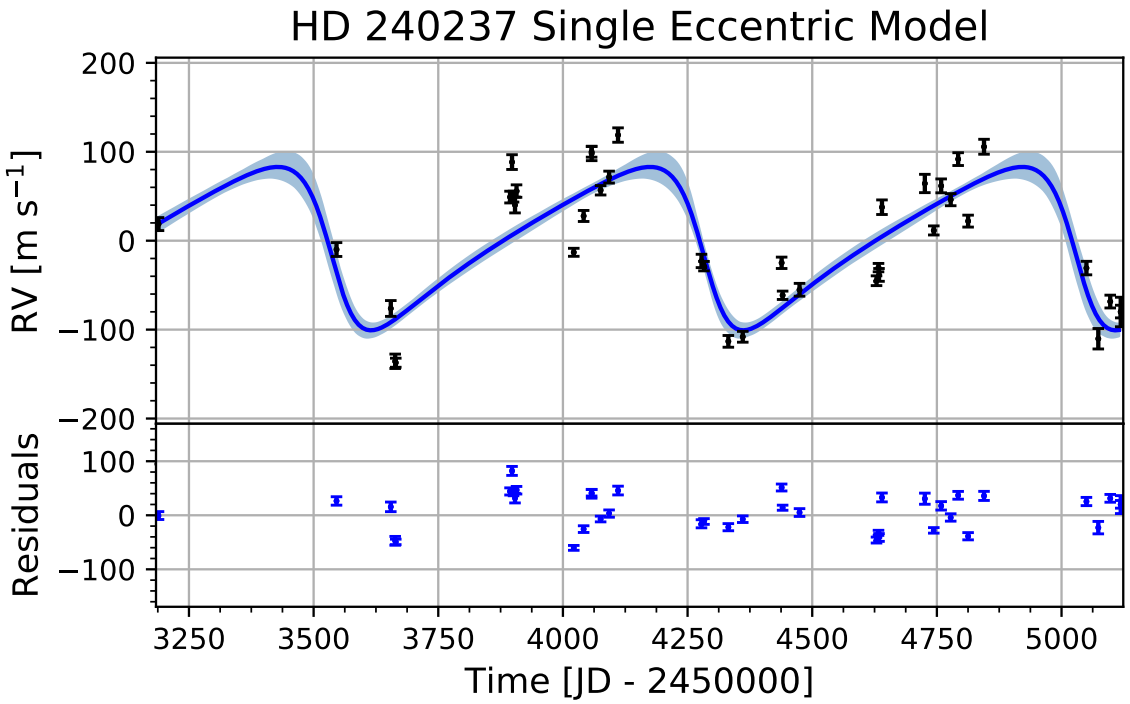

Figure 3.6: Time series for HD 240237. The offset data are in black, the single eccentric planet model in blue and the 68.2% error from the posterior distribution is in light blue.

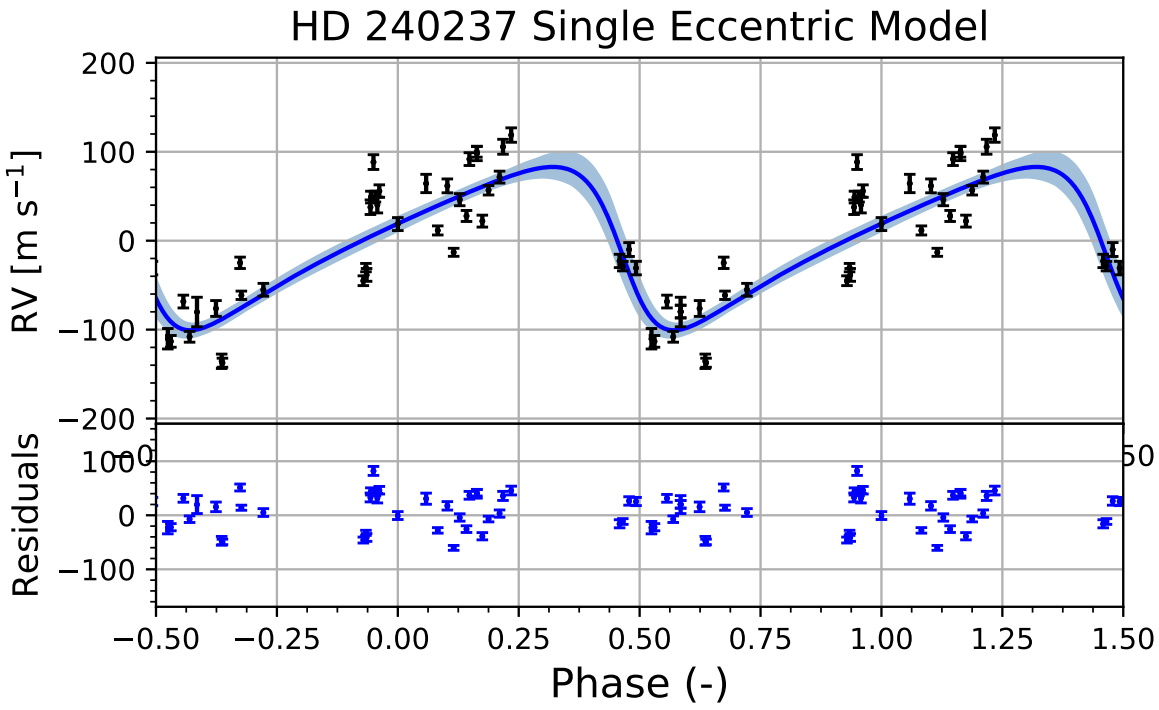

Figure 3.7: Phase plot for HD 240237. The offset data are in black, the single eccentric planet model in blue and the 68.2% error from the posterior distribution is in light blue.

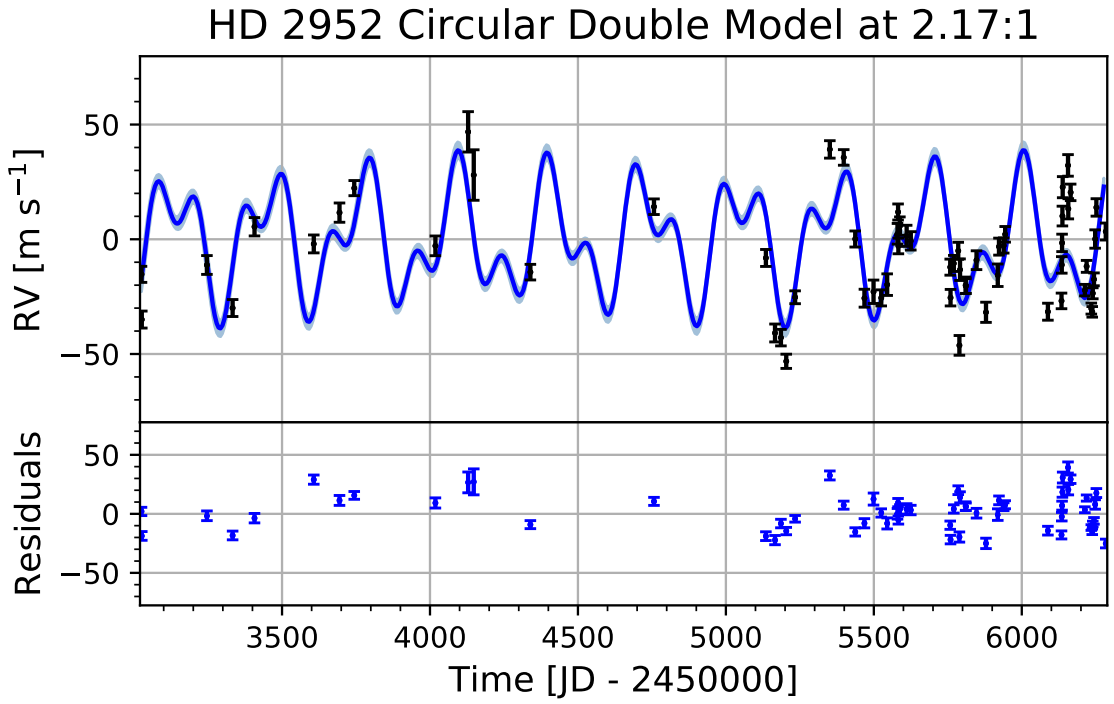

Figure 3.8: Time series for HD 2952. The offset data are in black, the circular double planet model fixed at 2.17:1 in blue and the 68.2% error from the posterior distribution is in light blue. This is the best fit out of the single eccentric, fixed circular double at 2:1, and fixed circular double at 2.17:1 models.

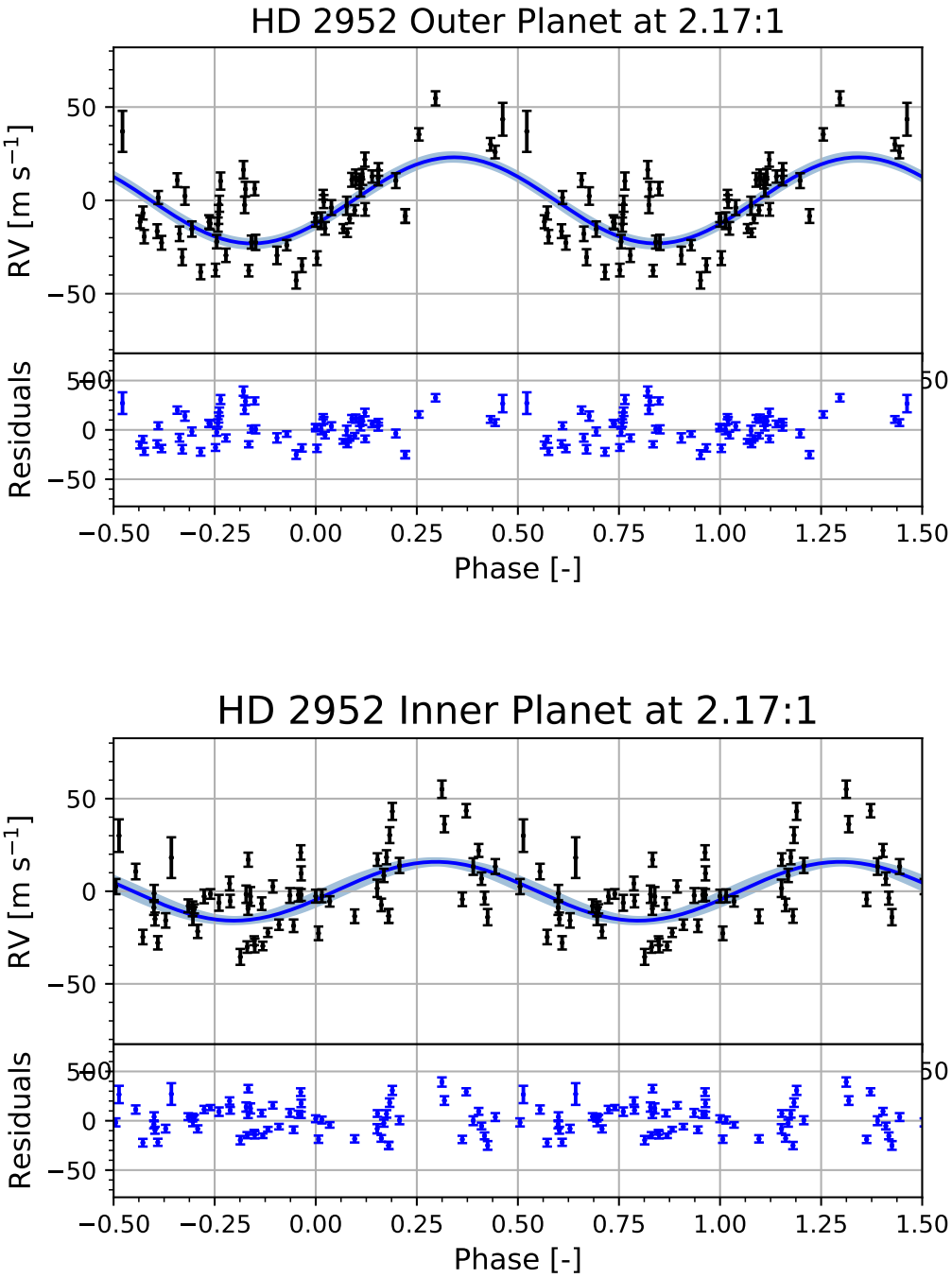

Figure 3.9: Phase plot for HD 2952. The offset data are in black, the circular double planet model fixed at 2.17:1 in blue and the 68.2% error from the posterior distribution is in light blue. This is the best fit out of the single eccentric, fixed circular double at 2:1, and fixed circular double at 2.17:1 models. The outer planet is on top and the inner planet is on the bottom.

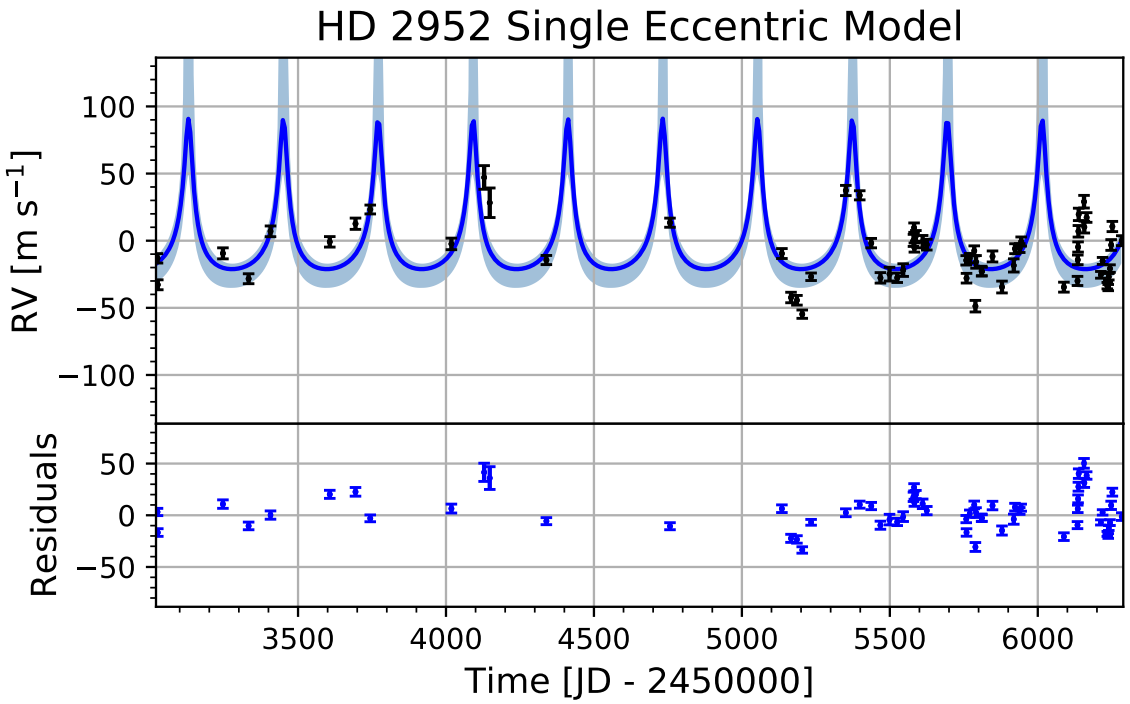

Figure 3.10: Time series for HD 2952. The offset data are in black, the single eccentric planet model in blue and the 68.2% error from the posterior distribution is in light blue.

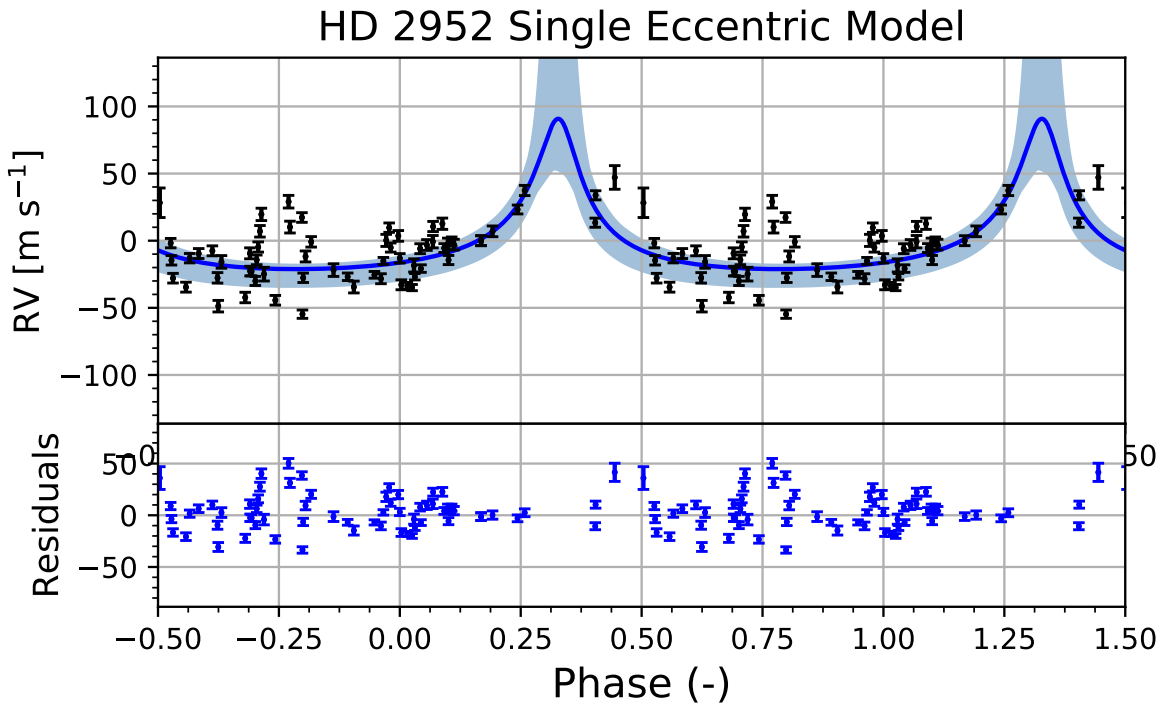

Figure 3.11: Phase plot for HD 2952. The offset data are in black, the single eccentric planet model in blue and the 68.2% error from the posterior distribution is in light blue.

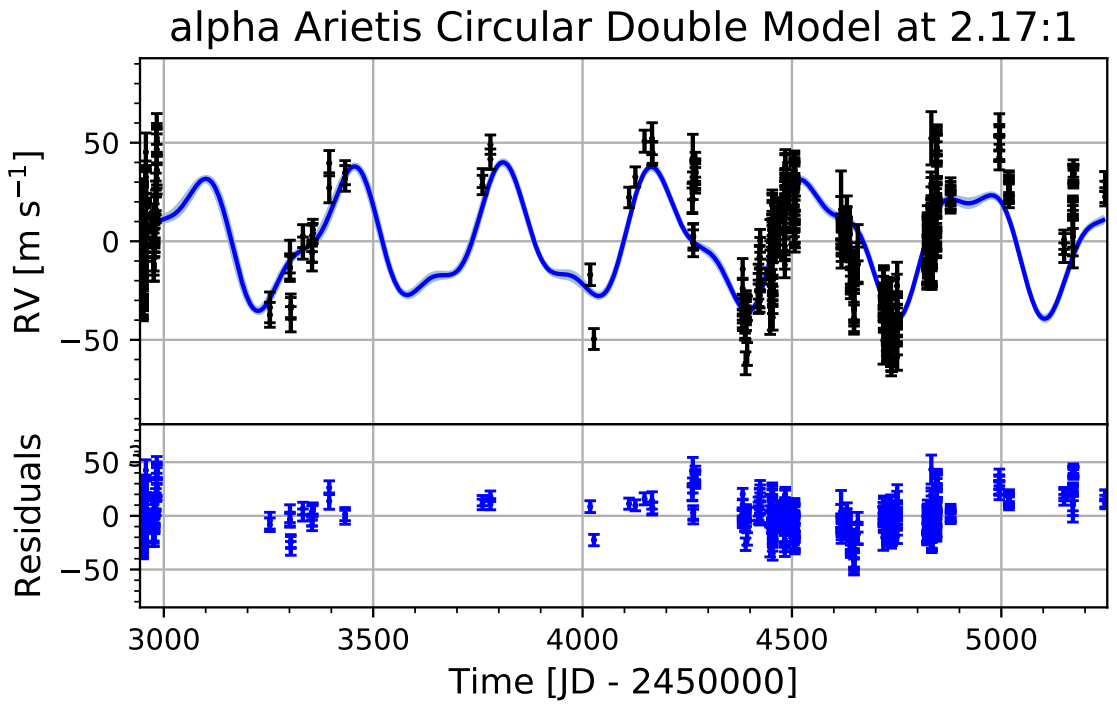

Figure 3.12: Time series for α Arietis. The offset data are in black, the circular double planet model fixed at 2.17:1 in blue and the 68.2% error from the posterior distribution is in light blue. This is the best fit out of the single eccentric, fixed circular double at 2:1, and fixed circular double at 2.17:1 models.

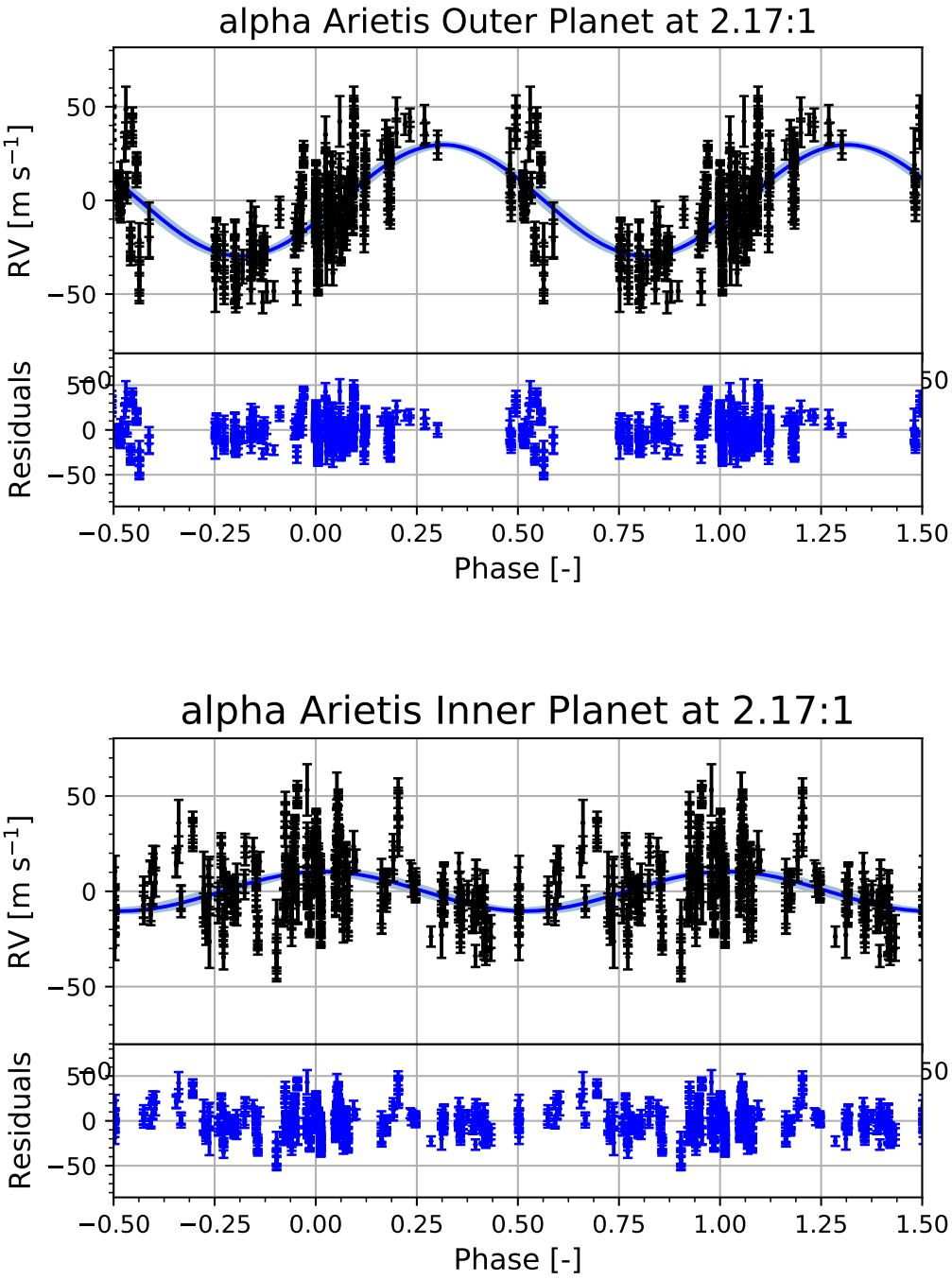

Figure 3.13: Phase plot for α Arietis. The offset data are in black, the circular double planet model fixed at 2.17:1 in blue and the 68.2% error from the posterior distribution is in light blue. This is the best fit out of the single eccentric, fixed circular double at 2:1, and fixed circular double at 2.17:1 models. The outer planet is on top and the inner planet is on the bottom.

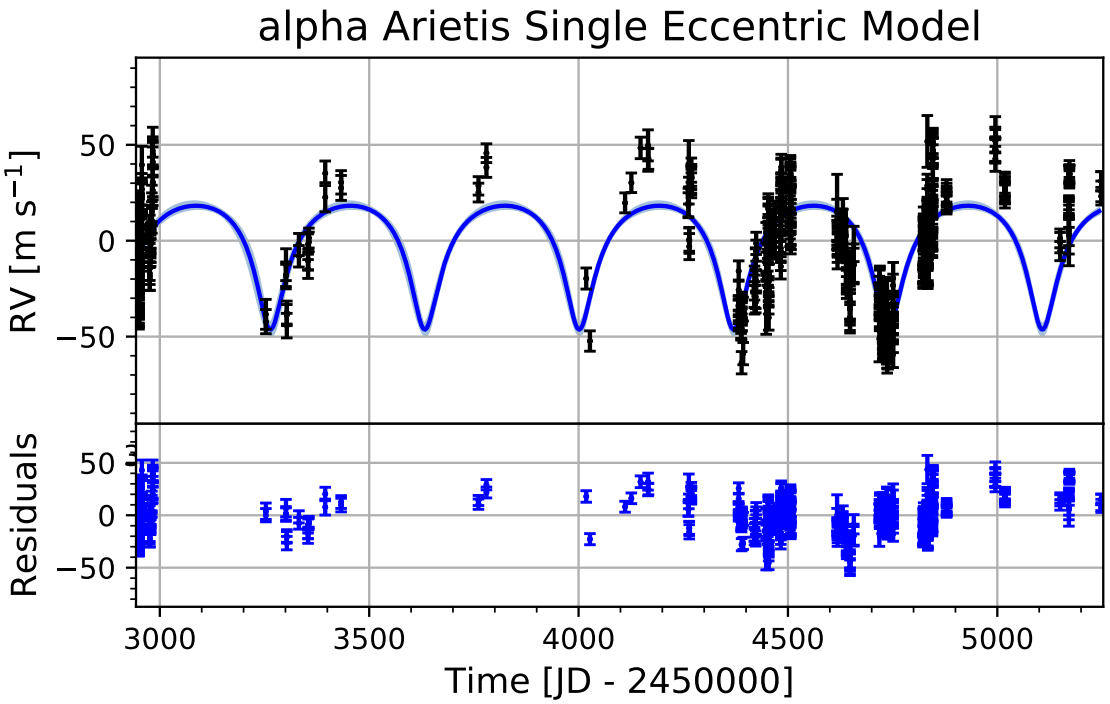

Figure 3.14: Time series for α Arietis. The offset data are in black, the single eccentric planet model in blue and the 68.2% error from the posterior distribution is in light blue.

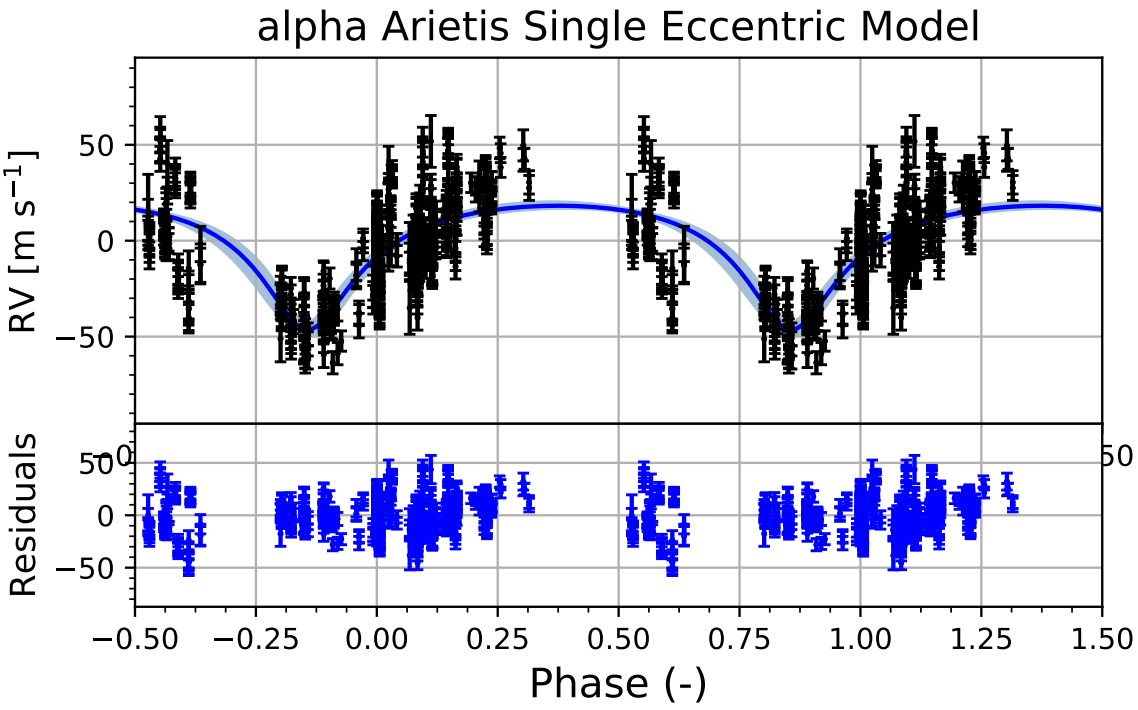

Figure 3.15: Phase plot for α Arietis. The offset data are in black, the single eccentric planet model in blue and the 68.2% error from the posterior distribution is in light blue.

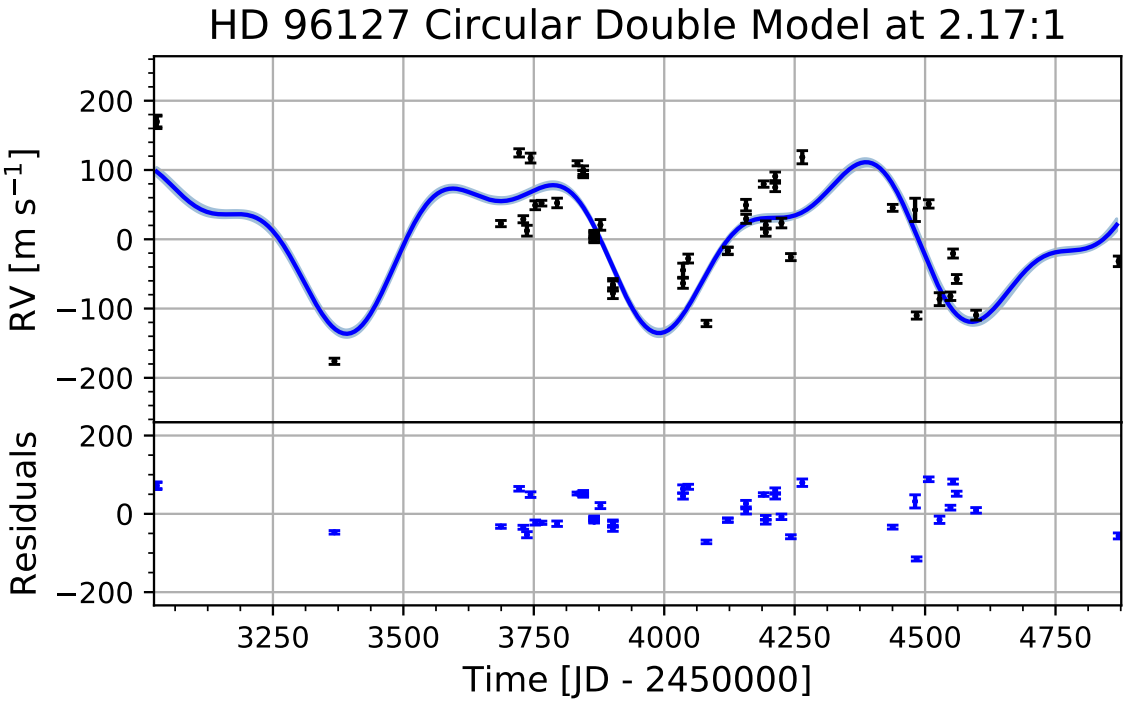

Figure 3.16: Time series for HD 96127. The offset data are in black, the circular double planet model fixed at 2.17:1 in blue and the 68.2% error from the posterior distribution is in light blue. This is the best fit out of the single eccentric, fixed circular double at 2:1, and fixed circular double at 2.17:1 models.

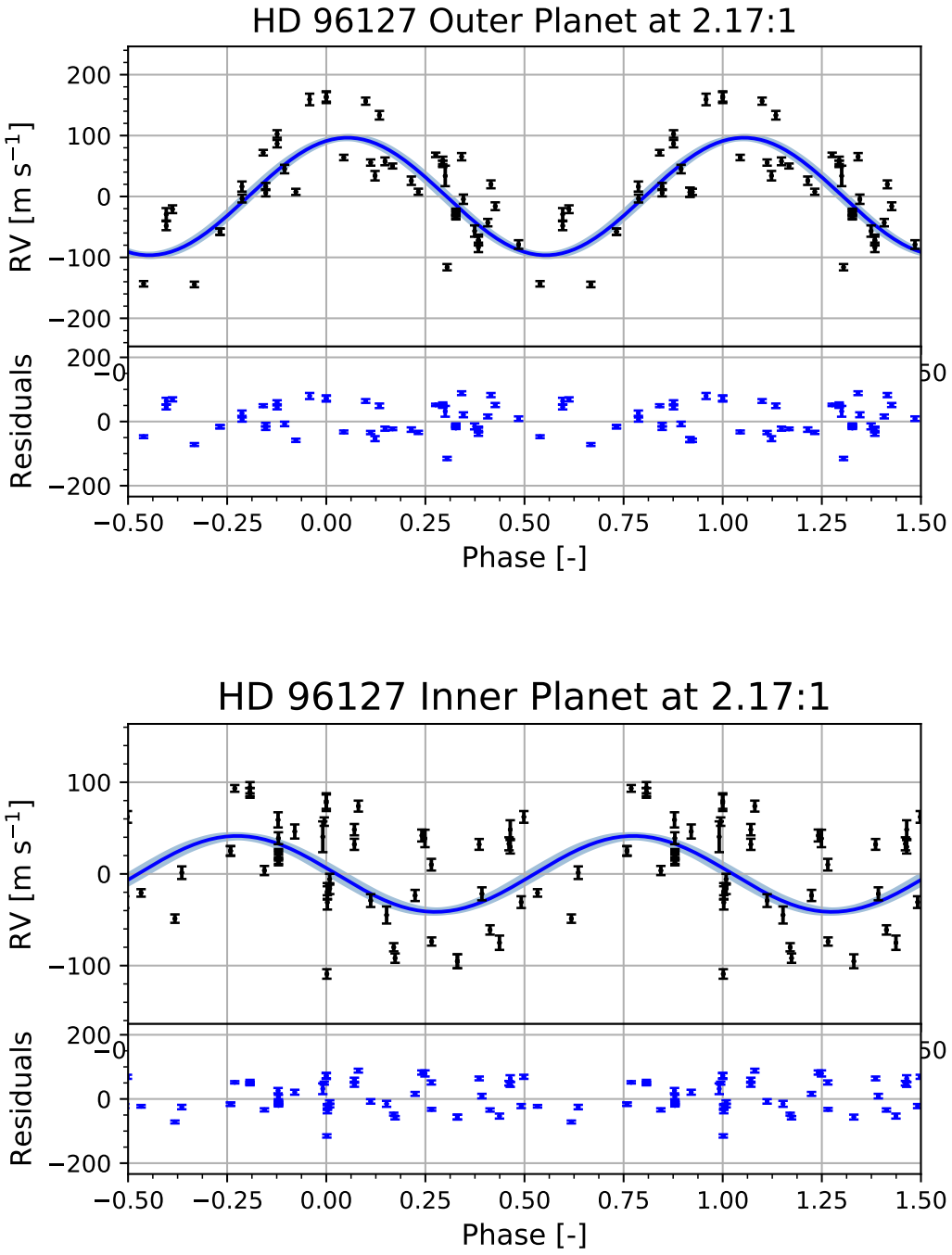

Figure 3.17: Phase plot for HD 96127. The offset data are in black, the circular double planet model fixed at 2.17:1 in blue and the 68.2% error from the posterior distribution is in light blue. This is the best fit out of the single eccentric, fixed circular double at 2:1, and fixed circular double at 2.17:1 models. The outer planet is on top and the inner planet is on the bottom.

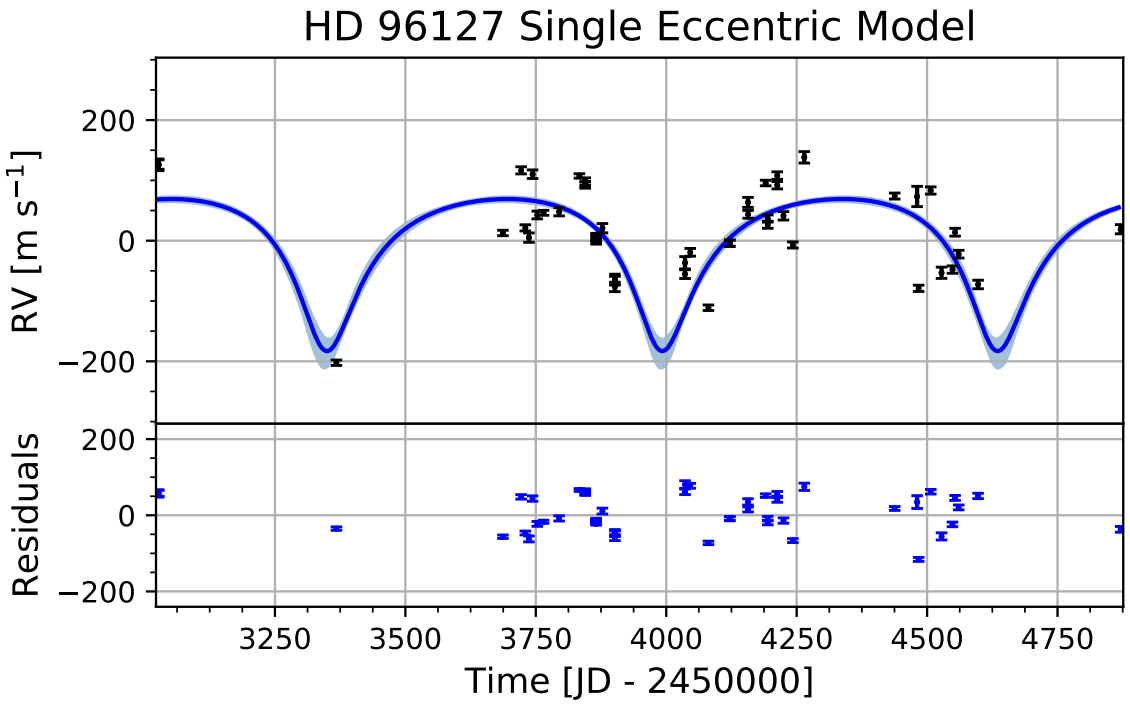

Figure 3.18: Time series for HD 96127. The offset data are in black, the single eccentric planet model in blue and the 68.2% error from the posterior distribution is in light blue.

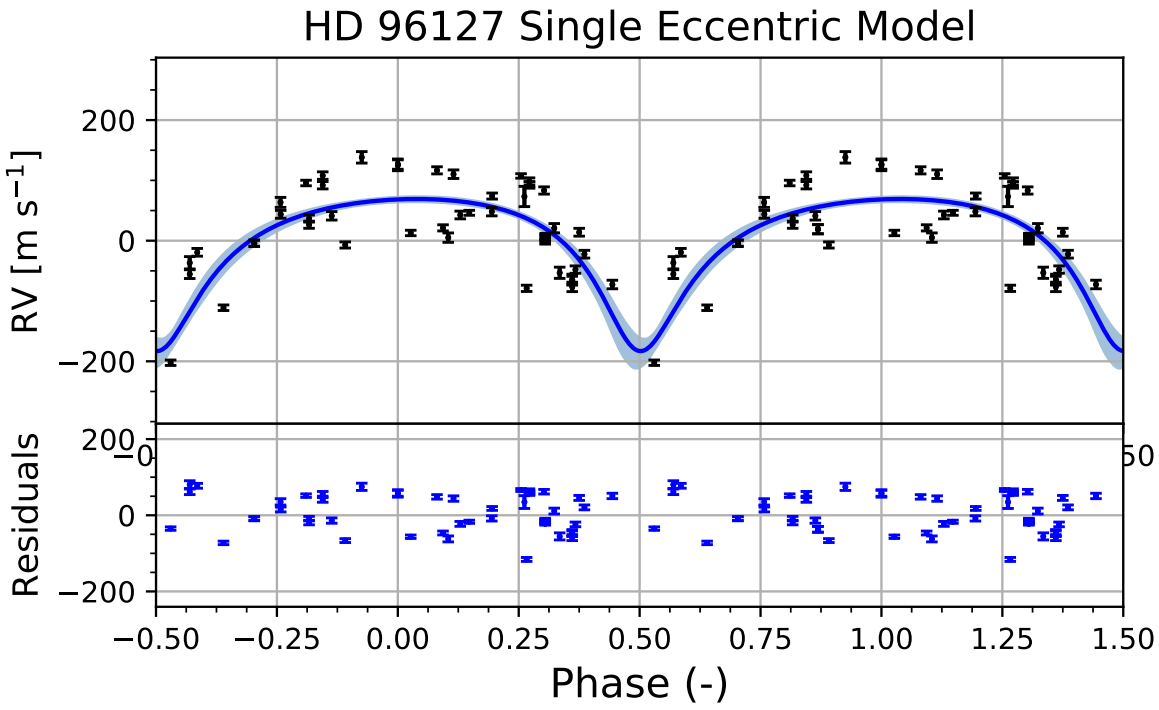

Figure 3.19: Phase plot for HD 96127. The offset data are in black, the single eccentric planet model in blue and the 68.2% error from the posterior distribution is in light blue.

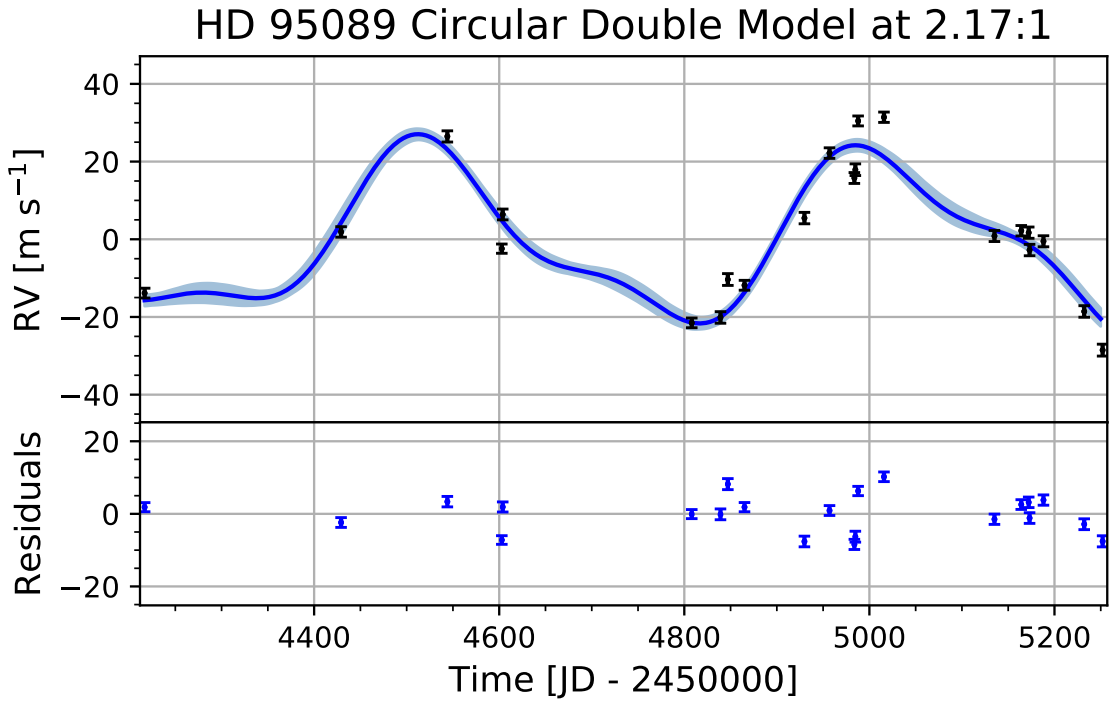

Figure 3.20: Time series for HD 95089. The offset data are in black, the circular double planet model fixed at 2.17:1 in blue and the 68.2% error from the posterior distribution is in light blue. This is the best fit out of the single eccentric, fixed circular double at 2:1, and fixed circular double at 2.17:1 models.

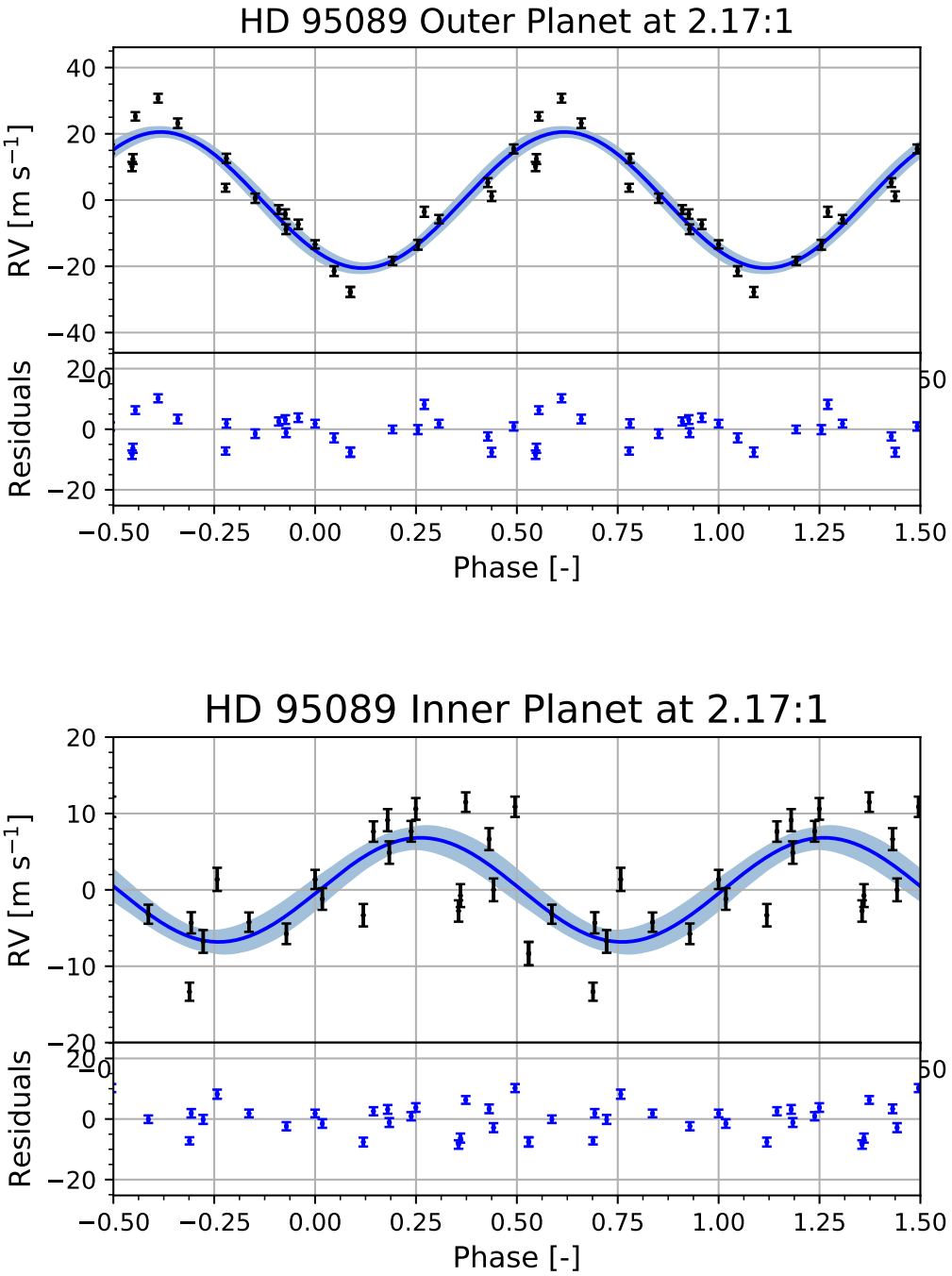

Figure 3.21: Phase plot for HD 95089. The offset data are in black, the circular double planet model fixed at 2.17:1 in blue and the 68.2% error from the posterior distribution is in light blue. This is the best fit out of the single eccentric, fixed circular double at 2:1, and fixed circular double at 2.17:1 models. The outer planet is on top and the inner planet is on the bottom.

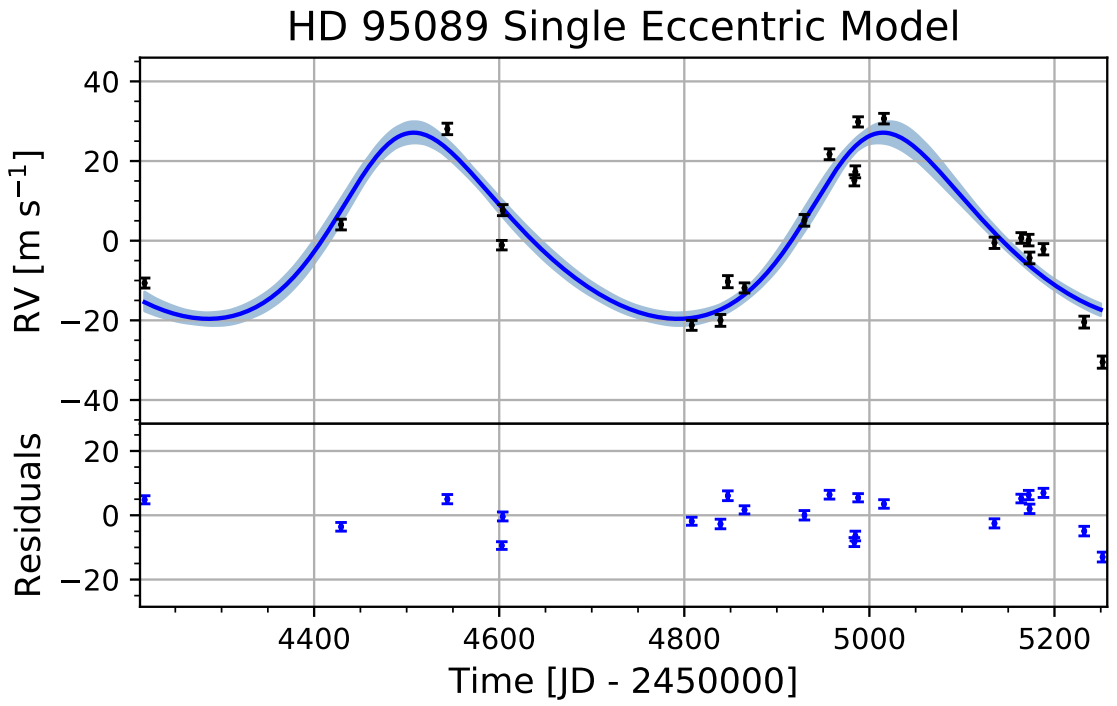

Figure 3.22: Time series for HD 95089. The offset data are in black, the single eccentric planet model in blue and the 68.2% error from the posterior distribution is in light blue.

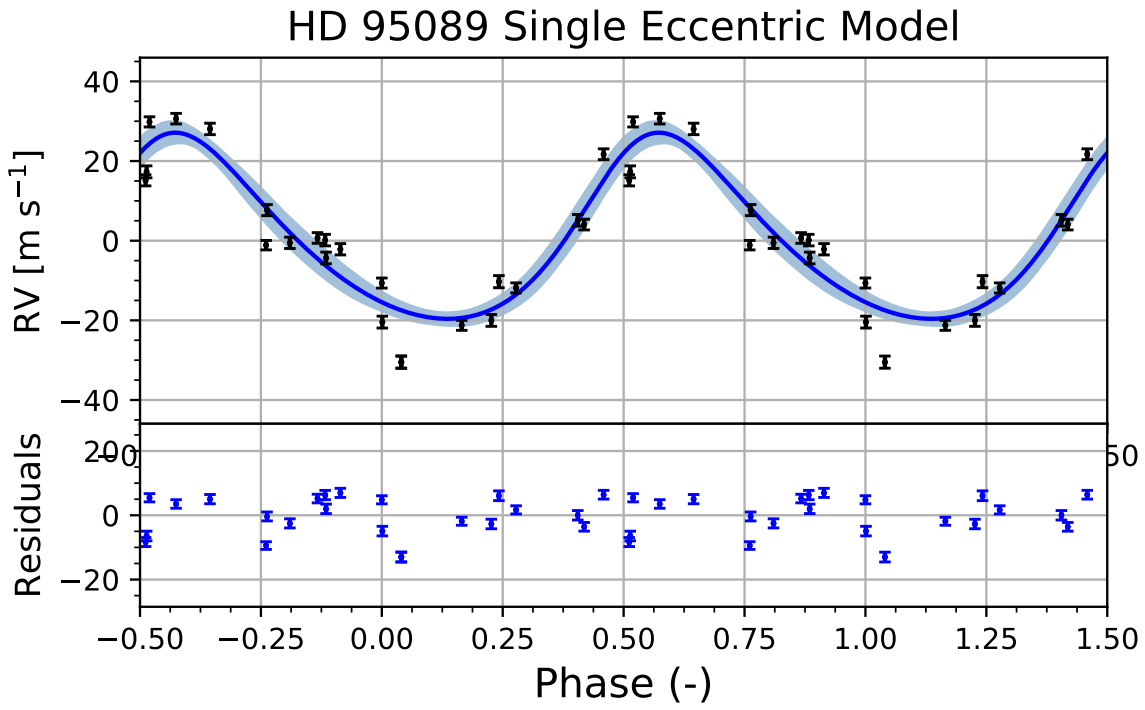

Figure 3.23: Phase plot for HD 95089. The offset data are in black, the single eccentric planet model in blue and the 68.2% error from the posterior distribution is in light blue.

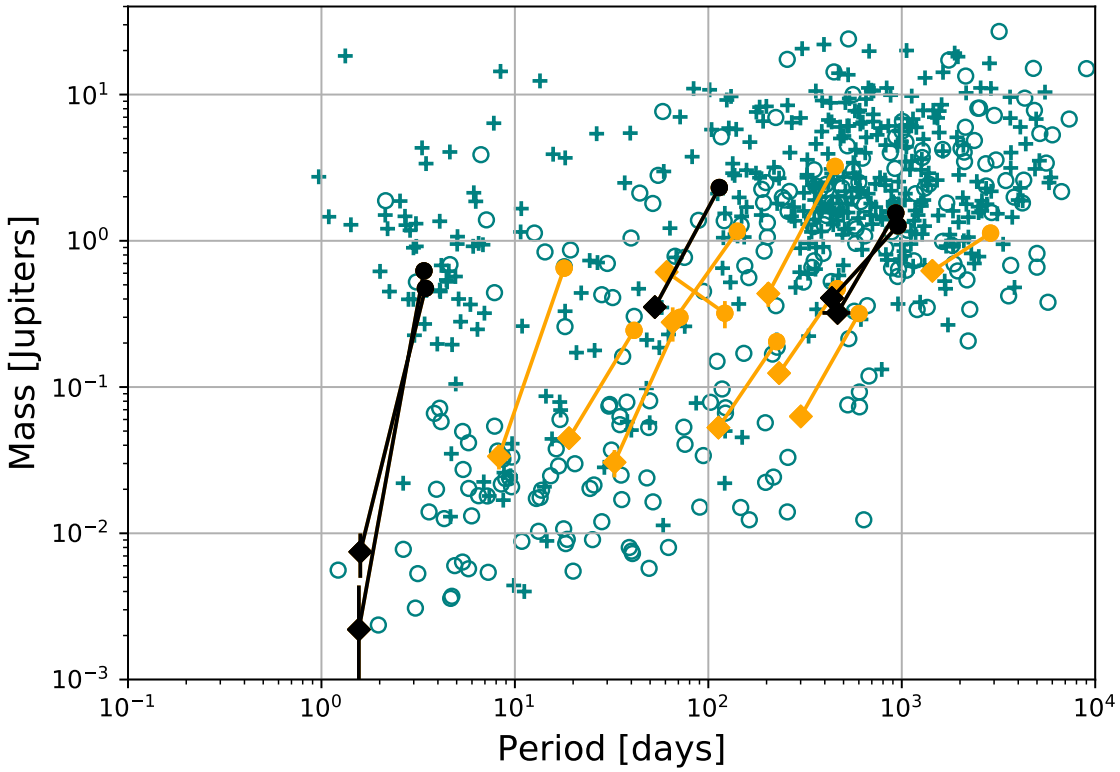

Figure 3.24: Orbital period vs. planetary mass for all RV planets. Systems from the NASA Exoplanet Archive are in teal, with multi- and single-planet systems as open circles and crosses, respectively. Each system orbiting a main sequence star with a measured Bayes factor larger than 95th percentile of the synthetic systems are plotted in orange. Each putative system is represented by a line on the plot, with the diamonds as the inner planet and the circles are the outer companion. Systems that remain in the 95th percentile after including a white noise stellar jitter term are in black. We note that these results lie well within normal parameter values of known systems.

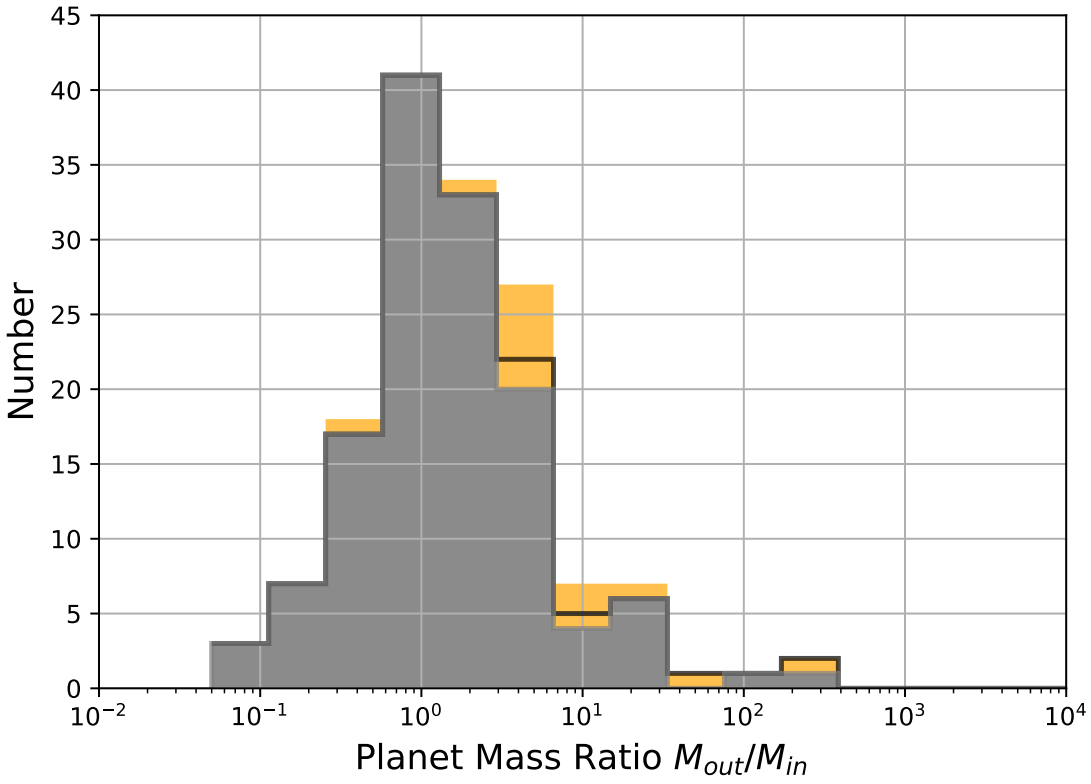

Figure 3.25: Mass ratio distribution for all RV adjacent planet pairs in grey. The stacked, orange distribution are our systems with Bayes factors larger than 24 (95th percentile) around main sequence stars. The nature of the signal favors more massive outer planets. The black outline shows the systems that have stellar jitter included and still had Bayes factors larger than the 95th percentile.

Star Name	Star Type	Model	$P_{\rm out}$ [days]	$K_{\rm out}$ [m s^{-1}]	$M_{\rm out}$ [rad]	$K_{\rm in}$ [m s^{-1}]	$M_{\rm in}$ [rad]	A [km s^{-1}day^{-1}]	C [m s^{-1}]	Bayes Factor
HD 240237	nMS	2.17:1	770.45 ± 3.09	69.75 ± 1.59	5.45 ± 0.04	57.52 ± 2.57	1.43 ± 0.08	9.96 ± 2.49	82.31 ± 1.54	1.09×10^{85}
HD 2952	nMS	2.17:1	318.83 ± 0.27	23.16 ± 0.94	4.13 ± 0.05	15.90 ± 0.71	4.41 ± 0.10	6.37 ± 0.62	3.15 ± 0.74	6.88×10^{57}
α Arietis	nMS	2.17:1	372.60 ± 0.52	29.67 ± 0.52	4.31 ± 0.04	10.25 ± 0.46	6.17 ± 0.10	-9.5 ± 0.64	0.91 ± 0.45	9.80×10^{49}
HD 96127	nMS	2.17:1	632.39 ± 2.19	96.66 ± 1.50	5.95 ± 0.04	41.02 ± 1.26	1.41 ± 0.07	22.83 ± 3.49	-923.95 ± 0.93	9.63×10^{43}
HD 95089	nMS	2.17:1	496.57 ± 2.88	20.65 ± 0.55	2.40 ± 0.05	6.73 ± 0.57	4.64 ± 0.11	-4.47 ± 1.39	0.23 ± 0.35	3.75×10^{16}
11 Ursae Minoris	nMS	2.17:1	513.22 ± 0.75	185.38 ± 1.15	4.30 ± 0.02	19.43 ± 1.15	4.80 ± 0.08	-11.14 ± 2.18	-10.39 ± 0.84	1.06×10^{15}
HD 136418	nMS	2.17:1	474.99 ± 1.26	41.62 ± 0.39	5.63 ± 0.03	10.41 ± 0.47	2.86 ± 0.08	-6.93 ± 1.1	-4.99 ± 0.42	8.03×10^{12}
HD 81688	nMS	2.17:1	184.08 ± 0.17	60.05 ± 0.92	5.47 ± 0.04	6.96 ± 0.97	5.44 ± 0.15	4.91 ± 1.53	-1.21 ± 0.67	1.41×10^{11}
HIP 57050	MS	2.17:1	41.31 ± 0.01	29.43 ± 0.62	2.60 ± 0.11	6.98 ± 0.64	3.86 ± 0.22	9.01 ± 0.53	-15.77 ± 0.83	4.27×10^{9}
HD 206610	nMS	2.17:1	628.84 ± 6.03	34.50 ± 0.55	1.67 ± 0.05	6.92 ± 0.79	3.06 ± 0.09	-17.3 ± 1.7	19.30 ± 0.47	1.59×10^{7}
HD 216770	MS	2:1	121.77 ± 0.38	14.05 ± 2.97	4.30 ± 0.33	33.86 ± 3.44	0.79 ± 0.18	36.91 ± 6.87	31146.80 ± 1.92	1.61×10^{5}
HD 114386	MS	2.17:1	955.46 ± 15.81	33.79 ± 1.55	3.92 ± 0.11	14.08 ± 1.96	2.23 ± 0.22	-5.24 ± 3.82	33367.71 ± 1.28	8.80×10^{4}
o Coronae Borealis	nMS	2:1	187.61 ± 0.13	33.64 ± 0.67	6.00 ± 0.05	8.64 ± 0.67	4.66 ± 0.10	1.92 ± 0.57	1.82 ± 0.49	8.66×10^{4}
HD 101930	MS	2.17:1	71.30 ± 0.17	17.93 ± 0.41	1.38 ± 0.04	2.37 ± 0.50	1.49 ± 0.20	-10.72 ± 3.43	18363.19 ± 0.35	3.00×10^{4}
14 Andomedae	nMS	2.17:1	185.89 ± 0.22	99.80 ± 1.41	4.90 ± 0.04	6.56 ± 1.61	4.60 ± 0.22	-7.64 ± 2.96	1.68 ± 1.24	2.59×10^{4}
HD 180902	nMS	2.17:1	482.22 ± 2.92	28.68 ± 1.16	0.30 ± 0.04	4.20 ± 0.88	1.05 ± 0.23	-3.74 ± 2.69	9.27 ± 0.54	1.83×10^{4}
HD 218566	MS	2:1	225.54 ± 0.14	7.60 ± 0.24	0.90 ± 0.07	2.47 ± 0.26	0.17 ± 0.17	0.61 ± 0.11	0.82 ± 0.19	1.70×10^{4}
GJ 649	MS	2:1	601.38 ± 2.17	11.55 ± 0.31	3.41 ± 0.11	2.88 ± 0.39	1.83 ± 0.32	0.86 ± 0.31	6.18 ± 0.46	1.10×10^{4}
75 Ceti	nMS	2.17:1	694.41 ± 1.40	37.12 ± 0.74	3.84 ± 0.04	3.31 ± 0.70	4.32 ± 0.21	4.59 ± 0.48	0.40 ± 0.47	1.04×10^{4}
HD 27894	MS	2.17:1	17.97 ± 0.01	58.40 ± 0.49	5.04 ± 0.07	3.91 ± 0.73	0.50 ± 0.17	-29.59 ± 10.58	82907.65 ± 1.94	6118.83
HD 32518	nMS	2.17:1	157.45 ± 0.19	117.90 ± 2.19	6.00 ± 0.03	9.90 ± 2.38	4.92 ± 0.17	13.84 ± 4.14	-11.90 ± 1.26	3259.36
HD 231701	MS	2.17:1	141.30 ± 0.35	41.56 ± 1.48	1.32 ± 0.10	12.81 ± 3.37	2.98 ± 0.23	-8.44 ± 6.3	-0.02 ± 1.81	993.37
γ_1 Leonis	nMS	2:1	428.87 ± 0.17	206.77 ± 0.62	1.09 ± 0.01	31.92 ± 0.70	4.99 ± 0.03	9.44 ± 0.86	178.66 ± 0.53	598.25
HD 2638	MS	2.17:1	3.45 ± 0.00	66.75 ± 0.45	5.26 ± 0.09	1.36 ± 0.46	4.44 ± 0.35	49.46 ± 13.72	9619.32 ± 2.32	407.32
HD 31253	MS	2:1	464.44 ± 0.64	10.75 ± 0.34	0.27 ± 0.06	3.58 ± 0.33	2.87 ± 0.13	0.59 ± 0.17	1.96 ± 0.25	295.22
HD 221287	MS	2.17:1	452.51 ± 1.00	73.50 ± 1.37	2.10 ± 0.04	12.95 ± 2.46	2.34 ± 0.05	10.72 ± 2.36	-21861.09 ± 1.29	124.62
HD 190647	MS	2:1	931.10 ± 76.66	30.25 ± 2.44	3.81 ± 0.44	7.90 ± 1.04	2.94 ± 0.93	-20.43 ± 5.67	-40266.66 ± 1.65	82.74
HD 220773	MS	2:1	2877.77 ± 87.74	14.60 ± 1.51	1.63 ± 0.12	10.15 ± 1.42	4.80 ± 0.25	2.89 ± 0.72	-4.96 ± 1.05	54.06
HD 330075	MS	2.17:1	3.39 ± 0.00	106.81 ± 0.73	5.87 ± 0.01	0.49 ± 0.49	3.22 ± 1.21	-13.2 ± 4.57	61278.58 ± 0.40	37.21
HIP 79431	MS	2.17:1	113.99 ± 0.40	155.87 ± 2.20	2.19 ± 0.02	30.65 ± 1.82	4.00 ± 0.09	-385.58 ± 24.21	10.53 ± 1.90	26.99

Table 3.1: The preferred fixed circular double models of the extended sample that has Bayes factor larger than the 95th percentile of the synthetic systems. Under star type, MS and nMS refer to Main Sequence and non Main Sequence stars, respectively. The order of the table is by Bayes factor.

3.3.1 Real Results

Our primary analysis does not include stellar jitter (even though our synthetic data has jitter added to the simulated data). We made this choice for a number of reasons. One is that since we are considering a fixed period ratio, only noise that occurs at that specific frequency could produce a spurious signal. Most sources of stellar noise occur on much different time-scales. The stellar rotation periods (typically ranging from 4–40 days [45]) are shorter than the inner planet periods for most of these systems. Stellar p-mode oscillations have typical time-scales of 5-15 minutes [46]. And, surface granulation variations last minutes to hours, with the largest granules remaining on the surface of stars for about a day [47, 46]. The time-scales of long term stellar activity arising from the cyclical appearance of starspots are on the order of years to decades [48].

These facts support the interpretation that stellar noise is not the cause of the inner companion signal for the majority of our systems. Nevertheless, we did a separate analysis that included a white noise jitter term to all models and found that 5 of the 15 systems still remain in the 95th percentile of likely two-planet systems, 4 of which prefer the 2.17:1 architecture. Thus, even if we adopt the much more conservative approach—which assumes stellar jitter does indeed affect our data at precisely the relevant time scales—we still see a number of systems that favor the two-planet models.

While our results are primarily from the fixed circular double models, we examined the results of a floating circular double model in order to estimate the likely distribution of orbital periods for the inner companion. We analyzed the real and synthetic systems with the floating circular double model and find an even larger portion of the systems that have Bayes factors above the 95th percentile—19 systems, 32% of the main sequence sample, (41 systems, 43%, of the extended sample) with an estimated false positive rate of 0.95 ± 0.97 (2.1 ± 1.4 for the extended sample). Of these 19 systems, 9 prefer the floating circular double model, 6 prefer the fixed 2.17:1 model, and the remaining 4 prefer the 2:1 double circular model. 4 systems remain in the 95th percentile when including stellar jitter in the model as a white noise term.

We show the period ratio posteriors that result from this analysis for these 19 systems and the synthetic systems in Figure 3.26. These histograms show the combined, period ratio posterior distribution from fitting the circular double model without a constraint on orbital periods to the 19 systems and to the synthetic systems. The distribution for the synthetic systems clearly shows the degeneracy at the location of the 2:1 MMR. If the real systems (in orange) were single-planet systems, then the expected distribution should be the same as for the synthetics. However, the two distributions differ significantly. In fact, the distribution for the real systems mirrors the period ratio distribution from the *Kepler* data [27]. Most of the combined posteriors favor period ratios just wide of the 2:1 or between 2.15 and 2.2. Only a few systems preferred the circular double model near the 2:1 because the degeneracy is located at the 2:1 and the power to distinguish between the models diminishes. Thus, in that regime, more data with appropriate phase-sampling is essential to distinguish between the models.

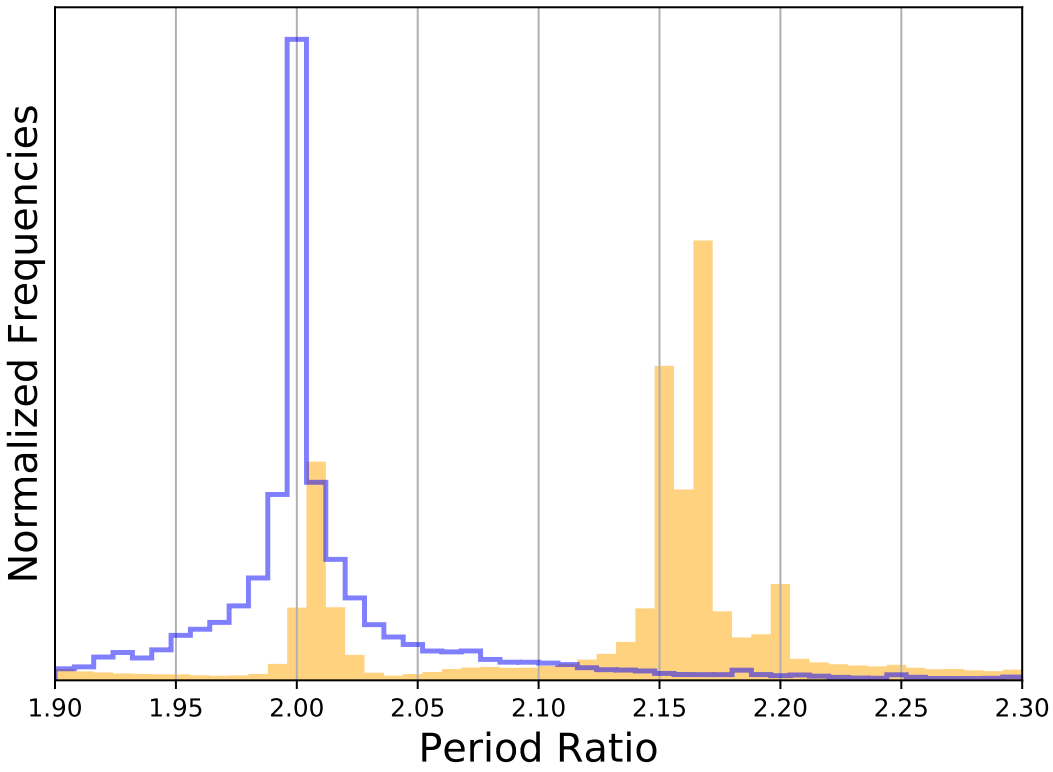

Figure 3.26: The posterior distributions for the period ratio when considering the inner planet period as a free parameter. The blue distribution is the 1,000 synthetic single eccentric systems. This distribution peaks at 2:1—the location of the degeneracy. The orange distribution is the 19 systems with Bayes factors larger than the 95th percentile thresholds that are hosted by main sequence stars.

3.4 Discussion

There are currently 395 confirmed solitary RV planets, and we reanalyzed about fifteen percent of them. Our extended sample contains nearly a quarter of the confirmed RV planets. The distribution of eccentricities, periods, velocity semi-amplitudes, etc. of our extended sample is shown in Figure 3.2. If the 15 systems in the main sequence sample (30 systems in the extended sample) that we identify are indeed circular doubles, then they would increase the number of RV multi-planet systems by ∼12.5% (∼25%) since there are 120 confirmed systems reported with at least two planets discovered by RV. They would also significantly alter the estimated mixture of these two architectures—shifting the relative importance of their implied dynamical histories.

If the fraction of misidentified single eccentrics in the entire NASA Archive is similar to the misidentification fraction seen in our sample, then there could be as many as ∼100 planets missing, or ∼15% of the overall confirmed RV planets (∼120 in the extended sample, or ∼18% of the overall confirmed RV planets). Moreover, the apparent propensity for some systems to cluster around period ratios near 2.17 is a further indication that there is something fundamental, but still unknown, that attracts planet pairs into this period ratio. We encourage observers to consider planning follow-up observations of these systems and make additional measurements at phases where the degeneracy is at its weakest. New observations near these phases could confirm or refute the existence of these putative interior companions. The success of such a campaign opens the door to identifying the architectures of the systems where the preferred model is still ambiguous. I go into more detail in this in Chapter 5.

3.4.1 Follow-up Observations with the Automated Planet Finder

A subset of the systems that show the best evidence for being a double planet system are being observed by the *Automated Planet Finder* (*APF*, Howard Isaacson, [49]. *APF* is restricted to stars that have a declination larger than -22^o and focuses on stars brighter a V-band magnitude of 9, i.e. $V < 9$. The target list contains 30 stars. Their stellar and circular double model properties, are listed on Table 3.2. The systems are listed in order of largest BF to smallest. The current number of data points and the observation cadence is also included. Observational cadence is the number of days between a single observation. It is set to a tenth of the proposed outer planet orbit, but not less than every two weeks—hence some systems have a cadence of 14 days.

Thus far, we have obtained additional data for seven systems: HD 2952, Hamal, HD 32518, HD 81688, 14 Andromedæ, γ_1 Leonis, and HD 86081. These systems are all non-main sequence stars except for HD 86081. The new analysis for system is shown in Table 3.3. The largest of the newly measured BFs for the circular double models were larger than its original for two of the systems, α Arietis and γ_1 Leonis. HD 86081 is a main sequence star that our preliminary results indicate the fixed 2:1 circular double model fits more likely than the rest. Every calculated Bayes factor is larger than with the archival data. The Bayes factor between the fixed 2:1 circular double and the single eccentric models is still the largest.

Despite the Bayes factors going down for four systems, three of them still show overwhelming evidence of an additional planet. Most of these stars are not on the main sequence,

and thus their stellar surfaces are more active than main sequence stars. As mentioned in Section 3.3.1, this activity could be misinterpreted as a planet signal, but it is difficult to conceive of stellar activity remaining at the frequency interval where the new planets are for the entire range of the timeseries. The analysis of these new fits seem to indicate that there is indeed a planet present in a third to six-sevenths of the newly observed sample, but the result it would be more robust if there was a more realistic noise model that takes into account the stellar properties. The timeseries for the best circular double model along with the single eccentric model for all seven systems are shown in Figures 3.27–3.54.

Star Name	P_{out} [day]	K_{out} [m s^{-1}]	K_{in} [m s^{-1}]	Preferred Fixed CD Model	Stellar Type	V-mag	N_{data}	Cadence [day]
HD 240237	770.449	69.754	57.522	2.17:1	nonMS	8.19	40	77.0449
HD 2952	318.826	23.159	15.895	2.17:1	nonMS	5.922	63	31.8826
Hamal	372.596	29.666	10.251	2.17:1	nonMS	2	678	37.2596
HD 96127	632.389	96.664	41.016	2.17:1	nonMS	7.43	50	63.2389
HD 95089	496.568	20.652	6.731	2.17:1	nonMS	7.946	22	49.6568
11 UrsæMinoris	513.221	185.384	19.425	2.17:1	nonMS	5.016	58	51.3221
HD 136418	474.991	41.621	10.413	2.17:1	nonMS	7.879	24	47.4991
HD 81688	184.079	60.052	6.955	2.17:1	nonMS	5.393	81	18.4079
HD 206610	628.838	34.503	6.924	2.17:1	nonMS	8.346	24	62.8838
HD 216770	121.769	14.047	33.86	2:1	MS	8.094	16	14
HD 114386	955.463	33.788	14.078	2.17:1	MS	8.73	58	95.5463
o CoronæBorealis	187.613	33.637	8.642	2:1	nonMS	5.51	85	18.7613
HD 101930	71.301	17.929	2.373	2.17:1	MS	8.21	16	14
14 Andromedæ	185.886	99.797	6.555	2.17:1	nonMS	5.22	34	18.5886
HD 180902	482.218	28.676	4.197	2.17:1	nonMS	7.778	12	48.2218
HD 218566	225.543	7.599	2.473	2:1	MS	8.63	56	22.5543
75 Ceti	694.412	37.124	3.306	2.17:1	nonMS	5.35	74	69.4412
HD 32518	157.452	117.9	9.897	2.17:1	nonMS	6.421	58	15.7452
HD 231701	141.299	41.561	12.813	2.17:1	MS	8.97	17	14.1299
γ_1 Leonis	428.868	206.773	31.921	2:1	nonMS	1.98	205	42.8868
HD 31253	464.439	10.748	3.581	2:1	MS	7.133	39	46.4439
HD 221287	452.506	73.499	12.947	2.17:1	MS	7.807	26	45.2506
HD 190647	931.104	30.254	7.898	2:1	MS	7.775	21	93.1104
HD 220773	2877.765	14.598	10.153	2:1	MS	7.096	43	287.7765
HD 86081	2.136	207.661	2.519	2:1	MS	8.73	26	14
HD 179079	14.474	6.411	0.795	2.17:1	MS	7.95	74	14.474
18 Delphini	986.940	119.243	5.995	2:1	nonMS	5.506	51	98.694
HD 139357	1016.707	143.187	10.776	2:1	nonMS	5.964	49	101.671
ϵ Tauri	585.700	94.330	13.707	2.17:1	nonMS	3.54	20	58.570
ϵ CoronæBorealis	417.810	113.120	10.695	2:1	nonMS	4.15	52	41.781

Table 3.2: The target list submitted to *APF*. These systems show the best evidence for the circular double case. They are bright enough—V-mag$<$ 9—and in the right spot of the sky for observation. The proposed outer planet orbital period and velocity semi-amplitude is listed, along with the velocity semi-amplitude of the proposed inner planet. The current number of data points and the observational cadence is also shown.

Star Name		N_{data}	BF 2:1 | SE	BF 2.17:1 | SE	BF CD | SE
HD 2952	Original	63	10	57	**69**
	New	91	-3	37	*55*
	Δ Trend	+30.8%	−	−	− [2.21:1]
HD 12929 Hamal α Arietis	Original	678	11	49	53
	New	699	-6	225	**226**
	Δ Trend	+3.1%	−	+	+ [2.18:1]
HD 32518	Original	58	0	3	**6**
	New	84	*-2*	-12	-3
	Δ Trend	+44.8%	−	−	− [1.99:1]
HD 81688	Original	81	1	11	**11**
	New	104	1	4	*9*
	Δ Trend	+28.4%	+	−	− [2.16:1]
HD 221345 14 Andromedae	Original	34	1	4	**14**
	New	59	*4*	-17	3
	Δ Trend	+74%	+	−	− [1.99:1]
HIP 50583 γ_1 Leonis	Original	205	2	$-\infty$	37
	New	231	-43	$-\infty$	**423**
	Δ Trend	+12.7%	−	−	+ [2.02:1]
HD 86081	Original	26	1	0	0
	New	33	**1**	1	1
	Δ Trend	+26.9%	+	+	+ [1.64:1⋆]

Table 3.3: The results of combining the new observations from *APF* with the old and sending them through the model comparison pipeline. The new number of data points is shown. Only the order of magnitude of the Bayes factor is shown. The Δ trend indicates if the Bayes factor got larger (+) or smaller (−). The bold element represents the largest Bayes factor. The italic element indicates the largest Bayes factor using just the new data with the old. ⋆ indicates that the model has a large variance in P_{in}.

(This page is intentionally left blank so that Figure 3.27 is on the back-side of the page.)

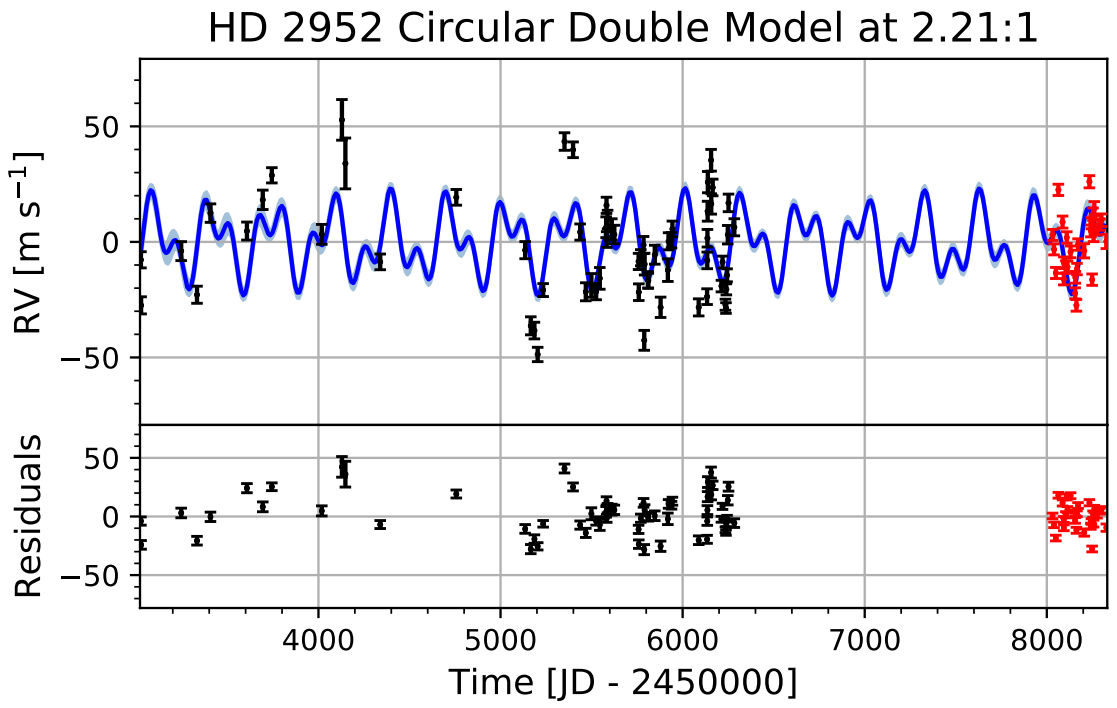

Figure 3.27: New time series for HD 2952. The archival offset data are in black, the new offset data from *APF* are in red, the circular double planet model in blue and the 68.2% error from the posterior distribution is in light blue. This is the best fit out of the single eccentric, fixed circular double at 2:1, fixed circular double at 2.17:1, and free circular double models. The period ratio for this model is 2.21:1.

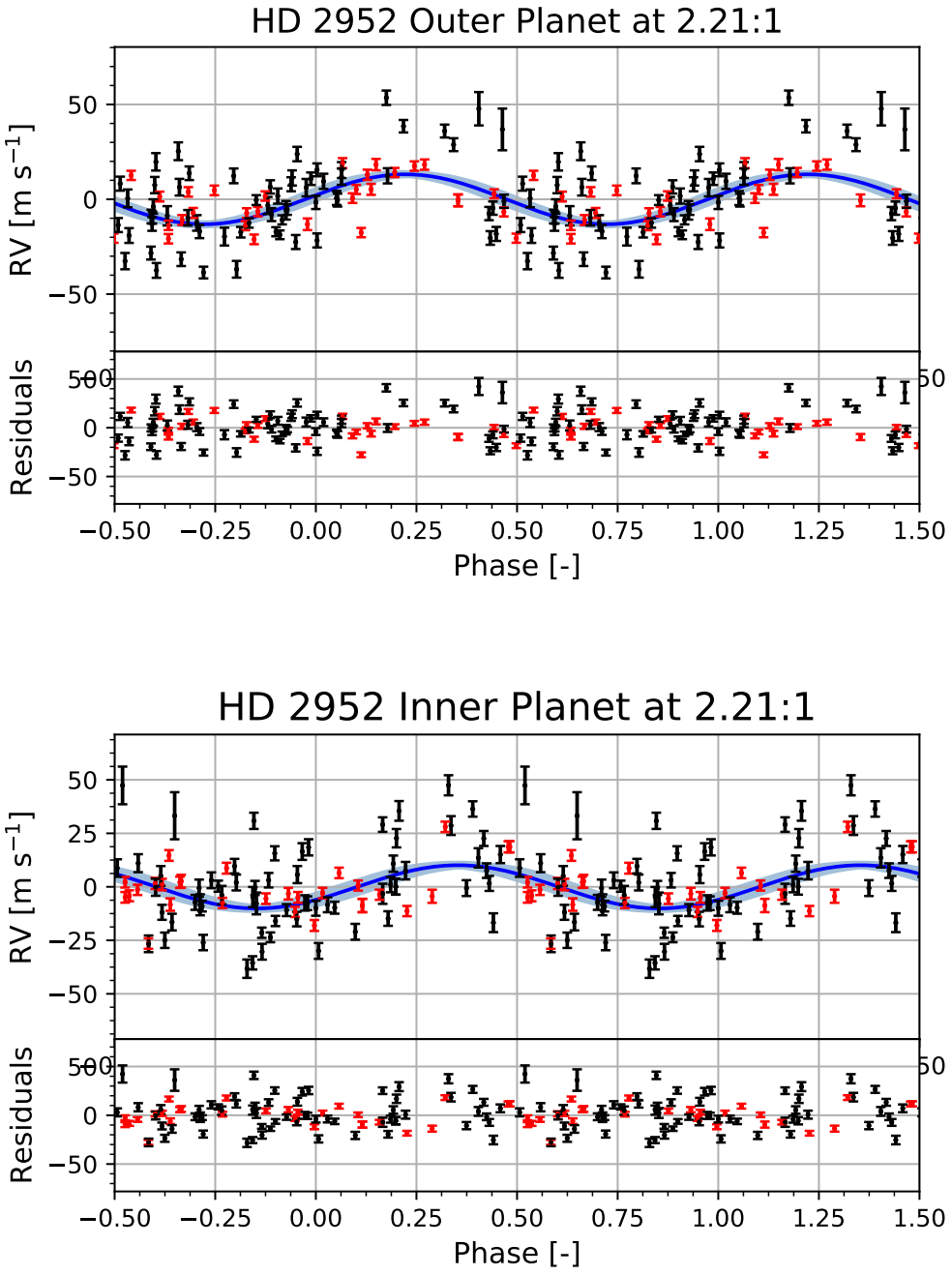

Figure 3.28: Phase plot for HD 2952. The archival offset data are in black, the new offset data from *APF* are in red, the circular double planet model in blue and the 68.2% error from the posterior distribution is in light blue. This is the best fit out of the single eccentric, fixed circular double at 2:1, fixed circular double at 2.17:1, and free circular double models. The outer planet is on top and the inner planet is on the bottom. The period ratio for this model is 2.21:1.

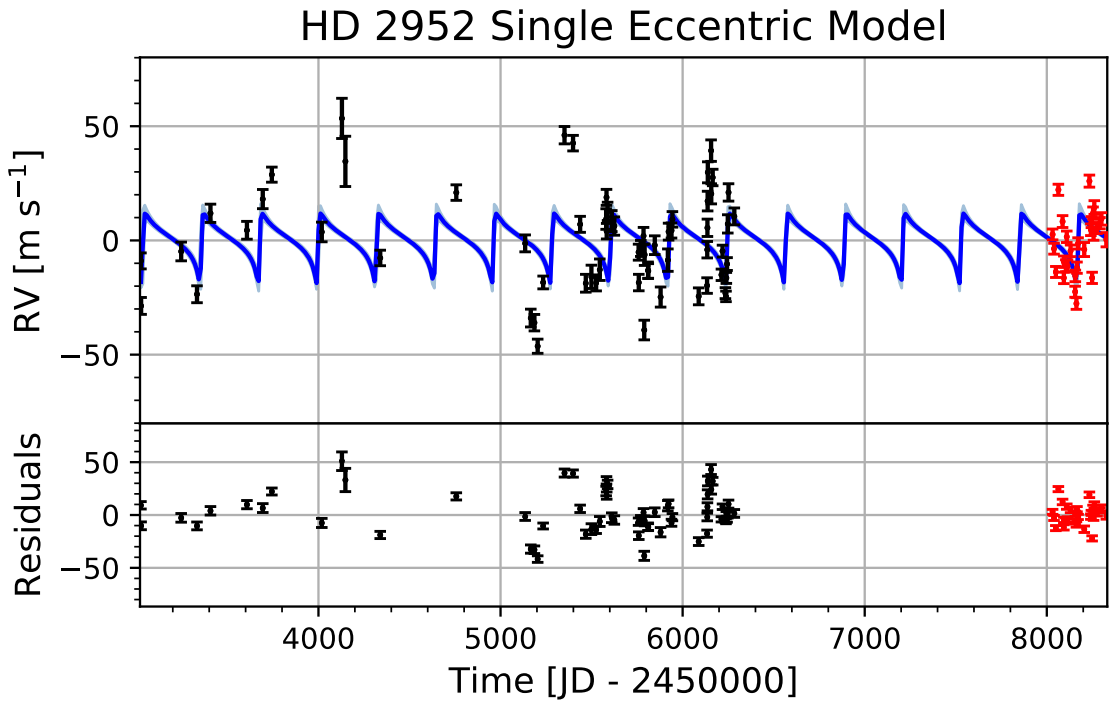

Figure 3.29: Time series for HD 2952. The offset data are in black, the new offset data from *APF* are in red, the single eccentric planet model in blue and the 68.2% error from the posterior distribution is in light blue.

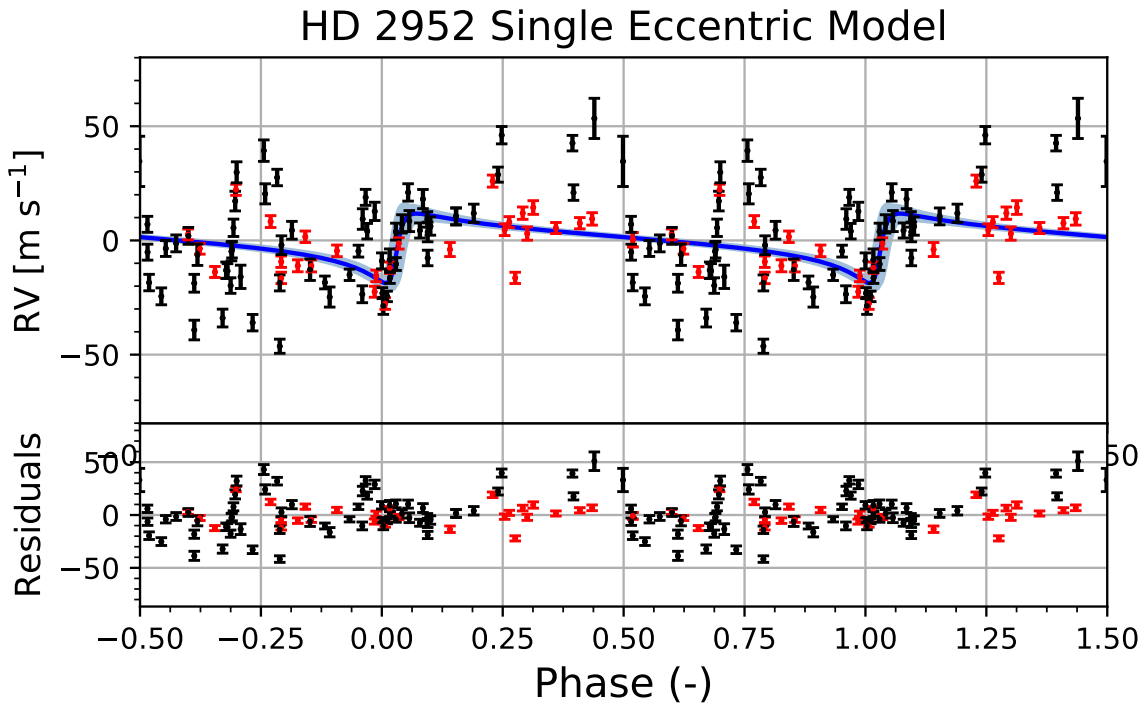

Figure 3.30: Phase plot for HD 2952. The offset data are in black, the new offset data from *APF* are in red, the single eccentric planet model in blue and the 68.2% error from the posterior distribution is in light blue.

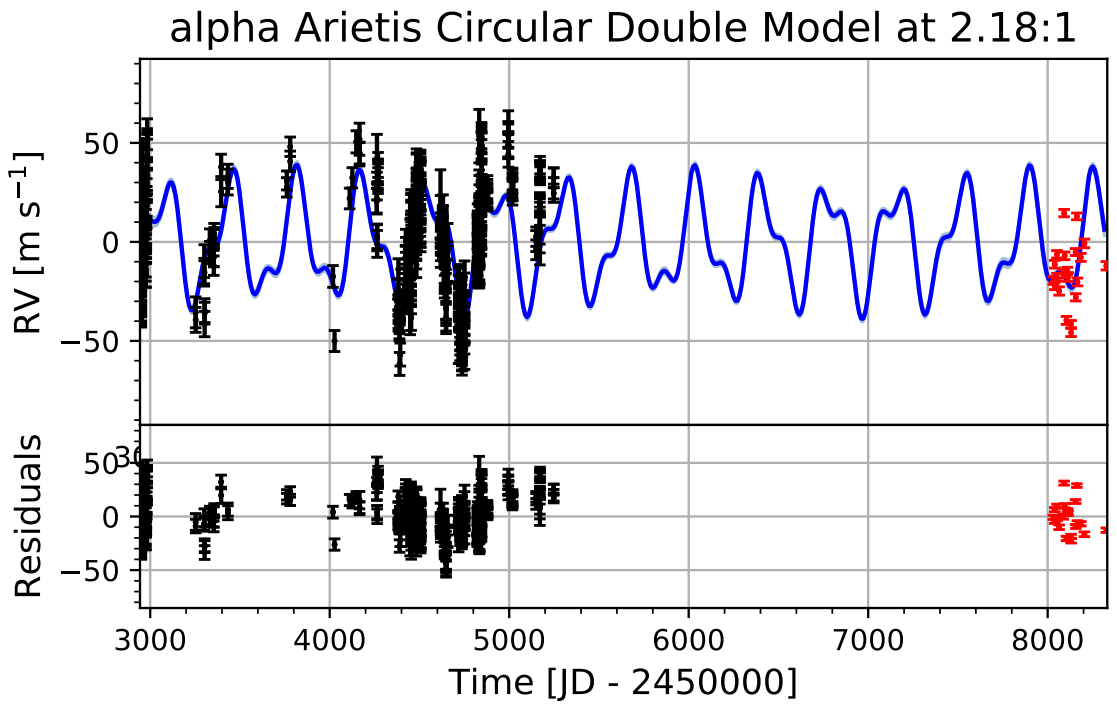

Figure 3.31: New time series for α Arietis. The archival offset data are in black, the new offset data from *APF* are in red, the circular double planet model in blue and the 68.2% error from the posterior distribution is in light blue. This is the best fit out of the single eccentric, fixed circular double at 2:1, fixed circular double at 2.17:1, and free circular double models. The period ratio for this model is 2.18:1.

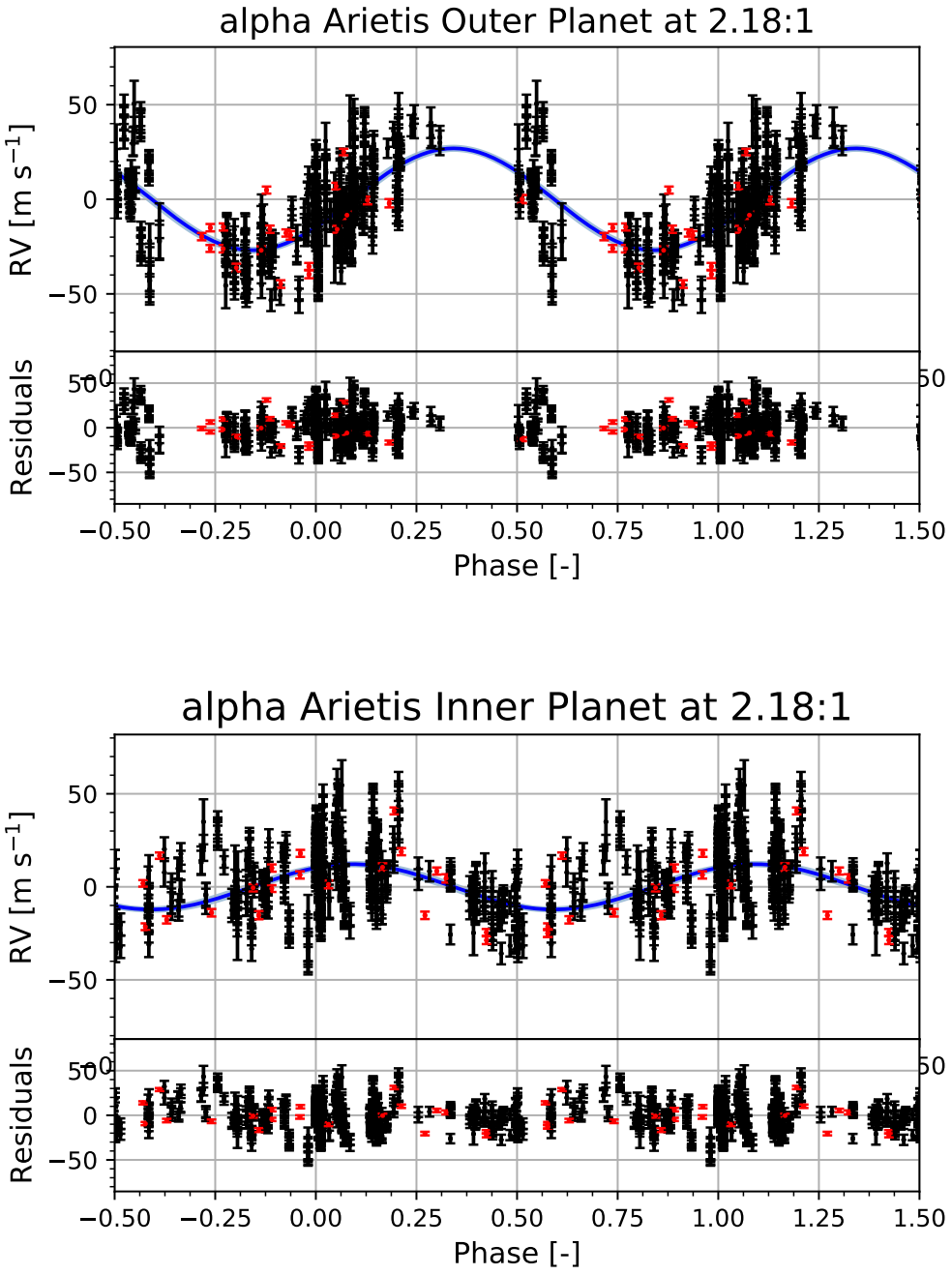

Figure 3.32: Phase plot for α Arietis. The archival offset data are in black, the new offset data from *APF* are in red, the circular double planet model in blue and the 68.2% error from the posterior distribution is in light blue. This is the best fit out of the single eccentric, fixed circular double at 2:1, fixed circular double at 2.17:1, and free circular double models. The outer planet is on top and the inner planet is on the bottom. The period ratio for this model is 2.18:1.

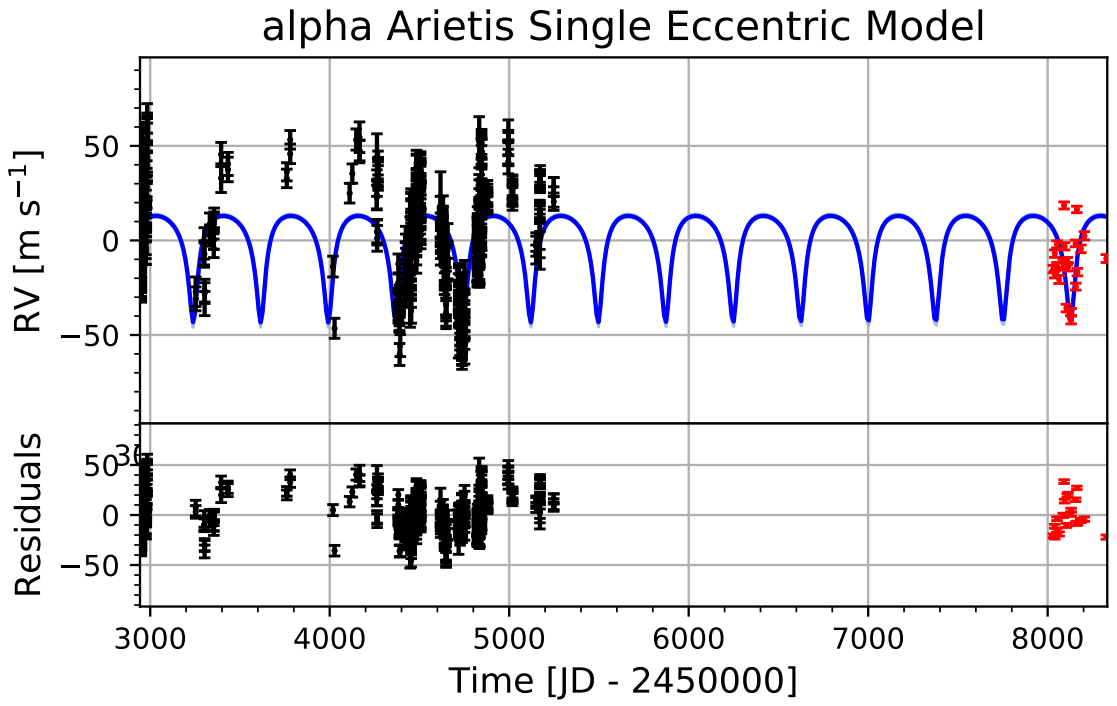

Figure 3.33: Time series for α Arietis. The offset data are in black, the new offset data from *APF* are in red, the single eccentric planet model in blue and the 68.2% error from the posterior distribution is in light blue.

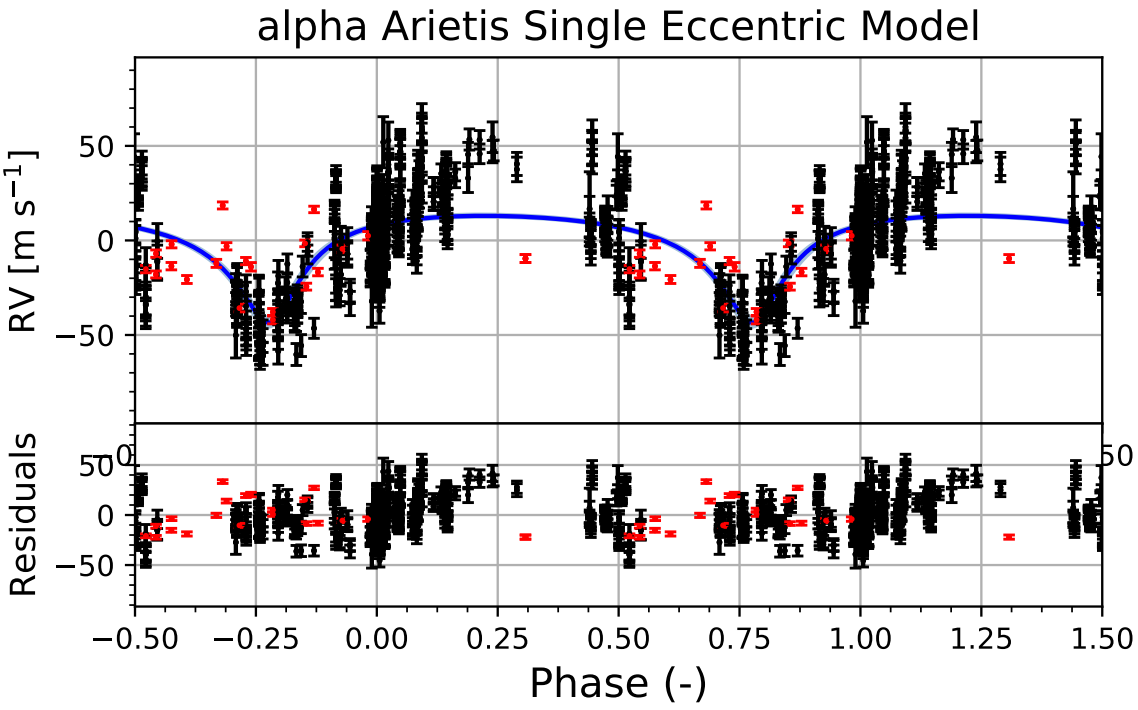

Figure 3.34: Phase plot for α Arietis. The offset data are in black, the new offset data from *APF* are in red, the single eccentric planet model in blue and the 68.2% error from the posterior distribution is in light blue.

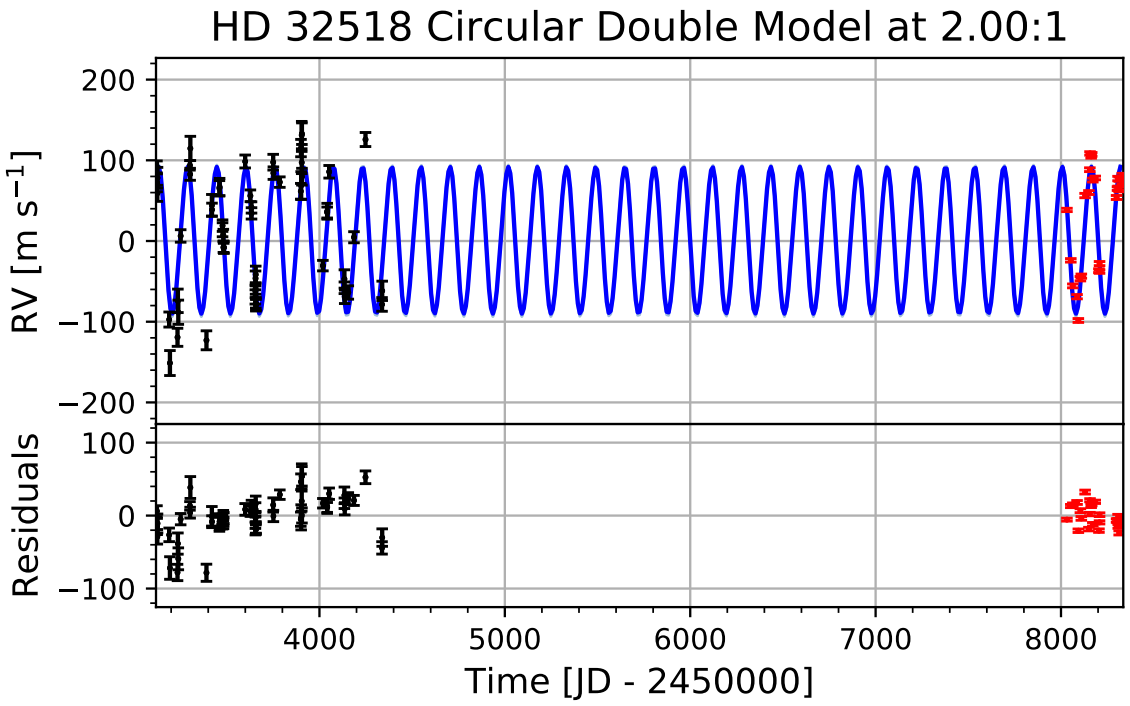

Figure 3.35: New time series for HD 32518. The archival offset data are in black, the new offset data from *APF* are in red, the circular double planet model in blue and the 68.2% error from the posterior distribution is in light blue. This is the best fit out of the single eccentric, fixed circular double at 2:1, fixed circular double at 2.17:1, and free circular double models.

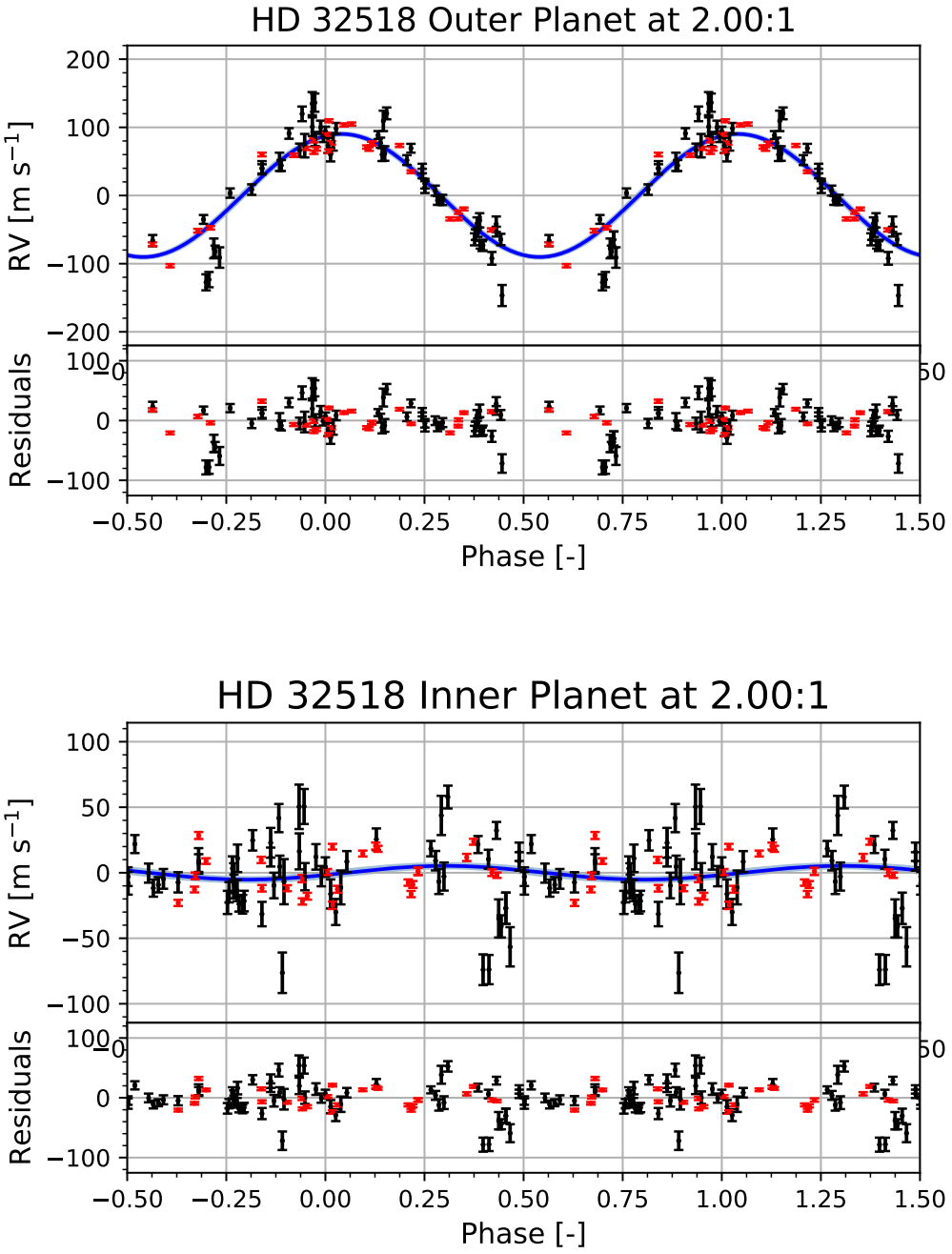

Figure 3.36: Phase plot for HD 32518. The archival offset data are in black, the new offset data from *APF* are in red, the circular double planet model in blue and the 68.2% error from the posterior distribution is in light blue. This is the best fit out of the single eccentric, fixed circular double at 2:1, fixed circular double at 2.17:1, and free circular double models. The outer planet is on top and the inner planet is on the bottom.

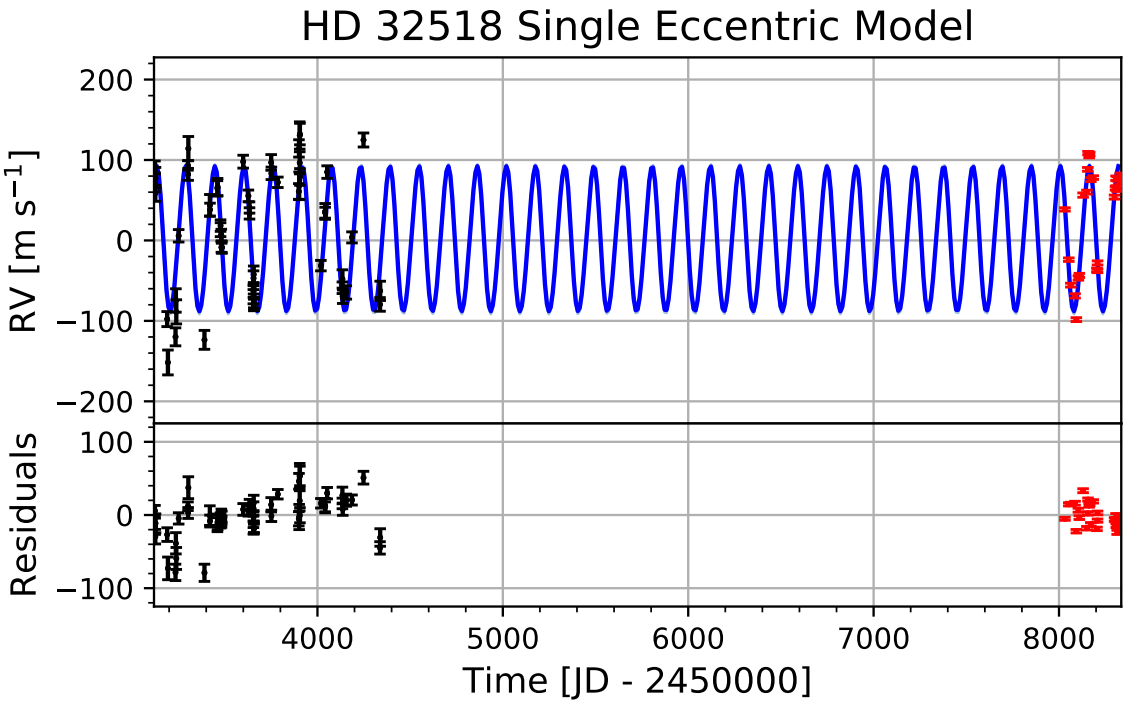

Figure 3.37: Time series for HD 32518. The offset data are in black, the new offset data from *APF* are in red, the single eccentric planet model in blue and the 68.2% error from the posterior distribution is in light blue.

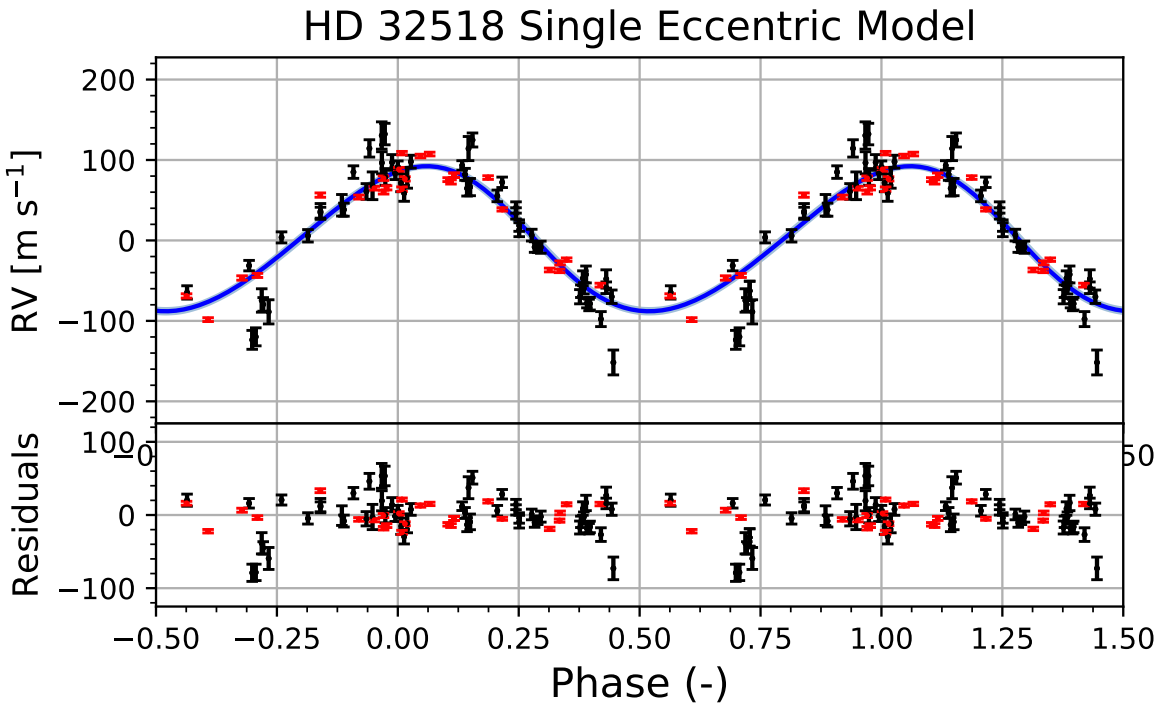

Figure 3.38: Phase plot for HD 32518. The offset data are in black, the new offset data from *APF* are in red, the single eccentric planet model in blue and the 68.2% error from the posterior distribution is in light blue.

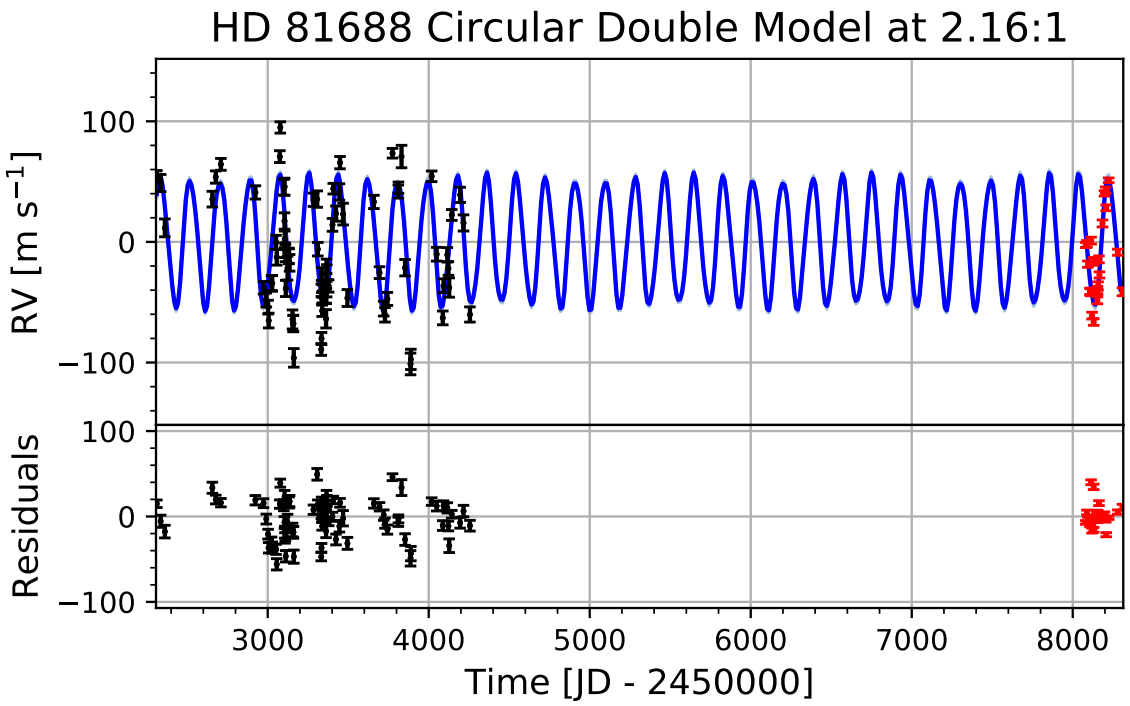

Figure 3.39: New time series for HD 81688. The archival offset data are in black, the new offset data from *APF* are in red, the circular double planet model in blue and the 68.2% error from the posterior distribution is in light blue. This is the best fit out of the single eccentric, fixed circular double at 2:1, fixed circular double at 2.17:1, and free circular double models. The period ratio for this model is 2.16:1.

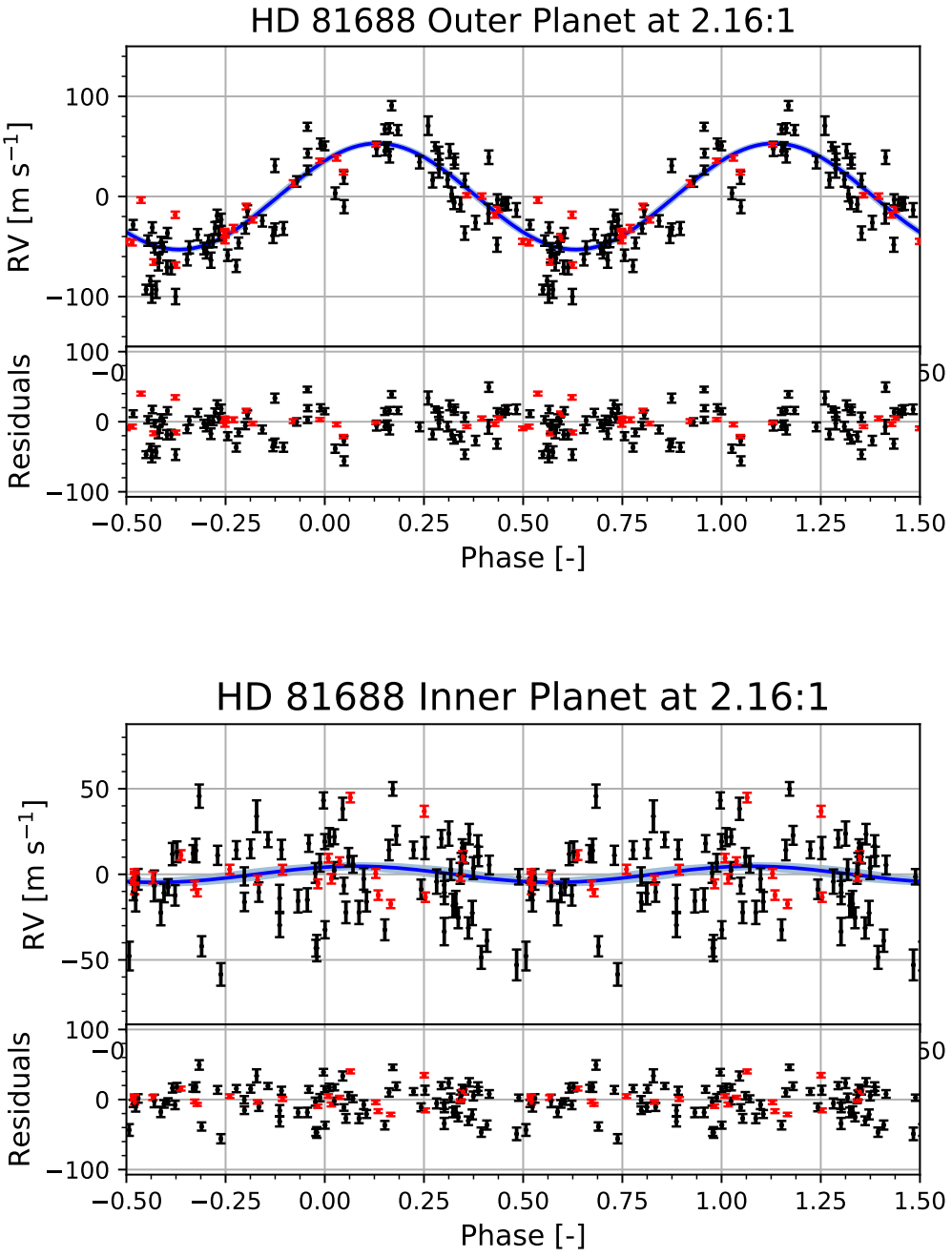

Figure 3.40: Phase plot for HD 81688. The archival offset data are in black, the new offset data from *APF* are in red, the circular double planet model in blue and the 68.2% error from the posterior distribution is in light blue. This is the best fit out of the single eccentric, fixed circular double at 2:1, fixed circular double at 2.17:1, and free circular double models. The outer planet is on top and the inner planet is on the bottom. The period ratio for this model is 2.16:1.

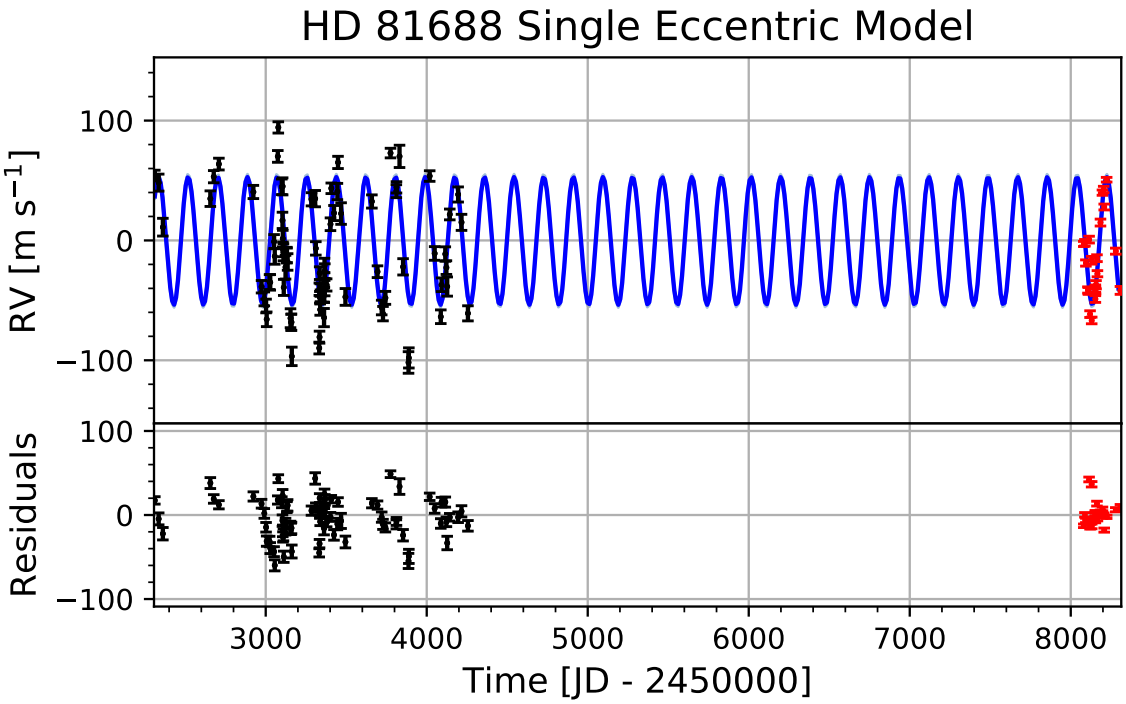

Figure 3.41: Time series for HD 81688. The offset data are in black, the new offset data from *APF* are in red, the single eccentric planet model in blue and the 68.2% error from the posterior distribution is in light blue.

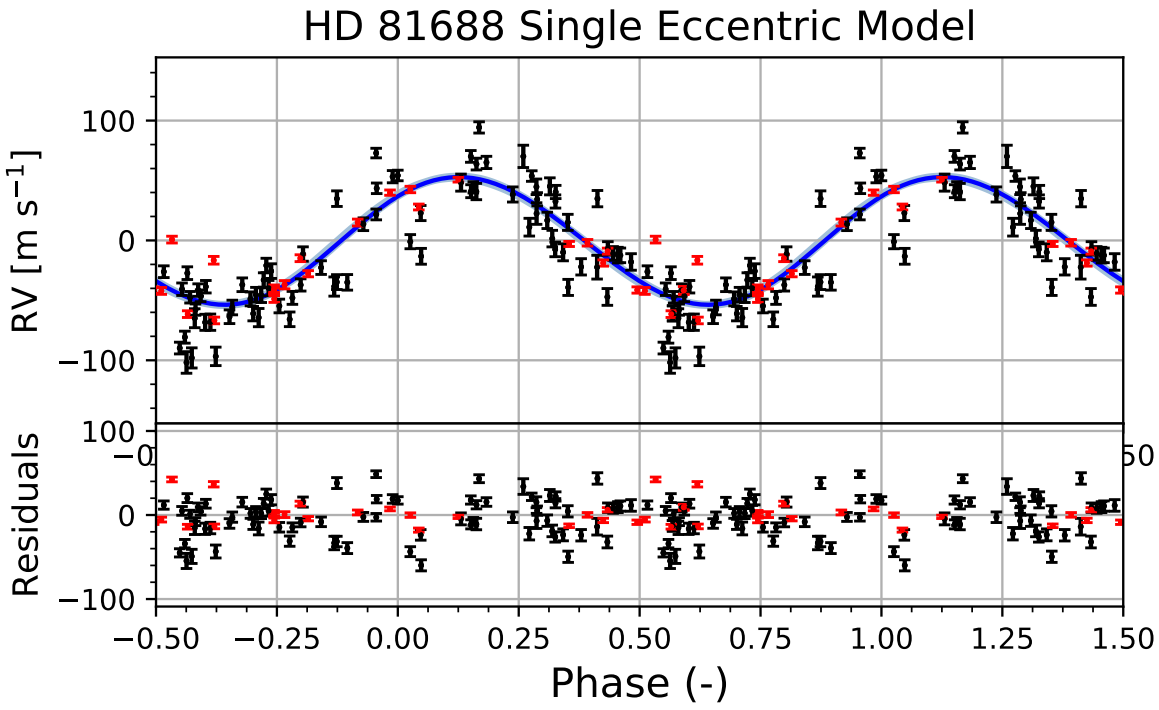

Figure 3.42: Phase plot for HD 81688. The offset data are in black, the new offset data from *APF* are in red, the single eccentric planet model in blue and the 68.2% error from the posterior distribution is in light blue.

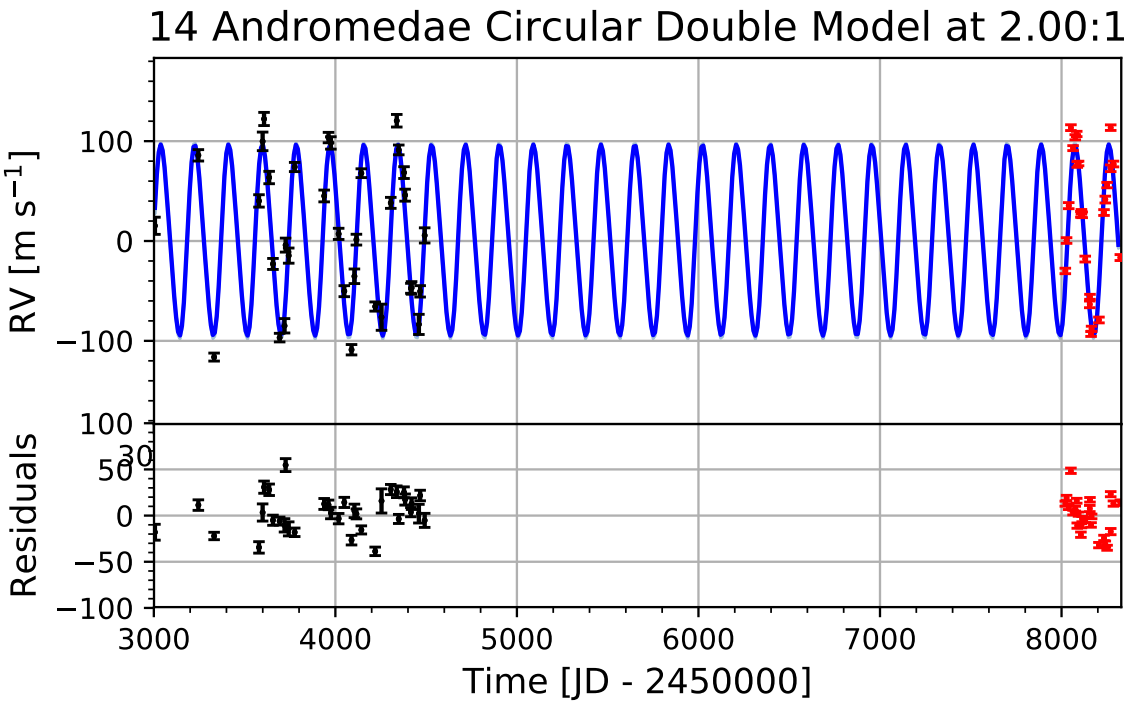

Figure 3.43: New time series for 14 Andromedae. The archival offset data are in black, the new offset data from *APF* are in red, the circular double planet model in blue and the 68.2% error from the posterior distribution is in light blue. This is the best fit out of the single eccentric, fixed circular double at 2:1, fixed circular double at 2.17:1, and free circular double models.

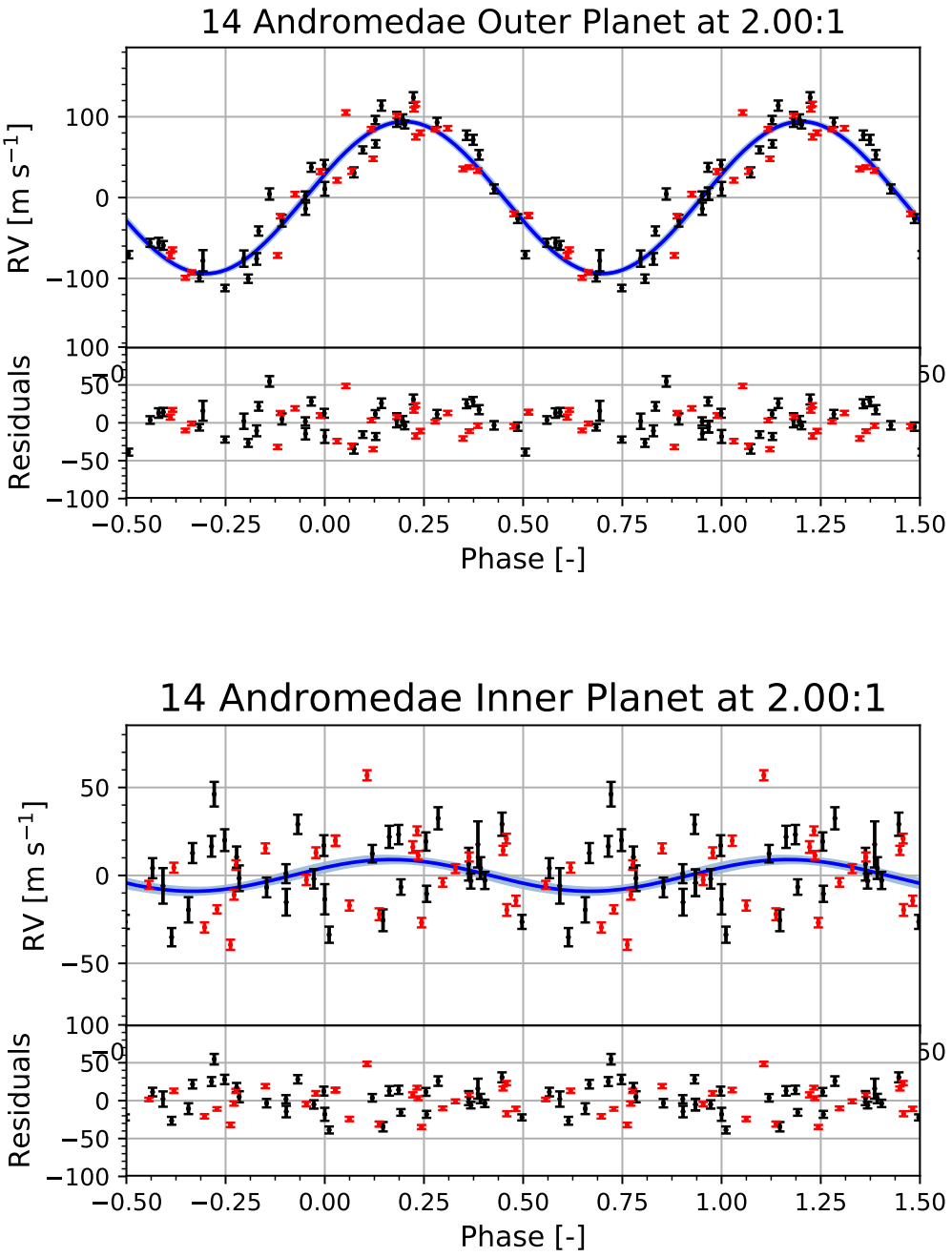

Figure 3.44: Phase plot for 14 Andromedae. The archival offset data are in black, the new offset data from *APF* are in red, the circular double planet model in blue and the 68.2% error from the posterior distribution is in light blue. This is the best fit out of the single eccentric, fixed circular double at 2:1, fixed circular double at 2.17:1, and free circular double models. The outer planet is on top and the inner planet is on the bottom.

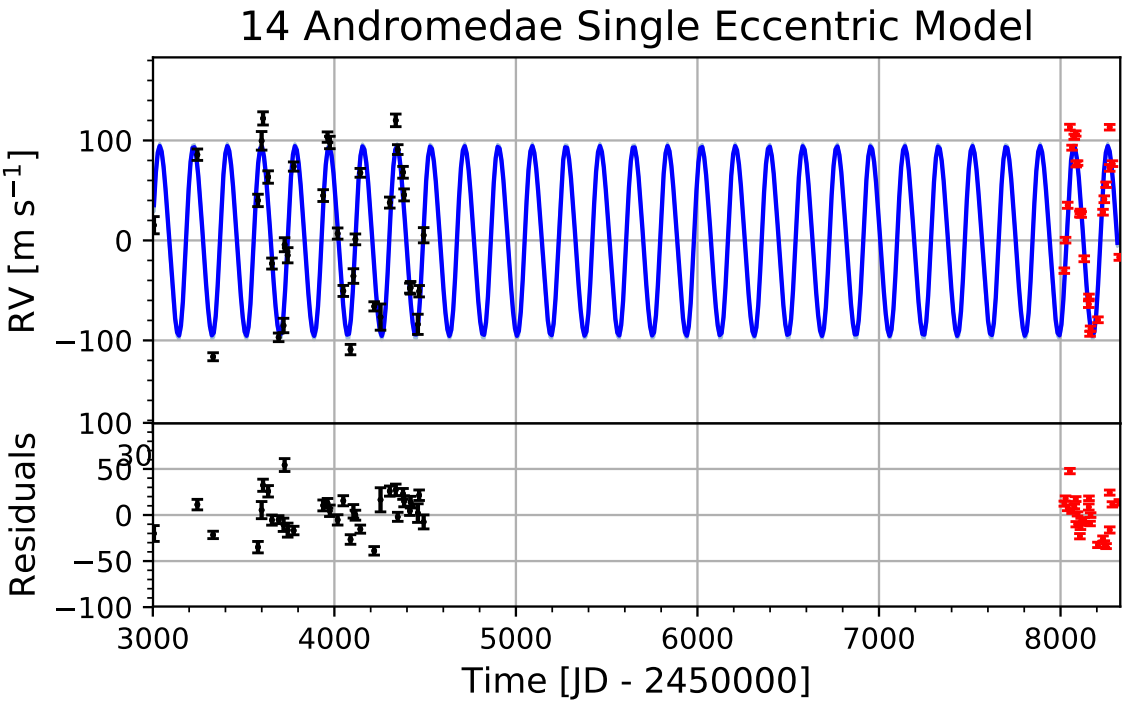

Figure 3.45: Time series for 14 Andromedae. The offset data are in black, the new offset data from *APF* are in red, the single eccentric planet model in blue and the 68.2% error from the posterior distribution is in light blue.

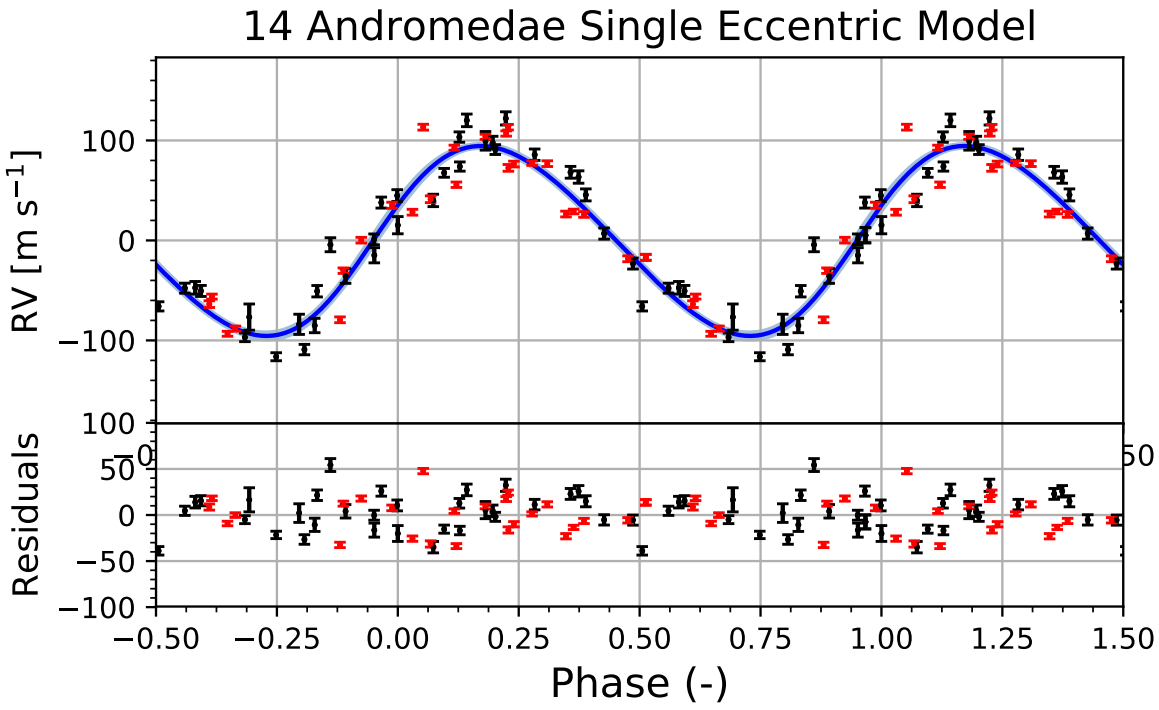

Figure 3.46: Phase plot for 14 Andromedae. The offset data are in black, the new offset data from *APF* are in red, the single eccentric planet model in blue and the 68.2% error from the posterior distribution is in light blue.

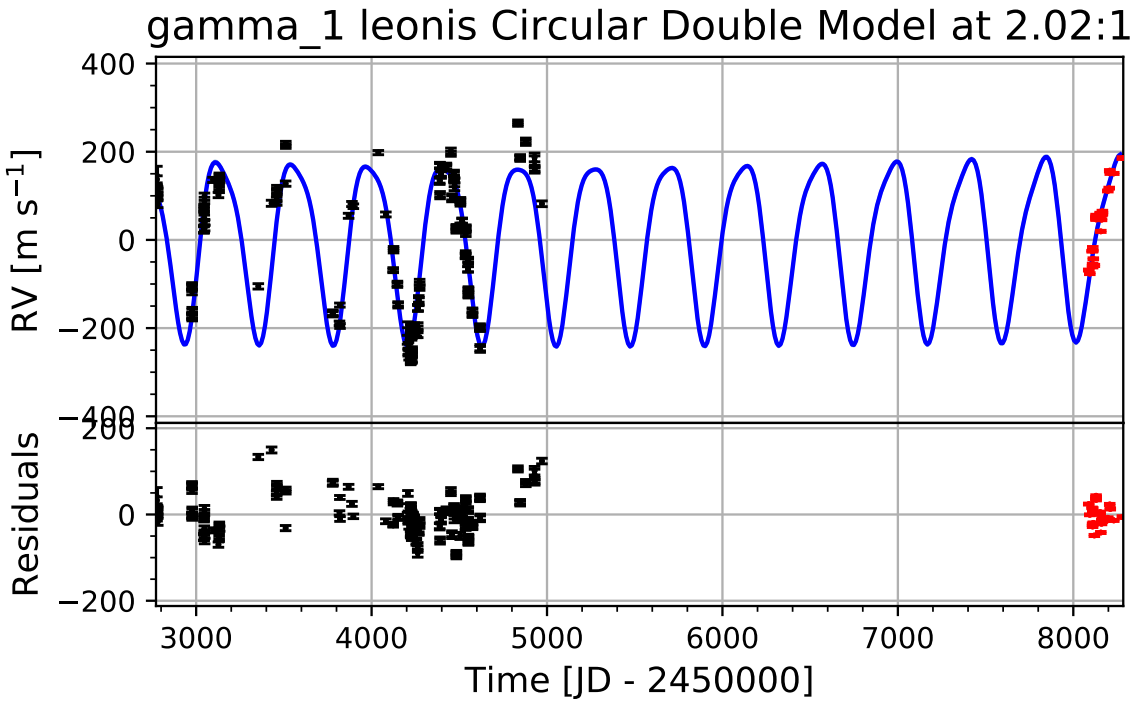

Figure 3.47: New time series for γ_1 Leonis. The archival offset data are in black, the new offset data from *APF* are in red, the circular double planet model in blue and the 68.2% error from the posterior distribution is in light blue. This is the best fit out of the single eccentric, fixed circular double at 2:1, fixed circular double at 2.17:1, and free circular double models. The period ratio for this model is 2.16:1.

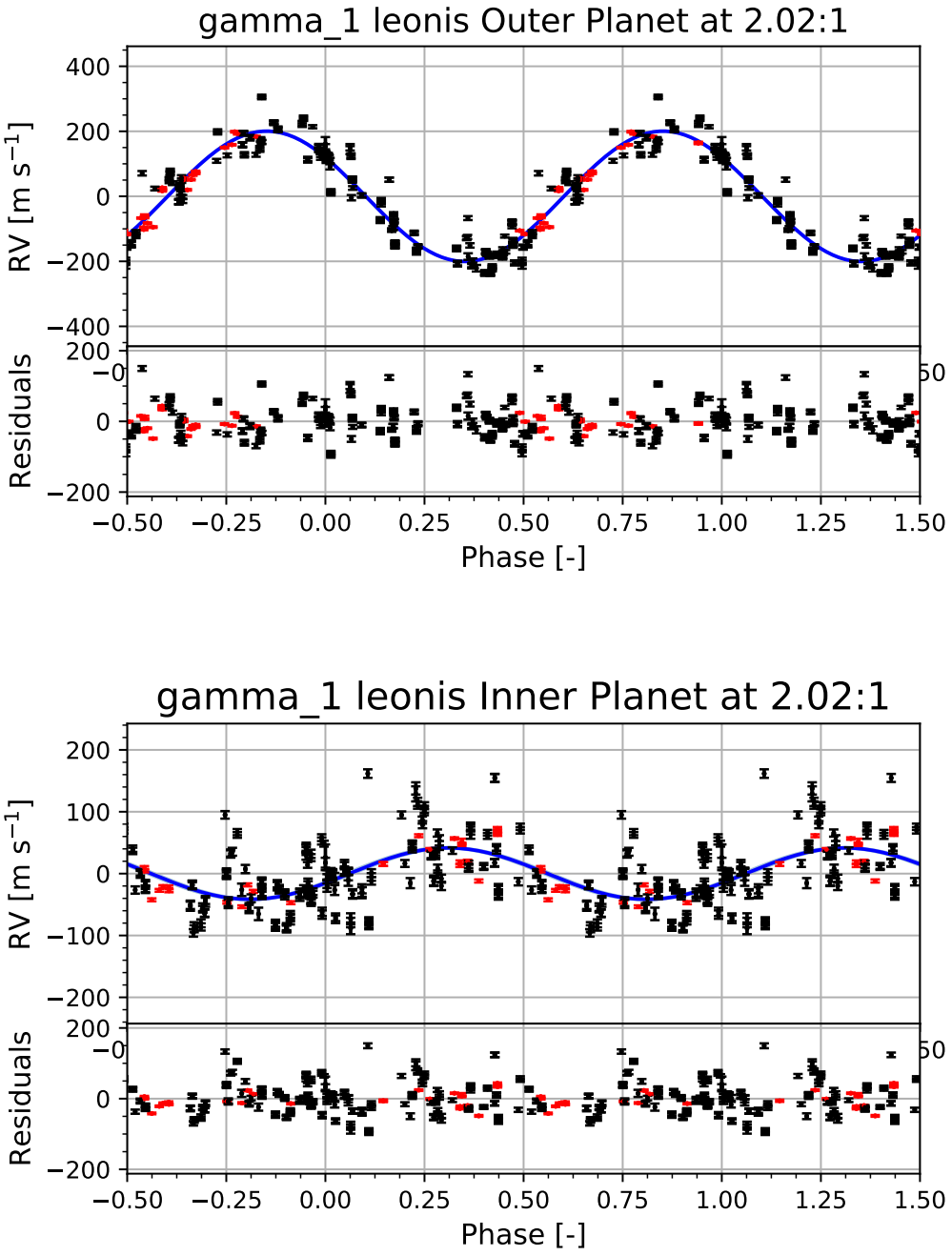

Figure 3.48: Phase plot for γ_1 Leonis. The archival offset data are in black, the new offset data from *APF* are in red, the circular double planet model in blue and the 68.2% error from the posterior distribution is in light blue. This is the best fit out of the single eccentric, fixed circular double at 2:1, fixed circular double at 2.17:1, and free circular double models. The outer planet is on top and the inner planet is on the bottom. The period ratio for this model is 2.16:1.

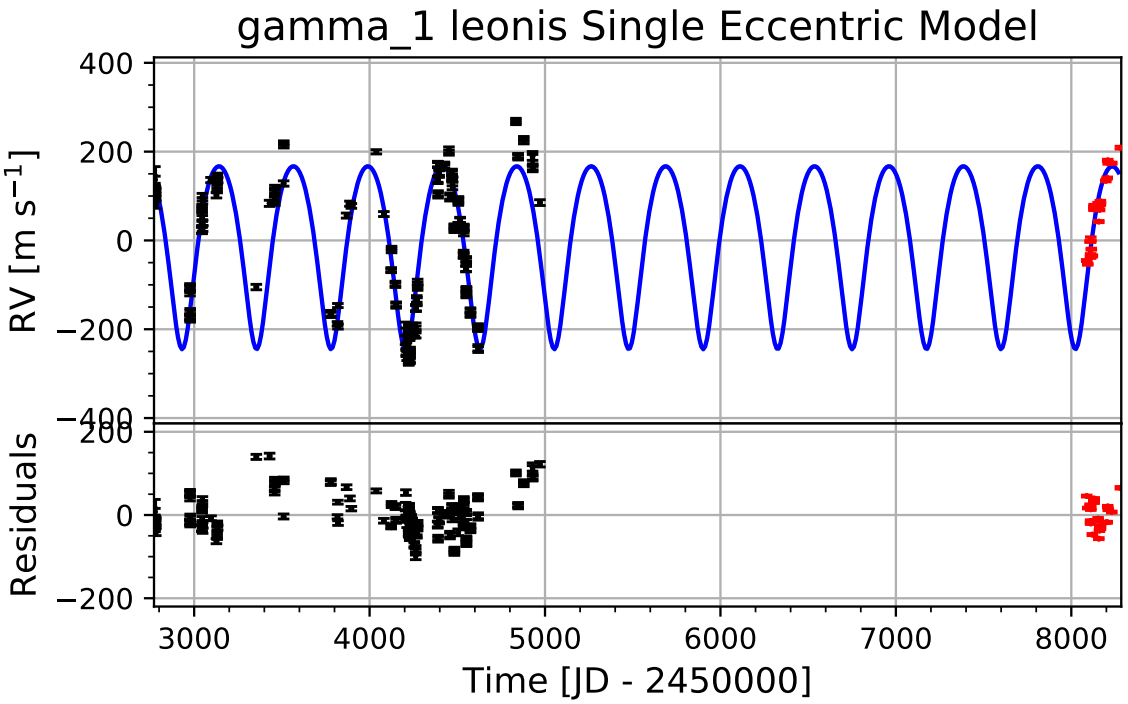

Figure 3.49: Time series for γ_1 Leonis. The offset data are in black, the new offset data from *APF* are in red, the single eccentric planet model in blue and the 68.2% error from the posterior distribution is in light blue.

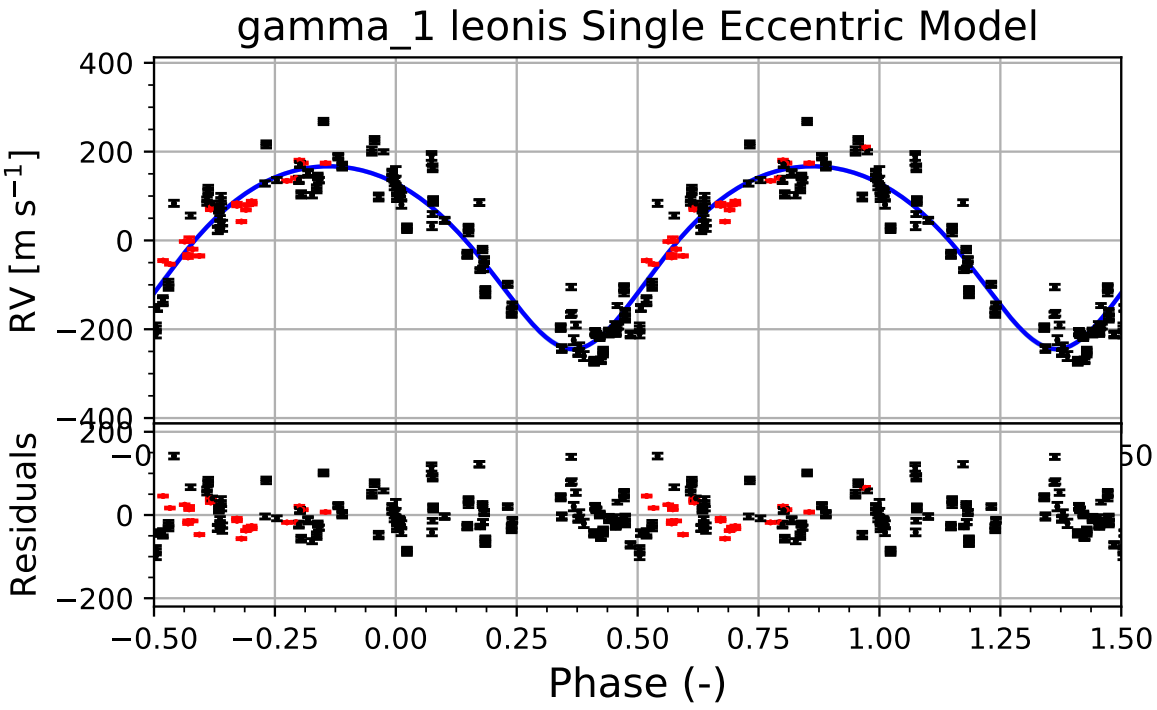

Figure 3.50: Phase plot for γ_1 Leonis. The offset data are in black, the new offset data from *APF* are in red, the single eccentric planet model in blue and the 68.2% error from the posterior distribution is in light blue.

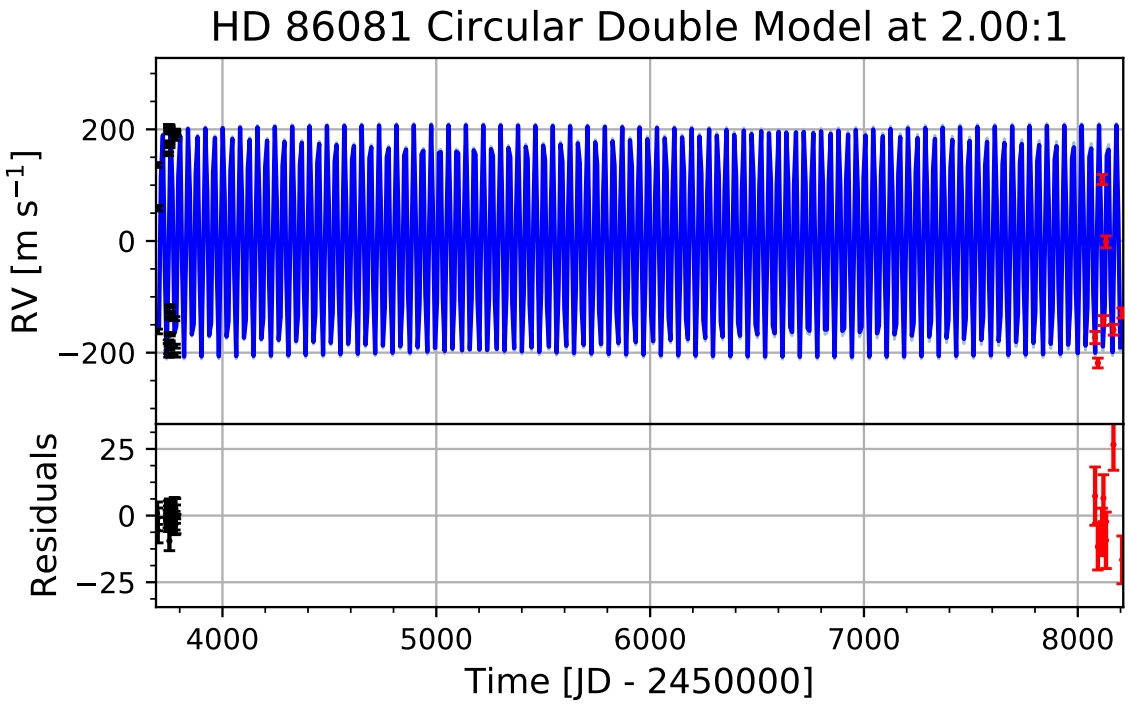

Figure 3.51: New time series for HD 86081. The archival offset data are in black, the new offset data from *APF* are in red, the circular double planet model in blue and the 68.2% error from the posterior distribution is in light blue. This is the best fit out of the single eccentric, fixed circular double at 2:1, fixed circular double at 2.17:1, and free circular double models.

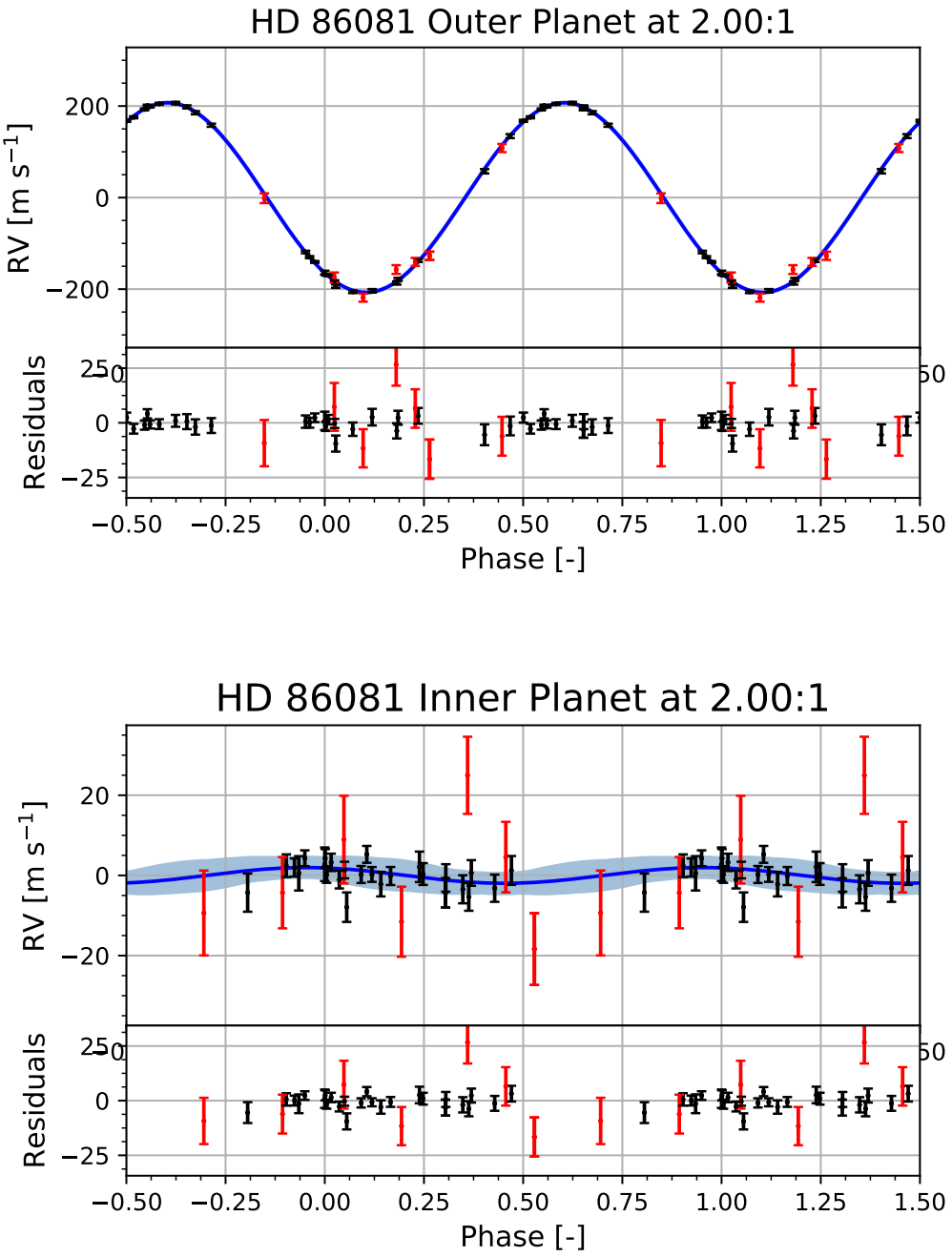

Figure 3.52: Phase plot for HD 86081. The archival offset data are in black, the new offset data from *APF* are in red, the circular double planet model in blue and the 68.2% error from the posterior distribution is in light blue. This is the best fit out of the single eccentric, fixed circular double at 2:1, fixed circular double at 2.17:1, and free circular double models. The outer planet is on top and the inner planet is on the bottom.

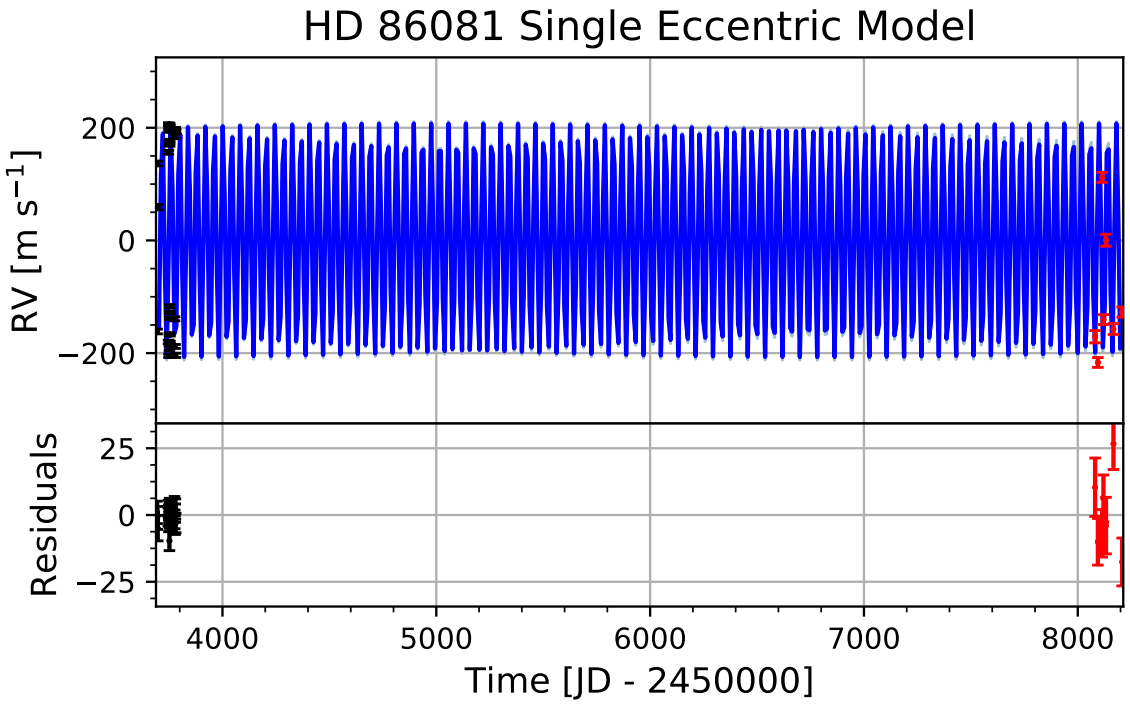

Figure 3.53: Time series for HD 86081. The offset data are in black, the new offset data from *APF* are in red, the single eccentric planet model in blue and the 68.2% error from the posterior distribution is in light blue.

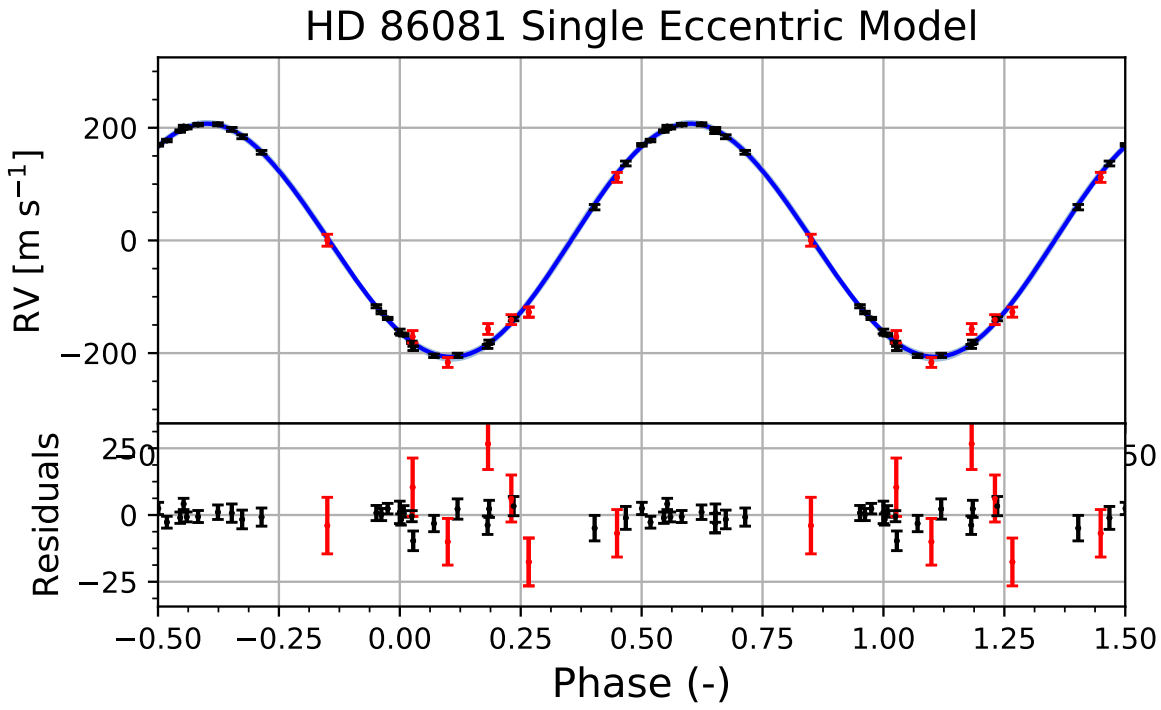

Figure 3.54: Phase plot for HD 86081. The offset data are in black, the new offset data from *APF* are in red, the single eccentric planet model in blue and the 68.2% error from the posterior distribution is in light blue.

Chapter 4

Modeling Stellar Noise

4.1 Stellar Activity

Section 3.3.1 mentions various stellar features and events that are a source of noise for RV observations. Haywood et al. [50] shows the observed RV signal can be thought of as the sum of the planet signal, the offset terms (the velocity offset and the linear trend), and the noise model:

$$RV_{\text{observed}} = \sum_{k=1}^{N_{\text{planets}}} RV_{\text{planet},k} + RV_{\text{o}} + RV_{\text{noise model}}. \tag{4.1.0.1}$$

Thus, the residuals of the planet fit, i.e. $RV_{\text{observed}} - \sum_{k=1}^{N_{\text{planets}}} RV_{\text{planet},k} - RV_{\text{o}}$, are a measurement of the stellar noise and the statistical/systematic errors associated with the observation.

4.2 Modeling Stellar Noise: Gaussian Processes

Many groups are starting to use Gaussian process (GP) regression [51] to model the stellar noise [50, 52, 53, 54]. GP regression modeling is a non-parametric way to model a set of data points using a covariance matrix kernel that describes the correlations between each pair of data points. The functional form of the kernel reflects the nature of the problem. For the case of stellar noise, the quasi-periodic kernel is a good choice. The kernel is described by hyperparameters that correspond to some physically meaningful quantity of the data [55].

4.2.1 Kernel Selection and New Likelihood Function

$$k(t_i, t_j) = h^2 \exp\left[-\frac{(t_i - t_j)^2}{2\lambda^2} - \frac{\sin^2(\pi(t_i - t_j)/P_{\text{rot}})}{2\omega^2}\right] + \sqrt{\sigma_{\text{jitter}}^2 + \sigma_{\text{error}}^2(t_i)}\,\delta_{ij}, \tag{4.2.1.1}$$

where h^2 is the amplitude of the noise, λ^2 is the timescale for growth and decay of active regions, ω controls the importance of the periodic and non-periodic terms, P_{rot} is the stellar rotation period, σ_{jitter} is the instrumental nose from the spectrograph, σ_{error} is is error in the RV data, and δ_{ij} is the Kronecker delta, indicating that the second term is only present in the diagonal of the covariance matrix [50, 53].

In practice, the hyperparameters are fit using S-index measurements and the same hyperparameters are used when doing the RV analysis. They use S-index measurements to constrain the noise model, and then use it to model the RV residuals [50, 52, 53, 54]. The S-index is an emission measurement defined in Vaughan, Preston, and Wilson [56] from the stellar counts of the H and K spectral lines (Ca II):

$$S\text{-index} = \alpha \frac{N_{\rm H} + N_{\rm K}}{N_{\rm R} + N_{\rm V}}, \tag{4.2.1.2}$$

where $N_{\rm H}$ and $N_{\rm K}$ are the number of counts in the H- and K-band. $N_{\rm R}$ and $N_{\rm V}$ are the number of counts the nearby continuum of the red and violet side of the H-K region. α is a normalization factor derived for each night of observation [56]. Some groups fix the hyperparameters to the values found from the S-index analysis, while others set Gaussian priors on the hyperparameters—with widths equal to the hyperparameter error—and fit the hyperparameters and the planet parameters simultaneously [50, 52, 53, 54].

The likelihood function takes on a different form when doing GP regression modeling [53, 51, 50]:

$$\ln \mathcal{P}(\mathrm{d}|\vartheta, \mathcal{M}) = -\frac{1}{2}\mathbf{r}^T\mathbf{K}^{-1}\mathbf{r} - \frac{1}{2}\ln\det\mathbf{K} - \frac{n}{2}\ln 2\pi, \tag{4.2.1.3}$$

where $\mathbf{K}$ is the covariance matrix constructed from Equation 4.2.1.1 described by hyperparameters a part of the vector ϑ, $\mathbf{r}$ is the residual vector— the difference between the data and the sum of the offset parameters and the planetary architecture model, and n is the number of data points.

4.3 New Synthetic Sample

We want to test the 2:1 degeneracy with realistic stellar noise in our synthetic systems. We generate a new sample of 2,000 synthetic systems: 1,000 single eccentrics and 1,000 circular doubles. Each system has GP noise included by drawing one realization from a quasi-periodic kernel and adding it to the RV signal. The pipeline will be tested to see if it can recover the correct architecture. Similar to Chapter 3, each system is fit with four models; single eccentric, circular double fixed to 2:1, circular double fixed to 2.17:1, and circular double with inner period free. The noise is modeled using a white-noise parameter—$\sigma_{\rm jitter}$—a common tactic of observers. This is the same noise model that is used in Chapter 3. This is done mainly because S-index measurements are often unavailable in the archive.

4.3.1 The Synthetic Recipe

The 2,000 synthetic RV time series are generated in a similar fashion as in Section 3.2.3. Their system and data properties are built to match the real systems. The Monte Carlo simulations were initialized as follows.

100 sky-positions were generated by uniformly drawing the cosine of the declination between 0–1 and uniformly drawing the right ascension between 0–24 hours. Then, each location was mock observed every hour, for 16 years, at Lick Observatory (37°20'29"N,

121°38'34"W) and times where the airmass between 1.3–2.2 using the *astroplan* and *astropy* codes [57]. The airmass is defined:

$$AM = \sec(90^{\rm o} - ALT), \tag{4.3.1.1}$$

where ALT is the altitude of the system—the angle between the horizon and the system—and AM is the airmass. The airmass is an approximate measure of the amount of atmosphere in a given line-of-sight. The direction of zenith—the location directly above the observer—at sea level is defined as an airmass of one. The first decade of the resulting time series are used to generate ten separate observational realizations. Each of the ten systems will have a different set of orbital parameters and GP hyperparameters according to the following paragraphs.

The start time (t_0) is a uniform random draw between the first 24 times. The observation time series is produced by selecting a set of observation differences ($t_i - t_{i-1}$) from the real distribution of observation differences with a similar, normally-distributed 10% variation added to each difference. These systems will have a timespan of ten years.

We determine the orbital period (P) using the selected number of orbits and the observation time series. The velocity semi-amplitude (K) and the eccentricity (e) are separate random draws from the real systems, with minimum eccentricity of 0.1. The mean anomaly of the start time (M_0) and argument of periastron (ω) are randomly drawn between 0 and 2π. The linear trend (A) is a 10% variation to a uniform random draw between -0.1 to 0.1. The offset (C) is a random draw between -1000 m s^{-1} and 1000 m s^{-1}. RV error bars ($\sigma_{\rm RV}$) are drawn randomly from the RV errors of the real systems and our error bars are assigned to this value. Stellar jitter is selected from a log uniform distribution between 0.5 and 5 m s^{-1}. h is a uniform random draw between 0.5 and 20. λ is a uniform random draw between 20 and 400. ω is a uniform random draw between 0.1 and 0.7. $P_{\rm rot}$ is a uniform random draw between 1 day and 150 days. A single GP model is drawn using a quasi-period kernel, with the generated hyperparameters, and is then added to each RV measurement.

The set of 1,000 circular double systems are generated using the same set of orbital parameters and hyperparameters as the 1,000 single eccentrics. They are converted into their first-order equivalent for the circular double case. This is done with the following:

$$\begin{aligned}
P_{\rm out} &= P \\
P_{\rm in} &= P_{\rm out}/(2+\epsilon) \\
K_{\rm out} &= K \\
K_{\rm in} &= K\ e \\
M_{\rm o,out} &= [\omega + M_{\rm o}] \bmod 2\pi \\
M_{\rm o,in} &= \left[\omega + 2M_{\rm o} - \tfrac{2\pi\,\epsilon\,(t-t_{\rm o})}{P}\right] \bmod 2\pi,
\end{aligned} \tag{4.3.1.2}$$

where P, K, e, $M_{\rm o}$ and ω are the orbital period, velocity semi-amplitude, eccentricity, mean anomaly of earliest observation, and argument of periastron passage for the single eccentric model, respectively. And $P_{\rm out}$, $K_{\rm out}$, $M_{\rm o,out}$ are the orbital period, velocity semi-amplitude, and mean anomaly for the outer planet in the circular double model equivalent to the single eccentric model. $P_{\rm in}$, $K_{\rm in}$, $M_{\rm o,in}$ is the same, but for the inner planet. ϵ is a parameter defined as the outer-to-inner period ratio minus two. For completeness, their inverse relationships

are:

$$\begin{aligned} P &= P_{\rm out} \\ K &= K_{\rm out} \\ e &= K_{\rm in}/K \\ M_{\rm o} &= [M_{\rm o,in} - M_{\rm o,out} + \tfrac{2\pi\,\epsilon\,(t-t_{\rm o})}{P}] \bmod 2\pi \\ \omega &= [M_{\rm o,out} - M_{\rm o}] \bmod 2\pi. \end{aligned} \tag{4.3.1.3}$$

Each circular double system is randomly separated into four even groups and assigned one of the following schemes for determining the outer-to-inner period ratio. The first scheme is a random uniform draw between 1.9 and 2.25. The remaining schemes are random normal draws centered on 1.95, 2.05, and 2.17. The width of the distribution for these schemes is 0.05. The hyperparameters for the circular doubles are the same

Finally, a single GP model is drawn using a quasi-period kernel, with the generated hyperparameters, and added to each RV measurement. Even though the same hyperparameters are used for the single eccentrics and circular doubles, there will be a different noise model added to each synthetic system. This is because the kernel defined by the hyperparameters is a family of infinite functions, so the chances of drawing the same exact model is very small. We make use of the *george* code to add the GP [55]. The pipeline is largely set-up the same exact way as in Section 3.2.2. The prior bounds for $\sigma_{\rm jitter}$ are between 0.5 m s^{-1} and 5 m s^{-1}.

4.4 Preliminary Results

Figure 4.1 shows the resulting Bayes factors for the 2,000 synthetic systems when fit with the white-noise model. The Bayes factors plotted correspond to the best fixed circular double model given the single eccentric model. The Bayes factor distribution for the single eccentrics is similar to Figure 3.3. Figure 4.2 shows the Bayes factors corresponding to the best circular double model given the single eccentric model. The pipeline was able to recover most of the synthetic parameters, however, there is still some architecture misidentification. The pipeline correctly identifies more circular double than single eccentrics. This result is seen by comparing the location of the 98$^{\rm th}$ percentiles. These differences are expected because the degeneracy is between single eccentrics and circular doubles occurs at 2:1. Since our synthetic circular double sample contains many systems that are not at the 2:1, it is understandable that the pipeline would do a better job at picking those kind of architectures out. The problem lies with breaking the degeneracy when the system is either close to the 2:1 or is a single eccentric. Here, the pipeline shows that the degeneracy is present in our synthetic sample. For the synthetic single eccentrics, 25% of the systems in the 98th percentile in the fixed circular doubles versus single eccentric case favored the 2:1 model. In the case of fixed and free circular doubles versus single eccentric, almost half of the 98th percentile had period ratios smaller than 1 and/or had $K_{\rm out}/K_{\rm out}$-ratios larger than a few hundred. The former suggests that the inner planet is orbiting much farther from the host star than the outer planet. The latter suggests that the mass of the inner planet is very small. Both of these are ways to obtain false positives because they are in places in the parameter-space where their contribution to the overall signal is negligible compared to the massive outer planet and, more importantly, the stellar signal.

In the next chapter we will investigate ways to alter the shape of Bayes factor distribution to decrease the ambiguity near the 2:1 degeneracy by including the remaining six-years of the generated time series.

(This page is intentionally left blank so that Figure 4.1 is on the back-side of the page.)

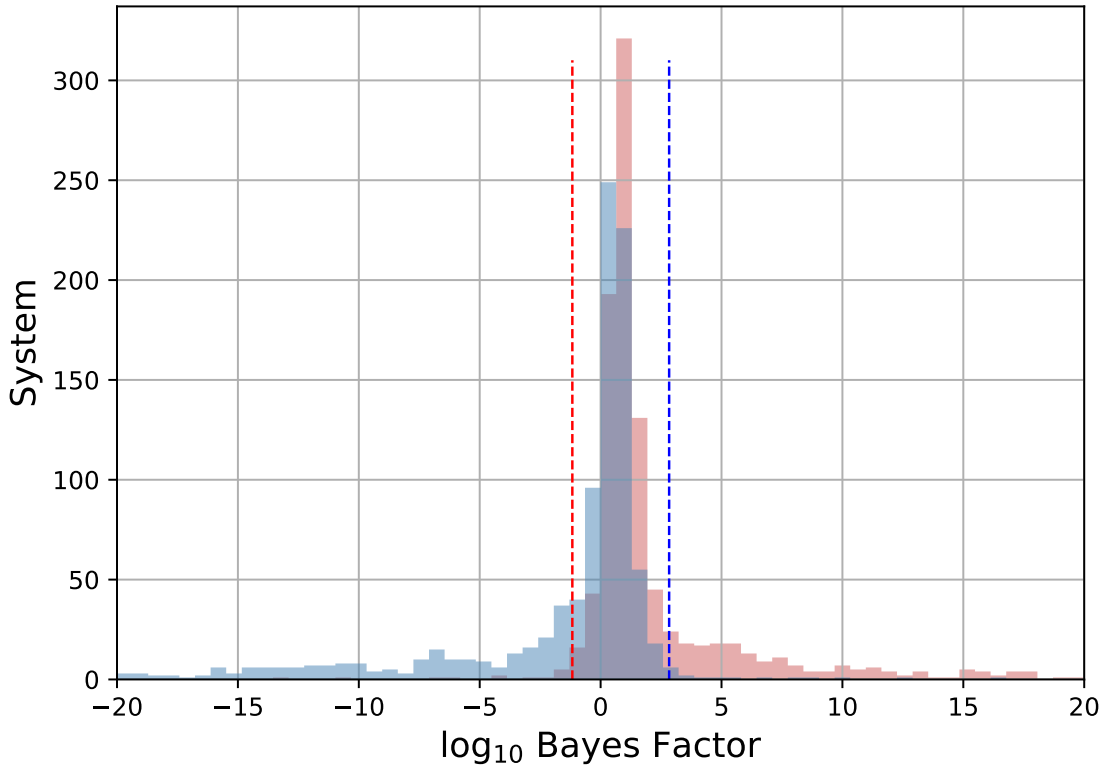

Figure 4.1: The log Bayes factor distribution for the 1,000 synthetic single eccentric time series in blue and the 1,000 synthetic circular double time series in red. Here, we compare only the single eccentric model to the fixed circular double model with the largest Bayes factor. The 98th percentiles are indicated with the dotted lines for each set of systems: log(Bayes factors) of 2.83 and -1.17 for the single eccentrics and circular doubles.

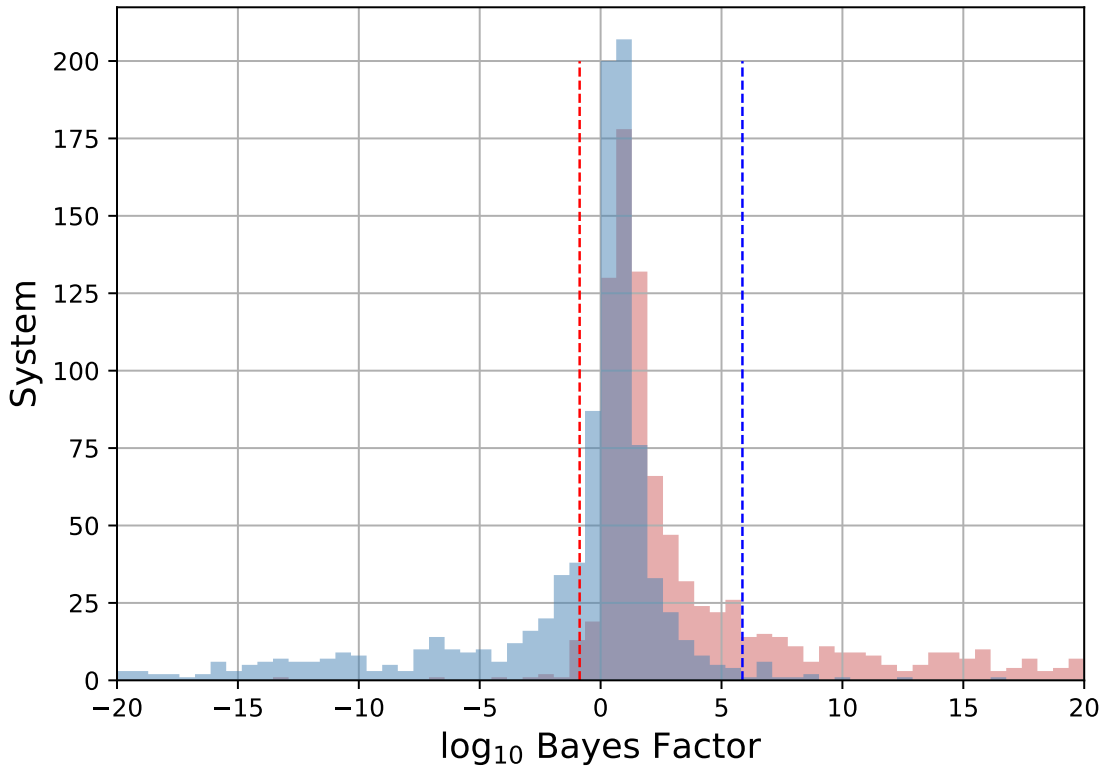

Figure 4.2: The log Bayes factor distribution for the 1,000 synthetic single eccentric time series in blue and the 1,000 synthetic circular double time series in red. Here, we compare the single eccentric model to the circular double model with the largest Bayes factor. The 98th percentiles are indicated with the dotted lines for each set of systems. log(Bayes factors) of 5.87 and -0.87 for the single eccentrics and circular doubles.

Chapter 5

Observation Strategy: Optimum Phases or More Data?

The best way to confirm that a system is a circular double is to obtain additional data, repeat the model comparison analysis, and see the effect on the newly computed Bayes factor. However, it may not be as fruitful to simply obtain data whenever the system is able to be observed. We want to alter the confidence as much as possible, as efficiently possible. The system needs to be observed near key phases, where the 2:1 degeneracy is at its weakest, to do this.

5.1 Breaking the Degeneracy: Optimum Phases

The 2:1 degeneracy can be broken two ways. First, when expanded to the second-order in e, the RV signal of a single eccentric has a third harmonic term ($3M$). We adopt the convention that the second harmonic is at twice the fundamental frequency, and so forth. Following Murray and Dermott [40], the second-order expansion of the RV signal of a single eccentric is:

$$RV_{\rm single} \approx K \left[1 - \frac{6e^2}{8}\right] \cos(\omega + M) + K\, e \cos(\omega + 2M) + K\, \frac{9e^2}{8} \cos(\omega + 3M) - K\, \frac{e^2}{8} \cos(\omega - M) + \mathcal{O}(e^3). \quad (5.1.0.1)$$

If one can measure the third harmonic term, then the degeneracy is broken since no equivalent term exists for the two-planet case. Therefore, a successful measurement of this term is an indication that the system is a single eccentric. However, the amplitude of the third harmonic signal is small compared to the amplitude of the second harmonic. So the system needs to be observed when the third harmonic signal is strong whilst the second harmonic signal is weak. There are key phases (henceforth, optimum phases) in the orbit where this occurs. They are found by solving for the difference between the second and third harmonics, finding all the local minima, and numerically solving the resulting transcendental equation:

$$2\, e \sin(\omega + 2\, M) = \frac{27\, e^2}{8} \sin(\omega + 3\, M). \quad (5.1.0.2)$$

Equation 5.1.0.2 only depends on e and ω. The mean anomalies (M) that satisfy this equation correspond to the optimum phases that the system should be observed.

Figures 5.1 and 5.2 show where the optimum phases are in relation to the exact single eccentric model and the equivalent circular double model. The locations of the optimum phases are not where the difference between the single eccentric and the 2:1 circular double signals differ the most.

The degeneracy can also be broken if the system is slightly out of resonance, $P_{\rm in} \neq 2P_{\rm out}$, because the single planet case can only have a second harmonic signal at exactly 2:1. If the second harmonic term is not seen in the data, then there may be enough period resolution to show that the system is not exactly at the 2:1, i.e. $\rm P_{out} \neq 2\,P_{in}$. This is enough break the degeneracy between the models because the single planet case only has a harmonic term at exactly 2:1.

5.1.1 Optimum Phase Parameter-Space

We explore the $\{e,\, \omega\}$-parameter space of Equation 5.1.0.2 and found that for any set of e and ω there will be either two or three solutions to the equation. Figures 5.3 – 5.6 show the possible solutions with the solution number for a pair of (e, ω) as different colors. The third solution is possible when the $e \gtrsim 0.4$, as seen in Figure 5.4. The solutions repeat every cycle, as expected. The repetition is seen when extending the phase to -0.2 and 1.2 in Figure 5.6.

This result can be used to inform observers of the optimum time to observe a system in order to diminish the ambiguity of RV models near the 2:1 degeneracy. Data gathered near these key phases should have more weight in determining the correct architecture then at any other phase. We test this strategy in the next section.

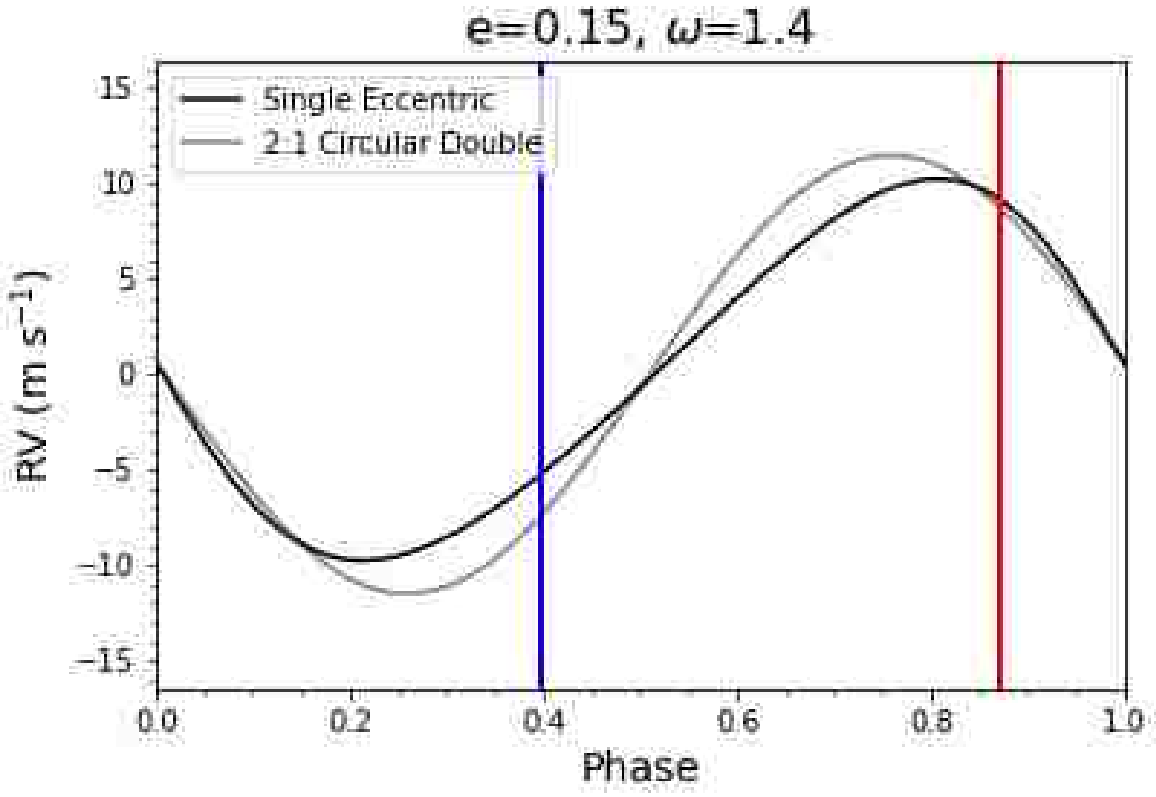

Figure 5.1: Single eccentric model with $K = 10$ m s^{-1}, $e = 0.15$, $\omega = 1.4$ rad in black and its equivalent 2:1 circular double model in grey. The locations of where the third harmonic signal of the single eccentric model is stronger than the second harmonic signal are the vertical lines. There are two solutions for this set of (e, ω).

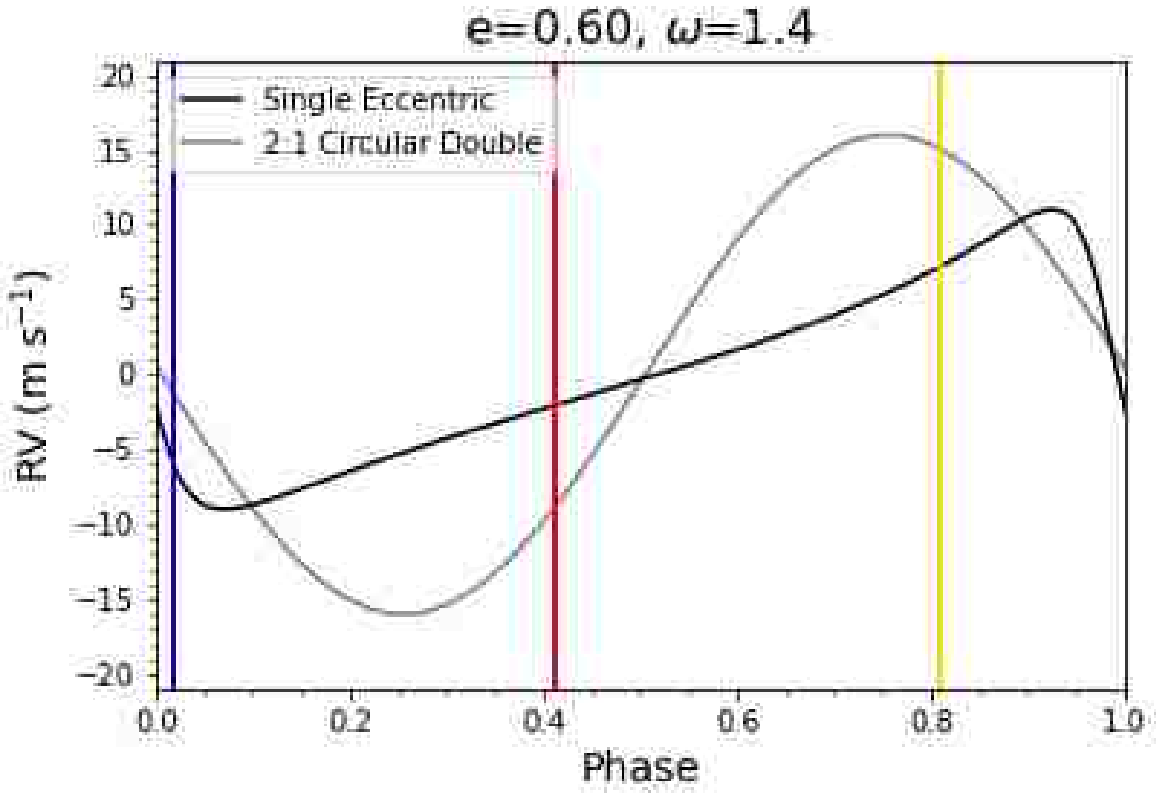

Figure 5.2: Single eccentric model with $K = 10$ m s^{-1}, $e = 0.6$, $\omega = 1.4$ rad in black and its equivalent 2:1 circular double model in grey. The locations of where the third harmonic signal of the single eccentric model is stronger than the second harmonic signal are the vertical lines. There are three solutions for this set of (e, ω).

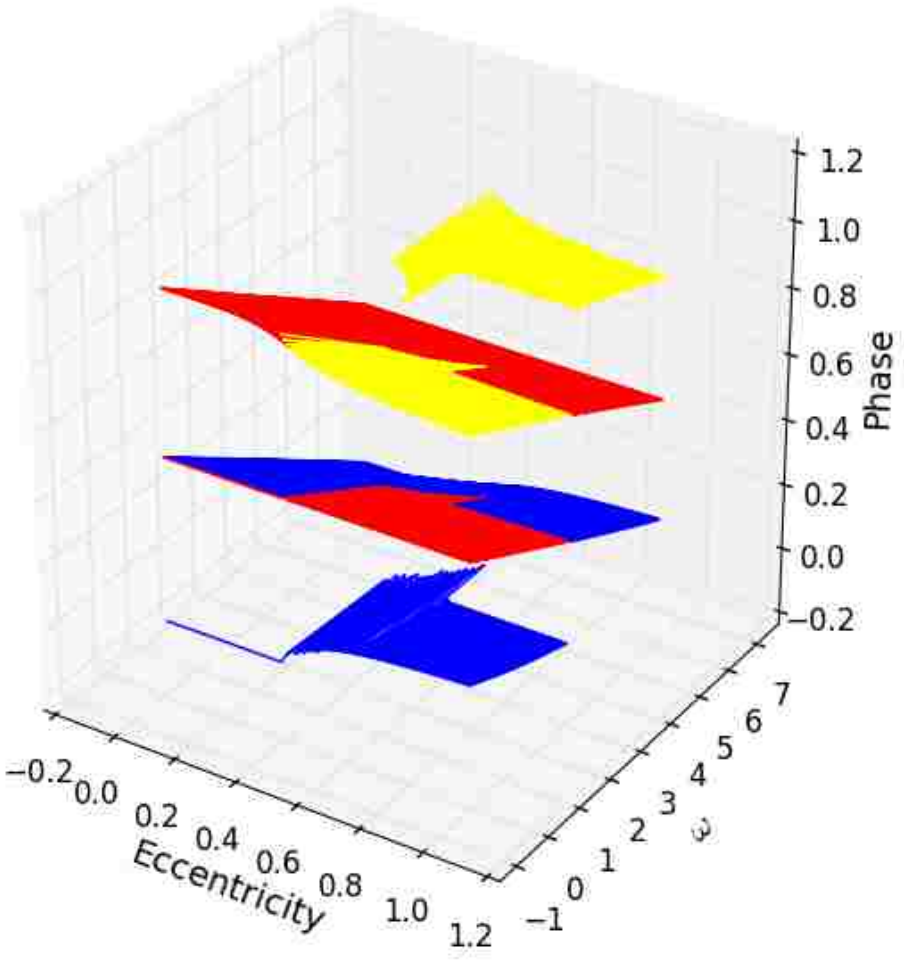

Figure 5.3: The solutions to Equation 5.1.0.2. The color indicates the solution number for the single set of $\{e, \omega\}$. Yellow is three solutions, red is two solutions, and blue is one solution. For example, $\{e{=}1, \omega{=}2\pi\}$ has three solutions, as indicated by the three colored surfaces on the plot.

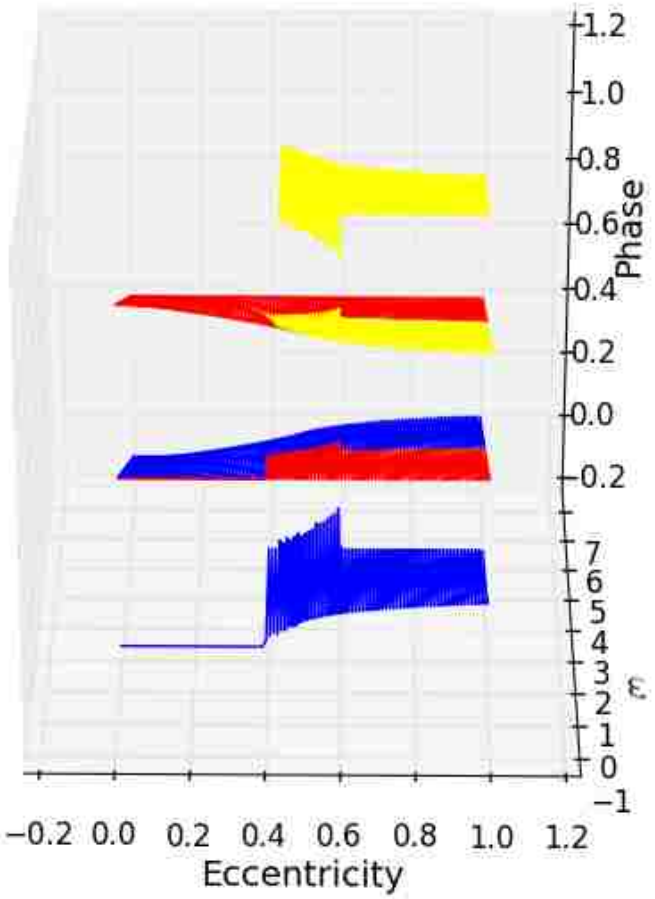

Figure 5.4: Same as Figure 5.3, but showing that the third solution emerges around $e \gtrsim 0.4$.

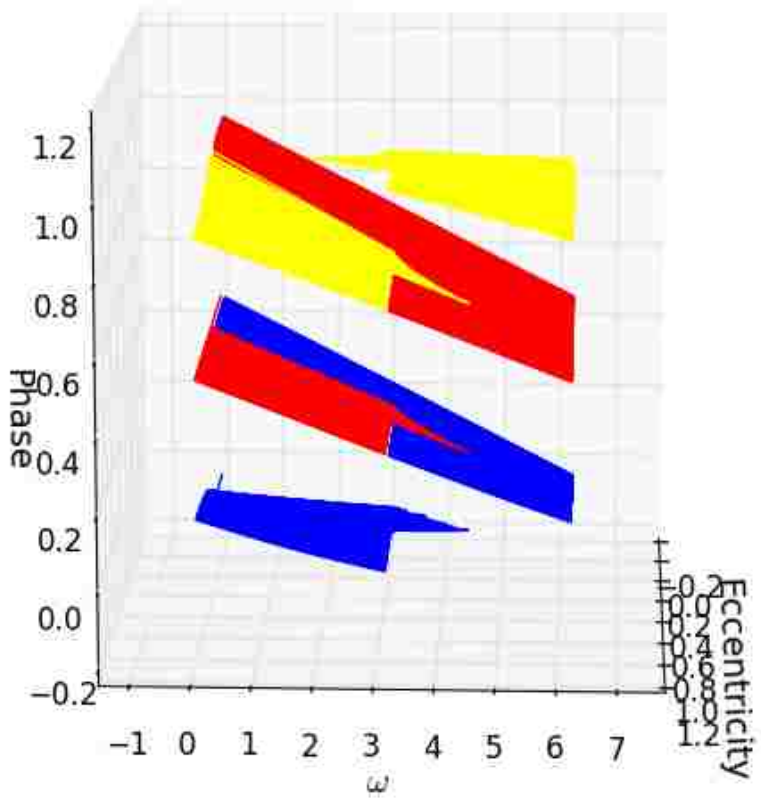

Figure 5.5: Same as Figure 5.3, but showing a different projection.

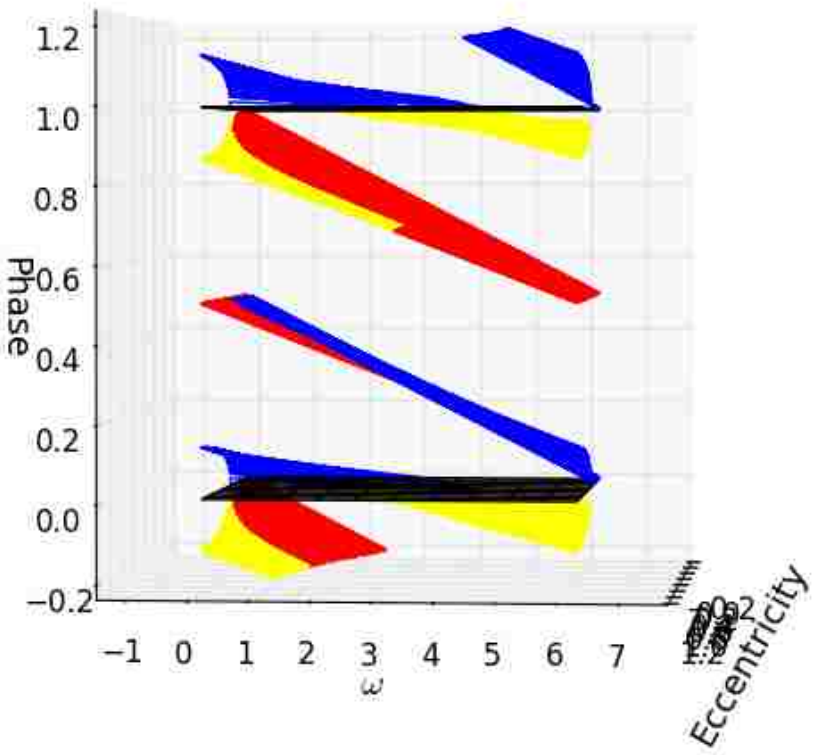

Figure 5.6: Same as Figure 5.5, but showing solutions extending past phases 0 and 1, which are indicated by the black plane.

5.2 Set-up

We use the remaining 6 years of the 16 year timeseries generated in Section 4.3.1 to test two observational strategies. The 6 year data is combined with the 10 year data to form a 16 year timeseries. These timeseries are analyzed with our the pipeline and the same models are fit as defined in Chapter 4.

The first strategy focuses on gathering data the same way as the first 10 years, i.e. without considering the system phase. The second strategy gathers 50% of the data from near the optimum phase—defined as ± 0.05 phase—and the remaining 50% of the data at random phases. The latter criterion simulates the observer not passing on an opportunity to observe the system if the sky conditions are right, but still focuses on gathering data when the system is near the optimum phase.

We test three schemes for the number of additional data in each 6 year observations. The first scheme is simply doubling the number of observations. The second is brings the number of observations in each system equal to the largest number of observations out of the initial 1,000 systems, 108 observations. The third scheme brings the number of observations in each system equal to twice the largest number of observations out of the initial 1,000 systems, 216 observations.

5.3 Results

The Bayes factors distributions for each scheme, observational strategy, and synthetic system type are shown in Figures 5.7–5.18. The Bayes factor of the 98th percentile is indicated in every plot. It is the location of where 98% of the sample is when integrated from the left/right for the single eccentric/circular double synthetics respectively. It is a measure of how well the sample is fit as a whole. Thus, comparing the locations of the 98th percentile for the two observational strategies is a way to determine which one is more effective.

In every case, the strategy of focusing on the optimum phase has better Bayes factors that favor the true synthetic architecture than compared to the strategy of adding data the random phases. In general, the ambiguity between the single eccentric and circular double model decreased. This is seen by the distribution near Bayes factor of unity decreasing throughout all strategies and schemes. The distribution decreases more for the optimum phase strategy than the random phase strategy in every synthetic circular double distribution and a third of the synthetic single eccentric distributions. The 98th percentile for the synthetic circular doubles are much closer to each other because these systems are not built exactly at the 2:1 degeneracy, so it is more likely that the circular double case would fit better.

These distributions suffer from the same false positive issues as in Section 4.4. In some cases, running the code with more data exacerbates the rate of these fictitious false positives. I call these fictitious because they are the result of the fitting procedure finding an unrealistic place of the parameter space that happens to have good likelihood value. Instead of a true false positive, that is a physically meaningful fit that is incorrect.

(This page is intentionally left blank so that Figure 5.7 is on the back-side of the page.)

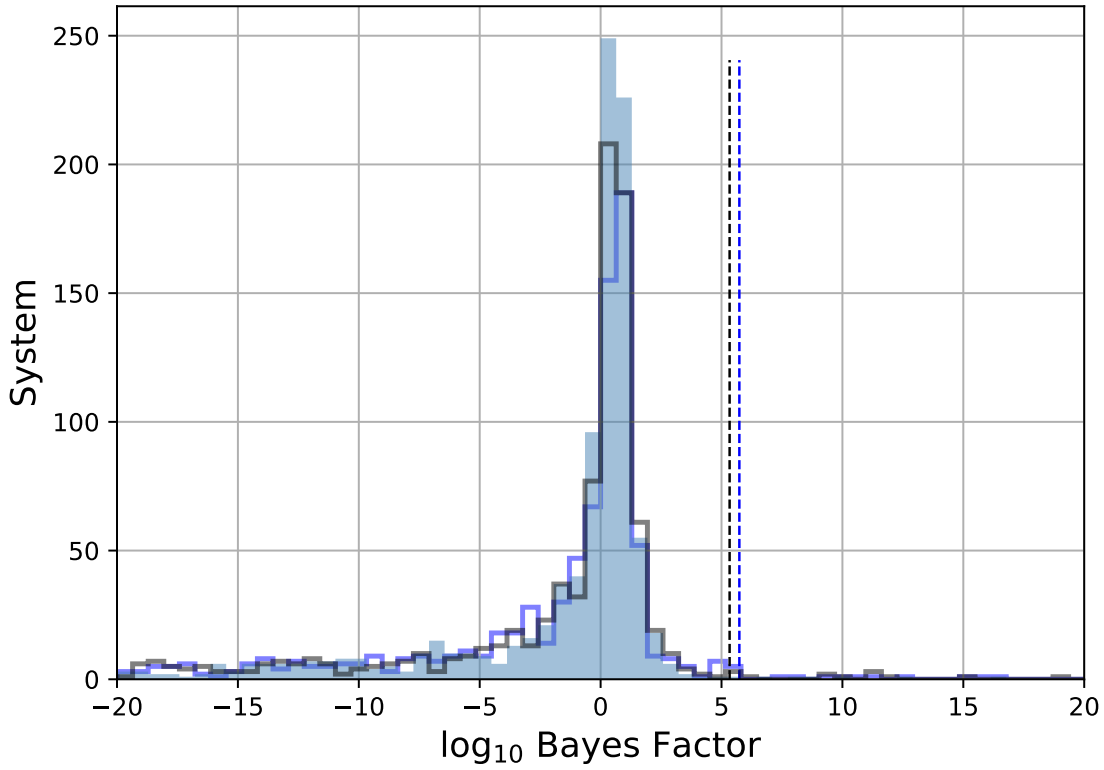

Figure 5.7: The log Bayes factor distributions for the scheme that doubles the number of data for the 1,000 synthetic single eccentric time series. The initial ten-year distribution is in light blue. The sixteen-year distributions with the last six years obtaining data at random phases is the blue wire. The sixteen-year distributions with the last six years obtaining data at 50% near the optimum phase and 50% random phases is the black wire. Here, we compare the single eccentric model to the fixed circular double model with the largest Bayes factor. The 98th percentiles are indicated with the dotted lines for each set of systems. The log(Bayes factors) for the six-years of data at random phases and six-years of data with 50% near the optimum phase and 50% random phases are 5.74 and 5.33, respectively.

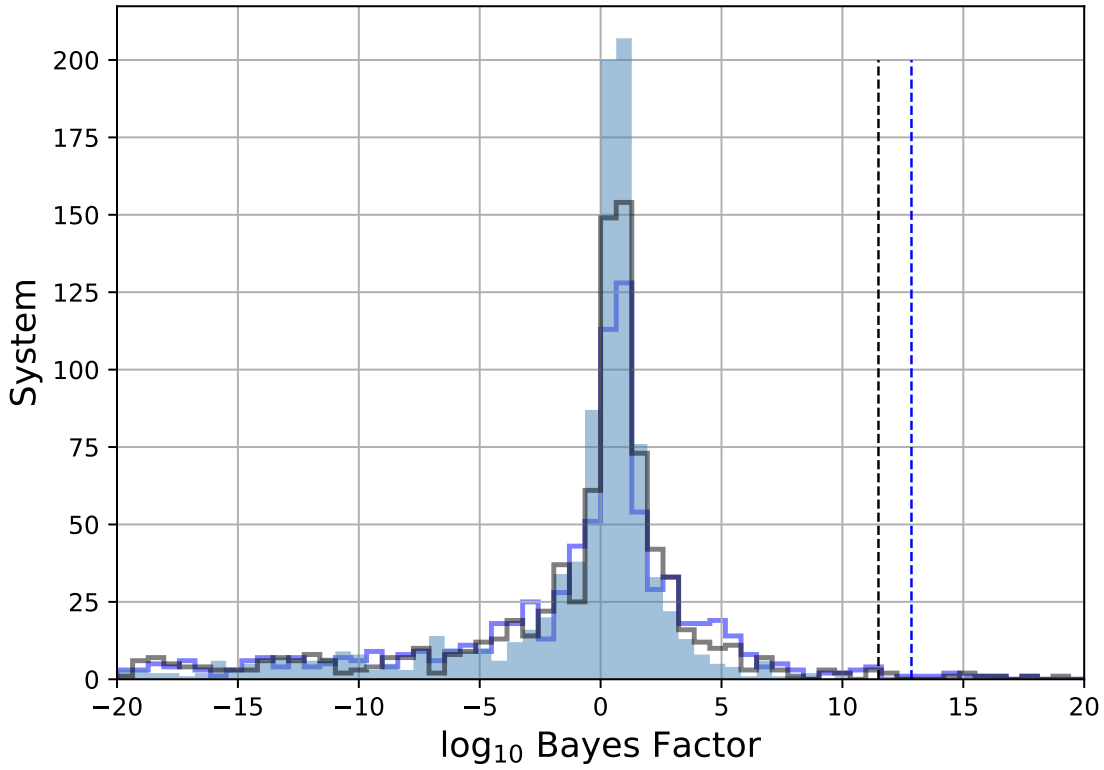

Figure 5.8: The log Bayes factor distributions for the scheme that doubles the number of data for the 1,000 synthetic single eccentric time series. The initial ten-year distribution is in light blue. The sixteen-year distributions with the last six years obtaining data at random phases is the blue wire. The sixteen-year distributions with the last six years obtaining data at 50% near the optimum phase and 50% random phases is the black wire. Here, we compare the single eccentric model to the circular double model with the largest Bayes factor. The 98th percentiles are indicated with the dotted lines for each set of systems. The log(Bayes factors) for the six-years of data at random phases and six-years of data with 50% near the optimum phase and 50% random phases are 12.86 and 11.49, respectively.

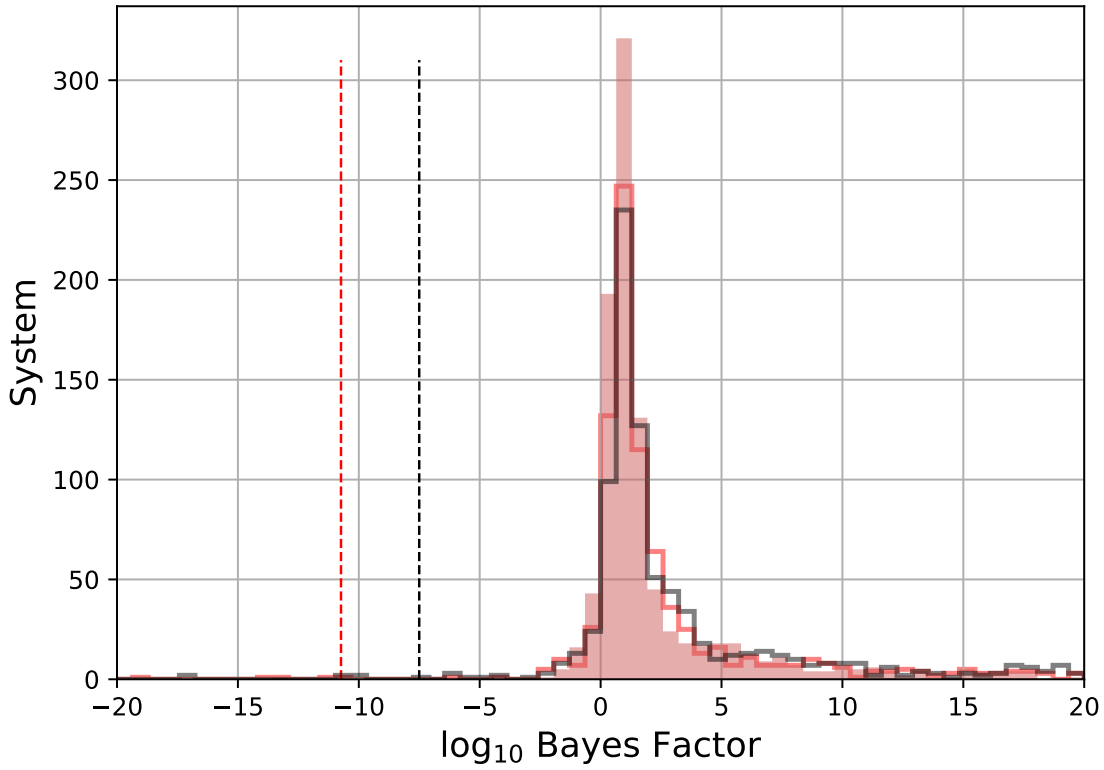

Figure 5.9: The log Bayes factor distributions for the scheme that doubles the number of data for the 1,000 synthetic circular double time series. The initial ten-year distribution is in light red. The sixteen-year distributions with the last six years obtaining data at random phases is the red wire. The sixteen-year distributions with the last six years obtaining data at 50% near the optimum phase and 50% random phases is the black wire. Here, we compare the single eccentric model to the fixed circular double model with the largest Bayes factor. The 98th percentiles are indicated with the dotted lines for each set of systems. The log(Bayes factors) for the six-years of data at random phases and six-years of data with 50% near the optimum phase and 50% random phases are -3.21 and -2.47, respectively.

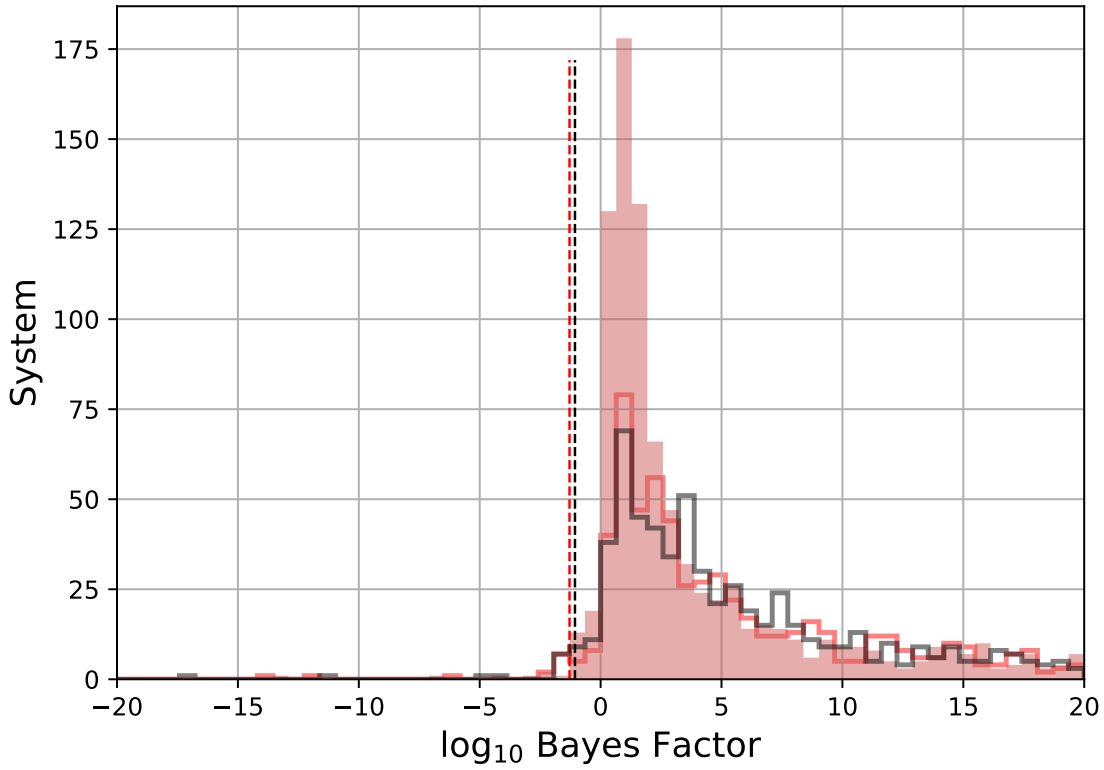

Figure 5.10: The log Bayes factor distributions for the scheme that doubles the number of data for the 1,000 synthetic circular double time series. The initial ten-year distribution is in light red. The sixteen-year distributions with the last six years obtaining data at random phases is the red wire. The sixteen-year distributions with the last six years obtaining data at 50% near the optimum phase and 50% random phases is the black wire. Here, we compare the single eccentric model to the circular double model with the largest Bayes factor. The 98th percentiles are indicated with the dotted lines for each set of systems. The log(Bayes factors) for the six-years of data at random phases and six-years of data with 50% near the optimum phase and 50% random phases are -0.67 and -0.59, respectively.

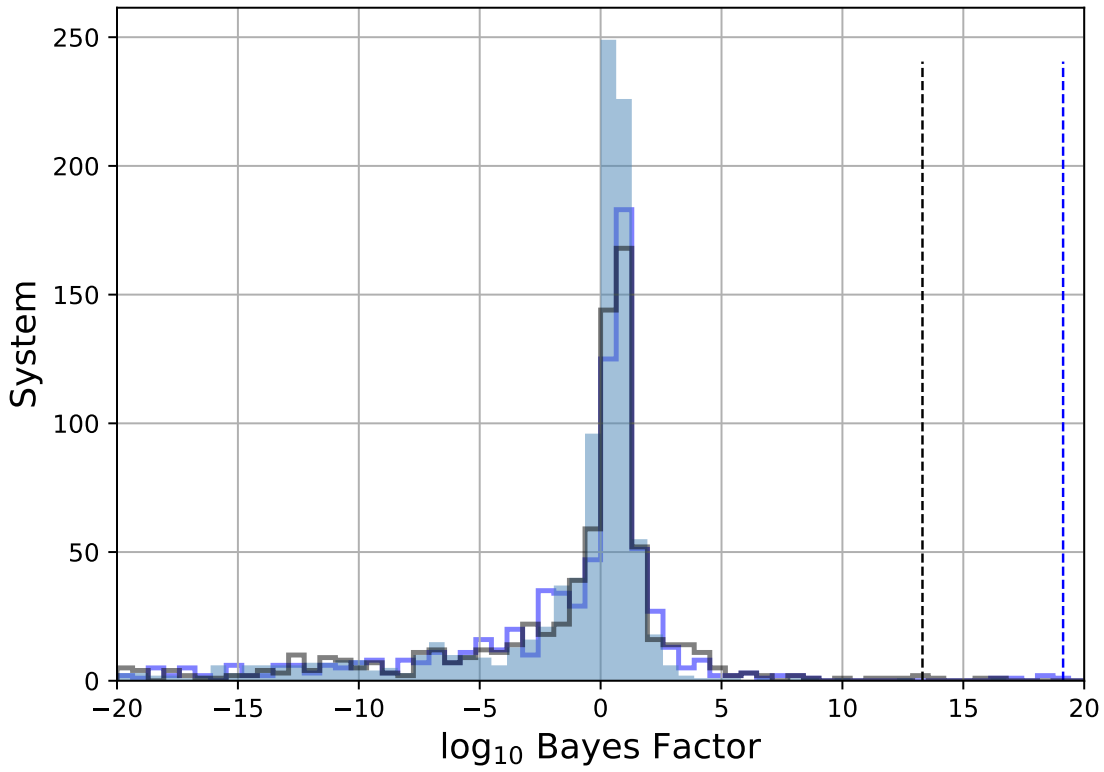

Figure 5.11: The log Bayes factor distributions for the scheme that brings the number of data for the 1,000 synthetic single eccentric time series up to twice the maximum number of data in the initial ten-year set, i.e. every system has 216 datapoints. The initial ten-year distribution is in light blue. The sixteen-year distributions with the last six years obtaining data at random phases is the blue wire. The sixteen-year distributions with the last six years obtaining data at 50% near the optimum phase and 50% random phases is the black wire. Here, we compare the single eccentric model to the fixed circular double model with the largest Bayes factor. The 98th percentiles are indicated with the dotted lines for each set of systems. The log(Bayes factors) for the six-years of data at random phases and six-years of data with 50% near the optimum phase and 50% random phases are 19.13 and 13.31, respectively.

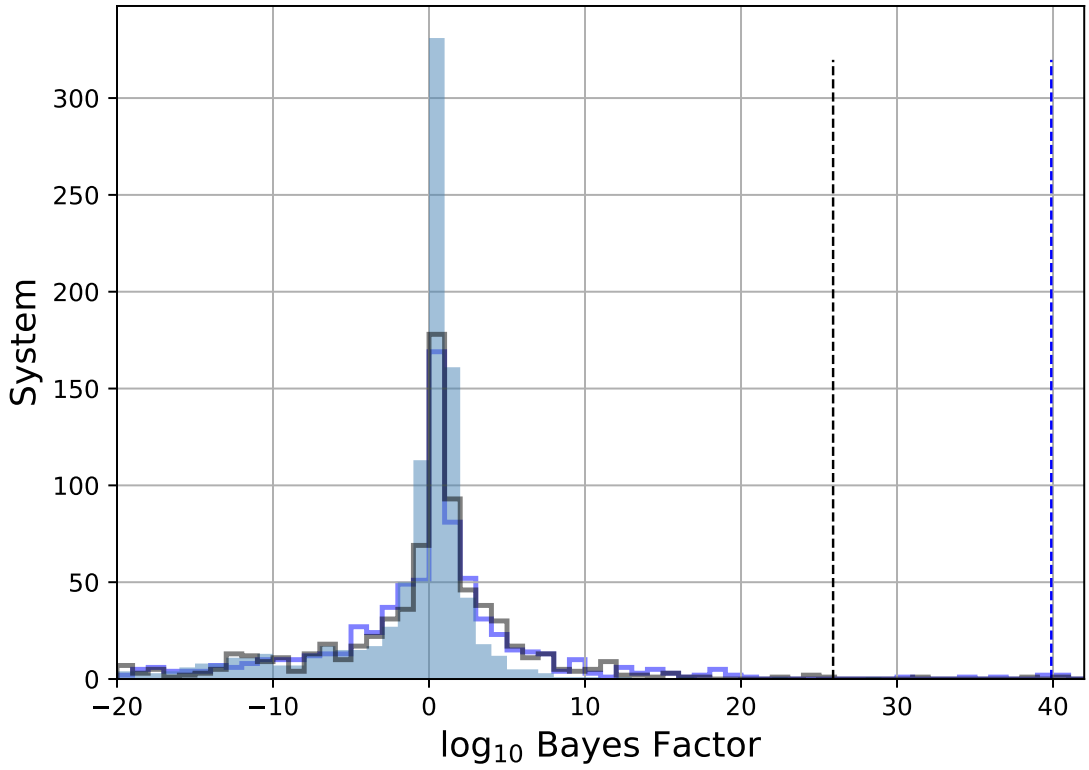

Figure 5.12: The log Bayes factor distributions for the scheme that brings the number of data for the 1,000 synthetic single eccentric time series up to twice the maximum number of data in the initial ten-year set, i.e. every system has 216 datapoints. The initial ten-year distribution is in light blue. The sixteen-year distributions with the last six years obtaining data at random phases is the blue wire. The sixteen-year distributions with the last six years obtaining data at 50% near the optimum phase and 50% random phases is the black wire. Here, we compare the single eccentric model to the circular double model with the largest Bayes factor. The 98th percentiles are indicated with the dotted lines for each set of systems. The log(Bayes factors) for the six-years of data at random phases and six-years of data with 50% near the optimum phase and 50% random phases are 39.89 and 25.90, respectively.

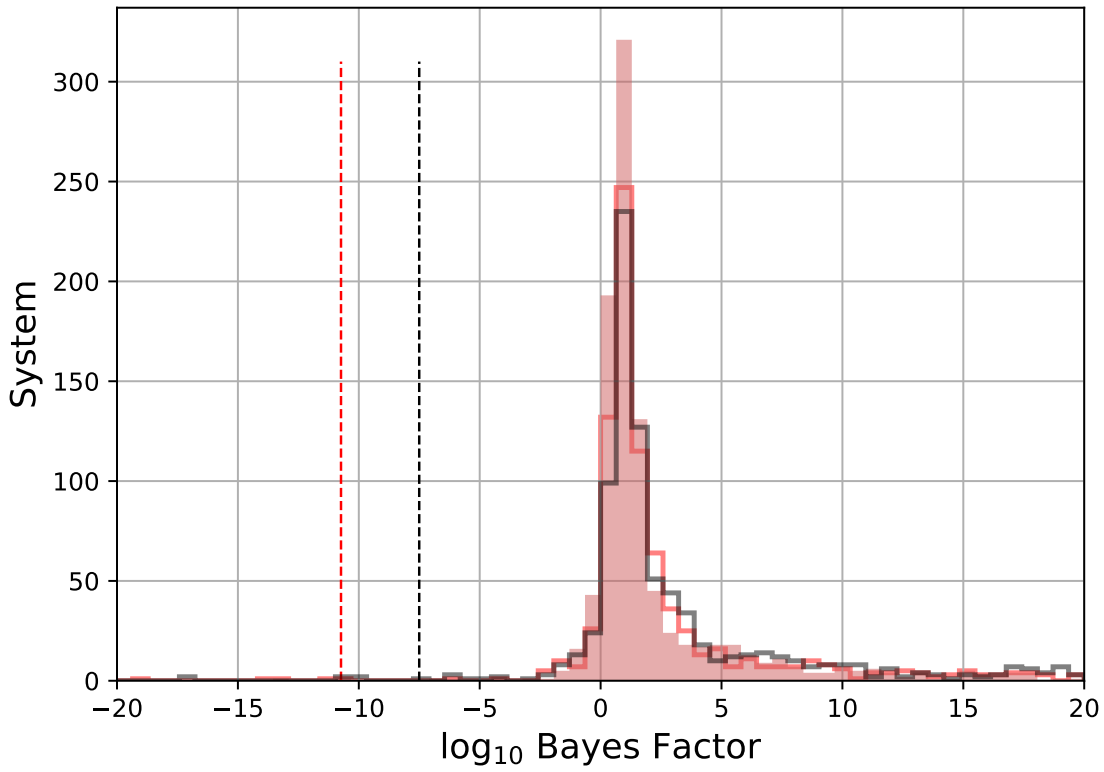

Figure 5.13: The log Bayes factor distributions for the scheme that brings the number of data for the 1,000 synthetic circular double time series up to twice the maximum number of data in the initial ten-year set, i.e. every system has 216 datapoints. The initial ten-year distribution is in light red. The sixteen-year distributions with the last six years obtaining data at random phases is the red wire. The sixteen-year distributions with the last six years obtaining data at 50% near the optimum phase and 50% random phases is the black wire. Here, we compare the single eccentric model to the fixed circular double model with the largest Bayes factor. The 98th percentiles are indicated with the dotted lines for each set of systems. The log(Bayes factors) for the six-years of data at random phases and six-years of data with 50% near the optimum phase and 50% random phases are -10.74 and -7.50, respectively.

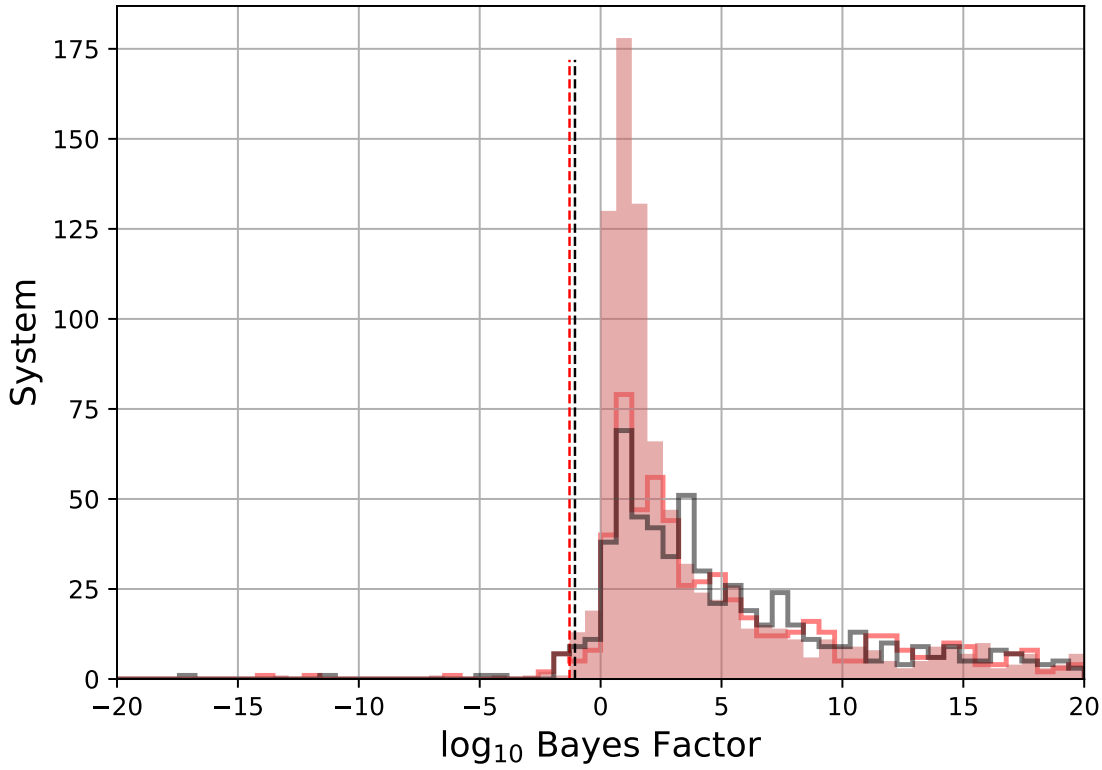

Figure 5.14: The log Bayes factor distributions for the scheme that brings the number of data for the 1,000 synthetic circular double time series up to twice the maximum number of data in the initial ten-year set, i.e. every system has 216 datapoints. The initial ten-year distribution is in light red. The sixteen-year distributions with the last six years obtaining data at random phases is the red wire. The sixteen-year distributions with the last six years obtaining data at 50% near the optimum phase and 50% random phases is the black wire. Here, we compare the single eccentric model to the circular double model with the largest Bayes factor. The 98th percentiles are indicated with the dotted lines for each set of systems. The log(Bayes factors) for the six-years of data at random phases and six-years of data with 50% near the optimum phase and 50% random phases are -1.29 and -1.06, respectively.

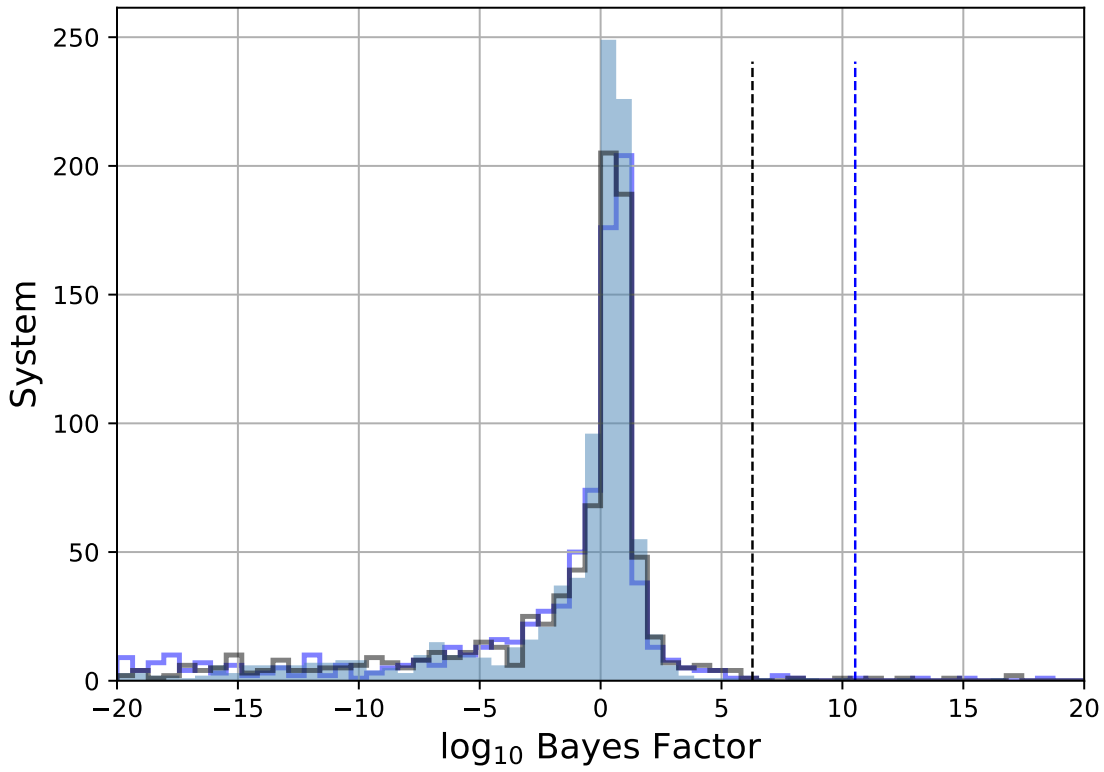

Figure 5.15: The log Bayes factor distributions for the scheme that brings the number of data for the 1,000 synthetic single eccentric time series up to match the maximum number of data in the initial ten-year set, i.e. every system has 108 datapoints. The initial ten-year distribution is in light blue. The sixteen-year distributions with the last six years obtaining data at random phases is the blue wire. The sixteen-year distributions with the last six years obtaining data at 50% near the optimum phase and 50% random phases is the black wire. Here, we compare the single eccentric model to the fixed circular double model with the largest Bayes factor. The 98th percentiles are indicated with the dotted lines for each set of systems. The log(Bayes factors) for the six-years of data at random phases and six-years of data with 50% near the optimum phase and 50% random phases are 10.53 and 6.28, respectively.

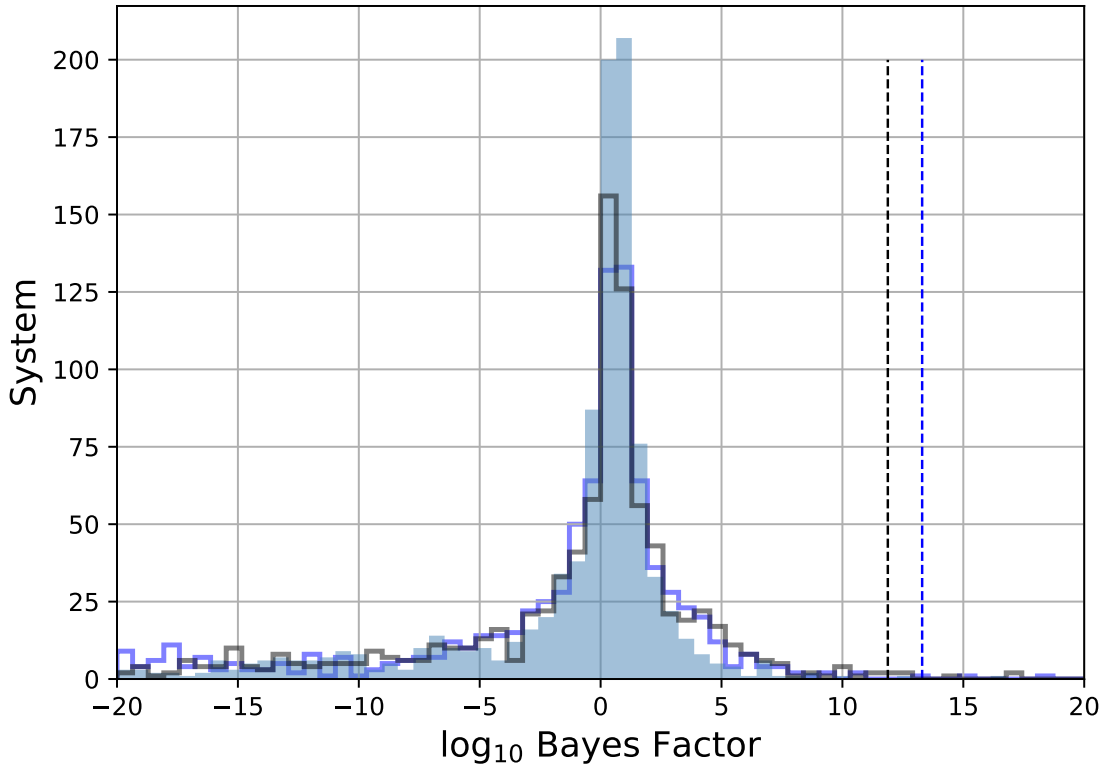

Figure 5.16: The log Bayes factor distributions for the scheme that brings the number of data for the 1,000 synthetic single eccentric time series up to match the maximum number of data in the initial ten-year set, i.e. every system has 108 datapoints. The initial ten-year distribution is in light blue. The sixteen-year distributions with the last six years obtaining data at random phases is the blue wire. The sixteen-year distributions with the last six years obtaining data at 50% near the optimum phase and 50% random phases is the black wire. Here, we compare the single eccentric model to the circular double model with the largest Bayes factor. The 98th percentiles are indicated with the dotted lines for each set of systems. The log(Bayes factors) for the six-years of data at random phases and six-years of data with 50% near the optimum phase and 50% random phases are 13.30 and 11.88, respectively.

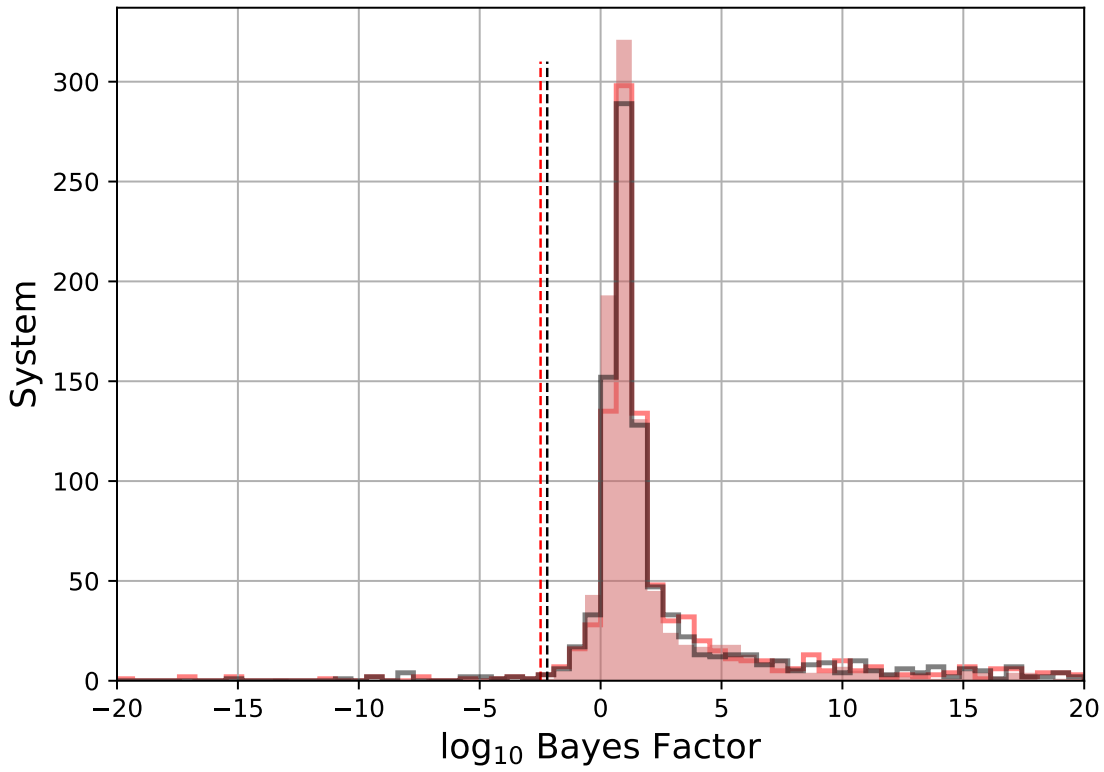

Figure 5.17: The log Bayes factor distributions for the scheme that brings the number of data for the 1,000 synthetic circular double time series up to match the maximum number of data in the initial ten-year set, i.e. every system has 108 datapoints. The initial ten-year distribution is in light red. The sixteen-year distributions with the last six years obtaining data at random phases is the red wire. The sixteen-year distributions with the last six years obtaining data at 50% near the optimum phase and 50% random phases is the black wire. Here, we compare the single eccentric model to the fixed circular double model with the largest Bayes factor. The 98th percentiles are indicated with the dotted lines for each set of systems. The log(Bayes factors) for the six-years of data at random phases and six-years of data with 50% near the optimum phase and 50% random phases are -2.48 and -2.21, respectively.

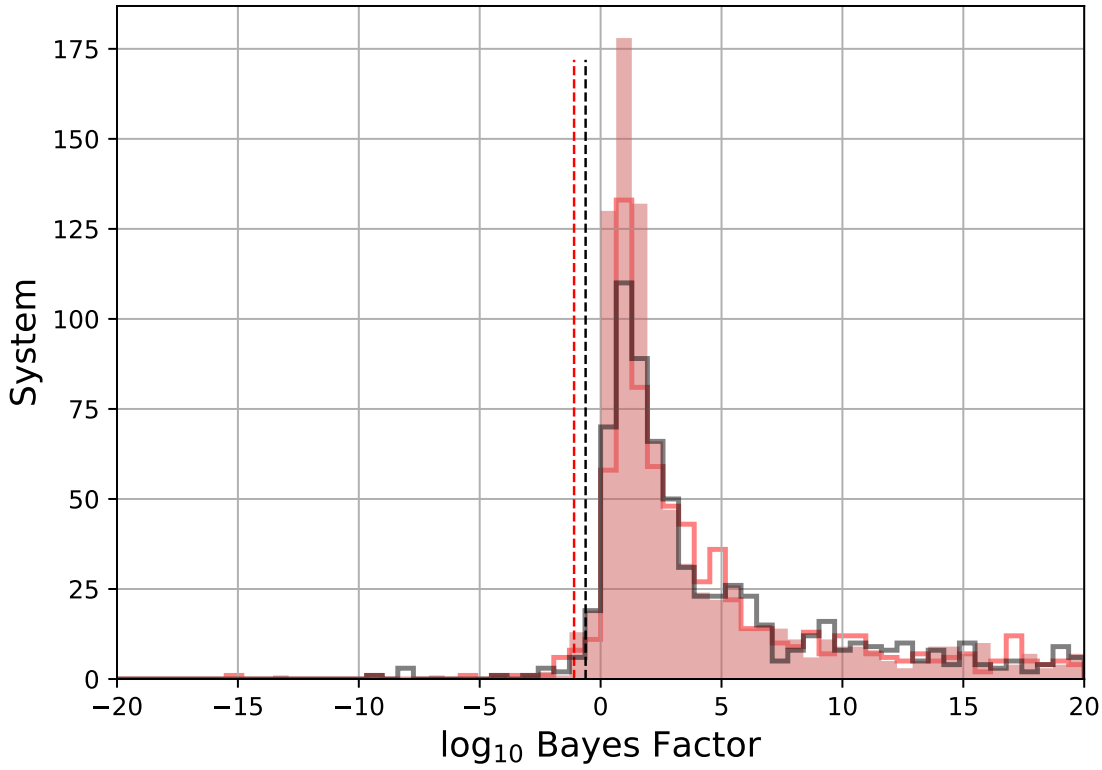

Figure 5.18: The log Bayes factor distributions for the scheme that brings the number of data for the 1,000 synthetic circular double time series up to match the maximum number of data in the initial ten-year set, i.e. every system has 108 datapoints. The initial ten-year distribution is in light red. The sixteen-year distributions with the last six years obtaining data at random phases is the red wire. The sixteen-year distributions with the last six years obtaining data at 50% near the optimum phase and 50% random phases is the black wire. Here, we compare the single eccentric model to the circular double model with the largest Bayes factor. The 98th percentiles are indicated with the dotted lines for each set of systems. The log(Bayes factors) for the six-years of data at random phases and six-years of data with 50% near the optimum phase and 50% random phases are -1.10 and -0.62, respectively.

5.4 Conclusions

The answer to question in the title of this chapter is clear: gathering data when the system is near the optimum phases is more efficient than blindly gathering data when the system is visible. This result should help observers plan better RV observations for a sample of stars. An efficient observational campaign can be geared to maximize the confidence of the resulting observations by giving preference to the systems that are near the optimum phases. Coordinated observations at specific times should reduce the operation time (and therefore costs) required to determine the correct architecture of planetary systems near the 2:1 degeneracy. This conclusion is seen in our Bayes factor distributions comparing the optimum phase strategy with the random phase strategy.

There are false positives in our synthetic samples. Finding false positives in generated datasets are a natural part of pipelines. As mentioned earlier, there could be ways where an inner planet can have very little contribution to the signal, but yet be considered a viable fit. For example, the ratio of the K-amplitudes between the inner and outer planet is near zero or if the orbital period of the "inner" period is driven to periods hundreds of times larger than the "outer" planet. In this scenario, the "inner" planet is actually orbiting farther from the host star than the "outer" planet. In other cases, every model fit was bad and so the comparison of those bad FMLs yielded a large Bayes factor. These issues will be addressed before the next step in this long-term project. Reader, please find the publication where the entirety of the project is presented.

Here, we fit the architecture models with a white noise model, which is a standard way that observers fit RV data. What happens if the pipeline fit a realistic noise model, as a Gaussian process? And, what happens if the pipeline fit without a noise model? Can the pipeline still find the correct architecture through the stellar noise? These questions will be answered in the next publication from Dr. Steffen and myself.

Chapter 6

The Shape of Water: High–Pressure/High–Temperature Data Modeling

Our model comparison code is used to analyze monochromatic powder x-ray diffraction data taken at Sector 16 ID-B, HPCAT, of the Advanced Photon Source at Argonne National Lab by the research group headed by Dr. Ashkan Salamat. Their data are taken using a new laser annealing method to achieve better precision than earlier studies. This precision allows us to investigate phase transitions in a region where it previously has been difficult. We see phase transitions in two separate datasets.

6.1 Bragg's Law

In 1913, father-son team William Henry and Lawrence Bragg [58] noticed bright X-ray reflections were produced when crystals were bombarded with certain wavelengths and at specific incident angles. They modeled the crystalline structure as a set of parallel planes, separated by a constant spacing d. The measured intensity is related to the constructive/destructive interference of the incident X-rays and the plane spacing. This lead to the formulation of Bragg's law:

$$\begin{aligned} \lambda &= 2d\sin(\theta) \\ d &= \begin{cases} \dfrac{a}{\sqrt{h^2+k^2+l^2}}, & \text{if Cubic} \\ \left[\dfrac{h^2}{a^2}+\dfrac{k^2+l^2}{c^2}\right]^{-\frac{1}{2}}, & \text{if Tetragonal} \end{cases}, \end{aligned} \tag{6.1.0.1}$$

where λ is set to 0.406626Å, the wavelength of incident X-rays; θ is the angle where the reflected radiation is the most intense; h, k, and l are the Miller indices, the notation system used in crystallography for the different planes in the crystal lattice; and a/c are the unit cell dimensions. The cubic model only requires one cell length—a—while the tetragonal model requires two cell lengths—a and c. The shape of the unit cell (cubic or tetragonal) will vary

the location of the reflected emission. Salamat et al. obtained X-ray diffraction data from these reflections off of ice.

6.1.1 Pipeline Set-Up

We use uniform priors for the model parameters because they are normalizable and objective. The only difference between the cubic and tetragonal models are the number of lattice parameters; a and a/c. The prior bounds for each lattice parameter is between 1 and 10 Å.

6.1.2 Results

Figure 6.1 shows the Bayes factors between the tetragonal and cubic models. Table 6.1 shows the logarithm of the Bayes factors (LBF) at each pressure. Positive LBFs prefer the tetragonal model, while negative LBFs prefer the cubic model. For a majority of the range of pressures, the tetragonal is preferred over the cubic model. Ice-VII is not solidified until 2.75 GPa. The cell is cubic until 5.42 GPa, where the cell changes to tetragonal. At large pressures, the cell returns to a cubic cell. There are only three or four data points (peak measurements) at 18.77 GPa and above. Thus, while the LBFs indicate that the tetragonal model fits better at large pressures, the lack of sufficient number of data points compared to the number of model parameters (two) needs to be noted. Another consideration is the peak fitting algorithm used could be refined to yield more accurate θs.

6.2 Equation of State

They obtained pressure and volume data for their sample. This data is described by an equation of state (EOS). We fit eight EOS to the data: a single-phase Birch-Murnaghan [59], a single-phase Rose-Vinet [60], a two-phase Birch-Murnaghan, and a two-phase Rose-Vinet, a three-phase Birch-Murnaghan with the derivative of the bulk modulus with respect to pressure (B'_{o}) set to 4, a three-phase Rose-Vinet with B'_{o} set to 4, a three-phase Birch-Murnaghan, and a three-phase Rose-Vinet. The two-phase fits are modeling cubic ice-VII and cubic ice-X. The three-phase fits are modeling cubic ice-VII, tetragonal ice-VII$_{\mathrm{t}}$, and cubic ice-X. Each model includes two additional parameters, β and γ to model the pressure dependence of our systematic uncertainties. The uncertainties for points that were not laser heated are inflated by the function,

$$\alpha(P) = \sigma_P(\beta + \gamma P), \tag{6.2.0.1}$$

where σ_P is the estimated pressure uncertainty, which comes from the distortions in the volume of our Au pressure marker. This weights the laser annealed points in the EOS fitting more than the non-laser annealed ones.

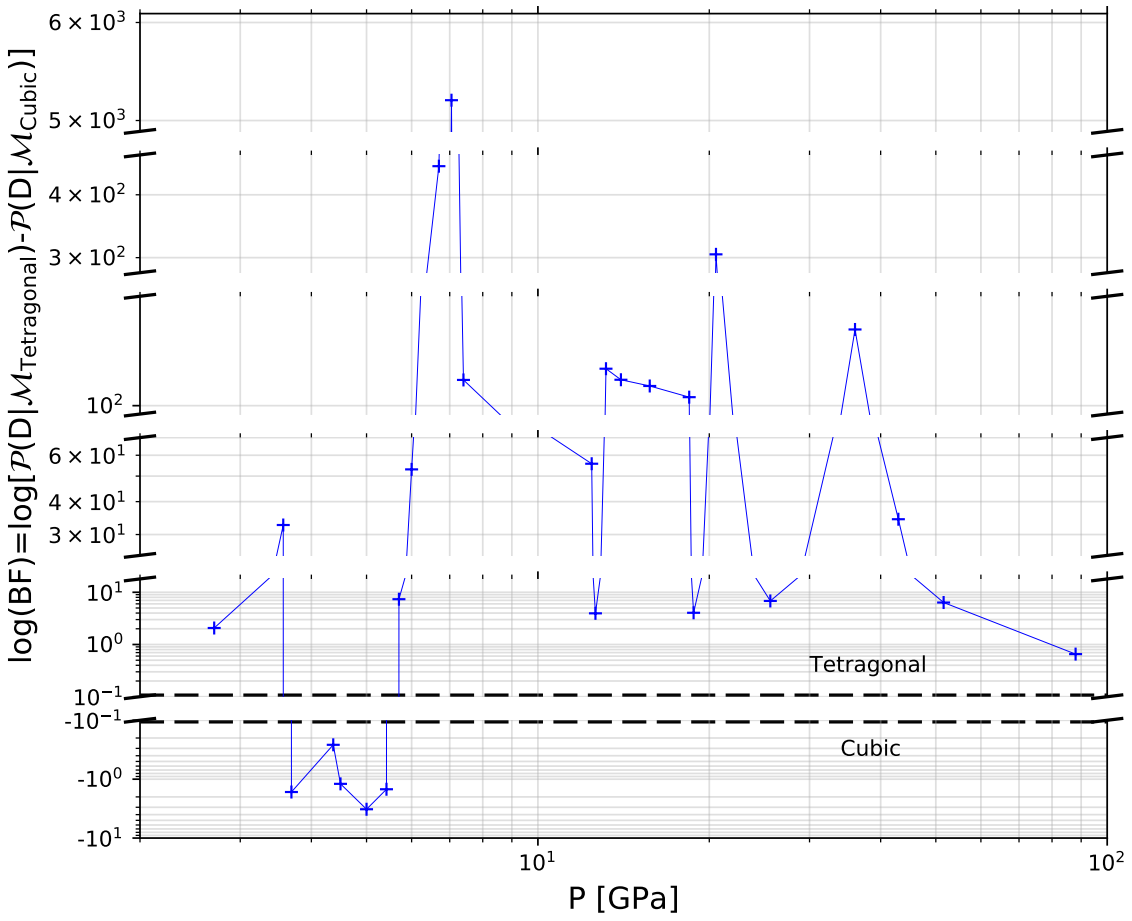

Figure 6.1: This plot is showing the log Bayes factor for the tetragonal unit cell compared to the cubit unit cell. For a majority of the pressure parameter space, the tetragonal cell is clearly preferred.

Pressure (GPa)	$\log_{10}$(Bayes factor)	N_{data}
2.70	2	6
3.57	33	6
3.69	-2	6
4.37	-0	6
4.50	-1	7
5.00	-3	7
5.42	-2	6
5.70	7	6
6.00	53	7
6.25	228	6
6.70	456	6
7.05	5193	6
7.40	117	6
12.43	56	6
12.62	4	6
13.17	125	7
13.99	117	7
15.72	113	6
18.44	105	7
18.77	4	3
20.55	305	4
25.60	7	4
36.06	159	3
42.97	34	3
51.60	6	3
88.00	1	3

Table 6.1: The Bayes factors from the model comparison between the cubic and tetragonal unit cell, rounded to the nearest order of magnitude. There also lists the number of data points that went in to the unit cell fit.

6.2.1 Single-Phase Birch-Murnaghan

The single-phase Birch-Murnaghan EOS is given by the following:

$$P(V) = \frac{3B_{\mathrm{o}}}{2}\left[\left(\frac{V_{\mathrm{o}}}{V}\right)^{\frac{7}{3}} - \left(\frac{V_{\mathrm{o}}}{V}\right)^{\frac{5}{3}}\right]\left\{1 + \frac{3}{4}(B'_{\mathrm{o}} - 4)\left[\left(\frac{V_{\mathrm{o}}}{V}\right)^{\frac{2}{3}} - 1\right]\right\} \tag{6.2.1.1}$$

where V are the measured volumes, σ_V are the measured volume errors, P are the measured pressures, σ_P are the measured pressure errors, V_{o} is the reference volume, B_{o} is the bulk modulus, and B'_{o} is the derivative of the bulk modulus with respect to pressure.

6.2.1.1 Pipeline Set-Up

We use uniform priors for the model parameters because they are normalizable and objective. The prior bounds for V_{o} is between 5 and 60 cm^3 mol^{-1}. The prior bounds for B_{o} is between 4 and 100 GPa. The prior bounds for B'_{o} is between 1 and 10. We use a uniform prior for β and a Rayleigh prior for γ because they are normalizable and objective. The Rayleigh parameter is set to 0.009 to allow $\alpha(P)$ to go from 1 (no change) to 4. The slope of $\alpha(P)$ must be positive and the value of $\alpha(P)$ is forced to be ≥ 1 at pressures greater than 2.71371 GPa—ensuring the errorbars do not shrink.

6.2.2 Single-Phase Rose-Vinet

The single-phase Rose-Vinet EOS is given by the following:

$$\begin{aligned} \eta &= \sqrt[3]{\frac{V}{V_{\mathrm{o}}}} \\ P(V) &= 3\,B_{\mathrm{o}}\left(\frac{1-\eta}{\eta^2}\right)\,\exp\left[\frac{3}{2}(B'_{\mathrm{o}} - 1)\,(1-\eta)\right] \end{aligned} \tag{6.2.2.1}$$

where V are the measured volumes, σ_V are the measured volume errors, P are the measured pressures, σ_P are the measured pressure errors, V_{o} is the reference volume, B_{o} is the bulk modulus, and B'_{o} is the derivative of the bulk modulus with respect to pressure.

6.2.2.1 Pipeline Set-Up

We use uniform priors for the model parameters because they are normalizable and objective. The prior bounds for V_{o} is between 5 and 60 cm^3 mol^{-1}. The prior bounds for B_{o} is between 4 and 100 GPa. The prior bounds for B'_{o} is between 1 and 10. We use a uniform prior for β and a Rayleigh prior for γ because they are normalizable and objective. The Rayleigh parameter is set to 0.009 to allow $\alpha(P)$ to go from 1 (no change) to 4. The slope of $\alpha(P)$ must be positive and the value of $\alpha(P)$ is forced to be ≥ 1 at pressures greater than 2.71371 GPa—ensuring the errorbars do not shrink.

6.2.3 Two-Phase Birch-Murnaghan

The two-phase Birch-Murnaghan EOS is composed of two Birch-Murnaghan EOS with one transition pressure. The two-phase Birch-Murnaghan is given by the following:

$$EOS(V_{\rm o}, B_{\rm o}, B'_{\rm o}) = \frac{3B_{\rm o}}{2}\left[\left(\frac{V_{\rm o}}{V}\right)^{\frac{7}{3}} - \left(\frac{V_{\rm o}}{V}\right)^{\frac{5}{3}}\right]\left\{1 + \frac{3}{4}(B'_{\rm o} - 4)\left[\left(\frac{V_{\rm o}}{V}\right)^{\frac{2}{3}} - 1\right]\right\} \quad (6.2.3.1)$$

$$P(V) = \begin{cases} EOS(V_{\rm o,VII}, B_{\rm o,VII}, B'_{\rm o,VII}), & \text{if } P < \mathbb{TP}_{\rm VII\rightarrow X} \\ EOS(V_{\rm o,X}, B_{\rm o,X}, B'_{\rm o,X}), & \text{if } \mathbb{TP}_{\rm VII\rightarrow X} < P \end{cases}, \quad (6.2.3.2)$$

where V are the measured volumes, σ_V are the measured volume errors, P are the measured pressures, σ_P and are the measured pressure errors. $V_{{\rm o},i}$ is the reference volume, $B_{{\rm o},i}$ is the bulk modulus, and $B'_{{\rm o},i}$ is the derivative of the bulk modulus with respect to pressure for phase i where the phases are ice-VII (cubic cell) and ice-X (cubic cell). $\mathbb{TP}_{\rm VII\rightarrow X}$ is the transition pressure from ice-VII to ice-X.

6.2.3.1 Pipeline Set-Up

We use uniform priors for the model parameters because they are normalizable and objective. The prior bounds for $V_{\rm o,VII}$ and $V_{\rm o,X}$ is between 5 and 60 cm^3 mol^{-1}. The prior bounds for $K_{\rm o,VII}$ and $K_{\rm o,X}$ is between 4 and 100 GPa. The prior bounds for transition pressure between ice-VII and ice-X is between 28 and 55 GPa. We use a uniform prior for β and a Rayleigh prior for γ because they are normalizable and objective. The Rayleigh parameter is set to 0.009 to allow $\alpha(P)$ to go from 1 (no change) to 4. The slope of $\alpha(P)$ must be positive and the value of $\alpha(P)$ is forced to be ≥ 1 at pressures greater than 2.71371 GPa—ensuring the errorbars do not shrink.

6.2.4 Two-Phase Rose-Vinet

The two-phase EOS is composed of two Rose-Vinet EOS with one transition pressure. The EOS is given by the following:

$$EOS(V_{\rm o}, B_{\rm o}, B'_{\rm o}) = 3\,B_{\rm o}\left(\frac{1-\eta}{\eta^2}\right)\,\exp\left[\frac{3}{2}(B'_{\rm o} - 1)\,(1-\eta)\right] \quad (6.2.4.1)$$

$$\eta = \sqrt[3]{\frac{V}{V_{\rm o}}}$$

$$P(V) = \begin{cases} EOS(V_{\rm o,VII}, B_{\rm o,VII}, B'_{\rm o,VII}), & \text{if } P < \mathbb{TP}_{\rm VII\rightarrow X} \\ EOS(V_{\rm o,X}, B_{\rm o,X}, B'_{\rm o,X}), & \text{if } \mathbb{TP}_{\rm VII\rightarrow X} < P \end{cases}, \quad (6.2.4.2)$$

where V are the measured volumes, σ_V are the measured volume errors, P are the measured pressures, σ_P and are the measured pressure errors. $V_{{\rm o},i}$ is the reference volume, $B_{{\rm o},i}$ is the bulk modulus, and $B'_{{\rm o},i}$ is the derivative of the bulk modulus with respect to pressure

for phase i where the phases are ice-VII (cubic cell) and ice-X (cubic cell). $\mathbb{TP}_{\mathrm{VII}\rightarrow\mathrm{X}}$ is the transition pressure from ice-VII to ice-X.

6.2.4.1 Pipeline Set-Up

We use uniform priors for the model parameters because they are normalizable and objective. The prior bounds for $V_{\mathrm{o,VII}}$ and $V_{\mathrm{o,X}}$ is between 5 and 60 cm^3 mol^{-1}. The prior bounds for $K_{\mathrm{o,VII}}$ and $K_{\mathrm{o,X}}$ is between 4 and 100 GPa. The prior bounds for transition pressure between ice-VII and ice-X is between 28 and 55 GPa. We use a uniform prior for β and a Rayleigh prior for γ because they are normalizable and objective. The Rayleigh parameter is set to 0.009 to allow $\alpha(P)$ to go from 1 (no change) to 4. The slope of $\alpha(P)$ must be positive and the value of $\alpha(P)$ is forced to be ≥ 1 at pressures greater than 2.71371 GPa—ensuring the errorbars do not shrink.

6.2.5 Triple-Phase Birch-Murnaghan

The triple-phase Birch-Murnaghan EOS is composed of three Birch-Murnaghan EOS with two transition pressures. The triple-phase Birch-Murnaghan is given by the following:

$$EOS(V_{\mathrm{o}}, B_{\mathrm{o}}, B'_{\mathrm{o}}) = \frac{3B_{\mathrm{o}}}{2}\left[\left(\frac{V_{\mathrm{o}}}{V}\right)^{\frac{7}{3}} - \left(\frac{V_{\mathrm{o}}}{V}\right)^{\frac{5}{3}}\right]\left\{1 + \frac{3}{4}(B'_{\mathrm{o}} - 4)\left[\left(\frac{V_{\mathrm{o}}}{V}\right)^{\frac{2}{3}} - 1\right]\right\} \quad (6.2.5.1)$$

$$P(V) = \begin{cases} EOS(V_{\mathrm{o,VII}}, B_{\mathrm{o,VII}}, B'_{\mathrm{o,VII}}), & \text{if } P < \mathbb{TP}_{\mathrm{VII}\rightarrow\mathrm{VII_t}} \\ EOS(V_{\mathrm{o,VII_t}}, B_{\mathrm{o,VII_t}}, B'_{\mathrm{o,VII_t}}), & \text{if } \mathbb{TP}_{\mathrm{VII}\rightarrow\mathrm{VII_t}} < P < \mathbb{TP}_{\mathrm{VII_t}\rightarrow\mathrm{X}} \\ EOS(V_{\mathrm{o,X}}, B_{\mathrm{o,X}}, B'_{\mathrm{o,X}}), & \text{if } \mathbb{TP}_{\mathrm{VII_t}\rightarrow\mathrm{X}} < P \end{cases}, \quad (6.2.5.2)$$

where V are the measured volumes, σ_V are the measured volume errors, P are the measured pressures, σ_P and are the measured pressure errors. $V_{\mathrm{o},i}$ is the reference volume, $B_{\mathrm{o},i}$ is the bulk modulus, and $B'_{\mathrm{o},i}$ is the derivative of the bulk modulus with respect to pressure for phase i where the phases are ice-VII (cubic cell), ice-VII$_{\mathrm{t}}$ (distorted VII, tetragonal cell), ice-X (cubic cell). $\mathbb{TP}_{\mathrm{VII}\rightarrow\mathrm{VII_t}}$ and $\mathbb{TP}_{\mathrm{VII_t}\rightarrow\mathrm{X}}$ are the transition pressure from ice-VII to ice-VII$_{\mathrm{t}}$ and ice-VII$_{\mathrm{t}}$ to ice-X, respectively.

6.2.5.1 Pipeline Set-Up

We use uniform priors for the model parameters because they are normalizable and objective. The prior bounds for $V_{\mathrm{o,VII}}$, $V_{\mathrm{o,VII_t}}$, and $V_{\mathrm{o,X}}$ is between 5 and 60 cm^3 mol^{-1}. The prior bounds for $K_{\mathrm{o,VII}}$, $K_{\mathrm{o,VII_t}}$, $K_{\mathrm{o,X}}$ is between 4 and 100 GPa. The prior bounds for transition pressure between ice-VII and ice-VII$_{\mathrm{t}}$ is between 3 and 11.75 GPa. The prior bounds for transition pressure between ice-VII$_{\mathrm{t}}$ and ice-X is between 28 and 55 GPa. We use a uniform prior for β and a Rayleigh prior for γ because they are normalizable and objective. The Rayleigh parameter is set to 0.009 to allow $\alpha(P)$ to go from 1 (no change) to 4. The slope of $\alpha(P)$ must be positive and the value of $\alpha(P)$ is forced to be ≥ 1 at pressures greater than 2.71371 GPa—ensuring the errorbars do not shrink.

6.2.6 Triple-Phase Rose-Vinet

The triple-phase EOS is composed of three second-order Rose-Vinet EOS with two transition pressures. The EOS is given by the following:

$$EOS(V_{\rm o}, B_{\rm o}, B'_{\rm o}) = 3\,B_{\rm o}\left(\frac{1-\eta}{\eta^2}\right)\,\exp\left[\frac{3}{2}(B'_{\rm o}-1)\,(1-\eta)\right] \qquad (6.2.6.1)$$
$$\eta = \sqrt[3]{\frac{V}{V_{\rm o}}}$$

$$P(V) = \begin{cases} EOS(V_{\rm o,VII}, B_{\rm o,VII}, B'_{\rm o,VII}), & \text{if } P < \mathbb{TP}_{\rm VII\rightarrow VII_t} \\ EOS(V_{\rm o,VII_t}, B_{\rm o,VII_t}, B'_{\rm o,VII_t}), & \text{if } \mathbb{TP}_{\rm VII_t\rightarrow VII_t} < P < \mathbb{TP}_{\rm VII_t\rightarrow X}\ , \\ EOS(V_{\rm o,X}, B_{\rm o,X}, B'_{\rm o,X}), & \text{if } \mathbb{TP}_{\rm VII_t\rightarrow X} < P \end{cases} \qquad (6.2.6.2)$$

where V are the measured volumes, σ_V are the measured volume errors, P are the measured pressures, σ_P and are the measured pressure errors. $V_{{\rm o},i}$ is the reference volume, $B_{{\rm o},i}$ is the bulk modulus, and $B'_{{\rm o},i}$ is the derivative of the bulk modulus with respect to pressure for phase i where the phases are ice-VII (cubic cell), ice-VII$_{\rm t}$ (distorted VII, tetragonal cell), ice-X (cubic cell). $\mathbb{TP}_{\rm VII\rightarrow VII_t}$ and $\mathbb{TP}_{\rm VII_t\rightarrow X}$ are the transition pressure from ice-VII to ice-VII$_{\rm t}$ and ice-VII$_{\rm t}$ to ice-X, respectively.

6.2.6.1 Pipeline Set-Up

We use uniform priors for the model parameters because they are normalizable and objective. The prior bounds for $V_{\rm o,VII}$, $V_{\rm o,VII_t}$, and $V_{\rm o,X}$ is between 5 and 60 cm^3 mol^{-1}. The prior bounds for $K_{\rm o,VII}$, $K_{\rm o,VII_t}$, $K_{\rm o,X}$ is between 4 and 100 GPa. The prior bounds for transition pressure between ice-VII and ice-VII$_{\rm t}$ is between 3 and 11.75 GPa. The prior bounds for transition pressure between ice-VII$_{\rm t}$ and ice-X is between 28 and 55 GPa. We use a uniform prior for β and a Rayleigh prior for γ because they are normalizable and objective. The Rayleigh parameter is set to 0.009 to allow $\alpha(P)$ to go from 1 (no change) to 4. The slope of $\alpha(P)$ must be positive and the value of $\alpha(P)$ is forced to be ≥ 1 at pressures greater than 2.71371 GPa—ensuring the errorbars do not shrink.

6.2.7 Results

The fit parameters for each model are in Tables 6.2–6.9 and the pressure vs. volume plots are shown in Figures 6.2–6.9. For the multi-phase EOS, the error on the transition pressures are of similar length to the spacings between the pressure data. This is indicated by the small error in the transition from ice-VII to ice-VII$_{\rm t}$ as opposed to the larger error on the transition from ice-VII$_{\rm t}$ to ice-X. Additional sampling near the transition pressure are needed for a better constraint. A Bayes factor analysis is done on the models. The table showing the Bayes factors between each model is shown in Table 6.10. Here, we compare each pair of model and determine that the three-phase Rose-Vinet EOS fits better than the others. There is an indicator in the normalized pressure verses Eulerian strain that clearly shows

a transition occurring at low pressures. Despite the model comparison code favoring the two-phase fit, there is a physical reason to favor the three-phase fit over the two-phase fit.

6.3 Conclusions

In this chapter my model comparison pipeline was able to quantifiably show there is a previously unreported phase transition in ice-VII near 5 GPa from a cubic cell to a tetragonal cell. We used two methods to do this. The first was by measuring Bragg's Law—the diffraction of X-rays off of the crystal lattice. The second is by fitting a three-phase Vinet EOS to pressure-volume data. Both methods also show a transition to ice-VII_t and a return to ice-X at high pressures. The uncertainties in the parameters are largely due to a lack of data quality, thus more data is required to better constrain the models.

V_o [Å^3]	B_o [GPa]	B'_o [GPa]
42.20 ± 0.13	16.05 ± 0.41	4.97 ± 0.05
β = 3.51 ± 0.49, γ=0.00201 ± 0.00200		

Table 6.2: The parameters fit for the single-phase Burch-Murnaghan Equation of State.

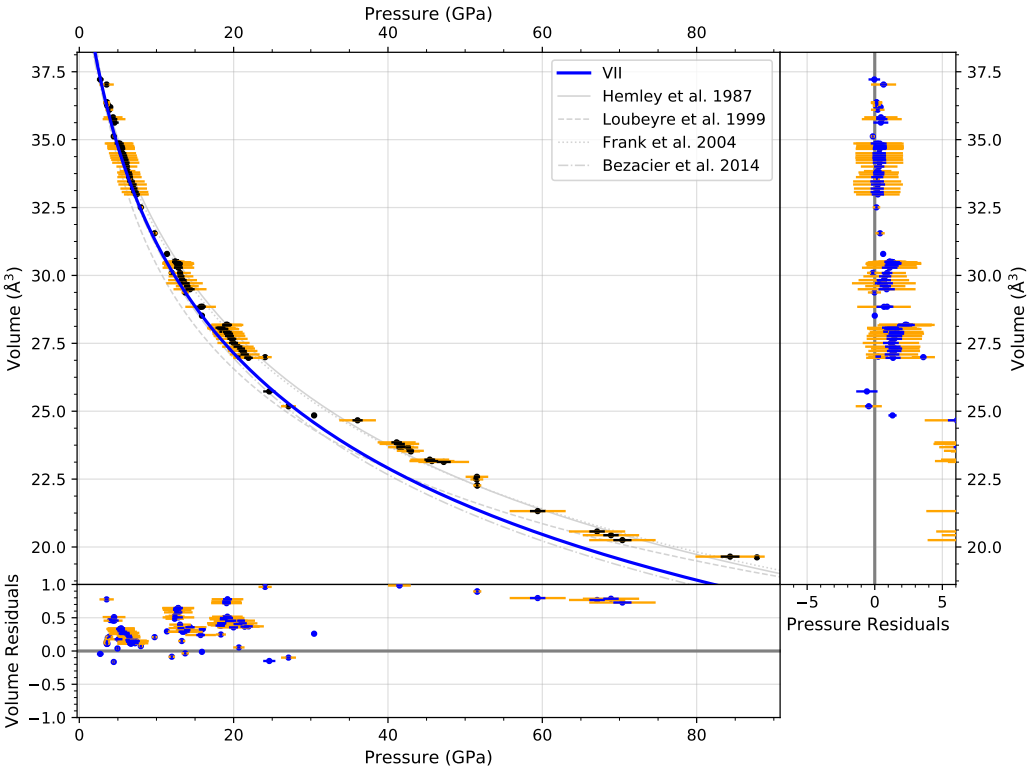

Figure 6.2: The resulting fit of the single-phase Burch-Murnaghan Equation of State. The fit is in blue and the data are in black. The orange error bars are the non-laser heater points and have their uncertainties inflated by Equation 6.2.0.1. The grey solid, dashed, dotted, and dot-dash lines come from Loubeyre et al. [61], Hemley et al. [62], Frank et al. [63], and Bezacier et al. [64], respectively.

V_o [Å^3]	B_o [GPa]	B'_o [GPa]
43.05 ± 0.20	12.57 ± 0.50	6.06 ± 0.07
$\beta = 3.57 \pm 0.43$, $\gamma = 0.00231 \pm 0.00230$		

Table 6.3: The parameters fit for the single-phase Rose-Vinet Equation of State.

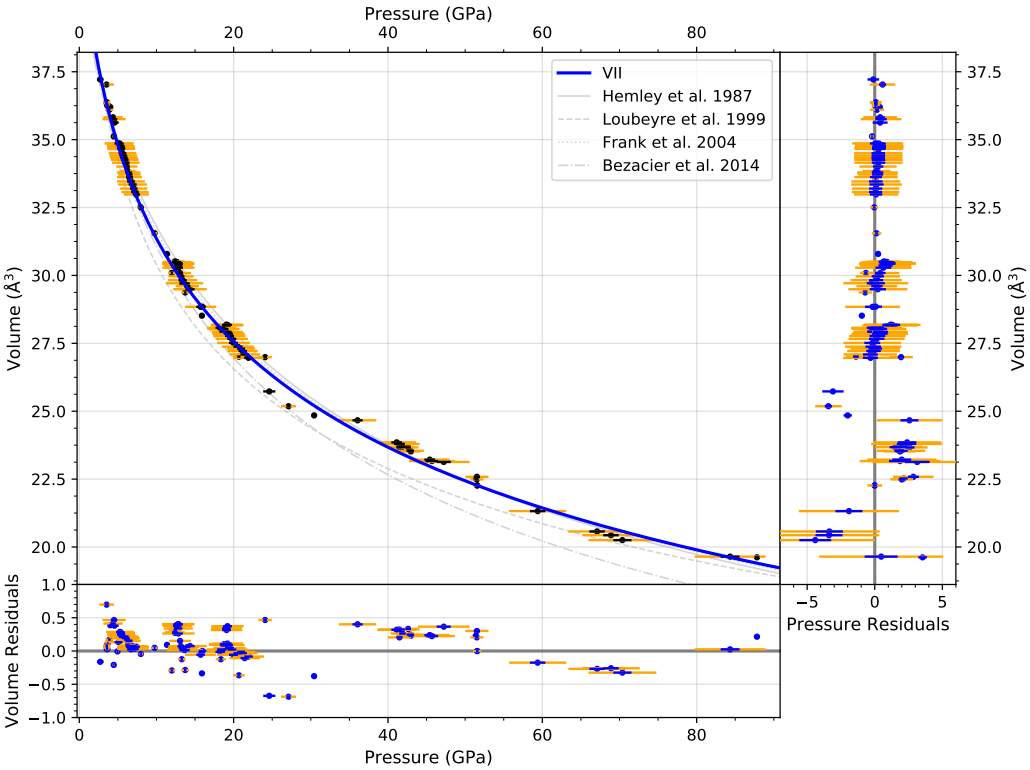

Figure 6.3: The resulting fit of the single-phase Rose-Vinet Equation of State. The fit is in blue and the data are in black. The orange error bars are the non-laser heater points and have their uncertainties inflated by Equation 6.2.0.1. The grey solid, dashed, dotted, and dot-dash lines come from Loubeyre et al. [61], Hemley et al. [62], Frank et al. [63], and Bezacier et al. [64], respectively.

Phase	V_o [Å^3]	B_o [GPa]	B'$_o$
VII	41.57 ± 0.29	19.61 ± 1.32	4.21 ± 0.17
Transition Pressure from Ice VII to Ice X at 33.17 ± 2.89 GPa			
X	35.01 ± 0.92	41.92 ± 6.48	4.26 ± 0.16
β = 1.96 ± 0.24, γ= 0.02235 ± 0.00285			

Table 6.4: The parameters fit for the two-phase Burch-Murnaghan Equation of State.

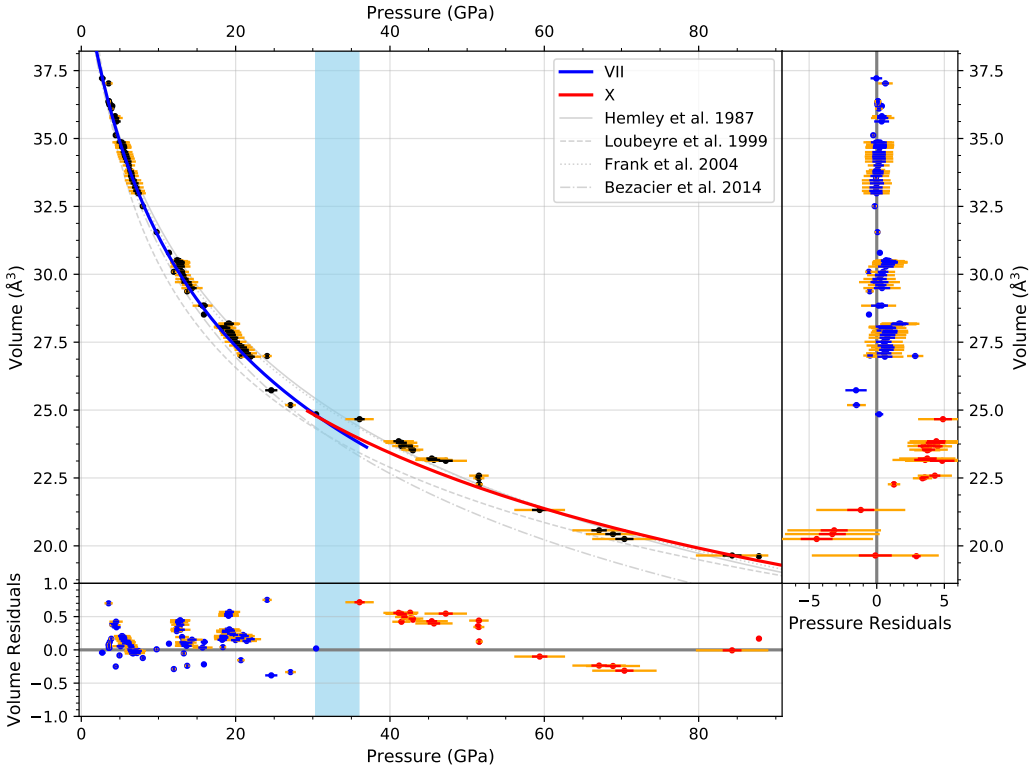

Figure 6.4: The resulting fit of the two-phase Burch-Murnaghan Equation of State. The blue curve is the VII cubic cell and the red is the X cubic cell. The transition pressures are indicated by the shaded light blue region. The width of the region is the error in that transition pressure. The orange error bars are the non-laser heater points and have their uncertainties inflated by Equation 6.2.0.1. The grey solid, dashed, dotted, and dot-dash lines come from Loubeyre et al. [61], Hemley et al. [62], Frank et al. [63], and Bezacier et al. [64], respectively.

Phase	V_o [Å^3]	B_o [GPa]	B'_o
VII	41.82 ± 0.35	17.86 ± 1.57	4.83 ± 0.27
Transition Pressure from Ice VII to Ice X at 31.03 ± 3.03 GPa			
X	34.73 ± 0.76	42.08 ± 6.12	4.86 ± 0.24
β = 1.97 ± 0.27, γ= 0.02255 ± 0.00327			

Table 6.5: The parameters fit for the two-phase Rose-Vinet Equation of State.

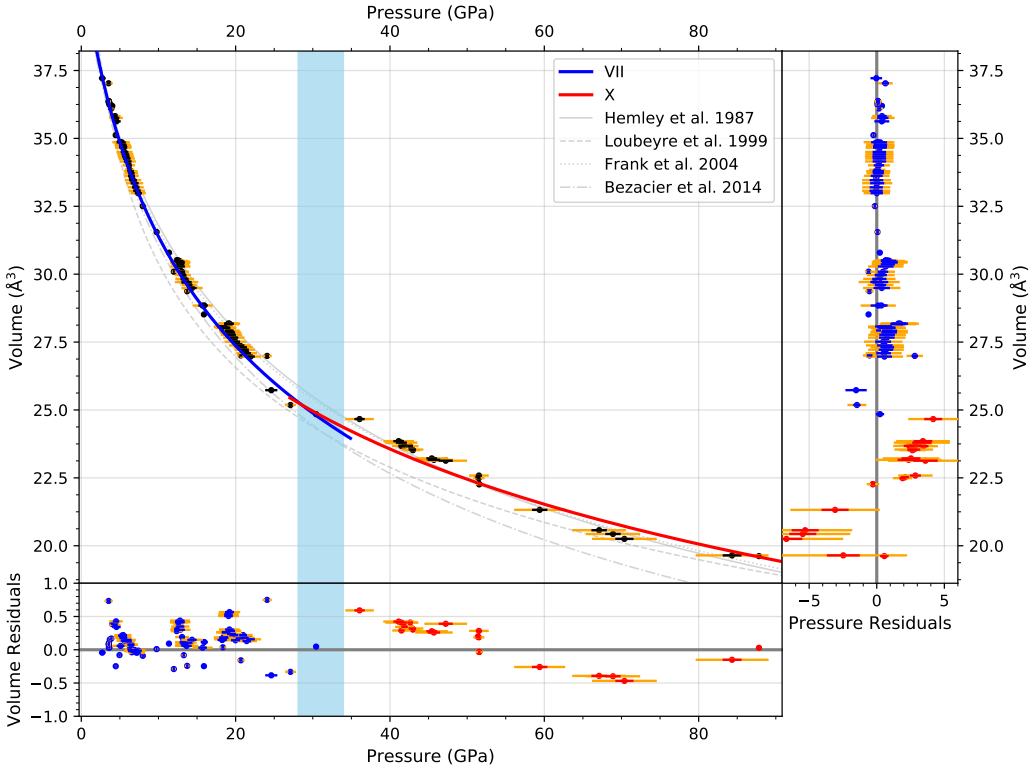

Figure 6.5: The resulting fit of the two-phase Rose-Vinet Equation of State. The blue curve is the VII cubic cell and the red is the X cubic cell. The transition pressure is indicated by the shaded light blue region. The width of the region is the error in that transition pressure. The orange error bars are the non-laser heater points and have their uncertainties inflated by Equation 6.2.0.1. The grey solid, dashed, dotted, and dot-dash lines come from Loubeyre et al. [61], Hemley et al. [62], Frank et al. [63], and Bezacier et al. [64], respectively.

Phase	V_o [Å^3]	B_o [GPa]
VII	44.98 ± 0.93	10.93 ± 1.43
Transition Pressure from Ice VII to Ice VII_t at 4.76 ± 0.17 GPa		
VII_t	40.84 ± 0.10	22.55 ± 0.29
Transition Pressure from Ice VII_t to Ice X at 32.36 ± 1.92 GPa		
X	44.19 ± 1.35	18.88 ± 2.02
β = 3.95 ± 0.05, γ= 0.00055 ± 0.00055		

Table 6.6: The parameters fit for the three-phase Burch Murnaghan Equation of State, fixing B'_o to 4.

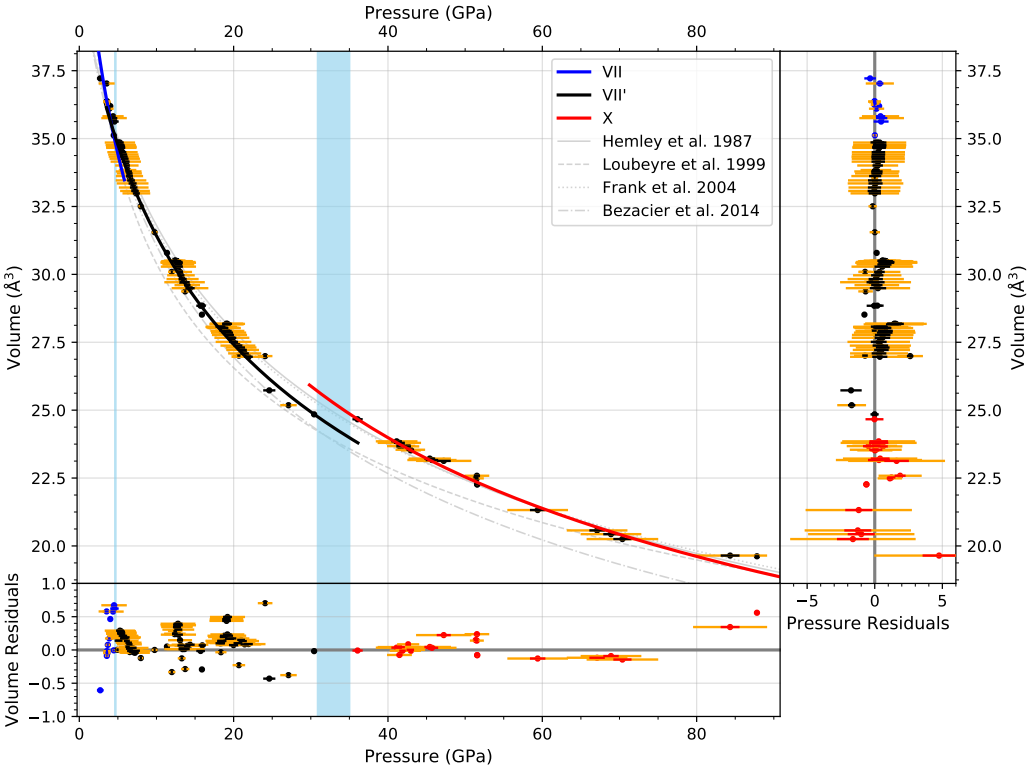

Figure 6.6: The resulting fit of the three-phase Burch-Murnaghan Equation of State fixing B_o' to 4. The blue curve is the VII cubic cell, the black is the VII_t tetragonal cell, and the red is the X cubic cell. The transition pressures are indicated by the shaded light blue region. The width of the region is the error in that transition pressure. The orange error bars are the non-laser heater points and have their uncertainties inflated by Equation 6.2.0.1. The grey solid, dashed, dotted, and dot-dash lines come from Loubeyre et al. [61], Hemley et al. [62], Frank et al. [63], and Bezacier et al. [64], respectively.

Phase	V_o [Å^3]	B_o [GPa]
VII	44.92 ± 1.07	11.14 ± 1.70
Transition Pressure from Ice VII to Ice VII_t at 4.74 ± 0.21 GPa		
VII_t	40.40 ± 0.08	24.61 ± 0.28
Transition Pressure from Ice VII_t to Ice X at 32.36 ± 1.92 GPa		
X	40.25 ± 0.53	28.94 ± 1.30
β = 3.96 ± 0.04, γ= 0.00055 ± 0.00055		

Table 6.7: The parameters fit for the three-phase Rose-Vinet Equation of State, fixing B'_o to 4.

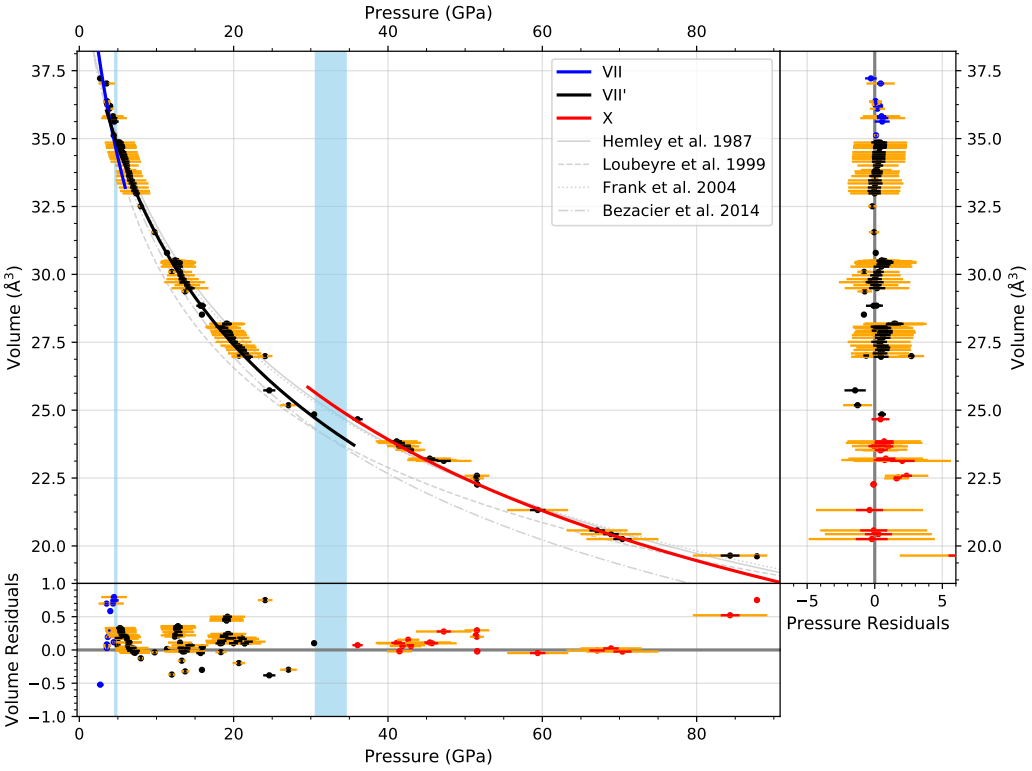

Figure 6.7: The resulting fit of the three-phase Rose-Vinet Equation of State fixing B'_o to 4. The blue curve is the VII cubic cell, the black is the VII_t tetragonal cell, and the red is the X cubic cell. The transition pressures are indicated by the shaded light blue region. The width of the region is the error in that transition pressure. The orange error bars are the non-laser heater points and have their uncertainties inflated by Equation 6.2.0.1. The grey solid, dashed, dotted, and dot-dash lines come from Loubeyre et al. [61], Hemley et al. [62], Frank et al. [63], and Bezacier et al. [64], respectively.

Phase	V_o [Å^3]	B_o [GPa]	B'$_o$
VII	42.17 ± 0.88	15.52 ± 2.99	5.09 ± 1.39
Transition Pressure from Ice VII to Ice VII$_t$ at 4.49 ± 0.49 GPa			
VII$_t$	40.88 ± 0.44	22.56 ± 2.09	4.00 ± 0.20
Transition Pressure from Ice VII$_t$ to Ice X at 32.77 ± 2.25 GPa			
X	34.76 ± 0.52	45.93 ± 4.16	4.15 ± 0.10
$\beta = 1.81 \pm 0.29$, $\gamma = 0.02437 \pm 0.00363$			

Table 6.8: The parameters fit for the three-phase Burch Murnaghan Equation of State.

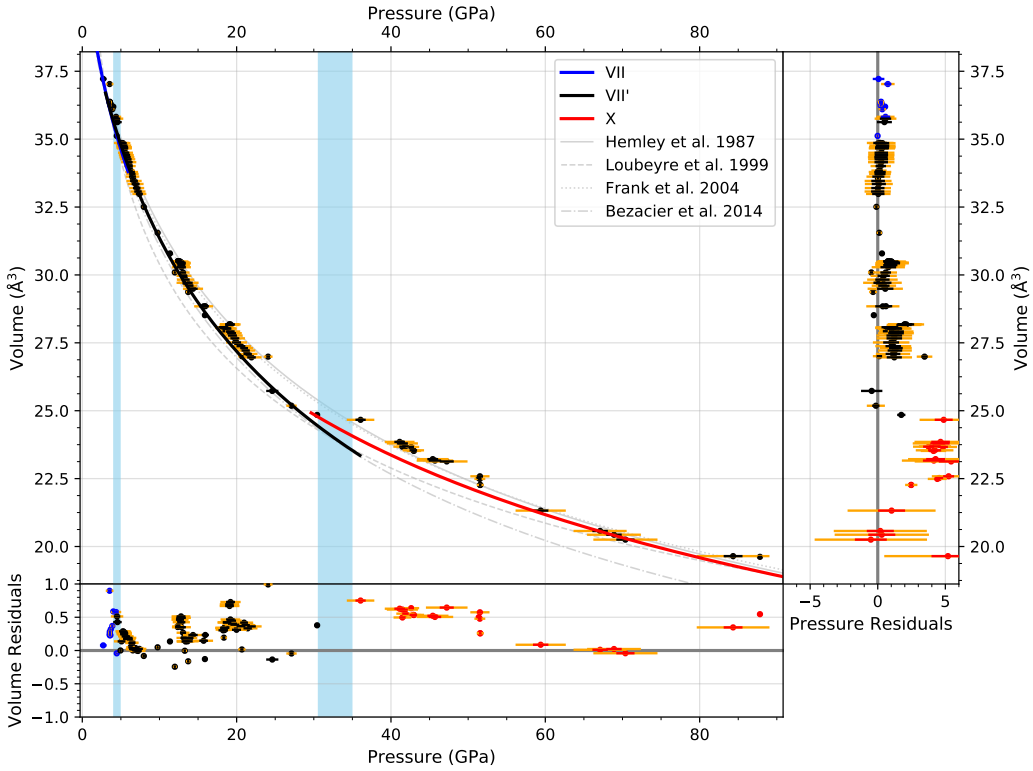

Figure 6.8: The resulting fit of the three-phase Burch-Murnaghan Equation of State. The blue curve is the VII cubic cell, the black is the VII$_t$ tetragonal cell, and the red is the X cubic cell. The transition pressures are indicated by the shaded light blue region. The width of the region is the error in that transition pressure. The orange error bars are the non-laser heater points and have their uncertainties inflated by Equation 6.2.0.1. The grey solid, dashed, dotted, and dot-dash lines come from Loubeyre et al. [61], Hemley et al. [62], Frank et al. [63], and Bezacier et al. [64], respectively.

Phase	V_o [Å^3]	B_o [GPa]	B'$_o$
VII	42.50 ± 0.88	18.47 ± 4.00	2.51 ± 1.51
Transition Pressure from Ice VII to Ice VII$_t$ at 4.78 ± 0.76 GPa			
VII$_t$	41.11 ± 0.53	20.76 ± 2.46	4.49 ± 0.35
Transition Pressure from Ice VII$_t$ to Ice X at 30.91 ± 2.90 GPa			
X	33.82 ± 0.43	50.52 ± 4.16	4.50 ± 0.15
$\beta = 1.83 \pm 0.30$, $\gamma = 0.02312 \pm 0.00381$			

Table 6.9: The parameters fit for the three-phase Rose-Vinet Equation of State.

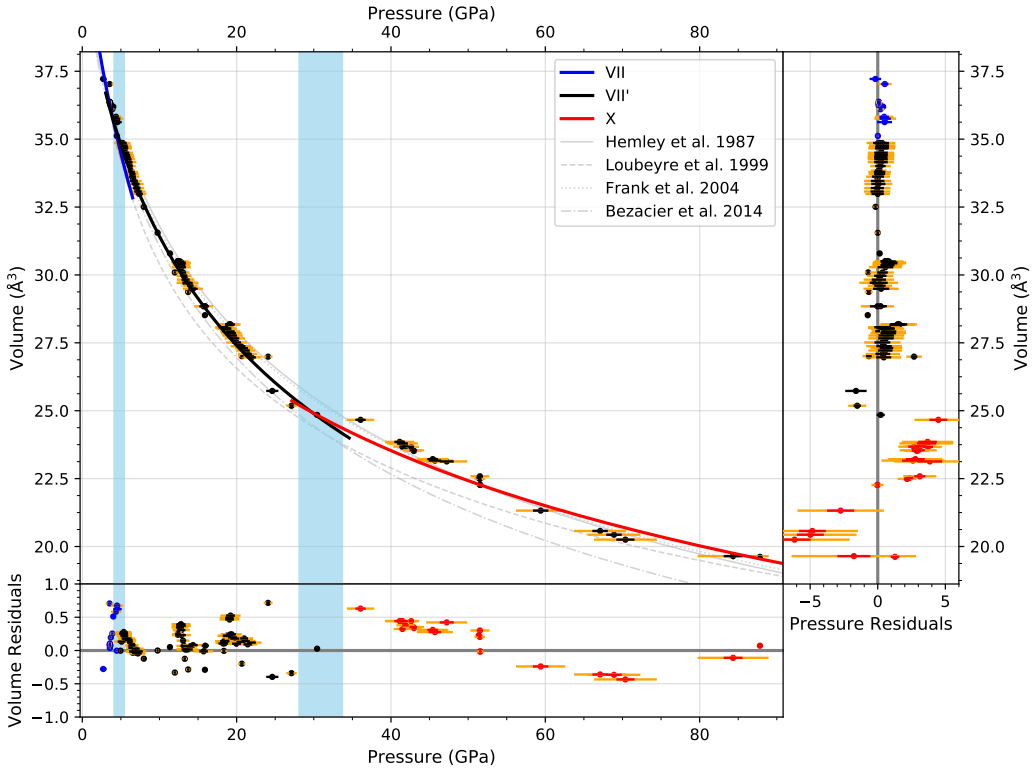

Figure 6.9: The resulting fit of the three-phase Rose-Vinet Equation of State. The blue curve is the VII cubic cell, the black is the VII$_t$ tetragonal cell, and the red is the X cubic cell. The transition pressures are indicated by the shaded light blue region. The width of the region is the error in that transition pressure. The orange error bars are the non-laser heater points and have their uncertainties inflated by Equation 6.2.0.1. The grey solid, dashed, dotted, and dot-dash lines come from Loubeyre et al. [61], Hemley et al. [62], Frank et al. [63], and Bezacier et al. [64], respectively.

	1 BM	**1 RV**	**2 BM**	**2 RV**	**3 BM F**	**3 RV F**	**3 BM**	**3 RV**
1 BM	-	6	98	96	65	61	94	93
1 RV	-7	-	91	90	59	54	88	86
2 BM	-98	-92	-	-2	-33	-37	-4	-5
2 RV	-97	-90	1	-	-31	-36	-2	-4
3 BM F	-65	-59	33	31	-	-4	29	28
3 RV F	-62	-55	37	35	4	-	33	31
3 BM	-95	-88	4	2	-29	-34	-	-2
3 RV	-93	-87	5	3	-28	-32	1	-

Table 6.10: The Bayes factors between all EOS models fit to the pressure/volume data, rounded to the nearest order of magnitude. The value is the EOS of the **bold** EOS versus the non-bold EOS.

Chapter 7

Discussion and Conclusions

The theme of this dissertation, and my time working under Dr. Jason Steffen, is data modeling in planetary science. The state-of-the-art model comparison code that I developed is useful, not only in RV exoplanet studies, but everywhere model fitting and comparison is needed. It has been used to solve five problems. Three of them relating to RV exoplanet detection and two relating to high-pressure/high-temperature phase data of H_2O-ice.

My code was used to distinguish between RV models near the 2:1 degeneracy and identify systems from the NASA Archive where there may be another planet. We find that 15 out of the 60 main sequence systems show compelling evidence for an additional planet with a confidence level of 95%. Our findings imply that there are hundreds of missing planets in NASA's archive. The best potential candidates of those systems had more data taken from *APF* and the Bayes factors computed with the new data continue to shows evidence for an additional planet for all but one system. To test the effectiveness of different observational campaigns, the code was used to test different sixteen year observational schemes. This undertaking contained fourteen thousand pipeline runs, totaling fifty-six thousand fits. We found that focusing on obtaining data when the system is near phases where the 2:1 degeneracy is the weakest decreases the ambiguity caused by the degeneracy more than obtaining data at random phases. The observational campaigns also highlight the need for accurate noise modeling in order to distinguish between a planet and a stellar source via S-index measurements. In the future, this code will be used with additional noise models than fit in this dissertation.

The high-pressure/high-temperature projects are useful for modeling planet interiors, as the internal pressures of planets are very strong and their descriptions require accurate equation of state measurements. Our code was able to identify two phase transitions in Dr. Salamat's pressure and temperature water-ice data taken at Argonne National Lab. One of the phase transitions we found—cubic ice-VII to tetragonal ice-VII_t—was previously unreported until now. Our code saw these phase transitions in X-ray diffraction data, which uses Bragg's law to peer into the crystal lattice of water, and in pressure–volume equation of state fits.

The Bayesian approach to model comparison is useful in any situation where there are two competing models. The end goal for this code is to be open source. The code will be allowed to be edited and adjusted to the user's needs. For example, there will be a way for additional priors and models can be added. The computational power that is achievable

now allows for these sophisticated statistical methods to be easily implemented for every situation that calls for model fitting.

References

[1] K. F. Pearson, X: On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 50 (302) (1900) 157–175. doi:10.1080/14786440009463897.

[2] T. Bayes, R. Price, J. Canton, An essay towards solving a problem in the doctrine of chances.

[3] A. Wolszczan, D. A. Frail, A planetary system around the millisecond pulsar PSR1257 + 12, Nature355 (1992) 145–147. doi:10.1038/355145a0.

[4] J. T. Wright, Radial Velocities as an Exoplanet Discovery Method, 2017, p. 4. doi:10.1007/978-3-319-30648-3_4-1.

[5] M. Mayor, D. Queloz, A Jupiter-mass companion to a solar-type star, Nature378 (1995) 355–359. doi:10.1038/378355a0.

[6] R. P. Butler, G. W. Marcy, D. A. Fischer, T. M. Brown, A. R. Contos, S. G. Korzennik, P. Nisenson, R. W. Noyes, Evidence for Multiple Companions to υ Andromedae, ApJ526 (1999) 916–927. doi:10.1086/308035.

[7] R. P. Butler, G. W. Marcy, E. Williams, H. Hauser, P. Shirts, Three New "51 Pegasi-Type" Planets, ApJ474 (1997) L115–L118. doi:10.1086/310444.

[8] S. Curiel, J. Cantó, L. Georgiev, C. E. Chávez, A. Poveda, A fourth planet orbiting υ Andromedae, A&A525 (2011) A78. doi:10.1051/0004-6361/201015693.

[9] P. D. Sackett, Searching for Unseen Planets via Occultation and Microlensing, in: J.-M. Mariotti, D. Alloin (Eds.), NATO Advanced Science Institutes (ASI) Series C, Vol. 532 of NATO Advanced Science Institutes (ASI) Series C, 1999, p. 189. arXiv:astro-ph/9811269.

[10] D. Charbonneau, T. M. Brown, D. W. Latham, M. Mayor, Detection of Planetary Transits Across a Sun-like Star, ApJ529 (2000) L45–L48. arXiv:astro-ph/9911436, doi:10.1086/312457.

[11] G. W. Henry, G. W. Marcy, R. P. Butler, S. S. Vogt, A Transiting "51 Peg-like" Planet, ApJ529 (2000) L41–L44. doi:10.1086/312458.

[12] A. Einstein, Lens-Like Action of a Star by the Deviation of Light in the Gravitational Field, Science 84 (1936) 506–507. doi:10.1126/science.84.2188.506.

[13] S. Mao, B. Paczynski, Gravitational microlensing by double stars and planetary systems, ApJ374 (1991) L37–L40. doi:10.1086/186066.

[14] J.-P. Beaulieu, D. P. Bennett, P. Fouqué, A. Williams, M. Dominik, U. G. Jørgensen, D. Kubas, A. Cassan, C. Coutures, J. Greenhill, K. Hill, J. Menzies, P. D. Sackett, M. Albrow, S. Brillant, J. A. R. Caldwell, J. J. Calitz, K. H. Cook, E. Corrales, M. Desort, S. Dieters, D. Dominis, J. Donatowicz, M. Hoffman, S. Kane, J.-B. Marquette, R. Martin, P. Meintjes, K. Pollard, K. Sahu, C. Vinter, J. Wambsganss, K. Woller, K. Horne, I. Steele, D. M. Bramich, M. Burgdorf, C. Snodgrass, M. Bode, A. Udalski, M. K. Szymański, M. Kubiak, T. Więckowski, G. Pietrzyński, I. Soszyński, O. Szewczyk, Ł. Wyrzykowski, B. Paczyński, F. Abe, I. A. Bond, T. R. Britton, A. C. Gilmore, J. B. Hearnshaw, Y. Itow, K. Kamiya, P. M. Kilmartin, A. V. Korpela, K. Masuda, Y. Matsubara, M. Motomura, Y. Muraki, S. Nakamura, C. Okada, K. Ohnishi, N. J. Rattenbury, T. Sako, S. Sato, M. Sasaki, T. Sekiguchi, D. J. Sullivan, P. J. Tristram, P. C. M. Yock, T. Yoshioka, Discovery of a cool planet of 5.5 Earth masses through gravitational microlensing, Nature439 (2006) 437–440. arXiv:astro-ph/0601563, doi:10.1038/nature04441.

[15] W. A. Traub, B. R. Oppenheimer, Direct Imaging of Exoplanets, 2010, pp. 111–156.

[16] P. Kalas, J. R. Graham, E. Chiang, M. P. Fitzgerald, M. Clampin, E. S. Kite, K. Stapelfeldt, C. Marois, J. Krist, Optical Images of an Exosolar Planet 25 Light-Years from Earth, Science 322 (2008) 1345. arXiv:0811.1994, doi:10.1126/science.1166609.

[17] M. H. Lee, S. J. Peale, Dynamics and Origin of the 2:1 Orbital Resonances of the GJ 876 Planets, ApJ567 (2002) 596.

[18] C. G. Tinney, et al., The 2:1 Resonant Exoplanetary System Orbiting HD 73526, ApJ647 (2006) 594. arXiv:astro-ph/0602557.

[19] S. Chatterjee, E. B. Ford, S. Matsumura, F. A. Rasio, Dynamical Outcomes of Planet-Planet Scattering, ApJ686 (2008) 580. arXiv:astro-ph/0703166.

[20] E. B. Ford, F. A. Rasio, Origins of Eccentric Extrasolar Planets: Testing the Planet-Planet Scattering Model, ApJ686 (2008) 621. arXiv:astro-ph/0703163.

[21] Y. Kozai, Secular perturbations of asteroids with high inclination and eccentricity, AJ67 (1962) 591.

[22] M. L. Lidov, The evolution of orbits of artificial satellites of planets under the action of gravitational perturbations of external bodies, Planet. Space Sci.9 (1962) 719.

[23] D. Fabrycky, S. Tremaine, Shrinking Binary and Planetary Orbits by Kozai Cycles with Tidal Friction, ApJ669 (2007) 1298. arXiv:0705.4285.

[24] G. Anglada-Escudé, et al., How Eccentric Orbital Solutions Can Hide Planetary Systems in 2:1 Resonant Orbits, ApJ709 (2010) 168. arXiv:0809.1275.

[25] R. A. Wittenmyer, et al., Forever Alone? Testing Single Eccentric Planetary Systems for Multiple Companions, ApJS208 (2013) 2. arXiv:1307.0894.

[26] M. Kürster, et al., Disentangling 2:1 resonant radial velocity orbits from eccentric ones and a case study for HD 27894, A&A577 (2015) A103. arXiv:1503.07769.

[27] J. H. Steffen, J. A. Hwang, The period ratio distribution of Kepler's candidate multi-planet systems, MNRAS448 (2015) 1956. arXiv:1409.3320.

[28] S. E. Thompson, et al., Planetary Candidates Observed by Kepler. VIII. A Fully Automated Catalog With Measured Completeness and Reliability Based on Data Release 25, ArXiv e-prints:1710.06758arXiv:1710.06758.

[29] T. Guillot, THE INTERIORS OF GIANT PLANETS: Models and Outstanding Questions, Annual Review of Earth and Planetary Sciences 33 (2005) 493–530. arXiv:astro-ph/0502068, doi:10.1146/annurev.earth.32.101802.120325.

[30] J. H. Boisvert, B. E. Nelson, J. H. Steffen, Systematic mischaracterization of exoplanetary system dynamical histories from a model degeneracy near mean-motion resonance, MNRAS480 (2018) 2846–2852. arXiv:1804.10143, doi:10.1093/mnras/sty2023.

[31] D. W. Hogg, J. Bovy, D. Lang, Data analysis recipes: Fitting a model to data, ArXiv e-printsarXiv:1008.4686.

[32] H. Jeffreys, An Invariant Form for the Prior Probability in Estimation Problems, Proceedings of the Royal Society of London Series A 186 (1946) 453–461. doi:10.1098/rspa.1946.0056.

[33] J. O. Berger, L. R. Pericchi, The intrinsic bayes factor for linear models, Bayesian statistics 5 (1996) 25–44.

[34] A. Achim, E. E. Kuruoglu, J. Zerubia, Sar image filtering based on the heavy-tailed rayleigh model, IEEE Transactions on Image Processing 15 (9) (2006) 2686–2693. doi:10.1109/TIP.2006.877362.

[35] M. Parchami, W.-P. Zhu, B. Champagne, E. Plourde, Bayesian stsa estimation using masking properties and generalized gamma prior for speech enhancement, EURASIP Journal on Advances in Signal Processing 2015 (1) (2015) 87. doi:10.1186/s13634-015-0270-6.
URL https://doi.org/10.1186/s13634-015-0270-6

[36] D. Foreman-Mackey, D. W. Hogg, D. Lang, J. Goodman, emcee: The MCMC Hammer, PASP125 (2013) 306. arXiv:1202.3665.

[37] B. E. Nelson, et al., An empirically derived three-dimensional Laplace resonance in the Gliese 876 planetary system, MNRAS455 (2016) 2484. arXiv:1504.07995.

[38] P. C. Guo, Orbital Characterization of Multi-Object Exoplanetary Systems with Radial Velocity Observation, Ph.D. thesis, University of Florida (2012).

[39] M. D. Weinberg, I. Yoon, N. Katz, A remarkably simple and accurate method for computing the Bayes Factor from a Markov chain Monte Carlo Simulation of the Posterior Distribution in high dimension, ArXiv e-printsarXiv:1301.3156.

[40] C. D. Murray, S. F. Dermott, Solar system dynamics, 1999.

[41] R. L. Akeson, et al., The NASA Exoplanet Archive: Data and Tools for Exoplanet Research, PASP125 (2013) 989. arXiv:1307.2944.

[42] D. Foreman-Mackey, B. E. Nelson, Private communication (2017).

[43] J. E. Chambers, F. Migliorini, Mercury - A New Software Package for Orbital Integrations, in: AAS/Division for Planetary Sciences Meeting Abstracts #29, Vol. 29 of Bulletin of the American Astronomical Society, 1997, p. 1024.

[44] J. C. Becker, et al., WASP-47: A Hot Jupiter System with Two Additional Planets Discovered by K2, ApJ812 (2015) L18. arXiv:1508.02411.

[45] A. McQuillan, T. Mazeh, S. Aigrain, Rotation Periods of 34,030 Kepler Main-sequence Stars: The Full Autocorrelation Sample, ApJS211 (2014) 24. arXiv:1402.5694.

[46] R. D. Haywood, Hide and Seek: Radial-Velocity Searches for Planets around Active Stars, Ph.D. thesis, University of St Andrews (2015).

[47] D. Del Moro, F. Berrilli, T. L. Duvall, Jr., A. G. Kosovichev, Dynamics and Structure of Supergranulation, Sol. Phys.221 (2004) 23.

[48] K. G. Strassmeier, Starspots, A&A Rev.17 (2009) 251.

[49] S. S. Vogt, M. Radovan, R. Kibrick, R. P. Butler, B. Alcott, S. Allen, P. Arriagada, M. Bolte, J. Burt, J. Cabak, K. Chloros, D. Cowley, W. Deich, B. Dupraw, W. Earthman, H. Epps, S. Faber, D. Fischer, E. Gates, D. Hilyard, B. Holden, K. Johnston, S. Keiser, D. Kanto, M. Katsuki, L. Laiterman, K. Lanclos, G. Laughlin, J. Lewis, C. Lockwood, P. Lynam, G. Marcy, M. McLean, J. Miller, T. Misch, M. Peck, T. Pfister, A. Phillips, E. Rivera, D. Sandford, M. Saylor, R. Stover, M. Thompson, B. Walp, J. Ward, J. Wareham, M. Wei, C. Wright, APF—The Lick Observatory Automated Planet Finder, PASP126 (2014) 359. arXiv:1402.6684, doi:10.1086/676120.

[50] R. D. Haywood, A. Collier Cameron, D. Queloz, S. C. C. Barros, M. Deleuil, R. Fares, M. Gillon, A. F. Lanza, C. Lovis, C. Moutou, F. Pepe, D. Pollacco, A. Santerne, D. Ségransan, Y. C. Unruh, Planets and stellar activity: hide and seek in the CoRoT-7 system, MNRAS443 (2014) 2517–2531. arXiv:1407.1044, doi:10.1093/mnras/stu1320.

[51] C. E. Rasmussen, K. I. Williams, Gaussian Processes for Machine Learning, 2006.

[52] S. K. Grunblatt, A. W. Howard, R. D. Haywood, Determining the Mass of Kepler-78b with Nonparametric Gaussian Process Estimation, ApJ808 (2015) 127. arXiv:1501.00369, doi:10.1088/0004-637X/808/2/127.

[53] S. Millholland, G. Laughlin, J. Teske, R. P. Butler, J. Burt, B. Holden, S. Vogt, J. Crane, S. Shectman, I. Thompson, New Constraints on Gliese 876—Exemplar of Mean-motion Resonance, AJ155 (2018) 106. arXiv:1801.07831, doi:10.3847/1538-3881/aaa894.

[54] B. E. Nelson, E. B. Ford, J. Buchner, R. Cloutier, R. F. Díaz, J. P. Faria, V. M. Rajpaul, S. Rukdee, Quantifying the Evidence for a Planet in Radial Velocity Data, ArXiv e-printsarXiv:1806.04683.

[55] S. Ambikasaran, D. Foreman-Mackey, L. Greengard, D. W. Hogg, M. O'Neil, Fast Direct Methods for Gaussian Processes.

[56] A. H. Vaughan, G. W. Preston, O. C. Wilson, Flux measurements of CA II H and K emission, PASP90 (1978) 267–274. doi:10.1086/130324.

[57] Astropy Collaboration, A. M. Price-Whelan, B. M. Sipőcz, H. M. Günther, P. L. Lim, S. M. Crawford, S. Conseil, D. L. Shupe, M. W. Craig, N. Dencheva, A. Ginsburg, J. T. VanderPlas, L. D. Bradley, D. Pérez-Suárez, M. de Val-Borro, T. L. Aldcroft, K. L. Cruz, T. P. Robitaille, E. J. Tollerud, C. Ardelean, T. Babej, Y. P. Bach, M. Bachetti, A. V. Bakanov, S. P. Bamford, G. Barentsen, P. Barmby, A. Baumbach, K. L. Berry, F. Biscani, M. Boquien, K. A. Bostroem, L. G. Bouma, G. B. Brammer, E. M. Bray, H. Breytenbach, H. Buddelmeijer, D. J. Burke, G. Calderone, J. L. Cano Rodríguez, M. Cara, J. V. M. Cardoso, S. Cheedella, Y. Copin, L. Corrales, D. Crichton, D. D'Avella, C. Deil, É. Depagne, J. P. Dietrich, A. Donath, M. Droettboom, N. Earl, T. Erben, S. Fabbro, L. A. Ferreira, T. Finethy, R. T. Fox, L. H. Garrison, S. L. J. Gibbons, D. A. Goldstein, R. Gommers, J. P. Greco, P. Greenfield, A. M. Groener, F. Grollier, A. Hagen, P. Hirst, D. Homeier, A. J. Horton, G. Hosseinzadeh, L. Hu, J. S. Hunkeler, Ž. Ivezić, A. Jain, T. Jenness, G. Kanarek, S. Kendrew, N. S. Kern, W. E. Kerzendorf, A. Khvalko, J. King, D. Kirkby, A. M. Kulkarni, A. Kumar, A. Lee, D. Lenz, S. P. Littlefair, Z. Ma, D. M. Macleod, M. Mastropietro, C. McCully, S. Montagnac, B. M. Morris, M. Mueller, S. J. Mumford, D. Muna, N. A. Murphy, S. Nelson, G. H. Nguyen, J. P. Ninan, M. Nöthe, S. Ogaz, S. Oh, J. K. Parejko, N. Parley, S. Pascual, R. Patil, A. A. Patil, A. L. Plunkett, J. X. Prochaska, T. Rastogi, V. Reddy Janga, J. Sabater, P. Sakurikar, M. Seifert, L. E. Sherbert, H. Sherwood-Taylor, A. Y. Shih, J. Sick, M. T. Silbiger, S. Singanamalla, L. P. Singer, P. H. Sladen, K. A. Sooley, S. Sornarajah, O. Streicher, P. Teuben, S. W. Thomas, G. R. Tremblay, J. E. H. Turner, V. Terrón, M. H. van Kerkwijk, A. de la Vega, L. L. Watkins, B. A. Weaver, J. B. Whitmore, J. Woillez, V. Zabalza, Astropy Contributors, The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package, AJ156 (2018) 123. arXiv:1801.02634, doi:10.3847/1538-3881/aabc4f.

[58] W. H. Bragg, W. L. Bragg, The Reflection of X-rays by Crystals, Proceedings of the Royal Society of London Series A 88 (1913) 428–438. doi:10.1098/rspa.1913.0040.

[59] F. Birch, Finite Elastic Strain of Cubic Crystals, Physical Review 71 (1947) 809–824. doi:10.1103/PhysRev.71.809.

[60] P. Vinet, J. R. Smith, J. Ferrante, J. H. Rose, Temperature effects on the universal equation of state of solids, Phys. Rev. B 35 (1987) 1945–1953. doi:10.1103/PhysRevB.35.1945.
URL https://link.aps.org/doi/10.1103/PhysRevB.35.1945

[61] P. Loubeyre, R. Letoullec, E. Wolanin, M. Hanfland, D. Hausermann, Modulated phases and proton centring in ice observed by X-ray diffraction up to 170GPa, Nature397 (1999) 503–506. doi:10.1038/17300.

[62] R. J. Hemley, A. P. Jephcoat, H. K. Mao, C. S. Zha, L. W. Finger, D. E. Cox, Static compression of H_2O-ice to 128 GPa (1.28 Mbar), Nature330 (1987) 737–740. doi:10.1038/330737a0.

[63] M. R. Frank, Y. Fei, J. Hu, Constraining the equation of state of fluid H $_2$O to 80 GPa using the melting curve, bulk modulus, and thermal expansivity of Ice VII, Geochim. Cosmochim. Acta68 (2004) 2781–2790. doi:10.1016/j.gca.2003.12.007.

[64] L. Bezacier, B. Journaux, J.-P. Perrillat, H. Cardon, M. Hanfland, I. Daniel, Equations of state of ice VI and ice VII at high pressure and high temperature, J. Chem. Phys.141 (10) (2014) 104505. doi:10.1063/1.4894421.

Curriculum Vitae

John Henry Boisvert

E-mail: johnhboisvert@gmail.com
https://www.linkedin.com/in/johnhboisvert

Education

August 2013 - December 2018 Ph.D., Astronomy *University of Nevada, Las Vegas*
Advisor: *Jason H. Steffen, Ph.D.*

January 2011 - August 2013 M.S., Physics *University of Nevada, Las Vegas*
Advisor: *George Rhee, Ph.D.*

August 2005 - December 2010 B.S., Physics *University of Nevada, Las Vegas*
Advisor: *Len Zane, Ph.D.*

Research interests

- Model comparison within a Bayesian framework
- Exoplanet detection and characterization
- Architecture identification of exoplanetary systems, especially near the 2:1 degeneracy
- Observational methods to break degeneracies between competing planetary models
- Exoplanet characterizations at or near their habitable zones
- Exoplanet dynamics and their relationship to the mean motion resonances

Publications

1. **Boisvert, J. H.**; Nelson, B. E; and Steffen, J. H., *Systematic mischaracterization of exoplanetary system dynamical histories from a model degeneracy near mean-motion resonance*, Monthly Notices of the Royal Astronomical Society, **480**, 2846 (2018)
2. Grande, Z. M.; Huang, C.; Smith, D.; Smith, J. S.; **Boisvert, J. H.**, Tschauner, O.; Steffen, J. H.; Salamat, A. *Bond strengthening in dense H_2O influences Earth's water-cycle and water-rich planets*, (submitted, 2018)
3. **Boisvert, J. H.** and Rhee, G., *Finding the center: An analysis of the tilted ring model fits to the inner and outer parts of six dwarf galaxies*, The Astronomical Journal, **152**, 15 (2016)
4. Li, Y.; Yuan, W.; Zhou, H. Y.; Komossa, S.; Ai, Y. L.; Liu, W. J.; and **Boisvert, J. H.**, *An Unobscured Type II Quasar Candidate: SDSS J012032.19-005501.9*, The Astronomical Journal, **149**, 75 (2015)

Skills

- Software development of data analysis algorithms
- Programming languages: Python, Cython, Fortran and C
- Bayesian analysis, running MCMCs, measuring the fully marginalized likelihood function
- Operation of various telescopes, e.g. 16-in Schmidt-Cassegrain at UNLV, Celestron, Meade
- Telescope collimation and alignment
- CCD camera calibration and operation
- Development and instruction of lectures and laboratory exercises in astronomy, physics, and mathematics
- Miscellaneous software: SAOImage ds9, Canvas, LaTeX, UNIX, and Microsoft Office
- Web-cam operation and streaming to various social media

Honors and awards

- Sagan Summer Workshop Travel Award (2016)
- UNLV Graduate and Professional Student Association Travel Grant (2014)
- UNLV Foundation Bigelow Travel Grant (2014)
- Nevada Millennium Scholarship (2005)
- UNLV Undergraduate Dean's Honor List (2009-2010)

Conferences

- 231^{st} Meeting of the American Astronomical Society 2018

  Contributed talk: *"System Architectures Near the 2:1 Resonance "*

- Aspen Winter Conference: Formation and Dynamical Evolution of Exoplanets 2017

  Contributed talk: *"Uncovering System Architectures Near the 2:1 Resonance"*

- 17^{th} Sagan Summer Workshop: Is There a Planet in My Data? Statistical Approaches to Finding and Characterizing Planets in Astronomical Data 2016

  Poster: *"Uncovering System Architectures Near the 2:1 Resonance"*
  Workshop Talk: *"Effect of Planetary Amplitude on Planet Detectability"*

- CGE and HiPACC The Near-Field Deep-Field Connection 2014

  Poster: *"How Accurately Can the Inclination Angle, Position Angle, and Location of the Dynamical Center be Measured from the Neutral Hydrogen Disk in the Central Regions of Dwarf Galaxies?"*

Teaching

- Adjunct Professor, College of Southern Nevada 2018 – Present

  AST 105: *"Introductory Astronomy Laboratory"*

- Substitute Lecturer, College of Southern Nevada Spring 2018

  AST 103: *"Solar System,"* AST 104: *"Stars & Galaxies,"* and AST 105: *"Introductory Astronomy Laboratory"*

Assistant teaching

- Graduate Teaching Assistant, UNLV 2011 – 2016, 2018

  AST 105: *"Introduction to Observational Astronomy"* Including Summers, 2013 – 2016, 2018

- Assistant to Lab Instructor, UNLV 2010

  AST 105: *"Introduction to Observational Astronomy"*

Outreach

- Telescope Coordinator, HOA Star Party, Mojave Origins Institute (UNLV), Las Vegas, NV October, November 2018
- Telescope Coordinator, Total Solar Eclipse, UNLV, Las Vegas, NV 2017
- Telescope Coordinator, Transit of Mercury, Cox Elementary School, Henderson, NV 2016
- AstroCoffee Coordinator, UNLV, Las Vegas, NV 2015-2016
- Astronomy Journal Club Coordinator, UNLV, Las Vegas, NV 2015-2016
- Science Fair Judge, Southern Regional Science & Engineering Fair, Las Vegas, NV 2013
- Telescope Coordinator, Transit of Venus, UNLV, Las Vegas, NV 2012
- Vice President, Society of Physics Students, UNLV chapter 2010
- Secretary, Society of Physics Students, UNLV chapter 2009

Professional references

1. Jason H. Steffen, Ph.D., M.S., Assistant Professor of Astrophysics, UNLV

   E-mail: jason.steffen@unlv.edu
   Phone: +1 702/895-3485
   Website: https://www.physics.unlv.edu/~jsteffen/

2. Benjamin E. Nelson, Ph.D., M.S., Data Science Fellow, Insight Data Science

   E-mail: mailto:benelson7@gmail.com
   Phone: +1 415/246-7861
   Website: https://www.linkedin.com/in/benelson7

3. Stephanie Fiorenza, Ph.D., Professor of Physics and Astronomy, CSN

   E-mail: stephanie.fiorenza@csn.edu
   Phone: +1 702/651-4456
   Website: https://www.csn.edu/directory/stephanie-fiorenza

4. Ashkan Salamat, Ph.D., M.S., Assistant Professor of Physics, UNLV

   E-mail: ashkan.salamat@unlv.edu
   Phone: +1 702/895-1716
   Website: https://scholar.harvard.edu/salamat/

5. Bing Zhang, Ph.D., Professor of Astrophysics, UNLV

   E-mail: zhang@physics.unlv.edu
   Phone: +1 702/895-4050

Website: http://www.physics.unlv.edu/~bzhang/

6. Zhaohuan Zhu, Ph.D., Assistant Professor of Astrophysics, UNLV

   E-mail: zhaohuan.zhu@unlv.edu
   Phone: +1 702/895-3755
   Website: http://www.physics.unlv.edu/~zhzhu/