

EXOSYSTEM MODELING FOR
MISSION SIMULATION AND SURVEY ANALYSIS

DMITRY SAVRANSKY

A DISSERTATION
PRESENTED TO THE FACULTY
OF PRINCETON UNIVERSITY
IN CANDIDACY FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE
BY THE DEPARTMENT OF
MECHANICAL AND AEROSPACE ENGINEERING
ADVISER: N. JEREMY KASDIN

SEPTEMBER 2011

© Copyright by Dmitry Savransky, 2019.

All Rights Reserved

Abstract

In the last twenty years, the existence of exoplanets (planets orbiting stars other than our own sun) has gone from conjecture to established fact. The accelerating rate of exoplanet discovery has generated a wealth of important new knowledge, and is due mainly to the development and maturation of a large number of technologies that drive a variety of planet detection and observation methods. The overall goal of the exoplanet community is to study planets around all types of stars, and across all ranges of planetary mass and orbit size. With this capability we will be able to build confidence in planet formation and evolution theories and learn how our solar system came to exist.

Achieving this goal requires creating dedicated instrumentation capable of detecting signals that are a small fraction of the magnitude of signals we can observe today. It also requires analyzing highly noisy data sets for the faint patterns that represent the presence of planets. Accurate modeling and simulation are necessary for both these tasks. With detailed planetary and observation models we can predict the type of data that will be generated when a specific instrument observes a specific planetary system. This allows us to evaluate the performance of both the instrument and the data analysis methods used to extract planet signals from observational data. The same simulations can help optimize observation scheduling and statistical analysis of data sets. The purpose of this thesis is to lay down the groundwork necessary for building simulations of this type, and to demonstrate a few of their many possible applications.

First, we show how each of four different detection methods (astrometry, doppler spectroscopy, transit photometry and direct imaging) can be described using a common parameter set which also encodes sufficient information to propagate the described exosystem in time. We analyze this parameter set and derive the distribution functions of several of its elements. These are shown to be useful in calculating the probabilities of planetary detection. Using this capability, we create a framework for simulating whole direct imaging planet-finding missions, incorporating detailed instrument models, observatory operations,

and an automated algorithm for observation scheduling. This framework is used in a series of case studies to evaluate the capabilities of multiple proposed missions. Finally, we show how the same modeling framework used to generate the mission simulations can also be used, with the formalism of dynamic filtering, for data analysis and data set synthesis.

Acknowledgements

I owe a great deal to everyone who has influenced and supported me—without you this thesis wouldn’t exist. My parents, in particular, have always served as an example of who I want to be. Thank you for bringing me to this country, for always encouraging me, and for your overall overachievement that serves as an ever-moving goal to shoot for. The rest of my family have also been good examples: my grandparents embodying quiet intelligence, iron will, artistic sensibility, and pure kindness; my aunt and uncle displaying a great talent for élan. Thank you for raising and teaching me. To my cousins, who are as close as siblings: than you for being a constant source of surprise and pride. Many thanks to all of my friends, who ensure that life is never dull.

I have also had the rare privilege of working with two of the finest researchers that I have ever encountered. Jim Bell and Jeremy Kasdin both share a clear love and enthusiasm for their chosen fields and a drive to expand and share their knowledge. I cannot thank you enough for your guidance and support.

Those portions of this text derived from published papers benefited greatly from the input of my various co-authors, editors, and collaborators; any remaining mistakes are solely my own. Jeremy Kasdin deserves a medal for reading every draft of every paper I have written in the last five years. Bob Vanderbei’s comments and suggestions have vastly improved several of my papers and Eric Cady’s erudite knowledge of arcane integrals allowed us to find some of the most interesting results presented here. Stephanie Goldfarb’s basic algorithm for generating reversion coefficients allowed me to write the code used to prove those results. Tyler Groff deserves many thanks for patiently explaining optics to me (over and over again), and not over-using a certain file folder. My readers, Ed Turner and Michael Littman, provided invaluable input into both this thesis and my research in general.

Many thanks go to Jessica O’Leary, who made navigating graduate school infinitely easier, and who doesn’t spend nearly as much time on Facebook as I tell people. Since her departure, Candy Reed and Jill Ray have done an excellent job in assuming both her

official duties, and maintaining the community feeling that makes the MAE department such a great place to work. I also owe much to Donald Knuth, Leslie Lamport, and the entire online L^AT_EXcommunity, who have freed me from the tyranny of Word forever.

The only remaining thanks are the most difficult, as they go to the one person whose contribution is incalculable. Melanie, I dedicate this thesis to you. Without you, life would be nowhere near as sweet.

This dissertation carries the number T-3230 in the records of the Department of Mechanical and Aerospace Engineering.

Come, my friends,
'Tis not too late to seek a newer world.
— Alfred, Lord Tennyson, *Ulysses*

We doctors know
a hopeless case if — listen: there's a hell
of a good universe next door; let's go
— E. E. Cummings,
Pity This Busy Monster, Manunkind

И слушает всяк, кто достоен,
Как чудо над миром парит:
Как тихо звезда со звездою
Во мраке ночном говорит.
— Марлена Рахлина,
„*В своей одинокой квартире...*”

Contents

Abstract	iii
Acknowledgements	v
List of Tables	xi
List of Figures	xii
List of Symbols	xviii
Notation	xix
1 Introduction	1
1.1 Searching for Exoplanets	1
1.2 Exosystem Parameters	5
1.3 The Point Spread Function	17
2 Observation Methods	21
2.1 Direct Detection	21
2.1.1 Imaging Observables and Constraints	23
2.1.2 Observation Modeling	31
2.2 Astrometry and Doppler Spectroscopy	38
2.2.1 Astrometric Observables and Constraints	39
2.2.2 Doppler Spectroscopy Observables and Constraints	46
2.3 Transit Photometry	48

3 Parameter Distributions and Detection Probabilities	50
3.1 Current State of Knowledge	50
3.1.1 Biases	54
3.2 Observational Completeness	56
3.2.1 Applications	58
3.3 Analytical Distributions	62
3.3.1 Derivation	64
3.3.2 Validation	72
3.3.3 Applications	77
4 Observation Simulation	83
4.1 Mission Planning and Analysis	84
4.2 Generating and Propagating Planetary Systems	85
4.2.1 System Generation	85
4.2.2 System Propagation	89
4.3 Target Selection and Availability	99
4.4 Decision Modeling	104
4.5 Simulation Methodology	115
4.6 Case Studies	119
4.6.1 Validation	125
4.6.2 Results	133
5 Integration of Data Streams	156
5.1 Exosystems as Hidden Markov Models	156
5.2 Initializing the State	169
5.2.1 Selecting the State Vector	169
5.2.2 Incorporating Prior Knowledge	172
5.3 Updating the State	176

5.3.1	State Dynamics	176
5.3.2	Observation Equations	178
5.3.3	Astrometric Observation Function Linearization	180
5.4	Applications	194
5.4.1	Combined Astrometry and Doppler Spectroscopy	195
5.4.2	Filter Applications for Direct Detection	198
6	Conclusions and Future Work	208
Bibliography		210
A	Code Listings	227
vec2orbElem		227
calcRevCoefs		230
invKepler		232
keplerSTM		233
nbodyVect		236
genCostFunTerm1		237
hybridEKF		238
hybridUKF		240
LombScargle		242
B	Mission Timeline Encoding	243
C	Default Simulation Settings	245

List of Tables

2.1	Zero Magnitude Flux Points	34
4.1	Occulter Designs.	124
4.2	O ₃ filters.	125
5.1	Typical values of terms in astrometric observation equation	182

List of Figures

1.1	History of exoplanet discovery	2
1.2	Inertial barycentric/ecliptic frame	6
1.3	Anomaly Diagram	9
1.4	Eccentricity variation	15
2.1	Starshade	22
2.2	Orbit Diagram	24
2.3	Imaging constraints diagram	27
2.4	Reflection Diagram	28
2.5	Solar system planet phase functions	30
2.6	Integration time vs. PSF sampling	35
2.7	Historical distribution of solar system zodi	37
2.8	Astrometric observation schematic	39
2.9	Narrow angle astrometry schematic	43
2.10	Transit schematic	48
3.1	$m_P \sin I - P_{orb}$ Data	51
3.2	$m_P \sin I - P_{orb}$ power-law fits	52
3.3	Rocky planet densities	53
3.4	Planet occurrence vs metallicity	54
3.5	Orbital parameter distributions of currently known exoplanets	55

3.6	Earth-twin observation PDF	57
3.7	Earth-twin observation CDF	58
3.8	Planet detection probability over multiple visits	60
3.9	Undersampled observation PDF	63
3.10	Validation of analytical ν PDF	74
3.11	Validation of analytical r PDFs	75
3.12	$f_{\bar{\beta}}$ Q-Q plots	76
3.13	Validation of analytical m PDF.	77
3.14	Validation of analytical s PDF.	78
3.15	Maximum likelihood for a given s	80
3.16	Portion of observable transits detected over one period.	81
4.1	Sample Multi-body exosystem	87
4.2	Integrator Comparison	97
4.3	keep-out Zones	104
4.4	Cost Function Matrix	112
4.5	Revisit timing strategies	113
4.6	Simulation Flowchart	116
4.7	Sample auto-generated DRM	119
4.8	DRMconstructor	120
4.9	Halo orbit	121
4.10	Throughput Curves	126
4.11	Science yield from automated and randomized visit order	128
4.12	THEIA occulter propellant use	129
4.13	Science yield metric statistics as a function of ensemble size	130
4.14	Validation of Bayesian unique planet classification	131
4.15	4 m telescope comparison	134
4.16	8 m telescope comparison	135

4.17	16 m telescope comparison	136
4.18	MDO comparison	138
4.19	Coronagraph comparison	140
4.20	Partial spectra	141
4.21	Total observation times	142
4.22	Orbital fits	143
4.23	First detection histograms	144
4.24	4 m telescope comparison for super-Earths	145
4.25	Effects of varying target pools	146
4.26	O_3 simulation results	147
4.27	Effects of varying burn portion	148
4.28	Occulter propellant use	149
4.29	THEIA Extended Mission	150
4.30	Integration times for varying exo-zodi levels	151
4.31	Spectral characterizations at varying exo-zodi levels	152
4.32	Effects of varying exo-zodi levels	153
4.33	Effects of varying false positive rates.	155
5.1	Solar system dynamic state covariance estimate	173
5.2	Astrometric wobble due to Earth mass planet	181
5.3	First order expansion of astrometric observation	184
5.4	First order expansion of astrometric observation without radial velocity terms	185
5.5	Second order expansion of astrometric observation equation	187
5.6	Second order expansion of astrometric observation equation assuming prior knowledge	191
5.7	Numerical encoding precision test	193
5.8	Two Jupiter system	202
5.9	Two Jupiter system periods	202

5.10	Astrometric observation noise	203
5.11	Filtered results for Earth mass planet	203
5.12	Filter estimated error for Earth mass planet	204
5.13	Mean filter variance for increasing planet mass error	204
5.14	Filter estimated error for a Jupiter mass planet	205
5.15	Filter output for RV	205
5.16	Filter output for RV and imaging	206
5.17	Filter and smoother output for RV and imaging	206
5.18	State variance estimates for confused data set	207

List of Symbols

S	Star position	5
sc	Observer position	5
\mathcal{I}	Inertial barycentric/ecliptic solar system frame.	5
O	Solar system barycenter	5
d	Star distance with respect to solar system barycenter	6
P	Planet position	6
G	Exosystem barycenter	6
r	Orbital radius	7
p	Geometric albedo	7
R	Planet radius	7
F_p	Planet flux	7
F_s	Star flux	7
L	Stellar luminosity	7
a	Orbital semi-major axis	8
e	Orbital eccentricity	8
ψ, θ, ϕ	Euler angles orienting orbit in space	8
M	Mean anomaly	8
E	Eccentric anomaly	8
ν	True anomaly	8
μ_P	Gravitational parameter of planet	9

μ_S	Gravitational parameter of star	9
m_S	Mass of star	9
m_P	Mass of planet	9
P_{orb}	Orbital period	10
t	Time	10
Ω	Longitude of the ascending node	12
I	Inclination	12
ω	Argument of periapsis	12
c	Speed of light	16
PSF	Point spread function	17
λ	Wavelength	18
\mathcal{S}	Sky frame	23
\mathbf{s}	Sky star-planet vector	23
s	Apparent separation	24
F_R	Ratio of planet to star fluxes	25
Φ	Planet phase function	25
β	Phase angle	25
Δmag	Difference in magnitude between star and planet	25
IWA	Inner working angle	25
Δmag_0	Limiting Δmag	26
d_s	Occulter separation distance	27
C_p	Mean photon count at pixel centered on planet PSF	32
C_b	Mean background photon count	32
Q	Ratio of planet to background photon counts	33
FAP	False Alarm Probability	33
MDP	Missed Detection Probability	33
Ψ	Sharpness	33

QE	Quantum efficiency	34
M_V	Absolute magnitude (V band)	36
μ	Brightness of exo-zodi	36
S/N	Signal to noise ratio	38
\mathcal{B}	Tangent frame	40
ϖ	Parallax	40
\mathbf{r}_μ	Exosystem barycenter motion vector	41
V_R	Radial velocity	41
\mathbf{V}_T	Transverse velocity	41
OPD	Optical path delay	42
\mathbf{B}	Observatory orientation vector	42
$F_S^{(e)}$	Obscured stellar flux	49
R_S	Star radius	49
I_{sp}	Specific Impulse	107
SDO	Single-Distance Occulter	122
MDO	Multiple-Distance Occulter	122

Notation

Unless otherwise noted, the mathematical notation will use the following conventions:

Scalar, function, or set element (when indexed)	x or x_i	lowercase regular
Random variable	\bar{x}	overbar lowercase regular
Vector	\mathbf{x}	lowercase bold
Space or frame reference point	X	uppercase regular
Matrix	\mathbf{X}	uppercase bold
Reference frame	\mathcal{I}	calligraphic
Enumerated set	$\{x\}_{i=1}^n$	

Frames will be defined by an origin and three orthogonal unit vectors and will be dextral unless otherwise noted:

$$\mathcal{I} = (O, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$$

$$\mathbf{e}_3 = \mathbf{e}_1 \times \mathbf{e}_2 \quad \mathbf{e}_2 = \mathbf{e}_3 \times \mathbf{e}_1 \quad -\mathbf{e}_1 = \mathbf{e}_3 \times \mathbf{e}_2.$$

Vectors describing the positions of point p with respect to reference point O will be written as $\mathbf{r}_{p/O}$ and as $[\mathbf{r}_{p/O}]_{\mathcal{I}}$ when expressed in the coordinates of frame \mathcal{I} .

Vector derivatives in frame \mathcal{I} will be written as:

$$\frac{^{\mathcal{I}}d}{dt} \mathbf{r}_{p/O} = {}^{\mathcal{I}}\mathbf{v}_{p/O}.$$

Unit vectors will be written as:

$$\hat{\mathbf{x}} \triangleq \frac{\mathbf{x}}{\|\mathbf{x}\|} \triangleq \langle \mathbf{x} \rangle$$

where $\langle \cdot \rangle$ represents the normalization operator.

Matrix transpose will be represented via $[...]^T$:

$$\begin{bmatrix} a & b & c \end{bmatrix}^T \equiv \begin{bmatrix} a \\ b \\ c \end{bmatrix}.$$

\mathbf{I}_n will represent the $n \times n$ identity matrix and $\mathbf{0}_{m \times n}$ will represent the $m \times n$ matrix of zeros.

The Fourier transform of a complex function will be represented via $\hat{\cdot}$:

$$\mathcal{F}\{g(x, y)\} \equiv \hat{g}(u, v) \triangleq \iint_{-\infty}^{\infty} g(x, y) e^{-i2\pi(ux+vy)} dx dy.$$

The symbol \otimes will denote convolution:

$$f \otimes g \triangleq \int_{-\infty}^{\infty} f(\tau) g(t - \tau) d\tau.$$

The Binomial coefficient will be written as:

$$\binom{n}{k} \equiv \prod_{i=1}^k \frac{n - (k - i)}{i}, \quad k \in \mathbb{N}.$$

Probabilities will be denoted via P , with conditional probabilities written as $P[X|Y]$ and joint probabilities as $P[X, Y]$ so that:

$$P[X, Y] = P[X|Y]P[Y].$$

Probability density functions will be written as $f_{\bar{x}}(x)$ and expectation values will be denoted by the symbol E as:

$$E[x] = \int_{-\infty}^{\infty} x f_{\bar{x}}(x) dx.$$

Chapter 1

Introduction

1.1 Searching for Exoplanets

The idea that there could be planets orbiting stars other than our own sun has been around for at least four hundred years, with historical proponents ranging from Giordano Bruno to Isaac Newton. It was not until 1988, however, that the first confirmed discovery of an extra-solar planet (or exoplanet) was made Campbell et al. [17], and not until 1995 that a planet was discovered about a star similar to the sun Mayor and Queloz [89]. In the fifteen years since then, new planets have been discovered at an ever accelerating rate, as evidenced by Figure 1.1.

Our ability to find large numbers of exoplanets is the key to our eventual understanding of planet formation and evolution. To understand why, we need only look at the history of astrophysics and our knowledge of the formation and evolution of stars and galaxies. Stellar and galactic evolution take place on time scales so vast that not even multiple generations of astronomers could hope to see many changes in the stars in our sky. However, despite having only a static view of these dynamic processes, we are lucky enough to have an incredibly large sample of observable objects, all at various stages of development. This large data set, along with the convenient fact that looking further into the universe is equivalent to looking

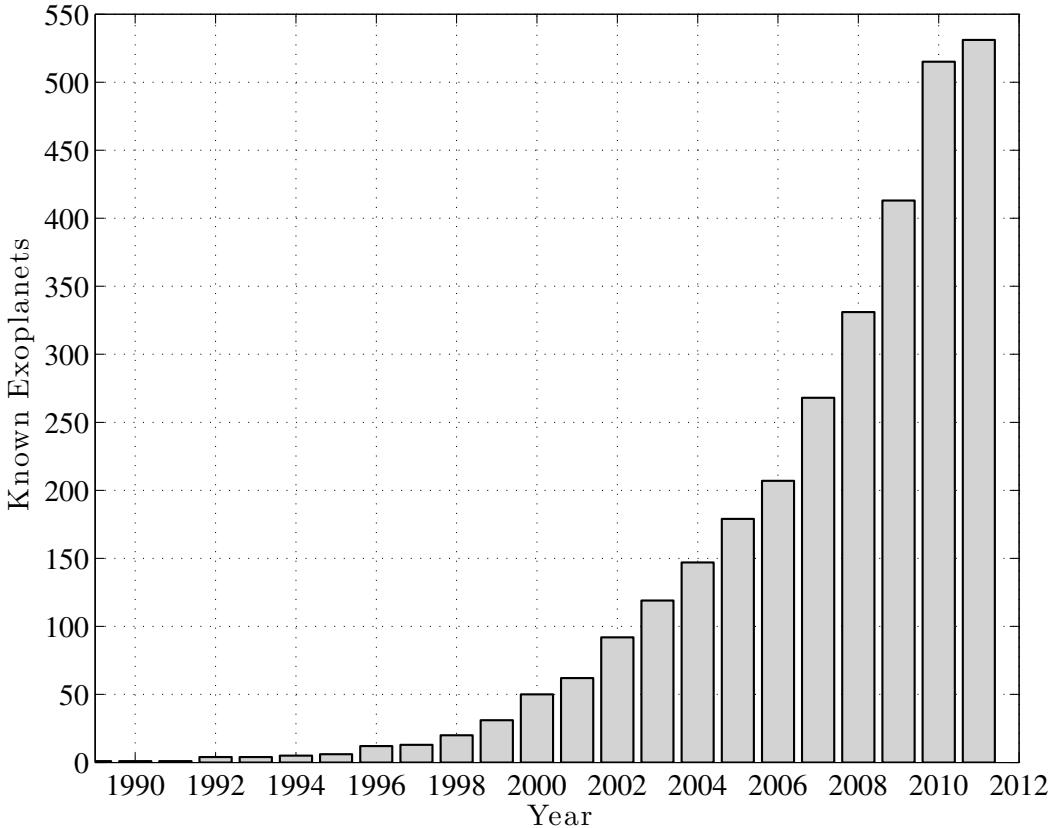


Figure 1.1: Number of confirmed exoplanets as a function of year. There are currently 531 known exoplanets, and over 1200 planetary candidates. Data retrieved from the NASA/I-PAC/NExSci Star and Exoplanet Database <http://nsted.ipac.caltech.edu> on April 17th, 2011.

further back in time, has allowed us to deduce the processes by which stars are formed, evolve, and eventually transform into a variety of other interesting objects. We've gained confidence in these models based on our ability to locate stars at each stage of the process. With a sufficiently large pool of observed exoplanets, we can expect the same outcome for planet science.

Unfortunately, observing exoplanets is significantly harder than observing stars. There are no examples of bright exoplanets that can be seen with the naked eye or simple instrumentation, and so we must find each exoplanet before studying it. In many ways, this searching for exoplanets is closer to particle physics than to traditional observational astronomy. In each case, we search for something that may not be there, guided only by constantly

evolving theory. Similarly, both new planets and new particles are often detected indirectly, based on their effects on more easily observed objects. Finally, both pursuits require complex dedicated equipment, which often represents the extreme borders of current technology and large investments of resources. To answer these challenges, the exoplanet community has developed new instrumentation and data analysis methods and even whole new methods for detecting the presence of exoplanets, leading to the accelerating rate of discoveries described above.

Alongside the various planet-finding surveys, there has been a concurrent effort to develop detailed models and simulations of observatory operations and planet observations. These simulations serve three main purposes: first, they allow the community to directly compare the expected scientific returns of proposed observatory and mission concepts. Second, they can be used to optimize target selection and observation scheduling for existing facilities, while providing a controlled testing environment for data processing pipelines. Finally, the ability to calculate the probability that a planet of a given type will be observed with a given frequency, conditional on a planet formation model, means that observations can be used to choose between competing planet evolution theories and closely estimate absolute planet frequencies.

These applications have already been proven to be helpful in both the mission planning and the data analysis. For example, the Kepler Mission, which has more than doubled the number of exoplanet candidates since its launch in 2009, utilizes the Kepler end-to-end model—an integrated simulation used to test the data processing pipeline and simulate observatory operations [12]. Similarly, there have been many statistical analyses of the existing exoplanet data set that relied heavily on observation modeling, including the analysis of radial velocity surveys in Cumming et al. [26] and a microlensing study described in Gould et al. [43].

This thesis will explore exoplanet observation modeling in relation to all of these applications. We will start by formulating a standardized set of parameters for use throughout the

rest of the chapters. The remainder of this chapter will introduce the variables required to describe how planets orbit their parent (or host) stars and how their position and motion in space can be defined with respect to a specific observer. We will also detail how light from stars and planets interacts with imaging systems to produce the images that form the raw data for many planet-finding methods.

This description will be continued in Chapter 2 with a detailed discussion of four methods used to study exoplanets. These are: doppler spectroscopy, which produced the majority of exoplanet detections before 2010, precision astrometry, which has the potential of allowing us to detect planets much smaller than those currently being found, transit photometry, which, thanks to the Kepler mission, has more than doubled the number of exoplanet candidates, and finally direct detection, which promises to become a very powerful tool in the near future, and has already produced a number of fascinating discoveries. The first three of these methods, doppler spectroscopy, astrometry, and transit photometry, are indirect—they rely on observations of the star in order to infer the presence of planets, tracking the star’s motion in the case of astrometry and doppler spectroscopy, and the amount of light coming from the target system in the case of transit photometry. Direct detection is the only method that attempts to image the exoplanets themselves. Each of these methods has specific strengths and limitations, all of which will be discussed in detail.

Chapter 3 will review what we currently know about exoplanets, and introduce the idea of a statistical approach to exoplanet study, focusing on properties of populations of planets rather than specific examples. Based on the formalism laid out in the first two chapters, we will develop a statistical description of planetary orbits, which will be applied to the optimization of planet-finding surveys. In the course of this discussion we will introduce the concept of completeness, which allows for a systematic way of incorporating instrument biases into the analysis of generated data.

Chapter 4 will discuss the various components required to build complete simulations of exoplanet observations. Starting with the generation of sample exoplanetary systems, we

will track each step of how a realistic data set can be generated for each observation method. We will also explore how to propagate systems forward in time so that we can simulate entire surveys in addition to individual observations. These techniques will be applied to the construction of an integrated simulation capability for direct detection planet-finders, which will be employed in a series of case studies to explore the performance of multiple proposed missions.

Finally, Chapter 5 will consider the data analysis problem, and will demonstrate how the formalism built up throughout the thesis, along with the well-developed field of optimal estimation, can be used to systematically generate complex planetary system models from multiple varied data streams. Taking into account the accelerating rate of discovery and the proliferation of observation methods, it is highly likely that these are the kinds of data sets the exoplanet community will have at its disposal in the near future, and the applications described in this chapter are a first step towards taking advantage of the increased information that can be extracted thanks to detailed modeling and simulation.

1.2 Exosystem Parameters

Before proceeding, it is important to define a consistent description of all of the physical quantities that will be treated in later chapters. This section defines every variable that will be used in multiple contexts so as to preserve a uniform notation throughout.

The position of an exosystem primary star (or ‘target’ star) is denoted by S , and, when expressed with respect to some specific point P , is given by the vector $\mathbf{r}_{S/P}$. The observer location is denoted by point sc , which may represent a space or ground-based observatory. We sometimes find it useful to define the star position via a celestial coordinate system. Typically, we use a barycentric/ecliptic frame $\mathcal{I} = (O, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ whose origin O is at the solar system barycenter (see Figure 1.2). In this frame, coordinates (λ, β) along with the

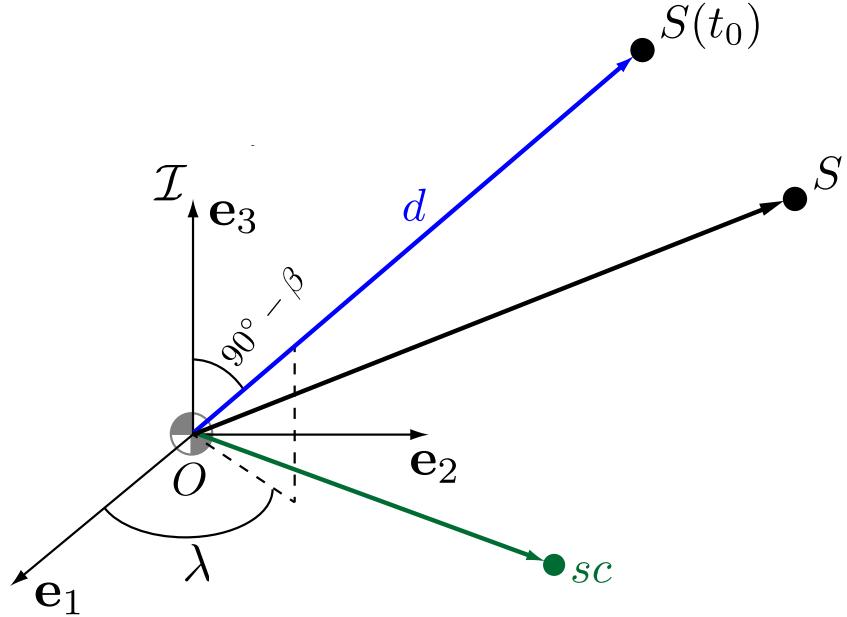


Figure 1.2: The inertial barycentric/ecliptic frame \mathcal{I} with origin O at the solar system barycenter. Star S is at coordinates λ, β and distance d at reference epoch t_0 .

target star distance d define the star position as:

$$[\mathbf{r}_{S/O}]_{\mathcal{I}} = d \begin{bmatrix} \cos \lambda \cos \beta \\ \sin \lambda \cos \beta \\ \sin \beta \end{bmatrix}, \quad (1.2.1)$$

where

$$d \triangleq \|\mathbf{r}_{S/O}\|. \quad (1.2.2)$$

It is important to note that (λ, β) and d are typically catalogued values, and thus define the star position at a specific epoch (t_0), whereas the star-pointing vector with respect to the solar system barycenter and the observer is constantly changing.

Similarly, the position of a planet is given by point P and the exosystem barycenter is defined as G . In cases with multiple planets in one exosystem, the planet positions are

enumerated as P_i . The star-planet separation vector is given by:

$$\mathbf{r}_{P/S} \triangleq \mathbf{r}_{P/G} - \mathbf{r}_{S/G}. \quad (1.2.3)$$

The time varying orbital radius, r is equal to the magnitude of the star-planet vector:

$$r \triangleq \|\mathbf{r}_{P/S}\|. \quad (1.2.4)$$

The planet's geometric albedo and radius are given by p and R , respectively, and the total flux of the planet and star are given by F_p and F_s , respectively. The intrinsic luminosity of the star is given by L measured in L_\odot (solar luminosity).

It is frequently useful to describe planetary orbits via Keplerian orbital elements. These are derived from the solution of the two-body problem, and are an exact description of planetary motion in the case of a single-planet exosystem. In systems with multiple planets, the orbits may still be approximated as Keplerian, although this ignores all planet-planet interactions and therefore cannot exactly predict future positions of all of the system bodies. This can still be a useful approximation, however, as no analytic solution exists to the general n-body problem. As we will see, there are actually two different ways of approximating n-body systems via Keplerian orbital elements: we can either fit an average ellipse to multiple orbits worth of data, producing an average set of Keplerian elements, or we can calculate time-varying Keplerian elements to reflect the perturbations due to the other bodies in the system. In the latter case, we typically use osculating Keplerian elements—i.e., the orbital elements of the Keplerian orbit tangent to the true orbit at that point in time.

The difference between constant, osculating, and averaged Keplerian orbital elements is very important as they describe very different things. With a fixed Keplerian orbit, you can predict the position and velocity of the planet for all time. If there are no perturbations on the planet, this position and velocity will be correct. With perturbations, however, the true position and velocity will progressively deviate from the ones predicted by the Keplerian

orbit. An averaged set of orbital elements will never exactly match the planet’s true position and velocity, but will minimize the position error over an entire data set. Osculating orbital elements will always give you the correct position and velocity, but require knowledge of the perturbations acting on the planet.

Furthermore, while a constant set of Keplerian elements describes a particular ellipse corresponding to the planet’s orbit, osculating elements are much more difficult to interpret, as they only correspond to the instantaneously tangential orbit. This means that in certain cases the osculating orbit will be highly eccentric while the planet’s orbit is actually nearly circular*. Throughout this thesis, we will employ both types of orbital elements. When dealing with single planet system simulations, as in §4.2.2, we will employ constant Keplerian orbital elements. When using orbital elements to talk about multi-planet systems, as we do later in this section, we will use osculating orbital elements. Finally, when discussing how certain detection methods that rely on finding periodic signals to detect planets, we will employ averaged orbital elements (see Chapter 2 and §5.2.2).

We define a Keplerian orbit via the parameter set $(a, e, \psi, \theta, \phi)$ where a is the semi-major axis, e is the eccentricity, and ψ, θ, ϕ are Euler angles determining the orientation of the orbit in the observer’s reference frame†. The position of the planet on its orbit at the time of observation is given by either the mean anomaly M , eccentric anomaly E , or the true anomaly ν . In a 2-body system, the planet’s orbit lies in a single plane, and thus can be fully described using only scalar values, as shown in Figure 1.3. Keplerian orbits are conic sections—closed planetary orbits are ellipses with the star located at one focus. The orbital radius is defined in terms of the eccentricity, semi-major axis, and true anomaly as:

$$r = \frac{a(1 - e^2)}{e \cos(\nu) + 1}. \quad (1.2.5)$$

The orbital velocity and radius vary as functions of the true anomaly, which is defined

*This is particularly true for resonant orbits [135].

†Of course, there exist many other ways of orienting the planar orbit in three dimensions. For further discussions on some alternatives to the Euler angles, see §5.2.1.

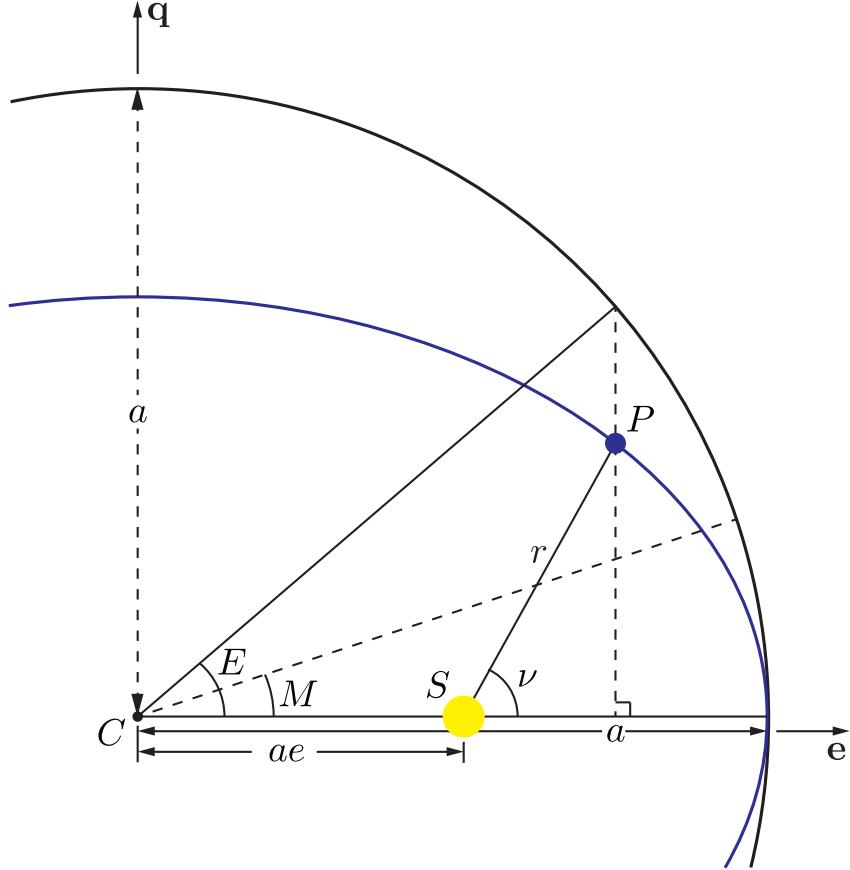


Figure 1.3: Schematic of a Keplerian orbit (blue ellipse) and auxiliary circle (black ellipse). The auxiliary circle has a radius equal to the orbit's semi-major axis, with center C .

as the angle between the current star-planet vector and the star-planet vector at the time of periapsis passage (see Figure 1.3). Using the planar orbital representation, we can write the planet position and velocity with respect to the star as:

$$\mathbf{r}_{P/S} = r (\cos \nu \hat{\mathbf{e}} + \sin \nu \hat{\mathbf{q}}) \quad (1.2.6)$$

$$\tau \mathbf{v}_{P/S} = \frac{d}{dt} \mathbf{r}_{P/S} = \sqrt{\frac{\mu_P + \mu_S}{\ell}} (-\sin \nu \hat{\mathbf{e}} + (e + \cos \nu) \hat{\mathbf{q}}) \quad (1.2.7)$$

where μ_P and μ_S are the gravitational parameters of the star and planet, respectively, equal to the product of their masses (m_S and m_P) and the universal gravitational constant G . The

semi-latus rectum of the ellipse, ℓ , is given geometrically as:

$$\ell = a(1 - e^2). \quad (1.2.8)$$

The unit vector $\hat{\mathbf{e}}$ (known as the normalized eccentricity vector) can be expressed as:

$$\hat{\mathbf{e}} = \left\langle \frac{1}{\mu_S + \mu_P} \left[\left(\|\mathbf{v}_{P/S}\|^2 - \frac{\mu_S + \mu_P}{r} \right) \mathbf{r}_{P/S} - (\mathbf{r}_{P/S} \cdot \mathbf{v}_{P/S}) \mathbf{v}_{P/S} \right] \right\rangle. \quad (1.2.9)$$

The orthogonal unit vector $\hat{\mathbf{q}}$ can be written as:

$$\hat{\mathbf{q}} = \langle \mathbf{r}_{P/S} \times \mathbf{v}_{P/S} \rangle \times \hat{\mathbf{e}}. \quad (1.2.10)$$

These two unit vectors, $\hat{\mathbf{e}}$ and $\hat{\mathbf{q}}$, define the perifocal reference frame. From Equation (1.2.10) we can see that the third vector of this frame, $\hat{\mathbf{h}} = \hat{\mathbf{e}} \times \hat{\mathbf{q}}$, is the direction of the orbit's normalized angular momentum:

$$\mathbf{h}_{P/S} = \mathbf{r}_{P/S} \times \mathbf{v}_{P/S}, \quad (1.2.11)$$

which is always orthogonal to the plane of the orbit.

To aid in calculating the planet position, it is useful to define an auxiliary circle with a radius equal to the orbit's semi-major axis and centered at the orbit's center. This allows us to define a mean anomaly, a central angle describing the mean planetary motion, given by:

$$M \triangleq 2\pi \frac{t - t_p}{P_{orb}}, \quad (1.2.12)$$

where P_{orb} is the orbital period, t is the current time, and t_p is the time of the last periapsis

passage (closest planet approach to the star). The orbital period is:

$$P_{orb} = 2\pi \sqrt{\frac{a^3}{\mu_S + \mu_P}}. \quad (1.2.13)$$

The mean anomaly is proportional to the area covered by $\mathbf{r}_{P/S}$ since periapsis passage, and is thus linear in time. The eccentric anomaly is related to the mean anomaly via the Kepler equation:

$$M = E - e \sin(E), \quad (1.2.14)$$

and related to the true anomaly as:

$$\tan\left(\frac{\nu}{2}\right) = \sqrt{\frac{1+e}{1-e}} \tan\left(\frac{E}{2}\right). \quad (1.2.15)$$

To orient the orbit in some observer's reference frame \mathcal{A} , we apply a 3-1-3 rotation using the angle set $(\psi, \theta, \phi)_{\mathcal{A}}$ [66]:

$$[\mathbf{r}_{P/S}]_{\mathcal{A}} = \begin{bmatrix} \cos \phi & \sin \phi & 0 \\ -\sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & \sin \theta \\ 0 & -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} \cos \psi & \sin \psi & 0 \\ -\sin \psi & \cos \psi & 0 \\ 0 & 0 & 1 \end{bmatrix} [\mathbf{r}_{P/S}]_{\mathcal{P}}, \quad (1.2.16)$$

where \mathcal{P} is a reference frame whose first two unit vectors lie in the $\hat{\mathbf{e}}, \hat{\mathbf{q}}$ plane and whose third unit vector is in the $\hat{\mathbf{h}}$ direction (see §2.1 for an example of one such frame). The \mathcal{P} frame is inertial only in the case of a non-accelerating frame origin, which results in another difficulty in using the Keplerian description as this makes \mathcal{P} a heliocentric frame. Note that we have used left-handed rotations here to conform to conventions in the literature (cf. Vinti [144]). We can use these rotations to convert between the Keplerian orbital elements and

the heliocentric planet position and velocity, simply by expressing the planar orbit as:

$$[\mathbf{r}_{P/S}]_{\mathcal{P}} = \begin{bmatrix} r \cos \nu \\ r \sin \nu \\ 0 \end{bmatrix} = \begin{bmatrix} a(\cos E - e) \\ b \sin E \\ 0 \end{bmatrix}, \quad (1.2.17)$$

where b is the semi-minor axis of the orbit, given by:

$$b = a\sqrt{1 - e}, \quad (1.2.18)$$

and we take the angle set $(\psi, \theta, \phi) = (\Omega, I, \omega)$ where Ω is the longitude of the ascending node, I is the inclination and ω is the argument of periapsis. The velocity vector is found by differentiating the position, yielding:

$${}^{\mathcal{P}}\mathbf{v}_{P/S} = {}^{\mathcal{P}}\frac{d}{dt}[\mathbf{r}_{P/S}]_{\mathcal{P}} = \sqrt{\frac{\mu_S + \mu_P}{a}} \frac{1}{1 - e \cos E} \begin{bmatrix} -\sin E \\ \sqrt{1 - e} \cos E \\ 0 \end{bmatrix}. \quad (1.2.19)$$

There are multiple algorithms for converting heliocentric position and velocity vectors to Keplerian orbital elements. One of the most straight-forward, described in Vinti [144], assumes a planar orientation as in Equation (1.2.16) and then calculates the semi-major axis and semi-latus rectum from the orbital energy and angular momentum:

$$a = -\frac{\mu_S + \mu_P}{2E} = -\frac{\mu_S + \mu_P}{2} \left[\frac{\|{}^{\mathcal{T}}\mathbf{v}_{P/S}\|^2}{2} - \frac{\mu_S + \mu_P}{\|\mathbf{r}_{P/S}\|} \right]^{-1}, \quad (1.2.20)$$

$$\ell = \frac{{}^{\mathcal{T}}\mathbf{h}_{P/S}}{\mu_S + \mu_P} \frac{\|\mathbf{r}_{P/S} \times {}^{\mathcal{T}}\mathbf{v}_{P/S}\|}{\mu_S + \mu_P}. \quad (1.2.21)$$

The eccentricity can then be calculated via Equation (1.2.8):

$$e = \sqrt{1 - \frac{\ell}{a}}. \quad (1.2.22)$$

Given inertial frame unit vectors $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$, we can write:

$$\cos I = \frac{(\mathbf{r}_{P/S} \times {}^T\mathbf{v}_{P/S}) \cdot \mathbf{e}_3}{\|\mathbf{r}_{P/S} \times {}^T\mathbf{v}_{P/S}\|}, \quad (1.2.23)$$

$$\sin I = \frac{\|\mathbf{e}_3 \times (\mathbf{r}_{P/S} \times {}^T\mathbf{v}_{P/S})\|}{\|\mathbf{r}_{P/S} \times {}^T\mathbf{v}_{P/S}\|}, \quad (1.2.24)$$

$$\cos \Omega = -\frac{(\mathbf{r}_{P/S} \times {}^T\mathbf{v}_{P/S}) \cdot \mathbf{e}_2}{\|\mathbf{r}_{P/S} \times {}^T\mathbf{v}_{P/S}\| \sin I}, \quad (1.2.25)$$

$$\sin \Omega = \frac{(\mathbf{r}_{P/S} \times {}^T\mathbf{v}_{P/S}) \cdot \mathbf{e}_1}{\|\mathbf{r}_{P/S} \times {}^T\mathbf{v}_{P/S}\| \sin I}, \quad (1.2.26)$$

$$\cos \omega = \frac{1}{e \sin I} \left(\frac{\|\mathbf{r}_{P/S} \times {}^T\mathbf{v}_{P/S}\| ({}^T\mathbf{v}_{P/S} \cdot \mathbf{e}_3)}{\mu_S + \mu_P} - \frac{\|\mathbf{r}_{P/S} \times {}^T\mathbf{v}_{P/S} \times \mathbf{r}_{P/S}\|}{\|\mathbf{r}_{P/S} \times {}^T\mathbf{v}_{P/S}\| \|\mathbf{r}_{P/S}\|} \right), \quad (1.2.27)$$

$$\sin \omega = \frac{1}{e \sin I} \left(\frac{\|\mathbf{r}_{P/S} \times \mathbf{r}_{P/S} \times {}^T\mathbf{v}_{P/S}\|}{\mu_S + \mu_P} - \frac{\mathbf{r}_{P/S} \cdot \mathbf{e}_3}{\|\mathbf{r}_{P/S}\|} \right). \quad (1.2.28)$$

With both sine and cosine expressions for an angle, the angle may be reconstructed over the full range of $[0, 2\pi]$ with the two argument arctangent function:

$$\text{atan2}(y, x) = \begin{cases} \tan^{-1}(y/x) & x > 0 \\ \pi + \tan^{-1}(y/x) & x < 0, y \geq 0 \\ -\pi + \tan^{-1}(y/x) & x < 0, y < 0 \\ \pi/2 & x = 0, y > 0 \\ -\pi/2 & x = 0, y < 0 \end{cases} \quad (1.2.29)$$

where $y = \cos \theta$ and $x = \sin \theta$. Thus Equations (1.2.23) through (1.2.28) can be used to precisely orient the orbit in the unit sphere given a single measurement of the position and velocity vectors. The time varying Keplerian orbital element (anomaly) can similarly be found by using:

$$\cos E = \frac{1}{e} \left(1 - \frac{\|\mathbf{r}_{P/S}\|}{a} \right), \quad (1.2.30)$$

$$\sin E = \frac{{}^T\mathbf{v}_{P/S} \cdot \mathbf{r}_{P/S}}{e \sqrt{a(\mu_S + \mu_P)}}. \quad (1.2.31)$$

Because this algorithm requires no iteration and the equations can be fully vectorized, the transformation can be implemented very efficiently (see Listing A.1).

As previously mentioned, the Keplerian description is exact only for two body systems and can only be used as an approximation when a system contains more than two massive bodies (especially if the planetary bodies have substantial masses, as in cases of systems with multiple Jupiter-size planets). A much more exact approximation of the motion of n bodies utilises Newton's law of gravity to express the dynamics of the planetary system via the second-order, ordinary differential equation:

$$\frac{^I d^2}{dt^2} \mathbf{r}_{j/O} = \sum_{k \in Y} \mu_k \frac{\mathbf{r}_{k/O} - \mathbf{r}_{j/O}}{\|\mathbf{r}_{k/O} - \mathbf{r}_{j/O}\|^3}, \quad j \in X \quad (1.2.32)$$

where the set X contains all of the bodies in the system:

$$X = \{P_1, P_2, \dots, P_n, S\}, \quad (1.2.33)$$

while the Y is the subset of X not containing j :

$$Y \subset X : j \notin Y. \quad (1.2.34)$$

The total energy, which is conserved in the absence of external perturbations, can be expressed as:

$$E_{tot} = \sum_{j \in X} \left(\frac{m_j}{2} ({}^I \mathbf{v}_{j/O} \cdot {}^I \mathbf{v}_{j/O}) - \sum_{k \in Y} \frac{m_j \mu_k}{\|\mathbf{r}_{k/O} - \mathbf{r}_{j/O}\|} \right). \quad (1.2.35)$$

In the case where there are only two bodies (say, a star and one planet), Equation (1.2.32) can be rewritten in terms of the separation vector between the pair, rather than with respect to some inertial origin:

$$\frac{^I d^2}{dt^2} \mathbf{r}_{P/S} + (\mu_S + \mu_P) \frac{\mathbf{r}_{P/S}}{\|\mathbf{r}_{P/S}\|^3} = 0, \quad (1.2.36)$$

from which all of Kepler's laws and orbital equations can be derived. Using Equations

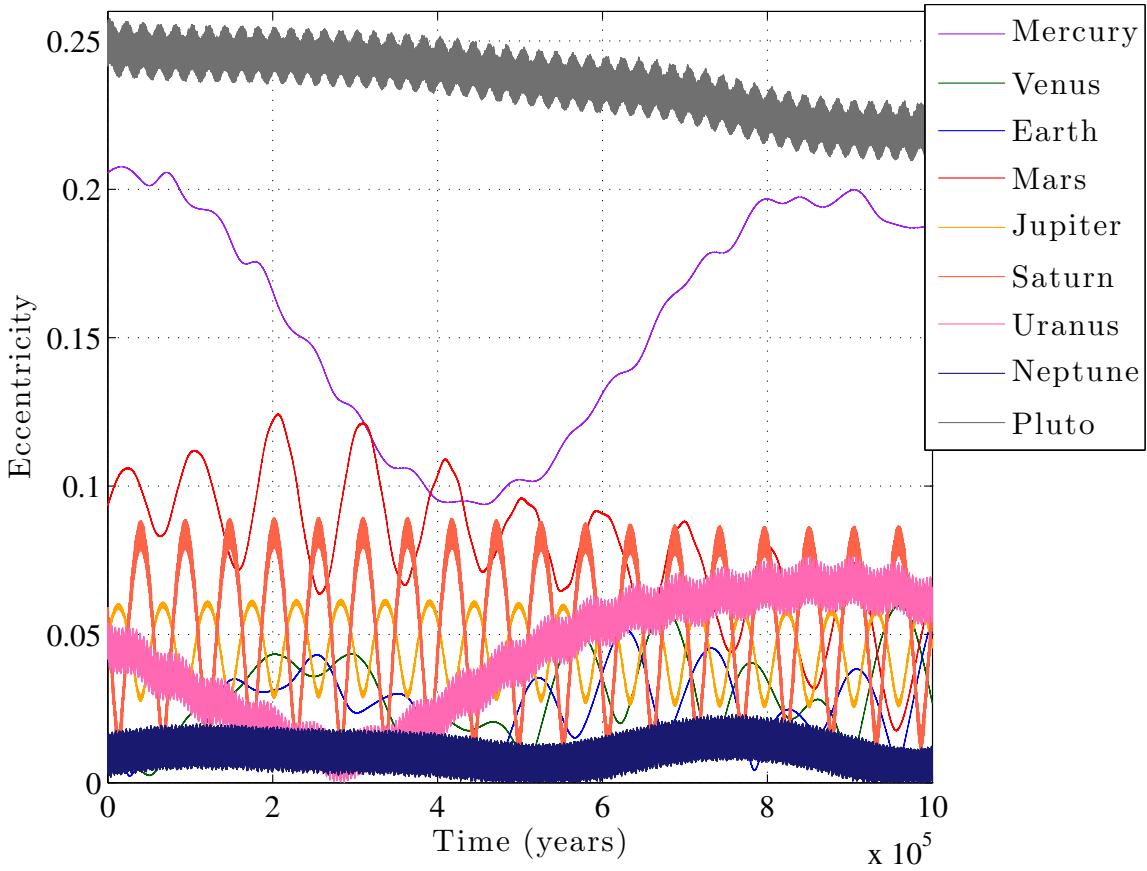


Figure 1.4: Variation in orbital eccentricities of solar system bodies over one million years.

(1.2.20) through (1.2.31), we can see that in the multi-planet case, the Keplerian orbital elements represent instantaneous fits to Equation (1.2.36), which approximates the true system dynamics assuming that the combined effects due to all other bodies in the system are small compared to those due to the star. Alternatively, we can find osculating Keplerian elements for multi-body systems that are time-varying, as in Figure 1.4.

Of course, the Newtonian formulation is itself an approximation, with the true expressions for the motions of bodies in gravitational fields given by the Einstein field equations. There are almost no cases in the modeling of planetary systems where it is necessary to use the field equations to calculate the exact metric tensor, but there may be observable general relativity effects in bodies orbiting closely to their parent stars (as happens, for example, with Mercury in our solar system [35]). In these cases, it is possible to model some of the

‘post-Newtonian’ effects by expanding the field equations to first order in the ratio of orbital velocity and the speed of light. One example of this is the Einstein-Infeld-Hofmann equations of motion [34, 33]:

$$\begin{aligned}
{}^{\mathcal{I}} \frac{d^2}{dt^2} \mathbf{r}_{j/O} = & \sum_{k \in Y} \frac{\mu_k (\mathbf{r}_{k/O} - \mathbf{r}_{j/O})}{\|\mathbf{r}_{k/O} - \mathbf{r}_{j/O}\|^3} + \frac{1}{c^2} \sum_{k \in Y} \frac{\mu_k (\mathbf{r}_{k/O} - \mathbf{r}_{j/O})}{\|\mathbf{r}_{k/O} - \mathbf{r}_{j/O}\|^3} \left[{}^I \mathbf{v}_{j/O} \cdot {}^I \mathbf{v}_{j/O} \right. \\
& + 2 {}^I \mathbf{v}_{k/O} \cdot {}^I \mathbf{v}_{k/O} - 4 {}^I \mathbf{v}_{j/O} \cdot {}^I \mathbf{v}_{k/O} - \frac{3}{2} \left(\frac{(\mathbf{r}_{j/O} - \mathbf{r}_{k/O}) \cdot {}^I \mathbf{v}_{k/O}}{\|\mathbf{r}_{j/O} - \mathbf{r}_{k/O}\|} \right)^2 \\
& - 4 \sum_{i \in Y} \frac{\mu_i}{\|\mathbf{r}_{j/O} - \mathbf{r}_{i/O}\|} - \sum_{i \in Z} \frac{\mu_i}{\|\mathbf{r}_{k/O} - \mathbf{r}_{i/O}\|} + \frac{1}{2} \left((\mathbf{r}_{k/O} - \mathbf{r}_{j/O}) \cdot {}^{\mathcal{I}} \frac{d^2}{dt^2} \mathbf{r}_{k/O} \right) \Big] \\
& + \frac{1}{c^2} \sum_{k \in Y} \frac{\mu_k}{\|\mathbf{r}_{k/O} - \mathbf{r}_{j/O}\|^3} [(\mathbf{r}_{j/O} - \mathbf{r}_{k/O}) \cdot (4 {}^I \mathbf{v}_{j/O} - 3 {}^I \mathbf{v}_{k/O})] ({}^I \mathbf{v}_{j/O} - {}^I \mathbf{v}_{k/O}) \\
& + \frac{7}{2c^2} \sum_{k \in Y} \frac{\mu_k}{\|\mathbf{r}_{k/O} - \mathbf{r}_{j/O}\|} {}^{\mathcal{I}} \frac{d^2}{dt^2} \mathbf{r}_{k/O}, \quad j \in X
\end{aligned} \tag{1.2.37}$$

where c is the speed of light and Z is the subset of X not containing k :

$$Z \subset X : k \notin Z. \tag{1.2.38}$$

Besides being functionally more complex than Equation (1.2.32), this equation is also significantly more difficult to integrate as it contains velocity terms (invalidating a whole class of second-order integrators that are commonly used to integrate n-body systems) as well as acceleration terms, which require a fully implicit integrator. Because of this, it is common practice to integrate only the Newtonian formulation, and to apply the post-Newtonian correction only when deviations due to general relativity effects become too pronounced to ignore[†].

[†]For further discussion on numerical integration see §4.2.1.

1.3 The Point Spread Function

Before proceeding to any discussion of specific detection methods, we need a standardized description of an imaging system, which will be used to develop the models for direct detection, doppler spectroscopy, and transit photometry (and is also applicable to some interferometers used for astrometry). Following Goodman [40], we assume the imaging system to act as a linear operator \mathcal{F} such that,

$$\mathcal{F} : E_o(u, v) \rightarrow E_i(x, y) \quad (1.3.1)$$

where $E_o(u, v)$ is the electric field transmitted by the object and $E_i(x, y)$ is the image electric field (with image coordinates x, y). The input can be decomposed by taking advantage of the Dirac δ function as:

$$E_o(u, v) = \iint_{-\infty}^{\infty} E_o(\xi, \eta) \delta(u - \xi, v - \eta) d\xi d\eta, \quad (1.3.2)$$

so that we can write:

$$E_i(x, y) = \iint_{-\infty}^{\infty} E_o(\xi, \eta) h(x, y, \xi, \eta) d\xi d\eta, \quad (1.3.3)$$

where

$$h(x, y, \xi, \eta) = \mathcal{F} \{ \delta(u - \xi, v - \eta) \}. \quad (1.3.4)$$

The superposition integral in Equation (1.3.3) can be used to characterize an imaging system by its response to unit impulses via the impulse response function h , known as the point spread function (PSF). This function can be interpreted as the resulting image of an electric field emitted by a point source; any image can thus be built up from images of point sources distributed throughout the field of view. With some assumptions, it can be shown that the function \mathcal{F} is a Fourier transform, allowing us to express h in terms of specific instrument properties [54, 6].

Starting with an aperture A in the (u, v) plane, we form another plane (x, y) , a distance z away. By the Huygens-Fresnel principle,

$$E_z(x, y) = \frac{1}{i\lambda} \iint_{A(u,v)} E_o(u, v) e^{i\frac{2\pi}{\lambda} \|\mathbf{r}\|} \frac{\cos \theta}{\|\mathbf{r}\|} du dv \quad (1.3.5)$$

where λ is the wavelength, \mathbf{r} is the vector between the point in (x, y) and the point in (u, v) , and θ is the angle between \mathbf{r} and the vector normal to the two planes, such that $\cos \theta = z/\|\mathbf{r}\|$. This formulation is based on scalar diffraction theory and assumes that $\|\mathbf{r}\| \gg \lambda$, i.e., that the image is many wavelengths from the aperture.

Applying the binomial expansion to the magnitude of \mathbf{r} , we have:

$$\|\mathbf{r}\| = z \sqrt{1 + \left(\frac{x-u}{z}\right)^2 + \left(\frac{y-v}{z}\right)^2} \quad (1.3.6)$$

$$= z + \frac{z}{2} \left(\frac{x-u}{z}\right)^2 + \frac{z}{2} \left(\frac{y-v}{z}\right)^2 + \dots \quad (1.3.7)$$

so we make the approximation:

$$E_z(x, y) \approx \frac{e^{i\frac{2\pi}{\lambda} z}}{i\lambda z} \iint_{A(u,v)} E_o(u, v) e^{i\frac{\pi}{\lambda z} [(x-u)^2 + (y-v)^2]} du dv. \quad (1.3.8)$$

This expression, known as the Fresnel integral, can be seen to be a convolution[§] between $E_o(u, v)$ and the kernel:

$$h(x, y) = \frac{e^{i\frac{2\pi}{\lambda} z}}{i\lambda z} e^{i\frac{\pi}{\lambda z} (x^2 + y^2)}. \quad (1.3.9)$$

Because this is a Gaussian Kernel, the Fourier conjugate of E_z can be written as:

$$\hat{E}_z(\xi, \eta) = e^{i\frac{2\pi}{\lambda} z} \hat{E}_o(\xi, \eta) e^{-i\pi\lambda z (\xi^2 + \eta^2)}. \quad (1.3.10)$$

[§]Strictly, the integral should be written with limits from $-\infty$ to ∞ , in which case it is understood that $E_o(u, v)$ is zero outside of $A(u, v)$.

Similarly, Equation (1.3.8) can be rewritten as:

$$E_z(x, y) = \frac{e^{i\frac{2\pi}{\lambda}z}}{i\lambda z} e^{i\frac{\pi}{\lambda z}(x^2+y^2)} \iint_{-\infty}^{\infty} E_o(u, v) e^{i\frac{\pi}{\lambda z}(u^2+v^2)} e^{-i\frac{2\pi}{\lambda z}(xu+yv)} du dv, \quad (1.3.11)$$

which is a scaled Fourier transform of $E_o(u, v) e^{i\frac{\pi}{\lambda z}(u^2+v^2)}$.

Assuming that there is a lens at the aperture plane with focal length f , Equation (1.3.8) can be used to express the field at the focal plane as:

$$E_f(x, y) = \frac{e^{i\frac{2\pi}{\lambda}f}}{i\lambda f} e^{i\frac{\pi}{\lambda z}(x^2+y^2)} \hat{E}_A \left(\frac{x}{\lambda f}, \frac{y}{\lambda f} \right) \quad (1.3.12)$$

for

$$E_o(u, v) = E_A(u, v) e^{i\frac{\pi}{\lambda f}(u^2+v^2)}. \quad (1.3.13)$$

Now, if the free space propagation before the lens is equal to f , then Equations (1.3.12) and (1.3.10) combine to produce:

$$E_i(x, y) = \frac{1}{i\lambda f} e^{i\frac{\pi}{\lambda f}(\frac{1}{f}-\frac{1}{f})(x^2+y^2)} \hat{E}_o \left(\frac{x}{\lambda f}, \frac{y}{\lambda f} \right) = \frac{1}{i\lambda f} \hat{E}_o \left(\frac{x}{\lambda f}, \frac{y}{\lambda f} \right) \quad (1.3.14)$$

so that the electric field at the focal plane is simply the Fourier transform of the incoming field.

Returning to Equation (1.3.3), the impulse response function can now be written as:

$$h(x, y) = \frac{1}{\lambda f} \iint_{-\infty}^{\infty} A(u, v) e^{-i\frac{2\pi}{\lambda f}(xu+yv)} du dv. \quad (1.3.15)$$

Since the effect of a lens is to take the Fourier transform of the incoming field, the intensity distribution of the image is given by

$$I_i(x, y) = |\hat{E}_0|^2, \quad (1.3.16)$$

and so we can write

$$I_i(x, y) = \kappa \iint_{-\infty}^{\infty} |h(\xi - u, v - \eta)|^2 I_o(u, v) \, du \, dv, \quad (1.3.17)$$

for real constant κ . We can thus define the intensity impulse response, $|h|^2$, or point spread function:

$$\text{PSF}(x, y) = \frac{1}{\lambda^2 f^2} \left| \iint_{-\infty}^{\infty} A(u, v) e^{-i \frac{2\pi}{\lambda f} (xu + yv)} \, du \, dv \right|^2. \quad (1.3.18)$$

Equation (1.3.18) provides us with a way of characterizing the effects of the optical system, and thereby the ability to disentangle these from the true signal being measured.

The descriptions of planetary orbits, physical parameters and imaging systems laid out in the previous two sections make it possible to rigorously define exactly how the different detection methods find and observe planets, using a unified syntax which will prove very important in subsequent chapters. In the next chapter, we will introduce the four detection method studied in this thesis and, using the groundwork laid here, we will be able to describe what form the data from each detection method will take, and develop expressions for the observables and constraints of each method.

Chapter 2

Observation Methods

This chapter will describe four methods currently in use for the detection of exoplanets. While these do not represent all of the ways that have been proposed or successfully employed to study exoplanets, three of these (doppler spectroscopy, transit photometry, and astrometry) have provided us with the bulk of the currently known exoplanets, while direct detection promises to be a very powerful tool for future exoplanet study, and a subject of much interest in the community. These four methods also fit well into the common parameter set described in §1.2, which makes it simpler to discuss applications incorporating data derived from more than one of these methods.

2.1 Direct Detection

Direct detection (or imaging) instrumentation can be split roughly into two classes: internal coronagraphs and external occulters. Broadly speaking, internal coronagraphs seek to create regions of high contrast in the image plane by generating specialized PSFs. This is achieved with various combinations of shaped pupils [142, 141, 68], apodized pupils [62, 97], Lyot stops [75, 125], pupil mapping systems [48, 140], etc. As these systems are highly sensitive to minute surface errors on any of the optical elements, they often include deformable mirrors (DMs) as part of active wave-front control systems to correct for any errors introduced in the

wavefront by internal optics [104]. This adds a great deal of complexity to these systems and requires precise wavefront estimation and broadband wavefront correction, both of which are extremely difficult [45].

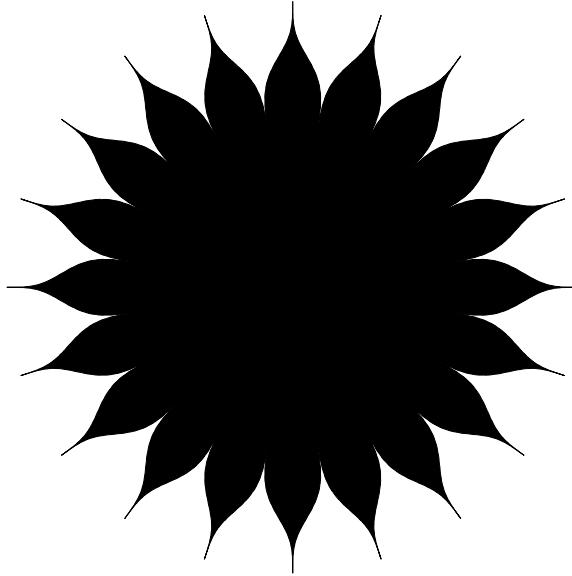


Figure 2.1: A starshade. The ‘petals’ are designed to prevent diffraction of light about the central disk, thereby creating a dark region at the aperture of the telescope.

Alternatively, external coronagraphs use a separate spacecraft (an ‘occultor’ or ‘starshade’; see Figure 2.1) between the target and the telescope to block most of the starlight from entering the telescope pupil [18, 143, 16]. This means that the telescope itself can be fairly conventional, as the contrast level in the image plane no longer needs to be large to detect planets. This also means that active wavefront control is not required, as the system is no longer highly sensitive to small errors in the optical surfaces. However, the occultor must be flown in exact formation with the telescope throughout observations, at separations of tens of thousands of kilometers and required alignment precisions of tens of centimeters, yielding a very difficult control problem [123]. The starshade must also be manufactured to very strict tolerances [117]. The modeling of a direct planetary observation will follow the descriptions in Brown [8] and Savransky et al. [113], and will be applicable to both types of imaging systems.

2.1.1 Imaging Observables and Constraints

An observation of a planet in reflected or emitted light produces two pieces of information: the planet's position and brightness. In most cases, these values will be relative to the primary star, providing us with the planet's position relative to the star in the plane of the sky (with respect to the observer) and the ratio of fluxes between the planet and star. This makes it useful to define a ‘sky’ frame, $\mathcal{S} = (S, \mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3)$ where the the observer line of sight to the target star is along \mathbf{s}_3 , and the plane of the sky (from the viewpoint of the observer) lies in $(\mathbf{s}_1, \mathbf{s}_2)$. That is,

$$\frac{\mathbf{r}_{S/sc}}{\|\mathbf{r}_{S/sc}\|} \equiv \mathbf{s}_3, \quad (2.1.1)$$

as shown in Figure 2.2. In contrast to Equation (1.2.17), we begin with the planar orbit orientation in $(\mathbf{s}_2, \mathbf{s}_3)$:

$$[\mathbf{r}_{P/S}]_{\mathcal{P}} = \begin{bmatrix} 0 \\ r \sin \nu \\ r \cos \nu \end{bmatrix}. \quad (2.1.2)$$

Since the 3-1-3 rotation can orient the orbit in all directions throughout the unit sphere, the initial orientation can be in any plane. The $(\mathbf{s}_2, \mathbf{s}_3)$ plane is used here to conform to earlier studies and allow simpler comparison of the results. Following Equation (1.2.16), we use the Euler angle set $(\psi, \theta, \phi)_{\mathcal{S}}$ to write:

$$[\mathbf{r}_{P/S}]_{\mathcal{S}} = \begin{bmatrix} r(\cos \nu \sin \theta \sin \phi + \sin \nu (\cos \theta \cos \psi \sin \phi + \cos \phi \sin \psi)) \\ r(\cos \nu \cos \phi \sin \theta + \sin \nu (\cos \theta \cos \phi \cos \psi - \sin \phi \sin \psi)) \\ r(\cos \theta \cos \nu - \cos \psi \sin \theta \sin \nu) \end{bmatrix}. \quad (2.1.3)$$

We define the separation \mathbf{s} as the projection of the star-planet vector into the plane of

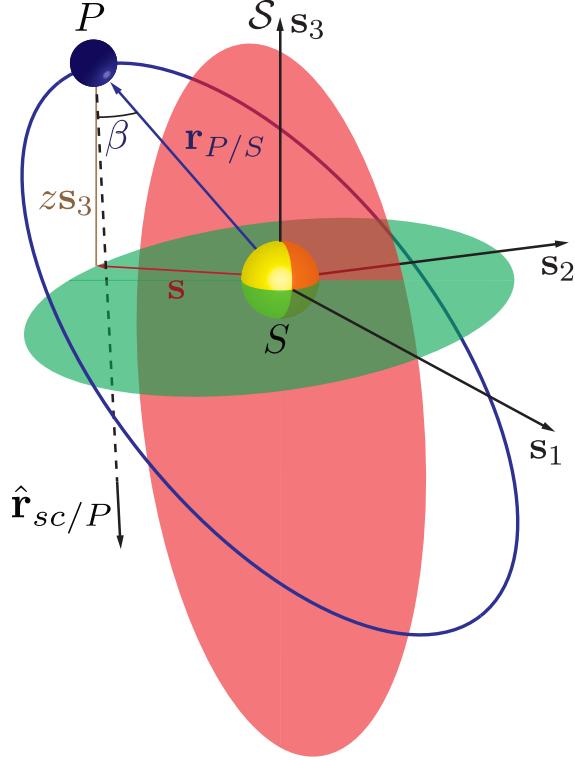


Figure 2.2: Schematic of a planetary orbit. The green shaded ellipse represents the projection of the orbit into the plane of the sky ($\mathbf{s}_1, \mathbf{s}_2$), while the red shaded ellipse represents the planar orbit in ($\mathbf{s}_2, \mathbf{s}_3$).

the sky:

$$[\mathbf{s}]_S = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} [\mathbf{r}_{p/s}]_S, \quad (2.1.4)$$

as in Figure 2.2. This projection also leads us to define the apparent separation - the magnitude of the star-planet vector projected into the plane of the sky:

$$s = \|\mathbf{s}\|. \quad (2.1.5)$$

The apparent separation may be either a derived quantity (when the instrument is capable of resolving the star-planet separation in two dimensions, or can measure the absolute planet position with respect to another reference like the solar system barycenter) or a primary

observable (when the instrument can only measure the separation in one dimension).

The relative flux between the planet and its parent star is a function of the orbital radius and the planet's albedo and radius:

$$F_R \triangleq \frac{F_p}{F_s} = p\Phi(\beta) \left(\frac{R}{r}\right)^2, \quad (2.1.6)$$

where Φ is the planet's phase function [124], and β is the star-planet-observer (phase) angle, given by:

$$\beta = \cos^{-1} \left(\frac{\mathbf{r}_{P/sc} \cdot \mathbf{r}_{P/S}}{\|\mathbf{r}_{P/sc}\| \|\mathbf{r}_{P/S}\|} \right), \quad (2.1.7)$$

where $\mathbf{r}_{P/sc}$ is the position of the planet relative to the observer:

$$\|\mathbf{r}_{P/sc} - \mathbf{r}_{sc/O} - \mathbf{r}_{P/S}\| = d. \quad (2.1.8)$$

However, since $d \gg r$, we can make the approximation:

$$\beta \approx \cos^{-1} \left(\frac{z}{r} \right). \quad (2.1.9)$$

where z is the \mathbf{s}_3 axis component of $[\mathbf{r}_{p/s}]_S$. For the closest star to Earth (approximately 1.3 pc away), this approximation produces errors of less than 0.001%. The flux ratio is often expressed on a logarithmic scale, in which case it becomes the difference in magnitude between the star and planet:

$$\Delta\text{mag} \triangleq -2.5 \log_{10} F_R. \quad (2.1.10)$$

Observable limits on s and F_R , or functions thereof, can be mapped to instrument specifications for a fairly broad variety of direct imaging instruments by assuming that a planet will be observable if its angular separation from the star is greater than the observatory's inner working angle (IWA), and illuminated such that the Δmag is below a threshold value,

called the limiting Δmag , or Δmag_0 . The IWA represents the minimum angular separation between the telescope’s central axis (line of sight) and detectable objects on the sky. The specific determination of the IWA depends on the instrument design, and may be affected by the size of a central obscuration, the capability of an adaptive optics systems to remove light from certain areas of the image plane, or the size and geometry of external occulting optics.

The limiting Δmag represents the point where systematic errors produce unresolvable confusion between planet signal and background noise [8]. It is important to note that while this value is often equated with the designed contrast of an instrument (i.e., the ratio of the core to halo of a coronagraph’s point spread function), they are not necessarily the same. Depending on the nature of the background and instrument systematics, Δmag_0 may actually be less than the designed contrast. In fact, it is becoming standard practice to set the design contrast at several magnitudes below the Δmag_0 required by the population of interest, based on a detailed error budget (see, for example, Shaklan et al. [117]). Figure 2.3 shows a schematic of the instrument constraints on direct imaging.

A significant difference between occulters and coronagraphs is the dependence of their IWA on wavelength. The extent of the coronagraph’s produced region of high contrast depends completely on the geometry of the internal optics. If the region extends to some location n in the image plane, then it represents an aperture plane location of $n/f\lambda/D$ where D is the aperture diameter and f is the focal length. Thus, the IWA for the coronagraph is commonly given in units of λ/D , and changes with wavelength. An occulter, on the other hand creates a region of high contrast over a specific wavelength band. In this case, the light from the star is removed before it ever gets to the telescope, so the dark region extends throughout the entire image plane. On the other hand, if the angular separation of the planet from the star is less than the angular size of the occulter on the sky, then the light from the planet will also be blocked. Therefore, the IWA (often called the geometric IWA)

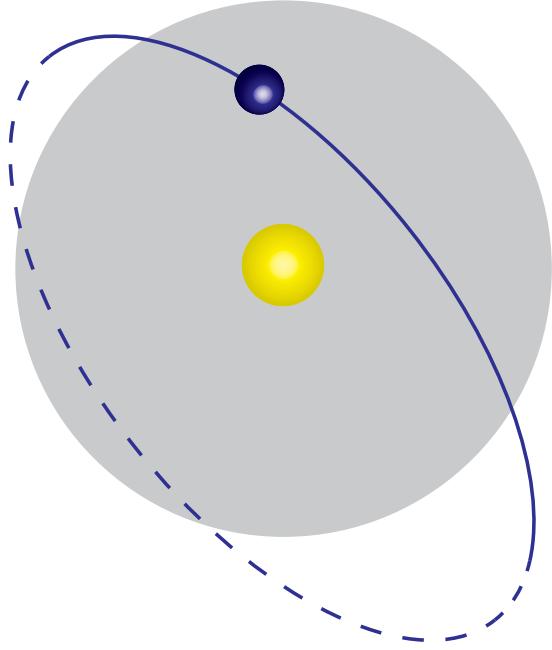


Figure 2.3: Schematic of the instrument constraints on direct imaging. The shaded circle centered on the star represents the projected IWA (no observations can be made within this angle of the star) and the planet is sufficiently illuminated ($\Delta\text{mag} < \Delta\text{mag}_0$) on the solid portion of the orbit (when the planet is between the observer and the star along the line of sight). In this case, the planet is only observable on two small portions of its orbit near periapsis and apoapsis.

for the occulter is set by its radius R_O and separation distance from the telescope d_s :

$$\text{IWA}_{\text{geom}} = \sin^{-1} \left(\frac{R_O}{d_s} \right). \quad (2.1.11)$$

Because occulters are designed with petals (as in Figure 2.1), it is actually possible to see some of the planet light coming through the gaps between the petals when the planet is inside the geometric IWA, albeit at a significantly decreased system throughput.

We have so far defined all of the elements that are used to calculate s and F_R , except for the phase function. Unfortunately, there can be no single definition for Φ , as it depends on the characteristics of a given planet. One way to get around this difficulty is to assume that the planet can be modeled as a Lambertian reflector, i.e., that the planet's surface luminance is isotropic. This leads to the commonly used Lambert phase function. Because

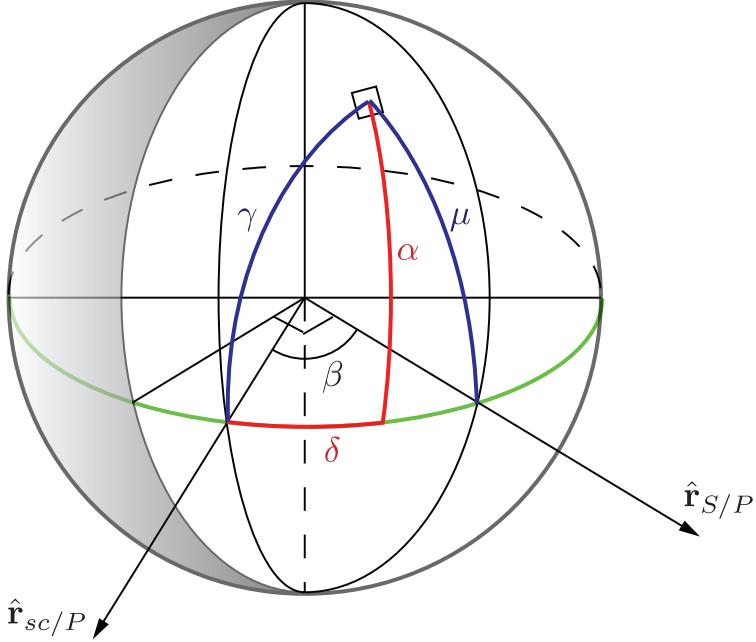


Figure 2.4: Schematic of a reflecting spherical body. The angles (α, δ) are the coordinates of an arbitrary point on the surface, μ is the angle of incident radiation, and γ is the angle of emergent radiation.

of the importance of this function, it is useful to briefly derive it here. Following Sobolev [124], we let (α, δ) be the planetary latitude and longitude of an arbitrary point on the planet's surface, μ the angle of incident radiation, γ the angle of emergent radiation at a surface point, and ξ the azimuthal angle between μ and γ such that:

$$\cos \xi = \frac{\cos \mu \cos \gamma - \cos \beta}{\sin \mu \sin \gamma}. \quad (2.1.12)$$

These definitions are shown schematically in Figure 2.4 and yield the relationships:

$$\cos \mu = \cos \alpha \cos(\beta - \delta) \triangleq C_\mu, \quad (2.1.13)$$

$$\cos \gamma = \cos \alpha \cos \delta \triangleq C_\gamma. \quad (2.1.14)$$

The emergent intensity is given by:

$$I = C_\gamma F \rho(C_\mu, C_\gamma, \xi), \quad (2.1.15)$$

where $\pi C_\gamma F$ is the flux on a patch of the surface and ρ is the reflection coefficient. The energy crossing a surface patch into unit solid angle is $C_\gamma C_\mu F \rho dA$ where

$$dA = R^2 \cos \alpha d\alpha d\delta. \quad (2.1.16)$$

The energy per second per unit area per unit solid angle received by an observer is thus:

$$E(\beta) = \frac{FR^2}{r^2} \int_{\beta-\pi/2}^{\pi/2} \cos(\beta - \delta) \cos \delta d\delta \int_{-\pi/2}^{\pi/2} \rho(C_\mu, C_\gamma, \xi) \cos^3 \alpha d\alpha \quad (2.1.17)$$

The phase function is defined as the ratio of $E(\beta)/E(0)$. For isotropic scattering, ρ is constant so:

$$\Phi_L(\beta) = \frac{\int_{\beta-\pi/2}^{\pi/2} \cos(\beta - \delta) \cos \delta d\delta}{\int_{-\pi/2}^{\pi/2} \cos(-\delta) \cos \delta d\delta} = \frac{\sin(\beta) + (\pi - \beta) \cos(\beta)}{\pi} \quad (2.1.18)$$

Figure 2.5 compares the Lambert phase function with empirical values found for the various planetary bodies in the solar system. It is immediately evident that the assumption of isotropic scattering is not a good one for most of these, as it overestimates the true phase function in almost all cases. More precise modeling can be achieved by using empirically derived lookup tables, whenever possible.

Having a functional form for the phase function does have certain advantages, however. For example, we can quantify the relationships between the direct detection observables such as the extrema in planet to star flux ratios as functions of the apparent separation. By

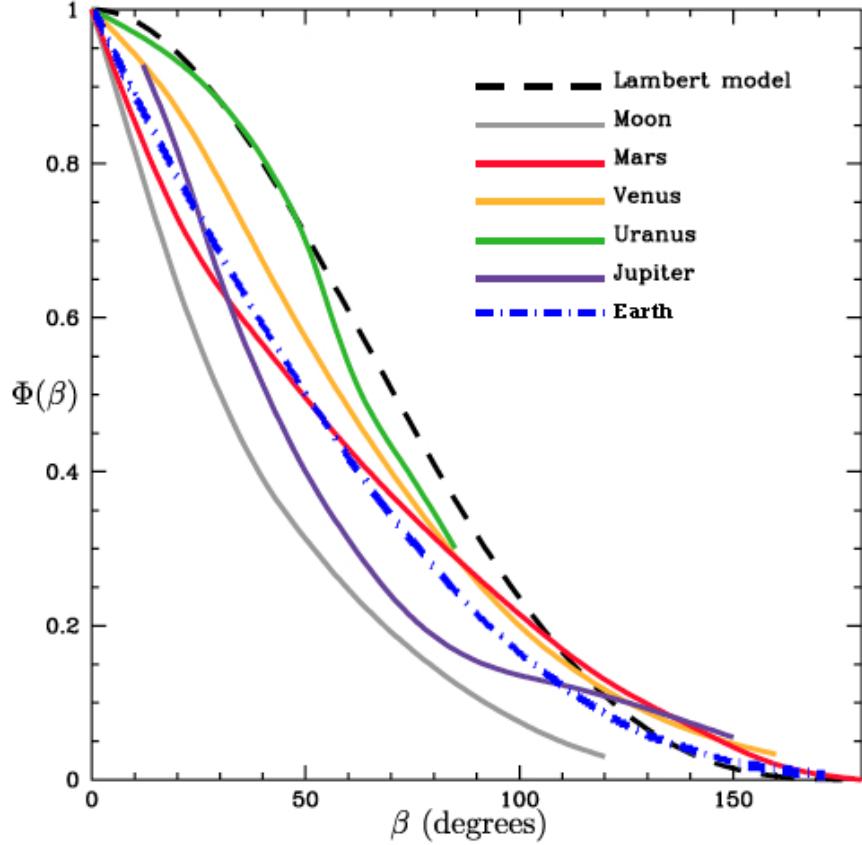


Figure 2.5: Empirically determined phase functions of solar system bodies plotted with the Lambert phase function. Data from Sudarsky et al. [133] and De Vaucouleurs [29].

Equation (2.1.9), we can also write:

$$\beta = \sin^{-1} \left(\frac{s}{r} \right) \Rightarrow r = \frac{s}{\sin \beta}, \quad (2.1.19)$$

so that Equation (2.1.6), assuming $\Phi = \Phi_L$, becomes:

$$F_R = \frac{pR^2}{s^2} \sin^2 \beta [(\pi - \beta) \cos \beta + \sin \beta]. \quad (2.1.20)$$

Following Brown [8], we differentiate with respect to β to find that the extrema of F_R for fixed p, R , and s occur when:

$$\sin \beta [(\pi - \beta) (1 + 3 \cos 2\beta) + 2 \sin 2\beta] = 0. \quad (2.1.21)$$

We can thus write the maximum flux ratio as:

$$\max F_R = \frac{pR^2}{s^2} \Phi_L(\beta^*) \quad (2.1.22)$$

where β^* is the solution to Equation (2.1.21) and is approximately equal to 1.1047 rad (63.296°). The other two extrema of F_R given by Equation (2.1.21) are at $\beta = 0$ and $\beta = \pi$, corresponding to the planet directly in front or behind the star along the observer's line of sight. As this does not occur for all orbital orientations, it is useful to consider the case of the minimum apparent separation (the planet's closest approach to the star in the plane of the sky) when the planet is between the star and the observer. This occurs when:

$$\beta = \pi - \sin^{-1} \left(\frac{s}{r_a} \right), \quad (2.1.23)$$

where r_a is the orbital radius at apoapsis, given (for a Keplerian orbit) by $r_a = a(1 + e)$.

The minimum flux ratio is thus:

$$\min F_R = \frac{pR^2}{r_a^3} \left[s - \sqrt{r_a^2 - s^2} \sin^{-1} \left(\frac{s}{r_a} \right) \right]. \quad (2.1.24)$$

These limits are useful in determining whether an observation belongs to a specific population of planets, or when bounding the expected performance of an instrument (see §3.2.1).

2.1.2 Observation Modeling

When using a direct detection system, an observation of (or ‘visit’ to) a target system can be broken down into two main activities: first, it is necessary to determine whether a planet has been detected; second, in the event of a detection, the planet may be characterized using the specific capabilities of the instrument, which may cover various wavelength ranges with varying spectral resolutions. Orbital characterization can only be obtained as a result of multiple detections of the same planet and depends on the spatial resolution of the instrument

and the ability to accurately centroid in the produced data. We can model the success or failure of both these tasks based on the time required compared with the time available.

Every visit results in one of four possible outcomes: detection, missed detection, null detection or false alarm. A null detection occurs when there are no observable planets at the time of observation. A missed detection signifies that an observable planet does exist, but was not found (i.e., if the integration time was insufficient to achieve a detection). On the other hand, a false alarm means that a detection is assumed when there is actually no corresponding planet in the target system (i.e., when noise or other objects in the field of view are mistaken for a planet). In the case of false alarms, it is assumed that followup spectroscopy will be able to resolve these as null detections (at the cost of additional integration time). In statistical terms, a null detection is a true negative, a missed detection is a false negative, a false alarm is a false positive, and a detection is a true positive.

Following Kasdin and Braems [67], we assume that the instrument produces images with sufficient sampling so that matched filtering (or some other probabilistic detection algorithm) can be applied, which significantly reduces the integration time required to decide whether a planet is present in the field of view. We treat the photons received at pixel j of the detector array as a random variable of the form

$$z_j = C_p \bar{P}_j + C_b + \nu \quad (2.1.25)$$

where C_p and C_b are the mean photon count at the pixel centered on the point spread function (PSF) and mean background photon count at all pixels, respectively, \bar{P} is the normalized, non-dimensionalized PSF, and ν is the photon and readout noise. This allows us to construct a signal to noise metric as the random variable \hat{C}_p/σ_b where \hat{C}_p is the linear, unbiased estimate of the planet signal, and σ_b is the variance of this estimate for pixels without a planet. The assumption made in this metric is that the background contribution is known, or can be estimated during the integration. We define the ratio of planet to

background counts as the parameter Q :

$$Q \triangleq \frac{C_p}{C_b}. \quad (2.1.26)$$

Simplifying the statistics by assuming the photon arrival rate can be approximated as a Gaussian distribution, and noting that both C_p and C_b are linear in integration time (t), we can write an expression for t based on a desired false alarm probability (FAP) and missed detection probability (MDP). Kasdin and Braems [67] derive this expression as:

$$t = \frac{1}{b} \frac{(K - \gamma \sqrt{1 + \tilde{Q}\Xi/\Psi})^2}{\tilde{Q}T_A\Psi}, \quad (2.1.27)$$

where \tilde{Q} is equal to Q scaled by the sum of \bar{P} ,

$$\Xi = \frac{\sum_j \bar{P}_j^3}{(\sum_j \bar{P}_j)^3}, \quad \text{and} \quad \Psi = \frac{\sum_j \bar{P}_j^2}{(\sum_j \bar{P}_j)^2}. \quad (2.1.28)$$

This last parameter, Ψ , is often referred to as the telescope ‘sharpness’. The threshold value K is determined by the required FAP, such that:

$$\Phi(K) = 1 - \text{FAP}, \quad (2.1.29)$$

where $\Phi(\cdot)$ represents the Gaussian distribution function. Similarly, γ is the threshold determined by the required MDP, such that:

$$\Phi(\gamma) = \text{MDP}. \quad (2.1.30)$$

The final two parameters, T_A and b , are derived from properties of the optical system and target star. The ‘Airy Throughput’, T_A , is the starlight suppression system throughput (T) multiplied by the non-dimensionalized pixel area and the sum of the normalized PSF [141].

The parameter b is defined as:

$$b \triangleq \text{QE} v I_p(\lambda_r) \Delta\lambda A \quad (2.1.31)$$

where QE is the detector's quantum efficiency, v is the attenuation of optical elements up to the exit pupil, and A is the entrance aperture area. The term $I_p(\lambda_r) \Delta\lambda$ represents the total irradiance of the planet in the detection band (centered at λ_r with bandwidth $\Delta\lambda$). $I_p(\lambda_r)$ is the average irradiance in this band, and is calculated as

$$I_p(\lambda_r) = \mathcal{F}_0 10^{-(V_s + \Delta\text{mag})/2.5} \quad (2.1.32)$$

where V_s is the target star's apparent magnitude in the detection band and \mathcal{F}_0 is the band specific flux for a zero magnitude star (see Table 2.1). Alternatively, if a model spectrum is

Table 2.1: Zero Magnitude Flux Points. Data from Colina et al. [23]

Filter	Central Wavelength (nm)	FWHM (nm)	\mathcal{F}_0 (photons cm $^{-2}$ nm $^{-1}$ s $^{-1}$)
U	373.5	48.5	8160.25
V	444.3	83.1	14314.61
B	548.3	82.7	9798.73
R	685.5	174.2	6625.70
I	863.7	197.0	4082.74

available for the target star, this term can be replaced by the integral of the spectrum over $\Delta\lambda$.

Because the basic detection approach used here is one of Bayesian signal processing, the sampling of the image plane has a large effect on the required integration time. In particular, it is important to have a sufficiently small pixel size, and to use a large enough portion of the PSF in the signal analysis. Figure 2.6 shows the normalized integration times for constant values of b and T as a function of size of the PSF used and the pixel area for an open circular aperture. We see that there exist critical values for each of these beyond which

the integration time is nearly constant. In fact, as long as at least $1.5 \lambda/D$ of the PSF are sampled, changes in integration time due to different pixel sizes are relatively minor. The same effect is found when considering non-Airy PSFs.

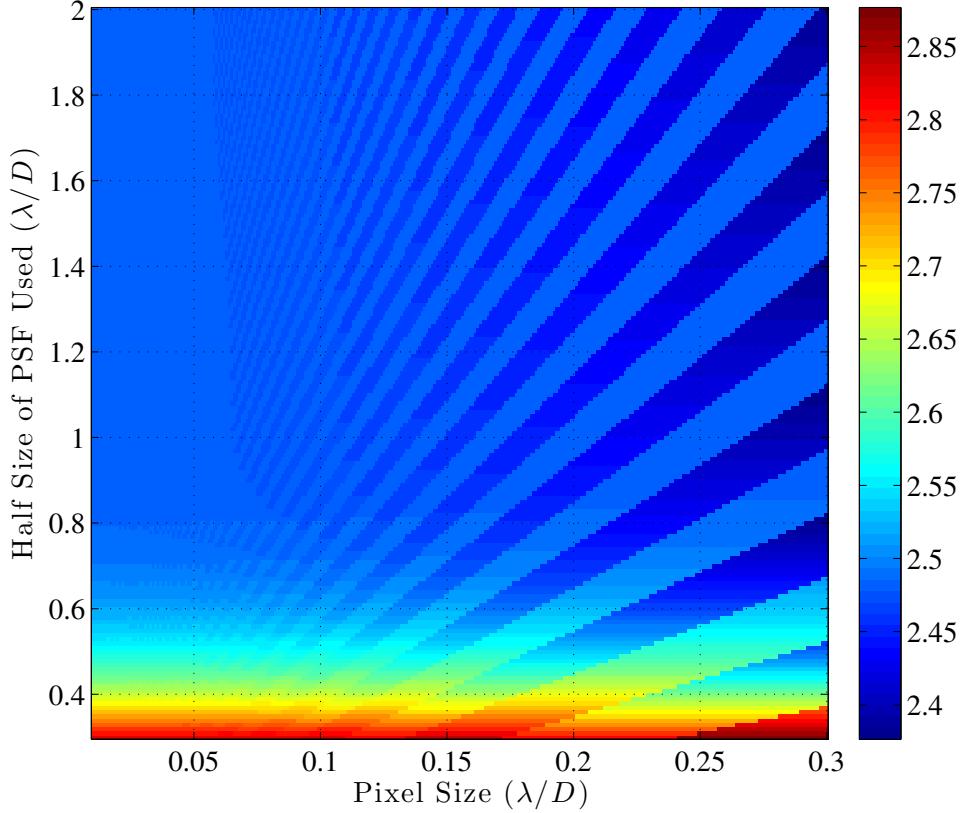


Figure 2.6: Normalized integration time ($\beta T t$) as a function of half size PSF used and pixel area for an open circular aperture. The color bar is log-scale, in powers of 10. Note that insufficient sampling of the PSF can significantly increase the required integration time.

To calculate the value of Q (defined in Equation (2.1.26)) we first define the value:

$$c_1 \triangleq \text{QE} v \Delta \lambda A T t . \quad (2.1.33)$$

This allows us to write:

$$C_p = \mathcal{F}_0 10^{-(V_s + \Delta \text{mag})/2.5} c_1 , \quad (2.1.34)$$

and

$$C_b = \mathcal{F}_0 10^{-\Omega_{zodi}/2.5} (f(\bar{\beta}) + \mu f(I) 2.5^{4.78-M_V}) \Delta\alpha c_1 + \mathcal{F}_0 10^{-V_s/2.5} C c_1 + DRt + \frac{\sigma_r^2}{t_r} t, \quad (2.1.35)$$

where C is the instrument's designed contrast, Ω_{zodi} is the detection band intensity of local zodiacal light in magnitudes per square arcsecond at the ecliptic pole (nominally, 23.34 for V band), M_V is the absolute magnitude of the target star, μ is the brightness of the extrasolar zodiacal light (exo-zodi) in units of local zodi, σ_r is the detector read noise, t_r is the exposure time per readout, DR is the detector dark current rate per pixel, and $\Delta\alpha$ is the pixel area on the sky in square arcseconds. The pixel area can be written as:

$$\Delta\alpha = \Omega \left(\frac{180 * 3600}{\pi} \right)^2, \quad (2.1.36)$$

where Ω is the solid angle of a detector pixel in steradians and equals $(\lambda/2D)^2$ for critically sampled systems with circular apertures [8]. The function f is the empirically derived variation of zodiacal light with viewing angle given by:

$$f(\theta) = 2.44 - 0.0403\theta + 0.000269\theta^2 \quad (2.1.37)$$

with θ in degrees, in the range $[0, 90]$ (the relation is mirrored for $\theta \in (90, 180]$). The function is applied to $\bar{\beta}$ - the absolute value of the ecliptic latitude of the target star, and I - the target star system's inclination* (D. Lindler, Personal Communication, 2008).

It should be noted that a major assumption being made here is that the exo-zodi is uniform and of constant magnitude in the region being observed. Recent results indicate that this is not an ideal assumption as exozodiacal dust will clump due to the gravitational effects of any planets and will always have color and brightness variations [74]. If we do select a single value for exo-zodi brightness, then it makes sense to use the historical average for our

*For multi-planet systems, this refers to the inclination of the ecliptic plane.

own solar system. Based on measurements of seafloor sediment isotope concentrations, the historical distribution of zodiacal dust in the solar system (over the last 80 Myr) is roughly log normal, with a mean of 1.55 zodi (Figure 2.7). It is predicted that 95% of solar analogs will have zodiacal dust levels in the range of 0.7 to 3.3 zodi, under the assumption that their asteroid and comet populations are similar to those of our solar system [73].

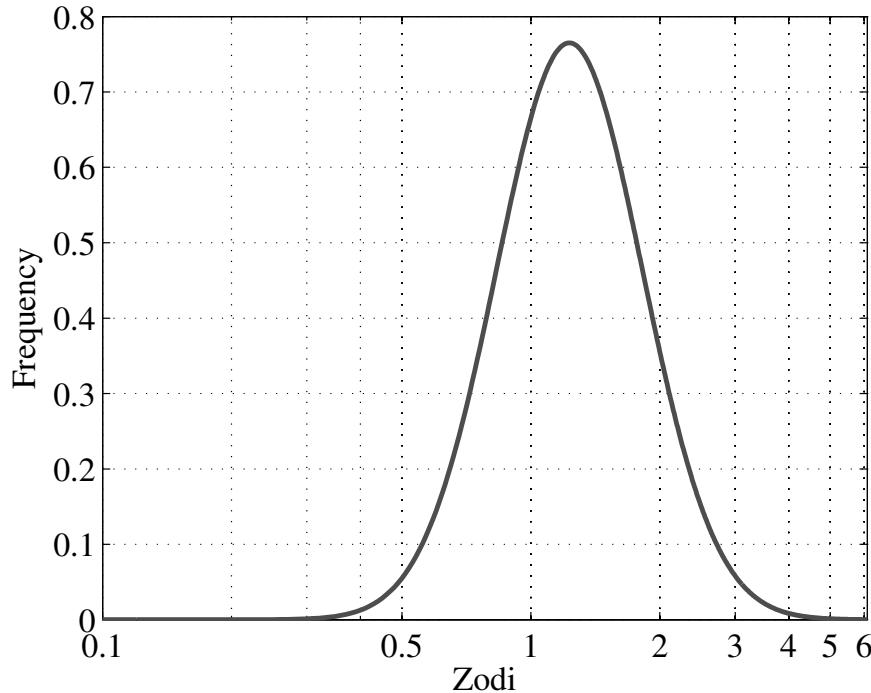


Figure 2.7: Historical distribution of solar system zodiacal dust levels. The distribution is log-normal with geometric mean of 1.55 [73].

Returning to Equations (2.1.34) and (2.1.35) we evaluate Q as:

$$Q = \left[10^{\Delta\text{mag}/2.5} C + 10^{(V_s + \Delta\text{mag} - \Omega_{zodi})/2.5} Z \Delta \alpha + \frac{10^{(V_s + \Delta\text{mag})/2.5}}{\mathcal{F}_0 \text{QE} v \Delta \lambda AT} \left(\text{DR} + \frac{\sigma_r^2}{t_r} \right) \right]^{-1}, \quad (2.1.38)$$

where

$$Z = (f(\bar{\beta}) + \mu f(I) 2.5^{4.78 - M_V}). \quad (2.1.39)$$

In the event of a detection, we calculate the time required for spectral characterization, assuming that the detection integration provides us the magnitude of all background sources.

Following [80], we will represent spectral characterization requirements as minimum signal to noise (S/N) values, using the metric

$$\text{S/N} = \frac{C_p}{\sqrt{V}} \quad (2.1.40)$$

where V is the total variance due to all noise sources. This value can be written as a function of the square root of the integration time as

$$\text{S/N} = \frac{\tilde{C}_p \sqrt{t}}{N_{pix} \left(\left(\frac{\sigma_r^2}{t_r} + \text{DR} \right) \left(1 + \frac{1}{N_{dark}} \right) + \mathcal{F}_0 10^{-\Omega_{zodi}/2.5} Z \Delta \alpha \right) + \tilde{C}_p + \tilde{C}_s} \quad (2.1.41)$$

where \tilde{C}_p and \tilde{C}_s are the planet and star counts per unit time, N_{dark} is the number of dark frames used, and N_{pix} is the number of pixels in the detection box[†].

The specific S/N value used depends on the spectral feature of interest. For terrestrial planets, we are especially interested in the possible discovery of biomarkers such as oxygen or ozone. Several of these features are identified in Heap [52], along with their S/N requirements; for example, a S/N of 11 for a resolving power of $R = 70$ will produce detections of oxygen at Earth atmospheric levels (21%) via the feature at 760nm with a confidence level of over 99.9999%. For 20% Earth O₂ abundance, the same signal to noise will still give confidence levels of over 99% [30].

2.2 Astrometry and Doppler Spectroscopy

While astrometry and doppler spectroscopy employ different technologies and methods to detect planets, they share one thing in common: namely, these methods infer the presence of planets by tracking the effects of gravitational interactions on the motions of the stars

[†]This value is often set to the reciprocal of the sharpness (i.e., $N_{pix} = \Psi^{-1}$) [80].

(known as the ‘astrometric wobble’). The data collected by doppler spectroscopy describes the motion of the target star along the line of sight (the radial velocity), while the data collected by precision astrometry tracks the star’s position in the plane of the sky. Both of these methods therefore require enough observations to fit planetary orbits, which means that the required observing time for confirmed discoveries is directly proportional to the period of planets discovered. For this reason, there is a strong bias towards short period planets in the early radial velocity data sets [15, 25, 26]. Similarly, since more massive planets exert a greater effect on their parent stars, these methods are also biased towards larger planets.

2.2.1 Astrometric Observables and Constraints

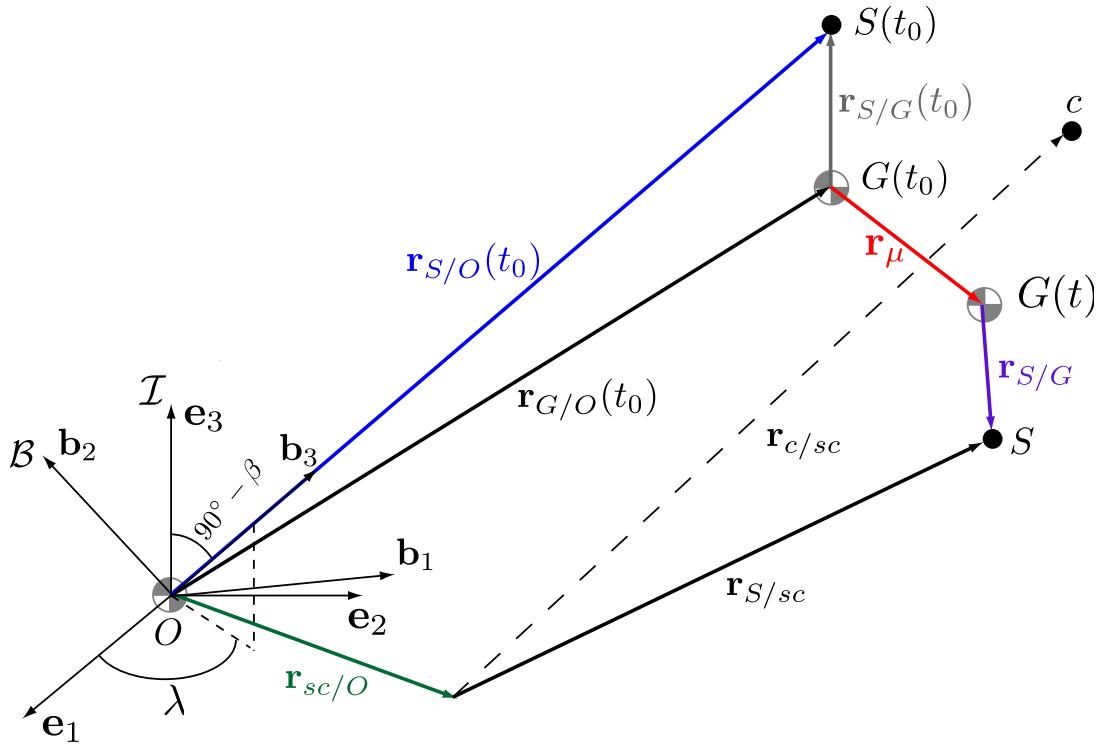


Figure 2.8: A schematic of an astrometric observation. The solar system barycenter is placed at the origin of frame \mathcal{I} , while the observed star system’s barycenter at epoch t_0 is at $G(t_0)$ and at $G(t)$ at observation time t . S represents the star position at time t and point c represents the (possibly time-varying) position of the centroid of a group of reference stars.

Figure 2.8 defines the vectors and reference frames used to describe an astrometric obser-

vation of an exosystem. Reference frame \mathcal{I} is the inertially fixed, barycentric/ ecliptic frame located at the solar system barycenter with the unit vector \mathbf{e}_3 directed perpendicular to the ecliptic plane (the \mathbf{e}_1 and \mathbf{e}_2 unit vectors are arbitrary). We define a second inertially fixed frame, called the tangent frame, \mathcal{B} , with \mathbf{b}_3 axis aligned with $\hat{\mathbf{r}}_{S/O}(t_0)$, the unit vector to the star at the reference time t_0 . The \mathbf{b}_1 axis is perpendicular to the plane containing $\mathbf{r}_{S/O}(t_0)$ and \mathbf{e}_3 . The final unit vector direction, \mathbf{b}_2 , is mutually perpendicular to \mathbf{b}_1 and $\mathbf{r}_{S/O}(t_0)$. Given these definitions, we can express the three unit vectors defining \mathcal{B} using coordinates in \mathcal{I} as:

$$\begin{aligned}\hat{\mathbf{r}}_{S/O}(t_0) \equiv \mathbf{b}_3 &= \left[\cos \lambda \cos \beta \quad \sin \lambda \cos \beta \quad \sin \beta \right]_{\mathcal{I}}^T \\ \mathbf{b}_1 &= \left[\begin{array}{ccc} 0 & 0 & 1 \end{array} \right]_{\mathcal{I}}^T \times \frac{\hat{\mathbf{r}}_s(t_0)}{\cos \beta} = \left[-\sin \lambda \quad \cos \lambda \quad 0 \right]_{\mathcal{I}}^T \\ \mathbf{b}_2 &= \hat{\mathbf{r}}_{S/O}(t_0) \times \mathbf{b}_1 = \left[-\cos \lambda \sin \beta \quad -\sin \lambda \sin \beta \quad \cos \beta \right]_{\mathcal{I}}^T\end{aligned}\tag{2.2.1}$$

where $\hat{\mathbf{r}}_{S/O}(t_0)$ is the unit vector to the star as given by its ecliptic coordinates at epoch t_0 : (λ, β) , measured in the \mathcal{I} frame. This definition of frames assumes constant relative velocities between exosystem and local barycenters (i.e., constant proper motions). For the time scales on which astrometric observations are taken, this is reasonable, but this derivation may have to be expanded if dealing with an accelerating exosystem.

We define the parallax by the small quantity:

$$\varpi \triangleq \frac{a_0}{\|\mathbf{r}_{S/O}(t_0)\|},\tag{2.2.2}$$

where a_0 is a distance constant such that ϖ is in units of radians. Thus, $\mathbf{r}_{S/O}(t_0) = (a/\varpi)\hat{\mathbf{r}}_{S/O}(t_0)$ for $a_0 = 1$ AU, with $\mathbf{r}_s(t_0)$ expressed in AU. An estimate of ϖ provides the distance to the target star at epoch t_0 . The motion of the barycenter of the target star

system, \mathbf{r}_μ , is considered as motion in the tangent reference frame,

$$\mathbf{r}_\mu(t) = \sigma_x(t - t_0)\mathbf{b}_1 + \sigma_y(t - t_0)\mathbf{b}_2 + \sigma_z(t - t_0)\hat{\mathbf{r}}_s(t_0), \quad (2.2.3)$$

where σ_i are the components of barycenter velocity at epoch t_0 . We approximate this velocity to be constant, as is usually done when considering short time spans such as a space-based observatory's lifetime. This expression is generally split into two components: the transverse and radial velocities. Following Green [44], we can write:

$$\begin{aligned} \frac{d}{dt}\mathbf{r}_{G/O}(t) &\equiv \frac{d}{dt}(\mathbf{r}_{G/O}(t_0) + \mathbf{r}_\mu(t)) = \frac{d}{dt}\mathbf{r}_\mu(t) \\ &= V_R\hat{\mathbf{r}}_{S/O}(t_0) + \mathbf{V}_T \end{aligned} \quad (2.2.4)$$

where $\mathbf{V}_T = \hat{\mathbf{r}}_{S/O}(t_0) \times \left(\frac{d}{dt}\mathbf{r}_{G/O}(t) \times \hat{\mathbf{r}}_{S/O}(t_0) \right).$

In our notation, V_R is just σ_z , whereas \mathbf{V}_T is the vector $\begin{bmatrix} \sigma_x & \sigma_y & 0 \end{bmatrix}^T$. The transverse velocity (multiplied by time) thus gives the proper motion of the barycenter (motion in the plane of the sky), which has the largest effect on the position of the target system at the time of observation. However, the radial velocity also has a measurable effect on the astrometric observation. Because motion of the target system in the radial direction causes the distance to the system to change, we observe a difference in the direction to the target star, known as ‘perspective acceleration’. The third component of \mathbf{r}_μ (and thus σ_z) interacts in non-negligible ways with other values in the measurement, and is thus observable. However, when dealing with *a priori* estimates of these velocity components it is important to note that the transverse and radial velocities are measured in different ways (i.e., astrometry vs. doppler spectroscopy). We will also find it convenient to have an expression for the normalized barycenter motion,

$$\bar{\mathbf{r}}_\mu \triangleq \frac{\mathbf{r}_\mu}{\|\mathbf{r}_{S/O}(t_0)\|} = \bar{\sigma}_x(t - t_0)\mathbf{b}_1 + \bar{\sigma}_y(t - t_0)\mathbf{b}_2 + \bar{\sigma}_z(t - t_0)\hat{\mathbf{r}}_{S/O}(t_0), \quad (2.2.5)$$

where $\bar{\sigma}_x$, $\bar{\sigma}_y$, and $\bar{\sigma}_z$ are the nondimensional barycenter velocities in radians/time unit.

Finally, we must consider what exactly is measured during an astrometric observation. While we will focus on astrometric methods that employ interferometers, it is important to note that there are other astrometric instrument designs. All of these, however, make some measurement of the observer-target star unit vector, and so the analysis above is fully applicable. An interferometer, in particular, measures the projection of a target direction onto the instrument's baseline vector, recorded as the optical path-length delay (OPD) between the two interferometer detectors:

$$\text{OPD} = \mathbf{B} \cdot \hat{\mathbf{r}}_{S/sc} + k + n, \quad (2.2.6)$$

where \mathbf{B} is the orientation vector of an interferometer of baseline length $B = \|\mathbf{B}\|$, k is a constant term representing the offset of the optical path differences, and n is the measurement noise. Planet-finding is generally proposed in a narrow angle mode, where the measurement is the relative OPD between two sources in a field of view, performed in quick succession such that k remains nearly constant, and is assumed to cancel in subtraction. Typically each target star has multiple reference sources, so the combined differential OPD (ΔOPD) is with respect to a centroid position. The measurement is also repeated for two (preferably orthogonal) interferometer baseline orientations to track the 2D position of the target in the plane of the sky. Assuming that the interferometer baselines are taken as directions \mathbf{b}_1 and \mathbf{b}_2 of our body frame, then the astrometric measurement becomes:

$$\mathbf{d} = B \begin{bmatrix} \mathbf{b}_1 \cdot (\hat{\mathbf{r}}_{S/sc} - \hat{\mathbf{r}}_{c/sc}) \\ \mathbf{b}_2 \cdot (\hat{\mathbf{r}}_{S/sc} - \hat{\mathbf{r}}_{c/sc}) \end{bmatrix} + \mathbf{n}, \quad (2.2.7)$$

where B is the size of the interferometer baseline and \mathbf{n} is an additive noise vector due to measurement error.

In the literature, the measurement in Equation (2.2.7) is often described as the angular

separation between the two sources, projected onto the baseline [72, 126, 127]; it is assumed that \mathbf{d} scaled by B is a radian measure that can be converted to other angular units such as arcseconds. It is this assumption that leads to the terminology of μas -precise astrometry, as the required sensitivity of the instrument to changes in the OPDs, when treated as an angle, evaluates to under 1 μas .

In fact, Colavita [22] points out that, using the definition for $\hat{\mathbf{r}}_{S/sc}$ in Equation (2.2.1), for a centroid separation of $(\Delta\lambda, \Delta\beta)$, the difference in unit vectors, to first order, can be written as,

$$\hat{\mathbf{r}}_{S/sc} - \hat{\mathbf{r}}_{c/sc} \approx \begin{bmatrix} \sin \beta \cos \lambda \Delta\beta + \cos \beta \sin \lambda \Delta\lambda \\ \sin \beta \sin \lambda \Delta\beta - \cos \beta \cos \lambda \Delta\lambda \\ -\cos \beta \Delta\beta \end{bmatrix}_{\mathcal{T}}. \quad (2.2.8)$$

In this way, knowledge of the baseline vector and the differential OPD allows you to calculate a vector which maps in a relatively simple fashion (to first order) to two spherical angles representing the separation between the target and centroid.

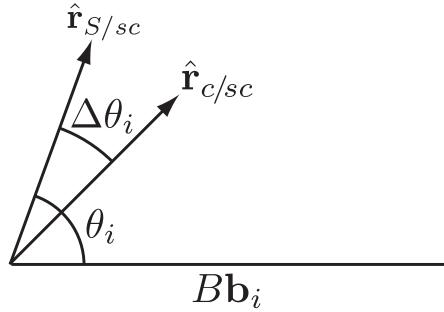


Figure 2.9: A schematic of a narrow-angle astrometric measurement. B is the size of the interferometer baseline.

Unfortunately, such a direct first order mapping between Equation (2.2.7) and the angle between the target and centroid is not very accurate. An OPD has units of distance or time (with the converting factor equal to the speed of light), and so a differential OPD will also have values equal to fractions of the interferometer baseline. Let us assume that the (flat) wavefront from the target star is incident on the interferometer baseline with angle θ_i , and

that the centroid is located at an angle $\Delta\theta_i$ from the target star, in the projection of the sky due to interferometer orientation \mathbf{b}_i (Figure 2.9 — $\Delta\theta_i$ is analogous to $\Delta\lambda$ and $\Delta\beta$ in Equation (2.2.8)). The components of the measurement (without the noise term) are then,

$$d_i = B (\cos \theta_i - \cos(\theta_i - \Delta\theta_i)) = B (\cos \theta_i(1 - \cos \Delta\theta_i) - \sin \theta_i \sin \Delta\theta_i) . \quad (2.2.9)$$

We are primarily interested in changes in the angle $\Delta\theta_i$, since this determines the movement of the target star with respect to the centroid in time (of course, this is further complicated by possible motion of the centroid itself). Assuming this to be a small angle, we can substitute the Taylor series expansions of the sine and cosine terms in $\Delta\theta_i$ to first order to find,

$$d_i \approx -B \sin \theta_i \Delta\theta_i . \quad (2.2.10)$$

Thus, by scaling the differential OPD by the interferometer baseline length and the direction on the sky of the target ($B \sin \theta_i$), we do get a first order approximation of the angular difference between the target and centroid (assuming perfect a priori knowledge of the target's (or equivalently centroid's) location with respect to the interferometer orientation). Unfortunately, if we assume the target star to be on the order of 1° from the centroid [126, 119], the first of the dropped terms ($\Delta\theta_i^2/2$) has a magnitude of 1.5×10^{-4} rad, or 31.4 arcseconds. When discussing ultra-precise applications, it is therefore inaccurate to treat the values produced by Equation (2.2.7) as directly mapping to angular measures. For these reasons, we treat all derived values as dimensionless for the remainder of this discussion, normalizing all distances and converting all angles to radians to remain consistent. It is also important to point out that in this discussion, we consistently assume zero pointing error. For a real instrument, accurate knowledge of the baseline vectors will be built up over many individual observations, each with an associated measurement error, but neither the interferometer orientation, nor the exact length of the baseline, can ever be known with perfect precision. In order to include a pointing and baseline length error, we would need to extend

the measurement equation, and update all subsequent calculations.

With these definitions and considerations, we can write the exact astrometric measurement in terms of known quantities and the quantities whose values we wish to determine. This is done by writing the vector from the spacecraft to the target star, $\mathbf{r}_{S/sc}$, in terms of the initial reference vector, $\mathbf{r}_{S/O}(t_0)$,

$$\begin{aligned}\mathbf{r}_{S/sc} &= \mathbf{r}_{S/O}(t_0) - \mathbf{r}_{S/G}(t_0) + \mathbf{r}_\mu + \mathbf{r}_{S/G} - \mathbf{r}_{sc/O} \\ &= \mathbf{r}_{S/O}(t_0) + \mathbf{r}_\mu + \Delta\mathbf{r}_{S/G} - \mathbf{r}_{sc/O},\end{aligned}\quad (2.2.11)$$

where $\mathbf{r}_{sc/O}$ is the spacecraft position vector relative to the solar system barycenter and $\Delta\mathbf{r}_{S/G}$ is the difference in the star's position relative to G between t_0 and epoch. We can then find the unit vector in the direction of $\mathbf{r}_{S/sc}$,

$$\begin{aligned}\hat{\mathbf{r}}_{S/sc} &= \frac{\mathbf{r}_{S/sc}}{\|\mathbf{r}_{S/sc}\|} \\ &= (\mathbf{r}_{S/O}(t_0) + \mathbf{r}_\mu + \Delta\mathbf{r}_{S/G} - \mathbf{r}_{sc/O}) \times \\ &\quad \{ \mathbf{r}_{S/O}(t_0) \cdot \mathbf{r}_{S/O}(t_0) + \mathbf{r}_\mu \cdot \mathbf{r}_\mu + \Delta\mathbf{r}_{S/G} \cdot \Delta\mathbf{r}_{S/G} + \mathbf{r}_{sc/O} \cdot \mathbf{r}_{sc/O} \\ &\quad + 2\mathbf{r}_{S/O}(t_0) \cdot \mathbf{r}_\mu + 2\mathbf{r}_{S/O}(t_0) \cdot \Delta\mathbf{r}_{S/G} - 2\mathbf{r}_{S/O}(t_0) \cdot \mathbf{r}_{sc/O} \\ &\quad + 2\mathbf{r}_\mu \cdot \Delta\mathbf{r}_{S/G} - 2\mathbf{r}_\mu \cdot \mathbf{r}_{sc/O} - 2\Delta\mathbf{r}_{S/G} \cdot \mathbf{r}_{sc/O} \}^{-\frac{1}{2}}.\end{aligned}\quad (2.2.12)$$

The spacecraft to centroid pointing unit vector, $\hat{\mathbf{r}}_{c/sc}$, can be similarly expressed by using Equation (2.2.12) for each reference source with the appropriate values for $\mathbf{r}_{S/O}(t_0)$ and \mathbf{r}_μ . When using this model for the reference sources, it is important to note that for a source n , $\hat{\mathbf{r}}_n(t_0)$ is not aligned with $\hat{\mathbf{r}}_{S/O}(t_0)$ and is thus not orthogonal to \mathbf{b}_1 and \mathbf{b}_2 . When evaluating Equation (2.2.7), this simply means that all vectors have to be expressed as components in the same reference frame, which requires a change of coordinates for the reference sources. If we assume extragalactic references with no companions or planets, the variation in the centroid position will be many orders of magnitudes below our desired level of accuracy.

Unfortunately, such references are not always available, and so these effects may become significant if one or more of the references is within our galaxy and has companions of its own.

2.2.2 Doppler Spectroscopy Observables and Constraints

Unlike astrometry, which uses interferometers to find the instantaneous position of the star in the plane of the sky, doppler spectroscopy utilizes high-resolution spectrometers to measure the instantaneous velocity of the star along the line of sight. Calibration sources such as cathode lamps and absorption cells are used to produce highly accurate spectra [146]. Following Marcy and Butler [87] and Butler et al. [14], we model the observation as starlight passing through an absorption cell (e.g., iodine), which creates a combined spectrum of absorption bands from the star’s spectrum and the cell. This combined spectrum is passed through an echelle spectrometer [145] and dispersed onto the detector as a nearly linear function in wavelength.

Any shift in the stellar spectrum must be modeled as due to Doppler effects and instrumental effects including mechanical instabilities, thermal effects, and the alignment of the optical system. The observed spectrum can then be written as:

$$I_{obs}(\lambda) = \kappa [I_S(\lambda + \Delta\lambda_S)T_C(\lambda + \Delta\lambda_C)] \otimes \text{PSF} \quad (2.2.13)$$

where I_S is the stellar spectrum, $\Delta\lambda_S$ is the shift of the stellar spectrum, T_C is the transmission function of the absorption cell, and $\Delta\lambda_C$ is the shift of the transmission function. The constant κ is directly proportional to the integration time of the observation. The transmission function is measured directly with the use of another spectrometer (for example a Fourier Transform Spectrometer [28]), while the stellar spectrum may be derived from high signal-to-noise observations of the target star by means of PSF deconvolution [39]. The remaining parameters are determined by fitting the observed spectrum to Equation (2.2.13).

The Doppler shift is parameterized as $\Delta\lambda/\lambda$, where:

$$\Delta\lambda \triangleq \Delta\lambda_S - \Delta\lambda_C. \quad (2.2.14)$$

This can be related to the star's velocity via the relativistic Doppler equation:

$$\frac{\Delta\lambda}{\lambda} = \frac{\left(1 + \left(\frac{v}{c}\right)^2 \cos\theta\right) (1 + \rho_g)}{n \sqrt{1 - \left(\frac{v}{c}\right)^2}} - 1 \quad (2.2.15)$$

where n is the index of refraction of the air column over the spectrograph and ρ_g is the gravitational redshift of the starlight. The velocity v is defined as the magnitude of the velocity of the target star with respect to the observer ($v = \|\mathbf{v}_{S/sc}\|$) while θ is the angle between the line of sight and the star's velocity vector:

$$\cos\theta = \hat{\mathbf{r}}_{S/sc} \cdot \hat{\mathbf{v}}_{S/sc}. \quad (2.2.16)$$

If we take $\theta = 0$ Equation (2.2.15) becomes:

$$\frac{\Delta\lambda}{\lambda} = \frac{(1 + \rho_g)}{n} \sqrt{\frac{\left(1 + \frac{v}{c}\right)}{1 - \frac{v}{c}}} - 1. \quad (2.2.17)$$

If $v \ll c$, the Taylor expansion of this expression gives us:

$$\frac{\Delta\lambda}{\lambda} = \frac{(1 + \rho_g)}{n} \left(1 + \frac{v}{c} + \mathcal{O}\left(\left(\frac{v}{c}\right)^2\right) - 1\right) \approx \frac{v}{c}. \quad (2.2.18)$$

We thus have an approximate linear mapping of the doppler spectroscopy observables to the star's line of sight velocity. Current ground-based observations have precisions of order 1 m s^{-1} [15], although improvements in instrumentation and calibration may provide us with precisions of order 1 cm s^{-1} in the near future [82]. Together with the astrometric measurement, doppler spectroscopy provides a third spatial component, making the full

three-dimensional orbit observable, given that sufficient measurement accuracy is achieved.

2.3 Transit Photometry

Of all of the methods so far discussed, transit photometry has the largest resource requirements as it demands constant monitoring of the target star. It also represents the least likely event, since the orientation of the orbit has to be such that the planet will obscure the star at some point, from the point of view of the observer.

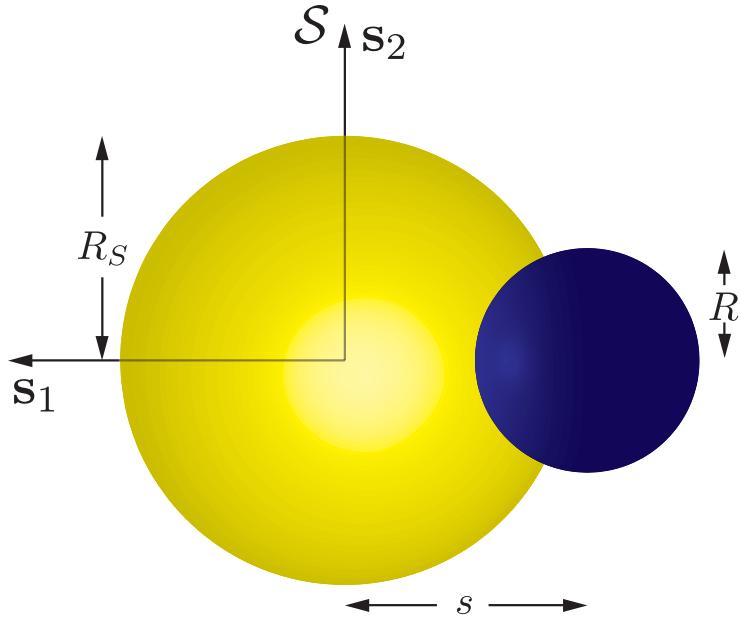


Figure 2.10: A schematic of a photometric transit observation. The amount of starlight occluded from the observer’s point of view is determined by the star and planet radii (R_S and R) and the apparent separation s .

Figure 2.10 presents a schematic view of a transit in the plane of the sky. The amount of starlight occluded can be expressed as a function of the star and planet radii and the apparent separation. Following Mandel and Agol [85], we can express the ratio of obscured

$(F_S^{(e)})$ to unobscured starlight, assuming a uniform source and an opaque eclipsing body, as:

$$\frac{F_S^{(e)}}{F_S} = \begin{cases} 1 & R_S + R < s \\ 1 - \frac{1}{\pi} \left[\frac{R^2}{R_S^2} \kappa_0 + \kappa_1 - \sqrt{\frac{s^2}{R_S^2} - \frac{(R_S^2 + s^2 - R^2)^2}{4R_S^4}} \right] & |R_S - R| < s \leq R_S + R \\ 1 - \left(\frac{R}{R_S} \right)^2 & s \leq R_S - R \\ 0 & s \leq R - R_S \end{cases} \quad (2.3.1)$$

where

$$\kappa_0 = \cos^{-1} \left(\frac{R^2 + s^2 - R_S^2}{2Rs} \right) \quad \text{and} \quad \kappa_1 = \cos^{-1} \left(\frac{R_S^2 - R^2 + s^2}{2R_S s} \right) \quad (2.3.2)$$

where R_S is the stellar radius.

These equations allow us to model the ‘dip’ in the photometric output of an exosystem associated with a transiting exoplanet, but there are many additional photometric effects that can be observed. Stars are not perfectly uniform sources, which produces limb darkening and causes more significant dimming during eclipse [85, 21]. Furthermore, effects due to atmospheric lensing and projected oblateness can be inferred from sufficiently precise transit photometry [59]. While all of these can improve our modeling of exosystems, they will not be explicitly considered in this thesis. However, as these effects may be modeled using the same parameter set employed throughout this chapter, they may be simulated and analyzed using the same methods that will be introduced in Chapters 3 and 4.

Chapter 3

Parameter Distributions and Detection Probabilities

3.1 Current State of Knowledge

The distributions of the physical and orbital parameters of exoplanets change rapidly as the pace of exoplanet discovery accelerates. As of this writing, this entire population is being transformed as the Kepler mission continuously introduces new candidates potentially doubling the total number of known exoplanets [4]. Multiple groups are maintaining and updating catalogues of exoplanets and are actively working on deriving the distributions of planetary and orbital parameters [86, 26, 92, 93]. One of the most important results of this work is the identification of a power-law trend in the relationship between projected mass ($m_P \sin I$) and period in radial velocity survey data. Figure 3.1 shows data from the Keck planet search, while Figure 3.2 shows the resulting power-law fit to the model:

$$dN = cm_P^\alpha P_{orb}^\beta d \log m_P d \log P_{orb}, \quad (3.1.1)$$

where c is a constant of proportionality. This work is very useful in generating sample planetary systems as in §4.2.1.

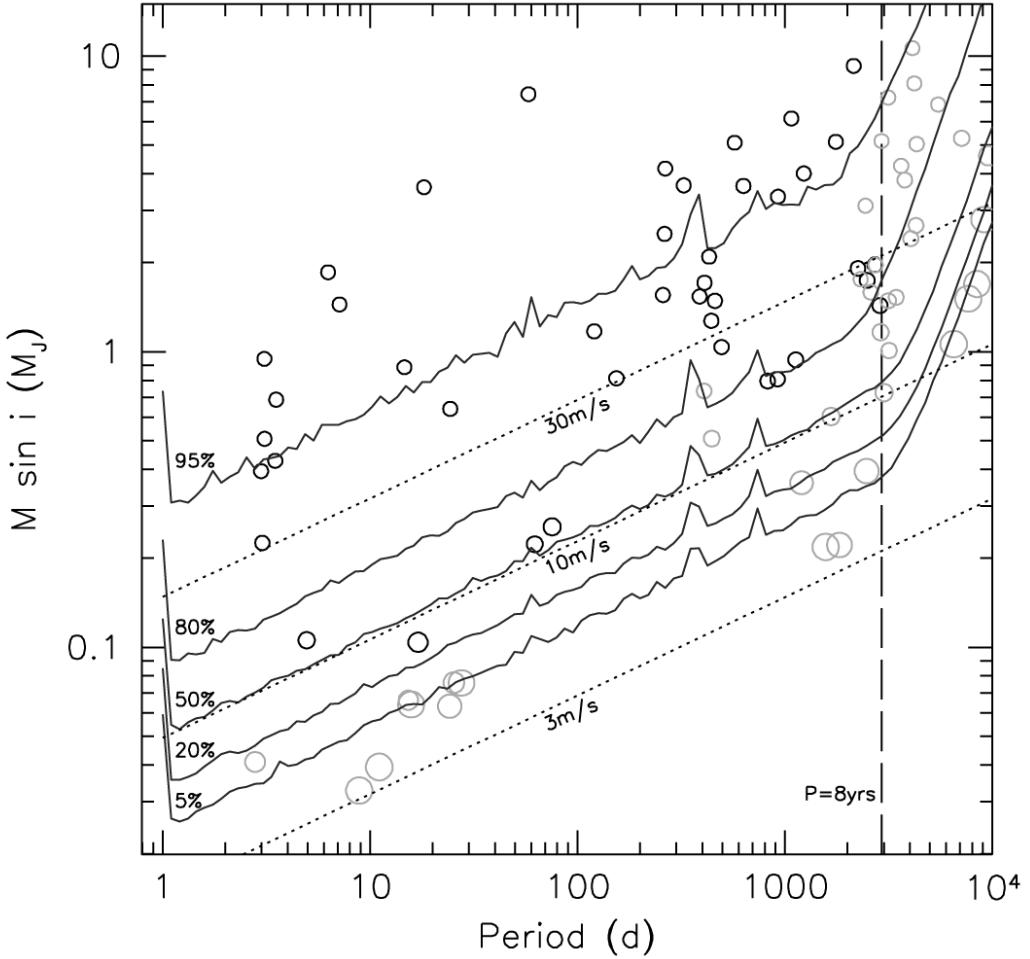


Figure 3.1: Detections for F, G, and K stars corrected for completeness of the survey. Black circles are confirmed planets circa May 2005 and gray circles are candidates with circle size indicating the magnitude of selection effects for each. The vertical dashed line indicates a period of 8 years and the dotted lines show velocity amplitudes of 3, 10, and 30 m/s for a $1 M_{\odot}$ star. For a given period, the solid curves show the mass that can be ruled out at the 99% level from 5%, 20%, 50%, 80%, and 95% of stars. From Cumming et al. [26].

Similarly, much work has been done to constrain the density distributions of known planets, relating planetary masses and radii. As shown in Fortney et al. [37], the densities of rocky planets can be modeled to a high degree of accuracy by assuming that the planets are composed of either a rock/ice or a rock/iron mixture. These assumptions lead to the

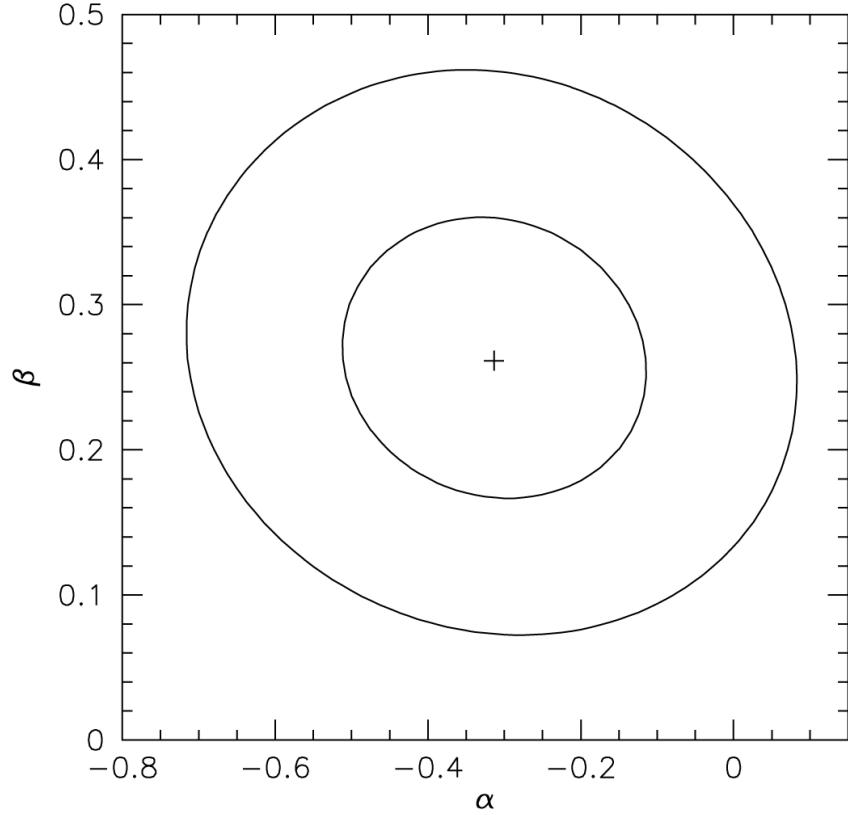


Figure 3.2: Contours of constant likelihood for a fit to Equation (3.1.1) . The contours correspond to the 68% and 95% confidence intervals with best-fit values of $\alpha = -0.31$ and $\beta = 0.26$. From Cumming et al. [26].

relations:

$$R = \begin{cases} (0.0912(1 - \text{rmf}) + 0.1603)(\log m_P)^2 \\ + (0.3330(1 - \text{rmf}) + 0.7387) \log m_P & \text{ice/rock} \\ + (0.4639(1 - \text{rmf}) + 1.1193) \\ (0.0592\text{rmf} + 0.0975)(\log m_P)^2 \\ + (0.2337\text{rmf} + 0.4938) \log m_P & \text{rock/iron} \\ + (0.3102\text{rmf} + 0.7932) \end{cases} \quad (3.1.2)$$

where the planetary radius is measured in R_{\oplus} , the planetary mass in M_{\oplus} and rmf is the fraction of rock. The resulting density curves are shown in Figure 3.3. For giant planets, radii values may be interpolated from a lookup table based on known values by assuming

the fraction of the planet's mass contained in the core [37]. Density modeling is especially important in simulating direct imaging as both the planet radius and planet mass must be defined to generate the observable quantities and to propagate the planet on its orbit.

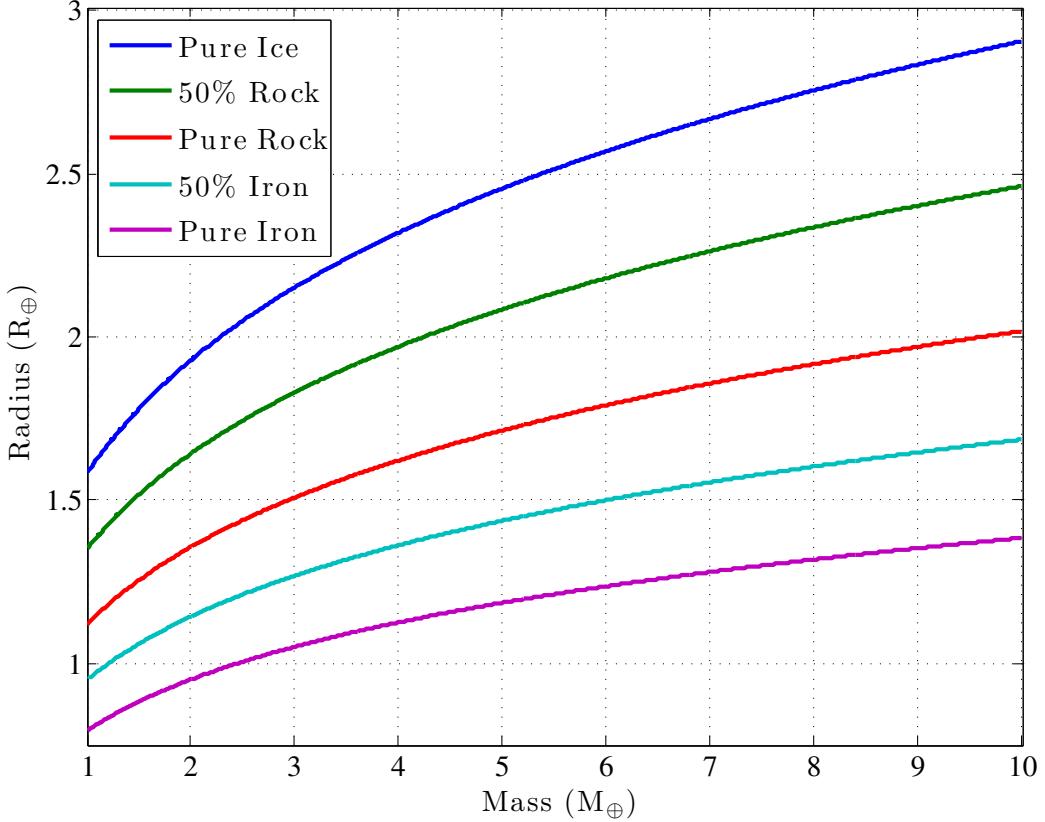


Figure 3.3: Densities of rocky planets modeled as rock/ice and rock/iron mixtures [37].

Another important observed correlation is between the occurrence of exoplanets and the metallicity of the host star. Surveying a population of FGK stars, Marcy et al. [86] found planets around approximately 25% of the most metal-rich stars (i.e., $[Fe/H] > 0.3$) and around only 3% of metal poor stars ($[Fe/H] < -0.5$). The occurrence of planets as a function of $[Fe/H]$ can also be fit to a power-law—gas giant detections are nearly proportional to the square of the number of iron atoms of the host star, with constant of proportionality of about 0.03 (Figure 3.4). This may be an important factor in the selection of a target list, especially as a final filter to pinpoint the highest detection probability targets (see §4.3).

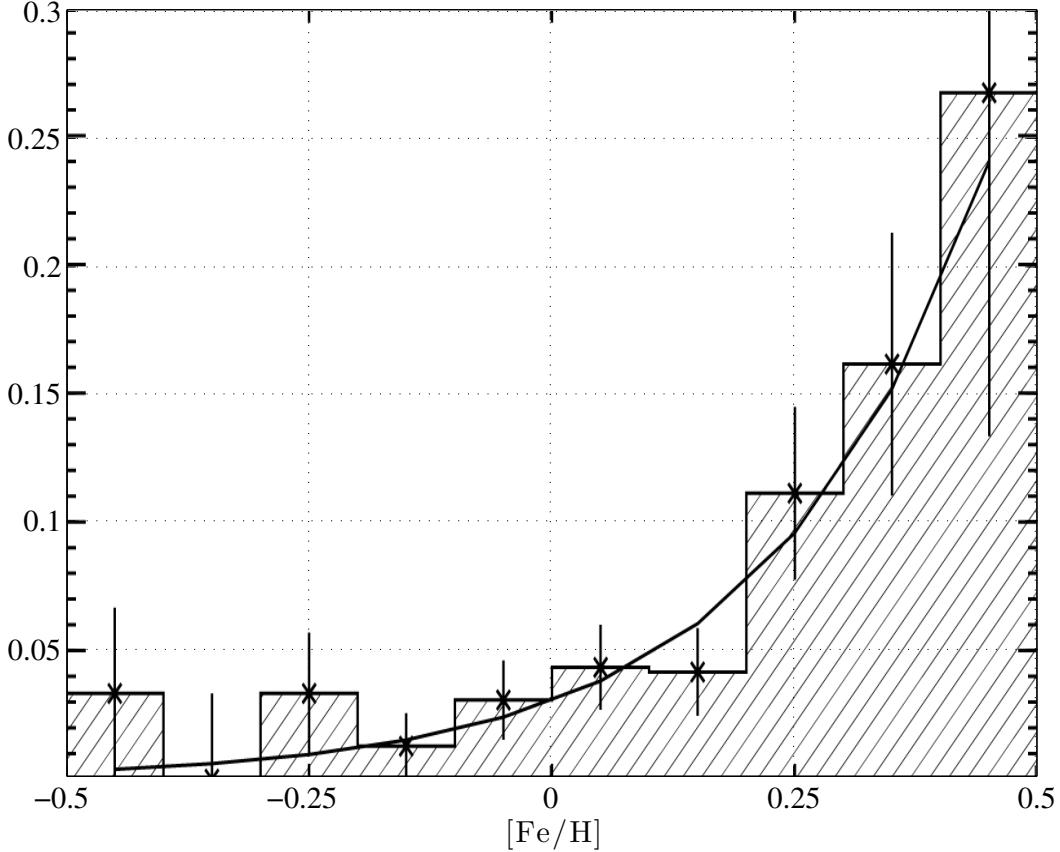


Figure 3.4: Portion of host stars with planets as function of iron abundance [Fe/H] star. Data from Marcy et al. [86].

3.1.1 Biases

If we plot a snapshot of the available orbital parameters for the currently known exoplanets, as in Figure 3.5, we immediately notice several prominent, and potentially misleading, features. There appears to be a large bias towards low eccentricity and semi-major axis. The low semi-major axes preference is due to the inherent bias in detection methods that require full (or even multiple) periods of data before a detection can be validated. These methods inherently detect more short period planets and as they include astrometry, radial velocity and transit photometry, which were used to find the majority of known exoplanets, the data set contains a disproportionately large number of short period (and thus low semi-major axis) planets. The low eccentricity accumulation is due to a specific population of planets:

tidally locked short period gas giants or ‘hot Jupiters’. These are easily detected due to their large size and short period, and all have near zero eccentricity due to orbital circularization.

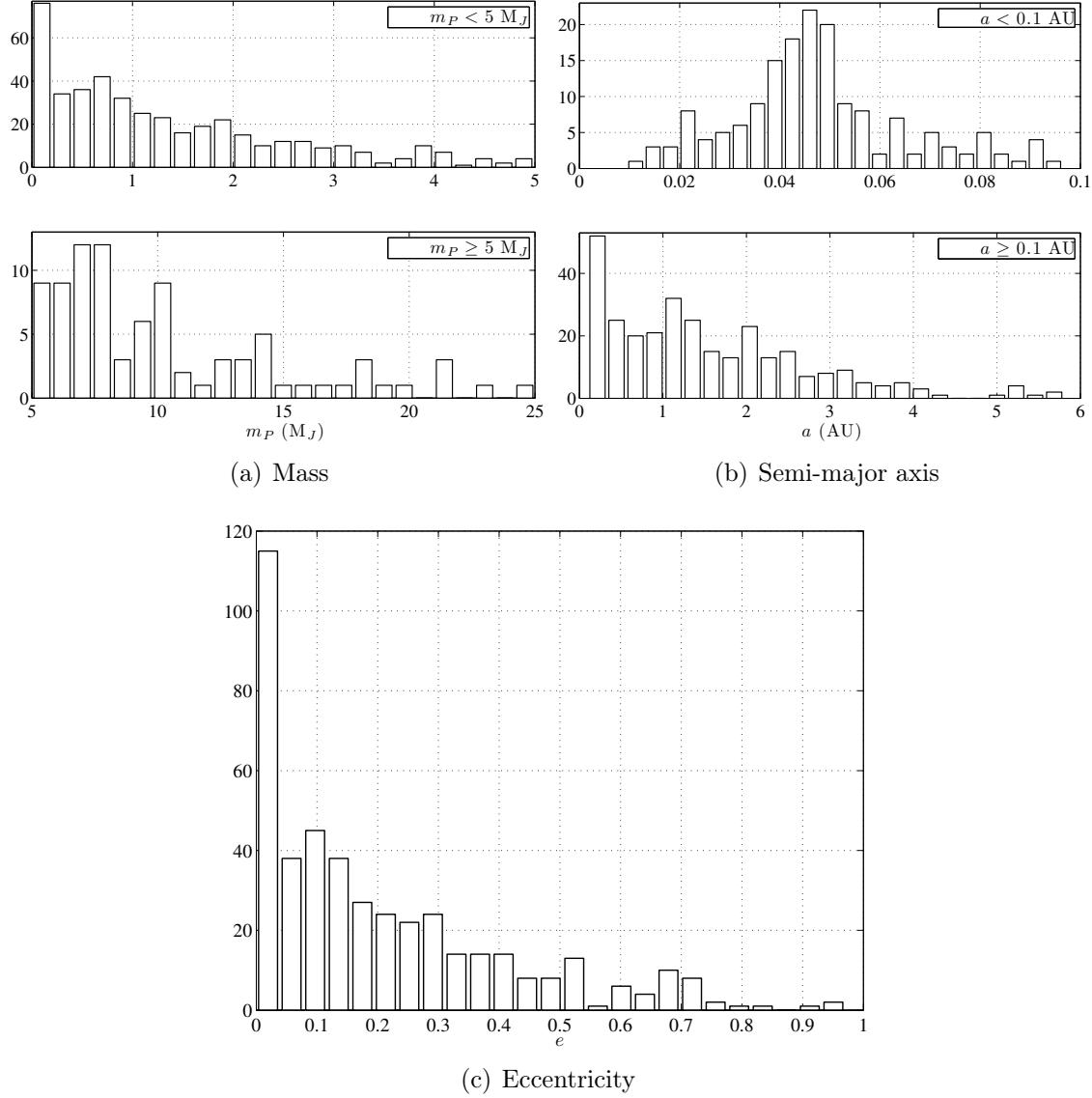


Figure 3.5: Histograms of orbital parameters of currently known exoplanets. Retrieved from the NASA/IPAC/NExSci Star and Exoplanet Database <http://nsted.ipac.caltech.edu> on April 17th, 2011.

Much work has been devoted to trying to find the true distributions of these parameters using both detailed modeling [27] and data analysis [57]. These efforts all seek to incorporate not just the raw data, but also the instrumental biases and the likelihoods of the individual detections. Another important point is to consider inherent biases in the parameters

themselves—for example, as eccentricity is defined within a strictly bounded range, it tends to be overestimated, introducing another source of bias. In the remainder of this chapter, we will consider how to model the likelihood of planetary detections, as well as the Keplerian elements used for describing planetary orbits.

3.2 Observational Completeness

Brown [7] introduced the concept of completeness to study the selection effects introduced by observatory architectures on direct detection searches for sub-stellar companions. Assuming distributions for the semi-major axis and eccentricity of planetary orbits, Brown calculated the probability that a companion’s angular separation would fall outside the telescope’s IWA during an observation of its parent star. Brown [8] subsequently expanded this concept to include the selection effects due to the telescope specific photometric restrictions on observability (i.e., Δmag_0), and Brown [9] demonstrated how completeness could be evaluated for indirect companion detection methods such as astrometry. Completeness has also been extended to account for multiple observations of one star at different times [11], and has been utilized in mission analysis and development for a variety of proposed exoplanet observatories [113, 9] as will be detailed in Chapter 4.

The direct detection completeness is evaluated by assuming that a companion will be observable if its angular separation from the star is greater than the observatory’s IWA, and illuminated such that the difference in brightness between star and companion is below the limiting Δmag . To calculate the completeness, probability distributions (or constant values) are assumed for planetary orbital elements and physical properties. A large, equal number of samples is generated from each distribution, and the star-planet angular separation and Δmag are calculated for each set of samples. When binned in a two-dimensional histogram, these generate a density function representing the probability that a planet drawn at random from the assumed population will have a given angular separation and Δmag (Figure 3.6).

Integrating over this density yields a cumulative distribution function (CDF). An instrument with a given Δmag_0 and IWA, observing a star with a planet from the assumed population, has a probability of detecting the planet upon its first observation equal to the corresponding CDF point (Figure 3.7).

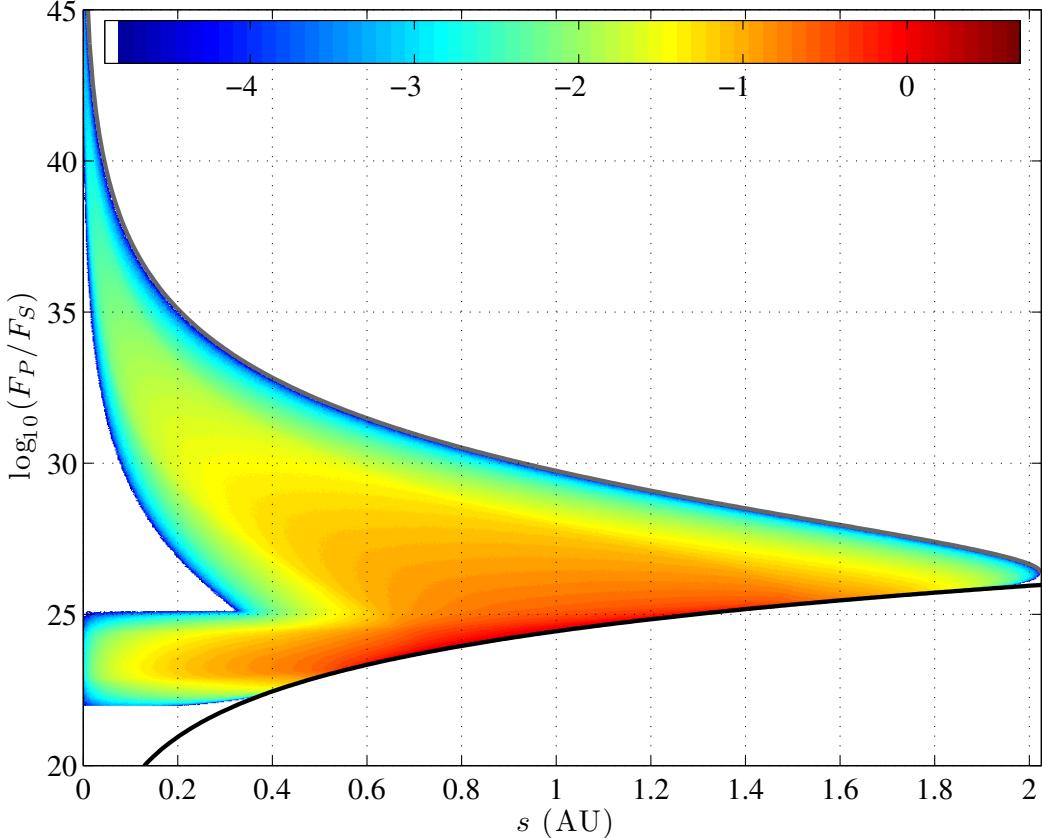


Figure 3.6: Observational and photometric completeness for a population of Earth-twin planets ($a \in [0.7, 1.5]$, $e \in [0, 0.35]$, $p = p_{\oplus}$, $R = R_{\oplus}$). The joint probability density function of ($\bar{s} = s$, $\bar{F}_P/F_S = F_P/F_S$) is sampled via 1 billion Monte Carlo trials, and the 2D histogram is constructed by counting the number of occurrences in each grid square scaled by total number of samples and the grid square area. The colorbar is log-scale, in powers of 10 and the bounding solid lines (black and gray) are calculated via Equations (2.1.22) and (2.1.24). The black line does not bound the histogram for all values of s because of the limits on a placed on the input distribution.

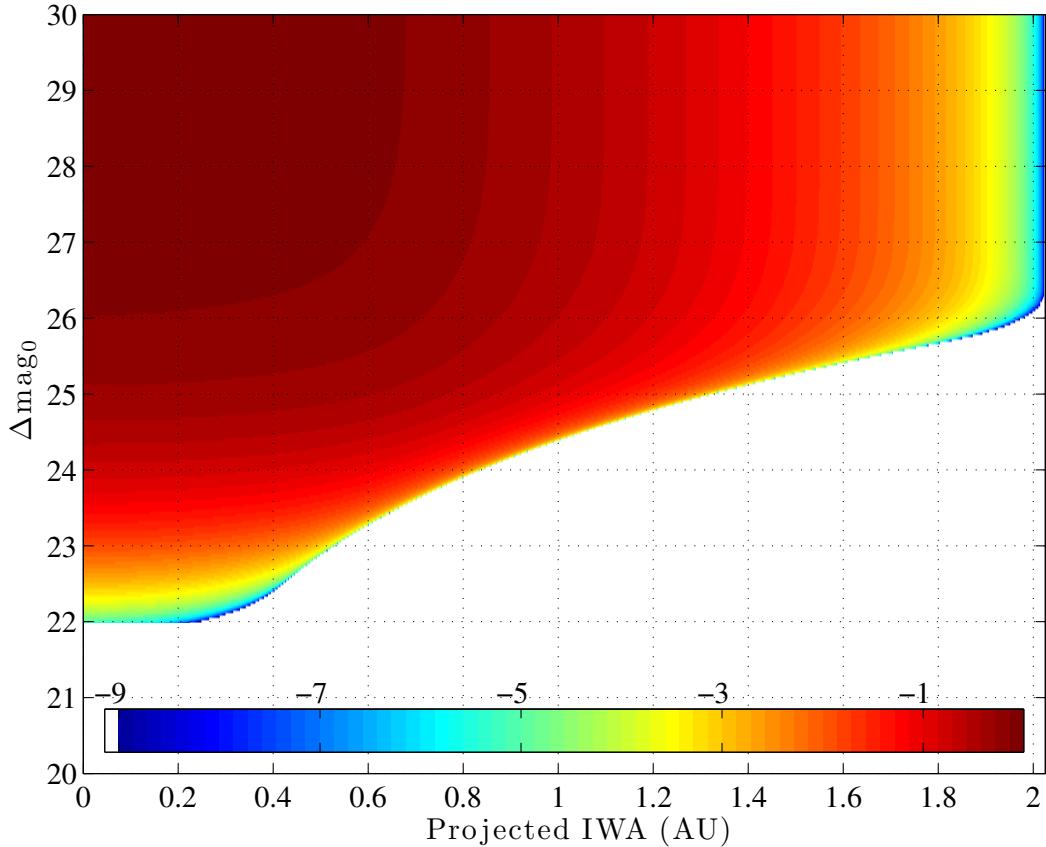


Figure 3.7: Integrated observational and photometric completeness from Figure 3.6. The colorbar is log-scale, in powers of 10, representing the cumulative distribution function of the random variable in Figure 3.6. The CDF is calculated by summing across the rows and columns of the (unscaled) PDF histogram at each grid square and scaling the result by the total number of samples. The histogram is truncated at Δmag_0 of 30, as the CDF becomes nearly constant for a given projected IWA for $\Delta\text{mag}_0 > 27$, as demonstrated by the vertical bands of nearly constant color. The CDF is unity at zero projected IWA and $\Delta\text{mag}_0 > 53$.

3.2.1 Applications

This definition of completeness describes the conditional detection probability (given the instrument capabilities and that a planet exists in the target system) only for the first target system observation. Subsequent observations of the same exosystem do not represent independent samples, and will thus have different detection probabilities. If each observation was fully independent of the others, with constant detection probability (p_c) equal to the single visit completeness, then the probability of k successful detections in n visits would be

given simply by the binomial theorem as:

$$P_n(k) = \binom{n}{k} p_c^k (1 - p_c)^{n-k}. \quad (3.2.1)$$

The probability of any (non-zero) number of detections in n visits would then be

$$P_n(k > 0) = 1 - \binom{n}{0} p_c^0 (1 - p_c)^n = 1 - (1 - p_c)^n. \quad (3.2.2)$$

In reality the probability of detecting a planet on subsequent visits is conditionally dependent on the planet's orbit and the times between observations, but this equation gives a good upper bound for the probability of detecting a planet (given that one exists) over multiple visits.

We can demonstrate this bound empirically by running Monte Carlo simulations of 1 million orbits in the habitable zone of a given star and tracking the aggregate number of detections over multiple visits. The results are shown in Figure 3.8. All of the planets are modeled as Earth-twins, and the instrument is assumed to have a limiting Δmag of 26. The star's luminosity and distance and instrument's IWA are chosen so that the projected IWA corresponds to the inner edge of the habitable zone (i.e., the projected IWA is at 0.70 AU for a star of $1 L_\odot$). The single visit completeness of this star for Earth-twins on habitable-zone orbits is approximately 0.5162. The plots show the results of returning every 1, 2, and 4 weeks, as well as daily revisits, randomly timed revisits, and the values of Equation (3.2.2). We assume that observations are instantaneous and do not add to the interim period between observations. In all cases, we eventually are able to observe over 99% of all planets and $P_n(k > 0)$ serves as an upper bound to the probability of planets found in n visits. It's also interesting to note that the randomized revisit interval most closely approximates the curve generated from Equation (3.2.2), giving the largest percentage of planets found in the smallest number of visits. This makes sense because randomization serves to weaken the dependence of subsequent visits on the time interval. In essence randomizing return times approximates the conceit of viewing a system over and over again 'for the first time'. Of

course, this breaks down after the initial few visits, and we again begin to perform more poorly than the bound.

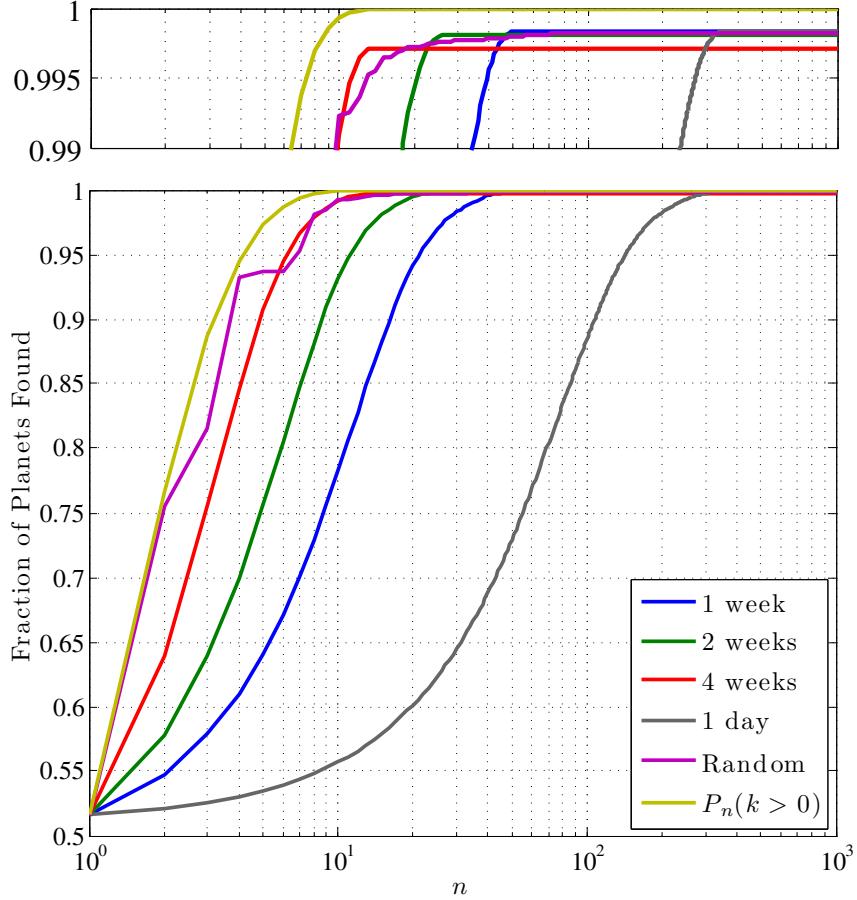


Figure 3.8: The fraction of planets found as a function of number of visits, with various intervals between observations, as well as the bounding probability given by Equation (3.2.2). The curve marked ‘Random’ represents randomized revisit intervals, uniformly distributed within one mean period of the population. The top panel is a zoomed view of the top 1% of the plot.

The exception to the bound given by Equation (3.2.2) would be the case of an optimal observing strategy (i.e., one where we only see previously unobservable portions of the habitable zone on each subsequent visit) on a fully observable system (i.e., one where all parts of the habitable zone are observable at some point). In the optimal scenario, a new portion of the habitable zone equal to the fraction represented by p_c is observed at each visit and the schedule is such that each portion is observed in sequence so no planets could be

bypassed (the equivalent of a continuous observation). If such an observing schedule could be achieved, the probability of detecting a planet after $1/p_c$ visits would be unity. For the majority of targets, however, it is quite difficult to follow the strict timing dictated by this strategy. Because target stars may only be observed for specific intervals of time at given times of the year*, it is often impossible to follow the optimal observing schedule in the limited period defined by the mission lifetime. It is because of these constraints that we often fail to detect existing planets even when the target system has been visited enough times to yield a high probability of detection. Equation (3.2.2) always holds as an upper bound for systems with permanently unobservable sections of the habitable zone (where the habitable zone is partially inside the projected IWA).

In the case of only one visit per target star, the calculation of overall detection probability is actually much simpler, and equals the sum of the targets' single visit completeness values. If all of the targets have the same completeness (fixed p_c), then the expected number of planets found (Y), given that each star has one ($\eta_{\oplus} = 1$), is:

$$E[Y] = \sum_{k=1}^n k \binom{n}{k} p_c^k (1 - p_c)^{n-k} = np_c . \quad (3.2.3)$$

Because Y is the sum of n Bernoulli random variables, each of which has expected value p_c , the expected value of Y must be np_c . If the target stars all have different single visit completeness values (non-constant p_c), the expected value of Y for n targets with completeness values of $\{p_i\}_{i=1}^n$ is:

$$E[Y] = \sum_{k=1}^n k \sum_{j \in {}_n C_k} \prod_{i \in j} p_i \prod_{i \notin j} (1 - p_i) = \sum_{i=1}^n p_i , \quad (3.2.4)$$

where ${}_n C_k$ is the set of combinations of the values from 1 to n taken k at a time.

*These observable time periods are known as ‘observing seasons’ and are determined by an observatory’s keep-out zone, which is discussed in §4.3

3.3 Analytical Distributions

The procedure described in the previous section for generating the completeness function is a Monte Carlo sampling of a bivariate distribution function of non-independent arguments (since star-planet separation and Δmag are functions of the same parameters). Thus in order to find any one point (or section) of the completeness distribution, it is necessary to sample it completely. Because of the relatively high dimensionality of the initial parameter space and wide range of values certain parameters can take, complete sampling requires a large number of Monte Carlo trials. For example, even with 100 million samples certain low probability areas of the function are significantly under-sampled, as demonstrated in Figure 3.9.

Any alternate method of sampling the completeness functions introduced in §3.2 requires at least some knowledge of their density functions. In particular, Markov chain methods such as Metropolis-Hastings [50] perform significantly better if the proposal distribution (a function used to ‘propose’ new samples that are then either accepted or rejected) closely approximates the target distribution. A special case of Metropolis-Hastings, known as Gibbs sampling, can be used to generate a sequence of samples from the joint distribution of two variables if their conditional probabilities are known. Our goal is to derive the distribution functions of the arguments to the completeness functions.

Following Savransky et al. [115], we start with an ensemble of Keplerian orbits whose orientation is uniformly distributed in space and derive the distribution functions of these orbits’ Keplerian parameters. We make no *a priori* assumptions as to the distribution of orbital semi-major axis and eccentricity, treating them instead as arbitrary functions. This allows us to derive the completely general distribution functions presented in §3.3.1. A key step here is the demonstration that the planetary phase angle (β) is independent of any of the orbital parameters. This makes it possible to write distribution functions for quantities directly related to the two parameters of the direct detection completeness function. It is important to note that, while the specific application explored here is direct imaging, these

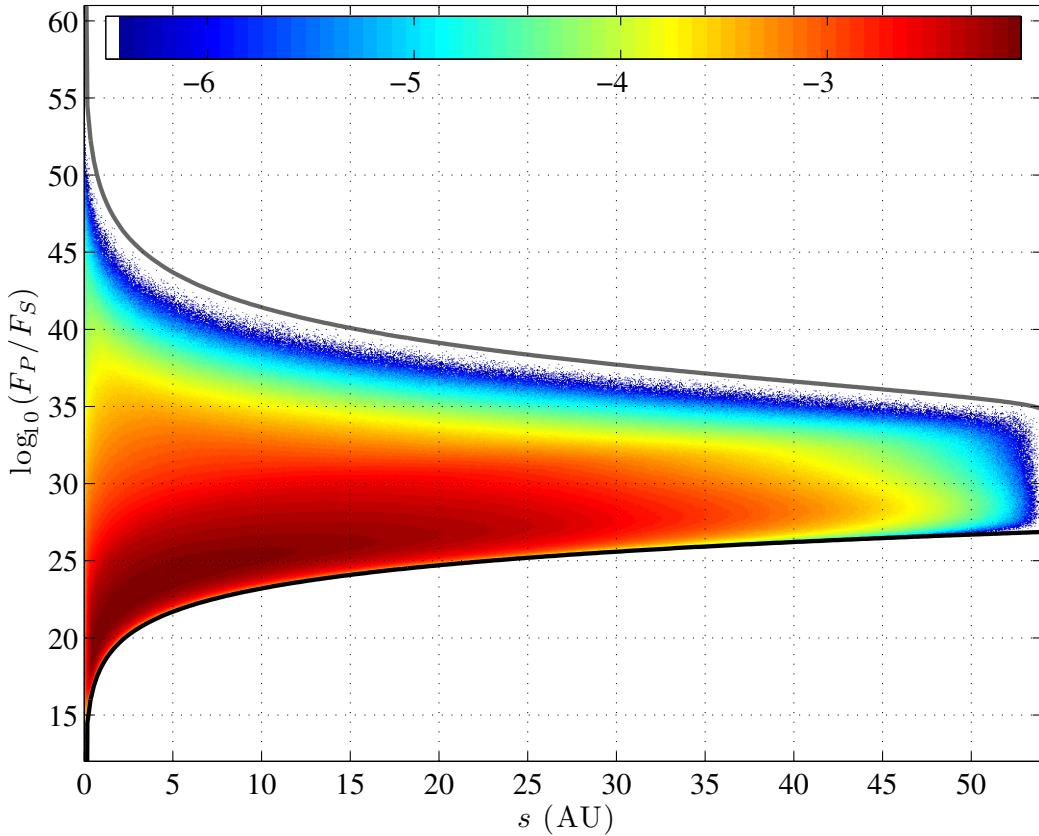


Figure 3.9: Observational and photometric completeness for a randomized planetary population ($a \in [0.4, 30]$, $e \in [0, 0.8]$, $p \in [0.1, 0.5]$, $R \in [0.7, 11.2]R_{\oplus}$). The colorbar values and bounding lines are calculated as in Figure 3.6. Note that the black line corresponds to high probability areas of the histogram and is thus a good bound, whereas the gray line corresponds to low probability areas and thus demonstrates the undersampling of the Monte Carlo procedure.

distributions are more broadly applicable to exoplanet studies in general. For example, Keplerian fits are often employed in doppler spectroscopy surveys, making these derivations useful for inferring the true distributions of orbital parameters derived from radial velocity data sets. Similarly, statistical analyses play an important role in other methods of exoplanet study, including transit photometry and microlensing surveys [43].

3.3.1 Derivation

The distributions of the orbit orienting Euler angles (see Equation (1.2.16)) are fixed by our assumption of uniform orbital orientation in space. The remaining two parameters, a and e , will be treated as unknowns, with probability density functions (PDFs) $f_{\bar{e}}(e)$ and $f_{\bar{a}}(a)$, respectively. Let \bar{M}, \bar{e} and $\bar{\nu}$ be random variables with the distributions of mean anomaly, eccentricity and true anomaly, respectively, and define two functions: $\nu = g(M, e)$ and $M = h(\nu, e)$. Following Larson and Shubert [78], we can write the CDF for $\bar{\nu}$ as a marginalization of a conditional probability,

$$\begin{aligned} F_{\bar{\nu}}(\nu) &= P[g(\bar{M}, \bar{e}) \leq \nu] = \int_{-\infty}^{\infty} P[g(\bar{M}, e) \leq \nu | \bar{e} = e] f_{\bar{e}}(e) de \\ &= \int_0^1 P[\bar{M} \leq h(\nu, e)] f_{\bar{e}}(e) de, \end{aligned} \quad (3.3.1)$$

where the last step assumes independence between \bar{M} and \bar{e} , and that the orbits are closed so that $e \in [0, 1]$. We note that:

$$P[\bar{M} \leq h(\nu, e)] = \int_0^{h(\nu, e)} f_{\bar{M}}(M) dM \quad (3.3.2)$$

where $f_{\bar{M}}(M)$ is the PDF of M :

$$f_{\bar{M}}(M) = \begin{cases} \frac{1}{2\pi} & M \in [0, 2\pi) \\ 0 & \text{else} \end{cases}. \quad (3.3.3)$$

Therefore the PDF of ν is:

$$f_{\bar{\nu}}(\nu) = \frac{d}{d\nu} F_{\bar{\nu}}(\nu) = \frac{1}{2\pi} \int_0^1 \frac{\partial h}{\partial \nu} f_{\bar{e}}(e) de = \frac{1}{2\pi} \int_0^1 \frac{(1 - e^2)^{\frac{3}{2}}}{(1 + e \cos \nu)^2} f_{\bar{e}}(e) de, \quad (3.3.4)$$

where h is given by equations (1.2.15) and (1.2.14).

Following the same procedure, but without substituting equation (1.2.15) into equation

(1.2.14), we find that the PDF of the eccentric anomaly is:

$$\begin{aligned} f_{\bar{E}}(E) &= \int_0^1 \frac{\partial}{\partial E} (E - e \sin(E)) f_{\bar{M}}(M) f_{\bar{e}}(e) de \\ &= \frac{1}{2\pi} \int_0^1 (1 - e \cos E) f_{\bar{e}}(e) de . \end{aligned} \quad (3.3.5)$$

Similarly, if we rewrite the Kepler equation as:

$$M = \cos^{-1} \left(\frac{a - r}{ea} \right) - e \sqrt{1 - \frac{(a - r)^2}{(ea)^2}} \quad (3.3.6)$$

and assume independence between \bar{a} and \bar{e} , we can write the PDF for the orbital radius as:

$$f_{\bar{r}}(r) = \frac{1}{\pi} \int_0^\infty \int_0^1 \frac{r}{a \sqrt{(ae)^2 - (a - r)^2}} f_{\bar{e}}(e) de f_{\bar{a}}(a) da . \quad (3.3.7)$$

Equations (3.3.4), (3.3.5), and (3.3.7) are the most complete descriptions possible for the distributions of the orbital position and anomaly, without assuming specific distributions for the eccentricity and semi-major axis.

We derived the distribution functions of the true anomaly and orbital radius by assuming independence between the Keplerian orbital elements from which the anomaly and radius are calculated. However, the quantities observed by direct searches (i.e., s and F_R) are nonlinear functions of the Keplerian elements and r , so we cannot make the assumption of independence between their functional arguments for either of these. Fortunately, inspection of the phase angle reveals something interesting. Using the approximation in equation (2.1.9), we can write:

$$\beta \approx \cos^{-1} (\cos \theta \cos \nu - \cos \psi \sin \theta \sin \nu) \quad (3.3.8)$$

from which we define a new variable,

$$x \triangleq \cos \beta = \cos \theta \cos \nu - \cos \psi \sin \theta \sin \nu . \quad (3.3.9)$$

If it can be shown that x is uniformly distributed for all distributions of eccentricity, then this would prove that β is sinusoidally distributed regardless of the distribution of any other orbital parameter, given only our assumption of uniform orbital orientation. We do so by considering the characteristic function of x ,

$$\varphi_{\bar{x}}(t) = \mathbb{E}(e^{it\bar{x}}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{itx} f_{\bar{\nu}}(\nu) f_{\bar{\psi}}(\psi) f_{\bar{\theta}}(\theta) d\psi d\theta d\nu, \quad (3.3.10)$$

where \mathbb{E} is the expectation value, and showing that it is equivalent to the characteristic function of a uniformly distributed variable.

The assumption of uniform orbital orientation yields:

$$f_{\bar{\psi}}(\psi) = \begin{cases} \frac{1}{2\pi} & \psi \in [0, 2\pi) \\ 0 & \text{else} \end{cases} \quad (3.3.11)$$

$$f_{\bar{\theta}}(\theta) = \begin{cases} \frac{\sin \theta}{2} & \theta \in [0, \pi) \\ 0 & \text{else} \end{cases} \quad (3.3.12)$$

so that the characteristic function of x becomes:

$$\varphi_{\bar{x}}(t) = \frac{1}{4\pi} \int_{-\infty}^{\infty} \int_0^{\pi} \int_0^{2\pi} e^{it(\cos \theta \cos \nu - \cos \psi \sin \theta \sin \nu)} d\psi \sin \theta d\theta f_{\bar{\nu}}(\nu) d\nu \quad (3.3.13)$$

Performing the integral over ψ ,

$$\varphi_{\bar{x}}(t) = \frac{1}{2} \int_{-\infty}^{\infty} \int_0^{\pi} e^{it \cos \theta \cos \nu} J_0(t \sin \theta \sin \nu) \sin \theta d\theta f_{\bar{\nu}}(\nu) d\nu \quad (3.3.14)$$

where J_0 is the zeroth-order Bessel function of the first kind. We next perform the integral

over θ using Gegenbauer's finite integral (see equation 12.14(1) in Watson [147]):

$$\begin{aligned} \int_0^\pi e^{it \cos \theta \cos \nu} J_{b-\frac{1}{2}}(t \sin \theta \sin \nu) C_a^b(\cos \theta) \sin^{b+\frac{1}{2}} \theta d\theta \\ = \sqrt{\frac{2\pi}{t}} i^a \sin^{b-\frac{1}{2}} \nu C_a^b(\cos \nu) J_{a+b}(t) \end{aligned} \quad (3.3.15)$$

where C_a^b is a Gegenbauer polynomial. Choosing $a = 0$ and $b = \frac{1}{2}$ yields $C_a^b(x) = 1$ and the integral becomes:

$$\int_0^\pi e^{it \cos \theta \cos \nu} J_0(t \sin \theta \sin \nu) \sin \theta d\theta = \sqrt{\frac{2\pi}{t}} J_{\frac{1}{2}}(t) \quad (3.3.16)$$

Since the half-order Bessel function is defined as:

$$J_{\frac{1}{2}}(t) = \sqrt{\frac{2}{\pi t}} \sin t \quad (3.3.17)$$

the characteristic function becomes:

$$\begin{aligned} \varphi_{\bar{x}}(t) &= \frac{1}{2} \int_{-\infty}^{\infty} \sqrt{\frac{2\pi}{t}} \sqrt{\frac{2}{\pi t}} \sin t f_{\bar{\nu}}(\nu) d\nu \\ &= \frac{\sin t}{t} \int_{-\infty}^{\infty} f_{\bar{\nu}}(\nu) d\nu. \end{aligned} \quad (3.3.18)$$

By definition,

$$\int_{-\infty}^{\infty} f_{\bar{\nu}}(\nu) d\nu = 1, \quad (3.3.19)$$

and thus:

$$\varphi_{\bar{x}}(t) = \frac{\sin t}{t}. \quad (3.3.20)$$

If we define a random variable $\bar{y} \sim U([-1, 1])$, its characteristic function will be:

$$\begin{aligned}\varphi_{\bar{y}}(t) &= \mathbb{E}(e^{it\bar{y}}) = \int_{-\infty}^{\infty} e^{ity} f_{\bar{y}}(y) dy \\ &= \frac{1}{2} \int_{-1}^1 e^{ity} dy = \frac{1}{2} \left(\frac{e^{it}}{it} - \frac{e^{-it}}{it} \right) \\ &= \frac{\sin t}{t}.\end{aligned}\tag{3.3.21}$$

As equations (3.3.20) and (3.3.21) are identical, \bar{x} and \bar{y} have the same characteristic function, and thus must have the same underlying distribution. Since \bar{x} is uniformly distributed on $[-1, 1]$, β is sinusoidally distributed in $[0, \pi)$ for any population of uniformly oriented orbits, independent of the specific distribution of \bar{e} and \bar{a} . This finding greatly simplifies our derivation of the distribution functions of the flux ratio and apparent separation.

Returning now to the definition of flux ratio in equation (2.1.6), and assuming that we are interested in a particular planet type such that pR^2 is approximately constant, we see that the flux ratio is a function of the product of two variables, $m = \Phi(\beta)$ and $n = r^{-2}$. If we define a random variable $\bar{k} \triangleq \bar{m}\bar{n}$, then it can be shown that:

$$f_{\bar{k}}(k) = \int_{-\infty}^{\infty} \frac{1}{n} f_{\bar{m}}\left(\frac{k}{n}\right) f_{\bar{n}}(n) dn\tag{3.3.22}$$

if \bar{m} and \bar{n} are independent [78]. As we have demonstrated the independence of $\bar{\beta}$ from \bar{r} , equation (3.3.22) is applicable. Using equation (3.3.7), the CDF of \bar{n} is given by:

$$\begin{aligned}F_{\bar{n}}(n) &= P[\bar{r}^{-2} \leq n] = P\left[-\frac{1}{\sqrt{n}} \leq \bar{r} \leq \frac{1}{\sqrt{n}}\right] \\ &= \int_{-1/\sqrt{n}}^{1/\sqrt{n}} \frac{1}{2\pi} \int_0^\infty \int_0^1 \frac{r}{a\sqrt{(ae)^2 - (a-r)^2}} f_{\bar{e}}(e) de f_{\bar{a}}(a) da dr \\ &= \frac{1}{2\pi} \int_0^\infty \int_0^1 \sqrt{(ae)^2 - (a-r)^2} + 2a \tan^{-1} \left(\sqrt{\frac{ae+a-r}{ae-a+r}} \right) \Big|_{-1/\sqrt{n}}^{1/\sqrt{n}} \\ &\quad \times f_{\bar{e}}(e) de \frac{f_{\bar{a}}(a)}{a} da.\end{aligned}\tag{3.3.23}$$

The PDF of n is thus:

$$f_{\bar{n}}(n) = \frac{1}{4\pi} \int_0^\infty \int_0^1 n^{-\frac{3}{2}} \left[(C_0 + 2a\sqrt{n} - 1)^{-\frac{1}{2}} - (C_0 - 2a\sqrt{n} - 1)^{-\frac{1}{2}} \right] \times f_{\bar{e}}(e) de \frac{f_{\bar{a}}(a)}{a} da \quad (3.3.24)$$

where $C_0 = a^2(e^2 - 1)n$. The PDF of \bar{m} is dependent on the exact form of Φ , but can also be evaluated analytically when Φ is invertible. In those cases,

$$f_{\bar{m}}(m) = f_{\bar{\beta}}(\Phi^{-1}(m)) \left| \frac{d}{dm} \Phi^{-1}(m) \right| \quad (3.3.25)$$

$$= \begin{cases} \frac{1}{2} \sin(\Phi^{-1}(m)) \left| \frac{d}{dm} \Phi^{-1}(m) \right| & \Phi^{-1}(m) \in [0, \pi) \\ 0 & \text{else} \end{cases}$$

This formulation is still problematic when dealing with transcendental functions such as the commonly used phase function of a Lambertian sphere (see Equation (2.1.18)):

$$\pi\Phi(\beta) = \sin\beta + (\pi - \beta)\cos\beta, \quad (3.3.26)$$

but, since the domain of the function is restricted to $\beta \in [0, \pi]$, the range of Φ is single-valued ($\in [1, 0]$), and the function is invertible. Expanding about $\pi/2$, equation (3.3.26) becomes:

$$\Phi(\beta) = \frac{1}{\pi} + \sum_{k=1}^{\infty} \alpha_k \left(\beta - \frac{\pi}{2} \right)^k \quad \text{where} \quad \alpha_k = \frac{d_k}{2k!} \left(\frac{2(k-1)}{\pi} \right)^{\frac{d_k+d_{k+1}}{2d_k}} \quad (3.3.27)$$

and $d_{1:3} = -1, 1, 1$, with $d_k = d_{k-1}d_{k-2}d_{k-3}$ for $k > 3$, such that you get the series $d_i = \{-1, 1, 1, -1, -1, 1, 1, -1, -1, \dots\}$. Following Morse and Feshbach [94], we can express the inverse function as the series:

$$\beta = \Phi^{-1}(w) = \frac{\pi}{2} + \sum_{k=1}^{\infty} b_k \left(w - \frac{1}{\pi} \right)^k \quad (3.3.28)$$

where

$$b_k = \frac{1}{k\alpha_1^k} \sum_{x \in X} (-1)^{|x|} \prod_{r=1}^{|x|} (k-1+r) \prod_{i=1}^{k-1} \frac{(\alpha_{i+1}/\alpha_1)^{x_i}}{x_i!} \quad (3.3.29)$$

and the space X of sets x is defined as:

$$X \triangleq \left\{ x \in \mathbb{N}^{k-1} : \sum_{i=1}^{k-1} ix_i = k-1 \right\} \quad \text{with} \quad |x| \triangleq \sum_i x_i. \quad (3.3.30)$$

The inverse series converges to machine double precision in a finite number of terms, except at the endpoints (0 and 1), which themselves do not need to be computed as they map exactly to $\beta = \pi$ and $\beta = 0$, respectively.

Since $\bar{F}_R = pR^2\bar{k}$, Equations (3.3.22) - (3.3.30) allow us to write:

$$\begin{aligned} f_{\bar{F}_R}(F_R) &= \frac{1}{pR^2} \int_{-\infty}^{\infty} \frac{f_{\bar{n}}(n)}{n} \cos \left(\sum_{k=1}^{\infty} b_k \left(\frac{F_R}{npR^2} - \frac{1}{\pi} \right)^k \right) \\ &\quad \times \left| \sum_{k=1}^{\infty} kb_k \left(\frac{F_R}{npR^2} - \frac{1}{\pi} \right)^{k-1} \right| dn, \end{aligned} \quad (3.3.31)$$

which represents the PDF of the planet-star flux ratio for a population of planets with constant pR^2 values (i.e., Earth-twins) modeled as Lambertian reflectors.

While Equation (3.3.31) is computationally tractable and produces accurate results, the coefficient sets in Equation (3.3.30) can be very time consuming to compute when a large number of terms is required for convergence. This leads us to consider a purely numerical solution that nonetheless produces equally accurate results. Rather than using series reversion, we can use an iterative numerical method to invert $\Phi(\beta)$ and then calculate the first order derivative of the result via finite differences. Recalling the definition of m , by which $\beta = \Phi^{-1}(m)$, we select an initial value for β by fitting and inverting a cubic polynomial to $m = \Phi(\beta)$, resulting in:

$$\beta_0 = -5.921m^3 + 9.825m^2 - 6.636m + 2.802 \quad (3.3.32)$$

when Φ is the Lambert phase function. Next, we use Newton-Raphson iteration to update the estimate as:

$$\beta_{n+1} = \beta + \frac{1}{\pi - \beta} \left(1 + \frac{\pi - \beta}{\tan \beta} - \frac{\pi m}{\sin \beta} \right). \quad (3.3.33)$$

This function, using the initial estimate in Equation (3.3.32), converges to machine double precision within 25 iterations for all values of β , except for the endpoints, which can be simply substituted for their correct values as described above. By using sufficiently small step sizes, the error produced by estimating the derivative via finite differences can similarly be constrained.

When the number of terms leading to convergence is relatively small (on the order of 50), the space X can be calculated in its entirety. This is done by first recursively generating all of the unique sets of natural numbers of length $k - 1$ that sum to $k - 1$. We generate all of the unique pairs of values that sum to $k - 1$, and then recursively augment these sets with all of the sets that sum to the value of the last set element, until no new combinations can be generated. These sets are encoded as an $N \times k - 1$ matrix, where N is the number of unique combinations. For each entry of this matrix of *value* i (where $i \neq 0$) in the j th row, the entry in the j th row and i th column of a matrix of equivalent size is iterated by one. The result is a matrix encoding the entire space X in its rows. This basic process is quite computationally intensive (of order n^n), but can be made significantly more efficient by removing redundant sets at every iteration of the recursive set generation, and by employing an efficient sort when filling the final matrix, rather than doing so recursively. These improvements greatly decrease the computational complexity, bringing the running time down to order 2^n . A sample implementation that is capable of generating X for $k - 1 = 45$ ($N = 89134$) in under 2 minutes on a single core is presented in Listing A.2.

Finally, we consider the apparent separation. The same approximation used in equation (2.1.9) allows us to write:

$$s = r \sin \beta. \quad (3.3.34)$$

If we let $\bar{l} = \sin \bar{\beta}$, following equation (3.3.25):

$$f_{\bar{l}}(l) = f_{\bar{\beta}}(\sin^{-1}(l)) \left| \frac{d}{dl} \sin^{-1}(l) \right| = \begin{cases} \frac{l}{\sqrt{1-l^2}} & l \in [0, 1] \\ 0 & \text{else} \end{cases} \quad (3.3.35)$$

Returning to equations (3.3.7) and (3.3.22) and now letting $\bar{m} = \bar{r}$ and $\bar{n} = \bar{l}$, we have:

$$f_{\bar{s}}(s) = \frac{1}{\pi} \int_0^1 \int_0^\infty \int_0^1 \frac{s}{a\sqrt{(1-l^2)[(ael)^2 - (al-s)^2]}} f_{\bar{e}}(e) de f_{\bar{a}}(a) da dl, \quad (3.3.36)$$

which represents the probability density function of the apparent separation. We now have expressions for the distribution functions of both of the observed quantities of direct imaging planet searches. While the expressions are quite complex, they are numerically integrable, and can be used to check and improve the efficiency of the sampling of the completeness function. With an assumed $f_{\bar{e}}(e)$ and $f_{\bar{a}}(a)$, equations (3.3.31) and (3.3.36), via marginalization and joint sampling, can directly produce the completeness distribution.

3.3.2 Validation

As a check on the derived distribution functions for the true anomaly and orbital radius, we first consider the simplest possible case, where both eccentricity and semi-major axis are uniformly distributed:

$$f_{\bar{e}}(e) = \begin{cases} 1 & e \in [0, 1] \\ 0 & \text{else} \end{cases} \quad (3.3.37)$$

$$f_{\bar{a}}(a) = \begin{cases} (a_{max} - a_{min})^{-1} & a \in [a_{max}, a_{min}] \\ 0 & \text{else} \end{cases} \quad (3.3.38)$$

Equation (3.3.4) then becomes:

$$\begin{aligned} f_{\bar{\nu}}(\nu) &= \frac{1}{2\pi} \int_0^1 \frac{(1-e^2)^{\frac{3}{2}}}{(1+e \cos \nu)^2} de \\ &= -{}_2F_1(1, 2; 7/2, \cos^2 \nu) \frac{\cos \nu}{5\pi} - \frac{3}{16 \cos^4 \nu} (4 \sin \nu + \cos 2\nu - 3) \end{aligned} \quad (3.3.39)$$

where ${}_2F_1$ is the Gaussian hypergeometric function:

$${}_2F_1(a, b; c; z) = \sum_{k=0}^{\infty} \frac{(a)_k (b)_k}{(c)_k} \frac{z^k}{k!}, \quad (3.3.40)$$

with the Pochhammer symbol $(\cdot)_k$ defined as:

$$(x)_0 = 1, \quad (3.3.41)$$

$$(x)_k = \prod_{n=0}^{k-1} (x+n), \quad (3.3.42)$$

and with the value of c limited to the set of natural numbers, i.e., $c \in \mathbb{Z}^+ \setminus \{0\}$. Equation (3.3.7) becomes:

$$\begin{aligned} f_{\bar{r}}(r) &= \frac{1}{\pi a (a_{max} - a_{min})} \left[2r \log \left(\frac{\sqrt{a-r} + \sqrt{r-a}}{\sqrt{2a-r} + \sqrt{r}} \right) \right. \\ &\quad + a \log(a-r) + 2a \log \left(-2 \left(r + \sqrt{2ar-r^2} \right) \right) \\ &\quad \left. + ia \log \left(\frac{32r \left(r - a + i\sqrt{(2a-r)r} \right)}{a} \right) \right] \Big|_{a_{min}}^{a_{max}}. \end{aligned} \quad (3.3.43)$$

To check these equations, we generate one million IID samples each from the uniform distributions in Equations (3.3.37), (3.3.38) and (3.3.3) [103]. From these, we calculate the eccentric anomalies via Newton-Raphson iteration applied to Equation (1.2.14), and then calculate ν and r via Equations (1.2.15) and (1.2.5), respectively (see §4.2.1, Equations (4.2.6) - (4.2.9)). The sample PDFs are calculated for these two parameters and compared

with the results of our closed-form solutions, as shown in Figures 3.10 and 3.11(a). These plots demonstrate excellent agreement, validating the equations.

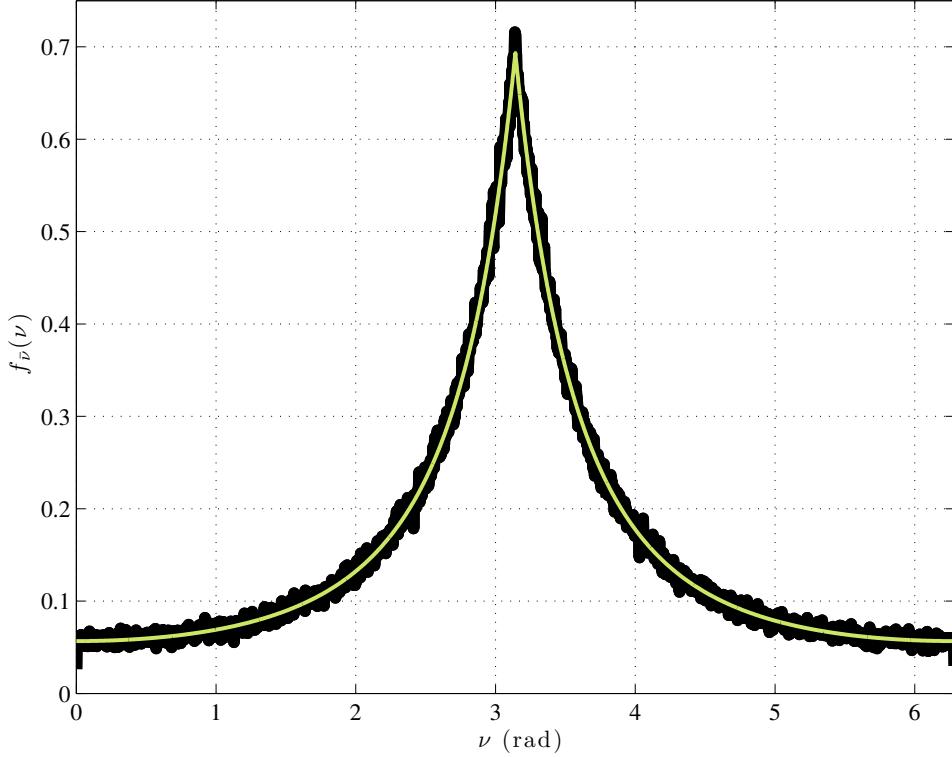


Figure 3.10: Comparison of empirical (black line) and derived (green line) values of $f_{\bar{\nu}}$ assuming a uniform distribution for \bar{e} .

While uniform distributions are a useful check, the semi-major axes of currently found planets appear more likely to be logarithmically distributed [27]. As an additional test, we will assume that \bar{a} is uniformly distributed in $\log_{10}(a)$ for $a \in [10^{c_{min}}, 10^{c_{max}}]$ so that its PDF is:

$$f_{\bar{a}}(a) = \begin{cases} (\Delta c \log(10)a)^{-1} & a \in [10^{c_{min}}, 10^{c_{max}}] \\ 0 & \text{else} \end{cases} \quad (3.3.44)$$

where $\Delta c = c_{max} - c_{min}$. If we retain the same uniform distribution for eccentricity, the true

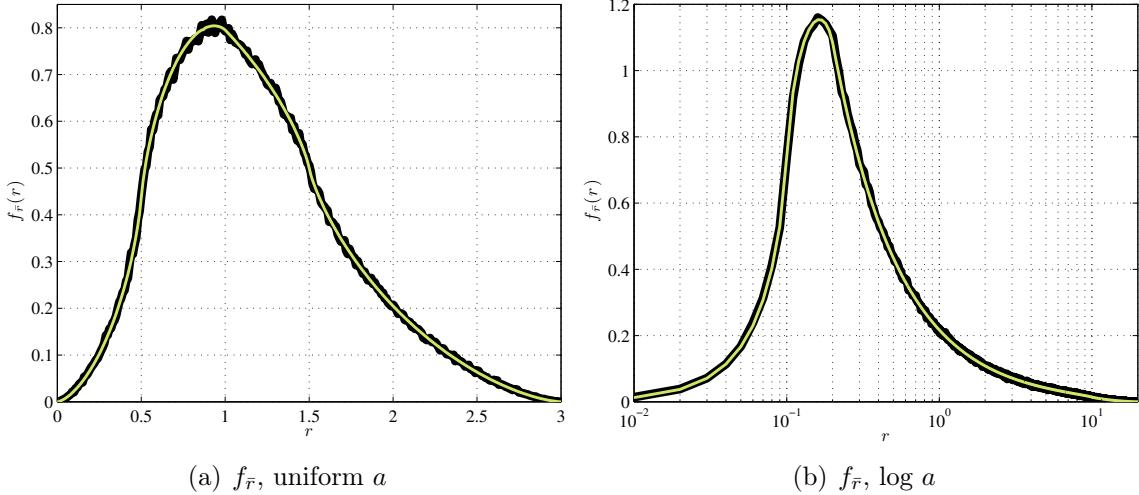


Figure 3.11: Comparison of empirical (black line) and derived (green line) values for $f_{\bar{r}}$, assuming a uniform distribution for \bar{e} and uniform and log distributions for \bar{a} . For uniformly distributed semi-major axis we have $a \in [0.5, 1.5]$, leading to $r \in (0, 3]$, and for \bar{a} uniform in $\log a$, we have $a \in [0.1, 10]$ leading to $r \in (0, 20]$.

anomaly distribution remains the same, but the orbital radius density becomes:

$$f_{\bar{r}}(r) = \frac{1}{2\pi \log(10)\Delta c a^2 r} \left[a\sqrt{(2a-r)r} - a^2 \log(a-r) - 2ia^2 \log\left(\frac{32r^3}{a}\left(r - ai\sqrt{(2a-r)r}\right)\right) + 2a^2 \log\left(-4\left(\sqrt{(2a-r)r^3} + r^2\right)\right) + 2r^2 \log\left(\frac{\sqrt{a-r} - \sqrt{r-a}}{\sqrt{2a-r} + \sqrt{r}}\right) \right] \Big|_{10^{c_{min}}}^{10^{c_{max}}}. \quad (3.3.45)$$

We repeat the simulation, this time generating our sample of semi-major axes by exponentiating 10 with a sample of one million uniformly distributed random values between -1 and 1 (thereby setting Δc to 2), and re-calculating the PDF for $f_{\bar{r}}$. This is compared with the results from Equation (3.3.45) in Figure 3.11(b), and shows excellent agreement.

We can also empirically check the assertion that the distribution of β is independent of the distribution of \bar{e} . We now generate two sets of IID samples, one using the uniform

distribution of \bar{e} , and one distributed via a step function,

$$f_{\bar{e}}(e) = \begin{cases} 1.3 & e \in [0, 0.7) \\ 0.3 & e \in [0.7, 1] \\ 0 & \text{else} \end{cases} \quad (3.3.46)$$

where accurate sampling is achieved via simple rejection sampling. The two sample sets are used to generate two different sample distributions of $\bar{\beta}$. Figure 3.12 shows Q-Q plots comparing the quantiles of these two samples to the theoretical quantiles of a sinusoidal distribution. Quantiles for the ideal sinusoidal distribution are calculated by evaluating its inverted CDF at regularly spaced intervals, while quantiles of the two simulated distributions are calculated by ordering and binning the generated values. As the resulting Q-Q plots follow the 45° line, this demonstrates that the simulated distributions are identical to the theoretical sinusoidal distribution [38].

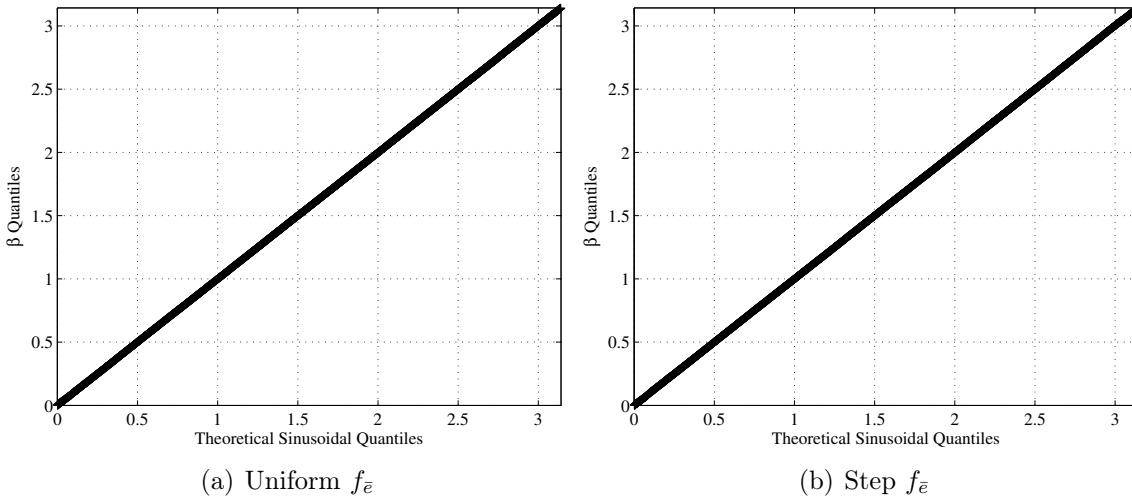


Figure 3.12: Q-Q plots of empirically derived $f_{\bar{\beta}}$ distributions compared with theoretical sinusoidal quantiles.

Continuing on to the analytical distributions of the direct detection observables, Equation (3.3.25) allows us to find the PDF of the Lambert Phase function, as shown in Figure 3.13. Similarly, we can use Equation (3.3.36) to find the PDF of the apparent separation via

quadrature as shown in Figure 3.14.

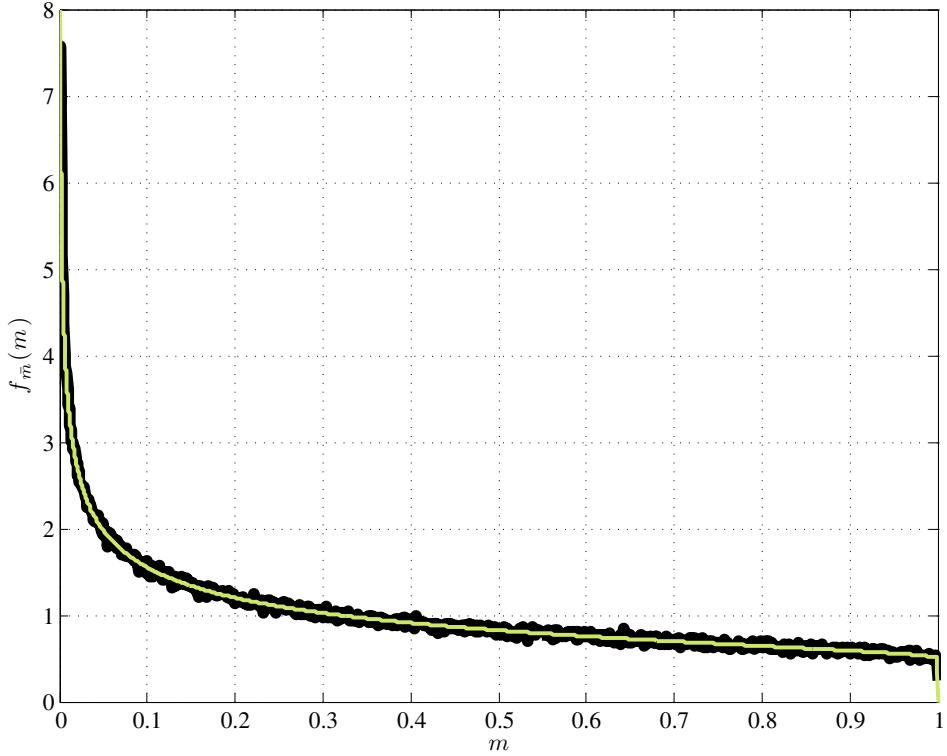


Figure 3.13: Comparison of empirical (black line) and derived (green line) distributions for the Lambert phase function $m = \Phi(\beta)$ for sinusoidally distributed β . Note that in the derived distribution, the endpoints go to zero and infinity, behavior that is not captured in the empirical distribution.

3.3.3 Applications

In addition to improving our ability to sample the completeness function, these analytical distributions have multiple other applications. One particularly interesting question is how we can use these forms to extract more information from observations and improve overall observing efficiency. For example, Equation (3.3.36) can be used to provide us with information about an orbit's semi-major axis from measurements of apparent separation. Given a set of measurements of the apparent separation $\{s_i\}_{i=1}^n$, the maximum likelihood estimate (MLE) for a is given by [98]:

$$\hat{a} = \arg \max_{a \in \hat{a}} \hat{\ell}(a | s_{1 \dots n}) \quad (3.3.47)$$

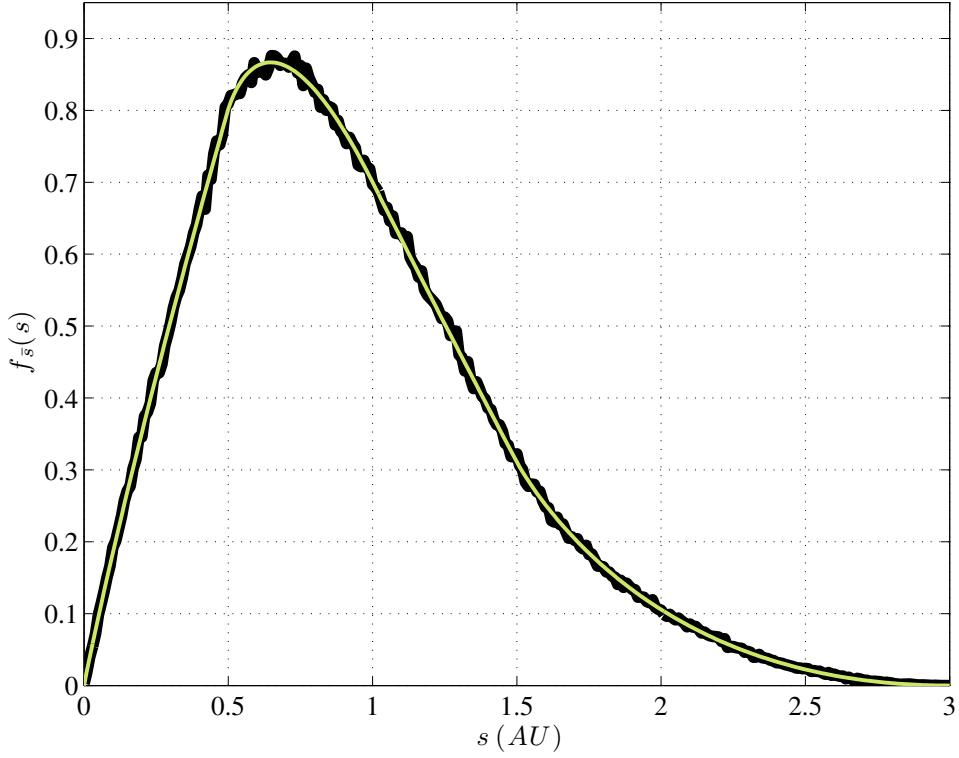


Figure 3.14: Comparison of empirical (black line) and derived (green line) distributions assuming the uniform distribution for \bar{e} and \bar{a} used in Figure 3.11(a).

where $\hat{\ell}$ is the likelihood function generally given as:

$$\hat{\ell}(\theta|x_{1\dots n}) = \log f(x_{1\dots n}|\theta)/n. \quad (3.3.48)$$

Rewriting Equation (3.3.36):

$$\begin{aligned} f_{\bar{s}}(s) &= \int_a f_{\bar{s}|\bar{a}}(s|a) f_{\bar{a}}(a) da \\ &= \int_0^\infty \frac{1}{\pi} \int_0^1 \int_0^1 \frac{s}{a\sqrt{(1-l^2)[(ael)^2 - (al-s)^2]}} f_{\bar{e}}(e) de dl f_{\bar{a}}(a) da \end{aligned} \quad (3.3.49)$$

so that

$$f_{\bar{s}|\bar{a}}(s|a) = \frac{1}{\pi} \int_0^1 \int_0^1 \frac{s}{a\sqrt{(1-l^2)[(ael)^2 - (al-s)^2]}} f_{\bar{e}}(e) de dl. \quad (3.3.50)$$

Since the logarithm is a monotonic function, for a single observation of the apparent separation $s = s_0$, the MLE for semi-major axis is therefore:

$$\hat{a} = \arg \max_{a \in \bar{a}} \frac{1}{\pi} \int_0^1 \int_0^1 \frac{s_0}{a \sqrt{(1-l^2) [(ael)^2 - (al-s_0)^2]}} f_{\bar{e}}(e) de dl . \quad (3.3.51)$$

It is clear that the value of a maximizing this equation is equal to s_0 —i.e., the likeliest value for the semi-major axis given a single observation of the apparent separation is the apparent separation itself, as demonstrated in Figure 3.15. Note that the apparent separation can be less than the semi-major axis due to either orientation of the orbit with respect to the line of sight or the orbit’s eccentricity. The apparent separation can be greater than the semi-major axis due only to eccentricity. As multiple observations of a single exosystem cannot be treated as independent, subsequent measurements of the apparent separation cannot be introduced into the likelihood estimator in Equation (3.3.51), but may be included in orbital fits. Nevertheless, the likelihood function can be used to weight each individual observation, and this estimator proves highly useful in the area of mission planning and observation scheduling, as will be discussed in §4.4.

Returning to the description of transit photometry in §2.3, we see from Equation (2.3.1) that a transit can also be defined in terms of the apparent separation. In fact, in the simplest possible treatment we can define a generalized transit condition of:

$$s < R_S + R . \quad (3.3.52)$$

Of course, this includes transits that are merely grazing and beyond the sensitivity of available instruments, so the subset of detectable transits is determined by the magnitude of the brightness variation ($F_S^{(e)}/F_S$), just as with direct detections. However, if we wish to calculate the overall probability of any transit occurring at any point in time, we can simply write:

$$P [s < R_S + R] = \int_0^{R_S+R} f_{\bar{s}}(s) ds . \quad (3.3.53)$$

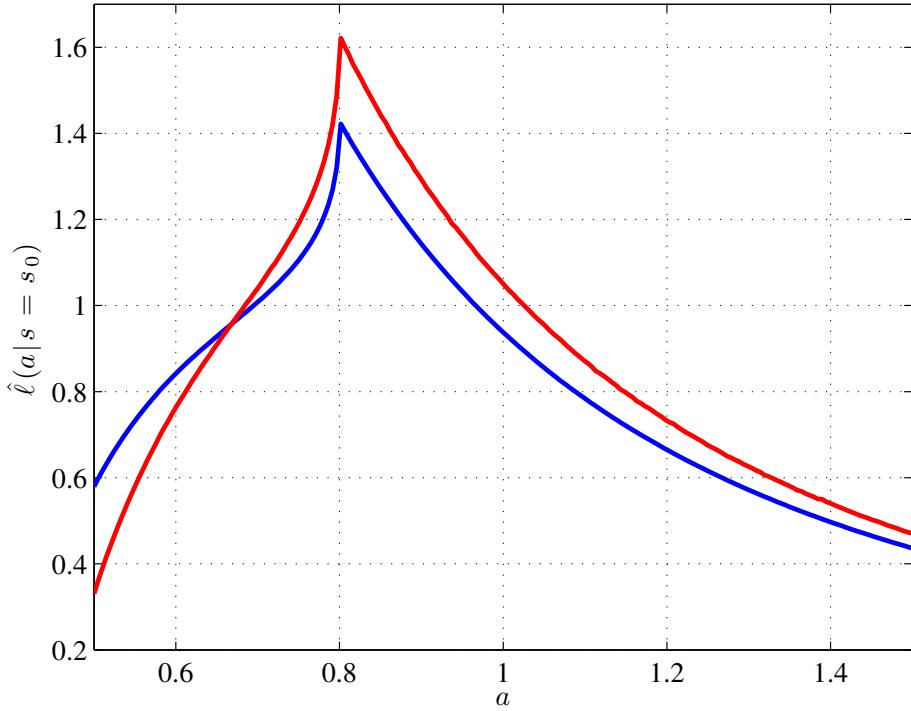


Figure 3.15: Likelihood function for semi-major axis given one observation of apparent separation $s_0 = 0.8$. a is uniformly distributed $\in [0.5, 1.5]$ and e is distributed either uniformly (blue line) or as in Equation (3.3.46) (red line). In each case, the value of a maximizing the likelihood function is equal to s_0 .

Furthermore, if we consider a particular population of orbits—for example, circular orbits with semi-major axis a_0 —this further simplifies to:

$$\begin{aligned}
 P[s < R_S + R | a = a_0, e = 0] &= \int_0^{R_S+R} f_{\bar{s}}(s) ds \\
 &= \frac{1}{\pi} \int_0^{R_S+R} \int_0^1 \frac{s}{a_0 \sqrt{(1 - l^2)(al - s)^2}} dl ds \quad (3.3.54) \\
 &= \frac{(R_S + R)^2}{2\pi a_0}.
 \end{aligned}$$

While an interesting result, Equation (3.3.54) is not very useful on its own, as it represents an instantaneous probability of transits, whereas we are generally interested in the probability of transits over a certain time period (for example, one of the Kepler quarters [4]). Every transit photometry survey has a certain cadence of observations as even instru-

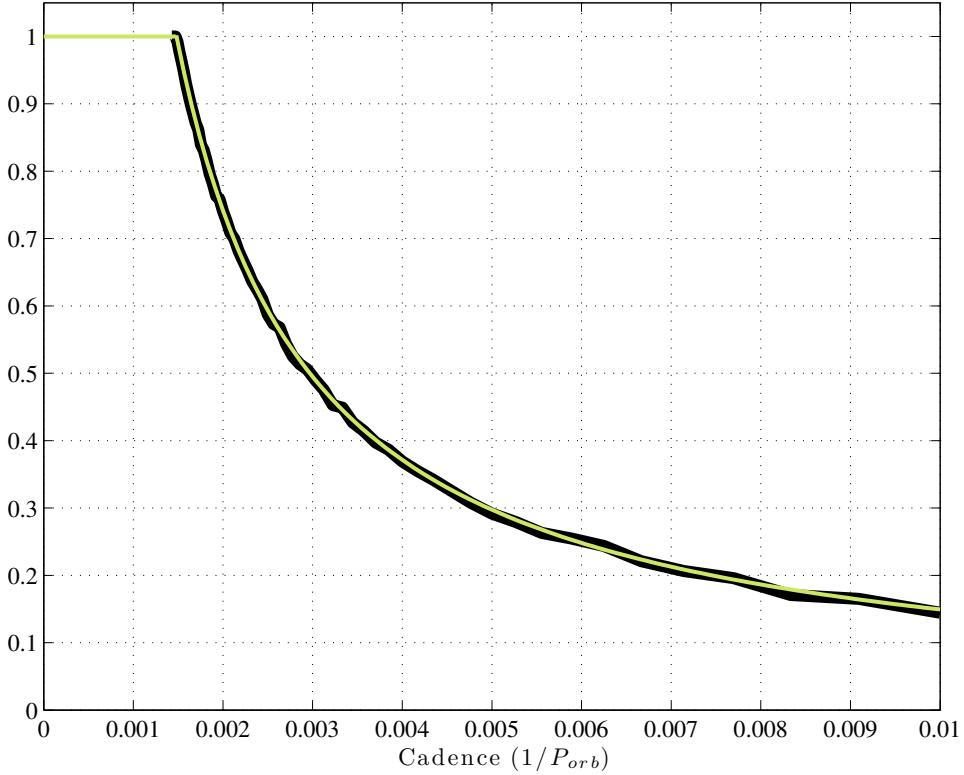


Figure 3.16: Portion of observable transits detected over one orbital period for $a_0 = 1$, $R_S = R_\odot$ and $R = 0$ as a function of observation cadence. The black line represents the results of simulating one million planets on orbits where transits would occur ($|\theta| < R_s/a_0$) and propagating them forward in time by varying time steps. The green line is defined by Equation (3.3.57).

ments that monitor stars continuously have a finite integration and readout time. Therefore, we can model transit observations on a large group of stars as a Poisson process. While the occurrence of transits about a single star is not independent, if we are merely interested in the portion of monitored stars that will have at least one transit in a given period of time, then we have a collection of random variables representing the number of events occurring throughout a given time interval, each of which is Poisson distributed. Thus, the portion of target stars with detected transits is a monotonically increasing function in time, with each specific interval having a probability of non-zero transits of:

$$P[N(\Delta t) > 0] = 1 - e^{-\lambda \Delta t} \quad (3.3.55)$$

where Δt is the time interval and λ is the expected number of occurrences per interval. This formulation, along with Equation (3.3.54), allows us to calculate the optimal cadence such that 100% of transits could be captured within one orbital period:

$$\Delta t_0 = \frac{1}{\pi} \frac{R_s + R}{a_0} \quad (3.3.56)$$

where this value is in units such that the orbital period associated with a_0 (for zero planet mass) is 1. For any greater time step, the portion of transits discovered within one orbital period is given by:

$$\frac{N(\Delta t)}{N(\Delta t_0)} = \frac{\Delta t_0}{\Delta t} + \Delta t. \quad (3.3.57)$$

Of course, the actual portion of transits detected will always be less than unity due to the photometric limitations of the instrument, but the sampling rate in Equation (3.3.56) represents the optimal cadence for a given semi-major axis. Since this interval is inversely proportional to semi-major axis, studies interested in closer-in planets can actually monitor stars fairly infrequently, whereas those looking for further out planets must monitor nearly continuously (as is evident from the geometry of the problem). Figure 3.16 shows a comparison of Equations (3.3.56) and (3.3.57) with results from simulation, demonstrating excellent agreement.

The specific detection method biases, the statistical modeling of exoplanet populations and the concept of completeness are all extremely important to both observation simulation and data analysis. The next two chapters will demonstrate this using the framework developed here, including the observation timing optimizations. Chapter 4 will present the details of constructing a planet-finding mission simulation, many of which deal with observation scheduling and are thus greatly informed by completeness and the distributions of planetary parameters.

Chapter 4

Observation Simulation

Along with improvements in instrumentation and data analysis, the planet-finding community has also devoted resources to the development of detailed models and simulations of the various instruments and surveys currently in operation. These simulations serve three main purposes: first, they allow the community to directly compare the expected scientific returns of proposed observatory and mission concepts [113]. Second, these simulations can be used to optimize target selection and observation scheduling for existing facilities, and provide a controlled testing environment for data processing pipelines [12]. Finally, by simulating a planet population derived from a specific formation/evolution theory, we can estimate the likelihood of survey results, and thus accept or reject the theory, while also closely estimating absolute planet frequencies [43]. This last capability translates to the ability to use surveys as controlled experiments to choose between competing exosystem formation and evolution theories. The ability to predict the expected science returns of a proposed architecture is particularly important given the resources required to operate a major ground-based observatory, or to launch a new space observatory.

A satellite along the lines of the Hubble Space Telescope or James Webb Space Telescope is a major undertaking and should only be built if it has a reasonable chance of succeeding in its stated mission. In terms of mission planning and design, this translates to formulating

a set of requirements which are reasonable given our current state of knowledge, and then designing hardware which will meet these requirements. This chapter will present the techniques required to model whole planet-finding missions and surveys, as well as a selection of case studies demonstrating various applications of these methods.

4.1 Mission Planning and Analysis

There are two possible approaches to predict scientific performance of potential planet finding missions. The first is to use a numerically evaluated statistical description of various quantities involved in a planetary detection and report results in terms of expected values of random variables, producing either an expected number of planets found as in Brown [8], or a total number of zones of interest probed* as in Lindler [80]. While these are powerful metrics for mission performance, they make it very difficult to evaluate highly specific mission requirements. For example, if we wish to require that our instrument produce orbital fits for $m\%$ of all planets found such that the fit eccentricities will be within $n\%$ of their true values, the statistical description of this requirement would be quite complex. The second option is to perform an analysis in which we simulate the entire mission, along with all detections, many times over, which allows us to not only evaluate the requirement above, but also actually produces values for n and m for every evaluated design. By performing such simulations many times over, with varying populations of planets, we produce an ensemble of mission timelines, giving us distributions of mission science yields. While this approach is significantly more computationally intensive than earlier treatments, it has the advantage of allowing us to answer a variety of questions about proposed mission architectures with a single ensemble of simulations.

In addition to helping us decide what exactly to fund, construct, and fly, this Monte Carlo approach has an additional benefit. A complete simulation capability of planet finding

*A zone of interest refers to the fractional volume of space that is observed in the course of the mission where a planet of interest might reside.

missions will not only allow us to compare the performance of different instrument designs, but also different mission designs for the same instrument. Assuming sufficient faith in the accuracy of the simulation, we can evaluate strategies and rules before any hardware is even assembled. Additionally, once a robust simulation framework has been constructed, it becomes quite simple to answer additional questions. A wholly analytical approach would require at least a partial reformulation of the problem before we could tell, for example, what effect knowing exactly which stars have planets would have on the mission results, but a Monte Carlo approach can give us the answer simply by running another set of simulations. The mission analysis method detailed in this chapter was first presented in Savransky and Kasdin [112] and further developed in Savransky et al. [113]. It involves generating randomized planetary systems assigned to stars from a specific target pool, propagating them forward in time, and simulating observations of these systems with specific instruments on spacecraft or on the ground.

4.2 Generating and Propagating Planetary Systems

4.2.1 System Generation

In generating randomized planetary systems, we can take one of three approaches:

1. Identify a planetary population of interest (for example, ‘Earth-like’ planets) which sets values or ranges for all orbital and physical planetary parameters.
2. Use the solar system as a model, and generate exosystems composed of a subset of the bodies in the solar system.
3. Attempt to model the actual distribution of physical and orbital parameters of known planets based on all available data (see §3.1).

Of these, the first approach is simplest as it calls for just one body in each simulated exosystem, and so this is the approach most often found in the literature. Frequently, the selected

population is ‘Earth-twins’ on habitable zone orbits. This means that all of the simulated planets have the mass and radius of the Earth as well as the Earth’s average albedo and reside on orbits with semi-major axes between 0.7 and 1.5 AU (scaled by the square root of their parent star’s luminosity) with eccentricities of less than 0.35 [69, 8]. It has become convention to use the parameter η_{\oplus} to indicate the expected number of Earth-like planets per star; that is, $\eta_{\oplus} = 1/3$ indicates that amongst random samples, one third of stars, on average, will have an Earth-like planet [?]. As many of the space-based planet-finding mission concepts have the stated goal of finding terrestrial planets, and as this usually represents their most difficult mission requirement, this simplified population is actually a very good test of mission capabilities.

The second approach provides a convenient shortcut to generating multi-planet exosystems. Rather than having to search for stable configurations of multiple planets, we can start with a system known to be stable and use the known positions and velocities of the solar system planets to generate exosystems with long-term stability. It is important to note that this approach does not require that we always simulate systems with 8 planets orbiting a star identical to the sun. If a proper subset of solar system planet are selected, then all that must be done is to subtract the scaled initial velocity of the resulting barycenter from the initial velocity of the star, i.e.:

$${}^T \mathbf{v}'_{S/G}(t_0) = {}^T \mathbf{v}_{S/G}(t_0) - \frac{1}{m_S} \left(\sum_{i \in Y} m_{P_i} {}^T \mathbf{v}_{P_i/G} + m_S {}^T \mathbf{v}_{S/G} \right) \quad (4.2.1)$$

where Y is the set of solar system planets to be included in the new ecosystem, ${}^T \mathbf{v}'_{S/G}(t_0)$ is the velocity of the star in the new ecosystem and ${}^T \mathbf{v}_{S/G}(t_0)$ is the velocity of the sun in the original initial conditions.

It is also possible to change the properties of the star in the simulated ecosystem and incorporate the effects of different stellar masses and luminosities on the orbits. Recalling equations (1.2.6) and (1.2.7) we note that the planet-star separation is directly proportional

to the semi-major axis, while a planet's velocity with respect to the star is inversely proportional to the square root of the semi-major axis. This means that if we wished to scale the entire exosystem by some size factor (such as \sqrt{L}), we could do so by multiplying the initial positions of the planets (but not the star) by \sqrt{L} and dividing the initial planet velocities by $L^{1/4}$. Figure 4.1 presents an example of one such system. While it is difficult to determine

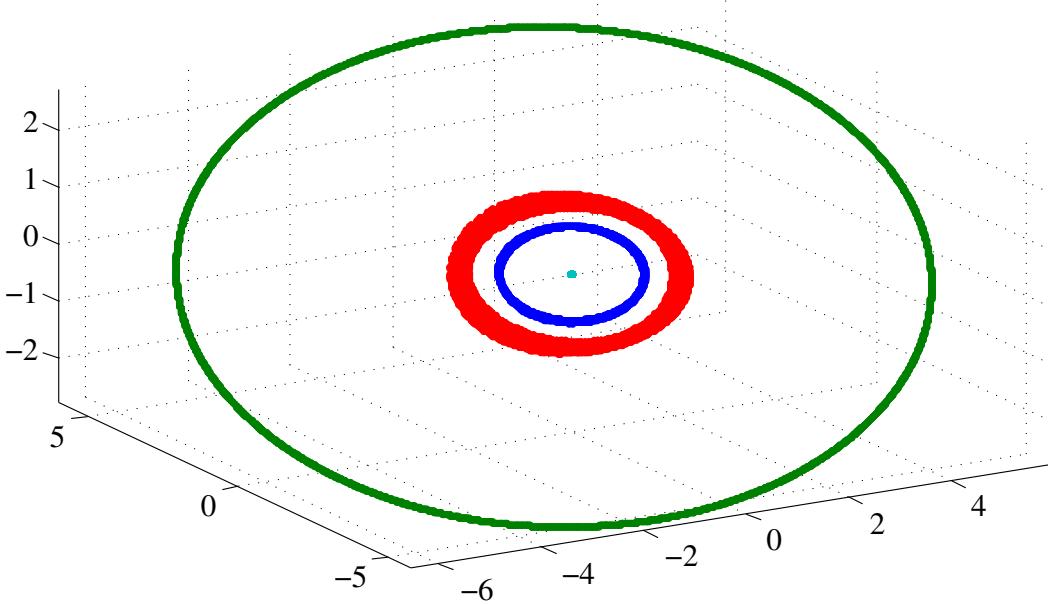


Figure 4.1: Orbits over 1 million years for exosystem composed of analogues of Earth (blue), Mars (red) and Jupiter (green), with a $1.5 L_\odot$ star and orbits scaled such that all planets receive the same flux as they would in our solar system. Note that over this time span, the Jupiter analogue causes significant deviations from the Keplerian approximation in the orbits of the two minor planets. Scale units are in AU.

global stability of n-body systems, it is possible to determine local stability by integrating over long periods of time (millions of years) and searching for collisions or ejections (see §4.2.2).

The third approach to generating planetary systems represents the widest possible range of possible exosystems and is also the most difficult to apply. One possibility is to ‘grow’ solar systems by starting with simulated nebulae and utilizing a formation model to develop planets. This is quite costly as it requires integrating many millions of years of the simulated system and often results in single-planet systems as bodies are ejected or destroyed.

An alternate method is to sample the known mass and semi-major axis distributions (see Figure 3.5) and generate systems composed of multiple such bodies with initial conditions determined by Keplerian approximations. However, using Keplerian approximations for initial conditions makes it difficult to simulate systems with non-Keplerian, n-body effects such as resonances [35] and long integrations are still required to ensure system stability.

The procedure for generating completely arbitrary planetary systems begins with the generation of N random mass values according to the best-fit mass distribution. From these masses the planetary radii are determined as follows: For mass ranges up to $10M_{\oplus}$, two random numbers are generated: a boolean value determining whether the planet will be a rock/iron or rock/ice mixture, and a random number uniformly distributed $\in [0, 1]$, determining the fraction of rock via Equation (3.1.2). For giant planets, a uniformly distributed random number is generated in the range of $[0, 1]$ to represent the portion of mass in the planet's core, and radii values are interpolated from a lookup table based on values in Fortney et al. [37].

Each planet is then associated with one of the target stars, with the likelihood of association given by the overall occurrence rate of planets and the iron abundance of the stars in the target list. Each planet is assigned an average albedo value (p) drawn from a uniform distribution in the range $[p_{min}, p_{max}]$. There is currently much work being done to model albedo values as functions of other planetary properties, but no one model is able to capture the entire parameter space [133]. For each planet, we generate initial positions and velocities by generating randomized average orbital elements via the distributions in §3.3. The mass of each target star is calculated from its absolute V magnitude (M_V) via the relation:

$$\log(m_S/m_{\odot}) = 0.002456M_V^2 - 0.09711M_V + 0.4365. \quad (4.2.2)$$

This relation has been shown empirically to be accurate to within 7% for mass ranges corresponding to visual magnitudes between 1.45 and 10.25 [55]. We treat the output of Equation

(4.2.2) as the ‘estimated’ mass and generate a ‘true’ stellar mass as

$$m_S^{est} = m_S^{true}(1 + \nu) \quad (4.2.3)$$

where ν is a random error term consistent with the known accuracy of Equation (4.2.2). The true masses are used for propagating the planets along their orbits, while the estimated masses of stars are needed for our return strategies, as described in §4.4. This relation is used since it conveniently covers the range of visual magnitudes and stellar masses considered here. However, newer mass-luminosity relations exist, with better accuracy for different mass ranges, and can be substituted for Equation (4.2.2) [106].

At the end of this process, we have a target list encoded as an array of N parameter sets

$$\{m_S^{est} \quad m_S^{true} \quad p_i \quad m_{P_i} \quad R_i \quad \mathbf{r}_{P_i/G} \quad {}^T\mathbf{v}_{P_i/G} \quad \mathbf{r}_{S/G} \quad {}^T\mathbf{v}_{S/G}\} \quad (4.2.4)$$

with $i = 1 \dots n$ for a system of n planets. This parameter set can now be propagated forward in time to test the system stability and to perform mission simulations. The set of parameter sets describing a group of simulated target systems will be referred to as a ‘universe’ throughout the remainder of the text.

4.2.2 System Propagation

Propagation of planets along their orbits can be done either by using the Keplerian formulation (which is exact only for a single-planet system) or by numerically integrating the Newtonian (Equation (1.2.32)) or post-Newtonian (Equation (1.2.37)) equations of motion. If using the Keplerian formulation, the location of the planet is determined by the (possibly osculating) Keplerian elements and the current anomaly. For the simulations in §4.6, the Keplerian propagation described here is used only for single-planet systems. All simulations with multiple-planet systems use the n-body numerical integration methods described below.

From Equations (1.2.12) and (1.2.13) we can write:

$$M = \sqrt{\frac{\mu_S + \mu_P}{a^3}}(t - t_p), \quad (4.2.5)$$

thereby directly relating the mean anomaly to the current time. As the Kepler equation (1.2.14) is non-invertible in E , we must use numerical methods in order to find the eccentric anomaly from the mean anomaly. The simplest approach is to use Newton-Raphson iteration:

$$E_{n+1} = E_n - \frac{M - E_n + e \sin E_n}{e \cos E_n - 1}. \quad (4.2.6)$$

This iterant is applied to some initial value E_0 until convergence:

$$E_{n+1} - E_n < \varepsilon \quad (4.2.7)$$

where ε is typically set to the machine precision value of the data type being used [91]. This method produces excellent convergence as long as E_0 is sufficiently close to the true value for E . For most values of eccentricity, it suffices to use M for E_0 , but in highly eccentric orbits, the mean and eccentric anomalies will differ greatly throughout the orbit. An alternate approach to selecting E_0 lies in examining the Taylor expansion of the Kepler equation:

$$M = E - e \sin E \approx (1 - e)E + \frac{eE^3}{6} + \mathcal{O}(E^4). \quad (4.2.8)$$

We note that for a fixed value of e and for $|M - E| >> \varepsilon$, one of the first two terms of the expansion is much larger than the other. We therefore select E_0 as:

$$E_0 = \begin{cases} \frac{M}{1-e} & \frac{M}{1-e} < \sqrt{\frac{6-6e}{e}} \\ \left(\frac{6M}{e}\right)^{\frac{1}{3}} & \text{else} \end{cases} \quad (4.2.9)$$

Because the Newton-Raphson algorithm is iterative, it is impossible to make it more time

efficient than the minimum number of iterations required to achieve the desired precision. However, using the initialization algorithm described above, machine double precision can be achieved for any eccentricity (up to $1 - \varepsilon$) in under 30 iterations for arbitrarily sized arrays of Mean anomalies and eccentricities (see Listing A.3).

While Equation (4.2.6) is guaranteed to result in the correct anomalies for closed orbits, it starts with orbital elements, meaning that if planetary positions are recorded using a position and velocity state, it would be necessary to first apply Equations (1.2.20) - (1.2.31) in order to find the osculating orbital elements. Then, an updated mean anomaly would be defined as:

$$M(t) = \left(M(t_0) + 2\pi \frac{\Delta t}{P_{orb}} \right) \bmod 2\pi, \quad (4.2.10)$$

where Δt is the desired time interval of the propagation ($\Delta t = t - t_0$) and mod represents the modulo operator. Equation (4.2.6) would then be applied to find $E(t)$, from which the updated position and velocity could be calculated via Equations (1.2.16), (1.2.17), and (1.2.19).

A more direct method, which efficiently combines the conversion and iteration and is applicable to all orbits (open or closed), is described by Shepperd [120], based on work by Goodyear [41, 42]. Following Sundman [134], the solution to Equation (1.2.36) can be written as a function of a single parameter w where:

$$\frac{dt}{dw} = \|\mathbf{r}_{P/S}\|. \quad (4.2.11)$$

This solution (sometimes referred to as the ‘f and g method’) allows us to update positions and velocities from their initial conditions:

$$\mathbf{r}_0 = \mathbf{r}_{P/S}(t_0) \quad \text{and} \quad \mathbf{v}_0 = {}^T \mathbf{v}_{P/S}(t_0) \quad (4.2.12)$$

by using the algorithm:

$$v_0 = \mathbf{r}_0 \cdot \mathbf{v}_0 , \quad (4.2.13)$$

$$\beta = \frac{2\mu}{\|\mathbf{r}_0\|} - \mathbf{v}_0 \cdot \mathbf{v}_0 , \quad (4.2.14)$$

$$r = \|\mathbf{r}_0\|U_0 + v_0U_1 + (\mu_S + \mu_P)U_2 , \quad (4.2.15)$$

$$t = \|\mathbf{r}_0\|U_1 + v_0U_2 + (\mu_S + \mu_P)U_3 , \quad (4.2.16)$$

$$\mathbf{r}(t) = f\mathbf{r}_0 + g\mathbf{v}_0 , \quad (4.2.17)$$

$$\mathbf{v}(t) = F\mathbf{r}_0 + G\mathbf{v}_0 , \quad (4.2.18)$$

where U_n are the universal functions $U_n(w, \beta)$ [2] and

$$\begin{aligned} f &= 1 - \frac{(\mu_S + \mu_P)U_2}{\|\mathbf{r}_0\|} & g &= \|\mathbf{r}_0\|U_1 + v_0U_2 \\ F &= \frac{(\mu_S + \mu_P)U_1}{r\|\mathbf{r}_0\|} & G &= 1 - \frac{(\mu_S + \mu_P)U_2}{r} . \end{aligned} \quad (4.2.19)$$

The f and g functions may be equivalently defined in terms of the Keplerian orbital elements as:

$$\begin{aligned} f &= 1 - \frac{\|\mathbf{r}\|}{\ell} + \frac{\|\mathbf{r}\|}{\ell} \cos(\nu - \nu_0) , \\ g &= \frac{\|\mathbf{r}\|\|\mathbf{r}_0\|}{\sqrt{(\mu_S + \mu_P)\ell}} \sin(\nu - \nu_0) , \\ F &= \sqrt{\frac{\mu_S + \mu_P}{\ell}} \tan \frac{\nu - \nu_0}{2} \left(\frac{1 - \cos(\nu - \nu_0)}{\ell} - \frac{1}{\|\mathbf{r}_0\|} - \frac{1}{\|\mathbf{r}\|} \right) , \\ G &= 1 - \frac{\|\mathbf{r}_0\|}{\|\mathbf{r}\|} (1 - \cos(\nu - \nu_0)) , \end{aligned} \quad (4.2.20)$$

where ν_0 is the true anomaly at time t_0 .

Note that β is proportional to the orbital energy and Equation (4.2.16) is a reformulation of Kepler's equation in terms of w , which itself can be related to the eccentric anomaly as:

$$w = \frac{E - E_0}{\sqrt{\beta}} , \quad (4.2.21)$$

where E_0 is the eccentric anomaly at t_0 .

The universal functions $U_n(w, \beta)$ may be defined implicitly as:

$$\begin{aligned} U_0 &= \begin{cases} \cos(w\sqrt{\beta}) & \beta > 0 \\ \cosh(w\sqrt{-\beta}) & \beta < 0 \end{cases} \\ U_1 &= \begin{cases} \sin(w\sqrt{\beta})/\sqrt{\beta} & \beta > 0 \\ \sinh(w\sqrt{-\beta})/\sqrt{-\beta} & \beta < 0 \end{cases} \\ U_{n+2} &= \frac{w^n}{\beta n!} - \frac{U_n}{\beta}. \end{aligned} \quad (4.2.22)$$

Unfortunately, this definition is not amenable to numerical implementation as the computation of the various trigonometric functions invariably introduces errors and often leads to singularities as various terms (β especially) tend to zero. The equivalent series representation:

$$U_n(w, \beta) = \sum_{k=0}^{\infty} \frac{(-\beta)^k w^{n+2k}}{(n+2k)!} \quad (4.2.23)$$

is more tractable, but may still have convergence issues in the general case. Shepperd [120] solves this problem by introducing a change of variable:

$$u = \frac{U_1(w/4, \beta)}{U_0(w/4, \beta)} \quad (4.2.24)$$

which falls strictly in the range $u \in [-|\beta|^{-\frac{1}{2}}, |\beta|^{-\frac{1}{2}}]$. Since the alternate formulation of Kepler's equation in Equation (4.2.16) requires only the first few universal functions, the problem may be greatly simplified by writing:

$$U_0\left(\frac{w}{2}, \beta\right) = 1 - 2q, \quad (4.2.25)$$

$$U_0(w, \beta) = 2U_0^2\left(\frac{w}{2}, \beta\right), \quad (4.2.26)$$

$$U_1\left(\frac{w}{2}, \beta\right) = 2(1 - q)u, \quad (4.2.27)$$

$$U_1(w, \beta) = 2U_0\left(\frac{w}{2}, \beta\right) U_1\left(\frac{w}{2}, \beta\right), \quad (4.2.28)$$

$$U_2(w, \beta) = 2U_1^2\left(\frac{w}{2}, \beta\right), \quad (4.2.29)$$

where

$$q = \frac{\beta u^2}{1 + \beta u^2}. \quad (4.2.30)$$

Furthermore,

$$U_3(w, \beta) = \frac{4}{3}U_1^3\left(\frac{w}{2}, \beta\right) G(3, 0, \frac{3}{2}, q), \quad (4.2.31)$$

where G is the continued fraction expansion of the ratio of the ratio of two hypergeometric series:

$$G(a, b; c; z) = \frac{{}_2F_1(a, b+1; c+1; z)}{{}_2F_1(a, b; c; z)}, \quad (4.2.32)$$

with the Gaussian hypergeometric function defined as in Equation (3.3.40).

Returning to Equations (4.2.17) and (4.2.18), we see that the position and velocity can be updated via a simple matrix multiplication:

$$\begin{bmatrix} \mathbf{r}(t) \\ \mathbf{v}(t) \end{bmatrix} = \begin{bmatrix} fI_3 & gI_3 \\ FI_3 & GI_3 \end{bmatrix} \begin{bmatrix} \mathbf{r}_0 \\ \mathbf{v}_0 \end{bmatrix}, \quad (4.2.33)$$

where I_3 represents the 3×3 identity matrix. Multiple bodies can be updated together by making the state transition matrix block diagonal and stacking positions and velocities into a single column vector:

$$\begin{bmatrix} \mathbf{r}_1(t) \\ \mathbf{v}_1(t) \\ \vdots \\ \mathbf{r}_n(t) \\ \mathbf{v}_n(t) \end{bmatrix} = \begin{bmatrix} f_1I_3 & g_1I_3 \\ F_1I_3 & G_1I_3 \\ & \ddots \\ & & f_nI_3 & g_nI_3 \\ & & F_nI_3 & G_nI_3 \end{bmatrix} \begin{bmatrix} \mathbf{r}_1(t_0) \\ \mathbf{v}_1(t_0) \\ \vdots \\ \mathbf{r}_n(t_0) \\ \mathbf{v}_n(t_0) \end{bmatrix}. \quad (4.2.34)$$

In this formulation, the Kepler equation is solved via Newton-Raphson iteration on the independent variable, u :

$$u_{n+1} = u_n - \frac{t - \Delta t}{4(1 - q)r}. \quad (4.2.35)$$

A secondary iteration is required to evaluate the function G . A convenient algorithm is Shepperd's modification of Gautschi's method for the evaluation of continued fractions. To evaluate the function $G(a, b; c; z)$, a set of intermediate variables is initialized as:

$$k_0 = 1 - 2(a - b), \quad (4.2.36)$$

$$l_0 = 2(c - 1), \quad (4.2.37)$$

$$d_0 = 4c(c - 1), \quad (4.2.38)$$

$$m_0 = 4b(c - a), \quad (4.2.39)$$

$$A_0 = B_0 = G_0 = 1, \quad (4.2.40)$$

and then iterated until convergence via:

$$k_{n+1} = -k_n, \quad (4.2.41)$$

$$l_{n+1} = l_n + 2, \quad (4.2.42)$$

$$d_{n+1} = d_n + 4l_{n+1}, \quad (4.2.43)$$

$$m_{n+1} = m_n + (1 + k_{n+1})l_{n+1}, \quad (4.2.44)$$

$$A_{n+1} = \frac{d_{n+1}}{d_{n+1} - m_{n+1}A_n z}, \quad (4.2.45)$$

$$B_{n+1} = (A_{n+1} - 1)B_n, \quad (4.2.46)$$

$$G_{n+1} = G_n + B_{n+1}. \quad (4.2.47)$$

An implementation of the full algorithm is shown in Listing A.4.

For multi-planet system, we abandon the Keplerian propagation in favor of numerical integration of Equation (1.2.32). While this is more computationally intensive, it produces

much more accurate results in systems with multiple planets, especially when one or more of these is Jupiter-size or larger. More importantly, the Keplerian formulation cannot be used to generate consistent stellar reflexes in multi-planet system and is therefore a poor model for studying precision astrometry and radial velocity. Because the Keplerian formulation places the star at the origin of a frame that is treated as inertial, the Keplerian formalism implies that the star does not move. Using the standard definition for center of mass, G :

$$\sum_{k \in X} m_k \mathbf{r}_k / G = 0, \quad (4.2.48)$$

where X is the set of all of the bodies in the exosystem (see Equation (1.2.33)), the position of the star with respect to the barycenter is:

$$\mathbf{r}_{S/G} = -\frac{\sum_{k \in Y} m_k \mathbf{r}_{k/S}}{\sum_{j \in X} m_j}, \quad (4.2.49)$$

where Y is now the set of all exosystem bodies excluding the star, $Y = X \setminus \{S\}$. If we consider the Newtonian formulation of the two body problem, as in Equation (1.2.36), with

$$\mathbf{r}_{P/S} = \mathbf{r}_{P/G} - \mathbf{r}_{S/G}, \quad (4.2.50)$$

where G is now the barycenter of the two-body system, we see that Equation (4.2.49) is the average of the star position with respect to the barycenter of each of the two-body systems composed of the star and each individual planet. This means that rather than modeling the true interactions of all of the bodies in the system, the best that can be accomplished via the Keplerian model is a first order approximation of the true stellar motions.

The specific method used for the numerical integration is very important. Because an orbital system with no external perturbations is a canonical Hamiltonian system, it is important to ensure that energy is being conserved. Low order Runge-Kutta methods will not do so, deviating quite quickly from the true system state. The solution is to use either high

order integrators, or one of a class of symplectic integration schemes. Of particular interest are the Runge-Kutta-Nyström (RKN) methods as they are specifically formulated to solve second order ODEs that are cyclic in the first derivative of the state:

$$\frac{d^2}{dt^2}y = f(y). \quad (4.2.51)$$

While these methods also do not guarantee that energy will be preserved, they are canonical, and sufficiently high-order implementations preserve energy to a very large degree [105]. Figure 4.2 shows a comparison of energy variation over time in numerical integrations of the solar system. Unlike the two RKN implementations, the Adams-Basforth-

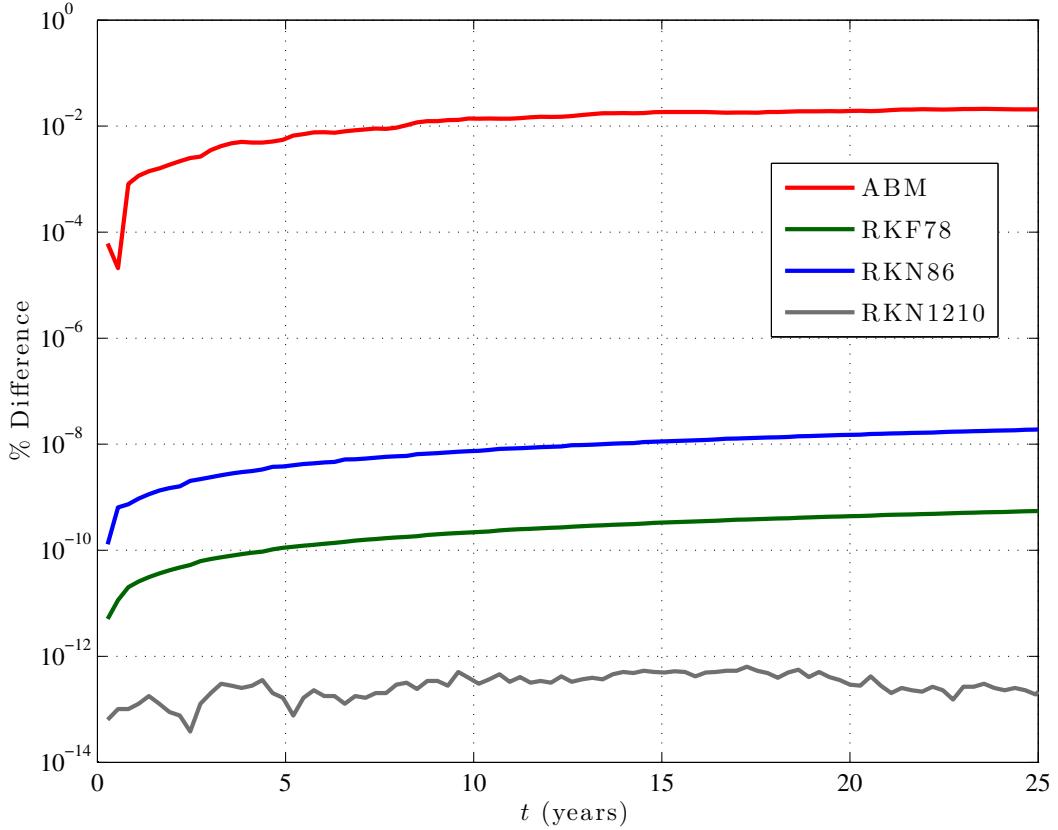


Figure 4.2: Change in total energy (Equation (1.2.35) over 25 years of integration of the solar system using different numerical solvers with identical initial conditions.

Moulton (ABM) PECE (prediction-evaluation-correction-evaluation) multi-step solver [118] and Runge-Kutta-Fehlberg (RKF) 7(8) solver [36], operate on first order ODEs. However,

the high order of the Runge-Kutta-Fehlberg (RKF) makes it outperform the (relatively) lower-order Runge-Kutta-Nyström 8(6) solver [99, 105] in terms of energy conservation. The overall best performance is achieved with the Runge-Kutta-Nyström 12(10) solver [31]. The ABM solver, while multistep, does not have volume or energy preserving qualities and is thus the poorest choice. Over relatively short periods (less than 1 million years), the 12/10th order RKN solver is energy-preserving to within machine double precision. For most system energy is stable on the order of $10^{-12}\%$ over billions of years.

It is also common to use hybrid-symplectic integrators which use symplectic integration schemes when masses interact closely, and non-symplectic schemes the rest of the time to greatly increase the speed of integration without losing the energy-conserving properties of fully symplectic integrators [20]. However, because of the amenability of the n-body equation to vectorization, and the constantly increasing speed of computer processors, very high-order methods such as the RKN 12(10) solver can be run with integration times of billions of years in feasible amounts of computer time by using variable time steps and efficient error estimation algorithms [136]. Equation (4.2.32) can be evaluated very efficiently by precomputing four indexing arrays. For a state vector x representing n bodies, with positions and velocities stacked in a column vector as in Equation (4.2.34), $[j]$ is a vector of $9n$ elements filled with indices from 1 to n each repeated $n - 1$ times:

$$[j] = [1, 1, \dots, 1, 2, 2, \dots, 2, \dots, n]. \quad (4.2.52)$$

The vector $[k]$ also has $9n$ elements, but is filled by circularly permuting the indices such that $k \neq j$:

$$[j] = [2, 3, \dots, n, 1, 3, \dots, n, 1, 2, 3, \dots, n, \dots, 1, \dots, n - 1]. \quad (4.2.53)$$

Finally, two indexing matrices are required: $[z1]$ contains indices $1 \dots 3n$, arranged in a matrix of dimension $3 \times n$ and $[z2]$ contains indices $1 \dots n(n - 1)$, arranged in a matrix of dimension $n \times n - 1$. In this way, every required combination of position differences can be

computed as by indexing the state with $[z1]$ recursively indexed by $[j]$ and $[k]$ and Equation (1.2.32) can be evaluated by summing over the columns of the matrix resulting from indexing the difference combinations with $[z2]$. An implementation of this algorithm is presented in Listing A.5.

Numerical integration also makes it simple to search for specific events, such as collisions (when $\|\mathbf{r}_{k/j}\| < R_k + R_j$) or ejections (when specific orbital energy changes sign). The basic approach is to define a set of continuous functions that cross zero when an event of interest occurs and then to use root finding methods such as Regula-Falsi to pinpoint the exact time of the event when a sign change occurs in the function evaluation between integrator iterations [103]. In this fashion the stability of generated exosystems can be evaluated by simply integrating for a period of one million years or more (depending on the time spans predicted by the planetary formation and evolution theories used).

4.3 Target Selection and Availability

A major factor in the planning and outcome of any planet-finding mission is the fact that we are limited to a very specific, non-isotropic population of possible target stars. Each detection method has limitations on what stars can be used, and a specific preferred type of star, limiting which stars can be added to the target pool. For example, for direct imaging of Earth-like planets, we need nearby stars so that we can observe large portions of their habitable zones. We would also like to have low exo-zodi contributions, making older stars preferential, and would likely concentrate on F, G, and K stars similar to the sun. On the other hand, if we were looking for self-luminous giant planets, we would preferentially target young systems with hot A and B stars. Instrument characteristics and mission goals will always drive the initial pool of stars, but the final target list has an equally dramatic effect on the mission science returns.

For direct imaging, we always start with a list of all of the stars about which a planet from

the planetary population of interest could, at some point in its orbit, be observable by our instrument. For example, if we are interested in Earth-twins in the habitable zone, we leave only those stars where a randomly oriented habitable zone orbit leaves the planet outside the instrument's projected IWA for a time longer than the required detection integration time for that star (as defined in §2.1.2). In order to decide whether a given star can have detectable planets in a region of interest (defined by some range of a and e), we need to find the smallest possible Δmag for a planet in this region that corresponds to a value of s greater than the projected IWA. This is done as follows:

1. Find β^* that minimizes $\Phi(\beta) \sin^2(\beta)$.
2. Define the range of orbital distances as $r = [a_{min}(1 - e_{max}), a_{max}(1 + e_{max})]$.
3. Define $s_{\beta^*} = r \sin \beta^*$ and $\hat{r} = \text{IWA} * d / \sqrt{L} \sin \beta^*$ - the minimum r such that $s_{\beta^*} > \text{IWA} * d / \sqrt{L}$.
4. If $\hat{r} \in r$ then $s = \text{IWA} * d / \sqrt{L}$ and $\beta = \beta^*$.
5. If $\hat{r} < r$ then $s = \min(r)$ and $\beta = \beta^*$.
6. if $\hat{r} > r$ then $s = \text{IWA} * d / \sqrt{L}$ and $\beta = \sin(\text{IWA} * d / \sqrt{L} / \max(r))$.

The s and β values produced by this algorithm represent the minimum value of Δmag for any planet in the region of interest, with the corresponding s value outside the projected IWA. If this Δmag value falls below the limit of the instrument's photometric sensitivity, then the star is considered a candidate.

Additional criteria for the target list may include:

- Stars on the main sequence
- Stars with high life expectancy (B-V Johnson magnitude > 0.3)
- No low metallicity stars (higher probability of planets, see Figure 3.4)

- No stars with high intrinsic variability (to remove major photometric variations)
- No binaries or stars with close companions or nearby bright background stars (since most starlight suppression systems cannot work on multiple sources)
- Maximum apparent magnitude (brighter stars only)
- No young stars (so that planets will have formed)

Young stars may be identified by X-ray emission for stars more distant than 15 pc (detected by ROSAT) [46], by chromospheric variability, by rotational periods of less than 10 days, or projected rotational velocity of greater than 5 km/s. For a typical imaging mission such as the ones in §4.6, we start with the 1612 non-binary stars within 30 pc of our solar system. For an IWA of 75 mas and Δmag_0 of 26, applying the procedure above to the subset of main sequence stars yields 372 possible target stars. With those instrument specifications, 144 of these targets have single-visit completeness values of less than 0.01, and 260 have completeness values of less than 0.1.

The biggest problem in selecting the target list comes from one of the inherent trade-offs in the stated science goals. The desire to visit as many unique targets as possible means that the overall probability of detecting planets is decreased by including lower completeness stars (where the probability of detection is smaller). On the other hand, only visiting a small number of stars automatically limits the number of unique detections that can be made, and if the frequency of the planet type you are searching for (or absolute frequency of all planets) happens to be low, having a large target pool may be the only way to detect any planets at all. An additional problem lies in trying to filter target stars by the amount of integration time required. Although we can calculate the integration time for any assumed Δmag value, in reality, planets from any population of interest can vary by several magnitudes at the time of observation. For example, an Earth-twin observed at one point on its (habitable zone) orbit at a Δmag of 26 will require 200 days for spectral characterization. The same planet, at a different orbital position can have a Δmag of 25, which would require only 30 days of

integration. In both cases, the time required to determine whether a detection has occurred is less than 2 days. Because of the great disparity in detection and characterization times, and since high ecliptic latitude stars have very long observing seasons, improper mission simulation logic can sometimes lead to extremely long integration times which adversely affect overall mission performance.

Because we cannot assume *a priori* knowledge of planet frequency, or what the brightness of planets will be at the moment of detection, it is quite difficult to address these issues using scheduling logic alone. Instead, we can introduce two additional parameters: a minimum target system completeness value, and a maximum single integration time. The minimum completeness gives us a systematic way to limit the size of target lists, mapping directly to the unique targets vs. total detections tradeoff. The maximum integration time allows us to include all potential targets, but not allow any one integration to significantly affect the mission outcome. This value can be used in two different ways: first, the target list is filtered so that no detection time for the worst case (highest zodi level and limiting Δmag) exceeds the minimum integration time. Second, in visits where a detection occurs, but the spectral characterization would take more than the maximum integration time, the spectral characterization is not attempted.

The proper value for each of these parameters is found by performing a line search over multiple simulations with these values smoothly varied to optimize specific target metrics. For all of the instruments and mission scenarios evaluated to date, there is improvement in the number of total and unique detections when the minimum completeness is raised from zero. Specific improvements vary between instrument and spacecraft designs, but all systems perform better with a minimum completeness of 0.1, as opposed to using the full target list. In cases where low completeness stars ($\text{completeness} < 0.1$) do provide detections, these are always very dim, require much longer integration times to acquire a spectrum, and are never detected again. The integration time cutoff is harder to optimize due to the larger search space and the computational costs associated with running many simulations. In certain

cases, the instrument design itself leads to a systematic limit on the length of observations, thereby automatically providing a maximum integration time. In other cases, good results were achieved by setting the maximum integration time based on the total observation time available and the average time to transfer between targets.

For ground based observatories, target availability is given strictly by the observatory location and the time of year. Space telescopes, however, can reside in a variety of orbits, and target availability may differ greatly from one observation to the next. Missions such as Kepler, whose target pool is always in the field of view do not have any scheduling considerations associated with target availability, so the discussion in this section is limited to direct imaging missions and hybrid missions with targeted transit photometry or other planet-finding observations.

In general, whether a target star is observable at a specific point in time can be determined by calculating the angle between the telescope-sun and telescope-target vectors:

$$\theta = \cos^{-1} (\hat{\mathbf{r}}_{S/sc} \cdot \hat{\mathbf{r}}_{S_0/sc}) \quad (4.3.1)$$

where S_0 refers to our sun. For any given telescope, the sun cannot be within some number of degrees of the line of sight ($\hat{\mathbf{r}}_{S/sc}$ - the spacecraft-target vector) or light from the sun would enter the telescope aperture. For a telescope with a completely unshielded primary, this value would be nearly 180° . However, most optical telescopes are designed with the primary enclosed within the spacecraft and many proposed designs (especially for direct imaging) have sun shields that further extend the permissible location of the sun. With these additions, the sun exclusion (or ‘keep-out’) zone can be as low as 45° .

If the optical system is self contained, the sun exclusion angle is the only factor determining whether a star is in the keep-out zone. Unfortunately, systems with external optical elements, such as free-flying occulters, have an additional keep-out zone so that sunlight isn’t reflected by the external elements into the telescope. Because of the nature of these systems,

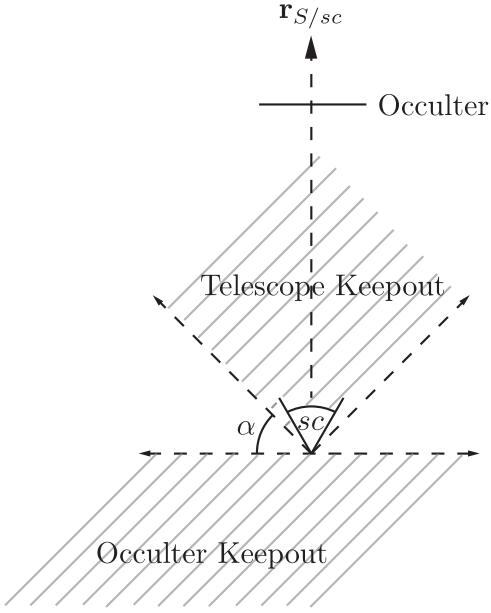


Figure 4.3: Schematic top-view of telescope and occulter spacecraft and associated keep-out zones. An observation can only be made when the sun (and any other sufficiently bright bodies such as the Earth or moon) is in the annulus extending α degrees (the complement to the sun-exclusion telescope angle) from the line orthogonal to $\mathbf{r}_{S/sc}$ at the telescope aperture. If there is no external occulter, then the annulus also includes the 180° below this line.

this keep-out region must cover the full 180° region below the line orthogonal to the line of sight (see Figure 4.3). It is possible to reduce this keep-out by tilting the occulter, at the expense of requiring larger starshades (as the effective shadow region is diminished) [11].

4.4 Decision Modeling

While the assumptions and definitions in §2.1.2 are sufficient to model any single observation with an imaging instrument, they are not enough to automatically generate an entire mission (or survey) timeline. This is primarily due to the limitation of imaging systems to only one target at a time, and the strong stochastic element in the outcome of any one observation. The amount of time spent on any one target will depend on whether an observable planet exists there, and also on how mission rules dictate that time be allocated to spectral characterization versus finding new planets. Because the length of every observation is variable, the

subset of available targets for the next observation (see §4.3) will be constantly changing, and it is thus impossible to assemble an exact observation schedule before the start of operations. For ground-based observatories, there is a natural cadence imposed by the day-night cycle, but it is still possible to optimize operations during a single night of observing, if real-time processing is employed. Our goal is to develop a scheduling algorithm that will be able to make decisions based on the time history of previous observations.

Our automatic scheduler is defined as an optimization problem wherein we seek to maximize the science yield of a mission by simultaneously maximizing the number of unique planets discovered, the number of spectral characterizations of discovered planets, the number of planets observed enough times to produce orbital fits, and the portion of the target list observed at least once. This last goal is included since there is scientific merit in any observation, whether or not a planet is detected (in the form of disk science and concurrent observations by other instruments on the observatory), and this provides a simple way of controlling the decision algorithm’s natural bias towards higher completeness stars. It must be noted that several of these science tasks are inherently antithetical—for example, spending time on finding more unique planets means less time devoted to scheduling additional observations for previously found planets, and thus fewer orbital fits. This conflict is addressed by introducing weights to our cost function, which can be used to tune the relative importance of each science goal.

Previous implementations of automated planning have modeled the order of observations as a traveling salesman problem (TSP) with repeated visits—a useful model as there exists a large body of work on methods for solving it [70]. Unfortunately, it is difficult to create a realistic mission simulation with this approach. Unlike the classical TSP, where path lengths between locations are fixed and the group of locations remains invariant, here, the cost of transferring between a pair of target stars is time dependent, and the amount of time required at each target is variable, and cannot be known *a priori*. For these reasons, even the time-dependent TSP description of the system is inadequate.

Instead, as with a TSP, we represent the sequence of visits as a directed graph with variable length edges. The graph is encoded as an $N \times N$ matrix (for N targets), where element ij represents the ‘cost’ associated with switching from target i to target j (which we shall refer to as ‘transiting’ between targets). We can use a variety of graph search techniques to determine the optimum (least cost) path, by assuming that this matrix will remain constant for k transits (where k is determined by the specifics of the system being simulated), and thus pick the next target as the node generating the least cost path. After a target is observed, however, the entire matrix is re-generated based on the current location of the spacecraft, and the process is repeated, thereby capturing the dynamic nature of the problem. We can also take the approach of simply always going to the ‘best’ (lowest cost) available target, by setting $k = 1$. In cases where the matrix is relatively stable over multiple steps (i.e., the time spent on each target is small), the re-evaluation of the costs does not produce significant changes in the planned path. However, in those cases where one observation takes a long time (i.e., when spectral characterization is required), the costs will change significantly, so it is very important to update the matrix to select the next best target.

The function determining the cost of each transit is a weighted linear combination of multiple factors,

$$A_{ij} = \left[a_1 \frac{\cos^{-1}(u_i \cdot u_j)}{2\pi} B_{inst} + a_2 \text{comp}_j - a_3 e^{t_c - t_f} B_{unvisited} + a_4 B_{visited} (1 - B_{revisit}) - a_5 B_{revisit} \left(\frac{N_j}{N_{req}} \right) (N_j < N_{req}) - a_6 \frac{\tau_j}{\text{vis}_j} \right] / (1 - B_{keep-out}). \quad (4.4.1)$$

The first term represents the cost associated with retargeting, with B_{inst} set equal to 0 for ground-based telescopes and space coronagraphs (since the amount of fuel and time used in retargeting should be approximately constant for an internal coronagraph for any pair of targets). B_{inst} is set to 1 for space telescopes with external occulters, or any other system

where a spacecraft needs to be physically repositioned in the line of sight from telescope to target. While this term is not the actual cost associated with each specific retargeting (which depends on the position of the spacecraft and its orbit), it serves as a good heuristic function. Because the orbits assumed for the telescope and occulter typically require active control to retarget in a reasonable amount of time, and the contribution of the orbital dynamics to the time and fuel costs of retargeting are comparatively small, in most cases this represents an admissible heuristic, and produces good results [70].

In calculating the actual transit time and fuel used, the full orbital dynamics and spacecraft masses are taken into account, and the masses are updated after each transit to reflect fuel use. Following Kolemen [71], transits are modeled either as impulsive thrusts or continuous point-to-point trajectories. In each case, the aim is to find an occulter trajectory such that the telescope-occulter separation is fixed at the end, with inertial velocities matched. In the impulsive case, this is achieved with two large maneuvers at the beginning and end of the trajectory which are taken to change the occulter's velocity while keeping its position nearly constant. This becomes a boundary value problem with the 3 dimensional velocity vectors at the trajectory endpoints as the unknowns, which can be solved via collocation.

In the case where longer thrust times are required (such as with electric propulsion), the strategy is to minimize the control effort given the differential equations of motion and fixed endpoints (a fixed transfer time is assumed), which results in a 12 dimensional boundary value problem. A detailed treatment of this problem type can be found in Stengel [132] and the particular problem is analyzed in Kolemen and Kasdin [70], Kolemen [71]. Fuel use is calculated by assuming that the spacecraft mass is nearly constant during any impulsive thrust maneuver and taking the product of the burn time with the propulsion system mass flow rate, which is determined by the system I_{sp} and thrust force.

The first term (when B_{inst} is not zero) can be efficiently constructed by first generating a

matrix of all of the combinations of the currently available target pool taken two at a time:

$$[c] = \left[\binom{N_t}{2} \right] \quad (4.4.2)$$

where N_t is the size of the currently available target pool. A vector of angular separations is calculated as in Equation (4.4.1):

$$[a] = \frac{\cos^{-1}(u_i \cdot u_j)}{2\pi} \quad (4.4.3)$$

and then indexed via a strictly upper triangular (zero diagonal) matrix of ones to generate an upper triangular matrix of angular distances. The result of this operation, when added to its transpose, generates the required matrix of the first term. A sample implementation is presented in Listing A.6.

For occulters designed to operate at multiple separation distances the slews required in order to complete spectral characterizations introduce high costs in terms of both fuel and time. It is important to remember, however, that there is no effective way of minimizing these costs other than by changing the occulter design, so they need not be included in the transition cost function. Since covering the whole spectral band is usually a mission requirement and observing seasons for most stars are relatively short, we have no choice but to move the occulter in at full thrust if characterizing the whole spectral band is to be attempted during a single visit. Thus, the fuel/time cost of this extra slew is essentially constant (changing only gradually with the decreasing spacecraft mass) and does not affect the transition cost calculation any more than spectral characterizations performed by a coronagraph. The only changes this system requires in the simulation logic are as follows: The occulter is moved for spectral characterization only if the star will remain observable for enough time to perform the slew and the second spectral integration. In cases where the first half of the spectrum is acquired, but not the second, on revisits the occulter is slewed directly to the closer separation distance. Because the smooth trajectories between the two

separation positions do not take significantly different amounts of time or fuel, the heuristic term in the original cost function still applies.

The second term of the cost function serves as a heuristic for the probability of a detection at the next target, with comp_j equal to one minus the completeness of the j th target. The third term is included to promote visits to as many unique targets as possible. Here, t_c is the current time, t_f is the mission lifetime, and $B_{unvisited}$ is a boolean equal to 1 if target j has not been visited. The fourth term creates a preference for targets that are scheduled for a revisit, with boolean $B_{visited}$ set to 1 if target j has been visited, and boolean $B_{revisit}$ set to 1 if target j is currently scheduled for a revisit ('currently', in this case, means \pm the average re-targeting time). The fifth term is included to improve the odds of good orbital fits. N_j is the number of previous detections of a planet at target j , and N_{req} is the number of detections required for an orbital fit. This term has the effect of biasing target selection towards those stars with more prior detections as they near the minimum required number of detections, when they are scheduled for revisits. The inequality expression represents a boolean that eliminates the effect of this term once the required number of detections has been achieved.

Nominally, we set N_{req} to be four, but depending on the system and desired accuracy of derived orbital elements, many more detections may be required. Since the method used to schedule future observations depends on an estimate of the semi-major axis, which is updated with each subsequent observation, we have found that four detections are generally enough to constrain the semi-major axes of planets in the limited population considered here to below 10% error (in some cases to within 1%). Four visits also constrain the possible ranges of the other orbital parameters, giving us the ability to say with high confidence whether the planet is in the habitable zone. Nevertheless, previous work has shown that orbital fitting accurate enough to predict the future positions of planets is very difficult, and may require many more observations, especially when considering more diverse populations of planets [101, 10].

The final term inside the brackets was included after a study of strategies employed in the manual scheduling of planet finding missions. The most valuable of these strategies is to identify the stars closest to ‘leaving’ a keep-out zone, and thus provide the maximum amount of possible integration time for the instrument. This strategy has the additional benefit of creating a schedule in which an occulter does not need to move from target to target at full thrust (since slew times are determined by how much time is left until the star is viewable), thus saving large amounts of fuel. This strategy is achieved by the sixth term, where τ_j is the amount of time before star j enters a viewable zone (this value is negative if the star is already outside of the keep-out zone) and vis_j is the total amount of time star j is continuously viewable. This term has the effect of making stars about to enter a viewable portion of the sky more likely to be selected, while penalizing those that are about to enter a keep-out zone.

The factors a_i are selected to weight the relative importance of the various terms in Equation (4.4.1). For the validation results presented in the next section, these terms were chosen manually, after experimenting with varying values. Obviously, these weights represent a great opportunity for optimization and possibly auto-tuning during the course of a single simulation. However, even with only six parameters, the search space of possible values is quite large, making an optimization remarkably computationally intensive. The weights are also fairly specific to individual instrument designs, meaning that they cannot be carried over to new configurations unless these are fairly similar to previously analyzed ones. Limited optimization, in the form of an unconstrained simplex search [77] over a scalar cost function was carried out for the designs in §4.6.

Finally, $B_{\text{keep-out}}$ is a boolean representing whether a given star is currently in the keep-out zone. Dividing by 1 minus this term leads to infinite costs for stars currently in the keep-out, so they are never selected; the same is done for the matrix diagonal so that the same target is not visited twice in a row. For an occulter system, stars are only considered to be in the keep-out zone if they will be unobservable when the occulter completes its transition

slew. That is, if the amount of time until they become viewable would result in a transit slew using less than some selected factor of the slew time for thrusting (typically, we use 50%), or if they could not be reached at full thrust before entering the keep-out zone. Another way of controlling this is to define a default burn portion—that is, the default portion of the transfer maneuver that is used for accelerating and decelerating (with the remainder used for coasting). This allows for an initial estimate of the slew times required to get to the next set of targets. After the next target is selected, the actual burn portion used can be adjusted as necessary to achieve the desired transfer time. Decreasing this parameter has the effect of saving fuel, as fuel use increases roughly as the square of the acceleration time, while potentially decreasing the number of targets visited. Figure 4.4 shows a graphical representation of a cost function matrix. In this case, all targets are included, leading to many whole columns of infinite values (for stars in the keep-out zone), however, to improve the efficiency of the simulation costs may be calculated only for the subset of stars not in the keep-out zone.

For the instruments considered in §4.6, those with an occulter will have transfer times between targets on the order of weeks, so that k must be quite small—just two or three steps. This makes it easy to compute all possible path costs and take the one representing the true minimum. Self-contained coronagraphs, on the other hand, can acquire a new target in as little as a single day, while ground-based systems can retarget in just minutes, so that it is possible to use a much larger k , which would require switching to a different selection method (such as an iterative deepening search). However, simulation shows that due to the relatively small target pools available, in most cases, just using the first 3 to 5 steps produces the same result as a more exhaustive search. To minimize computing time fixed k values of 3 are used for these instruments, except when directly testing the effects of varying this parameter.

The list of scheduled revisits (used to set the $B_{revisit}$ term in Equation (4.4.1)) must be updated after every observation and is based on the outcome of the current observation (i.e.,

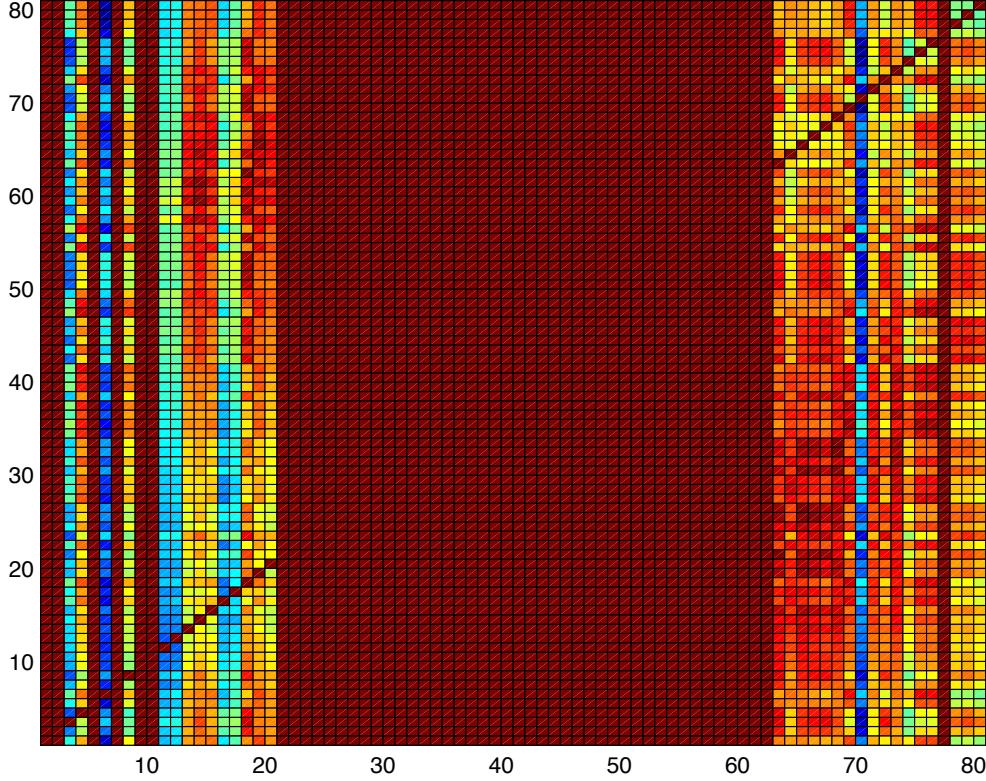


Figure 4.4: Visualization of cost function matrix. All stars (including those in the keep-out zone) were included, leading to multiple columns of infinite cost. The transition cost increases from blue to red.

whether a detection occurs), the previous contents of the list, and the current time. As shown by Equation (3.3.51), if a detection occurs, the most likely estimate for the planet's semi-major axis is the observed apparent separation. As shown in Equation (1.2.13), the orbital period (for a Keplerian orbit) is a function of the semi-major axis and the summed gravitational parameters of the star and planet. If we assume that the mass of the star is much greater than the mass of the planet, then $m_S + m_P \approx m_S$. The mass of the star can be approximated from its apparent magnitude via Equation (4.2.2), and so we can estimate the orbital period from a single detection of the planet.

Because of the geometry of the system (see Figure 2.2 and (2.3)), the best time to schedule a revisit (i.e., the most likely to result in a repeat detection) is either a full or half estimated orbital period after the initial detection. If half of the estimated orbital period is used, this

yields a higher probability that repeat detections will be on widely separated positions on the orbit, thereby improving the orbital fit. If no detection occurs, we must rely on the mean orbital period of the planet population of interest, and schedule the return visit for 3/4 of this mean period. Every subsequent detection can be used to improve our knowledge of the planet's orbital semi-major axis. The probability density function of the ratio of s/a , calculated from Equation (3.3.51) is used to weight each recorded s in order to update our estimate of a [111]. Figure 4.5 shows the revisit detection rates when using these period estimates compared with a constant interval between observations.

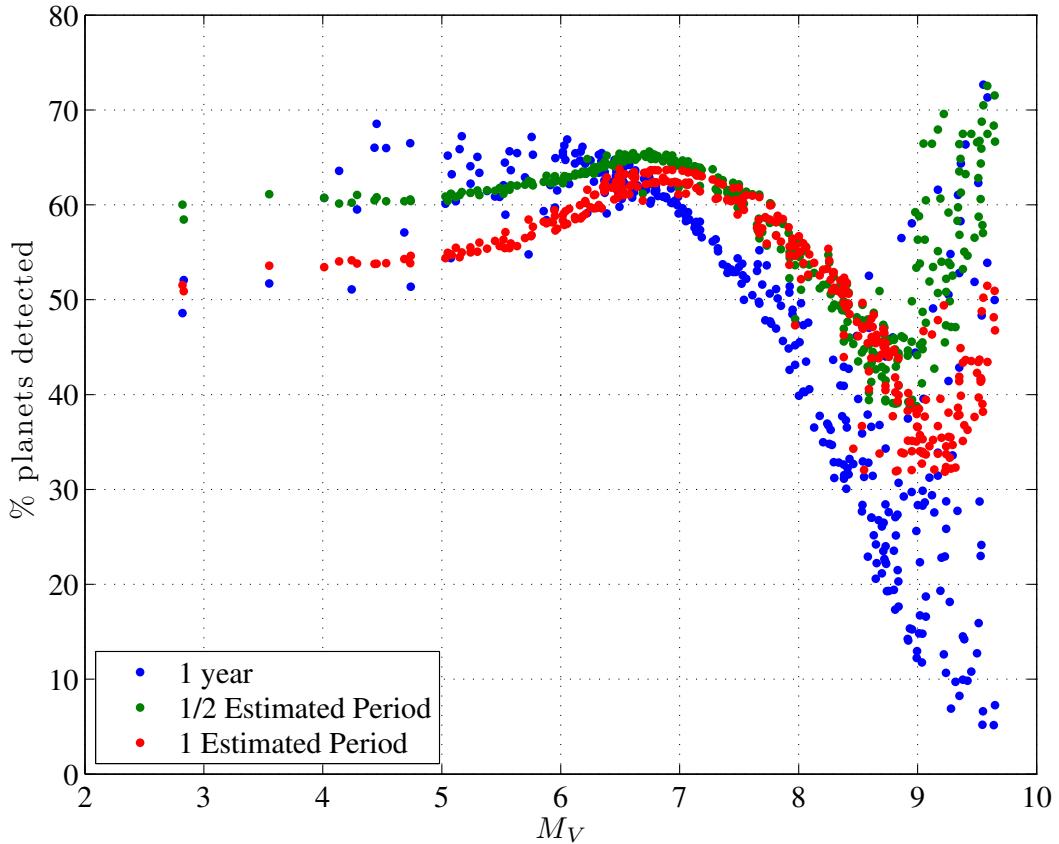


Figure 4.5: Comparison of revisit detection success rates using the estimated half orbital period, estimated full orbital period, and a constant value of one year, as a function of star absolute magnitude. The simulated population is $a \in [0.4, 30]$, $e \in [0, 0.8]$, $p \in [0.1, 0.5]$, $R \in [0.7, 11.2]R_{\oplus}$, with orbits scaled so that planets receive the equivalent flux as they would for a $1L_{\odot}$ star. The constant interval case outperforms the period estimate case only in the case of the brightest stars, where planets are most often visible.

It is important to point out that the return visit strategy may be significantly affected

by the detection of multiple planets during the same visit. If multiple planets are seen at wide angular separations during one observation, assuming that they have small mutual inclinations, it might be safe to assume that the exosystem was being viewed nearly pole-on, and thus would have already captured about as much of the habitable zone as possible, thereby negating the need for future visits if you were only interested in detection and spectral characterization (of course, multiple visits would still be needed for orbital characterizations). Attempting to estimate the frequency at which multiple detections will occur during single observations is fairly difficult, since it depends completely on the makeup of exosystems. Our currently small sample of multiple planet system contains a large fraction of ‘hot jupiters’ due to the biases of the detection methods used. These planets, due to their proximity to their primary stars, would not be detectable with the direct detection instruments studied here, and thus do not constitute a good test. Instead, we can take our own solar system, and simulate randomly timed observations with the plane of the ecliptic randomly oriented with respect to the line of sight, and the system barycenter located a random distance from the observer. Doing so for distances between 10 and 30 parsecs, with observations randomly placed in a 5 year window, and using a 75mas IWA instrument with a limiting Δmag of 26, produces slightly over 30% of observations in which more than one planet could be seen, with only 4% in which more than two are detectable.

Furthermore, because Jupiter and Saturn are so highly detectable, the system orientation was classified close to pole-on (less than 10 degrees rotation) in only 20% of these cases. Thus, while it is probable that multiple detections will occur within single observations of systems structured somewhat like our own, this does not guarantee that we will be able to discount systems after a single visit. Even in cases where we will have good constraints on the exosystem inclination (i.e., when at least one of the planets detected is clearly in the inner system), one could argue that repeat observations as scheduled by the algorithm used here can be useful. First, they will likely be required for confirmation—re-detecting a planet when it has moved noticeably on its orbit makes a much stronger case that the original

detection was not of a background object or structure in the zodiacal dust. Second, while our solar system has very small mutual inclinations, we cannot discount the possibility that there exist stable exosystems with high mutual inclinations [79].

Another very important question that cannot be addressed with the single-planet system model is: in cases where a repeat observation of a target system where a planet has been previously discovered yields another detection, how can we determine whether we are seeing the same planet as before, or a new one? This is often termed as ‘confusion’ and can lead to significant inaccuracies in the gathered data set. It is also extremely important to rule out confusion when attempting to confirm that initial detections really represent bodies that belong to the target star system, and to ensure that the data points we use for orbit fitting actually belong to the same data set.

One way to answer this question is to apply simple Bayesian theory. The probability that a current observation is of the same planet as a previous one is given by

$$P[\mathbf{p}_k | \mathbf{p}_{k-1}, \{\mathbf{p}_k, \mathbf{p}_{k-1}\} \in \gamma_{k-1}] = \frac{P[\mathbf{p}_{k-1}, \{\mathbf{p}_k, \mathbf{p}_{k-1}\} \in \gamma_{k-1} | \mathbf{p}_k] P[(\mathbf{p}_k)]}{P[\mathbf{p}_{k-1}]} \quad (4.4.4)$$

where \mathbf{p}_i is the parameter set recorded in the i th observation (i.e., star-planet separation and difference in magnitude), and γ_i is the space of all possible parameter sets belonging to a single planet that includes \mathbf{p}_i . The likelihood, $p(\mathbf{p}_{k-1}, \{\mathbf{p}_k, \mathbf{p}_{k-1}\} \in \gamma_{k-1} | \mathbf{p}_k)$, is approximately equivalent to the goodness-of-fit of a stable orbit to both data points, while the priors, $p(\mathbf{p}_k)$ and $p(\mathbf{p}_{k-1})$, can be calculated from the completeness distribution of the target star [114]. This technique is demonstrated in §4.6.1.

4.5 Simulation Methodology

Equation (2.1.27) allows us to calculate how much time we need to spend integrating before we can assume, with a specific confidence, that no planet exists (or is currently observable by our instrument). In the simulation framework, we calculate the required integration time

for each target star assuming a planet observed at the IWA and the instrument's limiting Δmag . For missed and null detections, this time value is used as the actual integration time (in a real mission, this would be the time spent integrating before moving on to the next target). For detections, the integration time is recalculated based on the simulated planet's orbital position, using the actual separation (and throughput) and Δmag of the planet at the time of the observation (in a real mission, this would be the point when a detection could be confirmed via a bayesian technique and integration would either be stopped, or switched to spectral characterization). False alarms (when no observable planet exists) and missed detections (when there is an observable planet) are generated at the rate determined by the FAP and MDP values used in calculating the integration time in Equation (2.1.27). Typically, the FAP and MDP are set to 1% divided by the expected number of observations for the whole mission, and 0.1%, respectively, which tends to yield zero missed detections and at most one or two false alarms per mission simulation for all of the cases considered in §4.6.

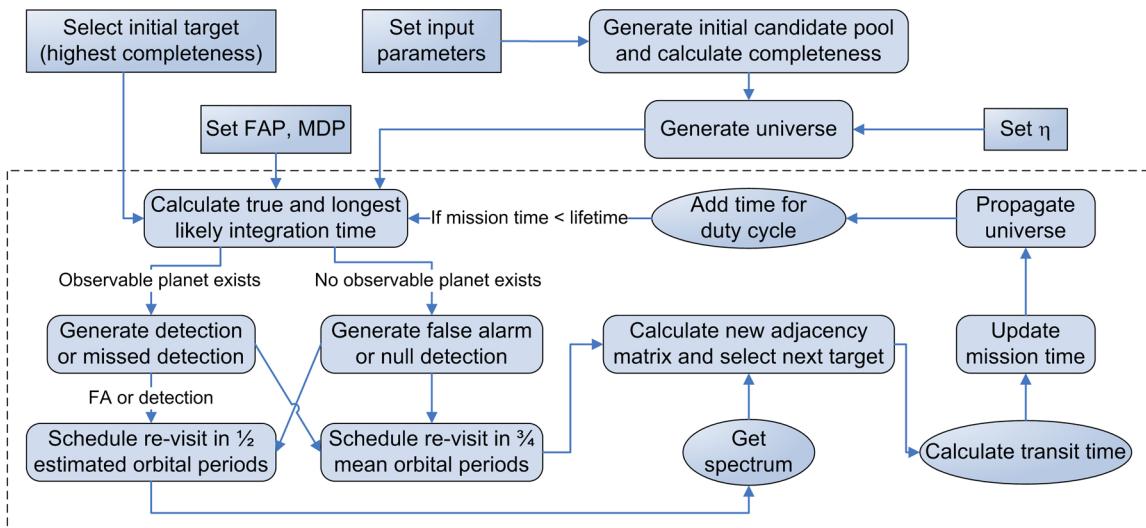


Figure 4.6: Flowchart of simulation framework. Ellipses represent optional steps determined by the specific instrument or rule set.

Figure 4.6 shows a flowchart for the full implementation of the simulation framework. The basic steps of the simulation are as follows:

1. Generate single visit completeness values for all stars for the assumed planetary population.
2. Generate target list based on given parameters and single visit completeness. Assign planets to some subset of the target list based on selected planet frequency.
3. Generate the observatory orbit (including Earth rotation for ground-based instruments) and find all times over mission lifetime when targets will be observable.
4. Calculate maximum integration times for each target assuming planets at the IWA and Δmag_0 , with the highest possible exo-zodi level.
5. Select initial target by choosing currently observable target with highest single visit completeness.
6. While the current time is less than the mission lifetime:
 - Update visited list with current target.
 - If there is an occulter, calculate the disturbance forces acting on it at its present location.
 - If a planet exists in the target system, find its current angle on the sky and Δmag , the amount of integration time required to detect it at the false alarm and missed detection probabilities being used. If the planet will remain outside the IWA and be brighter than Δmag_0 for this amount of time, it is observable. Otherwise, use the maximum integration time for this target, and increment the mission time by the integration time. If there is an occulter, decrement the spacecraft mass for stationkeeping for the integration.
 - If the planet is observable, generate a false negative (with probability of the missed detection rate) or a detection (with the complement probability 1). If the planet is not observable, generate a false positive (with probability of the false positive rate) or a non-detection.

- If a detection or false positive has occurred, and spectral characterization is called for, calculate the integration time required. If the target star will be out of the keep-out for this amount of time, increment the mission time (and, if there is an occulter, decrement the spacecraft mass). If the planet is observable for this whole time period, record a successful spectral characterization. For coronagraphs, calculate and record the maximum observable wavelength. If using a multiple distance occulter, attempt the second characterization in the same way.
- Add star and revisit time to the revisit list.
- Calculate new cost function matrix based on current time and select next target. If there is an occulter, calculate slew time and decrement spacecraft mass with the amount of fuel used.
- If a portion of the mission is devoted to some other science, increment mission time by proportional amount. Add spacecraft settling time before beginning next observation.
- Propagate all of the planets in all target systems along their orbits by the total amount of time that has elapsed in this iteration step.

At each step of this iteration, the state of the mission and all activities is encoded using the variables listed in Appendix B. To check whether the final path generated is reasonable, and to allow for easier visualization of a whole mission timeline, we can plot the sequence of observations on a map of the sky. A sample of this is shown in Figure 4.7. Because of the large number of inputs and parameters required to run a simulation, a graphical front-end was created to manage all of these. Figure 4.8 shows the main panel of the application, called ‘DRMconstructor’ used for setting up and running simulations.

Because this process must be repeated many times in order to generate the ensembles of mission timelines required to calculate the distributions of various performance metrics, it is reasonable to attempt to parallelize as much of it as possible. Fortunately, as each

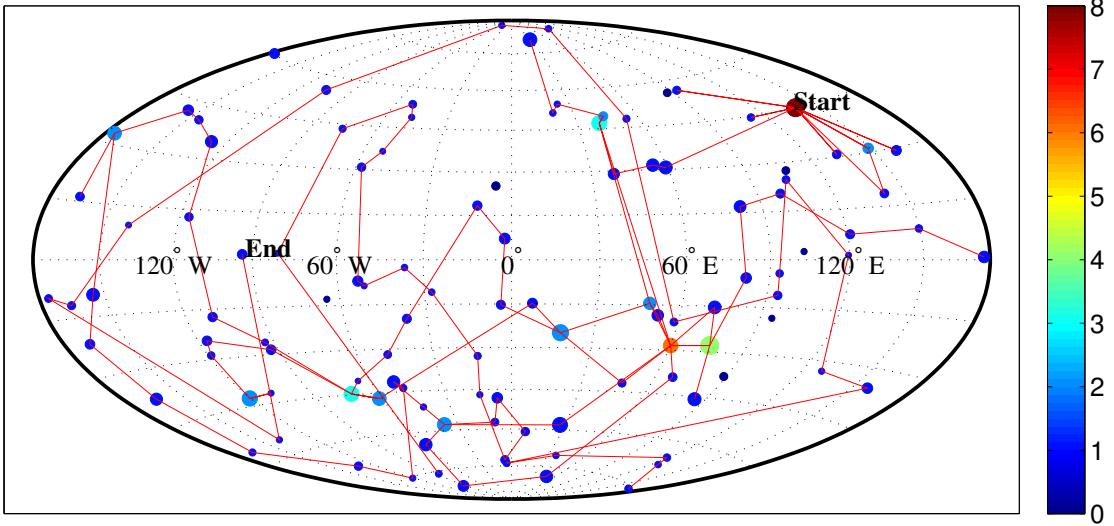


Figure 4.7: Visualization of an automatically generated sequence of observations. Points represent target stars, with size correlated with single visit completeness (i.e., the larger the point, the higher its single visit completeness). The color scale represents the number of observations made of a system in this simulation. Note that some systems get many visits not because of their high completeness, but because they are well located on the sky.

timeline can be evaluated independently of the others, generation of an ensemble is inherently parallelizable to an arbitrary number of processes by simply assigning each new simulation to the first available process thread, or by pre-assigning an equal number of simulations to each thread. The simulations in §4.6 were all carried out on a multi-core, multi-processor machine, utilizing a maximum of 8 concurrent threads and a single local scheduler, but the basic framework can be easily extended to large distributed computing environments by substituting a remote scheduler.

4.6 Case Studies

The purpose of this section is to demonstrate completed analyses of direct detection missions using the techniques described throughout this chapter. Because one of the most important goals of planet-finding is the discovery of Earth-like planets, we use, as the base case, populations of Earth-twins. Unless otherwise noted, the simulated systems in this sec-

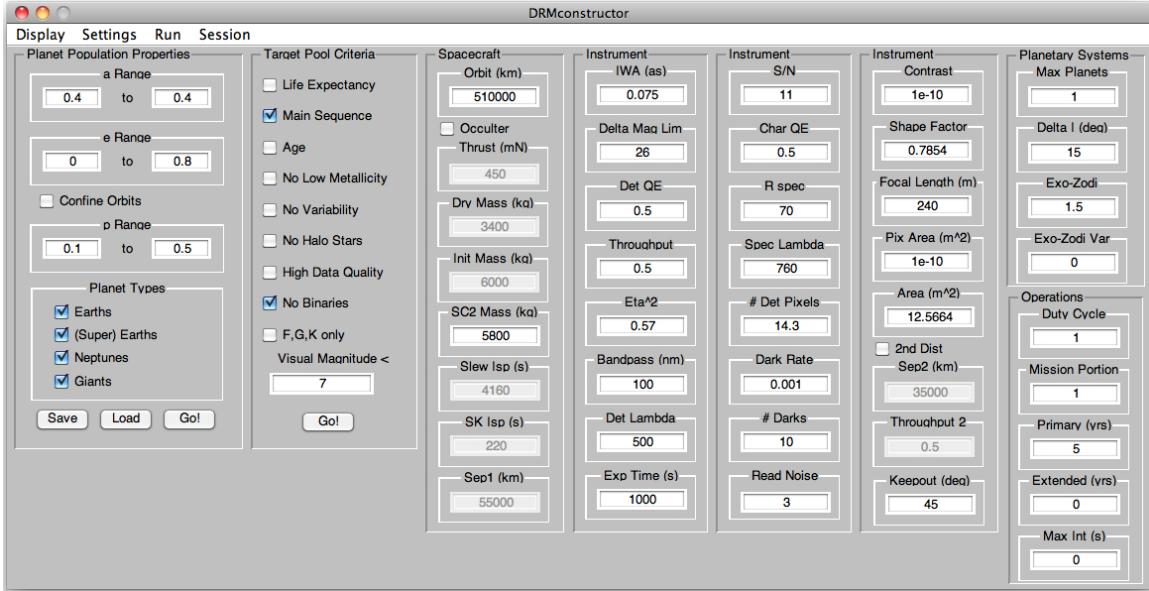


Figure 4.8: Graphical front-end of the DRMconstructor simulation utility.

tion are populated with Earth-twins distributed throughout their habitable zones (defined as $a \in [0.7, 1.5]\sqrt{L}$ and $e < 0.35$) with either zero or one planets per system. The term ‘Earth-twin’ indicates planets of the radius and mass of the Earth, with a constant albedo of 0.22 (this value was selected to conform with previous studies, including Lindler [80]) with isotropically distributed orbital orientations. Exo-zodi levels are taken to be constant, with $\mu = 1.55$ to correspond to the historical average zodi brightness [73]. When spectral characterization is possible, default mission rules call for one complete spectrum to be collected for each unique planet (meaning that spectral characterization is not performed during revisits unless a full spectrum was not taken at the initial detection, or the mission rules are modified). The settling time (i.e., the time to acquire a new target and begin observing) for all spacecraft is taken to be 24 hours. A full listing of the default parameters is available in Appendix C.

All of the spacecraft are assumed to reside on a halo orbit about the Earth-sun L2 point. For occulter systems (except in the case of O₃, see below) the telescope is placed on the halo orbit, which is assumed to be stable enough that only minor thrust corrections are required to keep the spacecraft on the orbit over the life of the mission. The orbit of the occulter

spacecraft, however, must be actively adjusted throughout the course of the mission, both to reposition the spacecraft for observing new targets (slewing), and to maintain precise alignment between the two spacecraft (stationkeeping). The methods for generating stable halo orbits, as well as techniques for optimal point-to-point transfers about these orbits are developed in Kolemen [71]. An example of the nominal halo orbit (510000 km in the azimuthal direction) is shown in Figure 4.9.

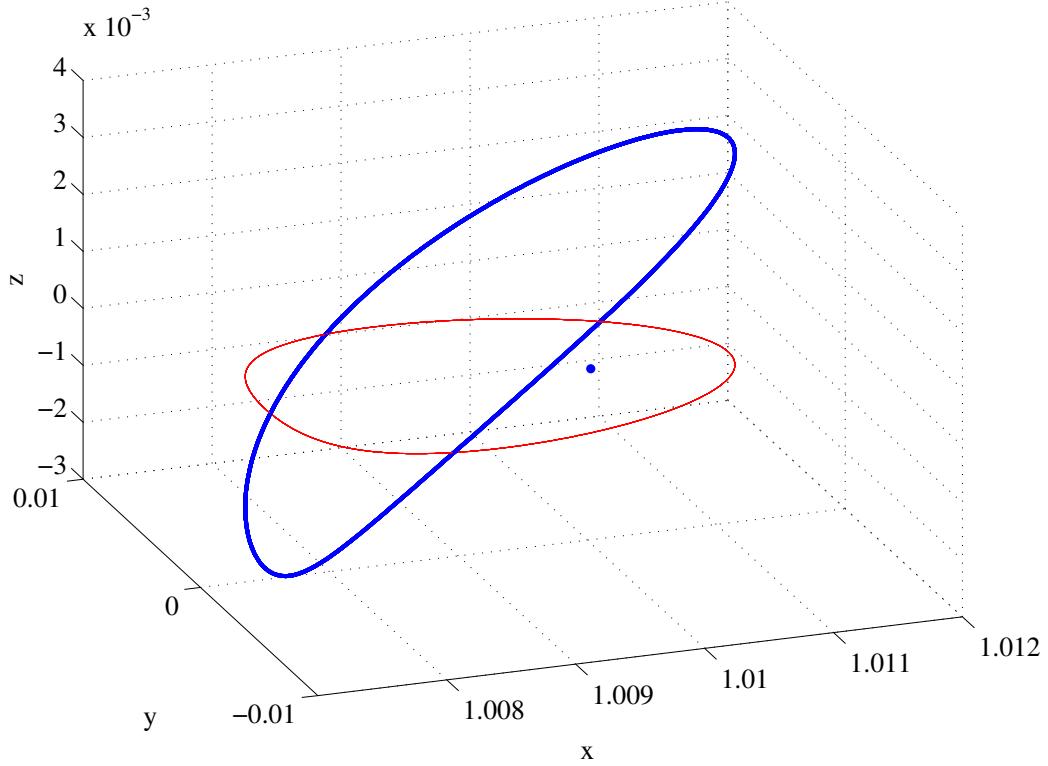


Figure 4.9: Nominal halo orbit (510000 km azimuthal radius). The blue point is the Earth-Sun L2 point, while the red curve is projection of the orbit into the plane of the ecliptic. Note that the out of plane component of the orbit is much smaller than the in-plane components. Units are AU and the coordinate system is heliocentric, with z perpendicular to the ecliptic, and x increasing away from the sun

Among the instruments and missions considered, some represent specific mission concepts that are relatively well developed (with a large number of design details and requirements available for the simulation), whereas some represent more generalized missions intended for high-level comparisons of different instrument types. The various concepts are listed below, along with the basic parameters used to describe each. Specific settings that were varied for

certain instruments for particular tests or to maximize different metrics are detailed along with the simulation results in the following sections.

1. Generalized Single-Distance Occulter (SDO) - The baseline occulter mission consisting of a circular aperture telescope and an occulter spacecraft operating at one separation distance from the telescope. This type of occulter is designed to provide high contrast across the entire spectral band and is therefore larger than occulters designed to provide equivalent contrast over smaller bands [16]. Different propulsion systems may be employed for different designs (especially at differing telescope scales), but the baseline uses the same propulsion system as THEIA (see below). Multiple designs have been created for telescopes of various aperture sizes, and are detailed in Table 4.1.
2. Generalized Multiple-Distance Occulter (MDO) - Essentially equivalent to the generalized SDO, the MDO is designed to achieve high contrast over two (roughly) equal bands at different distances. The majority of these designs use a second separation distance equal to half of the original, thereby operating at double the IWA for higher wavelengths. See Table 4.1 for details.
3. Generalized Coronagraph - The baseline design for self-contained coronagraphs. While not specifically modeling any one coronagraph type, this represents any of a number of high-throughput, low-IWA coronagraphs currently under development [49, 45]. Typical IWAs are either 2 or 3 λ/D , which corresponds to between 26 and 103 mas for a 4 m telescope—comparable to the geometric IWA of the baseline occulter for a 4 m aperture and detections in the V band. By default, the coronagraph missions devote 50% of their time to planet-finding activities (as opposed to coronagraphs, which typically use less than 33%).
4. Telescope for Habitable Exoplanets and Interstellar/Intergalactic Astronomy (THEIA) Exoplanet Characterizer (XPC) [129] - An MDO for a 4 m telescope developed as a flagship mission class concept study whose primary goal is the detection and spectral

characterization of Earth-like planets. The mission design yields high-contrast spectral capability over the 250 - 1000 nm band, split into two sub-bands, with geometric IWAs of 75 and 118 mas, respectively. This allows for the detection of molecular oxygen at high confidence levels. The mission rules developed to address the primary mission goal call for attempting to maximize the number of unique detections while heavily biasing the scheduling towards multiple revisits to a small candidate sub-pool of good targets (ones with previously discovered bright planets) to ensure at least a few orbit fits. A full spectrum (to resolution of $R=70$ at 760 nm) is also required for each unique detection. The THEIA occulter uses paired NEXT ion thrusters [100] to provide up to 450 mN of thrust at a specific impulse of 4160 s. Stationkeeping is performed with hydrazine thrusters with $I_{sp} = 220$ s. The separate stationkeeping system was added due to concerns that the plumes produced by the electric propulsion system would be bright enough to interfere with observations.

5. Occulting Ozone Observatory (O_3) - An SDO or MDO for a 1.1 - 2 m telescope developed as part of a mission concept study aimed at finding a less expensive (~\$1 billion) planet-finder. While O_3 can cover the same band as THEIA, the smaller aperture makes high resolution spectroscopy unreasonable, so O_3 focuses on band photometry, while retaining the ability to detect ozone, another biomarker [53]. When operating as an SDO, O_3 does not have photometric sensitivity beyond 550nm. As an MDO, O_3 can operate in the ‘blue’ band (250 - 550 nm) or the ‘red’ band (500 - 1100 nm). See Table 4.2 for details on the photometric bands [114]. Unlike the other occulters, O_3 ’s occulter spacecraft remains in an unperturbed orbit while the telescope spacecraft is repositioned between observations, using chemical thrusters providing a thrust of 500 mN at an I_{sp} of 600 s.

The point spread functions for the optical systems of these instruments are generated using different algorithms. Throughput curves for coronagraphs are created by sequentially generating tilted wavefronts at the telescope aperture and applying propagation algorithms

Table 4.1: Occulter Designs.

Telescope Diameter (m)	Occulter Type	50% Throughput IWA (mas)	Starshade Mass (kg)	Petal Length (m)
2	SDO/MDO	69.6	2300	7
	SDO	59	4200	19
4	MDO	57.5	3370	10
	SDO	56	7180	24
8	MDO	53	4915	13.5
	MDO	47	10022	21.5
16m				

Telescope Diameter (m)	Occulter Type	Starshade Radius (m)	Separation Distance (km) [†]
2	SDO/MDO*	17	38960/19480
	SDO	25.6	70400
4	MDO [‡]	20	55000/35000
	SDO	35.2	96800
8	MDO	27.2	74800/52360
	MDO	43.2	118800/83160
16m			

[†] For MDOs, the first separation distance is used for covering 250-700nm, and the second for 700-1000 nm, except for O₃, which is split into the ‘blue’ and ‘red’ bands in Table 4.2.

* This is the 1.5 - 2 m O₃ design.

[‡] This is the final design for THEIA. The same design can also be used with a 1 m telescope.

to these. For occulters, we use the Bessel expansion described in Vanderbei et al. [143] via a fast algorithm to calculate the electric field at the aperture of the telescope, assuming an incident plane wave on the occulter. The field at the telescope aperture following a tilted incident wave may be related to the field from an incident plane wave by:

$$E_{\text{tilt}}(x, y) = e^{-ik \sin \alpha x} e^{ik z \cos \alpha} E(x + \sin \alpha z, y) \quad (4.6.1)$$

which holds for small angles and allows the Bessel expansion for plane waves to be used. This field may then be fed into the appropriate propagation scheme. (Propagation through the telescope with no coronagraph is treated as just a Fourier transform.) In general, we prefer

Table 4.2: O₃ filters.

Filter #	Bandpass (μ)	
	Blue	Red
1	0.25 - 0.31	0.5 - 0.7
2	0.31 - 0.4	0.7 - 0.9
3	0.4 - 0.48	0.9 - 1.1
4	0.48 - 0.55	Narrowband [†]

[†] The specific narrowband filter selection will be informed by results from transit photometry, which should help pinpoint the most interesting and common giant planet near-IR features, but will most likely coincide with a methane absorption feature.

to examine throughput in terms of total energy in pupil planes rather than image planes, as the finite extent of the pupil planes ensures a more accurate measure of energy [67]. It is important to note that the throughput curves shown in Figure 4.10 are idealized, and achieving these levels of performance will require meeting exacting manufacturing and positioning tolerances for the occulters, and very precise wavefront control for the coronagraphs.

4.6.1 Validation

Before we can use the results from mission simulations based on the decision modeling described in §4.4, it is necessary to determine whether the cost function described by Equation (4.4.1) actually produces optimal (or close to optimal) results. Of course, the definition of optimality in this case depends entirely on what we want the mission to achieve, and how we prioritize the various science deliverables (i.e., how we set the weight factors a_i in the cost function). Still, if the cost function operates as designed, and all weights are set to be equal, we would expect the resulting simulations to maximize science yield while minimizing operational cost. Since there is a direct tradeoff between time spent on new detections and time spent on characterizations, and between fuel use and available observation time, we do not expect a globally optimal solution for any of these. Instead, we want to always be able to find the easily detectable planets in every universe, and to get orbits and spectra for as

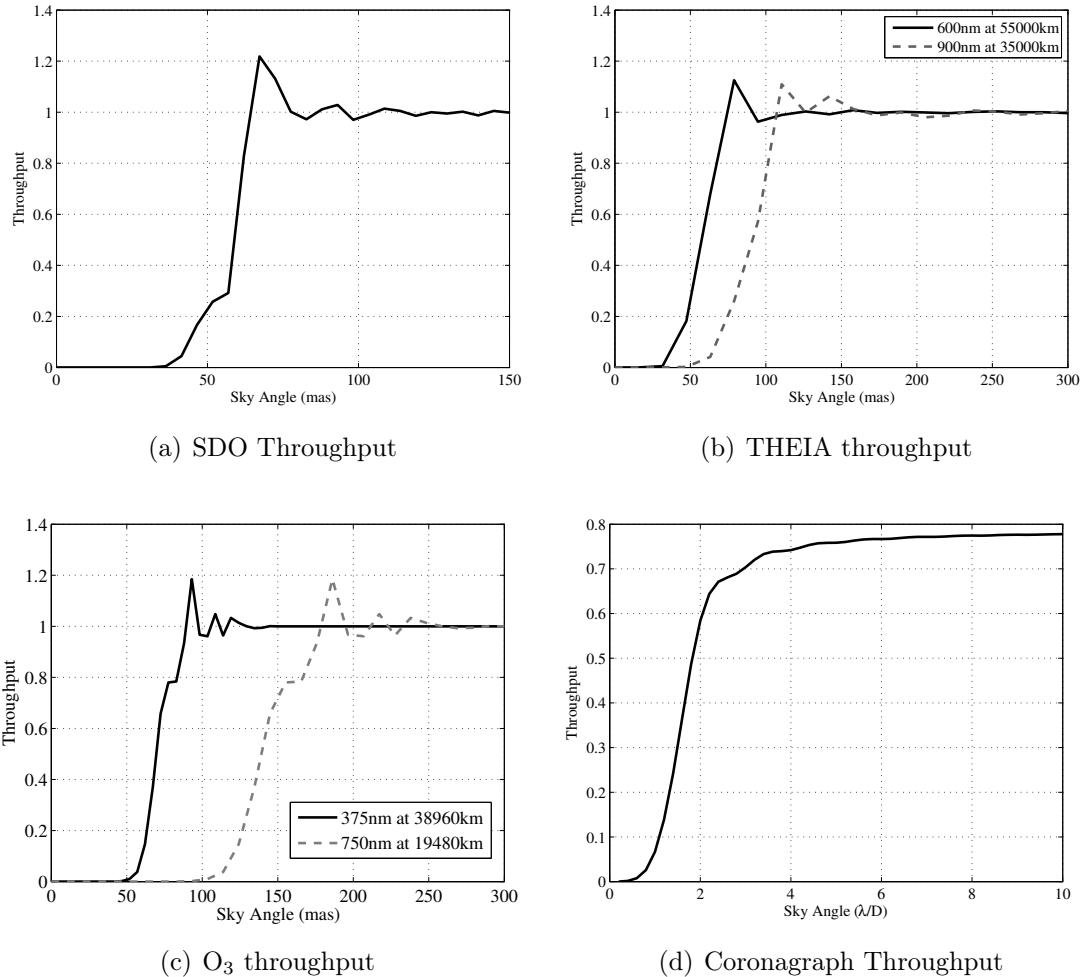


Figure 4.10: Throughputs as functions of angle on the sky (units of mas for occulters and λ/D for coronagraphs). The throughput may be greater than unity for occulters since at certain angular separations, the occulter petals will send planet light into the telescope aperture that would otherwise have bypassed it. Data for (d) was provided by L. Pueyo (Personal Communication, 2009).

many of these as time permits, while reducing fuel use in designs with occulters.

To test whether the framework described here accomplishes this, we can generate a ‘ground truth’ case in the form of the distribution of all possible mission scenarios with a fixed set of target systems. To do so, we generate one single universe, and compare the results of the mission simulation for this universe using our automated decision modeling to mission simulations with randomly generated visit orders. Specifically, rather than using our decision modeling algorithm to decide on the next target, we simply select one at random

from the list of targets currently outside of the keep-out zones. By running this randomized visit order mission on the same universe multiple times, we can map the space of possible observing sequences and find a distribution of mission results. We can also use a modified version of our framework that always selects the next available highest completeness target, rather than using the cost function in Equation (4.4.1).

Figure 4.11 shows the resulting histograms for the number of unique planet detections and successful spectral characterizations found from running 1000 randomized visit order missions on a single universe with $\eta_{\oplus} = 0.5$, using the THEIA and $2\lambda/D$ coronagraph mission concepts. On top of these, we have the results of the missions produced by the decision modeling algorithm with a depth of search (k) of either 3 or 1, and the results of the mission which only considers completeness in selecting visit order. For THEIA, using a k of 3 and 1 produces the same values for both these metrics, whereas the $k = 3$ case outperforms the $k = 1$ case in terms of unique detections for the coronagraph. All cases using the scheduling algorithm described in §4.4 outperform schedules based on just completeness.

Figure 4.12 shows a scatterplot of THEIA occulter fuel use vs. the number of unique planets found for the 1000 randomized visit order missions and the automated visit order missions. Here we do see a difference between the $k = 3$ and $k = 1$ simulations, with $k = 1$ producing a mission timeline where more fuel is used. We compare only the amount of fuel used in slews, rather than total fuel use, since the stationkeeping fuel is directly linked to the number of observations and spectra and is automatically maximized along with these metrics. This validation technique was applied to multiple randomly generated universes, with varying values of η_{\oplus} , producing similar results in each case.

These results show that the automated mission does not represent a global optimum for these metrics. There are a small number of missions which find more unique planets, or get more spectra. However, we can see that the automated mission does a very good job of balancing these metrics and maximizing (or minimizing, in the case of fuel use) them jointly. Allowing for the fact that the automated mission needs to be generated only once

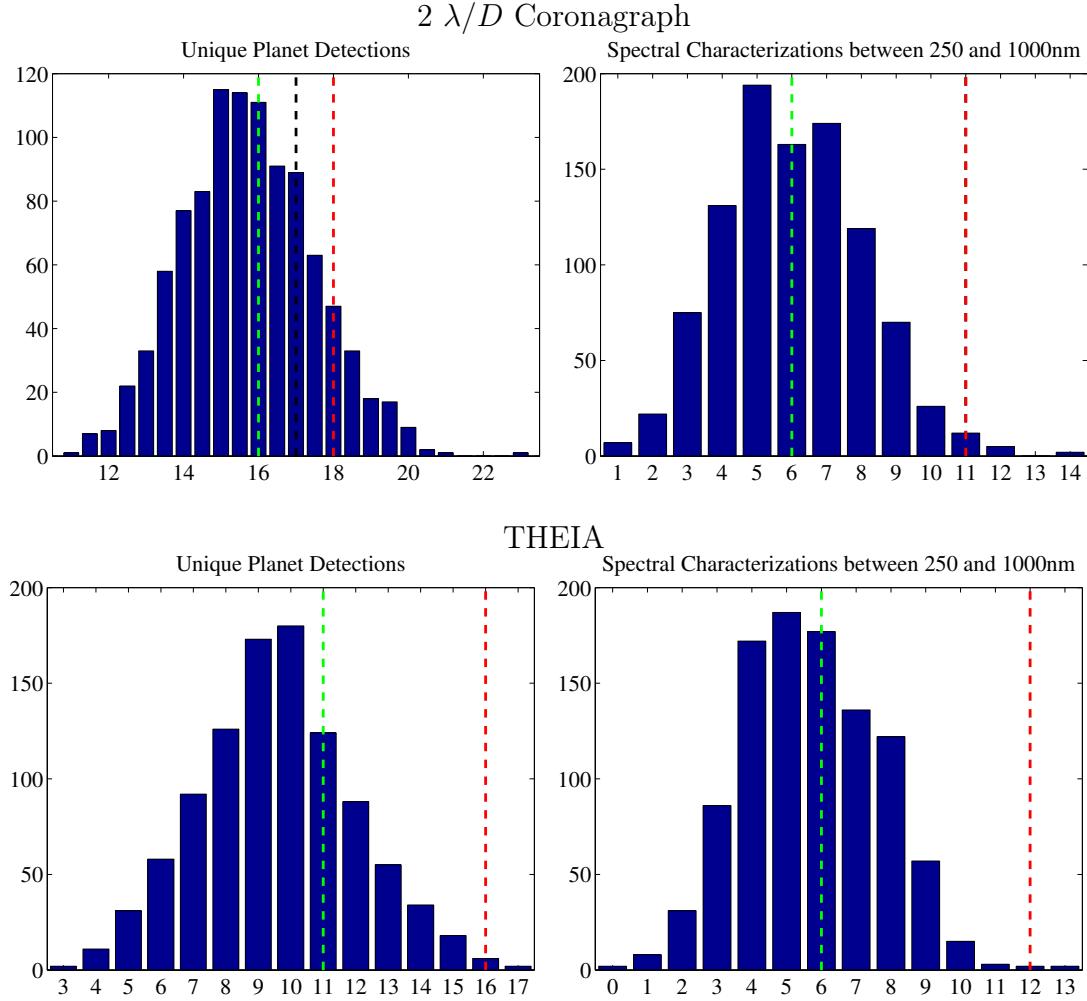


Figure 4.11: Comparison of science yield from automated and randomized visit order selection for $2 \lambda/D$ Coronagraph (top row) and THEIA (bottom row). Blue bars are histograms of results from 1000 mission simulations using randomized visit order. Black dashed lines are results from the automated visit order with depth of search $k = 1$, red lines are results for automated visit order with $k = 3$, and green lines represent timelines generated by selecting the next available highest completeness target. In the spectral characterizations for the coronagraph, and both metrics for THEIA, the $k = 1$ and $k = 3$ cases produced identical results.

per universe, whereas a random walk search for local optima takes many trials, the benefits and efficiency of this approach become quite clear. In terms of real mission planning, a random walk approach is simply not an option (since we only get one chance at scheduling the mission). We also see that even when using a $k > 1$ (looking multiple steps into the future) gives us the same science yield as only considering one step, we still end up with a

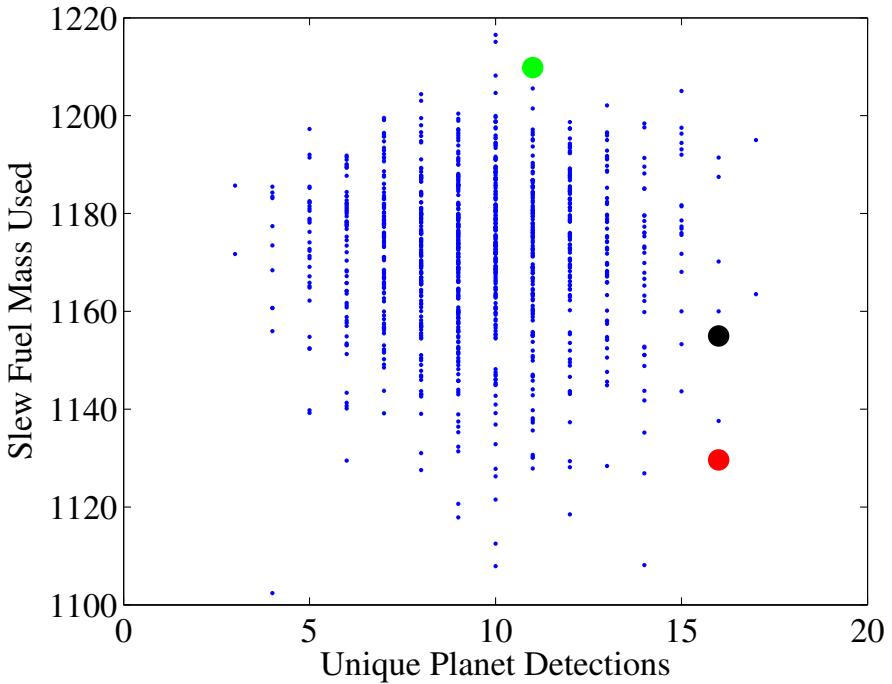


Figure 4.12: THEIA occulter propellant use (in kg) vs. the number of unique planet detections for 1000 mission simulations using randomized visit order in one simulated universe. The red point represents the mission timeline generated using the automated visit order for the same universe for $k = 3$, and the black point for $k = 1$. The green point represents the timeline generated by selecting the next available highest completeness target.

lower occulter fuel cost. Finally, it is important to note that the margin between the unique planet detections using our scheduling algorithm versus the selection of the next available highest completeness target is smaller for the coronagraph than for THEIA, and is further from the tail of the distribution of randomized visit orders. This is because the values for a_i were fixed across all of the simulations here, and the k values used are actually quite low for coronagraphs. Optimizing the weights specifically for coronagraphs and using higher k values produces improvements comparable to those seen in the occulter cases.

Another important question is how many mission simulations are required so that the statistics extracted from the ensemble are accurate with high confidence. It is difficult to make the case that the various metrics that we are interested in are approximately normally distributed, especially as many of them (such as the number of unique planets found) are dis-

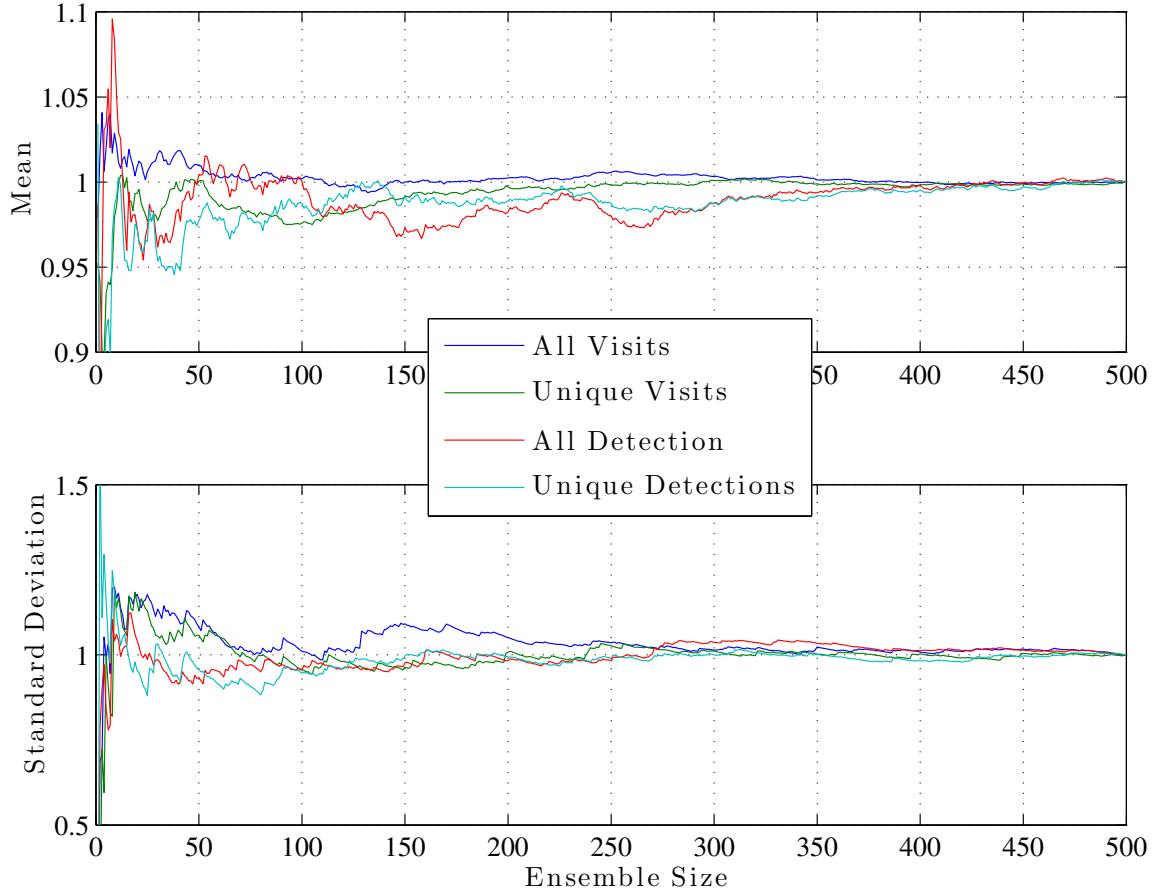


Figure 4.13: Means and standard deviations for four metrics (total number of stars observed, number of unique stars observed, total number of planetary detections, number of unique planets detected) as a function of ensemble size. These are normalized by the values for the full ensemble of 500 simulations. Each simulation is of a 5 year THEIA mission with the default parameter set, with a universe of Earth-twins with $\eta_{\oplus} = 0.5$.

crete values, making it difficult to select a sample size corresponding to a specific confidence level. However, this question can be answered empirically, by generating a large number of simulations and calculating distribution moments for the various metrics from subsets of these of varying size. Figure 4.13 does exactly this, using 500 simulations of a 5 year THEIA mission with an Earth-twin population at a constant η_{\oplus} of 0.5. For all of these metrics (and others not plotted here), the sample mean and standard deviation converge to within 5% of their respective values for the 500 simulation ensemble at around 100 samples. This holds for a variety of configurations and instrument types, and so 100 samples is taken to be the minimum number required by the simulation framework.

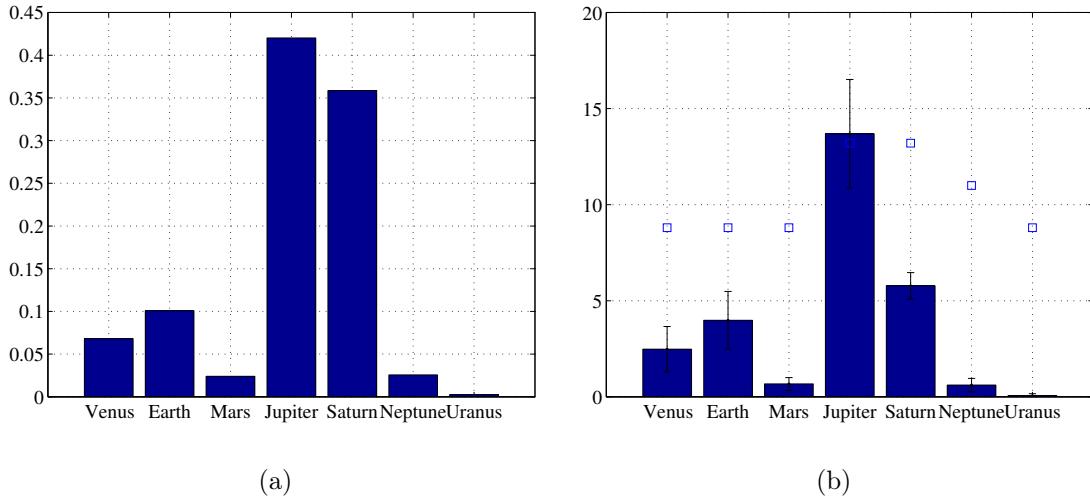


Figure 4.14: Results for 100 mission simulations using the 2 band photometry, 2 m O₃ configuration. (a) Total number of detections of each solar system planet analogue, normalized by the frequency with which they were simulated. This is used primarily as a test to ensure the multi-planet simulation capability is working as expected, but can also be useful in determining the relative detectability of the analogues - i.e., if Jupiters and Earths are distributed with equal frequency, O₃ should detect 4 times as many Jupiters as Earths, on average. (b) Average number of unique analogues found. Error-bars represent the standard deviation of the distribution. The squares indicate the average number of each analogue type that was included in the simulations.

Finally, in order to validate the multiple planet classification algorithm in Equation (4.4.4), we can generate systems populated with solar system planet analogues (as described in §4.2.1) and have the simulation algorithm classify detections made during repeat observations of targets where planets had previously been detected as newly discovered planets whenever the probability of a repeat detection is less than 50%. The test was run using a 44 star target pool with a 2 m telescope O₃ configuration and 2 band photometry required for each planet classified as a new detection. Figure 4.14(a) shows a histogram of the raw number of detections of each solar system planet analogue, normalized by the frequency with which they were simulated. This is essentially the same as counting the total number of detections of each planet type in simulations where all of the analogues are selected with equal frequencies, and gives a sense of the relative detectability of the various planets in our solar system.

As expected, Jupiter and Saturn-like planets, due to their size, relatively high albedo and placement, will be detected much more frequently than any other planet type. Earths will be detected more frequently than Venus or Mars analogues due to their size and placement, and ice giants like Neptune will be difficult to detect unless they orbit much closer to their stars than Neptune does to the sun. This histogram also implies that if Earth-twins and Jupiter-twins occur at the same rate in our universe, an observatory like O₃ should detect roughly four times as many Jupiters as Earths. We can also use this data set to estimate the probability that a habitable zone candidate (i.e., a detection with an apparent separation less than the outer edge of the target star's HZ) actually is a habitable zone planet. We find that if every exosystem was a scaled copy of the solar system, HZ candidates would represent about 22% of all detections, and would actually be Earth analogues 43% of the time. They would be rocky planet analogues (Venus, Earth, or Mars) 83% of the time, and giant planet analogues (Jupiter, Saturn, or Uranus) 17% of the time.

Figure 4.14(b) shows the average number of each planet analogue classified as unique over 100 simulations. The algorithm correctly identified planets as new detections in approximately 60% of applicable cases (detections during repeat observations of targets with prior detections). This led to the apparent discovery (according to the algorithm's internal bookkeeping) of approximately 45% of included Earth-like planets, whereas earlier single-planet system studies indicate that this value should be less than 33%. At the same time, according to the algorithm's classification, it discovered slightly more Jupiters, on average, than were included in the simulation, while in reality the true average rate of discovery was closer to 85%. Because of the limited target pool size, each star in the pool was observed at least once in most of the simulation instances, and the high visibility of Jupiter analogues made them easy to detect.

4.6.2 Results

For each instrument an ensemble of mission simulations was generated over simulated universes containing only Earth-twin planets, a fixed mission length of 5 years, and fixed target lists selected from a pool of real stars within 30pc of the solar system. Each ensemble consists of 1000 simulations with varying assumed frequencies of Earth-like planets (η_{\oplus}). Figures 4.15 to 4.17 show the median values of the four science yield metrics for the four instrument designs for telescopes with circular apertures of 4, 8 and 16m diameters. The error-bars represent the one-sided deviations of each distribution. No SDO was evaluated for the 16m telescope because the starshade is so large that little significant science can be achieved during a 5 year mission with the propulsion system used here. Results for O₃ are plotted separately since O₃ does not have a spectral characterization capability, making it difficult to compare directly with the other instruments that do. The lines in the various plots represent median values for 100 simulations at each value of η_{\oplus} . Error-bars represent the one-sided deviations of each distribution. ‘All Detections’ are the total number of planets found (including multiple detections of the same planet), ‘Unique Detections’ are the number of unique planets found, and ‘Unique Targets’ are the number of unique target systems visited during the mission.

These figures illustrate perfectly the inherent tradeoff between telescope size and achievable science with an occulter. As larger starshades become necessary, unless there is a significant improvement in the propulsion system, the science yield for a fixed mission length decreases. Although it is possible to achieve higher thrust with more power this requires increasing the size and mass of the power subsystem, and leads to another tradeoff. This trade study is an important future task, and one for which this framework is well suited.

At the 4m scale, the occulter and coronagraph mission concepts are competitive in different science yield metrics. For total number of detections, the coronagraphs are the clear winners, with many times as many detections as either of occulter designs. These results may be slightly misleading, since they represent, for the most part, many repeated detections

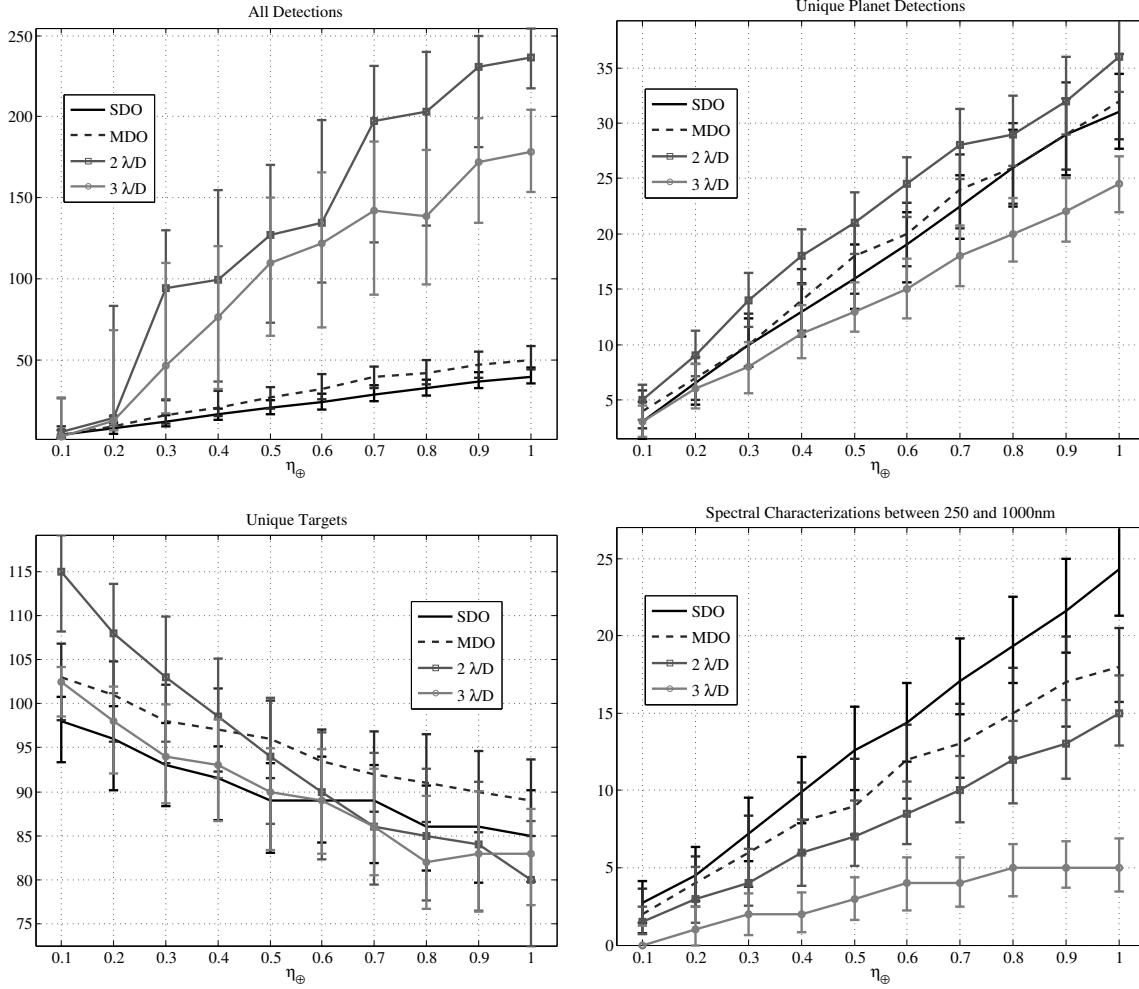


Figure 4.15: Simulation results for the SDO, MDO, and the 2 and 3 λ/D coronagraph mission concepts as functions of η_{\oplus} with a 4 m telescope for Earth-twin planetary populations.

of a small number (usually fewer than 5) of highly visible planets. However, this demonstrates the advantage of being able to perform many more observations over the course of the mission and, with improvements to the scheduling algorithm, should be translatable into a larger number of orbital fits for the coronagraphs. The two occulters average less than two detections per unique planet found, which means that they are generally unable to get more than one orbital fit in five years of mission time.

For unique planet detections, the two occulters perform almost identically, finding about 10% fewer planets than the 2 λ/D coronagraph and 25% more than the 3 λ/D corona-

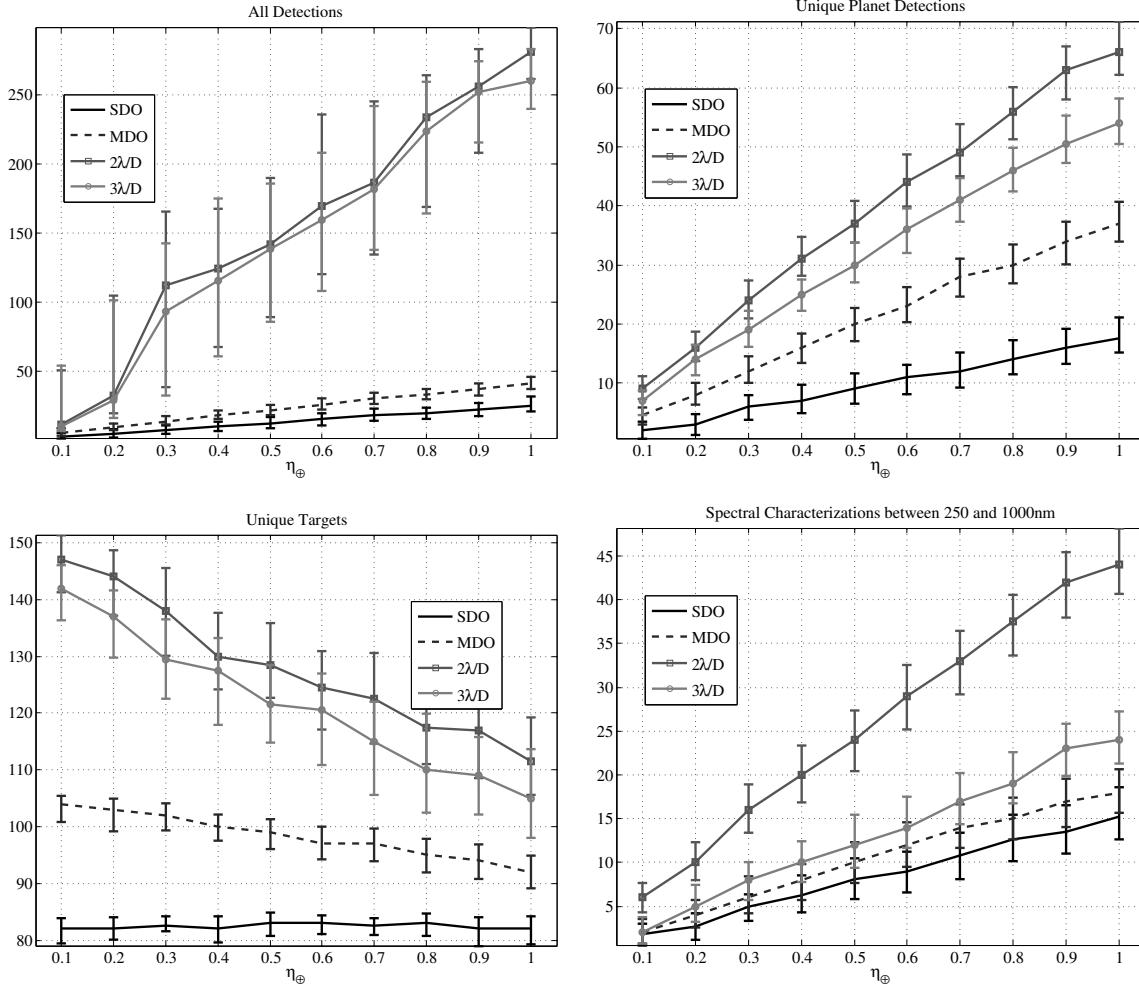


Figure 4.16: Simulation results for the SDO, MDO, and the 2 and 3 λ/D coronagraph mission concepts as functions of η_{\oplus} with an 8 m telescope for Earth-twin planetary populations.

graph at $\eta_{\oplus} = 1$. This result is almost purely a function of having a limited target list—in a universe with a uniform distribution of isotropically positioned stars, we would expect the coronagraphs’ larger total number of observations over the mission lifetime to produce proportionately more unique detections. As we are limited to existing stars, all of the instruments are able to find all of the easily observable planets within five years and the greater number of visits made by the coronagraphs translates into only a few more unique detections. At this telescope scale, all of the designs consistently find fewer than half of the available planets in a given universe.

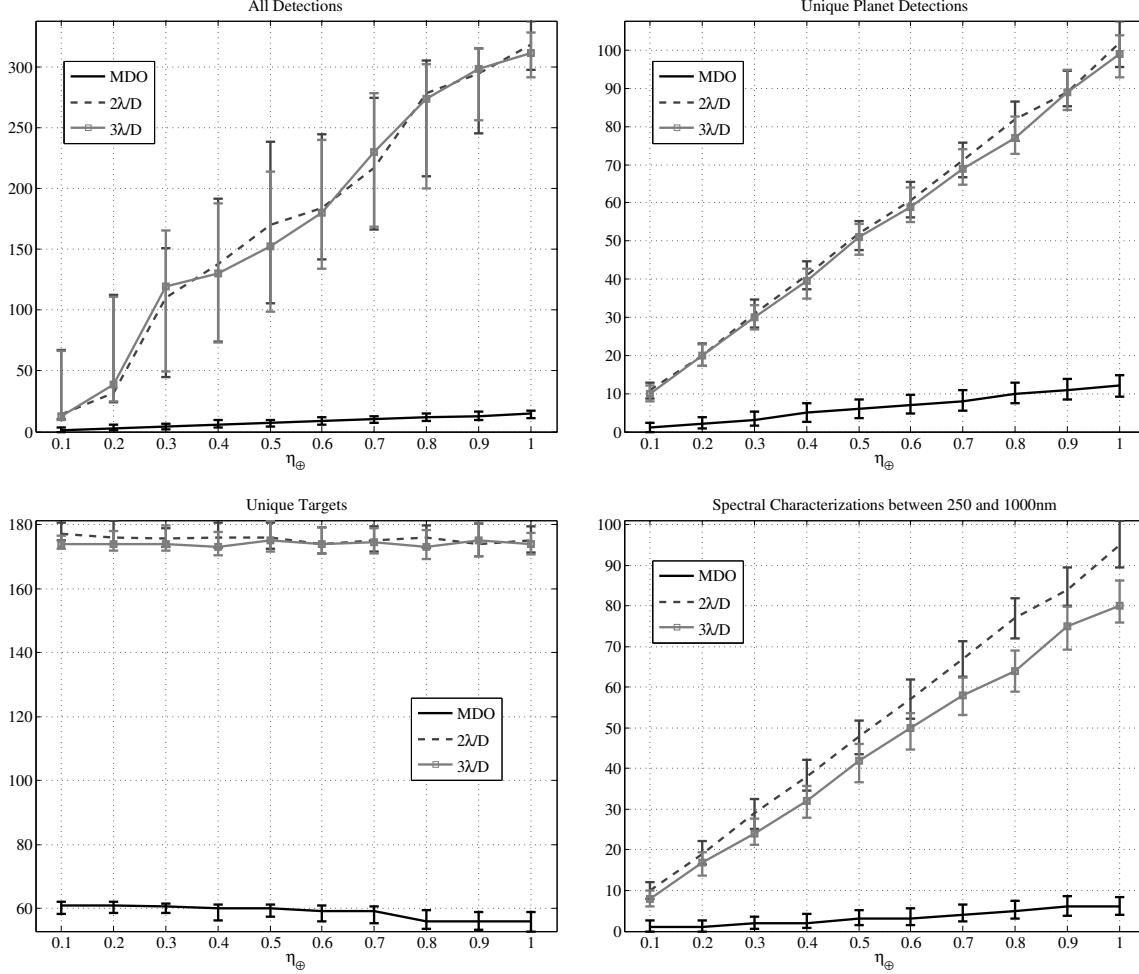


Figure 4.17: Simulation results for the MDO, and the 2 and 3 λ/D coronagraph mission concepts as functions of η_{\oplus} with a 16 m telescope for Earth-twin planetary populations.

The coronagraphs' large number of detections is balanced by their relatively poor performance in terms of spectral characterization. Because of their wavelength dependent IWA, and the fact that detections for Earth-like planets will occur at relatively low separations[8], in most cases the coronagraphs are capable of only a partial spectrum between 250 and 1000 nm. We see that the $2\lambda/D$ gets slightly fewer full spectra than the MDO (which has difficulties getting full spectra due to the extra required slew for the longer wavelength characterizations), but both are outperformed by the single distance occulter. The $3\lambda/D$ coronagraph is highly IWA limited, and gets many fewer full spectra than any of the other designs. From these metrics, we can conclude that for a 4 m telescope, coronagraphs and

occulters are both viable options, as long as a $2 \lambda/D$ IWA is achievable, and that the decision as to which instrument to build should be made based on other considerations, including the relative difficulties of wavefront control versus starshade manufacturing and positioning.

For the 8 m telescope, we start to see significant disparity between occulter and coronagraph performance, due completely to the smaller amount of observation time available to the occulters as more time is spent slewing between targets. Even so, the occulter designs still produce non-negligible science yield, with the MDO finding as many as 37 unique planets at $\eta_{\oplus} = 1$. The MDO is now outperforming the SDO, since the starshade required to cover the entire spectral band at one separation distance is much larger than the MDO. At this scale, coronagraphs clearly produce higher yields in all four science metrics. However, the problems of telescope stability and wavefront control become more difficult with larger telescopes. It may turn out that achieving even $3 \lambda/D$ is exceedingly difficult. If this is the case, then occulter are still a viable alternative, even if there are no improvements in propulsion technology beyond what is modeled here.

With the 16m telescope, the disparity between coronagraphs and occulters is so great that it becomes very difficult to argue for the benefits of an occulter, as modeled here. The extremely large starshade required even by the MDO makes transit slews take so long that less than 5% of a mission is typically spent on observation. This leads to multiple instances of zero detections at lower values of η_{\oplus} , meaning that this mission design could be a complete failure in terms of finding planets in this population. Of course, these results would be very different if we considered a system of multiple starshades operating with one telescope, as proposed in Hunyadi et al. [60]. With each additional starshade, the average transit slew time decreases, improving the science yield. In fact, one can design a constellation of starshades such that there is a guaranteed maximum retargeting time, in which case the occulter would operate much like a coronagraph system, with near constant overhead time per observation. Compared with an 8 or 16 m telescope, the cost of each starshade is relatively low, making multiple starshades a reasonable approach. Thus, we conclude that at current propulsion

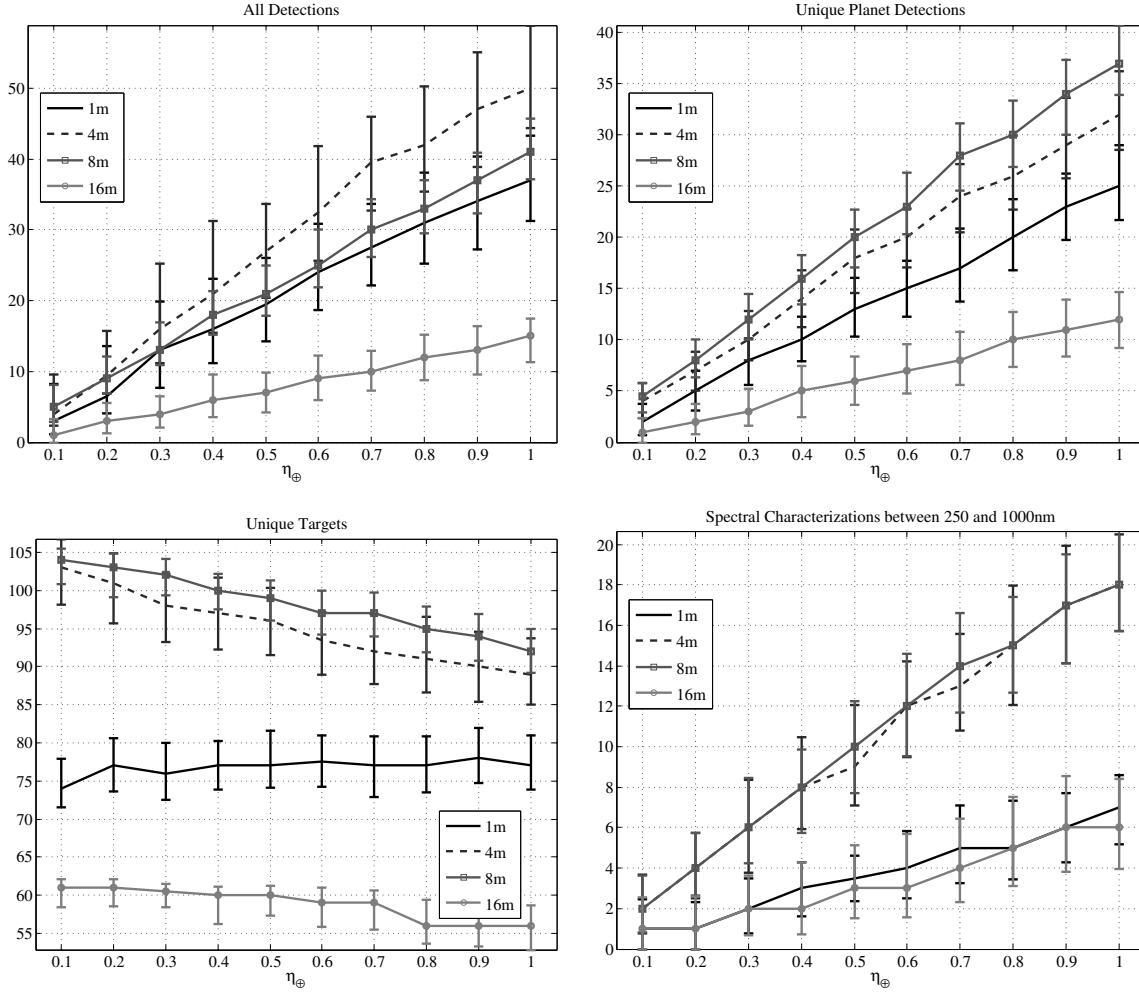


Figure 4.18: Simulation results for MDOs with 1, 4, 8, and 16 m telescopes for Earth-twin planetary populations.

technology levels, an occulter system with a single starshade is not a suitable starlight suppression system for telescopes greater than 8m in diameter. The coronagraphs continue to perform well at the larger scales, with the wavelength dependence of the IWA becoming much less of a problem for this specific planetary population

Figure 4.18 compares all of the MDO designs, including one for a 1m telescope. This 1 m telescope is different from O₃—it does not represent a developed mission concept, but rather what would happen if you were to use the THEIA occulter with a 1 m telescope and all of the same mission rules. It is not intended to represent a viable mission concept (unlike O₃)

but merely to test whether any useful science can be achieved with an occulter at aperture sizes where classical coronagraphs are essentially useless for terrestrial planet-finding. At 1 m, the IWA at even the telescope’s diffraction limit would completely cover the habitable zone for most target stars. Since the occulter’s IWA is determined solely by the starshade geometry, this problem does not exist for the occulter, although the smaller aperture size makes integration times much larger. The same size starshade is required for both the 1 and 4 m telescopes because the smaller telescope aperture spreads the point spread function, requiring a relatively larger level of suppression to produce the required contrast at the same IWA. The scheduling algorithm is able to produce a surprisingly large number of detections with the 1m MDO (about 20% fewer than the 4 m MDO) by visiting a smaller number of unique target stars during the mission. Of course, there is no way to get around the longer integration times required by the smaller telescope aperture, which means that the 1 m MDO gets relatively few spectra in the course of a mission, performing on par with the 16 m version. However, these results show that good science is possible with smaller telescopes, and led to the development of the O₃ concept.

Figure 4.19 compares all of the coronagraph designs, showing the progressive improvements in all metrics with telescope size, and the decreasing importance of the IWA wavelength dependence for this planetary population. Because both the MDOs split spectral characterizations into two sub-bands, and the largest observable wavelength for coronagraphs changes with distance for different targets, we can also track the number of spectra characterizations that cover less than the whole band. Figure 4.20 shows the total number of spectra to at least 700 nm collected by the various occulters and coronagraphs. The trends here are quite similar to those in the full spectra plots. As before, the 4 m occulter greatly outperforms its coronagraph counterpart.

We can also compare the total amount of time devoted to planet-finding observations by all of these instruments. Figure 4.21 demonstrates two clear trends: As the telescope aperture size increases, more photons are collected requiring less integration to achieve the same S/N,

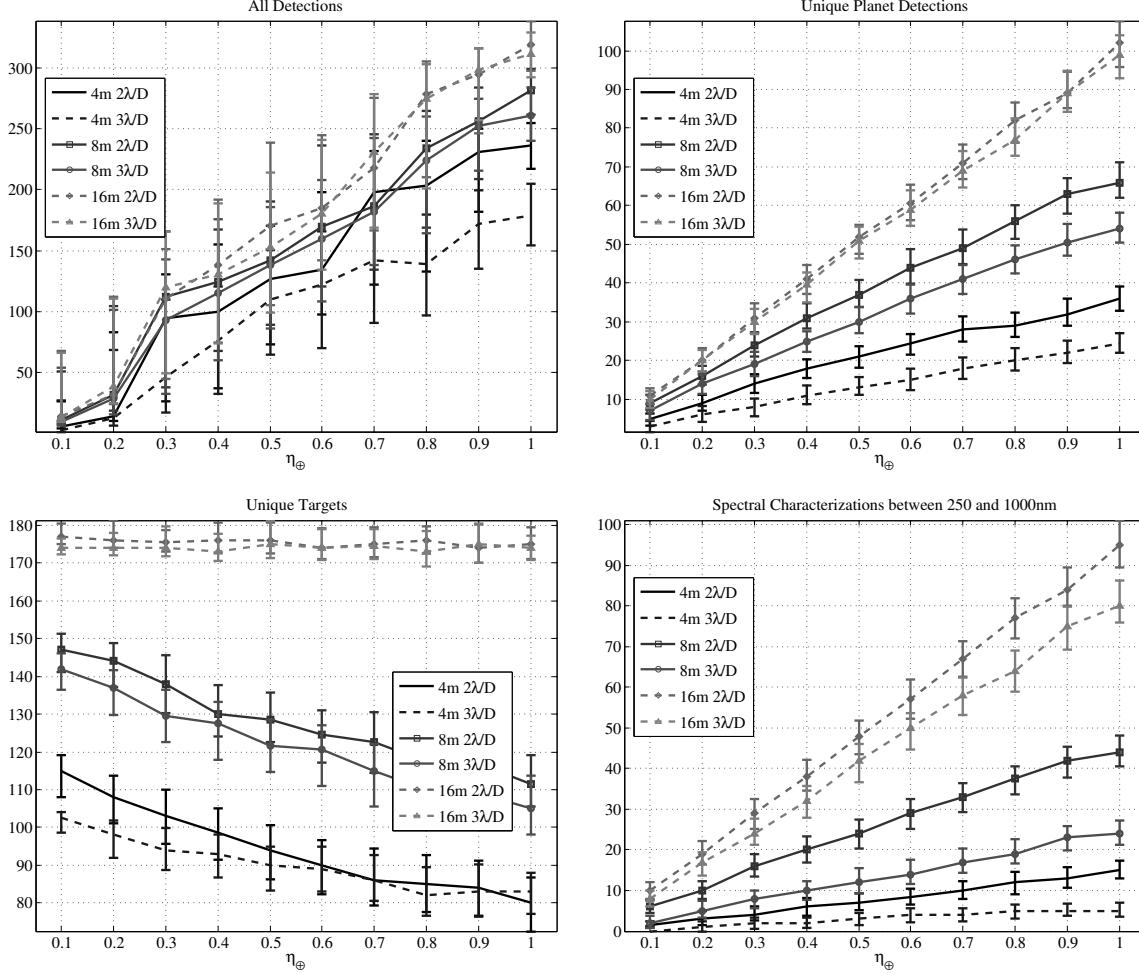


Figure 4.19: Simulation results for coronagraphs with 4, 8, and 16 m telescopes for Earth-twin planetary populations.

and so the total observation time generally decreases with aperture size. In addition, for occulters, as the telescope size grows, so does the size (and mass and separation distance) of the starshade, meaning that more time is needed to move the occulter between observations, so the total observation time is further decreased with increasing telescope aperture size. This is illustrated particularly well by the 1 m occulter, which uses the same starshade as the 4 m and has significantly larger total observation times. Another trend common to both instrument types is the increase in observation time with η_{\oplus} as higher occurrence rates of planets will lead to more spectral characterizations and longer total integration times. The one exception to this is the larger occulters, which actually have almost constant integration

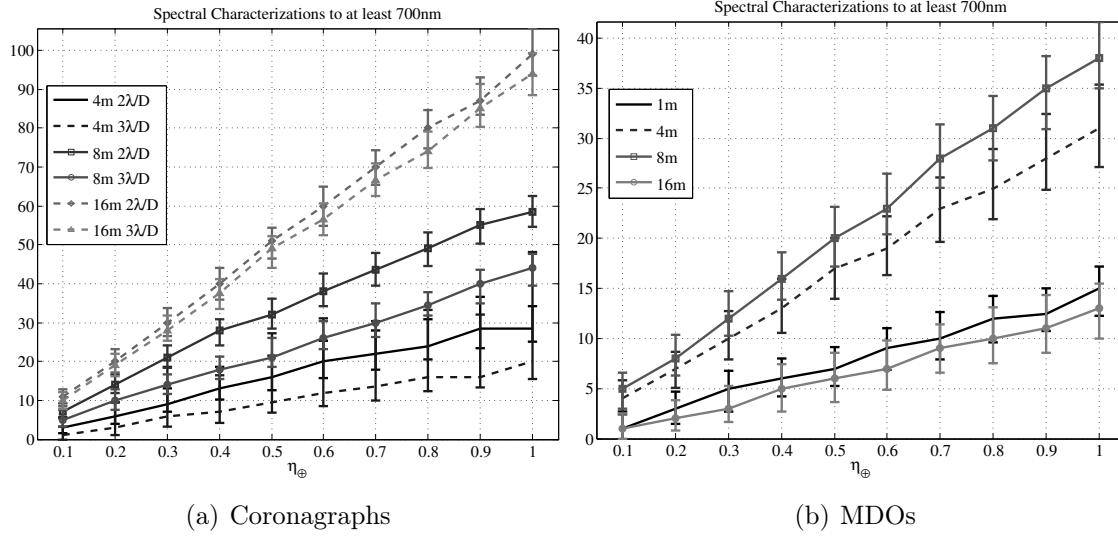


Figure 4.20: Total number of partial spectra (to at least 700 nm) collected for instruments at varying telescope scales.

times for all η_{\oplus} as so much of their mission time is devoted to slewing the occulter.

Similarly, we can compare the efficacy of the two instrument types at providing orbital fits (in this case, defined solely by the number of unique planets that are successfully detected more than 4 times). Figure 4.22 demonstrates that in this area, coronagraphs are clearly better than occulters. Due to the overall higher number of observations coronagraphs can achieve throughout a mission, they have much higher odds of re-detecting the same planet multiple times and can thus produce up to twice as many orbital fits as occulters.

Another point of interest is how quickly we will be able to tell whether an instrument is working as designed. Figure 4.23 shows histograms of first planet discoveries as a function of mission time for the 4 and 8 m telescope. For high values of η_{\oplus} , most of the instruments usually have their first detection in the first month of the mission, but as we see in the figure, at $\eta_{\oplus} = 0.3$, it may take many months before an initial detection. The two coronagraphs still have their first detections within the first month, but the SDO and MDO designs may take up to 6 months before having the same probability of a first detection. The results are essentially the same for the 4 and 8m telescope instruments with two interesting exceptions. The $2 \lambda/D$ coronagraph appears to actually have a lower probability of a first detection in the

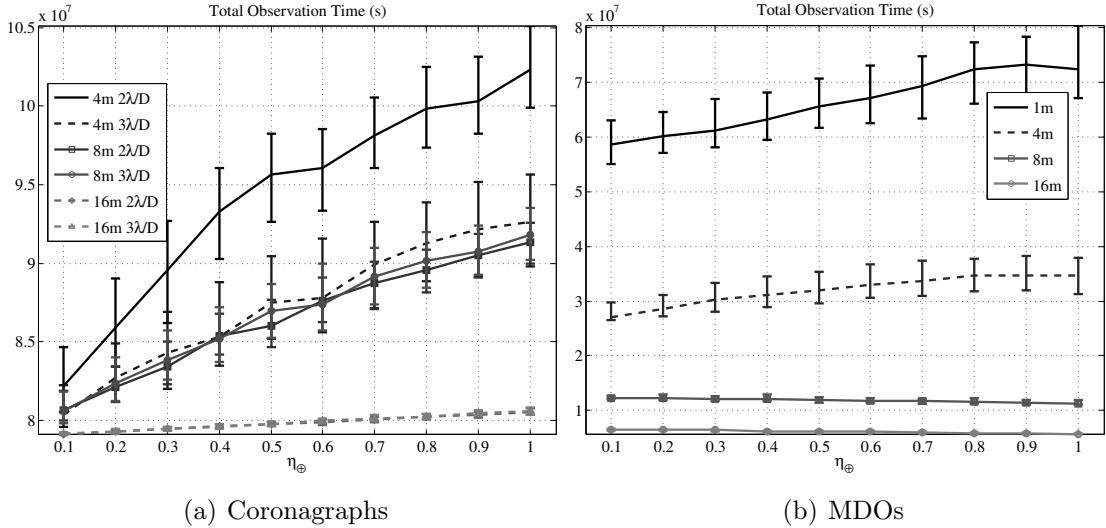


Figure 4.21: Total amount of time spent on planet-finding observations (detections and characterizations) over 5 years for instruments at varying telescope scales.

first month with an 8 m telescope than with a 4 m telescope. This is due to the expanded target list available to the 8 m telescope. Because there are now many more targets to consider, and we've assumed a constant rate of planet occurrence, it is unsurprising that it takes more visits, on average, before a detection is made. The second interesting feature of these histograms is that the SDO appears to have a higher probability of first detection in the first month than the MDO. Again, this is due to the available pool of targets and the number of visits made by each. At the 8 m scale, the MDO is able to make significantly more visits than the SDO. Since the SDO makes fewer total detections than the MDO, and visits to high completeness stars (which have highest probabilities of detections) occur early in the mission, the SDO at this scale either gets its first detection early on, or much later in the mission during revisits to these stars or at lower completeness targets. This is evidenced by the large percentage of first detections which occur more than 1 year into the mission for the 8 m SDO.

Besides varying telescope scales, we can also explore the performance of the same instruments on varying planetary populations. As a first step, we improve slightly on the overly simplistic Earth-twin only population by including habitable zone Super-Earths - rocky plan-

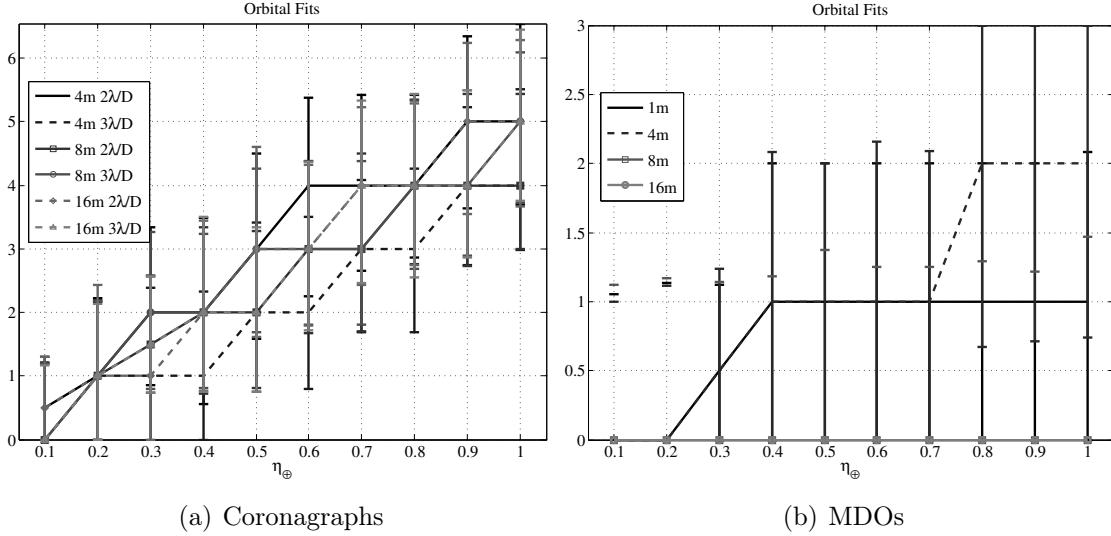


Figure 4.22: Total number of planets observed more than 4 times for instruments at varying telescope scales.

ets of 1 to $10 M_{\oplus}$ and radii determined by randomly generated ice-rock-iron proportions [37]. Unsurprisingly, the trends found with this planetary population (Figure 4.24) are, for the most part, the same as those seen in Figure 4.15 with an Earth-twin only population, but with larger values for most of the metrics, since Super-Earths are, on average, easier to detect than Earth-twins. With this population, in terms of unique planet detections, the two occulters and the $2 \lambda/D$ coronagraph are virtually indistinguishable, with the $3 \lambda/D$ coronagraph finding about 30% fewer unique planets. Due to the larger number of detections (which translates to longer observation times for characterization), a slightly smaller subset of the target list is typically visited than with the Earth-twin only population. For spectral characterizations, we find that the $3 \lambda/D$ coronagraph does much better than before, nearly matching the performance of the $2 \lambda/D$ coronagraph. Again, this is most likely due to the differing available targets for each, and suggests that the $2 \lambda/D$ coronagraph's performance in this metric could be improved with scheduling and target selection optimization.

Because O₃ does not perform spectral characterizations, but rather band photometry that requires much less integration time it is quite useful in illustrating the effects of certain parameters on the resulting science yield. As previously discussed, the selection of the target

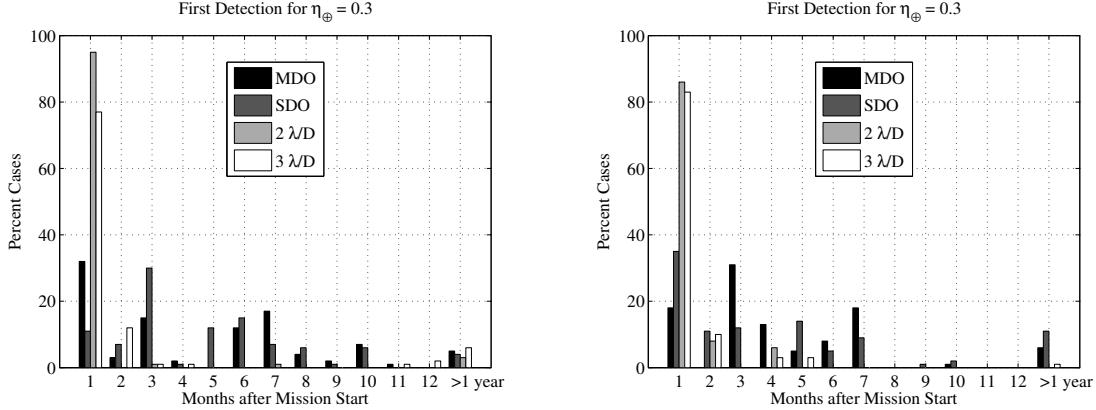


Figure 4.23: Histograms of first planet detections vs. months after mission start for the 4m (left) and 8m (right) telescope instrument designs.

pool is a major variable in the simulation. Fewer total targets generally increase the total number of detections, while reducing the number of unique detections. They also increase fuel use for slews, since targets are further apart, on average, and longer burns must be used to keep the mission time below the five year maximum. In simulations where mission rules only call for photometric characterization in the case of newly discovered planets, the amount of fuel used for stationkeeping is decreased due to the lower number of unique detections. Also, in cases where η_{\oplus} is low, the probability that no planets at all will be detected is increased by decreasing the target pool. These effects are shown for one of the O₃ configurations in Figure 4.25. Automatically generated target pools (based on minimum completeness) smaller than the ones shown here tend to perform worse in terms of both total and unique detections. However, it is possible to improve performance by considering not only the single visit completeness of the target systems, but also their placement. Adding strategically located systems with slightly lower completeness values can actually increase the overall expected number of detections as it leads to smaller average transition slews and a larger number of overall observations.

We can also use O₃ to analyze what happens when observations are allowed only when the planet is outside strict geometric IWA (90 mas, in this case) as opposed to allowing

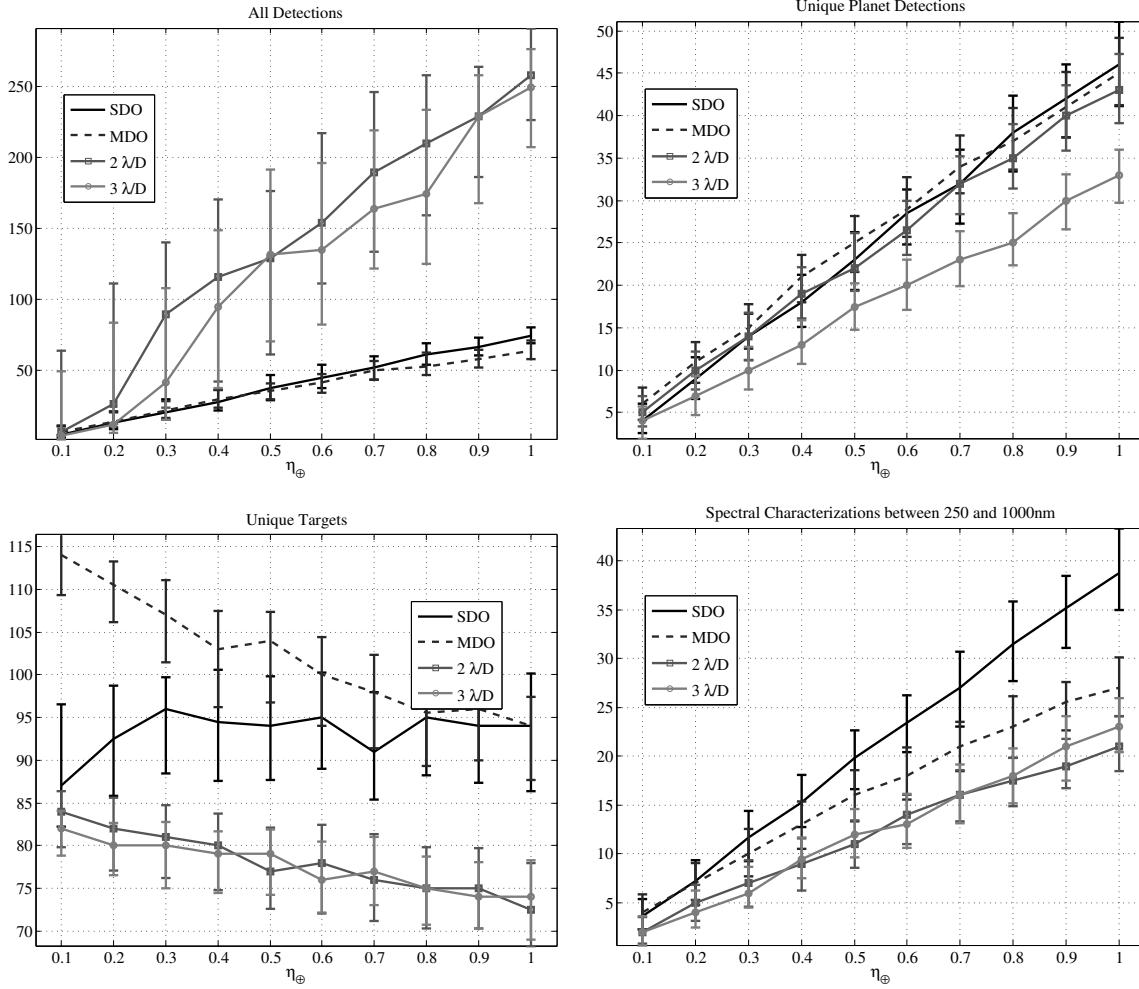


Figure 4.24: Simulation results for the SDO, MDO, and the 2 and 3 λ/D coronagraph mission concepts as functions of η_{\oplus} with a 4 m telescope for Super-Earth planetary populations..

observations between the petals of the occulter (down to 75 mas). We also observe the effects of performing photometry in only the blue band or in both bands. Figure 4.26 shows that, as expected, missions which attempt photometry in both bands will have fewer detections than mission characterizing only in the blue, due to the time required to change the telescope-occultor distance. Furthermore, missions which attempt to characterize planets through the occulter's petals do much better in getting red band photometry as the doubled inner working angle at 90 mas makes most HZ planets unobservable.

Figure 4.27 shows the effects of varying the default burn portion. Increasing this param-

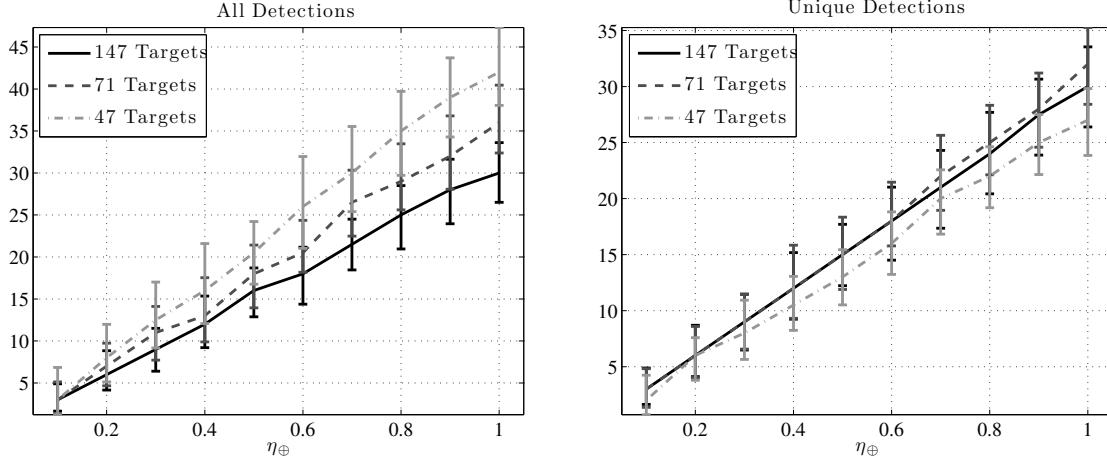


Figure 4.25: Simulation results for O₃ with different target pool sizes and Earth-twin systems. The three values correspond to minimum single visit completeness values of 0.1, 0.2 and 0.3. Mission rules for these simulations called for photometric characterization of newly discovered planets in the Blue band only.

eter has the predictable effect of decreasing the number of total detections, while greatly shortening the mission duration. While this would ordinarily be a reason to maintain a minimum default burn portion, a shorter mission lifetime with almost the same science return may actually be quite attractive when discussing smaller class missions. Both operating costs and the probability of single event failure decrease with the mission span, making shorter missions somewhat preferable, if we could expect an equivalent science return.

Figure 4.28 shows the average propellant use for the simulations of the 4 m SDO and THEIA mission concept. Both occulters were assumed to have the same initial mass (the payload capacity of an ATLAS V launch vehicle), which translates to a smaller total fuel capacity for the heavier SDO. As expected, the SDO, having the larger separation distance from the telescope, uses more fuel overall. It is also instructive to look at the breakdown of fuel use in terms of slewing and stationkeeping. Because characterizations for THEIA include an extra slew at full thrust, as η_{\oplus} increases so does the number of characterizations and the slew propellant usage. In the case of the single distance occulter, on the other hand, more time characterizing translates to less time slewing, and so its slew propellant use drops with

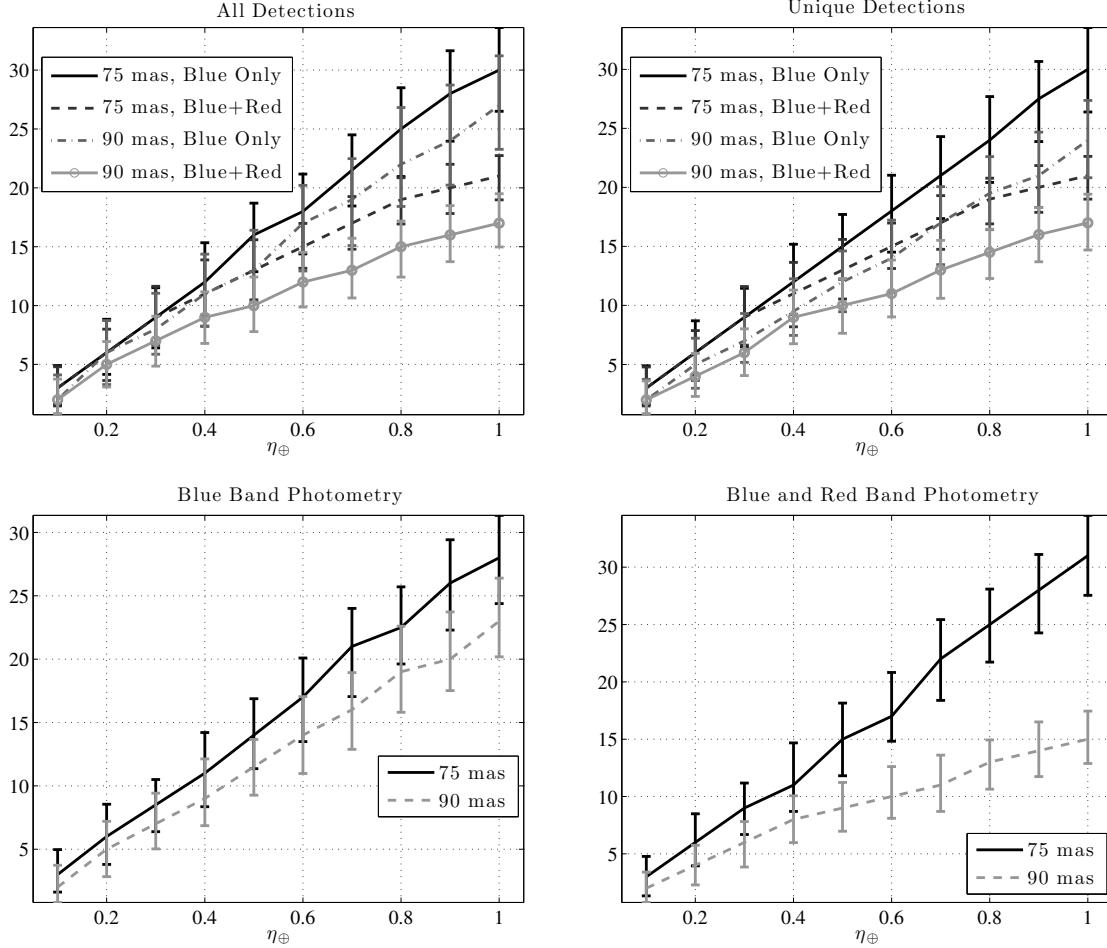


Figure 4.26: Simulation results for O₃ performing Blue or Red + Blue band photometry, and searching for planets down to angular separations of 90 or 75 mas. The much larger difference in the two curves in the final two plots is due to the extreme difficulty of getting complete photometry in the Red band where the geometric IWA is 180 mas for the 90 mas case.

increasing η_{\oplus} . A similar effect is seen in the stationkeeping propellant use—as η_{\oplus} increases, THEIA spends more and more time at the smaller separation distance, where stationkeeping fuel costs are lower, thereby slightly lowering overall stationkeeping fuel use. At the same time, the single distance occulter spends more time characterizing at its wider separation, increasing its stationkeeping fuel use. All of these effects, however, are relatively minor when compared with the difference in the two occulters' dry masses—the fuel remaining at the end of THEIA's primary mission is due almost completely to the fact that the occulter is

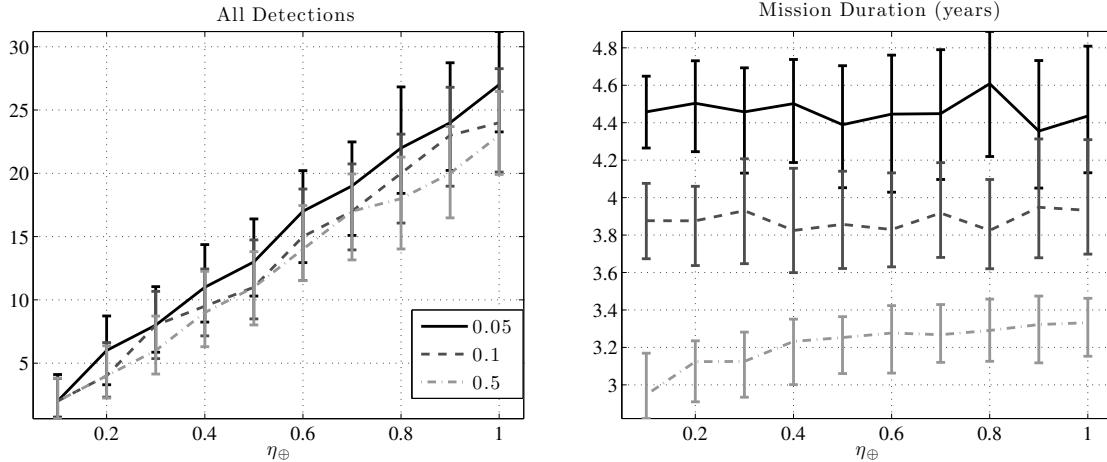


Figure 4.27: Simulation results for different default burn portions using the 2 band photometry, 75 mas IWA O₃ configuration.

smaller and lighter than its single distance counterpart.

We can take advantage of the extra fuel at THEIA’s disposal by adding an extended mission to make up for some of the differences in science yield between the two occulter designs. One option is to continue into the extended mission with the same mission rules as for the primary mission. However, because there is generally only about 2.5 years of fuel left, and the THEIA design is able to find all of the easily observable planets during the primary mission, this approach does not significantly increase overall science yield. Another alternative is to limit the target list in the extended mission to only those stars which had planets detected in the primary mission. This allows for many additional detections, and, in some cases, a few more complete spectra.

Figure 4.29 shows the results for the primary and primary + extended missions. Unfortunately, for this design, the extended mission does little to increase the number of complete spectra attained, mostly because the characterization times allow for full spectra only for planets that are observable for longer than the average observing season for this population. This means that if a full spectrum was achievable, it would generally have been acquired in the primary mission. However, the limited target pool with guaranteed observable planets allows us to significantly increase the total number of detections in a relatively short

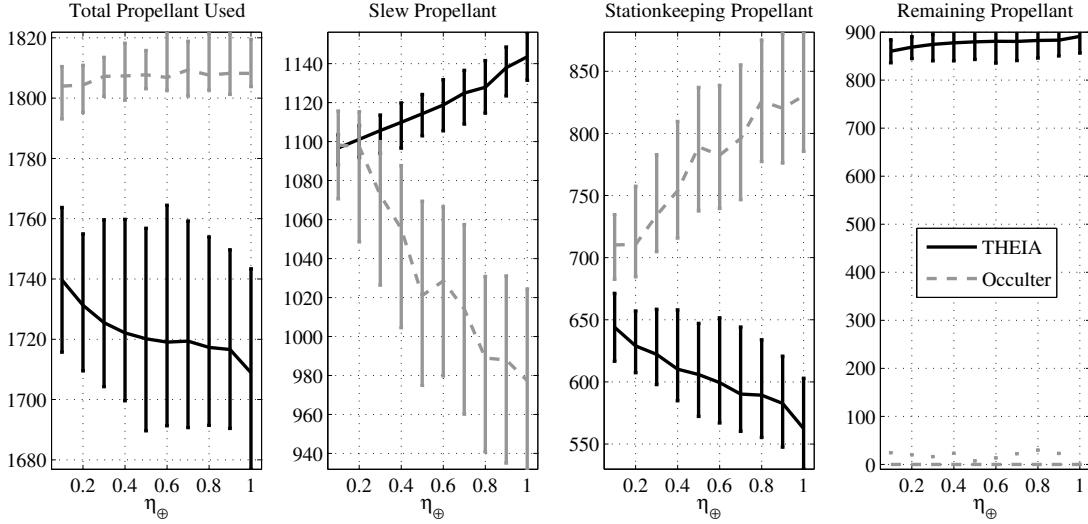


Figure 4.28: Occulter propellant use (in kg) for the 4 m SDO (gray) and THEIA (black). Although the assumed stationkeeping and slewing propulsion systems use different propellants, the simulation does not distinguish between the two fuel types and simply assumes that enough fuel is available, up to a maximum total fuel mass. When this mass is exceeded, the mission ends, regardless of mission time. The primary cause for the difference in the fuel remaining after 5 years is the difference in starshade dry masses.

time. Enough of these occur at varying points on the planets' orbits that this translates to a significant number of orbital fits for the total mission (up to 5, on average, for $\eta_{\oplus} = 1$). In this respect, THEIA is superior to the single distance occulter design. With the larger occulter, to get the same number of orbital fits, it would be necessary either to increase the fuel carried (thereby requiring a larger launch vehicle and increasing costs), or to partition the primary mission in this fashion, allocating the last one or two years to only revisiting systems with previous detections, which would sacrifice the total number of unique planets found.

Another area of interest is the effect of varying exo-zodi levels on mission science yield. Figure 4.30 shows the integration times required for THEIA to detect and spectrally characterize a 25 Δ mag Earth-twin about each of 117 target stars, with exo-zodi levels at the mean and 2σ levels of the historical zodi distribution for the solar system (see Figure 2.7). These plots illustrate quite well how the different approaches to detection and characterization (bayesian inference vs. integrating to a pre-determined signal-to-noise ratio) are affected

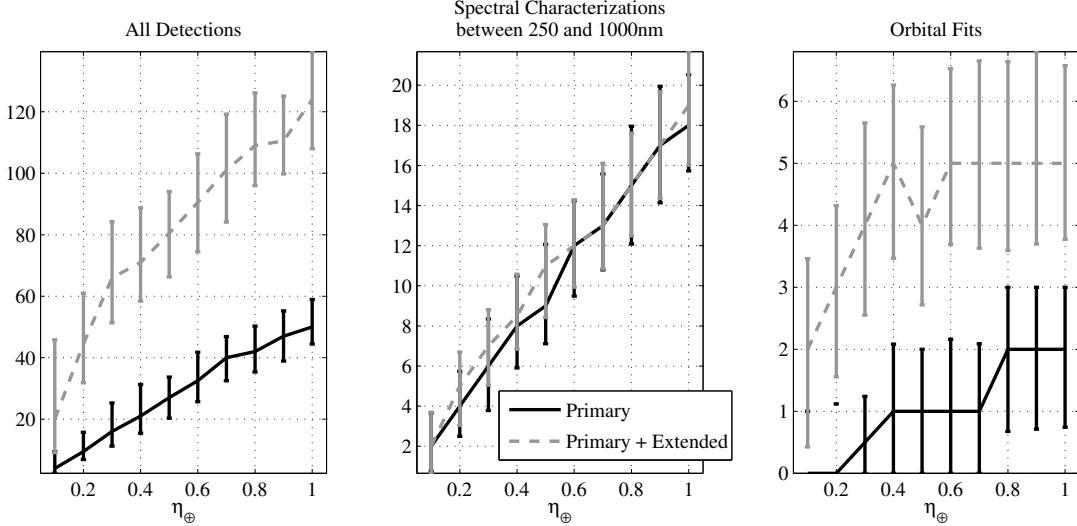


Figure 4.29: Simulation results for the primary THEIA mission shown and the primary + extended mission, where the occulter is allowed to use all of the available fuel.

by varying the exo-zodi magnitude with a fixed planet signal. Despite the proportionately larger increase in detection integration time for lower magnitude stars these all remain below 30 hours. On the other hand, the integration times required for characterization quickly stretch to many months, regardless of exo-zodi levels. From this, we can predict that while increased exo-zodi levels may decrease the number of detections, this effect probably won't be too dramatic unless the exo-zodi is present at much higher levels than found in the historical distribution of the solar system zodi over the last 80 Myr. At the same time, because the only detections that will lead to complete spectral characterizations will be of bright planets around stars with long observing seasons, the increased characterization integration time should similarly lead to a relatively small decrease in the number of full spectra attained throughout the mission.

We can test this hypothesis by running the mission simulations with varying exo-zodi levels. Doing so for 1.55 and 3.3 zodi produces a difference of 5%, on average, in the number unique planets found, and 3.5% in the number of total detections over the course of the mission. The number of spectra acquired also changes quite modestly, with a difference of 6%, on average, biased towards simulations with higher frequencies of planets, as shown in

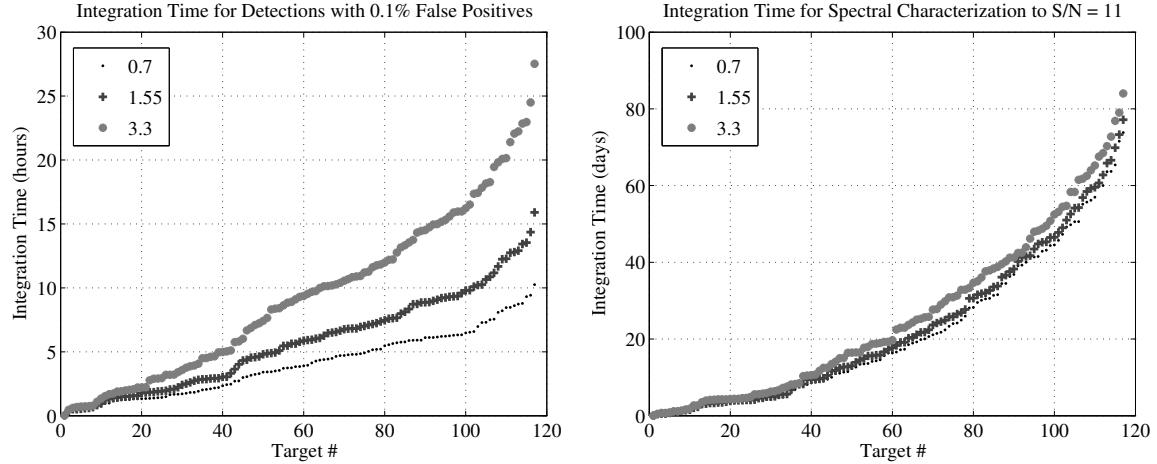


Figure 4.30: Integration times for planetary detection with 0.1% false positives, and spectral characterizations with $S/N = 11$ at the 760nm O₂ feature with resolving power $R = 70$. The planet is assumed to be 25 magnitudes fainter than the target star, and targets are ordered with decreasing magnitude. The legend numbers are the exo-zodi levels in zodi.

Figure 4.31.

The most significant differences in the two simulation ensembles occur in the amount of observation time and the number of targets visited. As shown in Figure 4.32, the amount of time spent on planet-finding/characterizing observations increases by two months, on average, between the 1.55 and 3.3 zodi simulations. While this represents only 3% of the allotted 5 years, we must keep in mind that for a design featuring a starshade, the majority of the mission time is spent repositioning the starshade between target stars. For THEIA, typically only 20 to 30% of the mission time is used planetary observations, making this a 10% increase. Because of this, the increase in observation time leads to a decrease in the number of targets that can be observed within the primary mission. Longer observation times for occulters also lead to longer station-keeping times versus transit slew times which can be a very important factor in the mission design. Since THEIA uses a separate propulsion system for station-keeping, which uses a different type of fuel from the slewing propulsion system, an incorrect assumption as to how much observation time is required could lead to major problems. Since the mission cannot proceed without adequate supplies of each

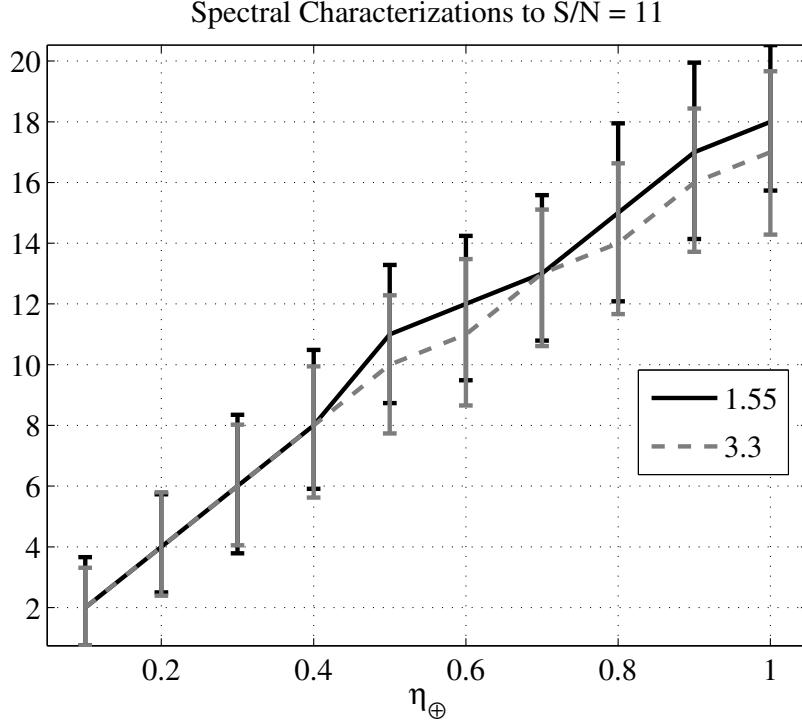


Figure 4.31: Simulation results of spectral characterizations with $S/N = 11$ at the 760nm O₂ feature with resolving power $R = 70$. The legend numbers are the exo-zodi levels in zodi.

propellant type, budgeting too much or too little observation time may lead to the mission ending early. The differences in observation times due to varying the assumed exo-zodi levels lead to differing requirements on the propellant carried by the starshade.

While it seems that as long as exo-zodi levels do not typically exceed those found in our own solar system's past, we should be able to achieve the primary goal of finding and characterizing planets, we will lose some of the science yield associated with observing a larger sample size if exo-zodi levels verge towards the high end of this distribution. At the same time, any large ambiguity in the expected exo-zodi levels will make mission design and planning much more difficult. If we consider exo-zodi levels outside the 2σ level of the solar system's historical distribution, as expected, all science yield metrics produce lower results. At 10 zodi concentrations, the mission simulations produce, on average, 19% fewer unique detections, 17% fewer total detections, and 26% fewer spectral characterizations (as compared with the 1.55 zodi simulations). More significantly, in simulations where Earth-like

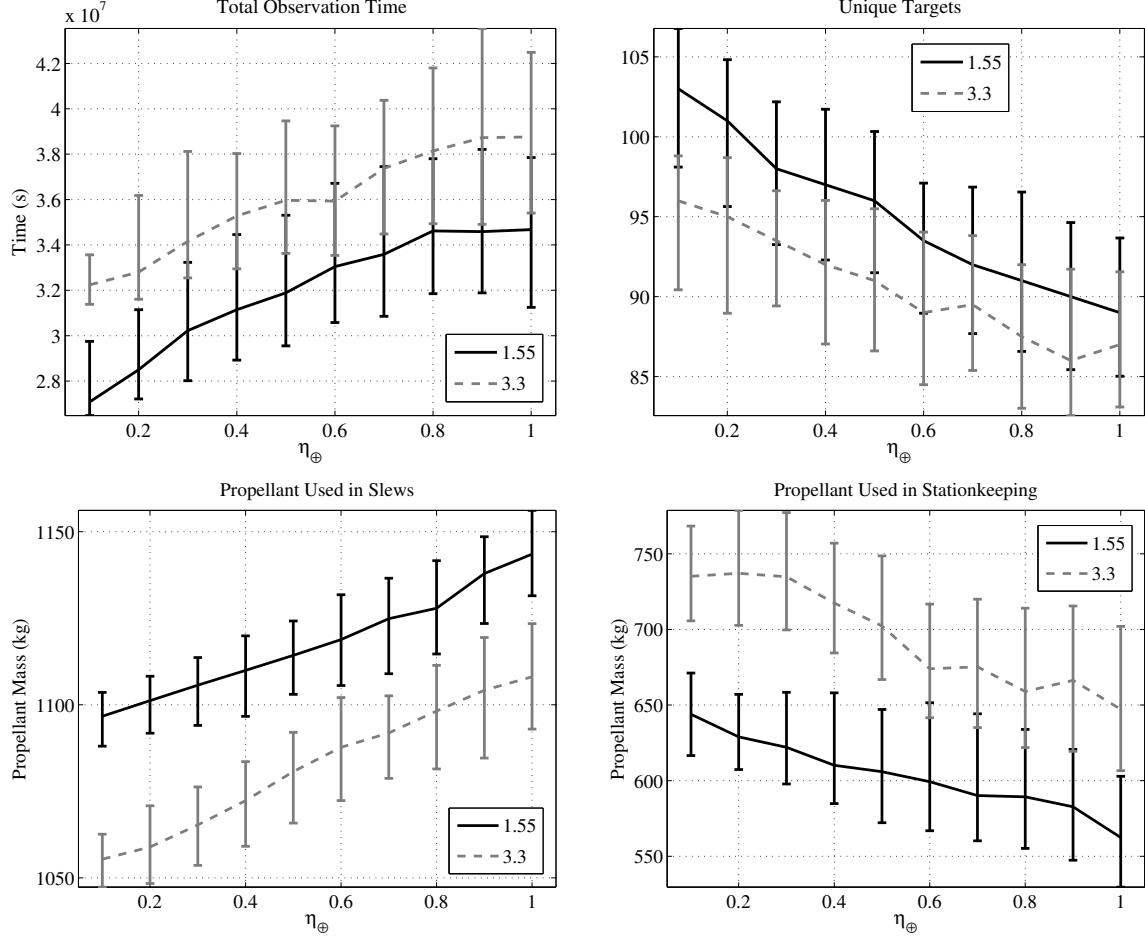


Figure 4.32: Total time for planet-finding/characterizing observations over the length of a mission, the number target stars observed at least once in the course of the mission, propellant used for transit slews, and propellant used for stationkeeping for THEIA with varying levels of exo-zodi.

planets are assumed to be rare ($\eta_{\oplus} < 0.2$), almost 18% of all mission simulations have zero spectral characterizations, which means that if high exo-zodi levels are common and planets are rare in our target pool, there is a real danger of being unable to meet a basic mission requirement.

In addition to the possibility of higher overall exo-zodi levels, we also know that dust disks may form concentrations (or ‘clumps’), especially in the presence of planetary bodies [131]. It is unreasonable to assume that our detection integration time algorithm can achieve arbitrarily low false positive rates regardless of any structure in the background, and it is

likely that there will be observations where C_b cannot be accurately estimated in the time given by the algorithm described above, or will vary significantly throughout the field of view. This will lead to an increase in the number of false positives, which will require additional integration time to resolve.

A simple approach to modeling these effects is to increase the number of false positives generated in our simulations, while leaving the same integration times for detections as before. This approximates the situation where a planet signal is found in the data due to mischaracterization of the background. In all previous simulations the false positive rate used to calculate detection integration times was also used to generate false positives in the simulations of observations. Now, we decouple these two values, leaving the same low rate for the detection time calculation, but increasing the rate at which false positives are generated in the simulation.

Since THEIA’s mission rules require spectral characterization after each new detection, it is assumed that this would resolve any false positives due to exo-zodiacal clumping, at the expense of the wasted integration time. With these simplifying assumptions, we have a system in which false positives generated by bright structures in the exo-zodiacal dust clouds do not directly compete with detections of real planets. Rather, the time spent spectrally characterizing these false alarms takes away from the time available to find real planets.

The effects of introducing an increased rate of false alarms are very similar to those of increasing the mean exo-zodi level. When 5% of observations which were formerly true negatives are made false positives, the number of total detections changes very little, while the number of unique planets found decreases by 3.4%, while the number of spectral characterizations decreases by about 7.4%. Setting the false positive rate at 10% decreases the number of true detections by 4%, and the number of unique planets and spectral characterizations by 8.7% and 16%, respectively (Figure 4.33). As before, these values reinforce the idea that a planet-finding mission, even for Earth-like planets, is not impossible without detailed prior knowledge of the exo-zodiacal dust distributions of our target stars, although any increase

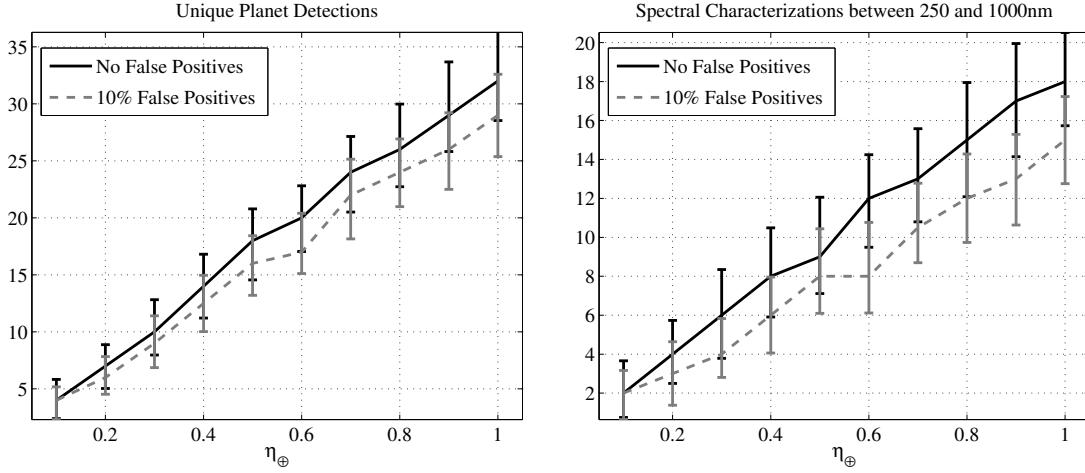


Figure 4.33: Comparison of simulations run with a false positive rate set sufficiently low to produce no false positives, and with a simulated 10% false positive rate.

in such knowledge will help greatly in mission design and planning.

The various results presented here demonstrate the various capabilities of the mission simulation framework described in this chapter. Most importantly, any new metric that one wishes to evaluate can be calculated from the data generated by these same simulations. It is not necessary to redo a simulation (unless one wishes to evaluate a new instrument, or a different planet population). Everything we need to know about a mission can be mined from the data generated in these ensembles.

The simulations developed in this chapter relied heavily upon the models of planetary systems and planet-finding instrumentation developed in the first three chapters. We can also apply the same models to the problem of data analysis. The next chapter will demonstrate how the modeling that allowed us to simulate planetary systems can also be used generate exosystem models from observed data.

Chapter 5

Integration of Data Streams

One rapidly emerging trend in the field of exoplanet studies is the application of multiple instruments and observatories to the study of a single target. After the initial discovery of planets about HR8799 in data collected at the Keck and Gemini observatories [88], evidence for these planets was also uncovered in archival Hubble Space Telescope (HST) data [76] and multiple other groups have made followup observations at a variety of other observatories [56, 63]. Planetary candidates derived from Kepler Observations are all followed up with ground-based spectroscopic and photometric studies before they can be confirmed as new planets [4, 5]. Similarly, proposed planet-finding instruments using astrometric techniques usually assume a pre-existing (or concurrently collected) radial velocity data set to help with orbit fitting [19]. This trend points to the importance of developing systematic ways of combining disparate data sets of the same target stars into unified exosystem models. This chapter will describe one such method, drawing heavily on the modeling work developed throughout this thesis.

5.1 Exosystems as Hidden Markov Models

Regardless of how we observe an exosystem, we are making measurements of an underlying system with deterministic dynamics. If we had sufficient information about all of the bodies

in the exosystem (e.g., equal to the amount of data we have about our own solar system), then we could predict the positions and photometric properties of all of the bodies in that system far into the future. Unfortunately, our observation methods give us only brief snapshots of small portions of exosystem dynamics, so that we have to reconstruct the full picture from multiple observations and modeling. It is this property of planet-finding that makes Hidden Markov Models (HMM), Dynamic Bayesian Networks (DBN) and dynamic filtering so appropriate to describing exoplanet observations.

We assume a dynamic system with state variable X_t whose evolution in time is a Markov process, i.e.:

$$\begin{aligned} P[X_k = x_k | X_{k-1} = x_{k-1}, X_{k-2} = x_{k-2}, X_{k-3} = x_{k-3}, \dots] \\ = P[X_k = x_k | X_{k-1} = x_{k-1}]. \end{aligned} \quad (5.1.1)$$

An HMM is generated when some part of the state is unobserved. If our knowledge of the system comes from observations:

$$Y_t = f(X_t) \quad (5.1.2)$$

the state is partially observed when function f discards some portion of X_t . For example, if $X_t \in \mathbb{R}^n$, then the state is partially observed when $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ for $m < n$. A generalization of this formalism yields dynamic Bayesian networks, where the system evolution is governed by temporal probability models, and leads us to the idea of dynamic filtering and state estimation [107].

In general, if we know how X_t evolves in time and how Y_t depends on X_t , via either probabilities or flows, we can use time histories of observations to approximate the time history (and current value) of the state. More specifically, for a Markov process with state **x** and observation **z**:

$$P[\mathbf{x}_0, \mathbf{x}_1 \dots \mathbf{x}_n, \mathbf{z}_1, \mathbf{z}_2 \dots \mathbf{z}_n] = P[\mathbf{x}_0] \prod_{j=1}^n P[\mathbf{z}_j | \mathbf{x}_j] P[\mathbf{x}_j | \mathbf{x}_{j-1}], \quad (5.1.3)$$

the next state can be predicted from the observed history as:

$$P[\mathbf{x}_j | \mathbf{z}_{1:j-1}] = \int_{\mathbf{x}_{j-1}} P[\mathbf{x}_j | \mathbf{x}_{j-1}] P[\mathbf{x}_{j-1} | \mathbf{z}_{1:j-1}] d\mathbf{x}_{j-1}^*, \quad (5.1.4)$$

and updated with the current observation via Bayes rule:

$$P[\mathbf{x}_j | \mathbf{z}_{1:j}] = \frac{P[\mathbf{z}_j | \mathbf{x}_j] P[\mathbf{x}_j | \mathbf{z}_{1:j-1}]}{P[\mathbf{z}_j | \mathbf{z}_{1:j-1}]} . \quad (5.1.5)$$

Alternatively, we can write the state update and observation as functions with additive noise terms:

$$\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, k) + \mathbf{w}_k \quad (5.1.6)$$

$$\mathbf{z}_k = \mathbf{h}(\mathbf{x}_k, k) + \mathbf{n}_k \quad (5.1.7)$$

where \mathbf{w}_k and \mathbf{n}_k are the process noise and measurement noise, respectively. Note that while Equations (5.1.6) and (5.1.7) represent discrete time systems, an equivalent formalism exists for continuous time systems of the form:

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{f}(\mathbf{x}(t), t) + \mathbf{w}(t) \quad (5.1.8)$$

$$\mathbf{z}(t) = \mathbf{h}(\mathbf{x}(t), t) + \mathbf{n}(t) . \quad (5.1.9)$$

In this formulation, we can define a state estimate, $\hat{\mathbf{x}}$, that is optimal in some sense—for example minimizing mean-square error:

$$\hat{\mathbf{x}}_k = \arg \min E\{\|\mathbf{x}_k - \hat{\mathbf{x}}_k\|^2\} , \quad (5.1.10)$$

in which case the state estimate for a time history of N observations corresponds to the

*The subscript $1:j-1$ represents the joint distribution $\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, \dots, \mathbf{z}_{j-1}$

minimization (with respect to \mathbf{x}) of the quadratic cost function:

$$J = \sum_{k=1}^N [\mathbf{z}_k - \mathbf{f}(\mathbf{x}_k, k)]^T R^{-1} [\mathbf{z}_k - \mathbf{f}(\mathbf{x}_k, k)] \quad (5.1.11)$$

where R is a weighting matrix often taken to be the observation noise covariance. In many cases Equation (5.1.11) represents a non-convex optimization problem and is more tractable when formulated as a recursive filter [24].

As in Equations (5.1.4) and (5.1.5) the recursive filter predicts the next state based on the current state and the system dynamics (\mathbf{f}), and then updates this prediction by incorporating the next measurement and taking into account the spectral density of the measurement noise. The most well known of these filters is the Kalman filter, which assumes zero-mean, white, Gaussian process and measurement noise, and can be shown to be optimal (in the minimum least-squares error sense) for linear state and observation equations; i.e., where \mathbf{f} and \mathbf{h} can be written as:

$$\mathbf{f}(\mathbf{x}_k, k) = \mathbf{F}_k \mathbf{x}_k \quad (5.1.12)$$

$$\mathbf{h}(\mathbf{x}_k, k) = \mathbf{H}_k \mathbf{x}_k . \quad (5.1.13)$$

The Kalman filter has been successfully applied to a wide range of problems including navigation and tracking systems, computer vision, and radio and telecommunications systems [132, 24]. There is also an extensive body of literature on using Kalman filters for orbit and attitude estimation (see, for example Spilker [130]).

The Kalman filter can be thought of as estimating the change in the distribution function of the state by propagating its first two central moments: the mean and variance. This operation is greatly simplified by noting that if the current distribution of a state is Gaussian then the predicted state distribution using a linear Gaussian transition model $P[\mathbf{x}_j | \mathbf{x}_{j-1}]$ (as in Equation (5.1.4)) will also be a Gaussian distribution. Similarly, if the observation model $P(\mathbf{z}_j | \mathbf{x}_j)$ is also linear Gaussian then the updated state distribution will be Gaussian as well.

This makes it relatively simple to propagate linear Gaussian state distributions in time, as long as the initial conditions correspond to the true distribution.

For a linear state update (as in Equation (5.1.12)) the predicted state estimate $\hat{\mathbf{x}}_k$ given the previous estimate $\hat{\mathbf{x}}_{k-1}$ is:

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{F}_k \hat{\mathbf{x}}_{k-1|k-1} .^{\dagger} \quad (5.1.14)$$

From this, we find that the state covariance, defined as:

$$\mathbf{P}_k = E[(\mathbf{x}_k - E[\mathbf{x}_k])(\mathbf{x}_k - E[\mathbf{x}_k])^T], \quad (5.1.15)$$

is predicted as:

$$\mathbf{P}_{k|k-1} = \mathbf{F}_k \hat{\mathbf{x}}_{k-1|k-1} \mathbf{F}_k^T + \mathbf{Q}_k, \quad (5.1.16)$$

where \mathbf{Q}_k is the covariance of the process noise:

$$\mathbf{w}_k \sim N(0, \mathbf{Q}_k). \quad (5.1.17)$$

If the initial state estimate and covariance accurately reflect the state distribution:

$$\begin{aligned} \hat{\mathbf{x}}_0 &= E[\mathbf{x}_0] \\ \mathbf{P}_0 &= E[(\mathbf{x}_0 - \hat{\mathbf{x}}_0)(\mathbf{x}_0 - \hat{\mathbf{x}}_0)^T] \end{aligned} \quad (5.1.18)$$

then all subsequent state estimates should have zero mean error:

$$E[\mathbf{x}_k - \hat{\mathbf{x}}_k] = 0. \quad (5.1.19)$$

[†]The notation used here represents the predicted (*a priori*) values via the subscript $k|k-1$ and the updated (*a posteriori*) values via the subscript $k|k$. An alternate convention (c.f., Stengel [132]) is to label predicted values with a minus sign ($\hat{\mathbf{x}}_k(-)$ or $\hat{\mathbf{x}}_k^-$) and updated values with a plus sign ($\hat{\mathbf{x}}_k(+)$ or $\hat{\mathbf{x}}_k^+$).

This is achieved by updating the predicted state with the current measurement as:

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k (\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}) , \quad (5.1.20)$$

where \mathbf{K}_k is the optimal Kalman gain that minimizes the updated estimate mean square error, $E[(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k})^T(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k})]$, and has the form:

$$\mathbf{K}_k = \mathbf{P}_{k|k-1} \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \mathbf{R}_k)^{-1} , \quad (5.1.21)$$

where \mathbf{R}_k is the covariance of the observation noise:

$$\mathbf{n}_k \sim N(0, \mathbf{R}_k) . \quad (5.1.22)$$

Equation (5.1.21) can be derived in a variety of ways, the simplest of which is to write the cost function in Equation (5.1.11) for two subsequent measurements, substituting $\mathbf{H}_k \hat{\mathbf{x}}_k$ for $\mathbf{f}(\mathbf{x}_k, k)$, and to then minimize the resulting function. This then allows us to write the updated covariance as:

$$\mathbf{P}_{k|k} = \left(\mathbf{P}_{k|k-1}^{-1} + \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k \right)^{-1} . \quad (5.1.23)$$

Alternatively, we can take advantage of the fact that minimizing the mean-square error is equivalent to minimizing the trace of the updated covariance, which leads to the equivalent Kalman gain expression, but requires us to first define $\mathbf{P}_{k|k}$. This is done by noting that the updated covariance must equal the covariance of the error:

$$\mathbf{P}_{k|k} = E \left[((\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}) - E[(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k})]) ((\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}) - E[(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k})])^T \right] \quad (5.1.24)$$

$$= (I - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_{k|k-1} (I - \mathbf{K}_k \mathbf{H}_k)^T + \mathbf{K}_k \mathbf{R}_k \mathbf{K}_k^T , \quad (5.1.25)$$

where I is the identity matrix. When \mathbf{K}_k has the form in Equation (5.1.21), this further

simplifies to:

$$\mathbf{P}_{k|k} = (I - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_{k|k-1}, \quad (5.1.26)$$

which can be shown to be equivalent to Equation (5.1.23) via the matrix inversion lemma [132].

The Kalman-Bucy filter extends this formalism to continuous time systems, as in Equations (5.1.8) and (5.1.9), where again \mathbf{f} and \mathbf{h} are linear in \mathbf{x} (but possibly time varying):

$$\mathbf{f}(\mathbf{x}(t), t) = \mathbf{F}(t)\mathbf{x}(t) \quad (5.1.27)$$

$$\mathbf{h}(\mathbf{x}(t), t) = \mathbf{H}(t)\mathbf{x}(t). \quad (5.1.28)$$

The resulting filter is defined via the differential equations:

$$\frac{d\hat{\mathbf{x}}(t)}{dt} = \mathbf{F}(t)\hat{\mathbf{x}}(t) + \mathbf{K}(t)(\mathbf{z}(t) - \mathbf{H}(t)\hat{\mathbf{x}}(t)) \quad (5.1.29)$$

$$\frac{d\mathbf{P}(t)}{dt} = \mathbf{F}(t)\mathbf{P}(t) + \mathbf{P}(t)\mathbf{F}^T(t) + \mathbf{Q}(t) - \mathbf{K}(t)\mathbf{R}(t)\mathbf{K}^T(t) \quad (5.1.30)$$

$$\mathbf{K}(t) = \mathbf{P}(t)\mathbf{H}^T(t)\mathbf{R}^{-1}(t). \quad (5.1.31)$$

Most physical system, including those studied here, are not actually observed continuously, so that the fully continuous Kalman-Bucy equations are not particularly helpful. However, it is possible to write a hybrid filter for the system:

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{F}(t)\mathbf{x}(t) + \mathbf{w}(t) \quad (5.1.32)$$

$$\mathbf{z}_k = \mathbf{H}_k \mathbf{x}(t_k) + \mathbf{v}_k, \quad (5.1.33)$$

where t_k is an element of the set of observation times $\{t_i\}$. The resulting hybrid equations

are then:

$$\hat{\mathbf{x}}_{k|k-1} = \int_{t_{k-1}}^{t_k} \mathbf{F}(t) \hat{\mathbf{x}}(t) dt \quad (5.1.34)$$

$$\mathbf{P}_{k|k-1} = \int_{t_{k-1}}^{t_k} \mathbf{F}(t) \mathbf{P}(t) + \mathbf{P}(t) \mathbf{F}^T(t) + \mathbf{Q}(t) dt \quad (5.1.35)$$

$$\mathbf{K}_k = \mathbf{P}_{k|k-1} \mathbf{H}_k^T \left(\mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \mathbf{R}_k \right)^{-1} \quad (5.1.36)$$

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k (\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}) \quad (5.1.37)$$

$$\mathbf{P}_{k|k} = \left(\mathbf{P}_{k|k-1}^{-1} + \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k \right)^{-1}, \quad (5.1.38)$$

with the state and covariance initialized as in Equation (5.1.18) and $\hat{\mathbf{x}}_{k-1|k-1}$ and $\mathbf{P}_{k-1|k-1}$ used as the initial conditions for the integrations in Equations (5.1.34) and (5.1.35), respectively.

When the state and measurement equations are nonlinear, the Kalman formalism can still be applied to the linearized system, yielding the Extended Kalman Filter (EKF). Unlike the original Kalman filter, the extended filter is no longer guaranteed to be an optimal estimator, and can produce highly divergent estimates in cases where the filter pushes the state outside the neighborhood of the linearization [24]. Nevertheless, the EKF is widely applied and can produce good results with comparatively small computational costs. The equations for the hybrid EKF are the same as (5.1.34) - (5.1.38), except that \mathbf{F} and \mathbf{H} are now the linearizations of \mathbf{f} and \mathbf{h} :

$$\mathbf{F}(t) = \frac{\partial \mathbf{f}(\mathbf{x}(t), t)}{\partial \mathbf{x}} \Bigg|_{\hat{\mathbf{x}}(t)} \quad (5.1.39)$$

$$\mathbf{H}_k = \frac{\partial \mathbf{h}(\mathbf{x}(t), t)}{\partial \mathbf{x}} \Bigg|_{\hat{\mathbf{x}}_{k|k-1}} \quad (5.1.40)$$

and the state update equation uses the nonlinear measurement:

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k (\mathbf{z}_k - \mathbf{h}(\hat{\mathbf{x}}_{k|k-1}, t_k)) . \quad (5.1.41)$$

A sample implementation of the hybrid EKF is presented in Listing A.7.

Because the linearization of the system dynamics makes $\mathbf{F}(t)$ a function of $\hat{\mathbf{x}}(t)$, the differential equations (5.1.34) and (5.1.35) must be integrated simultaneously as a single system:

$$\begin{bmatrix} \dot{\hat{\mathbf{x}}}_{k|k-1} \\ \mathbf{P}_{k|k-1} \end{bmatrix} = \int_{t_{k-1}}^{t_k} \begin{bmatrix} \mathbf{F}(t)\hat{\mathbf{x}}(t) \\ \mathbf{F}(t)\mathbf{P}(t) + \mathbf{P}(t)\mathbf{F}^T(t) + \mathbf{Q}(t) \end{bmatrix} dt. \quad (5.1.42)$$

To deal with the nonlinearities introduced by nonlinear observation equations, it is common to iterate the state update equations as:

$$\mathbf{H}_{k_i} = \left. \frac{\partial \mathbf{h}(\mathbf{x}(t), t)}{\partial \mathbf{x}} \right|_{\hat{\mathbf{x}}_{k|k_i}} \quad (5.1.43)$$

$$\mathbf{K}_{k_i} = \mathbf{P}_{k|k-1} \mathbf{H}_{k_i}^T (\mathbf{H}_{k_i} \mathbf{P}_{k|k-1} \mathbf{H}_{k_i}^T + \mathbf{R}_k)^{-1} \quad (5.1.44)$$

$$\hat{\mathbf{x}}_{k|k_i} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_{k_i} (\mathbf{z}_k - \mathbf{h}(\hat{\mathbf{x}}_{k|k_{i-1}}, t_k) - \mathbf{H}_{k_i} (\hat{\mathbf{x}}_{k|k-1} - \hat{\mathbf{x}}_{k|k_{i-1}})) \quad (5.1.45)$$

where the iteration is initialized with $\hat{\mathbf{x}}_{k|k_0} = \hat{\mathbf{x}}_{k|k-1}$ and run until

$$\max |\hat{\mathbf{x}}_{k|k_f} - \hat{\mathbf{x}}_{k|k_{f-1}}| \leq \epsilon \quad (5.1.46)$$

for desired convergence level (ϵ). The covariance update is then calculated from $\hat{\mathbf{x}}_{k|k_f}$ using Equation (5.1.38) with \mathbf{H}_{k_f} .

Another modification that can be made is the introduction of constraints to the state. If the chosen state for a nonlinear system can describe physically impossible (or undesirable) configurations, then it makes sense to constrain the state at each filter iteration. For example, if we use position and velocity to describe planetary orbits, our state can describe open (hyperbolic) orbits just as easily as closed ones. Since it is highly unlikely that we will find any exoplanet on an escape trajectory from its parent star, we may wish to constrain the Keplerian orbital energy of the orbits described by our state vector to be negative or zero at each time step.

Following Simon and Simon [122], we introduce a set of inequality constraints of the form:

$$D\mathbf{x} \leq \mathbf{d} \quad (5.1.47)$$

where D is a known $m \times n$ matrix (for m constraints of n state variables, such that $m \leq n$), and \mathbf{d} is the vector of constraints. The filter state can then be constrained by introducing a quadratic programming problem into the filter iteration of the form:

$$\min_{\tilde{\mathbf{x}}} (\tilde{\mathbf{x}}^T W \tilde{\mathbf{x}} - 2\hat{\mathbf{x}}_{k|k}^T W \tilde{\mathbf{x}}) \quad \text{such that} \quad D\tilde{\mathbf{x}} \leq \mathbf{d} \quad (5.1.48)$$

where $\tilde{\mathbf{x}}$ is the constrained state estimate and W is a symmetric positive definite matrix. The solution to this problem is equivalent to the solution of the equivalent equality constraint problem [121], and is given by:

$$\tilde{\mathbf{x}} = \hat{\mathbf{x}}_{k|k} - W^{-1} D^T (DW^{-1} D^T)^{-1} (D\hat{\mathbf{x}}_{k|k} - \mathbf{d}) . \quad (5.1.49)$$

Selecting $W = \mathbf{P}_{k|k}^{-1}$ has the effect of maximizing the conditional probability $P(\tilde{\mathbf{x}}|\mathbf{z}_k)$. Alternatively, taking W to be the identity matrix is equivalent to minimizing the conditional mean square error:

$$E [\|\mathbf{x}_{k|k} - \tilde{\mathbf{x}}\|^2 | \mathbf{z}_k] . \quad (5.1.50)$$

In either case, using Equation (5.1.49) we can write the expected value of the constrained estimate error as:

$$E [\mathbf{x} - \tilde{\mathbf{x}}] = \left(I - W^{-1} D^T (DW^{-1} D^T)^{-1} D \right) E [\mathbf{x} - \hat{\mathbf{x}}_{k|k}] , \quad (5.1.51)$$

and since the expected value of the error associated with the Kalman filter state estimate is zero, so is the expected value of the constrained estimate error. Thus, the constrained state estimate is an unbiased estimator for the system, just like the original unconstrained state

estimate [122].

Because all of the equations used to formulate the filter are time reversible, every Kalman filter can also be used as a smoother, i.e., to find the most probable values of a state at some time t where $t_0 < t < t_f$ while using the whole data set from t_0 to t_f . The only exception to this generalization is if the numerical integrator used in the EKF is not time reversible, which holds true for a class of symplectic integrators [149], but is not an issue for the methods used here. It is common to perform multiple smoothing and filtering steps on noisy data although for nonlinear systems, there is no guarantee that the state estimates will ever converge to some final values.

There are also many other variations, refinements and extensions of the Kalman filter. One family of filters, called Unscented Kalman Filters (UKF), first introduced in Julier et al. [65], seeks to improve upon the EKF by taking into account the second order terms in the expansion of the system dynamics, thereby increasing the probability that the propagated state distribution remains in the neighborhood of the linearization. This is done by propagating a minimal set of points of the same mean and covariance as the state distribution and calculating the updated state and covariance estimates from their updated values. The set is selected so that the distribution can be modeled to second order accuracy with $2n$ points [65, 64, 138].

More specifically, for an n -dimensional (Gaussian) distribution of states, given a state mean and covariance estimate, we generate a set of points (often called ‘sigma’ points):

$$\chi_0 = \hat{\mathbf{x}}_{k|k} \quad (5.1.52)$$

$$\{\chi_{k|k}^i\}_{i=1}^{2n} = \hat{\mathbf{x}}_{k|k} \pm \text{col} \left[\sqrt{(n + \lambda) \mathbf{P}_{k|k}} \right], \quad (5.1.53)$$

where col represents the column operator (which equals the set of columns of the argument matrix) and the square root of the scaled covariance can be found using the Cholesky decomposition. The scaling parameter λ is defined via positive constant parameters α and γ

such that:

$$\gamma = \alpha^2(n + \lambda) - n. \quad (5.1.54)$$

The sigma points are propagated via the system dynamics:

$$\chi_{k|k-1}^i = \mathbf{f}(\chi_{k-1|k-1}^i, k), \quad (5.1.55)$$

and the predicted mean and covariance are computed as:

$$\hat{\mathbf{x}}_{k|k-1} = \sum_{i=0}^{2n} W_m^i \chi_{k|k-1}^i \quad (5.1.56)$$

$$\mathbf{P}_{k|k-1} = \sum_{i=0}^{2n} W_c^i (\chi_{k|k-1}^i - \hat{\mathbf{x}}_{k|k-1}) (\chi_{k|k-1}^i - \hat{\mathbf{x}}_{k|k-1})^T + \mathbf{Q}_k, \quad (5.1.57)$$

where $\{W_m\}$ and $\{W_c\}$ are weight sets defined as:

$$W_m^i = \begin{cases} \frac{\gamma}{n+\gamma} & i = 0 \\ \frac{1}{2(n+\gamma)} & \text{else} \end{cases} \quad (5.1.58)$$

$$W_c^i = \begin{cases} \frac{\gamma}{(n+\gamma)+(1-\alpha^2+\beta)} & i = 0 \\ \frac{1}{2(n+\gamma)} & \text{else} \end{cases} \quad (5.1.59)$$

with:

$$\kappa = \alpha^2(n + \lambda) - n \quad (5.1.60)$$

for positive constants β . A new set of sigma points $\{\chi_{k|k}^i\}$ is generated from the predicted covariance $\mathbf{P}_{k|k-1}$ and the predicted observations are calculated from the observation equation:

$$\zeta_k^i = \mathbf{h}(\chi_{k|k}^i, k). \quad (5.1.61)$$

The Kalman gain can then be calculated as:

$$\mathbf{K}_k = \left(\sum_{i=0}^{2n} \left(\chi_{k|k}^i - \sum_{j=0}^{2n} W_m^j \chi_{k|k-1}^j \right) W_c^i \left(\zeta_k^i - \sum_{j=0}^{2n} W_m^j \zeta_k^j \right)^T \right) \mathbf{S}_k^{-1} \quad (5.1.62)$$

where

$$\mathbf{S}_k = \sum_{i=0}^{2n} \left(\zeta_k^i - \sum_{j=0}^{2n} W_m^j \zeta_k^j \right) W_c^i \left(\zeta_k^i - \sum_{j=0}^{2n} W_m^j \zeta_k^j \right)^T - \mathbf{R}_k, \quad (5.1.63)$$

so that the state and covariance updates become:

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k \left(\mathbf{z}_k - \sum_{i=0}^{2n} W_m^i \zeta_k^i \right) \quad (5.1.64)$$

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}_k \mathbf{S}_k \mathbf{K}_k^T. \quad (5.1.65)$$

The literature varies in the particular values to be used for the various constant parameters. Van Der Merwe and Wan [138] suggest that α (which determines the spread of the sigma points about the state estimate) be taken from the range of $(1 \times 10^{-4}, 1)$, β (which determines how much prior knowledge of the state distribution should be incorporated) should be equal to 2 for Gaussian distributions, and that λ should be set to zero for state estimation and $n - 3$ for parameter estimation. The optimal values for a given system are often determined via empirical checks, rather than being derived rigorously from first principles. As with the EKF, the UKF can be extended to continuous time systems and hybrid systems [109]. A sample implementation of a hybrid UKF is presented in Listing A.8.

In addition to Kalman filters, there is also the family of particle filters, also known as sequential Monte Carlo (SMC) methods. These also model the state distribution via a population of points, but use much larger sets in an attempt to capture the true state distribution. This means that these filters can relax the assumptions of Gaussian noise at the cost of a significantly higher computational burden [139, 96]. Another recent innovation on the UKF—the j th moment extended Kalman filter (JMEKF)—involves the propagation of

higher order moments of the state distribution (instead of just the first two) and is described in Majji et al. [83]. This technique requires automated expansion of the system dynamics about the current state estimate and can be very computationally intensive. The latest refinements of this filter include incorporation of perturbation techniques to derive the time evolution of the higher order moments of the estimation error [84]. For the purposes of the analysis in this chapter, we assume that the process and observation noise are approximately Gaussian, and use the formalism of the EKF and UKF. However, as advances are made in the efficient computation of the various higher order filters, it may prove useful to apply these as well. The same descriptions of the state, system dynamics, and observation functions can be utilized, regardless of the specific filter update equations.

5.2 Initializing the State

5.2.1 Selecting the State Vector

In order to apply the formalism developed in the previous section, it is first necessary to define the state that will be used to describe the exosystem. The only requirements on this state are that it must be able to model the dynamics of the exosystem (i.e., the orbital motion) and that it must map to any observation method. The simpler (and more linear) these mappings, the better. In fact, we have already defined one such state when discussing the parameter sets necessary to encode simulated planetary systems in §4.2.1. We can define a ‘dynamic’ state from the parameters in Equation (4.2.4):

$$\mathbf{x}_d = \begin{bmatrix} \mathbf{r}_{i/G} & {}^T\mathbf{v}_{i/G} & \mu_i \end{bmatrix}^T, \quad i \in X \quad (5.2.1)$$

where X is the set of all exosystem bodies (see Equation (1.2.33)) and G is the exosystem barycenter[‡]. As discussed in §1.2 and §4.2.2, one instance of the elements of \mathbf{x}_d is sufficient

[‡]While the ordering of the specific elements is fairly arbitrary, the equations in the previous section all assume that the state is a column vector. By default, we will stack the posi-

to propagate the orbits of the exosystem indefinitely both forwards and backwards in time.

As shown in Chapter 2, \mathbf{x}_d can be mapped to a variety of observation methods, as long as we incorporate a few additional constant parameters:

$$\mathbf{x}_c = \begin{bmatrix} p_j & R_i & \mathbf{r}_\mu & \varpi \end{bmatrix}^T, \quad j \in Y \quad (5.2.2)$$

where Y is the set of all exosystem bodies excluding the star (see Equation (4.2.49)). We can also include the planets' effective temperatures, if we are interested in detailed modeling of the spectra, or using spectral information to improve orbital fits [1, 51, 13]. Our final state is the combination of the static and dynamic states:

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_d \\ \mathbf{x}_c \end{bmatrix}. \quad (5.2.3)$$

It should be noted that this state contains a combination of time varying and constant (but possibly unknown) elements. Such parameter-adaptive filtering is a fairly standard problem and has been extensively discussed in the literature [132, 24, 47, 90, 108]. In the EKF formulation, the filtering equations remain exactly the same as in the purely dynamic state case, but in the UKF, the values of the weight parameters are often varied for parameter estimation [64]. It must be pointed out that there is no guarantee that these filters will always converge on the true parameter values, and it is very helpful to start with initial parameter values that are as close to the true state as possible.

Before proceeding, it is also important to mention that \mathbf{x}_d is by no means the only way we have to describe the orbital dynamics. Rather than propagating positions, we can have a state composed of fixed or osculating orbital elements. These can be the Keplerian orbital elements discussed in §1.2 either using the usual three angles to describe orbit orientation, or any other method for orienting a body in space, such as quaternions. Alternatively, we can

tions and velocities of each body in turn, followed by all the gravitational parameters, i.e.: $\mathbf{x}_d = [\mathbf{r}_{1/G} \ \tau_{\mathbf{v}_{1/G}} \ \mathbf{r}_{2/G} \ \tau_{\mathbf{v}_{2/G}} \ \dots \ \mu_1 \ \mu_2 \ \dots]^T$.

encode the dynamic state using eccentricity and angular momentum vectors (see Equation (1.2.9)). If we do choose to encode positions and velocities, we are not restricted to the barycentric coordinates used here.

Many symplectic integrators designed to propagate systems where the majority of the mass is concentrated in a central body (i.e., the star) use mixed coordinate sets. Wisdom and Holman [148] use Jacobi coordinates, which cause the n-body Hamiltonian to be cyclic in the center of mass coordinate. This is accomplished by taking the first coordinate (\mathbf{q}_0) to be the system center of mass, and all subsequent coordinates to be the position of the i th body with respect to the center of mass of bodies 1 through $i - 1$:

$$\mathbf{q}_i = \mathbf{r}_i - \frac{\sum_{j=0}^{i-1} m_j \mathbf{r}_j}{\sum_{j=0}^{i-1} m_j}. \quad (5.2.4)$$

Similarly, many implementations use mixed-center (or ‘democratic heliocentric’ coordinates) where the first coordinate is again the position of the center of mass, and the remaining coordinates are the heliocentric positions:

$$\mathbf{q}_i = \mathbf{r}_{i/G} - \mathbf{r}_{S/G}. \quad (5.2.5)$$

The canonical momenta associated with these coordinates are simply the total momentum of the system and the barycentric momenta of the system bodies (not including the star). This coordinate set has been shown to be very useful for long term integration involving close encounters between bodies [32] and is used in John Chambers’ popular MERCURY integrator [20].

While possibly not the most efficient solution, the barycentric coordinates used throughout this thesis are useful because they are simple to interpret, very straightforward to integrate, and map easily to the various observation functions. They are preferable to orbital element descriptions in all cases except for the single planet exosystem and have worked best within the EKF/UKF framework of any of the dynamic states that were evaluated.

5.2.2 Incorporating Prior Knowledge

Having selected the form of the state vector, we must now select initial values for both the state estimate and covariance. These should be derived from:

- Previous measurements of the target system
- Relevant statistics for known exoplanet populations
- Expected values of the distribution functions derived in §3.3

with covariances determined by measurement uncertainty, population variance, and the specific distribution functions, respectively. In the absence of other information, we usually assume that there are no cross-correlations between the state elements. Unfortunately, the full state contains many interdependent parameters, so this assumption may be a poor one. In the case of directly measured values, cross-correlations may be derived from instrument models. For orbital positions we can use our analytic distribution functions and Monte Carlo simulation to estimate the covariance. Figure 5.1 shows the average covariance of \mathbf{x}_d (not including the gravitational parameters) for 10 solar system bodies, found by generating N samples of \mathbf{x}_d (packaged as the $60 \times N$ element matrix X) and calculating:

$$\mathbf{P} = \frac{1}{N}(X - \bar{X})(X - \bar{X})^T, \quad (5.2.6)$$

where \bar{X} represents the equivalent size matrix containing the column means of X . The resulting visualization clearly displays a structure composed of 6×6 blocks. The blocks along the diagonal represent the cross-correlations between state elements of each body (with 3×3 substructures for position and velocity), ordered with the Sun in the top, left corner, followed by the planets and with Pluto in the bottom, right corner. The off-diagonal blocks represent cross-correlations between separate bodies. As expected, the variances and cross-correlations increase with semi-major axis, as larger orbits represent much greater variations in position, over much longer periods.

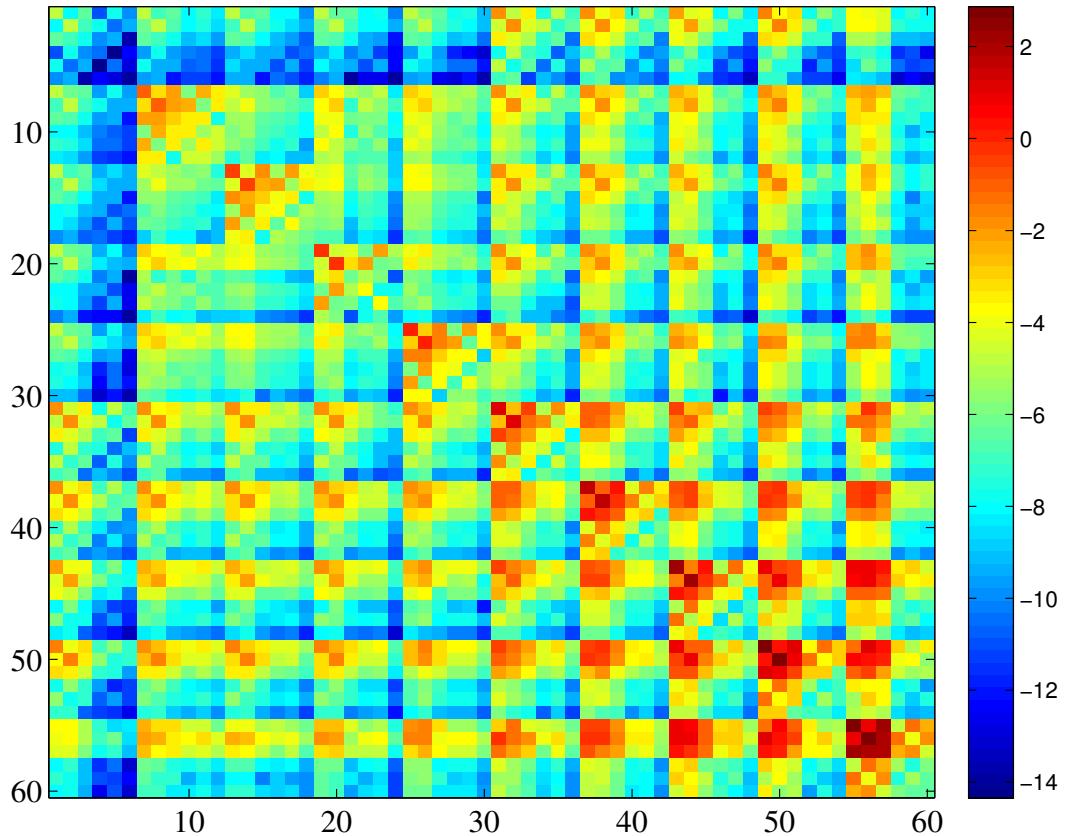


Figure 5.1: Covariance estimate for \mathbf{x}_d (excluding gravitational parameters) for the sun, solar system planets, and pluto, calculated as in Equation (5.2.6). The colorbar is in log scale (base 10), representing the absolute covariance magnitudes. The block structure arises from cross-correlations between parameters associated with specific bodies, creating 6×6 blocks with 3×3 sub-blocks for position and velocity states.

Because so much of the original (pre-Kepler) exoplanet data set is composed of doppler spectroscopy surveys it is useful to briefly discuss how exoplanet signals are extracted from these measurements. As detailed in §2.2.2 doppler spectroscopy measures the radial component of a star's velocity. Planet signals may be identified by finding periodicities in this velocity (or analogously, in the star's position as found by precision astrometry). This is usually accomplished via a periodogram (Fourier spectrum) analysis. For an evenly sampled set of data points $\{x_j\}_{j=1}^N$, the discrete Fourier transform (DFT) is given by:

$$\hat{\mathbf{x}} = \mathbf{F}\mathbf{x} \quad (5.2.7)$$

where \mathbf{x} is the column vector of x_i values and \mathbf{F} is the Fourier matrix defined as:

$$[\mathbf{F}]_{mn} = \frac{1}{\sqrt{N}} e^{-2\pi i mn/N}. \quad (5.2.8)$$

As long as the data is evenly sampled, the set $\{j/N\}_{j=1}^N$ represents the Fourier frequencies which will include all of the strong periodicities found in the data set and whose spectral power can be calculated as:

$$P_j = 2 \left| [\hat{\mathbf{x}}]_j \right|^2. \quad (5.2.9)$$

Since the DFT produces a vector $\hat{\mathbf{x}}$ such that:

$$[\hat{\mathbf{x}}]_j = [\hat{\mathbf{x}}]_{N-j}^* \quad (5.2.10)$$

where $[\cdot]^*$ represents the complex conjugate, the periodogram is only computed for the first half of the matrix.

Unfortunately, in most cases it is difficult to achieve an astronomical data set with completely even sampling due to all of the various target observability effects discussed in §4.3. One method for dealing with this, very commonly found in the literature [3, 116, 25, 128, 26] is the Lomb-Scargle periodogram. For a set of data points $\{x_j\}_{j=1}^N$ collected at times $\{t_j\}_{j=1}^N$ with sample mean \bar{x} and sample variance σ^2 , the Lomb-Scargle periodogram defines the spectral power (as a function of angular frequency ω) as:

$$P_N(\omega) = \frac{1}{2\sigma^2} \left[\frac{\left(\sum_{j=1}^N (x_j - \bar{x}) \cos(\omega(t_j - \tau)) \right)^2}{\sum_{j=1}^N \cos^2(\omega(t_j - \tau))} + \frac{\left(\sum_{j=1}^N (x_j - \bar{x}) \sin(\omega(t_j - \tau)) \right)^2}{\sum_{j=1}^N \sin^2(\omega(t_j - \tau))} \right], \quad (5.2.11)$$

where the time offset τ for each ω is determined by:

$$\tau = \frac{1}{2\omega} \tan^{-1} \left(\frac{\sum_{j=1}^N \sin(2\omega t_j)}{\sum_{j=1}^N \cos(2\omega t_j)} \right). \quad (5.2.12)$$

The calculation of the time offset has two effects. First, it makes the spectral power independent of any constant time shift. Second, it makes the calculation of the periodogram equivalent to performing a least squares fit on the data to the simple harmonic model [81]:

$$x(t) = A \cos \omega t + B \sin \omega t, \quad (5.2.13)$$

where the periodogram is the sum of the square of the coefficients:

$$P_N(\omega) = A^2 + B^2. \quad (5.2.14)$$

For any given angular frequency, the periodogram power, $P_N(\omega)$, has an exponential distribution function with unit mean for the case of the null hypothesis (i.e., that the data points represent independent Gaussian random values). If we test M independent frequencies then we can write the probability of finding a spectral power above some threshold value P_0 is given by:

$$P[P_N(\omega) > P_0] = 1 - (1 - e^{-P_0})^M. \quad (5.2.15)$$

This probability is equivalent to the false positive probability of the null hypothesis, and is therefore the significance level of $P_N(\omega)$. The lower this value, the more significant the associated frequency. This significance value scales approximately linearly with the number of independent frequencies M . This value may be determined by Monte Carlo simulations where synthetic data sets of Gaussian deviates are generated for a fixed number of data points and the resulting periodogram values are fit for M . Such experimentation has yielded the general result that M is approximately equal to N for approximately evenly sampled data, and when the range from zero to the Nyquist frequency is oversampled. The value of

M differs from N significantly only in cases where data points are closely clumped, in which case M is reduced by a factor equal to the average clump size [58, 103].

Much effort has been devoted to finding fast implementations of the Lomb-Scargle periodogram, with one very popular method based on the fast Fourier transform (FFT) described in Press and Rybicki [102]. While these methods can be made accurate to any desired precision, modern computers are fully capable of calculating the exact periodogram for data sets of several thousand points. A sample implementation of this algorithm is presented in Listing A.9. In cases where astrometric or radial velocity data is available, the periodogram can be very valuable in initializing the state by providing the number of planets in the system and their periods. Furthermore, it is common practice to fit orbits directly to the significant periodogram frequencies, generating even better starting points for the filter [126, 127].

5.3 Updating the State

5.3.1 State Dynamics

The system dynamics are governed by the orbital mechanics described in §1.2. Given the state defined in Equation (5.2.3), the system dynamics are governed by the ODE:

$$\dot{\mathbf{x}} = f(\mathbf{x}) = \begin{bmatrix} \mathcal{I}\mathbf{v}_{1/G} \\ -\sum_{k \neq 1} \frac{\mu_k \mathbf{r}_{k/j}}{\|\mathbf{r}_{k/j}\|^3} \\ 0 \\ \vdots \\ \mathcal{I}\mathbf{v}_{n/G} \\ -\sum_{k \neq n} \frac{\mu_k \mathbf{r}_{k/j}}{\|\mathbf{r}_{k/j}\|^3} \\ 0 \\ \mathbf{0}_{c \times 1} \end{bmatrix} \quad (5.3.1)$$

where $\mathbf{r}_{k/j} = \mathbf{r}_{k/G} - \mathbf{r}_{j/G}$, c is the length of \mathbf{x}_c and n is the size of set X .

The Jacobian of $f(\mathbf{x})$ can be decomposed into row blocks:

$$\mathbf{F} = \frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}} = \begin{bmatrix} \mathbf{F}_1 \\ \mathbf{0}_{1 \times 7n+c} \\ \mathbf{F}_2 \\ \mathbf{0}_{1 \times 7n+c} \\ \vdots \\ \mathbf{F}_n \\ \mathbf{0}_{c+1 \times 7n+c} \end{bmatrix} \quad (5.3.2)$$

where:

$$\mathbf{F}_j = \begin{bmatrix} \mathbf{0}_{3 \times 3} & \mathbf{I} & \mathbf{0}_{3 \times 1} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 1} & \cdots & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times c} \\ \mathbf{F}_{j1} & \mathbf{0}_{3 \times 3} & \mathbf{G}_{j1} & \mathbf{F}_{j2} & \mathbf{0}_{3 \times 3} & \mathbf{G}_{j2} & \cdots & \mathbf{F}_{jn} & \mathbf{G}_{jn} & \mathbf{0}_{3 \times c} \end{bmatrix}. \quad (5.3.3)$$

The submatrices \mathbf{F}_{jk} are the partial derivatives of 1.2.32 with respect to the position vectors and the submatrices \mathbf{G}_{jk} are the partial derivatives with respect to the gravitational parameters. Taking $\mathbf{r}_{k/G} = \begin{bmatrix} r_{k1} & r_{k2} & r_{k3} \end{bmatrix}^T$ and $\mathbf{r}_{j/G} = \begin{bmatrix} r_{j1} & r_{j2} & r_{j3} \end{bmatrix}^T$, we have:

$$\mathbf{r}_{k/j} = \begin{bmatrix} r_{k1} - r_{j1} \\ r_{k2} - r_{j2} \\ r_{k3} - r_{j3} \end{bmatrix} = \begin{bmatrix} r_{kj1} \\ r_{kj2} \\ r_{kj3} \end{bmatrix}. \quad (5.3.4)$$

This allows us to write:

$$\mathbf{F}_{jk} = \begin{cases} -\sum_{k \neq j} \mu_k \begin{bmatrix} \frac{3r_{kj_1}^2}{|\mathbf{r}_{k/j}|^5} - \frac{1}{|\mathbf{r}_{k/j}|^3} & \frac{3r_{kj_1}r_{kj_2}}{|\mathbf{r}_{k/j}|^5} & \frac{3r_{kj_1}r_{kj_3}}{|\mathbf{r}_{k/j}|^5} \\ \frac{3r_{kj_1}r_{kj_2}}{|\mathbf{r}_{k/j}|^5} & \frac{3r_{kj_2}^2}{|\mathbf{r}_{k/j}|^5} - \frac{1}{|\mathbf{r}_{k/j}|^3} & \frac{3r_{kj_2}r_{kj_3}}{|\mathbf{r}_{k/j}|^5} \\ \frac{3r_{kj_1}r_{kj_3}}{|\mathbf{r}_{k/j}|^5} & \frac{3r_{kj_2}r_{kj_3}}{|\mathbf{r}_{k/j}|^5} & \frac{3r_{kj_3}^2}{|\mathbf{r}_{k/j}|^5} - \frac{1}{|\mathbf{r}_{k/j}|^3} \end{bmatrix} & j \neq k \\ \mu_k \begin{bmatrix} \frac{3r_{kj_1}^2}{|\mathbf{r}_{k/j}|^5} - \frac{1}{|\mathbf{r}_{k/j}|^3} & \frac{3r_{kj_1}r_{kj_2}}{|\mathbf{r}_{k/j}|^5} & \frac{3r_{kj_1}r_{kj_3}}{|\mathbf{r}_{k/j}|^5} \\ \frac{3r_{kj_1}r_{kj_2}}{|\mathbf{r}_{k/j}|^5} & \frac{3r_{kj_2}^2}{|\mathbf{r}_{k/j}|^5} - \frac{1}{|\mathbf{r}_{k/j}|^3} & \frac{3r_{kj_2}r_{kj_3}}{|\mathbf{r}_{k/j}|^5} \\ \frac{3r_{kj_1}r_{kj_3}}{|\mathbf{r}_{k/j}|^5} & \frac{3r_{kj_2}r_{kj_3}}{|\mathbf{r}_{k/j}|^5} & \frac{3r_{kj_3}^2}{|\mathbf{r}_{k/j}|^5} - \frac{1}{|\mathbf{r}_{k/j}|^3} \end{bmatrix} & j = k \end{cases} \quad (5.3.5)$$

and

$$\mathbf{G}_{jk} = \begin{cases} -\frac{\mathbf{r}_{k/j}}{\|\mathbf{r}_{k/j}\|^3} & j \neq k \\ \mathbf{0}_{3 \times 1} & j = k. \end{cases} \quad (5.3.6)$$

Since numerical integration of Equation (5.1.42) requires a first order integrator, the Runge-Kutta-Nyström integrators cannot be used to propagate the state and covariance and must be replaced with a first order symplectic integrator. Because of the large quantities of zero entries in the Jacobians, a sparse matrix encoding is often used to increase filtering efficiency.

5.3.2 Observation Equations

In order to write the observation functions for the various detection methods discussed in Chapter 2, we must first decide what to count as the observation for each method. We could start with the absolute raw form of each measurement, such as pixel data numbers for imaging, or uncalibrated photon counts for transit photometry, but this would place a significant extra burden on the filter by making all of the observation equations highly

nonlinear and requiring an additional set of parameters (beyond the ones already defined) to map the observations to the state. More importantly, each detection method has a well-developed set of tools for processing the raw data. These have been developed and tested over many years, and so we can have a high degree of confidence in their outputs. Therefore, we will assume that the standard processing typically applied to each observation has been performed and use that as a starting point. This could potentially introduce additional noise sources into the data sets, but as long as these can be characterized and incorporated into the appropriate \mathbf{R} covariances this should not have a large effect on the filter performance.

Thus we take the observations to be:

$$\mathbf{z}_{dd} = \begin{bmatrix} \mathbf{s} \\ \frac{F_P}{F_S} \end{bmatrix}, \quad (5.3.7)$$

$$\mathbf{z}_{ast} = \begin{bmatrix} \mathbf{b}_1 \cdot (\hat{\mathbf{r}}_{S/sc} - \hat{\mathbf{r}}_{c/sc}) \\ \mathbf{b}_2 \cdot (\hat{\mathbf{r}}_{S/sc} - \hat{\mathbf{r}}_{c/sc}) \end{bmatrix} B, \quad (5.3.8)$$

$$\mathbf{z}_{rv} = \mathbf{r}_{S/G}, \quad (5.3.9)$$

$$\mathbf{z}_{phot} = \frac{F_P}{F_S}, \quad (5.3.10)$$

for direct detection, astrometry,[§] doppler spectroscopy, and transit photometry, respectively, with the appropriate noise terms added. Of these four, only doppler spectroscopy is fully linear in the state vector, whereas astrometry is the most highly nonlinear. The corresponding Jacobian for direct detection is:

$$\mathbf{H}_{dd} = \begin{bmatrix} \mathbf{I}_{2 \times 2} & \mathbf{0}_{2 \times 7} \\ \mathbf{H}_z \end{bmatrix} \quad (5.3.11)$$

[§]Since the baseline length, B , and orientations \mathbf{b}_i are properties of the instrument, we assume that these are known parameters and do not include them into the state. However, since the orientation knowledge for a real astrometric space-based observatory would be built up over the course of many observations [95] an alternate filter formulation designed to run before the mission had completed would need to include the orientation vectors in the state.

where

$$\mathbf{H}_z = \frac{p_j R_j^2 r_{j1}}{\pi r^4} \begin{bmatrix} -\left(\sqrt{r_{j1}^2 + r_{j2}^2} + \pi r_{j3} + 2r_{j3} \sin^{-1}(r_{j3}/r)\right) \\ -\left(\sqrt{r_{j1}^2 + r_{j2}^2} + \pi r_{j3} + 2r_{j3} \sin^{-1}(r_{j3}/r)\right) \\ \left(-\sqrt{r_{j1}^2 + r_{j2}^2} r_{j3} + (r_{j1}^2 + r_{j2}^2 - r_{j3}^2)(\pi - \cos^{-1}(r_{j3}/r))\right) \\ \mathbf{0}_{4 \times 1} \\ (r^2/p_j) \left(\sqrt{r_{j1}^2 + r_{j2}^2} + \pi r_{j3} - r_{j3} \cos^{-1}(r_{j3}/r)\right) \\ (2r^2/R_j) \left(\sqrt{r_{j1}^2 + r_{j2}^2} + \pi r_{j3} + r_{j3} \sin^{-1}(r_{j3}/r)\right) \end{bmatrix} \quad (5.3.12)$$

for the compact state $\mathbf{x} = \begin{bmatrix} \mathbf{r}_{j/G} & \mathcal{I} \mathbf{v}_{j/G} & \mu_j & p_j & R_j \end{bmatrix}$ with $r = \|\mathbf{r}_{j/G}\|$.

Because the astrometric observation is so highly nonlinear, it is useful to try to find a to find a linearization (or low order series solutions) consistent with the original observation equation to the required level of precision. The next section demonstrates the technique for doing so, using Earth-like planets to determine required precision.

5.3.3 Astrometric Observation Function Linearization

The classical approach to linearizing Equation (2.2.12) has been to expand about ϖ . As we will see below, for the purposes of finding Earth-mass planets, these must necessarily include nonlinear terms, but an expansion is still useful as a simplifying step in the data analysis. In order to decide which terms should be retained, we must first quantify the desired precision of the measurement. There are several ways to do this, but a relatively straightforward one is to calculate the magnitude of the smallest signal we wish to be able to measure. Following Savransky and Kasdin [110], we will take as our smallest signal the effect of a single Earth-mass planet residing on an Earth-like orbit on the astrometric signature of a sun-twin star.

Figure 5.2 shows the difference between the results of two simulations, each using Equation (2.2.12) and a fixed centroid position (with respect to the solar system barycenter). The

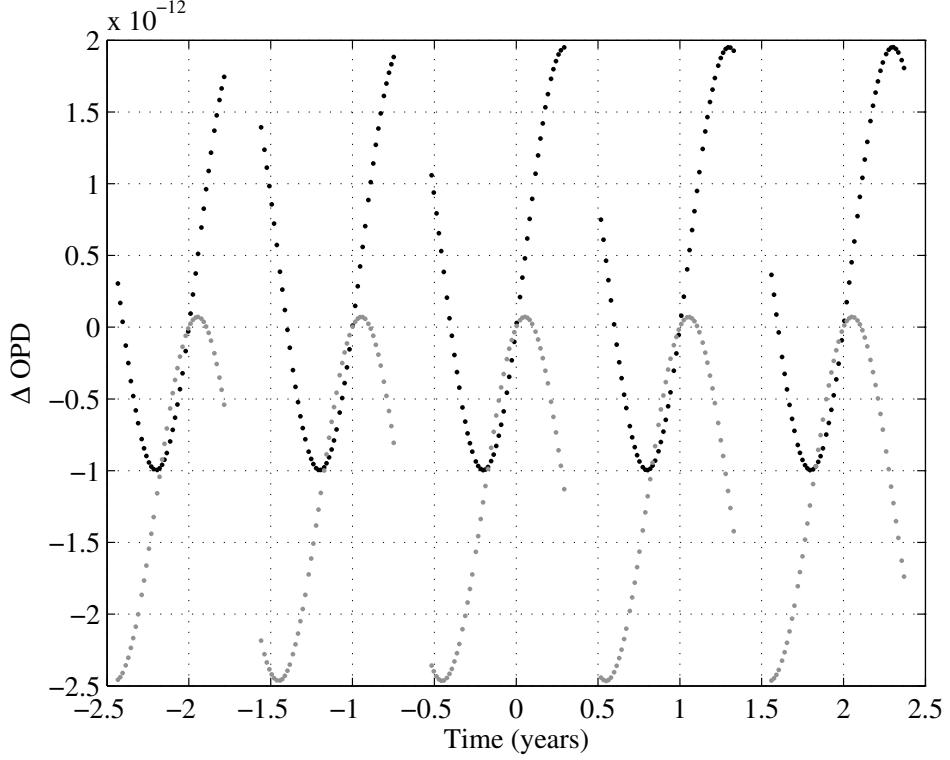


Figure 5.2: Difference between differential OPDs (between a sun-twin star and fixed centroid) using exact expressions for $\hat{\mathbf{r}}_{s/sc}$ with and without the effects of an Earth-mass planet on an Earth-like orbit. The two curves represent the measurements along two orthogonal interferometer baseline orientations.

simulations include identical values for parallax, barycenter motion, and assume a sun-twin target star lying exactly 10 parsecs from the solar system barycenter (at epoch t_0). One simulation includes the stellar jitter due to an Earth-mass planet while the other does not[†]. The difference between the two measurements is the most basic representation of the magnitude of the signal in which we are interested, and, in this case, shows a signal on the order of 1×10^{-12} . This value is in the units of Equation (2.2.7): it is the difference between fractions of the baseline distance.

This tells us that to be sensitive to the jitter due to an Earth-twin with any measure of confidence, we must retain all terms in the measurement of order 1×10^{-13} or greater (assuming that our analysis technique will be sensitive to structured signals up to one order

[†]The line of nodes of the planet's orbit is rotated by 45° with respect to the line of sight, producing the asymmetric magnitudes about 0 in the difference between differential OPDs

of magnitude below the target signal). To aid in this calculation, Table 5.1 enumerates the ranges of typical values for the various terms in Equation (2.2.12).

Table 5.1: Typical values for terms in Equation (2.2.12).

Term	Typical Order	Notes
a_0	1	All distances are assumed to be in AU
$\hat{\mathbf{r}}_s(t_0)$	1	
$\bar{\mathbf{r}}_\mu^*$	5×10^{-6} /year	Barnard's star is 5×10^{-5} /year
$\Delta\tilde{\mathbf{r}}_{s/G}$	1×10^{-6} to 1×10^{-2}	For solar system planets
$\tilde{\mathbf{r}}_{sc}$	1	Assuming a ground-based, Earth orbiting or Earth-trailing spacecraft
ϖ	5×10^{-6} to 5×10^{-8}	For target distance of 1-100 parsecs.

* While $\bar{\mathbf{r}}_\mu$ is a time dependent value, for measurements made over a span of 5 to 10 years it will change by less than one order, so we will assume a conservative, constant value when considering which terms to keep in our expansions.

We begin the expansion by dividing the numerator and denominator of Equation (2.2.12) by $\|\mathbf{r}_s(t_0)\|$ to write the normalized version of the equation:

$$\begin{aligned} \hat{\mathbf{r}}_{s/sc} = & (\hat{\mathbf{r}}_s(t_0) + \bar{\mathbf{r}}_\mu + \varpi \Delta\tilde{\mathbf{r}}_{s/G} - \varpi \tilde{\mathbf{r}}_{sc}) \times \\ & \left\{ 1 + \bar{\mathbf{r}}_\mu \cdot \bar{\mathbf{r}}_\mu + \varpi^2 \Delta\tilde{\mathbf{r}}_{s/G} \cdot \Delta\tilde{\mathbf{r}}_{s/G} + \varpi^2 \tilde{\mathbf{r}}_{sc} \cdot \tilde{\mathbf{r}}_{sc} + 2\hat{\mathbf{r}}_s(t_0) \cdot \bar{\mathbf{r}}_\mu \right. \\ & + 2\varpi \hat{\mathbf{r}}_s(t_0) \cdot \Delta\tilde{\mathbf{r}}_{s/G} - 2\varpi \hat{\mathbf{r}}_s(t_0) \cdot \tilde{\mathbf{r}}_{sc} + 2\varpi \bar{\mathbf{r}}_\mu \cdot \Delta\tilde{\mathbf{r}}_{s/G} \\ & \left. - 2\varpi \bar{\mathbf{r}}_\mu \cdot \tilde{\mathbf{r}}_{sc} - 2\varpi^2 \Delta\tilde{\mathbf{r}}_{s/G} \cdot \tilde{\mathbf{r}}_{sc} \right\}^{-\frac{1}{2}} \end{aligned} \quad (5.3.13)$$

where $\bar{\mathbf{r}}_\mu = \mathbf{r}_\mu / \|\mathbf{r}_s(t_0)\|$, $\tilde{\mathbf{r}}_{sc}$ is the spacecraft position normalized by a_0 and $\Delta\tilde{\mathbf{r}}_{s/G}$ is $\Delta\mathbf{r}_{s/G}$ normalized by a_0 .

We next apply a binomial expansion to the denominator of Equation (5.3.13) and retain terms explicitly in ϖ to first order. Since the signal we are looking for appears explicitly only as $\varpi \Delta\tilde{\mathbf{r}}_{s/G}$, any terms of proportional order must be retained. Thus, terms proportional to $\|\bar{\mathbf{r}}_\mu\|^2$ must be left in, but terms proportional to $\|\bar{\mathbf{r}}_\mu\|^n$ for $n > 2$ and terms proportional to $\|\bar{\mathbf{r}}_\mu\|^n \varpi$ for $n > 1$ are dropped. We also drop all terms proportional to $\hat{\mathbf{r}}_s(t_0)$, as dotting

$\hat{\mathbf{r}}_{s/sc}$ with \mathbf{b}_i as in the measurement Equation (2.2.7) causes these terms to equal zero. The resulting approximation is:

$$\begin{aligned}\hat{\mathbf{r}}_{s/sc} \cdot \mathbf{b}_i &\approx \left\{ \bar{\mathbf{r}}_\mu - (\hat{\mathbf{r}}_s(t_0) \cdot \bar{\mathbf{r}}_\mu) \bar{\mathbf{r}}_\mu + [\Delta \tilde{\mathbf{r}}_{s/G} - \tilde{\mathbf{r}}_{sc} - (\hat{\mathbf{r}}_s(t_0) \cdot \bar{\mathbf{r}}_\mu) \Delta \tilde{\mathbf{r}}_{s/G}\right. \\ &\quad \left. + (\hat{\mathbf{r}}_s(t_0) \cdot \tilde{\mathbf{r}}_{sc}) \bar{\mathbf{r}}_\mu - (\Delta \tilde{\mathbf{r}}_{s/G} \cdot \hat{\mathbf{r}}_s(t_0)) \bar{\mathbf{r}}_\mu + (\hat{\mathbf{r}}_s(t_0) \cdot \bar{\mathbf{r}}_\mu) \tilde{\mathbf{r}}_{sc}] \varpi \right\} \cdot \mathbf{b}_i.\end{aligned}\tag{5.3.14}$$

This is a linear expression in parallax and can be used to extract the stellar reflex signal, though it contains complicated couplings among the spacecraft position, parallax, barycenter motion, and stellar reflex. Note that it is not assumed that these other terms are known; rather, the barycenter motion and the parallax factor, ϖ , must also be estimated in the data analysis. This leads to the problem of attempting to fit the radial motion of the target from purely astrometric measurements. While the radial and proper motions of targets are approximated as completely separate in the classical treatment, the level of precision required for this application means that we must consider the effects of radial motion. The first term of Equation (5.3.14) is $\bar{\mathbf{r}}_\mu \cdot \mathbf{b}_i$. If \mathbf{b}_i is, as was assumed in §2.2.1, equivalent to \mathbf{b}_1 or \mathbf{b}_2 , then the σ_z dependence of this term is exactly zero. Similarly, the next two terms in the direction of $\bar{\mathbf{r}}_\mu$ will have no σ_z components when dotted with \mathbf{b}_i . However, the two final terms in this expansion have magnitudes given by $\varpi(\hat{\mathbf{r}}_s(t_0) \cdot \bar{\mathbf{r}}_\mu)$, which makes them proportional to σ_z . As these two terms are in the directions of $\tilde{\mathbf{r}}_{sc}$ and $\Delta \tilde{\mathbf{r}}_{s/G}$, which are arbitrarily oriented in the tangent frame, this σ_z dependence is not zeroed even when the baseline is perfectly aligned with \mathbf{b}_1 or \mathbf{b}_2 . Thus, the radial star motion explicitly enters our expression, and makes a significant contribution at the desired level of precision.

Figure 5.3 repeats the simulation from Figure 5.2, including the planet in both cases, but now comparing the exact expression for $\hat{\mathbf{r}}_{s/sc}$ and the first order expansion in Equation (5.3.14). The resulting error is less than one order of magnitude below the signal due to the planet, which means that it is of the same order as the desired sensitivity for all targets closer than 10 pc or with any combination of the other parameters which would cause the planet signal to decrease in magnitude. Additionally, the structure of the error is

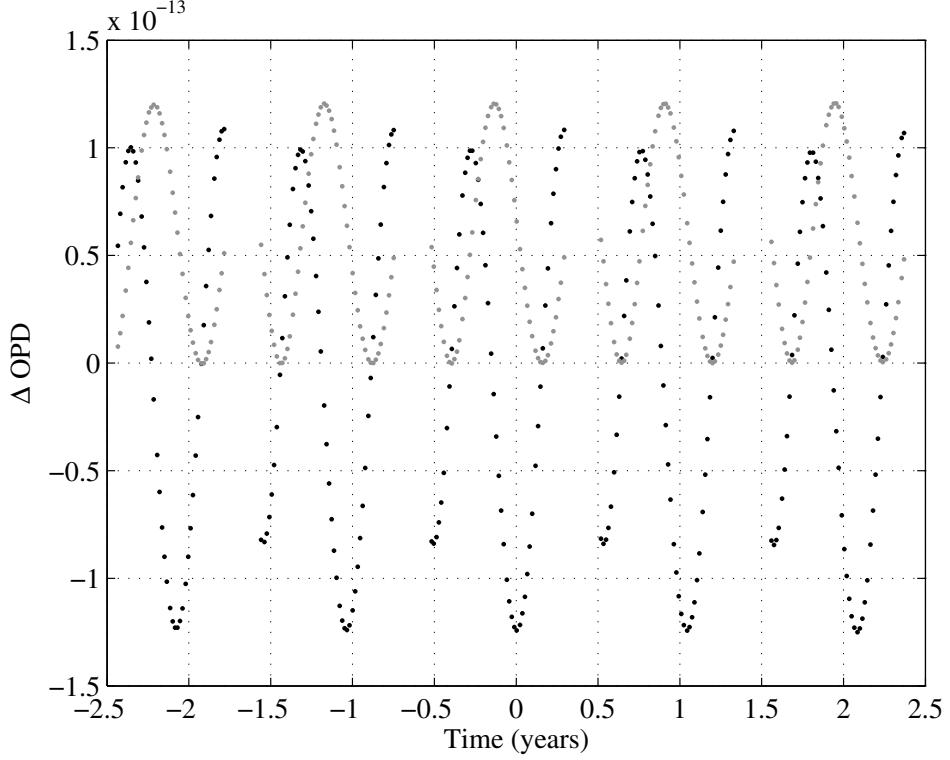


Figure 5.3: Difference between differential OPDs (between a sun-twin star and fixed centroid) using the exact expression for $\hat{\mathbf{r}}_{s/sc}$ and the first order expansion from Equation (5.3.14). The two curves represent the measurements along two orthogonal interferometer baseline orientations.

dominated by parallax terms, which produce periodic structure that could be mistaken for astrometric wobble due to another planet. We can compare Equation (5.3.14) with previously published first-order forms such as Equations 31 and 32 in Konacki et al. [72]. These are quite similar in form to the expression derived here, except that Konacki assumes that all radial components of motion are unobservable, and thus the equivalent vectors to $\Delta\tilde{\mathbf{r}}_{s/G}$ and $\bar{\mathbf{r}}_\mu$ are formed without the $\hat{\mathbf{r}}_s(t_0)$ components. Konacki's expression also does not subtract the initial displacement of the target star from its system's barycenter, thereby applying the proper motion to the star rather than the system barycenter. If we redo our simulation using the first order expansion with the radial terms dropped, we end up with residuals one order of magnitude higher than those in Figure 5.3. Thus, the exclusion of radial terms produces residuals that are one order of magnitude above the target signal, rather than one order of

magnitude below, as seen in Figure 5.4.

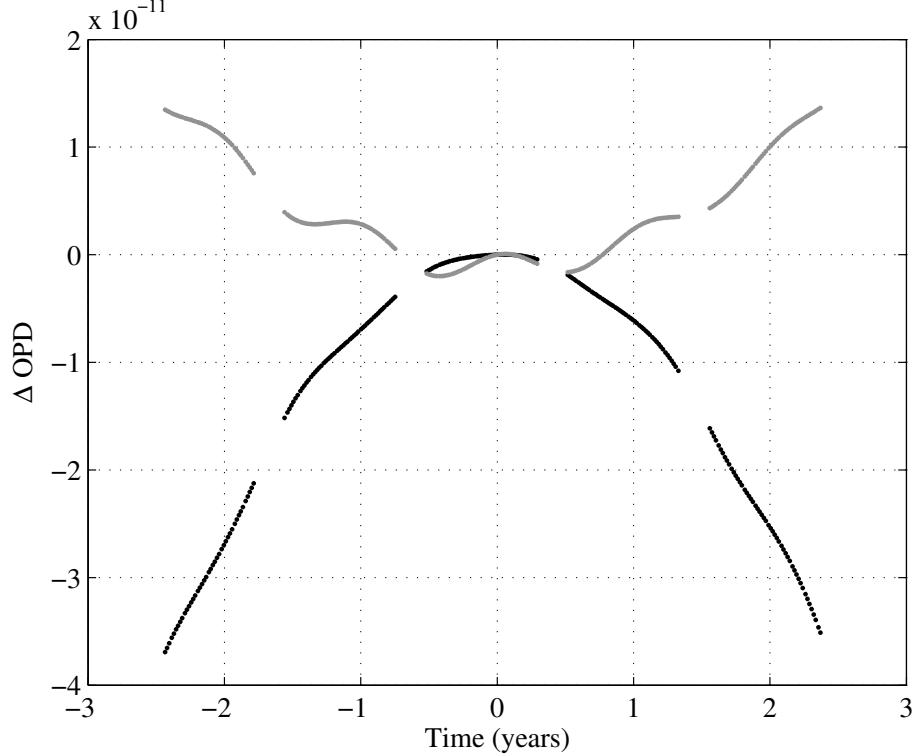


Figure 5.4: Difference between differential OPDs (between a sun-twin star and fixed centroid) using the exact expression for $\hat{\mathbf{r}}_{s/sc}$ and the first order expansion from Equation (5.3.14) with the radial terms neglected. The two curves represent the measurements along two orthogonal interferometer baseline orientations.

Furthermore, a number of the terms dropped in this expansion were proportional to ϖ^2 multiplied by quantities of order 1. The astrometric literature is quite clear that a first order expansion in ϖ is only good to milliarcsec accuracy [44]. If an astrometric instrument has a final precision of sub- μ arcsec, it is important to ask if measurable terms were dropped in the expansion. Such terms can get as large as 6 times ϖ^2 . Again, for a star at 10 pc, $\varpi \sim 5 \times 10^{-7}$, making the second order terms of order 2.5×10^{-13} . Thus, for most stars the approximation is a good one but it does raise concerns for the closest targets, or for analysis methods that are sensitive to non-random signals just below the level of the target signal.

To address this, we can repeat the expansion to second order in ϖ . However, for consistency, we must now retain some additional terms. The zeroth and first order terms (in ϖ)

remain the same (since $\|\bar{\mathbf{r}}_\mu\|^3$ is of order 1×10^{-16} and $\|\bar{\mathbf{r}}_\mu\|^2\varpi$ is of order 1×10^{-17}). Any terms in ϖ^2 with factors of $\bar{\mathbf{r}}_\mu$ can safely be dropped. Terms in ϖ^2 with factors of $\Delta\tilde{\mathbf{r}}_{s/G}$ should also be quite small, but we will leave these to first order since it is theoretically possible that there exist systems with both Earth-like planets and very large, widely separated super-Jupiters which have not yet been detected by RV (due to their very long periods), and which would cause a large stellar reflex to make these factors significant. These considerations leave only three new terms in the expansion:

$$\begin{aligned} \hat{\mathbf{r}}_{s/sc} \cdot \mathbf{b}_i \approx & \left\{ \bar{\mathbf{r}}_\mu - (\hat{\mathbf{r}}_s(t_0) \cdot \bar{\mathbf{r}}_\mu) \bar{\mathbf{r}}_\mu + [\Delta\tilde{\mathbf{r}}_{s/G} - \tilde{\mathbf{r}}_{sc} - (\hat{\mathbf{r}}_s(t_0) \cdot \bar{\mathbf{r}}_\mu) \Delta\tilde{\mathbf{r}}_{s/G} \right. \\ & + (\hat{\mathbf{r}}_s(t_0) \cdot \tilde{\mathbf{r}}_{sc}) \bar{\mathbf{r}}_\mu - (\Delta\tilde{\mathbf{r}}_{s/G} \cdot \hat{\mathbf{r}}_s(t_0)) \bar{\mathbf{r}}_\mu + (\hat{\mathbf{r}}_s(t_0) \cdot \bar{\mathbf{r}}_\mu) \tilde{\mathbf{r}}_{sc}] \varpi \\ & \left. + [\tilde{\mathbf{r}}_{sc}(\hat{\mathbf{r}}_s(t_0) \cdot \Delta\tilde{\mathbf{r}}_{s/G}) + \Delta\tilde{\mathbf{r}}_{s/G}(\hat{\mathbf{r}}_s(t_0) \cdot \tilde{\mathbf{r}}_{sc}) \right. \\ & \left. - \tilde{\mathbf{r}}_{sc}(\hat{\mathbf{r}}_s(t_0) \cdot \tilde{\mathbf{r}}_{sc})] \varpi^2 \right\} \cdot \mathbf{b}_i . \end{aligned} \quad (5.3.15)$$

Figure 5.5 shows the difference between the exact expression for $\hat{\mathbf{r}}_{s/sc}$ and the second order expansion in Equation (5.3.15). The resulting error is now two to three orders of magnitude below the target signal. The error structure is now dominated by proper motion terms, which are linear, not periodic. These could still be confused with the effects of long period planets, but the signals are so low compared to the target signal that this possibility is unlikely. It is important to note, however, the fitting problem has become significantly more challenging as the parameters enter nonlinearly in the measurement.

The practice of dropping terms that are in the direction of $\hat{\mathbf{r}}_s(t_0)$ breaks down in two cases: when there is pointing error such that the interferometer baseline is not aligned with \mathbf{b}_1 or \mathbf{b}_2 , or when expressing the direction to a different star (i.e., a non-ideal reference source) in the tangent frame defined by the target star. In these cases, a number of additional terms must be included in our expansion. Following the same procedure as above, we can rewrite

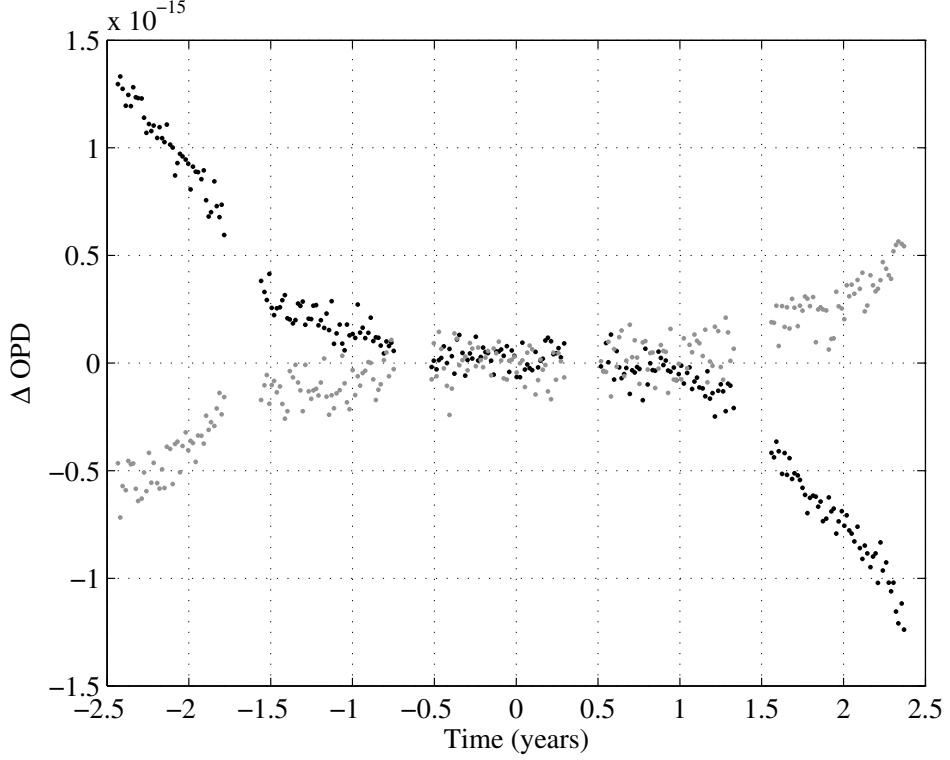


Figure 5.5: Difference between differential OPDs (between a sun-twin star and fixed centroid) using the exact expression for $\hat{\mathbf{r}}_{s/sc}$ and the second order expansion from Equation (5.3.15). The two curves represent the measurements along two orthogonal interferometer baseline orientations.

Equation (5.3.15) as:

$$\begin{aligned}
 \hat{\mathbf{r}}_{s/sc} \approx & \hat{\mathbf{r}}_s(t_0) + \bar{\mathbf{r}}_\mu - \hat{\mathbf{r}}_s(t_0)(\hat{\mathbf{r}}_s(t_0) \cdot \bar{\mathbf{r}}_\mu) - \bar{\mathbf{r}}_\mu(\hat{\mathbf{r}}_s(t_0) \cdot \bar{\mathbf{r}}_\mu) - \frac{1}{2}\hat{\mathbf{r}}_s(t_0)(\bar{\mathbf{r}}_\mu \cdot \bar{\mathbf{r}}_\mu) \\
 & + \frac{3}{2}\hat{\mathbf{r}}_s(t_0)(\hat{\mathbf{r}}_s(t_0) \cdot \bar{\mathbf{r}}_\mu)^2 + \left[\Delta\tilde{\mathbf{r}}_{s/G} - \tilde{\mathbf{r}}_{sc} - \Delta\tilde{\mathbf{r}}_{s/G}(\hat{\mathbf{r}}_s(t_0) \cdot \bar{\mathbf{r}}_\mu) + \tilde{\mathbf{r}}_{sc}(\hat{\mathbf{r}}_s(t_0) \cdot \bar{\mathbf{r}}_\mu) \right. \\
 & - \hat{\mathbf{r}}_s(t_0)(\hat{\mathbf{r}}_s(t_0) \cdot \Delta\tilde{\mathbf{r}}_{s/G}) - \bar{\mathbf{r}}_\mu(\hat{\mathbf{r}}_s(t_0) \cdot \Delta\tilde{\mathbf{r}}_{s/G}) + \hat{\mathbf{r}}_s(t_0)(\hat{\mathbf{r}}_s(t_0) \cdot \tilde{\mathbf{r}}_{sc}) \\
 & + 3\hat{\mathbf{r}}_s(t_0)(\hat{\mathbf{r}}_s(t_0) \cdot \bar{\mathbf{r}}_\mu)(\hat{\mathbf{r}}_s(t_0) \cdot \Delta\tilde{\mathbf{r}}_{s/G}) + \bar{\mathbf{r}}_\mu(\hat{\mathbf{r}}_s(t_0) \cdot \tilde{\mathbf{r}}_{sc}) \\
 & \left. - 3\hat{\mathbf{r}}_s(t_0)(\hat{\mathbf{r}}_s(t_0) \cdot \bar{\mathbf{r}}_\mu)(\hat{\mathbf{r}}_s(t_0) \cdot \tilde{\mathbf{r}}_{sc}) - \hat{\mathbf{r}}_s(t_0)(\bar{\mathbf{r}}_\mu \cdot \Delta\tilde{\mathbf{r}}_{s/G}) + \hat{\mathbf{r}}_s(t_0)(\bar{\mathbf{r}}_\mu \cdot \tilde{\mathbf{r}}_{sc}) \right] \varpi \\
 & + \left[\tilde{\mathbf{r}}_{sc}(\hat{\mathbf{r}}_s(t_0) \cdot \Delta\tilde{\mathbf{r}}_{s/G}) + \Delta\tilde{\mathbf{r}}_{s/G}(\hat{\mathbf{r}}_s(t_0) \cdot \tilde{\mathbf{r}}_{sc}) - \tilde{\mathbf{r}}_{sc}(\hat{\mathbf{r}}_s(t_0) \cdot \tilde{\mathbf{r}}_{sc}) \right. \\
 & - 3\hat{\mathbf{r}}_s(t_0)(\hat{\mathbf{r}}_s(t_0) \cdot \Delta\tilde{\mathbf{r}}_{s/G})(\hat{\mathbf{r}}_s(t_0) \cdot \tilde{\mathbf{r}}_{sc}) + \hat{\mathbf{r}}_s(t_0)(\Delta\tilde{\mathbf{r}}_{s/G} \cdot \tilde{\mathbf{r}}_{sc}) \\
 & \left. - \frac{1}{2}\hat{\mathbf{r}}_s(t_0)(\tilde{\mathbf{r}}_{sc} \cdot \tilde{\mathbf{r}}_{sc}) + \frac{3}{2}\hat{\mathbf{r}}_s(t_0)(\hat{\mathbf{r}}_s(t_0) \cdot \tilde{\mathbf{r}}_{sc})^2 \right] \varpi^2
 \end{aligned} \tag{5.3.16}$$

where the subscript s may now refer to either the target star, or any of its reference sources.

We now have a completely general second order expansion of the spacecraft to source unit vector, which applies when the source is different from the target star used to define the tangent frame, or when the spacecraft baseline pointing does not coincide with the vectors of the tangent frame. In the first case, we retain terms proportional to $\hat{\mathbf{r}}_s(t_0) \cdot \mathbf{b}_i$ because $\hat{\mathbf{r}}_s(t_0) \neq \mathbf{b}_3$ and so these terms do not automatically go to zero. In the second case, these terms do not go to zero because $\mathbf{b}_i \neq \mathbf{b}_1$ or \mathbf{b}_2 . This expression is even more complex than Equation (5.3.15), and even less suitable for use in analysis. However, this added complexity is actually mitigated by the same reasons that caused us to add in the extra terms. First, if the pointing error is small (i.e., the baseline is reasonably close to one of the two tangent frame unit vectors), then only the new terms in $\hat{\mathbf{r}}_s(t_0)$ whose coefficients are of order 1 should be kept, as $\hat{\mathbf{r}}_s(t_0) \cdot \mathbf{b}_i$ will still represent a small value. Second, even non-ideal reference sources will still usually be stars which are significantly further from the solar system than the target star, with much smaller parallax and barycenter motion values. As such, the first order form will typically be sufficiently precise for these sources, allowing most or all of the terms in ϖ^2 to be dropped.

Another way to simplify these equations is to note that we have so far been assuming no prior knowledge of the astrometric quantities $(\varpi, \bar{\mathbf{r}}_\mu)$ being measured. In reality we already have some knowledge of the barycenter motion and parallax of our target stars, and high precision astrometric observations will be used to update, rather than replace existing measurements. To represent this, we can rewrite Equation (5.3.13) with $\bar{\mathbf{r}}_\mu$ replaced by $\bar{\mathbf{r}}_{\mu 0} + \delta \bar{\mathbf{r}}_\mu$ and ϖ replaced by $\varpi_0 + \delta \varpi$, where $\bar{\mathbf{r}}_{\mu 0}$ and ϖ_0 are the currently known components of barycenter motion and parallax and $\delta \bar{\mathbf{r}}_\mu$ and $\delta \varpi$ are the (relatively small) correction terms.

The result is:

$$\begin{aligned}
\hat{\mathbf{r}}_{s/sc} = & (\hat{\mathbf{r}}_s(t_0) + \bar{\mathbf{r}}_{\mu 0} + \delta \bar{\mathbf{r}}_\mu + (\varpi_0 + \delta \varpi) \Delta \tilde{\mathbf{r}}_{s/G} - (\varpi_0 + \delta \varpi) \tilde{\mathbf{r}}_{sc}) \times \\
& \left\{ 1 + \bar{\mathbf{r}}_{\mu 0} \cdot \bar{\mathbf{r}}_{\mu 0} + \delta \bar{\mathbf{r}}_\mu \cdot \delta \bar{\mathbf{r}}_\mu + 2 \bar{\mathbf{r}}_{\mu 0} \cdot \delta \bar{\mathbf{r}}_\mu + (\varpi_0 + \delta \varpi)^2 \Delta \tilde{\mathbf{r}}_{s/G} \cdot \Delta \tilde{\mathbf{r}}_{s/G} \right. \\
& + (\varpi_0 + \delta \varpi)^2 \tilde{\mathbf{r}}_{sc} \cdot \tilde{\mathbf{r}}_{sc} + 2 \hat{\mathbf{r}}_s(t_0) \cdot \bar{\mathbf{r}}_{\mu 0} + 2 \hat{\mathbf{r}}_s(t_0) \cdot \delta \bar{\mathbf{r}}_\mu \\
& + 2(\varpi_0 + \delta \varpi) \hat{\mathbf{r}}_s(t_0) \cdot \Delta \tilde{\mathbf{r}}_{s/G} - 2(\varpi_0 + \delta \varpi) \hat{\mathbf{r}}_s(t_0) \cdot \tilde{\mathbf{r}}_{sc} \\
& + 2(\varpi_0 + \delta \varpi) (\bar{\mathbf{r}}_{\mu 0} \cdot \Delta \tilde{\mathbf{r}}_{s/G} + \delta \bar{\mathbf{r}}_\mu \cdot \Delta \tilde{\mathbf{r}}_{s/G}) \\
& \left. - 2(\varpi_0 + \delta \varpi) (\bar{\mathbf{r}}_{\mu 0} \cdot \tilde{\mathbf{r}}_{sc} + \delta \bar{\mathbf{r}}_\mu \cdot \tilde{\mathbf{r}}_{sc}) - 2(\varpi_0 + \delta \varpi)^2 \Delta \tilde{\mathbf{r}}_{s/G} \cdot \tilde{\mathbf{r}}_{sc} \right\}^{-\frac{1}{2}}
\end{aligned} \tag{5.3.17}$$

In order to carry out expansions as before, it is necessary to quantify the expected values of these new variables so that we can decide which terms are negligible and which are important. Since we are assuming that our a priori estimates will be reasonably good, $\bar{\mathbf{r}}_{\mu 0}$ and ϖ_0 should be of the same order as \mathbf{r}_μ and ϖ above. Currently, parallax and proper motion are known to at least 1 mas and 1 mas/year accuracy, respectively [137]. We will take $\delta \varpi$ to be of order 5×10^{-9} ; $\|\delta \bar{\mathbf{r}}_\mu\|$ will vary in magnitude in time, but could be of order 1×10^{-8} within five years of observations.

We again perform a binomial expansion on the denominator of equation (5.3.17) and retain terms explicitly in ϖ_0 to second order and explicitly in $\delta \varpi$ to first order. As before, terms proportional to $\|\bar{\mathbf{r}}_{\mu 0}\|^n$ for $n > 2$ and terms proportional to $\|\bar{\mathbf{r}}_{\mu 0}\|^n \varpi_0$ for $n > 1$ are dropped, as are terms proportional to $\|\delta \bar{\mathbf{r}}_\mu\|^n$ for $n > 1$, terms proportional to $\|\delta \bar{\mathbf{r}}_\mu\| \delta \varpi$, and terms including ϖ^2 that are proportional to $\|\Delta \tilde{\mathbf{r}}_{s/G}\|^n$ for $n > 1$.

Terms proportional to $\|\bar{\mathbf{r}}_{\mu 0}\| \delta \varpi$, $\|\bar{\mathbf{r}}_{\mu 0}\| \|\delta \bar{\mathbf{r}}_\mu\|$, and $\|\delta \bar{\mathbf{r}}_\mu\| \varpi_0$ are borderline—for the values assumed here, they have magnitudes of up to 1×10^{-14} , which means they should be neglected, but if the proper motions are larger than expected (or there is higher uncertainty in the a priori measurements), then these terms can become significant, so we will leave them. Similarly, terms proportional to $\varpi_0 \delta \varpi$ have magnitudes of 1×10^{-15} for the values assumed here, and are fairly unlikely to become significant unless uncertainties in parallax are much

greater than expected for the nearest stars. We include only one of these terms because leaving it out leaves a clear sinusoidal pattern in the signal of order 5×10^{-15} . If the chosen analysis method is not expected to be sensitive to such a signal, this term can safely be omitted. The resulting approximation is:

$$\begin{aligned}
\hat{\mathbf{r}}_{s/sc} \cdot \mathbf{b}_i &\approx \{ \bar{\mathbf{r}}_{\mu 0} + \delta \bar{\mathbf{r}}_{\mu} - \bar{\mathbf{r}}_{\mu 0}(\hat{\mathbf{r}}_s(t_0) \cdot \bar{\mathbf{r}}_{\mu 0}) - \delta \bar{\mathbf{r}}_{\mu}(\hat{\mathbf{r}}_s(t_0) \cdot \bar{\mathbf{r}}_{\mu 0}) \\
&\quad - \bar{\mathbf{r}}_{\mu 0}(\hat{\mathbf{r}}_s(t_0) \cdot \delta \bar{\mathbf{r}}_{\mu}) + [2\Delta \tilde{\mathbf{r}}_{s/G} - \tilde{\mathbf{r}}_{sc} - \Delta \tilde{\mathbf{r}}_{s/G}(\hat{\mathbf{r}}_s(t_0) \cdot \bar{\mathbf{r}}_{\mu 0}) \\
&\quad + \tilde{\mathbf{r}}_{sc}(\hat{\mathbf{r}}_s(t_0) \cdot \bar{\mathbf{r}}_{\mu 0}) + \bar{\mathbf{r}}_{\mu 0}(\hat{\mathbf{r}}_s(t_0) \cdot \tilde{\mathbf{r}}_{sc}) 2] (\delta \varpi + \varpi_0) \\
&\quad - \bar{\mathbf{r}}_{\mu 0}(\hat{\mathbf{r}}_s(t_0) \cdot \Delta \tilde{\mathbf{r}}_{s/G}) \delta \varpi + [\tilde{\mathbf{r}}_{sc}(\hat{\mathbf{r}}_s(t_0) \cdot \delta \bar{\mathbf{r}}_{\mu}) + \delta \bar{\mathbf{r}}_{\mu}(\hat{\mathbf{r}}_s(t_0) \cdot \tilde{\mathbf{r}}_{sc})] \varpi_0 \\
&\quad - 2\tilde{\mathbf{r}}_{sc}(\hat{\mathbf{r}}_s(t_0) \cdot \tilde{\mathbf{r}}_{sc}) \varpi_0 \delta \varpi + [\tilde{\mathbf{r}}_{sc}(\hat{\mathbf{r}}_s(t_0) \cdot \Delta \tilde{\mathbf{r}}_{s/G}) \\
&\quad + \Delta \tilde{\mathbf{r}}_{s/G}(\hat{\mathbf{r}}_s(t_0) \cdot \tilde{\mathbf{r}}_{sc}) - \tilde{\mathbf{r}}_{sc}(\hat{\mathbf{r}}_s(t_0) \cdot \tilde{\mathbf{r}}_{sc})] \varpi_0^2 \} \cdot \mathbf{b}_i
\end{aligned} \tag{5.3.18}$$

The difference between this equation and the OPD using the exact expression for $\hat{\mathbf{r}}_{s/sc}$ is almost exactly the same as for the second order expansion assuming no prior knowledge. However, unlike Equation (5.3.15), this equation is linear in the quantities to be estimated, thereby simplifying the filtering task. In fact, the difference between differential OPDs calculated using Equations (5.3.15) and (5.3.18) is actually at the level of the numerical precision of the data type used in these simulations, which is what produces the characteristic residual pattern in Figure 5.6. This figure highlights a very important point: due to the small magnitudes of many of the quantities utilized in this analysis the limitations of the specific numerical representation become a real concern when constructing simulations. Any computer generated value will have an associated finite precision, but specific representation schemas may actually introduce precision differences as functions of value magnitude. For example, in MATLAB, the default numerical data type is a double precision, floating point number, encoded as:

$$2^e \times s \tag{5.3.19}$$

where e is the exponent and s is the significand. These are stored as 11 bits and 53 bits,

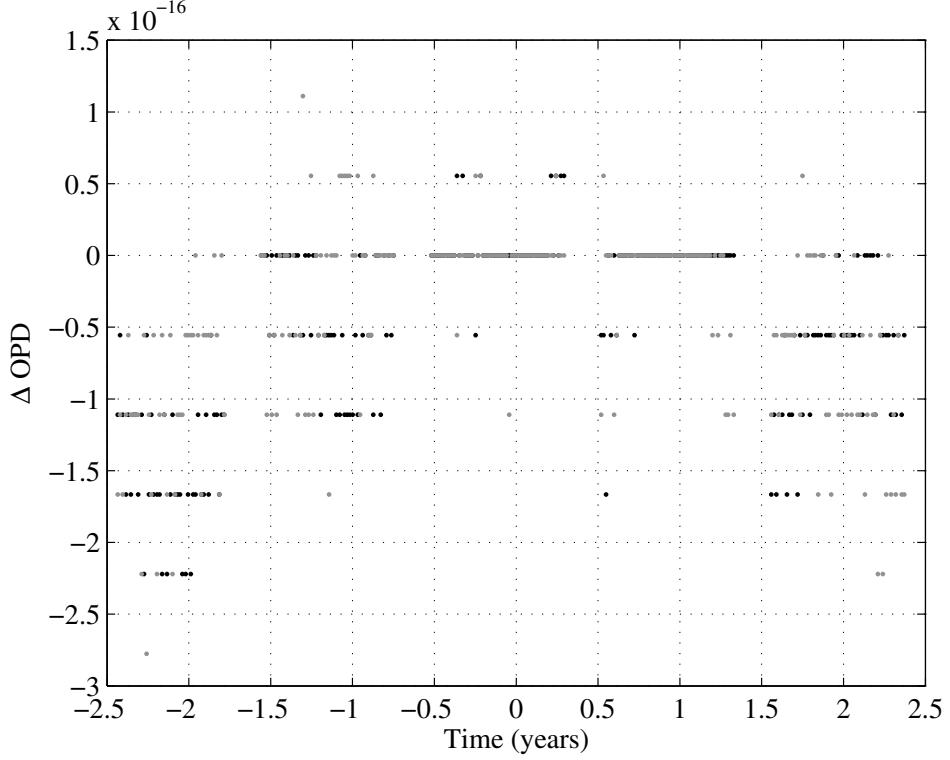


Figure 5.6: Difference between differential OPDs (between a sun-twin star and fixed centroid) using the second order expansion of $\hat{\mathbf{r}}_{s/sc}$ from Equation (5.3.15) and the expansion from Equation (5.3.18) with assumed errors in parallax and barycenter motion of 1 mas and 1 mas/year, respectively. The pattern in the residuals is due to the precision of the numerical data type used. The two curves represent the measurements along two orthogonal interferometer baseline orientations.

respectively, as per the IEEE 754 standard [61]. Because the exponent and significand are encoded with a finite number of bits, the minimum step size to the next available value in this schema is a function of the current value. Thus, differences between pairs of values of different orders will have different precisions [91]. For example, when encoding values of order 1 as double precision floats, the minimum step size to the next available value is on the order of 1×10^{-16} . However, when encoding values of order 1e6 (like, for example, the value of 10 pc in AU), the step size to the next available value is of order 1×10^{-10} .

The particulars of the encoding become important when one wants to generate ‘ground truth’ data, using, for example, the exact representation of the astrometric observation in Equation (2.2.12). In this case, certain values, such as $\mathbf{r}_s(t_0)$ are multiple orders of magnitude

greater than 1, whereas the final result is many orders of magnitude less. This means that numerical noise at the level of the desired signal could inadvertently be introduced into the simulated data. In the simulations presented here, this only became significant when converting $\mathbf{r}_{s/sc}$ to a unit vector. One solution (the one employed here) is to generate the large values with an arbitrary precision data type using the appropriate number of digits. Once the spacecraft to star unit vector is generated, all values are of sufficiently low order that the default fixed precision data type is sufficient for the remaining calculations. To do so, it is necessary to use an arbitrary-precision data type, which simply encodes values using variable length significands, rather than the fixed bit scheme described above. For the simulations here, sufficient precision was achieved by using the GNU multiple precision library (<http://gmplib.org/>), with the significand stored as 256 bits. The numerical effect is quite limited in the simulations above (with the target star at 10 pc), but becomes much more pronounced for targets at 100 pc. Figure 5.7 shows the difference between differential OPDs using Equation (2.2.12), with one simulation done using the default double precision data type for all values, and the other using the multiple precision data type. The error between the two is close to 1×10^{-8} in magnitude, making it a significant noise contribution.

An alternative to this approach would be to use Equation (5.3.13). All terms in this formulation (which is equivalent to Equation (2.2.12)) are of order 1 or less, so that the default double precision data type should be sufficient. Unfortunately, use of this equation can also introduce numerical noise, if one is not careful in how the various terms are defined. Terms such as \mathbf{r}_μ are essentially small values divided by very large values. If they are not exactly defined, but rather calculated in software, then they will have the same problems as shown above. Since this discussion applies only to the simulation or generated data, it is relatively simple to just assume values for barycenter motions that will give exact fractional values when scaled by the parallax, eliminating such concerns. If we wished to update our expansions such that the residual was below the precision of the default double precision

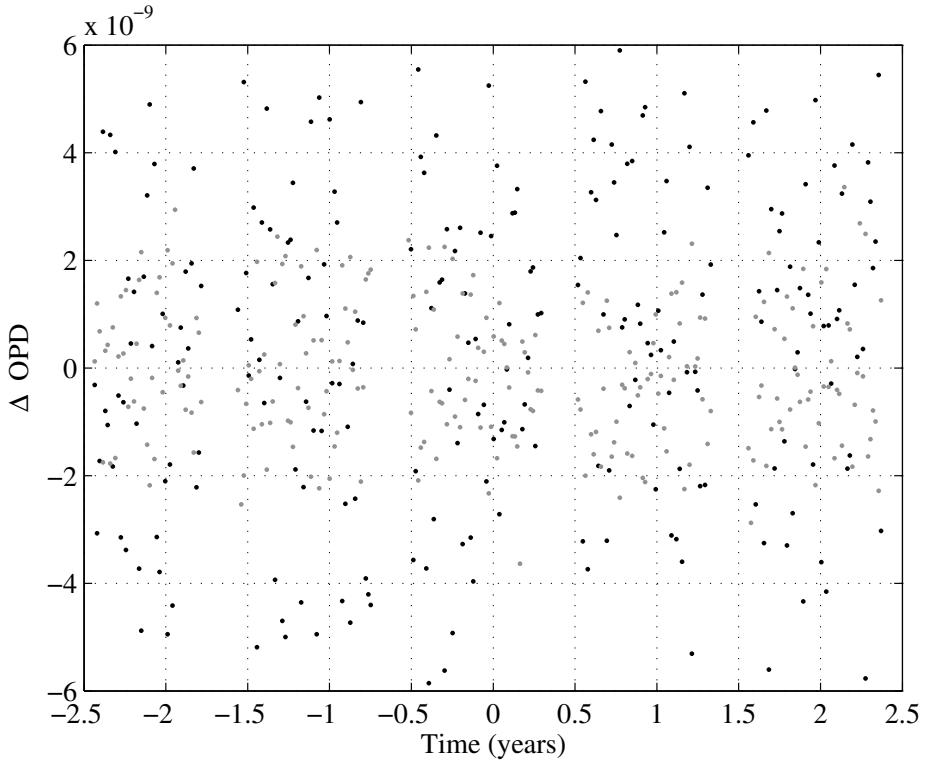


Figure 5.7: Difference between differential OPDs (between a sun-twin star and fixed centroid) using the exact formulation of $\hat{\mathbf{r}}_{s/sc}$ from Equation (2.2.12) with the default double precision data type and a multiple precision data type at 256 bits. All values are the same as in previous simulations, except that the target star is placed at 100 pc. The two curves represent the measurements along two orthogonal interferometer baseline orientations.

data type, we would need to retain terms proportional to ϖ^3 . However, in the case of the expansions assuming prior knowledge of parallax and barycenter motion, no higher order terms in $\delta\varpi$ would be needed, as seen in Equation (5.3.20).

One assumption made in this analysis is that the error is of the same magnitude for all three components of $\bar{\mathbf{r}}_\mu$. However, this ignores the fact that currently existing estimates for the first two components (the transverse motion) are derived from different sources than estimates for the third component (the radial motion). While proper motions have been measured by space-based astrometric instruments, radial velocities are generally provided by ground-based doppler spectroscopy, and are inherently less accurate. To address this, we

can modify Equation (5.3.18) by separating $\delta\bar{\mathbf{r}}_\mu$ into two separate errors so that:

$$\bar{\mathbf{r}}_\mu = \bar{\mathbf{r}}_{\mu 0} + \delta\mathbf{r}_R + \delta\mathbf{r}_T \quad (5.3.21)$$

where $\delta\mathbf{r}_R$ and $\delta\mathbf{r}_T$ are due to errors in our prior knowledge of V_R and \mathbf{V}_T , as defined in Equation (2.2.4).

We can now repeat the expansion performed to generate Equation (5.3.18)), substituting the right-hand side of Equation (5.3.21) for $\bar{\mathbf{r}}_\mu$. The rules for keeping terms containing $\delta\mathbf{r}_T$ remain the same as those for $\delta\bar{\mathbf{r}}_\mu$. If we assume that $\delta\mathbf{r}_R$ is less than three orders of magnitude higher than $\delta\mathbf{r}_T$ (i.e., m/s radial velocity precision), then the expansion in Equation (5.3.18) must also retain terms proportional to $\|\delta\mathbf{r}_R\|^n$ for $n <= 2$. The result is that only one term, equal to $-\delta\mathbf{r}_R(\hat{\mathbf{r}}_s(t_0) \cdot \delta\mathbf{r}_R) = -\delta\mathbf{r}_R\|\delta\mathbf{r}_R\|$, must be added to Equation (5.3.18) to account for the lower precision of the prior radial velocity measurement.

5.4 Applications

In this section we will explore some applications of the formalism built up throughout the chapter. In particular, we will discuss how to construct a filter to combine multiple observations of one planetary system, and examine whether the filtering framework can help to resolve confusion in the imaging of multi-planet systems. Because of the inherent synergies between doppler spectroscopy and astrometry (one essentially measures the projected component least observable by the other), there has always been significant interest in combining the two [128]. One commonly proposed method of combining these data sets (or just the two signal channels produced by astrometry) is the formulation of a joint periodogram [19], which is just a summation of independent Lomb-Scargle periodograms. This approach has the drawback of ignoring much of the known dynamics of orbital motion. In particular, as both eccentricity and projection effects cause deviations from pure harmonic motion in the observations, they are only separable by performing full orbital fits to the significant

periodicities found by the periodogram, which themselves will deviate from the true periods due to these same effects.

When there are significant n-body effects, the periodogram has the effect of averaging over any variations in orbital periods, which can lead to Keplerian fits that do not actually correspond to the best-fit orbit at any specific time in the data set. We can demonstrate this by simulating a system with two Jupiter-mass planets with initial conditions corresponding to a 1:3 mean motion resonance. As the initial conditions do not describe a stable resonance, the orbits do not maintain an exact 1:3 period ratio, but do exhibit significant deviations from the equivalent mean Keplerian fits, as demonstrated in Figure 5.8. Without any added noise, the Lomb-Scargle periodogram has no trouble pulling out two significant periodicities (Figure 5.9(a)) corresponding to the two planets using just one component of the star's velocity. However, if we compare the periods derived from the periodogram with the osculating periods, we immediately see the averaging effect described above (Figure 5.9(b)).

What's particularly interesting about this system is that the osculating arguments or periapsis for both planets remain very stable in time (within 5% of the initial value for the inner planet and within 7% of the initial value for the outer planet), and the rest of the orbital elements osculate on time frames longer than one orbit. This indicates that the instantaneous Keplerian fits do describe meaningful orbits (unlike the case of true resonances where elements osculate on time scales faster than a single orbital period), and information is being discarded by averaging over the entire data set. In the next section we will see how the filtering formalism allows us to fit the dynamic orbits, rather than average Keplerian elements.

5.4.1 Combined Astrometry and Doppler Spectroscopy

To test the filter's ability to fit orbits we use the same type of simulated data as shown in the previous section, except with added astrometric effects and instrument noise. We begin with a single planet of $1 M_{\oplus}$ on an Earth-like orbit (1 AU semi-major axis, but arbitrarily

oriented with respect to the instrument line of sight) and simulate the stellar reflex. To simulate the data set that would be produced by an astrometric space-based observatory, we take snapshots of the stellar reflex over a 15 year period, with the first 10 years composed solely of doppler spectroscopy data, and the final 5 years containing a mix of radial velocity and astrometric measurements. The doppler spectroscopy data is assumed to cover a longer period since it can be taken using a ground-based observatory, whereas the assumed precision of the astrometry requires a space-based instrument. The noise variance applied to the data has a magnitude of $0.82 \mu\text{as}$ for the astrometric data and 1 m/s for the doppler spectroscopy data, as shown in Figure 5.10.

Figure 5.11 shows the output of the hybrid, constrained EKF for the case of the single Earth-mass planet. The plots represent the estimates of the position and velocity (in AU and AU/day, respectively) plotted in components in the tangent frame. The time is in days, with the zero epoch coinciding with the first astrometric measurement (i.e., estimates at times $t < 0$ are based solely on radial velocity measurements). We can immediately see that the filter performs quite well, despite the very large noise level, except for expected errors in badly observed components. The state is initialized with the planet mass set to its correct value, but with randomized initial positions and velocities (consistent with the observations and a closed (negative Keplerian energy) orbit). The initial proper motion and parallax have errors of 1 mas/year and 1 mas, respectively. Although the system does not have true process noise, a diagonal \mathbf{Q} with values six orders of magnitude below the state estimate is introduced in order to keep the filter from falling into local minima. Initial states are also propagated via one UKF step to more closely model the state distribution, while initial covariances are produced via Monte Carlo simulation, as in Figure 5.1.

Figure 5.12 displays the same data as errors (true state minus the estimate), along with the variance estimated by the filter. It is quite encouraging to see that the filter is doing a credible job of estimating how well it is doing almost everywhere in the data set. There are, however, large spikes in the velocity error estimate after the astrometric data kicks in. These

are due to numerical effects in the current state prediction integration, and are removed in subsequent smoothing runs.

We can also study the effects of varying the initial mass estimate. Through a large number of simulations, it is found that the filter generally fails to converge for the entire data set when the assumed mass of the planet is more than 50% off from the true mass (in cases where convergence occurs, the described orbits are highly eccentric and do not correspond at all to the actual orbits). Unfortunately, for assumed masses within 50% of the true mass, the filter can actually converge to state estimates with covariances lower than that produced by the true mass. In cases where the mass is within 50% of the true mass, the constraints placed on the filter generally stop the mass estimate from diverging to any large extent from its initial value. Figure 5.13 shows the mean variances for position produced by the filter, as a function of error in the planetary mass.

These results are very promising, most especially because of how well the filter does at estimating its own error. The fact that the variance estimates bound the true error at all filter steps gives us great confidence in the filter's estimation of its performance. Furthermore, the filter does an excellent job of tracking the more observable states (the sky plane positions and radial velocities) and at least captures the proper periodicities in the remaining state. Unfortunately, the radial position estimate deviates at the end of the run, and the velocity states show signs of over-fitting to the observations. The fact that the initial mass estimate has to be within 50% of the true mass for the filter to perform well also indicates that this particular filter may be unsuitable for smaller planets with such low signal to noise values. It is possible that a large-scale particle filter more capable of modeling the state distribution accurately would perform better.

Turning to the multi-planet case, we repeat the simulation, now including the Earth-mass planet and a Jupiter mass planet on a Jupiter-like orbit. Figure 5.14 shows the filter estimate error for the Jupiter mass planet, which, as expected, is approximately a factor of 100 lower than that for the Earth-mass planet. Because it dominates the astrometric signal

to such a large extent, the filter has no problem estimating this planet’s state. Unfortunately, the Earth-mass signal becomes lost in this much larger one, and all states estimates have variances an order of magnitude higher than those produced in the case of the Earth-mass planet alone.

Due to the highly nonlinear observation functions and very high noise levels, the filter performance in terms of orbit fitting is relatively poor with this data set. However, several important things are demonstrated by these simulations. First, it is possible to swap out different observation functions at each time step without destabilizing the filter. Second, we have demonstrated the capability to treat mass as one of the state parameters, as long as the initial mass estimate is sufficiently close to the true mass. These estimates can be greatly constrained by using the periodogram analysis combined with the PDFs derived in §3.3. Finally, in cases where the filter does converge upon an orbit, it is tracking the actual positions and velocities of the planets, rather than averaged orbital elements, and can thus potentially tell us more about the exosystem than any Keplerian orbital fit. More work, however, is required on creating a filter capable of dealing with both the nonlinearities of the dynamics and observation functions as well as the high level of noise in the observations.

5.4.2 Filter Applications for Direct Detection

Following the simulation methodology adopted in the previous section, we can also consider what happens when long term doppler spectroscopy is followed by one single successful direct detection. In this case, we would like to also test whether imaging can help fit orbits in data sets where there is less than a whole orbital period of observations. To do so, we simulate 5 years (approximately half of one orbit) of doppler spectroscopy measurements of a sun mass star with a Jupiter mass planet, immediately followed by one image taken with an instrument identical to THEIA’s XPC, with an assumed centroiding accuracy of one tenth of one pixel.

Figure 5.15 shows the filter output when using only the doppler spectroscopy data set.

Regardless of the initial conditions and the number of smoothing steps, the filter is unable to track the true state values. It is interesting to note, however, that the instantaneous orbital elements produced by fitting Keplerian orbits to the filter state estimates do generate approximately correct orbital periods.

Figure 5.16 shows the results of filtering the same data set, but now adding one direct detection immediately after the last doppler spectroscopy measurement. The direct detection immediately transfers the position state elements much nearer to their true values (with a larger error on the third component due to the nonlinearity in that observation). The velocity states are still poorly estimated as they are not observed at all by the direct detection.

Finally, Figure 5.17 repeats the simulation but adds two additional smoothing and filtering runs (i.e., running the entire data set forwards and backwards in time three times). Both the positions and velocities are now much more closely constrained to the true states, although a negative consequence of the multiple smoothings is the over-fitting of the velocity state elements. Because of this, the instantaneous orbital elements generated from the filter outputs can have very high errors, although the average orbital elements are very close (within 5%) of their true values.

An alternate application for direct detection data returns to the question of possible confusion in multiple planet systems discussed in §4.4. We can now evaluate what happens when data from observations of two different planets is filtered with a state assuming that they are observations of the same planet at different times. Because stable orbits generally have to be sufficiently separated to remain stable, it is most likely that confusion will be due to projection effects from the inclination of the ecosystem. If we inadvertently create a data set composed of observations of two different planets, the filter will effectively be trying to estimate a state distribution composed of two separate distributions. While these are not fully independent (due to n-body effects and the requirements of system stability), they will tend to make the filter’s target distribution significantly bimodal. This bimodality should, given enough filtering steps, cause the state element variance to increase.

To test this application, we again simulate our two Jupiter system, generating one data set with five observations of the outer planet, and another data set with ten observations split evenly between the two planets. The system is highly inclined (70°) so that the apparent separations of the two planets can be very similar. Figure 5.18 shows the sum of the estimated variances (covariance diagonal elements) as a function of filter time step. As predicted, the inclusion of a mixed state does not initially have a great effect on the covariance estimate, but with a sufficient number of observations, the variance in the state distribution becomes much more apparent. These results indicate that the filtering framework can be a highly useful tool for validating data sets and picking out suspect observations.

The applications presented here represent just a tiny fraction of possible uses for the formalism developed in the first part of the chapter. What's most important is the fact that because the very essence of the dynamic filter is the possibility of time-varying observation functions, this is the ideal tool for putting together complicated data sets with multiple information sources. This framework also appears to be a good choice for tracking exosystem dynamics, although the inherent nonlinearities of these systems and observations, and the high measurement noises make this a very difficult task, and necessitate more work in filter development.

$$\begin{aligned}
\hat{\mathbf{r}}_{s/sc} \cdot \mathbf{b}_i &\approx \left\{ \bar{\mathbf{r}}_{\mu 0} + \delta \bar{\mathbf{r}}_{\mu} - \frac{1}{2} \bar{\mathbf{r}}_{\mu 0} (\bar{\mathbf{r}}_{\mu 0} \cdot \bar{\mathbf{r}}_{\mu 0}) - \bar{\mathbf{r}}_{\mu 0} (\hat{\mathbf{r}}_s(t_0) \cdot \bar{\mathbf{r}}_{\mu 0}) \right. \\
&\quad - \delta \bar{\mathbf{r}}_{\mu} (\hat{\mathbf{r}}_s(t_0) \cdot \bar{\mathbf{r}}_{\mu 0}) + \frac{3}{2} \bar{\mathbf{r}}_{\mu 0} (\hat{\mathbf{r}}_s(t_0) \cdot \bar{\mathbf{r}}_{\mu 0})^2 - \bar{\mathbf{r}}_{\mu 0} (\hat{\mathbf{r}}_s(t_0) \cdot \delta \bar{\mathbf{r}}_{\mu}) \\
&\quad + \left[\Delta \tilde{\mathbf{r}}_{s/G} - \tilde{\mathbf{r}}_{sc} + (\tilde{\mathbf{r}}_{sc} - \Delta \tilde{\mathbf{r}}_{s/G}) (\hat{\mathbf{r}}_s(t_0) \cdot \bar{\mathbf{r}}_{\mu 0}) \right. \\
&\quad \left. + \bar{\mathbf{r}}_{\mu 0} ((\hat{\mathbf{r}}_s(t_0) \cdot \tilde{\mathbf{r}}_{sc}) - (\hat{\mathbf{r}}_s(t_0) \cdot \Delta \tilde{\mathbf{r}}_{s/G})) \right] \delta \varpi \\
&\quad + 2 \left[(\tilde{\mathbf{r}}_{sc} - \Delta \tilde{\mathbf{r}}_{s/G}) (\hat{\mathbf{r}}_s(t_0) \cdot \Delta \tilde{\mathbf{r}}_{s/G}) \right. \\
&\quad \left. + (\Delta \tilde{\mathbf{r}}_{s/G} - \tilde{\mathbf{r}}_{sc}) (\hat{\mathbf{r}}_s(t_0) \cdot \tilde{\mathbf{r}}_{sc}) \right] \varpi_0 \delta \varpi \\
&\quad + \left[\Delta \tilde{\mathbf{r}}_{s/G} - \tilde{\mathbf{r}}_{sc} + \frac{1}{2} (\tilde{\mathbf{r}}_{sc} - \Delta \tilde{\mathbf{r}}_{s/G}) (\bar{\mathbf{r}}_{\mu 0} \cdot \bar{\mathbf{r}}_{\mu 0}) \right. \\
&\quad + \frac{3}{2} (\Delta \tilde{\mathbf{r}}_{s/G} - \tilde{\mathbf{r}}_{sc}) (\hat{\mathbf{r}}_s(t_0) \cdot \bar{\mathbf{r}}_{\mu 0})^2 + \bar{\mathbf{r}}_{\mu 0} ((\bar{\mathbf{r}}_{\mu 0} \cdot \tilde{\mathbf{r}}_{sc}) - (\bar{\mathbf{r}}_{\mu 0} \cdot \Delta \tilde{\mathbf{r}}_{s/G})) \\
&\quad + (\tilde{\mathbf{r}}_{sc} - \Delta \tilde{\mathbf{r}}_{s/G}) ((\hat{\mathbf{r}}_s(t_0) \cdot \bar{\mathbf{r}}_{\mu 0}) + (\hat{\mathbf{r}}_s(t_0) \cdot \delta \bar{\mathbf{r}}_{\mu})) \\
&\quad + (\bar{\mathbf{r}}_{\mu 0} + \delta \bar{\mathbf{r}}_{\mu}) ((\hat{\mathbf{r}}_s(t_0) \cdot \tilde{\mathbf{r}}_{sc}) - (\hat{\mathbf{r}}_s(t_0) \cdot \Delta \tilde{\mathbf{r}}_{s/G})) \\
&\quad \left. + 3 \bar{\mathbf{r}}_{\mu 0} (\hat{\mathbf{r}}_s(t_0) \cdot \bar{\mathbf{r}}_{\mu 0}) ((\hat{\mathbf{r}}_s(t_0) \cdot \Delta \tilde{\mathbf{r}}_{s/G}) - (\hat{\mathbf{r}}_s(t_0) \cdot \tilde{\mathbf{r}}_{sc})) \right] \varpi_0 \\
&\quad + \left[(\tilde{\mathbf{r}}_{sc} - \Delta \tilde{\mathbf{r}}_{s/G}) (\bar{\mathbf{r}}_{\mu 0} \cdot \Delta \tilde{\mathbf{r}}_{s/G}) + (\Delta \tilde{\mathbf{r}}_{s/G} - \tilde{\mathbf{r}}_{sc}) (\bar{\mathbf{r}}_{\mu 0} \cdot \tilde{\mathbf{r}}_{sc}) \right. \\
&\quad + \bar{\mathbf{r}}_{\mu 0} (\Delta \tilde{\mathbf{r}}_{s/G} \cdot \tilde{\mathbf{r}}_{sc}) - \frac{1}{2} \bar{\mathbf{r}}_{\mu 0} ((\Delta \tilde{\mathbf{r}}_{s/G} \cdot \Delta \tilde{\mathbf{r}}_{s/G}) + (\tilde{\mathbf{r}}_{sc} \cdot \tilde{\mathbf{r}}_{sc})) \\
&\quad + (\tilde{\mathbf{r}}_{sc} - \Delta \tilde{\mathbf{r}}_{s/G}) ((\hat{\mathbf{r}}_s(t_0) \cdot \Delta \tilde{\mathbf{r}}_{s/G}) - (\hat{\mathbf{r}}_s(t_0) \cdot \tilde{\mathbf{r}}_{sc})) \\
&\quad + 3 (\Delta \tilde{\mathbf{r}}_{s/G} (\hat{\mathbf{r}}_s(t_0) \cdot \bar{\mathbf{r}}_{\mu 0}) - \tilde{\mathbf{r}}_{sc} (\hat{\mathbf{r}}_s(t_0) \cdot \bar{\mathbf{r}}_{\mu 0})) (\hat{\mathbf{r}}_s(t_0) \cdot \Delta \tilde{\mathbf{r}}_{s/G}) \\
&\quad + 3 (\tilde{\mathbf{r}}_{sc} - \Delta \tilde{\mathbf{r}}_{s/G}) (\hat{\mathbf{r}}_s(t_0) \cdot \bar{\mathbf{r}}_{\mu 0}) (\hat{\mathbf{r}}_s(t_0) \cdot \tilde{\mathbf{r}}_{sc}) \\
&\quad - 3 \bar{\mathbf{r}}_{\mu 0} (\hat{\mathbf{r}}_s(t_0) \cdot \Delta \tilde{\mathbf{r}}_{s/G}) (\hat{\mathbf{r}}_s(t_0) \cdot \tilde{\mathbf{r}}_{sc}) \\
&\quad \left. + \frac{3}{2} \bar{\mathbf{r}}_{\mu 0} ((\hat{\mathbf{r}}_s(t_0) \cdot \Delta \tilde{\mathbf{r}}_{s/G})^2 + (\hat{\mathbf{r}}_s(t_0) \cdot \tilde{\mathbf{r}}_{sc})^2) \right] \varpi_0^2 \\
&\quad + \left[(\Delta \tilde{\mathbf{r}}_{s/G} - \tilde{\mathbf{r}}_{sc}) (\Delta \tilde{\mathbf{r}}_{s/G} \cdot \tilde{\mathbf{r}}_{sc}) - \frac{1}{2} (\Delta \tilde{\mathbf{r}}_{s/G} + \tilde{\mathbf{r}}_{sc}) (\Delta \tilde{\mathbf{r}}_{s/G} \cdot \Delta \tilde{\mathbf{r}}_{s/G}) \right. \\
&\quad + \frac{1}{2} (\tilde{\mathbf{r}}_{sc} - \Delta \tilde{\mathbf{r}}_{s/G}) (\tilde{\mathbf{r}}_{sc} \cdot \tilde{\mathbf{r}}_{sc}) + \frac{3}{2} (\Delta \tilde{\mathbf{r}}_{s/G} - \tilde{\mathbf{r}}_{sc}) (\hat{\mathbf{r}}_s(t_0) \cdot \Delta \tilde{\mathbf{r}}_{s/G})^2 \\
&\quad + 3 (\tilde{\mathbf{r}}_{sc} - \Delta \tilde{\mathbf{r}}_{s/G}) (\hat{\mathbf{r}}_s(t_0) \cdot \Delta \tilde{\mathbf{r}}_{s/G}) (\hat{\mathbf{r}}_s(t_0) \cdot \tilde{\mathbf{r}}_{sc}) \\
&\quad \left. + \frac{3}{2} (\Delta \tilde{\mathbf{r}}_{s/G} - \tilde{\mathbf{r}}_{sc}) (\hat{\mathbf{r}}_s(t_0) \cdot \tilde{\mathbf{r}}_{sc})^2 \right] \varpi_0^3 \Big\} \cdot \mathbf{b}_i
\end{aligned} \tag{5.3.20}$$

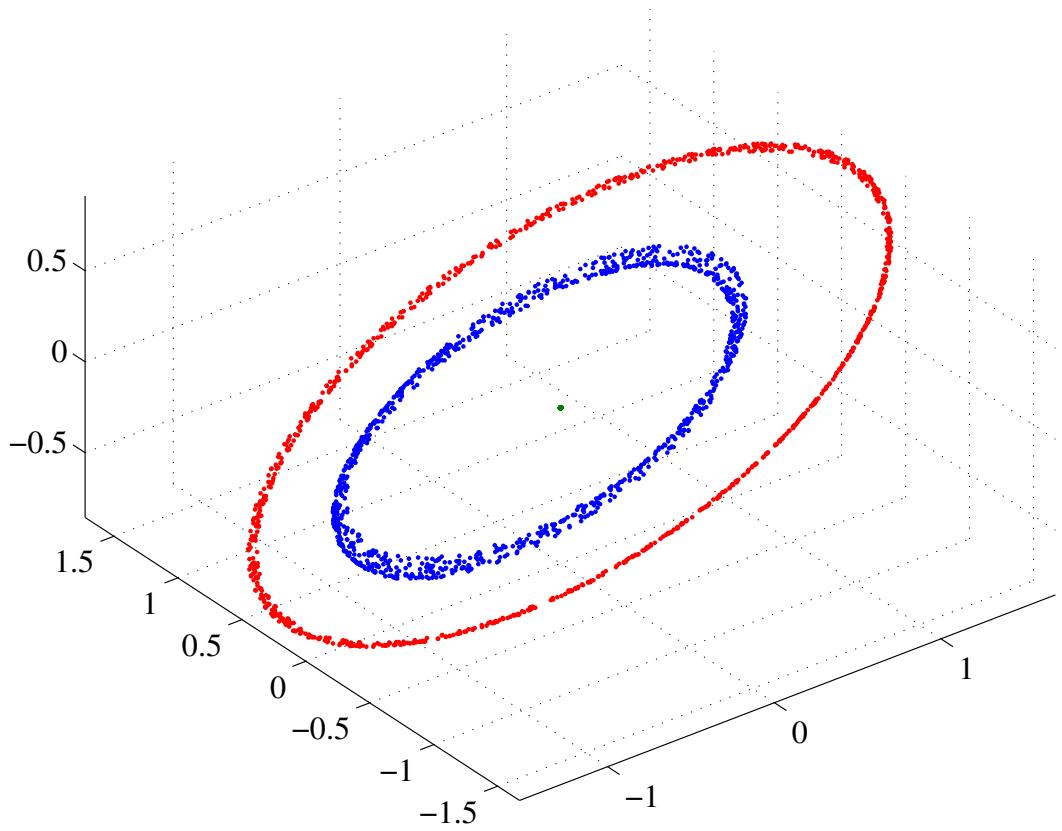


Figure 5.8: Integration of a system composed of a sun-twin and two Jupiters for 75 orbits of the innermost planet.

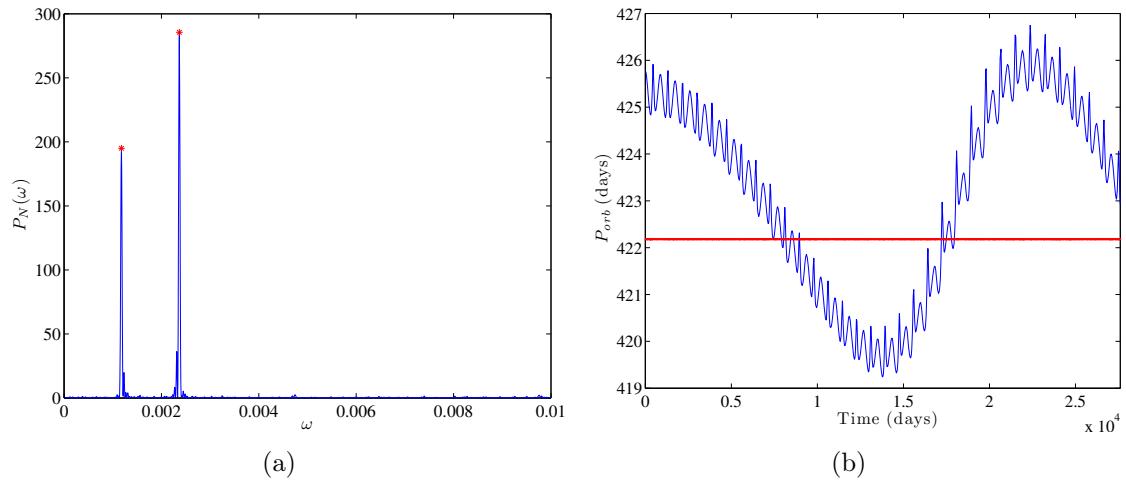


Figure 5.9: (a) Lomb-Scargle periodogram of \hat{z} -axis component of star velocity for system from Figure 5.8. The red ticks represent the significant frequencies. (b) The orbital period of the Keplerian orbital fit for every time step for the outer planet (blue) and the period from the periodogram (red).

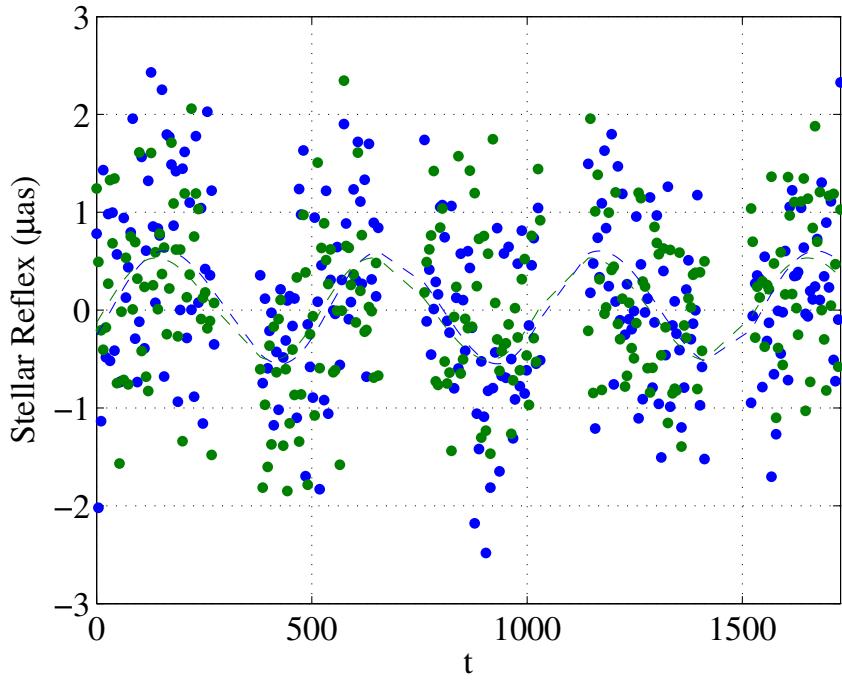


Figure 5.10: Two channels (orthogonal instrument baselines) of a simulated astrometric observation of an Earth-mass planet. The dashed lines represent the true signal, whereas the points are the noisy ($0.82 \mu\text{as}$ variance) measurement.

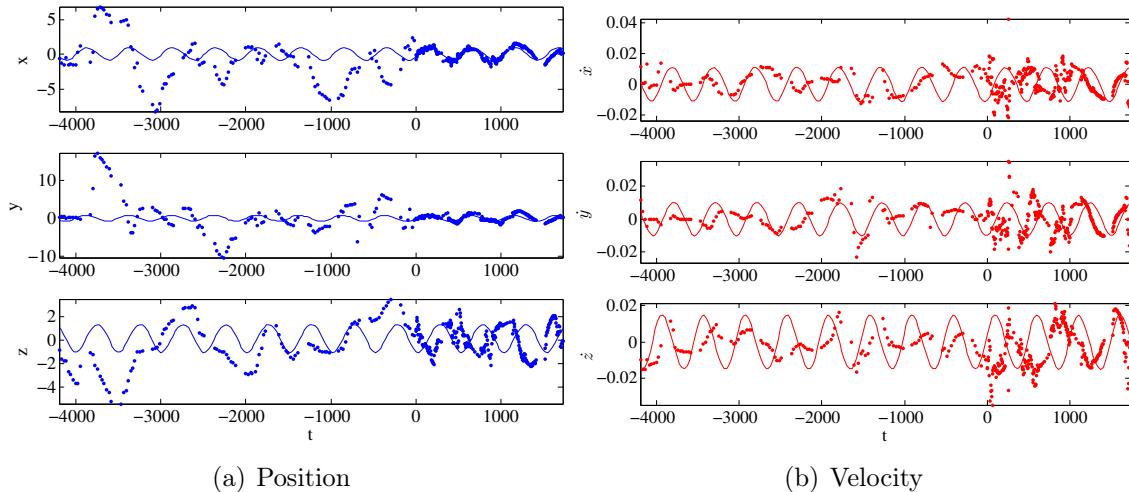


Figure 5.11: Filter outputs of planet position and velocity, plotted as points over solid lines representing the true positions for a single Earth-mass planet system.

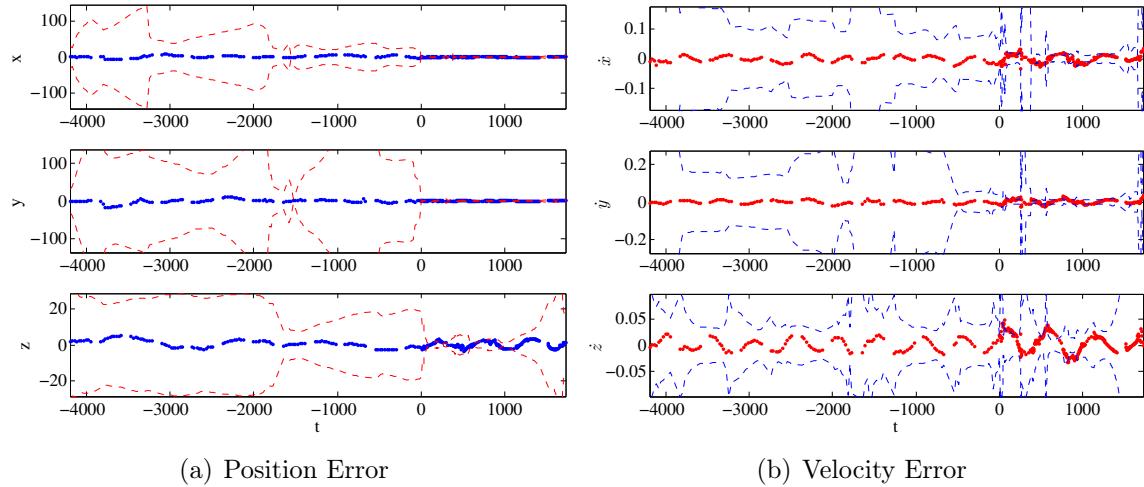


Figure 5.12: Error of state estimate in position and velocity plotted as points with the filter's variance estimate plotted as dashed lines for data in Figure 5.11.

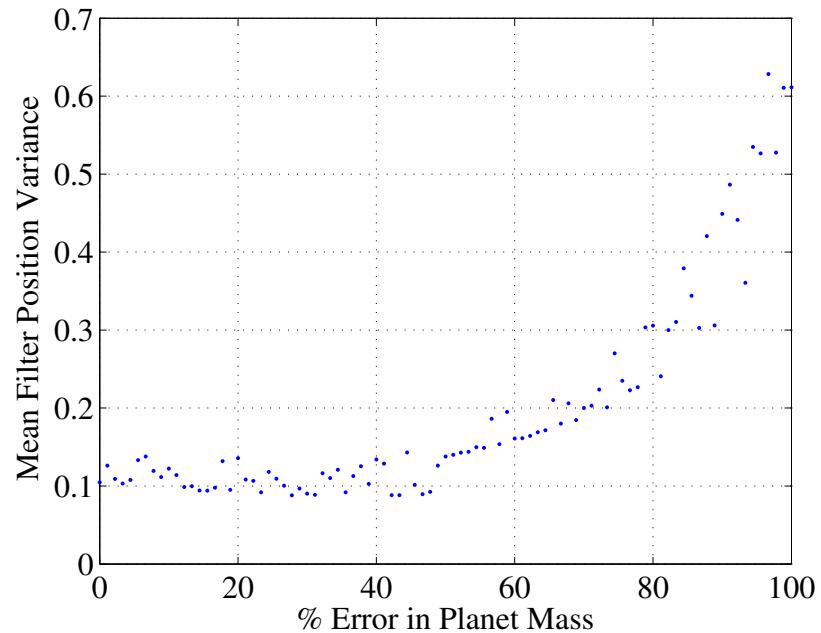


Figure 5.13: Mean filter variance estimate plotted as a function of error in initial planet mass.

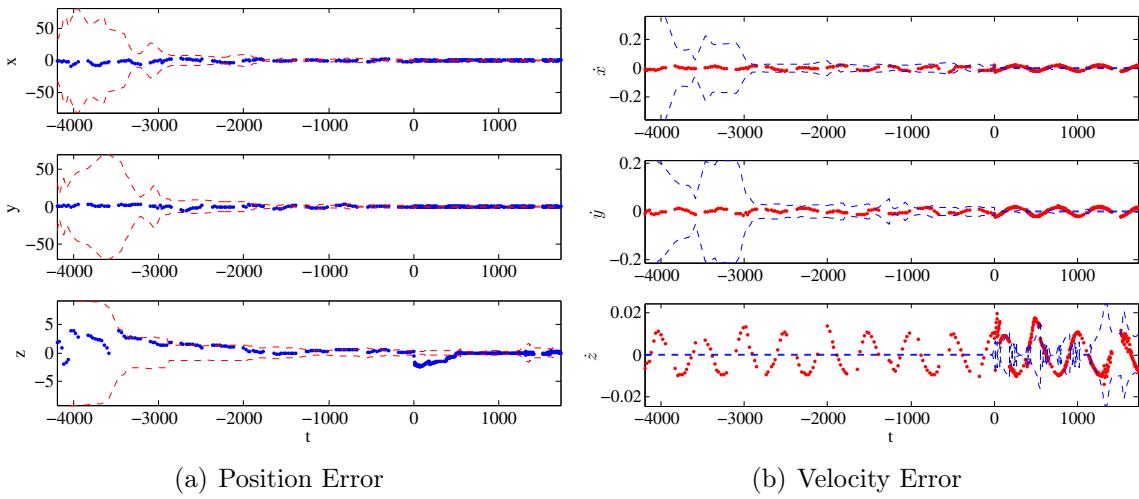


Figure 5.14: Error of state estimate in position and velocity plotted as points with the filter's variance estimate plotted as dashed lines for a Jupiter mass planet in multi-body system.

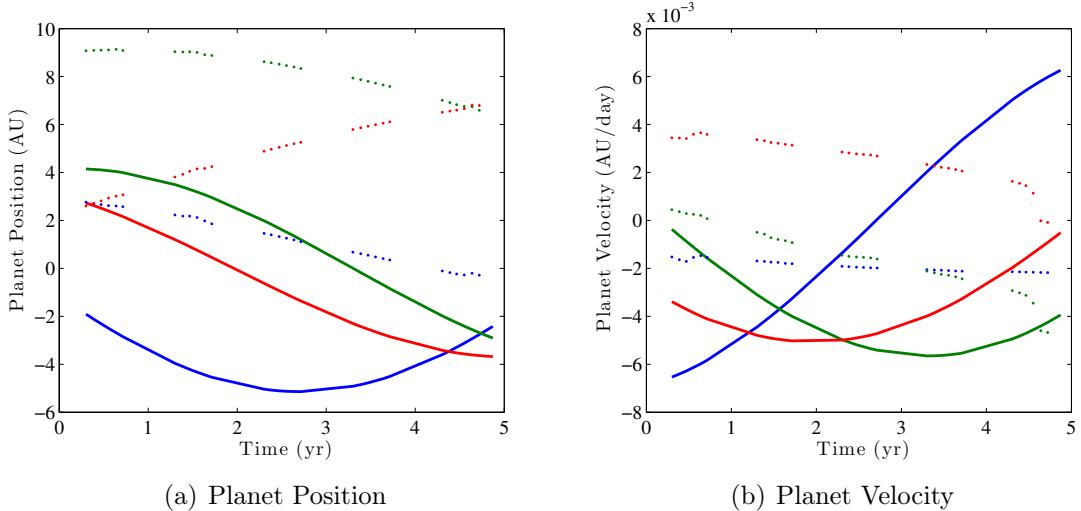


Figure 5.15: Planet position and velocity components (solid lines) and filter output (points) for half of one period of doppler spectroscopy observations of a Jupiter mass planet. Even with smoothing and multiple random re-initializations, the filter is unable to track the true state even approximately.

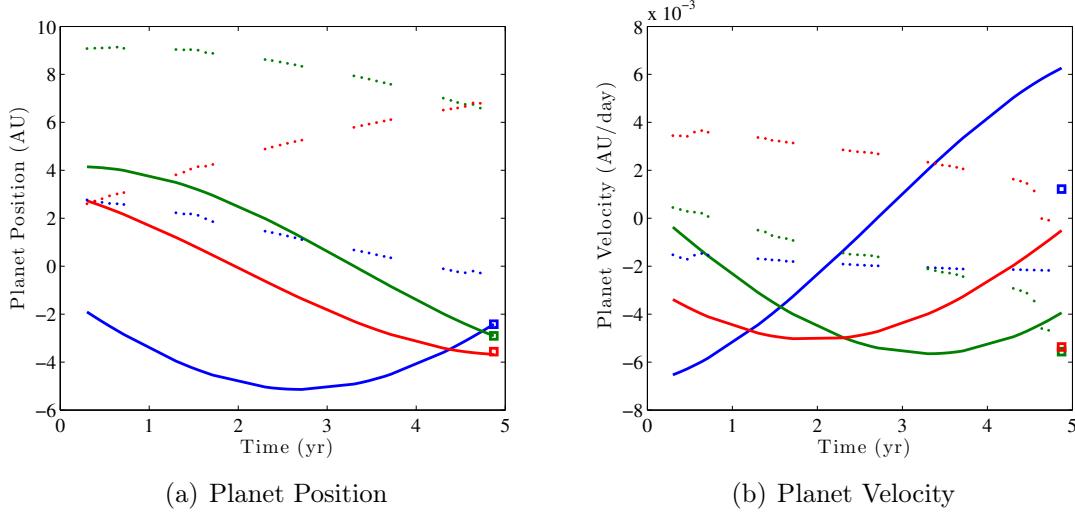


Figure 5.16: Planet position and velocity components (solid lines) and filter output (points) for half of one period of doppler spectroscopy observations of a Jupiter mass planet and one direct detection. The square marks represent the filter state after the inclusion of the direct detection.

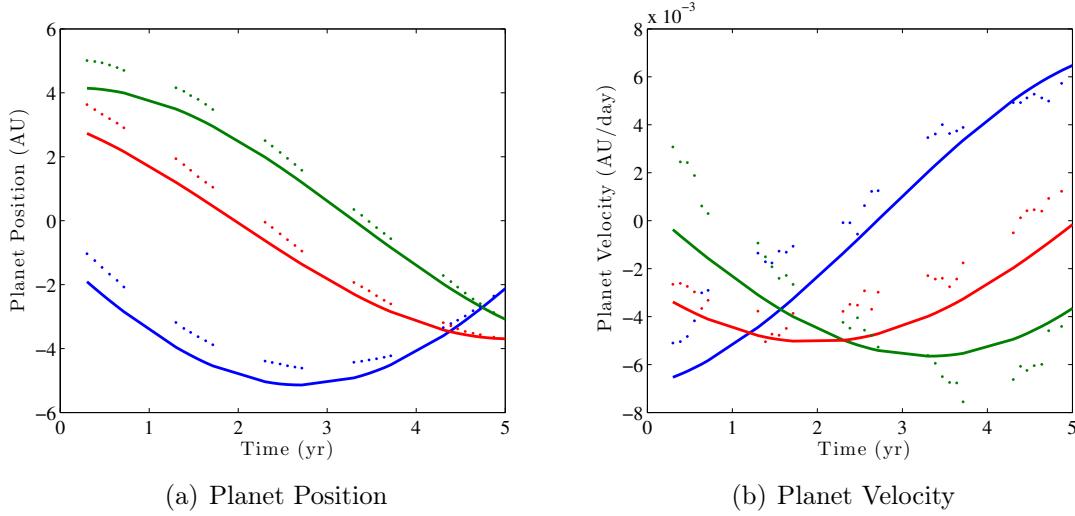


Figure 5.17: Planet position and velocity components (solid lines) and filter output (points) for half of one period of doppler spectroscopy observations of a Jupiter mass planet and one direct detection with multiple filtering and smoothing steps.

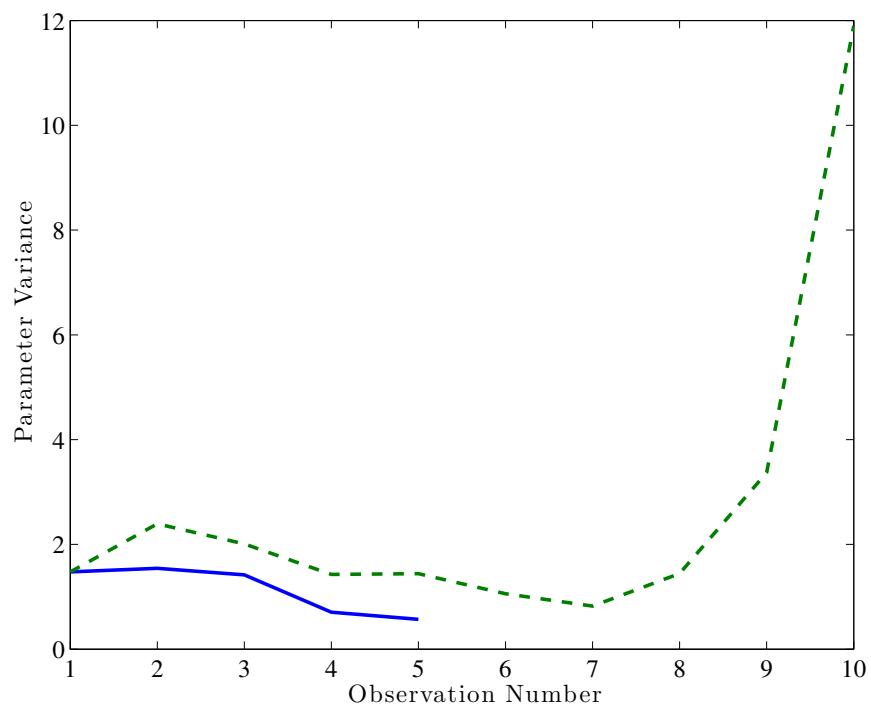


Figure 5.18: Estimated variance of planet state (sum of covariance diagonal) for data sets composed of measurements of one planet (solid lines) and two different planets (dashed lines).

Chapter 6

Conclusions and Future Work

In order to gain a better understanding of how planets form and develop we must improve our ability to observe planets of all types, and to extract information from such observations. The intent of this thesis was to create a unified formalism that could be applied towards both these goals. To do so, it was necessary to identify the set of variables that could describe the time varying behavior of an exosystem, and map directly to the observables of any method that could be used to find and study planets.

Once this set of variables was identified, and mappings were generated to the various observables, we could begin to develop applications in support of our goals. First, in order to improve the chances of future planet-finding instruments, we created a framework for simulating entire planet-finding surveys, taking into account instrument and observatory constraints. Along the way, it was necessary to implement a rudimentary form of decision modeling so that entire timelines of observations could be scheduled automatically, without human interference. The final utility that came from direct application of our instrument and exosystem models has proven to be an invaluable tool for evaluating proposed science instruments and survey plans.

Similarly, by taking advantage of the optimal estimation formalism used for dynamic filtering, the very same models were brought to bear on the data analysis problem, and

proved equally applicable. The fact that this formalism is capable of dealing with multiple changing observation functions implies that we already have a proven method for synthesizing multiple data sets, which is certain to prove useful in the near future.

The applications presented in this thesis are only the first step. As more detail is included in the simulations, they will become much more powerful tools for instrument evaluation and will help with setting up and running actual surveys. More importantly, as our confidence in these simulations grows, we will be able to compare them with actual survey results and thus judge whether the underlying assumptions in the simulations (i.e., planetary populations or evolution theories) correspond to reality. This final step will turn exoplanet finding from a blind search into a series of controlled experiments, and will greatly advance our knowledge of how we came to be here, and what the future has in store.

Bibliography

- [1] T. S. Barman, P. H. Hauschildt, and F. Allard. Irradiated planets. *The Astrophysical Journal*, 556:885, 2001.
- [2] R. H. Battin. *An introduction to the mathematics and methods of astrodynamics*. AIAA, New York, 1987. ISBN 1563473429.
- [3] D. C. Black and J. D. Scargle. On the detection of other planetary systems by astrometric techniques. *The Astrophysical Journal*, 263:854–869, 1982.
- [4] W. J. Borucki, D. Koch, G. Basri, N. Batalha, T. Brown, D. Caldwell, J. Caldwell, J. Christensen-Dalsgaard, W. D. Cochran, E. DeVore, et al. Kepler planet-detection mission: introduction and first results. *Science*, 327(5968):977, 2010.
- [5] W. J. Borucki, D. G. Koch, G. Basri, N. Batalha, A. Boss, T. M. Brown, D. Caldwell, J. Christensen-Dalsgaard, W. D. Cochran, E. DeVore, et al. Characteristics of kepler planetary candidates based on the first data set. *The Astrophysical Journal*, 728:117, 2011.
- [6] Ronald N. Bracewell. *The Fourier Transform and its Applications*. McGraw-Hill, New York, 3rd edition, 2000.
- [7] R. A. Brown. Obscurational completeness. *The Astrophysical Journal*, 607:1003–1017, 2004.

- [8] R. A. Brown. Single-visit photometric and obscurational completeness. *The Astrophysical Journal*, 624:1010–1024, 2005.
- [9] R. A. Brown. On the completeness of reflex astrometry on extrasolar planets near the sensitivity limit. *The Astrophysical Journal*, 699(1):711–715, 2009.
- [10] R. A. Brown, S. B. Shaklan, and S. L. Hunyadi. Science operations and the minimum scale of tpf. In *Coronagraph Workshop*, volume 07-02. JPL, 2007.
- [11] R.A. Brown and R. Soummer. New completeness methods for estimating exoplanet discoveries by direct detection. *The Astrophysical Journal*, 715:122, 2010.
- [12] S. T. Bryson, J. M. Jenkins, D. J. Peters, P. P. Tenenbaum, T. C. Klaus, J. P. Gunter, M. T. Cote, and D. A. Caldwell. The kepler end-to-end model: creating high-fidelity simulations to test kepler ground processing. In *Proc. SPIE*, volume 7738, 2010.
- [13] A. Burrows, D. Sudarsky, and J. I. Lunine. Beyond the t dwarfs: Theoretical spectra, colors, and detectability of the coolest brown dwarfs. *The Astrophysical Journal*, 596:587, 2003.
- [14] R. P. Butler, G. W. Marcy, E. Williams, C. McCarthy, P. Dosanjh, and S. S. Vogt. Attaining doppler precision of 3 m s-1. *Publications of the Astronomical Society of the Pacific*, 108:500, 1996.
- [15] R. P. Butler, J. T. Wright, G. W. Marcy, D. A Fischer, S. S. Vogt, C. G. Tinney, H. R. A. Jones, B. D. Carter, J. A. Johnson, C. McCarthy, and A. J. Penny. Catalog of nearby exoplanets. *The Astrophysical Journal*, 646:505, 2006.
- [16] Eric J. Cady. *Design, tolerancing, and experimental verification of occulters for finding extrasolar planets*. PhD thesis, Princeton University, 2010.
- [17] B. Campbell, GAH Walker, and S. Yang. A search for substellar companions to solar-type stars. *The Astrophysical Journal*, 331:902–921, 1988. ISSN 0004-637X.

- [18] W. Cash. Detection of earth-like planets around nearby stars using a petal-shaped occulter. *Nature*, 442(7098):51–53, 2006.
- [19] J. Catanzarite, M. Shao, A. Tanner, S. Unwin, and J. Yu. Astrometric detection of terrestrial planets in the habitable zones of nearby stars with sim planetquest. *Publications of the Astronomical Society of the Pacific*, 118:1319–1339, 2006. URL <http://adsseeks.cfa.harvard.edu/abs/2006PASP..118.1319C>.
- [20] J. E. Chambers. A hybrid symplectic integrator that permits close encounters between massive bodies. *Monthly Notices of the Royal Astronomical Society*, 304(4):793–799, 1999.
- [21] A. Claret. Testing the limb-darkening coefficients measured from eclipsing binaries. *Astronomy and Astrophysics*, 482(1):259–266, 2008. ISSN 0004-6361.
- [22] M. M. Colavita. Measurement of the atmospheric limit to narrow angle interferometric astrometry using the mark-iii stellar interferometer. *Astronomy and Astrophysics*, 283: 1027, 1994.
- [23] L. Colina, R. Bohlin, and F. Castelli. Absolute flux calibrated spectrum of vega. In J. E. Gunn and L. L. Stryker, editors, *Stellar Spectrophotometric Atlas*, volume 52. ApJ Supp. Ser., 1996.
- [24] J. L. Crassidis and J. L. Junkins. *Optimal Estimation of Dynamic Systems*. Chapman & Hall/CRC, 2004.
- [25] A. Cumming, G. W. Marcy, R. P. Butler, and S. S. Vogt. The statistics of of extrasolar planets: Results from the keck survey. *Scientific Frontiers in Research on Extrasolar Planets*, 2003.
- [26] A. Cumming, R. P. Butler, G. W. Marcy, S. S. Vogt, J. T. Wright, and D. A. Fischer. The Keck planet search: detectability and the minimum mass and orbital period dis-

tribution of extrasolar planets. *Publications of the Astronomical Society of the Pacific*, 120:531–554, 2008.

- [27] T. Currie. On the semimajor axis distribution of extrasolar gas giant planets: Why hot jupiters are rare around high-mass stars. *The Astrophysical Journal Letters*, 694:L171, 2009.
- [28] S. P. Davis, M. C. Abrams, and J. W. Brault. *Fourier transform spectrometry*. Academic Press, 2001.
- [29] G. De Vaucouleurs. Geometric and photometric parameters of the terrestrial planets. *Icarus*, 3(3):187–235, 1964.
- [30] D. J. Des Marais, M. O. Harwit, K. W. Jucks, J. F. Kasting, D. N. C. Lin, J. I. Lunine, J. Schneider, S. Seager, W. A. Traub, and N. J. Woolf. Remote sensing of planetary properties and biosignatures on extrasolar terrestrial planets. *Astrobiology*, 2(2):153–181, 2002.
- [31] J. R. Dormand, M. E. A. El-Mikkawy, and P. J. Prince. High-order embedded runge-kutta-nystrom formulae. *IMA Journal of Numerical Analysis*, 7(4):423, 1987. ISSN 0272-4979.
- [32] M. J. Duncan, H. F. Levison, and M. H. Lee. A multiple time step symplectic algorithm for integrating close encounters. *The Astronomical Journal*, 116:2067, 1998.
- [33] A. Einstein and L. Infeld. The gravitational equations and the problem of motion. ii. *Annals of Mathematics*, pages 455–464, 1940. ISSN 0003-486X.
- [34] A. Einstein, L. Infeld, and B. Hoffmann. The gravitational equations and the problem of motion. *Annals of mathematics*, 39(1):65–100, 1938. ISSN 0003-486X.
- [35] D. C. Fabrycky. Non-keplerian dynamics. *Arxiv preprint arXiv:1006.3834*, 2010.

- [36] E. Fehlberg. *Classical fifth-, sixth-, seventh-, and eighth-order Runge-Kutta formulas with stepsize control*. NASA, 1968.
- [37] J. J. Fortney, M. S. Marley, and J. W. Barnes. Planetary radii across five orders of magnitude in mass and stellar insolation: application to transits. *The Astrophysical Journal*, 659:1661–1672, 2007.
- [38] J. D. Gibbons and S. Chakraborti. *Nonparametric statistical inference*. CRC, 2003.
- [39] R. L. Gilliland, S. L. Morris, R. J. Weymann, D. C. Ebbets, and D. J. Lindler. Resolution and noise properties of the goddard high-resolution spectrograph. *Publications of the Astronomical Society of the Pacific*, 104:367–382, 1992.
- [40] J. W. Goodman. *Introduction to Fourier optics*. Roberts & Company Publishers, 2005.
- [41] W. H. Goodyear. Completely general closed-form solution for coordinates and partial derivative of the two-body problem. *The Astronomical Journal*, 70:189, 1965. ISSN 0004-6256.
- [42] W. H. Goodyear. A general method for the computation of cartesian coordinates and partial derivatives of the two-body problem. Technical report, National Aeronautics and Space Administration, 1966.
- [43] A. Gould, S. Dong, B. S. Gaudi, A. Udalski, I. A. Bond, J. Greenhill, R. A Street, M. Dominik, T. Sumi, M. K. Szymański, et al. Frequency of solar-like systems and of ice and gas giants beyond the snow line from high-magnification microlensing events in 2005-2008. *The Astrophysical Journal*, 720:1073, 2010.
- [44] R. M. Green. *Spherical astronomy*. Cambridge University Press, 1985.
- [45] T. D. Groff, A. Carlotti, and N. J. Kasdin. Progress on broadband control and deformable mirror tolerances in a 2-DM system. In *Proceedings of SPIE*, volume 7731, page 77314S, 2010.

- [46] P. Guillout, J. Schmitt, D. Egret, W. Voges, C. Motch, and M. F. Sterzik. The stellar content of soft X-ray surveys. II. Cross-correlation of the ROSAT All-Sky Survey with the Tycho and Hipparcos catalogs. *Astronomy and Astrophysics*, 351:1003–1015, 1999. ISSN 0004-6361.
- [47] F. Gustafsson. *Adaptive filtering and change detection*. John Wiley & Sons Inc, 2000. ISBN 0471492876.
- [48] O. Guyon. Phase-induced amplitude apodization of telescope pupils for extrasolar terrestrial planet imaging. *Astronomy and Astrophysics*, 404:379, 2003.
- [49] O. Guyon, EA Pluzhnik, MJ Kuchner, B. Collins, and ST Ridgway. Theoretical limits on extrasolar terrestrial planet detection with coronagraphs. *The Astrophysical Journal Supplement Series*, 167:81, 2006.
- [50] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.
- [51] P. H. Hauschildt, T. Barman, and E. Baron. Irradiated planets. *Physica Scripta*, 2008: 014033, 2008.
- [52] S. R. Heap. The terrestrial planet finder-occultor (TPF-O) science program. In *Proceedings of SPIE*, volume 6687, page 668713, 2007.
- [53] S. R. Heap. Detecting Biomarkers in ExoPlanetary Atmospheres with Terrestrial Planet Finder. In *EAS Publications Series*, volume 41, pages 517–520, 2010.
- [54] Eugene Hecht. *Optics*. Addison Wesley, New York, fourth edition, 2002.
- [55] T. J. Henry and D. W. McCarthy Jr. The mass-luminosity relation for stars of mass 1.0 to $0.08m_{\odot}$. *AJ*, 106:773–789, 1993.

- [56] P. M. Hinz, T. J. Rodigas, M. A. Kenworthy, S. Sivanandam, A. N. Heinze, E. E. Mamajek, and M. R. Meyer. Thermal infrared MMTAO observations of the HR 8799 planetary system. *The Astrophysical Journal*, 716:417, 2010.
- [57] D. W. Hogg, A. D. Myers, and J. Bovy. Inferring the eccentricity distribution. *The Astrophysical Journal*, 725(2):2166, 2010. URL <http://stacks.iop.org/0004-637X/725/i=2/a=2166>.
- [58] J. H. Horne and S. L. Baliunas. A prescription for period analysis of unevenly sampled time series. *The Astrophysical Journal*, 302:757–763, 1986. ISSN 0004-637X.
- [59] L. Hui and S. Seager. Atmospheric lensing and oblateness effects during an extrasolar planetary transit. *The Astrophysical Journal*, 572:540, 2002.
- [60] S. L. Hunyadi, A. S. Lo, and S. B. Shaklan. The dark side of TPF: detecting and characterizing extra-solar earthlike planets with one or two external occulters. In *Proc SPIE*, volume 6693, page 669303, 2007.
- [61] IEEE Computer Society. Ieee standard for floating-point arithmetic. *IEEE Std 754-2008*, pages 1 –58, 29 2008. doi: 10.1109/IEEESTD.2008.4610935.
- [62] P. Jacquinot and B. Roizen-Dossier. Apodisation. In E. Wolf, editor, *Progress in Optics*, volume 3, pages 29–186. Elsevier, 1964.
- [63] M. Janson, C. Bergfors, M. Goto, W. Brandner, and D. Lafreniere. Spatially resolved spectroscopy of the exoplanet hr 8799 c. *The Astrophysical Journal Letters*, 710:L35, 2010.
- [64] S. J. Julier and J. K. Uhlmann. Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92(3):401–422, 2004. ISSN 0018-9219.

- [65] S. J. Julier, J. K. Uhlmann, and H. F. Durrant-Whyte. A new approach for filtering nonlinear systems. In *American Control Conference, 1995. Proceedings of the*, volume 3, pages 1628–1632. IEEE, 1995. ISBN 0780324455.
- [66] T. R. Kane, P. W. Likins, and D. A. Levinson. *Spacecraft Dynamics*. McGraw-Hill Book Co., New York, 1983.
- [67] N. J. Kasdin and I. Braems. Linear and bayesian planet detection algorithms for the terrestrial planet finder. *The Astrophysical Journal*, 646:1260–1274, 2006.
- [68] N. J. Kasdin, R. J. Vanderbei, M. G. Littman, and D. N. Spergel. Optimal one-dimensional apodizations and shaped pupils for planet finding coronagraphy. *Applied optics*, 44(7):1117–1128, 2005.
- [69] J. F. Kasting, D. P. Whitmire, and R. T. Reynolds. Habitable zones around main sequence stars. *Icarus*, 101(1):108–128, 1993.
- [70] E. Kolemen and N. J. Kasdin. Optimal trajectory control of an occulter based planet finding telescope. *Proceedings of the AAS Guidance and Control Conference at Breckenridge, Colorado*, Feb 2007. AAS 07-037.
- [71] Egemen Kolemen. *Optimal configuration of a planet-finding mission consisting of a telescope and a constellation of occulters*. PhD thesis, Princeton University, 2008.
- [72] M. Konacki, A. J. Maciejewski, and A. Wolszczan. Frequency Decomposition of Astrometric Signature of Planetary Systems. *Astrophysical Journal*, 567:566–578, 2002.
- [73] M. J. Kuchner and K. A. Farley. The Historical Distribution of Zodiacal Dust Levels and its Implications for the Direct Detection of Exoearths. 2008.
- [74] M. J. Kuchner and C. C. Stark. Collisional grooming models of the kuiper belt dust cloud. *The Astronomical Journal*, 140:1007, 2010.

- [75] M. J. Kuchner and W. A. Traub. A coronagraph with a band-limited mask for finding terrestrial planets. *The Astrophysical Journal*, 570(2):900–908, 2002.
- [76] D. Lafrenière, C. Marois, R. Doyon, and T. Barman. HST/NICMOS Detection of HR 8799 b in 1998. *The Astrophysical Journal*, 694(2), 2009.
- [77] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright. Convergence properties of the nelder-mead simplex method in low dimensions. *SIAM Journal on Optimization*, 9(1):112–147, 1999. ISSN 1052-6234.
- [78] H. J. Larson and B. O. Shubert. *Probabilistic Models in Engineering Sciences Vol. 1*. John Wiley & Sons, New York, 1979.
- [79] A. S. Libert and J. Henrard. Exoplanetary systems: The role of an equilibrium at high mutual inclination in shaping the global behavior of the 3-d secular planetary three-body problem. *Icarus*, 191(2):469–485, 2007.
- [80] D. J. Lindler. TPF-O design reference mission. In *Proc. SPIE*, volume 6687, 2007.
- [81] N. R. Lomb. Least-squares frequency analysis of unequally spaced data. *Astrophysics and space science*, 39(2):447–462, 1976. ISSN 0004-640X.
- [82] C. Lovis, F. Pepe, F. Bouchy, G.L. Curto, M. Mayor, L. Pasquini, D. Queloz, G. Rupprecht, S. Udry, and S. Zuckere. The exoplanet hunter HARPS: unequalled accuracy and perspectives toward 1 cm s^{-1} precision. In *Proceedings of SPIE*, volume 6269, 2006.
- [83] M. Majji, J. L. Junkins, and J. D. Turner. A high order method for estimation of dynamic systems. *Journal of the Astronautical Sciences*, 56(3), 2008.
- [84] M. Majji, J. L. Junkins, and J. D. Turner. A perturbation method for estimation of dynamic systems. *Nonlinear Dynamics*, 60(3):303–325, 2010. ISSN 0924-090X.

- [85] K. Mandel and E. Agol. Analytic light curves for planetary transit searches. *The Astrophysical Journal Letters*, 580:L171, 2002.
- [86] G. Marcy, R. P. Butler, D. Fischer, S. Vogt, J. T. Wright, C. G. Tinney, and H. R. A. Jones. Observed properties of exoplanets: Masses, orbits, and metallicities. *Progress of Theoretical Physics Supplement*, 158:24–42, 2005. ISSN 0375-9687.
- [87] G. W. Marcy and R. P. Butler. Precision radial velocities with an iodine absorption cell. *Publications of the Astronomical Society of the Pacific*, 104:270–277, 1992.
- [88] C. Marois, B. Macintosh, T. Barman, B. Zuckerman, I. Song, J. Patience, D. Lafreniere, and R. Doyon. Direct imaging of multiple planets orbiting the star HR 8799. *Science*, 322(5906):1348, 2008.
- [89] M. Mayor and D. Queloz. A jupiter-mass companion to a solar-type star. *Nature*, 378 (6555):355–359, 1995. ISSN 0028-0836.
- [90] R. Mehra. Approaches to adaptive filtering. *Automatic Control, IEEE Transactions on*, 17(5):693–698, 1972. ISSN 0018-9286.
- [91] C. B. Moler. *Numerical computing with MATLAB*. Society for Industrial Mathematics, 2004.
- [92] C. Mordasini, Y. Alibert, and W. Benz. Extrasolar planet population synthesis. I. Method, formation tracks, and mass-distance distribution. *Astronomy and Astrophysics*, 501:1139–1160, 2009.
- [93] C. Mordasini, Y. Alibert, W. Benz, and D. Naef. Extrasolar planet population synthesis. II. Statistical comparison with observations. *Astronomy and Astrophysics*, 501: 1161–1184, 2009.
- [94] P. M. Morse and H. Feshbach. *Methods of Theoretical Physics*. McGraw-Hill New York, 1953.

- [95] D. W. Murphy and D. L. Meier. Simsim: an end-to-end simulator for the sim mission. In Simon C. Craig and Martin J. Cullum, editors, *Modeling and Systems Engineering for Astronomy*, volume 5497 of *Proceedings of SPIE*, pages 577–585. SPIE, 2004.
- [96] B. Ng, A. Pfeffer, and R. Dearden. Continuous time particle filtering. In *International Joint Conference on Artificial Intelligence*, volume 19, page 1360. Citeseer, 2005.
- [97] P. Nisenson and C. Papaliolios. Detection of earth-like planets using apodized telescopes. *The Astrophysical Journal Letters*, 548(2):201–205, 2001.
- [98] Babatunde A. Ogunnaike. *Random Phenomena: Fundamentals of Probability and Statistics for Engineers*. CRC Press, 2009.
- [99] S. N. Papakostas and C. Tsitouras. High phase-lag-order runge-kutta and nystrom pairs. *SIAM Journal on Scientific Computing*, 21(2):747–763, 2000.
- [100] M. J. Patterson, J. E. Foster, T. W. Haag, V. K. Rawlin, G. C. Soulard, and R. F. Roman. NEXT: NASAs Evolutionary Xenon Thruster. *AIAA Paper*, 3832, 2002.
- [101] S. H. Pravdo, S. Shaklan, R. A. Brown, and S. L. Hunyadi. Observation strategies and statistics of tpf and other planetary observations. In *Bulletin of the American Astronomical Society*, volume 38, page 333, 2007.
- [102] W. H. Press and G. B. Rybicki. Fast algorithm for spectral analysis of unevenly sampled data. *The Astrophysical Journal*, 338:277–280, 1989. ISSN 0004-637X.
- [103] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical recipes in C: The Art of Scientific Computing*. Cambridge University Press, 1992.
- [104] Laurent Pueyo. *Broadband contrast for exo-planet imaging : the impact of propagation effects*. PhD thesis, Princeton University, 2008.

- [105] Meng-Zhao Qin and Wen-Jie Zhu. Canonical Runge-Kutta-Nystrom (RKN) methods for second order ordinary differential equations. *Computers Math. Applic.*, 22(9):85–95, 1991.
- [106] I. N. Reid, J. E. Gizis, and S. L. Hawley. The Palomar/MSU Nearby Star Spectroscopic Survey. IV. The Luminosity Function in the Solar Neighborhood and M Dwarf Kinematics. *The Astronomical Journal*, 124(5):2721–2738, 2002.
- [107] S. J. Russell, P. Norvig, J. F. Canny, J. M. Malik, and D. D. Edwards. *Artificial Intelligence: A Modern Approach*, volume 74. Prentice Hall, Englewood Cliffs, NJ, 1995.
- [108] S. C. Rutan. Adaptive kalman filtering. *Analytical Chemistry*, 63(22):1103–1109, 1991.
ISSN 0003-2700.
- [109] S. Sarkka. On unscented kalman filtering for state estimation of continuous-time non-linear systems. *Automatic Control, IEEE Transactions on*, 52(9):1631–1641, 2007.
ISSN 0018-9286.
- [110] D. Savransky and N. J. Kasdin. Simulation and analysis of sub- μ as precision astrometric data for planet finding. *The Astrophysical Journal*, 721(2):1559, 2010. URL <http://stacks.iop.org/0004-637X/721/i=2/a=1559>.
- [111] D. Savransky and N. Jeremy Kasdin. Return visit optimization for planet finding missions. In *Bulletin of the American Astronomical Society*, volume 39, page 134. AAS, 2007.
- [112] D. Savransky and N. Jeremy Kasdin. Design reference mission construction for planet finders. In *Proc. SPIE*, volume 7010, 2008.
- [113] D. Savransky, N. J. Kasdin, and E. Cady. Analyzing the designs of planet finding

missions. *Publications of the Astronomical Society of the Pacific*, 122(890):401–419, April 2010.

- [114] D. Savransky, D. N. Spergel, N. J. Kasdin, E. J. Cady, P. D. Lisman, S. H. Pravdo, S. B. Shaklan, and Y. Fujii. Occulting ozone observatory science overview. In *Proc. SPIE*, volume 7731, page 77312H, 2010.
- [115] D. Savransky, E. Cady, and N .J. Kasdin. Parameter distributions of Keplerian orbits. *The Astrophysical Journal*, 728(1):66, 2011.
- [116] Jeffrey D. Scargle. Studies in astronomical time series analysis ii. statistical aspects of spectral analysis of unevenly spaced data. *The Astrophysical Journal*, 263:835–853, 1982.
- [117] S. B. Shaklan, M. C. Noecker, A. S. Lo, T. Glassman, P. J. Dumont, E. O. Jordan, N. J. Kasdin, Jr. W. C. Cash, E. J. Cady, and P. R. Lawson. Error budgeting and tolerancing of starshades for exoplanet detection. In *Proceedings of SPIE*, volume 7731, 2010.
- [118] L. F. Shampine and M. K. Gordon. *Computer solution of ordinary differential equations: the initial value problem*, volume 8. Freeman San Francisco, 1975.
- [119] M. Shao and M. M. Colavita. Potential of long-baseline infrared interferometry for narrow-angle astrometry. *Astronomy and Astrophysics*, 262:353–358, 1992.
- [120] S. W. Shepperd. Universal Keplerian state transition matrix. *Celestial Mechanics and Dynamical Astronomy*, 35(2):129–144, 1985. ISSN 0923-2958.
- [121] D. Simon and T. L. Chia. Kalman filtering with state equality constraints. *Aerospace and Electronic Systems, IEEE Transactions on*, 38(1):128–136, 2002. ISSN 0018-9251.
- [122] D. Simon and D. L. Simon. Kalman filtering with inequality constraints for turbofan

engine health estimation. In *Control Theory and Applications, IEE Proceedings*, volume 153, pages 371–378, 2006.

- [123] D. Sirbu, C. V. Karsten, and N. J. Kasdin. Dynamical performance for science-mode stationkeeping with an external occulter. In *Proceedings of SPIE*, volume 7731, page 773152, 2010.
- [124] V. V. Sobolev. *Light Scattering in Planetary Atmospheres*. Pergamon Press, New York, 1975.
- [125] R. Soummer. Apodized pupil lyot coronagraphs for arbitrary telescope apertures. *The Astrophysical Journal Letters*, 618(2):L161–L164, 2005.
- [126] A. Sozzetti, S. Casertano, R. A. Brown, and M. G. Lattanzi. Narrow-Angle Astrometry with the Space Interferometry Mission: The Search for Extrasolar Planets. I. Detection and Characterization of Single Planets. *Publications of the Astronomical Society of the Pacific*, 114:1173–1196, 2002. URL <http://adsseeks.cfa.harvard.edu/abs/2002PASP..114.1173S>.
- [127] A. Sozzetti, S. Casertano, R. A. Brown, and M. G. Lattanzi. Narrow-Angle Astrometry with the Space Interferometry Mission: The Search for Extrasolar Planets. II. Detection and Characterization of Planetary Systems. *Publications of the Astronomical Society of the Pacific*, 115:1072–1104, 2003. URL <http://adsseeks.cfa.harvard.edu/abs/2003PASP..115.1072S>.
- [128] Alessandro Sozzetti. Astrometric methods and instrumentation to identify and characterize extrasolar planets: A review. *Publications of the Astronomical Society of the Pacific*, 117:1021–1048, 2005. URL <http://adsseeks.cfa.harvard.edu/abs/2005PASP..117.1021S>.
- [129] D.N. Spergel, J. Kasdin, R. Belikov, P. Atcheson, M. Beasley, D. Calzetti, B. Cameron, C. Copi, S. Desch, A. Dressler, D. Ebbets, R. Egerman, A. Fullerton, J. Gallagher,

- J. Green, O. Guyon, S. Heap, R. Jansen, E. Jenkins, J. Kasting, R. Keski-Kuha, M. Kuchner, R. Lee, D. Lindler, R. Linfield, D. Lisman, R. Lyon, S. Malhotra, G. Matthews, M. McCaughean, J. Mentzel, M. Mountain, S. NIkzad, R. O'Connell, S. Oey, D. Padgett, B. Parvin, J. Procashka, W. Reeve, I. N. Reid, J. Rhoads, A. Roberge, B. Saif, P. Scowen, S. Seager, O. Seigmund, K. Sembach, S. Shaklan, M. Shull, and R. Soummer. THEIA: Telescope for habitable exoplanets and interstellar/intergalactic astronomy. In *AAS Meeting 213*, volume 458. American Astronomical Society, 2009.
- [130] J. J. Spilker. *Global positioning system: theory and applications*. AIAA, 1996. ISBN 1563471078.
- [131] C. C. Stark and M. J. Kuchner. The Detectability of Exo-Earths and Super-Earths Via Resonant Signatures in Exozodiacal Clouds. *The Astrophysical Journal*, 686:637–648, 2008.
- [132] R. F. Stengel. *Optimal control and estimation*. Dover Publications, 1994.
- [133] D. Sudarsky, A. Burrows, I. Hubeny, and A. Li. Phase Functions and Light Curves of Wide-Separation Extrasolar Giant Planets. *The Astrophysical Journal*, 627(1):520–533, 2005.
- [134] Karl Sundman. Mémoire sur le problème des trois corps. *Acta Mathematica*, 36: 105–179, 1913. ISSN 0001-5962. URL <http://dx.doi.org/10.1007/BF02422379>. 10.1007/BF02422379.
- [135] C. G. Tinney, R. P. Butler, G. W. Marcy, H. R. A. Jones, G. Laughlin, B. D. Carter, J. A. Bailey, and S. O'Toole. The 2: 1 resonant exoplanetary system orbiting hd 73526. *The Astrophysical Journal*, 647:594, 2006.
- [136] C. Tsitouras and S. N. Papakostas. Cheap error estimation for runge-kutta methods. *SIAM Journal on Scientific Computing*, 20(6):2067–2088, 1999. ISSN 1064-8275.

- [137] C. Turon, Y. Requiem, M. Grenon, A. Gomez, D. Morin, F. Crifo, F. Arenou, M. Froeschle, F. Mignard, M. A. C. Perryman, A. N. Argue, M. Creze, J. Dommangé, D. Egret, H. Jahreiss, M. O. Mennessier, A. Pham, D. W. Evans, F. van Leeuwen, and L. Lindegren. Properties of the HIPPARCOS Input Catalogue. *Astronomy and Astrophysics*, 304:82–+, December 1995.
- [138] R. Van Der Merwe and E. A. Wan. The square-root unscented kalman filter for state and parameter-estimation. In *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on*, volume 6, pages 3461–3464. IEEE, 2001. ISBN 0780370414.
- [139] R. Van Der Merwe, A. Doucet, N. De Freitas, and E. Wan. The unscented particle filter. *Advances in Neural Information Processing Systems*, pages 584–590, 2001. ISSN 1049-5258.
- [140] R. J. Vanderbei. Diffraction analysis of two-dimensional pupil mapping for high-contrast imaging. *The Astrophysical Journal*, 636(1):528–543, 2006.
- [141] R. J. Vanderbei, D. N. Spergel, and N. J. Kasdin. Circularly symmetric apodization via star-shaped masks. *The Astrophysical Journal*, 599(1):686–694, 2003.
- [142] R. J. Vanderbei, N. J. Kasdin, and D. N. Spergel. Checkerboard-mask coronagraphs for high-contrast imaging. *The Astrophysical Journal*, 615(1):555–561, 2004.
- [143] R. J. Vanderbei, E. Cady, and N. J. Kasdin. Optimal Occulter Design for Finding Extrasolar Planets. *The Astrophysical Journal*, 665(1):794–798, 2007.
- [144] J. P. Vinti. *Orbital and Celestial Mechanics*. Charles Stark Draper Laboratory, Inc., Cambridge, Mass., 1998.
- [145] S. S. Vogt. The lick observatory hamilton echelle spectrometer. *Publications of the Astronomical Society of the Pacific*, 99:1214–1228, 1987.

- [146] S. S. Vogt, S. L. Allen, B. C. Bigelow, L. Bresee, W. E. Brown, T. Cantrall, A. Conrad, M. Couture, C. Delaney, H. W. Epps, et al. HIRES: The high-resolution echelle spectrometer on the Keck 10-m telescope. In *Proceedings of SPIE*, volume 2198, page 362, 1994.
- [147] G. N. Watson. *A treatise on the theory of Bessel functions*. Cambridge University Press, 1944.
- [148] J. Wisdom and M. Holman. Symplectic maps for the n-body problem. *The Astronomical Journal*, 102:1528–1538, 1991. ISSN 0004-6256.
- [149] H. Yoshida. Recent progress in the theory and application of symplectic integrators. *Celestial mechanics and dynamical astronomy*, 56(1):27–43, 1993. ISSN 0923-2958.

Appendix A

Code Listings

Listing A.1: Implementation of Equations (1.2.20) - (1.2.31).

```
function [a,e,E,I,omega,Omega,P,tau,A,B] = vec2orbElem(rs,vs,mus)
% vec2orbElem(rs,vs,mus) converts positions (rs) and velocities (vs)
% of bodies with gravitational parameters (mus) to Keplerian orbital
% elements.
5
%
% INPUTS:
% rs      3n x 1 stacked initial position vectors:
%          [r1(1);r1(2);r1(3);r2(1);r2(2);r2(3);...;rn(1);rn(2);rn(3)]
%          or 3 x n matrix of position vecotrs.
10
% vs      3n x 1 stacked initial velocity vectors or 3 x n matrix
% mus    gravitational parameters ( $G*m_i$ ) where  $G$  is the
%         gravitational constant and  $m_i$  is the mass of the  $i$ th body.
%         if all vectors represent the same body, mus may be a
%         scalar.
15
%
% OUTPUTS:
% a        semi-major axes
% e        eccentricities
% E        eccentric anomalies
20
% I        inclinations
% omega   arguments of pericenter
% Omega    longitudes of ascending nodes
% P        orbital periods
% tau     time of periapsis crossing
25
% A, B    orientation matrices
%
%Example:
% %solar system orbital elements
% ssdat = load('solarSystemData.mat');
30
% rs = ssdat.p0(1:end-3) - repmat(ssdat.p0(end-2:end), 9, 1);
% vs = ssdat.v0(1:end-3) - repmat(ssdat.v0(end-2:end), 9, 1);
% mus = ssdat.mus(1:9) + ssdat.mus(10);
% [a,e,E,I,omega,Omega,P,tau,A,B] = vec2orbElem(rs,vs,mus);
% %convert back:
35
% r = A*diag(cos(E) - e) + B*diag(sin(E));
% rdot = (-A*diag(sin(E)) + B*diag(cos(E))) * ...
%         diag(sqrt(mus(:)).'*a.^-3)./(1 - e.*cos(E)));
%
```

```

%condition inputs
40 mus = mus(:)';
nplanets = numel(rs)/3;
if mod(nplanets,1) ~= 0 || numel(vs) ~= nplanets*3 || ...
    (length(mus) ~= nplanets && length(mus) ~= 1)
    error('vec2orbElem:inputError',[ 'rs and vs must contain 3n',...
        'elements and mus must have length n or 1 for n bodies']);
end
if length(rs) == numel(rs)
    rs = reshape(rs,3,nplanets);
end
50 if length(vs) == numel(vs)
    vs = reshape(vs,3,nplanets);
end

v2s = sum(vs.^2);
55 r = sqrt(sum(rs.^2)); %orbital separation
Ws = 0.5*v2s - mus./r;
a = -mus./Ws; %semi-major axis

L = [rs(2,:).*vs(3,:) - rs(3,:).*vs(2,:);...
60     rs(3,:).*vs(1,:) - rs(1,:).*vs(3,:);...
     rs(1,:).*vs(2,:) - rs(2,:).*vs(1,:)]; %angular momentum
L2s = sum(L.^2);
p = L2s./mus; %semi-latus rectum
e = sqrt(1 - p./a); %eccentricity
65

%eccentric anomaly
cosE = (1 - r./a)./e;
sinE = sum(rs.*vs)./(e.*sqrt(mus.*a));
E = atan2(sinE,cosE);

70 if nargout < 4, return; end

%inclination
sinI = sqrt(L(1,:).^2 + L(2,:).^2)./sqrt(L2s);
cosI = L(3,:)./sqrt(L2s);
I = atan2(sinI,cosI);

%argument of pericenter
sinw = ((vs(1,:).*L(2,:)-vs(2,:).*L(1,:))./mus - ...
80     rs(3,:)./r)./(e.*sinI);
cosw = ((sqrt(L2s).*vs(3,:))./mus - (L(1,:).*rs(2,:)-...
     L(2,:).*rs(1,:))./(sqrt(L2s).*r))./(e.*sinI);
omega = atan2(sinw,cosw);

85 %longitude of ascending node
cos0 = -L(2,:)./(sqrt(L2s).*sinI);
sin0 = L(1,:)./(sqrt(L2s).*sinI);
Omega = atan2(sin0,cos0);

90 if nargout < 7, return; end

```

```

%orbital periods
P = 2*pi*sqrt(a.^3./mus);

95 %time of periapsis crossing
tau = -(E - e.*sin(E))./sqrt(mus.*a.^-3);

if nargout < 9, return; end

100 %A and B vectors
A = [a.*cos(Omega).*cos(omega) - sin(Omega).*cos(I).*sin(omega)); ...
      a.*sin(Omega).*cos(omega) + cos(Omega).*cos(I).*sin(omega)); ...
      a.*sin(I).*sin(omega)]; 

105 B = [-a.*sqrt(1-e.^2).*(cos(Omega).*sin(omega) + ...
      sin(Omega).*cos(I).*cos(omega)); ...
      a.*sqrt(1-e.^2).*(-sin(Omega).*sin(omega) + ...
      cos(Omega).*cos(I).*cos(omega)); ...
      a.*sqrt(1-e.^2).*sin(I).*cos(omega)];
```

Listing A.2: Efficient generation of the coefficient space defined in Equation (3.3.30).

```

function xkmat = calcRevCoefs(n)
% calcRevCoefs - Find series reversion coefficient sets
%
% INPUTS
% calcRevCoefs(n) returns all sets of natural numbers such that
% sum(i*x_i) = n. n must be greater than 1.
%
% OUTPUT
% xkmat      N x n matrix with each row representing a unique set,
%             such that xkmat(j,:)*(1:n).' = n

%check input
if ~isa(n,'numeric') || floor(n) ~= n || n < 2
    error('calcRevCoefs:inputError',...
        'n must be a natural number greater than 1');
end

%generate original pairs (excluding (0,n))
matr = genBitPairs(n);

%expand
for k = 3:n
    %there will be this many new subsets:
    newCombs = floor(matr(:,end)/2);
    oDim = size(matr); %orig dimensions
    newDim = oDim + [sum(newCombs),1]; %new dimensions
    matr(newDim(1),newDim(2)) = 0; %resize matrix

    %only expand rows with new combinations
    for j = find(newCombs).'
        if matr(j,oDim(2)) > 1
            matr(oDim(1)+1+sum(newCombs(1:j-1)):...
                oDim(1)+sum(newCombs(1:j)),:) = ...
                repmat(matr(j,1:oDim(2)-1),newCombs(j),1),...
                genBitPairs(matr(j,oDim(2))]);
        end
    end
    %keep only unique combinations
    [~,i] = unique(sort(matr,2),'rows','first');
    matr = matr(i,:);
end
y = sort(matr,2);

%xkmat will be the same size as y
xkmat = zeros(size(y));
%for all non-zero entries in y, add 1 to xkmat
%in the same row as y, and column given by the y entry
%(i.e., if y has an i anywhere in row j, add 1
%to xkmat at index (i-1)*size(y,1) + j
inds = (y-1)*size(y,1);
good = inds >= 0;
inds = inds + repmat((1:size(y,1)).',1,n);

```

```

55 inds = inds(good);
uinds = unique(inds);
%returns number of all unique indices:
xkmat(uinds) = histc(inds,uinds);

%pad output with initial row and package
xkmat = [zeros(1,n);xkmat];
xkmat(1,n) = 1;
end

60 function bitpairs = genBitPairs(n)
%generate 2 x floor(n/2) matrix containing all
%integer pairs summing to n excluding (0,n)

65 if n <= 1
    bitpairs = [];
    return
end

70 bitpairs = zeros(floor(n/2),2);
bitpairs(:,1) = 1:floor(n/2);
bitpairs(:,2) = n - bitpairs(:,1);
end
75

```

Listing A.3: Implementation of Newton-Raphson iterative inversion of the Kepler equation.

```

function [E,nu] = invKepler(M,e)
%invKepler - Newton-Raphson inversion of the Kepler Equation
%
% INPUTS
% M      n x 1 or 1 x n vector of mean anomalies.
% e      vector of n eccentricity values, or scalar eccentricity
%
% OUTPUTS
% E      n x 1 vector of eccentric anomalies
% nu     n x 1 vector of true anomalies
%
% Example:
% one year of highly eccentric orbit
% [E, nu] = invKepler(0:pi/100:2*pi,0.8);

if (numel(M) ~= numel(e)) && numel(e) ~= 1
    error('invKepler:inputError',['For M of length n, e must',...
        ''have 1 or n elements']);
end

M = M(:); e = e(:);

initialize
counter = 0;
del = 1;
E = M./(1-e);
inds = E > sqrt(6*(1-e)./e);
if length(e) == 1
    einds = 1;
else
    einds = inds;
end
E(inds) = (6*M(inds)./e(einds)).^(1/3);

iterate
while ((del > eps(2*pi)) && (counter < 1000))
    E = E - (M - E + e.*sin(E))./(e.*cos(E)-1);
    del = max(abs(M - (E - e.*sin(E))));
    counter = counter+1;
end
if (counter == 1000) %convergence failure
    error('invKepler:overIteration',...
        ''Maximum number of iterations exceeded');
end

calculate true anomaly if needed
if (nargout > 1)
    nu = 2*atan(sqrt((1+e)./(1-e)).*tan(E/2));
    inds = nu < 0;
    nu(inds) = nu(inds)+2*pi;
end

```

Listing A.4: Implementation of multi-body Kepler state transition matrix [120].

```

function Phi = keplerSTM(x0,dt,mu)
% keplerSTM - State transition matrix for Keplerian orbits.
%
% INPUTS
5   % x0      6n x 1 vector of stacked positions and velocities:
%           [r1_1,r1_2,r1_3,v1_1,v1_2,v1_3,r2_1,r2_2,r2_3,...].'
%           positions and velocities are assumed heliocentric
% dt      Scalar representing propagation time
% mu      n x 1 vector of gravitational parameters (mu_S + mu_Pi)
10
%
% OUTPUT
% Phi     State transition matrix such that x = Phi*x0
%
% Example:
15   %propogate a planet 10 days along its orbit:
% x0 = [-0.8684, -0.6637, 0.7207, 0.0039, 0.0056, 0.0120].';
% x1 = keplerSTM(x0,10,2.6935e-04)*x0;

%determine number of planets and validate input
20 nplanets = length(x0)/6;
if nplanets - floor(nplanets) > 0
    error('keplerSTM:inputError',['The length of x0 ',...
        'must be a multiple of 6.']);
end
25 if length(mu) ~= nplanets
    error('keplerSTM:inputError',['The length of mu vector ',...
        'must be the length of x0 divided by 6']);
end

30 %orient mu properly
mu = mu(:).';

%create position and velocity matrices
x0 = reshape(x0,6,nplanets);
35 r0 = x0(1:3,:);
v0 = x0(4:6,:);

%constants
r0norm = sqrt(sum(r0.^2));
40 nu0 = dot(r0,v0);
bet = 2*mu./r0norm - dot(v0,v0);

%initialization
u = zeros(1,nplanets);

45 %For elliptic orbits, calculate period effects
deltaU = zeros(1,length(bet));
eorbs = bet > 0;
if any(eorbs)
    P = 2*pi*mu(eorbs).*bet(eorbs).^( -3/2 );
    n = floor((dt + P/2 - 2*nu0(eorbs)./bet(eorbs))./P);
    deltaU(eorbs) = 2*pi*n.*bet(eorbs).^( -5/2 );

```

```

end

55 %continued fraction constants
a = 5;
b = 0;
c = 5/2;

60 %kepler iteration loop
t = zeros(1,nplanets);
counter = 0;
%loop until convergence of the time array to the time step
while max(abs(t-dt)) > max(eps(t))*4 && counter < 1000
    q = bet.*u.^2./(1+bet.*u.^2);
    U0w2 = 1 - 2*q;
    U1w2 = 2*(1-q).*u;
    U = 16/15*U1w2.^5 .* contFrac(a,b,c,q) + deltaU;
    U0 = 2*U0w2.^2-1;
    U1 = 2*U0w2.*U1w2;
    U2 = 2*U1w2.^2;
    U3 = bet.*U + U1.*U2/3;
    r = r0norm.*U0 + nu0.*U1 + mu.*U2;
    t = r0norm.*U1 + nu0.*U2 + mu.*U3;
    u = u - (t-dt)./(4*(1-q).*r);
    counter = counter+1;
end
if counter == 1000
    error('keplerSTM:convError','Failed to converge on t.');
end

%kepler solution
f = 1 - mu./r0norm.*U2;
g = r0norm.*U1 + nu0.*U2;
F = -mu.*U1./r./r0norm;
G = 1 - mu./r.*U2;

%construct the state transition matrix
Phi = zeros(6*nplanets);
90 for j=1:nplanets
    st = (j-1)*6+1;
    Phi(st:st+5,st:st+5) = [eye(3)*f(j) eye(3)*g(j); ...
                           eye(3)*F(j) eye(3)*G(j)];
end

95 %calculate continued fraction
function G = contFrac(a,b,c,x)

    %initialize
100    k = 1 - 2*(a-b);
    l = 2*(c-1);
    d = 4*c*(c-1);
    n = 4*b*(c-a);
    A = ones(1,length(x));
    B = ones(1,length(x));
    G = ones(1,length(x));

```

```

    Gprev = zeros(1,length(x))+2;
    counter2 = 0;
    %loop until convergence of continued fraction
    while max(abs(G-Gprev)) > max(eps(G)) && counter2 < 1000
        k = -k;
        l = l+2;
        d = d+4*l;
        n = n+(1+k)*l;
        A = d./(d - n*A.*x);
        B = (A-1).*B;
        Gprev = G;
        G = G + B;
        counter2 = counter2+1;
    end
    if counter2 == 1000
        error('keplerSTM:convError','Failed to converge on G.');
    end
end
end

```

Listing A.5: Implementation of vectorized n-body equation.

```

function ddx = nbodyVect(x,mus)
%NBODYVECT - Fully vectorized n-body equation
%
% INPUTS
5   % x      3n x 1 stacked initial position vectors:
%       [r1(1);r1(2);r1(3);r2(1);r2(2);r2(3);...;rn(1);rn(2);rn(3)]
% mus    gravitational parameters ( $G*m_i$ ) where G is the
%         gravitational constant and  $m_i$  is the mass of the ith body.
% OUTPUTS
10  % ddx    second derivative of x

mus = mus(:); %make mus a column
n = length(mus); %this should be the number of planets

15  %input check
if (size(x,1) ~= 3*n)
    error('nbodyVect:inputError',[ 'x0 must be 3n x 1 vectors',...
        'and mus must have length n for n bodies']);
end

20  % precompute index arrays
% k index must be circularly permuted to fill 1 x 3n
% j index must be repeated to fill 1 x 3n
% inds is 1 - 3*n in 3 x n
25  % inds2 is 1 - 3*n in n x n-1
inds = reshape(1:3*n,3,n);
cols = repmat(1:n,1,n);
ks = cols(logical(ones(n) - eye(n)));
js = reshape(repmat(1:n,n-1,1),1,n*(n-1));
30  inds2 = reshape(1:n*(n-1),n-1,n).';

% r_k/j = r_k/o - r_j/o
rkj = x(inds(:,ks)) - x(inds(:,js));

35  %d^2/dt^2(r_k/j) = G*sum_{k ~= j} (mu_k*r_k/j)/(r_k/j^3)
rkj = rkj * diag(1./sqrt(sum(rkj.^2)).^3)*diag(mus(ks));
ddx = sum(reshape(rkj(:,inds2(:)),3*n,n-1),2);

```

Listing A.6: Efficient generation of the first cost function term from Equation (4.4.1).

```
function sdists = genCostFunTerm1(sinds, u_ts)
% genCostFunTerm1(sinds, u_ts) generates the first cost
% function term for available target indices (s_inds) and
% target unit look vectors (u_ts)

5 %combinations of visits
combs = nchoosek(sinds,2);

%angle between each pair of target stars
10 angdists = acos(dot(u_ts(combs(:,1),:),
    u_ts(combs(:,2),:),2));

%properly ordered cost function term
sdists = zeros(length(sinds));
sdists((tril(ones(length(sinds))) - ...
    eye(length(sinds)) == 1)) = angdists;
15 sdists = sdists+sdists.';
```

Listing A.7: Implementation of a generally applicable hybrid Extended Kalman Filter as described in Equations (5.1.34) through (5.1.41).

```

function [xhist,Phist] = hybridEKF(t,z,x0,P0,R,Q,f,F,h,H)
%hybridEKF - Hybrid Extended Kalman Filter
%
% INPUTS
% t      (N x 1) Array of times for N observations
% z      (N x m) matrix of observations for m element h function
% x0     (n x 1) vector of nitial conditions for n element state
% P0     (n x n) initial covariance matrix
% R      (m x m) observation noise covariance
% Q      (n x n) process noise covariance
% f      Function handle to (possible nonlinear) dynamics function.
%        Must accept 2 arguments: scalar time and (n x 1) state vector
%        so that f(t,x) = dx (n x 1)
% F      Function handle to Jacobian of f. Same inputs as f and
%        returns F(t,x) = df/dx (n x n)
% h      Function handle to (possibly nonlinear) observation function.
%        Input argument is an (n x 1) state vector so that
%        h(x) = y (m x 1)
% H      Function handle to Jacobian of h. Same inputs as h and
%        returns
%        H(x) = dh/dx (m x n)
%
% OUTPUTS
% xhist    (n x N) time history of filter state estimates
% Phist    (n x n x N) time history of filter covariance estimates

5   n = length(x0); %basic dimension
xminus = x0;
Pminus = P0;
10  xhist = zeros(n,length(t));
Phist = zeros(n,n,length(t));

%do one update to prime the filter
Hk = H(xminus);
15  K = (Pminus*Hk.')/(Hk*Pminus*Hk.' + R);
x = xminus + K*(z(1,:).' - h(xminus));
P = (eye(length(Pminus)) - K*Hk)*Pminus;
xhist(:,1) = x;
Phist(:,:,1) = P;

20  for k = 2:length(t)

    disp(k);
    %propogate
    [~,Zo] = ode45(@hybridEKF_deq,[t(k-1),t(k)],[x;P(:)]);
    25  xminus = Zo(end,1:n).';
    Pminus = reshape(Zo(end,n+1:end),n,n);

    %update
    Hk = H(xminus);
    K = (Pminus*Hk.')/(Hk*Pminus*Hk.' + R);

```

```

x = xminus + K*(z(k,:)' - h(xminus));
P = (eye(length(Pminus)) - K*Hk)*Pminus;

55 %add to history
xhist(:,k) = x;
Phist(:,:,k) = P;
end

60 %combined x and P differential equations
function dZ = hybridEKF_deq(T,Z)
    xi = Z(1:n);
    Pi = reshape(Z(n+1:end),n,n);
    dx = f(T,xi);
    Fi = F(T,xi);
    dPi = Fi*Pi + Pi*Fi.' + Q;
    dZ = [dx;dPi(:)];
end

```

Listing A.8: Implementation of a generally applicable hybrid Unscented Kalman Filter.

```

function [xhist,Phist] = hybridUKF(t,z,x0,P0,R,f,h)
%hybridEKF - Hybrid Unscented Kalman Filter
%
% INPUTS
5   % t      (N x 1) Array of times for N observations
% z      (N x m) matrix of observations for m element h function
% x0     (n x 1) vector of nitial conditions for n element state
% P0     (n x n) initial covariance matrix
% R      (m x m) observation noise covariance
10  % f      Function handle to (possible nonlinear) dynamics function.
% Must accept 2 arguments: scalar time and (n x 2n+1)
% matrix of sigma points so that f(t,x) = dx (n x 2n+1)
% h      Function handle to (possibly nonlinear) observation function.
% Input argument is an (n x 2n+1) matrix of sigma points
15  % so that h(x) = y (m x 2n+1)
%
% OUTPUTS
% xhist    (n x N) time history of filter state estimates
% Phist    (n x n x N) time history of filter covariance estimates
20
%
%basic dimensions
n = length(x0);
d = 2*n+1;
25
%
%prepare arrays
xminus = x0;
Pminus = P0;
xhist = zeros(n,length(t));
Phist = zeros(n,n,length(t));
30
%
%form weight matrices
%alph = 1e-3;
alph = 0.5;
bet = 2;
35
%kappa = 1;
kappa = 3-n;
lambda = alph^2*(n+kappa) - n;
c = alph^2*(n+kappa);
w_m = ones(d,1)*1/(2*(n+lambda));
40
w_c = w_m;
w_m(1) = lambda/(n+lambda);
w_c(1) = lambda/(n+lambda) + (1 - alph^2 + bet);
temp = eye(d) - repmat(w_m,1,d);
45
W = temp*diag(w_c)*temp.';

%
%do one update to prime the filter
A = chol(Pminus,'lower');
Xminus = repmat(xminus,1,d) + sqrt(c)*[zeros(n,1),A,-A];
50
Yminus = h(Xminus);
mu = Yminus*w_m;
S = Yminus*W*Yminus.' + R;

```

```

C = Xminus*W*Yminus.';
K = C/S;
x = xminus + K*(z(1,:).' - mu);
P = Pminus - K*S*K.';
xhist(:,1) = x;
Phist(:,:,1) = P;

60 %iterate
for k = 2:length(t)
    disp(k);
    %propagate
    [~,Zo] = ode45(@hybridUKF_deq,[t(k-1),t(k)], [x;P(:)]);
    65 xminus = Zo(end,1:n).';
    Pminus = reshape(Zo(end,n+1:end),n,n);

    %update
    A = chol(Pminus,'lower');
    70 Xminus = repmat(xminus,1,d) + sqrt(c)*[zeros(n,1),A,-A];
    Yminus = h(Xminus);
    mu = Yminus*w_m;
    S = Yminus*W*Yminus.' + R;
    C = Xminus*W*Yminus.';
    K = C/S;
    x = xminus + K*(z(k,:).' - mu);
    75 P = Pminus - K*S*K.';
end

80 function dZ = hybridUKF_deq(T,Z)
    %unpackaged input
    xi = Z(1:n);
    Pi = reshape(Z(n+1:end),n,n);

    85 %form sigma-point matrix
    A = chol(Pi);
    Xi = repmat(xi,1,d) + sqrt(c)*[zeros(n,1),A,-A];

    %integrate
    90 fXi = f(T,Xi);
    dx_i = fXi*w_m;
    dPi = Xi*W*fXi.' + fXi*W*Xi.';

    %repackage
    95 dZ = [dx_i;dPi(:)];
end

end

```

Listing A.9: Direct computation of the Lomb-Scargle periodogram.

```

function [f,P,prob] = LombScargle(t,h,ofac,hifac)
%LombScargle - Compute normalized Lomb-Scargle Periodogram
%
% INPUTS
5 % t (N x 1) Array of times for N observations
% h (N x 1) Array of observations
% ofac Frequency oversampling rate
% hifac Highest frequency to consider (multiple of Nyquist frequency)
%
10 % OUTPUTS
% f Array of M frequencies sampled
% P Array of M periodogram values
% prob Array of M probabilities of null hypothesis false alarm

15 %condition input
t = t(:);
h = h(:);

%sample length and time span
20 N = length(h);
T = max(t) - min(t);

%mean and variance
mu = mean(h);
25 s2 = var(h);

%calculate sampling frequencies
f = (1/(T*ofac):1/(T*ofac):hifac*N/(2*T)).';

30 %angular frequencies and constant offsets
w = 2*pi*f;
tau = atan2(sum(sin(2*w*t.'),2),sum(cos(2*w*t.'),2))/(2*w);

%spectral power
35 cterm = cos(w*t.' - repmat(w.*tau,1,length(t)));
sterm = sin(w*t.' - repmat(w.*tau,1,length(t)));
P = (sum(cterm*diag(h-mu),2).^2./sum(cterm.^2,2) + ...
      sum(sterm*diag(h-mu),2).^2./sum(sterm.^2,2))/(2*s2);

40 %estimate of the number of independent frequencies
M=2*length(f)/ofac;

%statistical significane of power
prob = M*exp(-P);
45 inds = prob > 0.01;
prob(inds) = 1-(1-exp(-P(inds))).^M;

```

Appendix B

Mission Timeline Encoding

At each step (every new target observation) of a mission simulation, the following values are recorded:

1. Target number (based on arbitrary ordering of the whole target list)
2. Arrival time (seconds after mission start)
3. Spacecraft mass on target arrival (kg)
4. Detection status (-2 = False Positive, -1 = False Negative, 0 = True Negative, 1 = True Positive)
5. Detection integration time (s)
6. Detection stationkeeping ΔV (m/s)
7. Detection stationkeeping fuel mass (kg)
8. Planet sky angle at time of detection (mas)
9. Planet Δmag at time of detection
10. First distance spectrum status (boolean: 0 = Failure, 1 = Success)*
11. First distance spectrum integration time (s)
12. First distance spectrum stationkeeping ΔV (m/s)
13. First distance spectrum stationkeeping fuel mass (kg)
14. Characterization slew ΔV (m/s)[†]
15. Characterization slew fuel mass (kg)

*First distance refers to the first separation distance for MDOs. For coronagraphs, this is the only spectrum taken.

[†]Only for MDOs - refers to the slew between the different separation distances.

16. Second distance spectrum success (boolean)
17. Second distance spectrum integration time (s)
18. Second distance spectrum stationkeeping ΔV (m/s)
19. Second distance spectrum stationkeeping fuel mass (kg)
20. Slew time to next target (s) [‡]
21. Slew ΔV to next target (m/s)
22. Slew fuel mass to next target (kg)
23. Slew angle to next target(rad)
24. Lateral disturbance force at first distance separation (N)
25. Axial disturbance force at first distance separation (N)
26. Lateral disturbance force at second distance separation (N)
27. Axial disturbance force at second distance separation (N)

[‡]Constant for coronagraphs.

Appendix C

Default Simulation Settings

```
% processing parameters
'FAP',0.01/1000;...           false alarm probability
'MDP',0.001;...               missed detection probability
'Nruns',100;...                number of monte carlo runs used
'minComp',0.1;...              minimum completeness for targets
'intCutoff',50;...             integration cutoff (days)
'maxSlewTime',30;...           maximum slewing time (days)
'defBurnPortion',0.05;...      default burn portion
'coeffs',ones(6,1)/6;...       Cost function coefficients
'missionStart',23.75;...       Mission start date, years after 1991.25
%instrument parameters
'dr',0.001;...                 dark rate
'Npix',14.3;...                # of noise pixels
'Ndark',10;...                  number of dark frames used
'lambda',500;...                detection wavelength (nm)
'sigma_r',3;...                  read noise (e/s)
't_exp',1000;...                exposure time per read (s)
'SNchar',11;...                  SNR for characterization
'deltaLambda',100;...            bandpass (nm)
'QE',0.5;...                   detection quantum efficiency
'QEchar',0.5;...                characterization quantum efficiency
'eta2',0.57;...                 post-supression system attenuation
'pupilArea',4*pi;...            entrance pupil area (m^2)
'throughput',0.5;...            supression system throughput
'haveOcculter',false;...         boolean
'occulterSep',55000;...          occulter-telescope distance (km)
'occulterSep2',35000;...         2nd occulter separation (km) (if MDO)
'throughput2',0.5;...           throughput at 2nd separation (if MDO)
'Rspec', 70;...                  spectral resolving power 1/Delta 1
'specLambda',760;...             spectral wavelength of interest
'shapeFac',pi/4;...              shapeFac*diameter^2 = Area
'contrast',1e-10;...             designed suppression level
```

```

'pixelArea',1e-10;...           pixel size (m^2)
'focalLength',240;...          (m)
'telescopeKeepout',45;...       degrees
%orbit parameters
'a_min',0.7;...                 semi-major axis min (AU)
'a_max',1.5;...                 semi-major axis max (AU)
'e_min',0;...                   eccentricity min
'e_max',0.35;...                eccentricity max
'p_min',0.26;...                albedo min
'p_max',0.26;...                albedo max
'conf',false;...                 orbits confined to [a_min a_max] range (boolean)
%target list filters (booleans)
'cS1',false;...                  1. Life Expectancy
'cS2',true;...                   2. Main Sequence
'cS3',false;...                  3. No Indication of Youth
'cS4',false;...                  4. No Indication of Low Metallicity
'cS5',false;...                  5. No Indication of Intrinsic Variability
'cS6',false;...                  6. Not a Halo Star
'cS7',false;...                  7. F,G,K only
'cE1',false;...                  1. Data Quality
'cE2',true;...                   2. Binary (or bright star within 10 as)
'cE3',7;...                      3. Max Apparent Visual Magnitude (integer)
'IWA',75/1000;...                inner working angle (as)
'dMagLim',26;...                 limiting delta mag
'haveTargList',false;...         target list has been calculated
%system properties
'dImax',15;...                  maximum mutual inclination (degrees)
'maxNplan',1;...                 maximum # of planets per system
'exozodi',1.5;...                exozodi level in zodi (geometric mean)
'exozodiVar',0;...                exozodi variation (variance of log normal dist)
%spacecraft parameters
'mission_life',5;...             primary mission lifetime (years)
'extended_life',0;...             extended mission (years)
'thrust',450;...                 maneuvering spacecraft slew thrust (mN)
'initMass',6000;...              maneuvering spacecraft initial (wet) mass (kg)
'dryMass',3400;...               maneuvering spacecraft dry mass (kg)
'coMass',5800;...                non-maneuvering spacecraft mass (kg)
'slewIsp',4160;...               specific impulse for slewing (s)
'skIsp',220;...                  specific impulse for stationkeeping (s)
'orbitSize',5.1e5;...            azimuthal size of telescope orbit (km)
'dutyCycle',1;...                 observation duty cycle
'missionPortion',1;...          portion of mission devoted to planets

```