THE UNIVERSITY OF CHICAGO

ACCURATE MODELING OF GRAZING TRANSITS USING UMBRELLA SAMPLING

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF ASTRONOMY AND ASTROPHYSICS

BY

GREGORY J. GILBERT

CHICAGO, ILLINOIS
DECEMBER 2021

To my students, who inspire me to keep learning

Nature study will show you

how full of beautiful and wonderful things

God has made the world

— Lord Robert Baden Powell

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

My time at the University of Chicago was set against a backdrop of persistent health problems, and so I must first and foremost thanks the friends, family, and colleagues who helped me through a difficult time in my life. Chief among these people was my advisor, Dan Fabrycky, who was an unfailing source of kindness, patience, and support. Throughout personal challenges, studying astrophysics remained a grounding endeavor for me, and so I am especially thankful to the many students, researchers, faculty, and staff who kept me engaged both personally and professionally during my time way from the university, particularly Megan Bedell, Laura Kreidberg, Mike Gladders, Laticia Rebeles, John Carlstrom, Emily Gilbert, Nora Bailey, and David Martin. Of my many wonderful friends outside the academy, I am particularly grateful to Doug Foote, Brian Roberston, and Spencer Toder, who kept me going though dark times. And of course I will always be thankful for my father, who has supported me unconditionally.

I thank the many, many incredible scientists I have had the pleasure of knowing and working with these past several years. There are too many of you to list by name, though I would like to highlight my thesis committee members, Fred Ciesla, Andrey Kravstov, and Leslie Rogers, as well as Dan Foreman-Mackey of the Center for Computational Astronomy.

Modern astrophysics is a collaborative effort, and this dissertation made extensive use of data products from the Kepler mission hosted on the NASA Exoplanet Archive. This research also made use of computational resources provided by the University of Chicago Research Computing Center. Computer code was developed using many open source libraries.

ABSTRACT

Grazing transits present a special problem for statistical studies of exoplanets. Even though grazing planetary orbits are rare (due to geometric selection effects), for many low to moderate signal-to-noise cases, a significant fraction of the posterior distribution is nonetheless consistent with a grazing geometry. A failure to accurately model grazing transits can therefore lead to biased inferences even for cases where the planet is not actually on a grazing trajectory. With recent advances in stellar characterization, the limiting factor for many scientific applications is now the quality of available transit fits themselves, and so the time is ripe to revisit the transit fitting problem. In this paper, we model exoplanet transits using a novel application of umbrella sampling and a geometry-dependent parameter basis that minimizes covariances between transit parameters. Our technique splits the transit fitting problem into independent Monte Carlo sampling runs for the grazing, non-grazing, and transition regions of the parameter space, which we then recombine into a single joint posterior probability distribution using a robust weighting scheme. Our method can be trivially parallelized and so requires no increase in the wall clock time needed for computations. Most importantly, our method produces accurate estimates of exoplanet properties for both grazing and non-grazing orbits, yielding more robust results than standard methods for many common star-planet configurations.

CHAPTER 1

INTRODUCTION

Roughly 75% of all known exoplanets have been discovered via transit surveys, giving transiting planets an outsized influence on our ability to constrain both exoplanet demographics and planet formation models. In most cases, our best answers remain data-limited: the precision of available transit measurements is insufficient to distinguish between theoretical models. Until recently, uncertainties on transit parameters were dominated by uncertainties on stellar quantities (especially stellar radii), but with recent advances in stellar characterization via *Gaia* DR2 astrometry [Gaia Collaboration et al., 2018, Berger et al., 2018] and via California Kepler Survey spectroscopy [Petigura et al., 2017, Johnson et al., 2017], the achievable precision on planetary radii and orbital elements derived from transit surveys is now limited predominantly by the quality of the transit fits themselves [Petigura, 2020].

The keystone transit parameter which must be accurately estimated in order to allow accurate estimates of all underlying planet properties is the impact parameter. Unfortunately, impact parameters are notoriously difficult to estimate and consequently have seldom been the focus of transit lightcurve analyses. Instead, most previous studies have focused on measuring transit depths and durations, both of which are usually well-constrained by observations [e.g. Mullally et al., 2015, Thompson et al., 2018]. Still, the limiting factor when converting transit observables into planetary radii, inclinations, and eccentricities is more often than not the least well constrained variable, so we must address the challenge of measuring impact parameters head-on. The problem is vital because transit observables can rarely be translated into planet properties on a one-to-one basis, and, furthermore, doing so usually requires at least some knowledge of the transit impact parameter. For planetary radii, this requirement arises because transit depths contain information about stellar limb darkening in addition to information about planetary radius, while for inclinations and eccentricities the requirement arises because transit durations contain information about transit chord

length as well as information about orbital velocity. In all cases then, accurately deriving planet parameters from transit observations requires knowing how far from the stellar center the planet transits, which is precisely what the impact parameter measures. The methods developed in this paper thus first and foremost represent a means of obtaining the highest quality impact parameter measurements possible as a stepping stone toward obtaining correspondingly high quality estimates of planetary radii and other orbital elements.

One particularly pressing open question in exoplanet science is whether the so-called radius valley is fully or only partially depleted of planets [Fulton et al., 2017, Fulton and Petigura, 2018, Van Eylen et al., 2018, Hardegree-Ullman et al., 2020, Petigura, 2020]. Answering this question demands a sharper picture of the exoplanet period-radius distribution, which can only be developed by making better measurements of planetary radii, which in turn depend on estimates of transit impact parameters. Better empirical constraints on the statistics of planets in and around the radius valley will illuminate how planets evolve over their lifetimes, whether by photoevaporation [Owen and Wu, 2017], by core-powered mass loss [Ginzburg et al., 2018, Gupta and Schlichting, 2019], or by some combination of mechanisms [Neil and Rogers, 2020].

In addition to revealing the properties of individual planets, better estimates of individual exoplanet properties will almost immediately lead to a better understanding of exoplanet demographics and exo-system architectures. Although 1D distributions of exoplanet sizes, spacings, inclinations, and eccentricities have been analyzed in detail [Howard et al., 2012, Fabrycky et al., 2014, Winn and Fabrycky, 2015, Millholland et al., 2017, Weiss et al., 2018, Mills et al., 2019], parameter uncertainties have thus far precluded a joint analysis which synthesizes all available data into a single coherent picture that captures trends both within and between systems. The methods for performing this sort of synthesis have already begun to be developed [He et al., 2019, Gilbert and Fabrycky, 2020], but progress has been stymied by insufficiently precise estimates of planet properties for a suitably large sample

of objects. As soon as such measurements become available, many of their implications for exoplanet demographics and exo-system architectures will almost immediately become apparent because we will be able to analyze the new data using existing, already mature techniques.

Because high-precision impact parameter measurements are rare, dynamical analyses of exo-systems have often relied on measurements of transit durations [e.g. Fabrycky et al., 2014] or transit duration variations [e.g. Hamann et al., 2019, Millholland et al., 2021] in order to make inferences into the structure of dynamically active systems. While transit durations and their variations present a rich source of information [Dawson, 2020], any conclusion reached via consideration of these quantities will implicitly depend on assumptions about the impact parameter (due to the aforementioned degeneracy between orbital speed and transit chord length). When stellar parameter uncertainties are large, the role of transit parameter degeneracies will be obscured and we can safely marginalize over uncertainties introduced by the transit model. But with estimates of stellar properties becoming more refined every day, we can no longer overlook the interaction between transit durations and impact parameters —we must consider all quantities simultaneously and in detail.

If we wish to make reliable transit measurements over the full range of exo-system architectures, we must overcome a trick of geometry that leads to strong degeneracies between transit parameters for planets on grazing trajectories. In particular, although the planet-to-star radius ratio, $r_p/R_\star$, and the normalized impact parameter, b, are largely uncorrelated for non-grazing orbits, these two parameters become highly correlated for grazing geometries. Because grazing orbits are rare [Kipping and Sandford, 2016], the standard approach has been to reject suspected grazing or near-grazing bodies from statistical studies all together [e.g. Petigura, 2020]. However, for many low to moderate signal-to-noise cases, a significant fraction of the posterior distribution is nonetheless consistent with a grazing geometry, even for planets which most likely orbit on non-grazing trajectories. A failure to accurately model

grazing transits can therefore lead to biased inferences even for cases where the planet is not actually on a grazing orbit. Although the grazing degeneracy has been known about for some time [Rowe et al., 2014, 2015], the severity of the problem as it pertains to near-grazing orbits - i.e. orbits with $0.7 \lesssim b < 1 - r_p/R_\star$, as compared to grazing orbits or non-grazing orbits - has only recently begun to be appreciated. Because planetary orbits are (probably) isotropically oriented (formally, $\cos i$ is isotropically distributed), the unavoidable conclusion is that as many as one third of all existing transit measurements may be corrupted by incomplete consideration of grazing transit geometries. The true situation is probably not so dire, but without a reliable way to distinguish grazing from near-grazing or non-grazing orbits, it is difficult to know which transit measurements should be trusted.

The goal of this paper is twofold: first, to find a way to reliably identify grazing and near-grazing transits, and second, to accurately fit these transits using a method that produces unbiased estimates of a exoplanet properties. The tool for the job is umbrella sampling [Torrie and Valleau, 1977], a statistical technique which is closely related to importance sampling. Although the application of umbrella sampling is standard practice is the field of molecular dynamics where it originated, umbrella sampling methods have only recently begun to be applied to astrophysical problems [Matthews et al., 2018]. Umbrella sampling is a powerful tool for sampling multimodal and other complicated posterior distributions and is suitable for many astrophysical applications. So, in addition to addressing the specific problem of transit lightcurve fitting, we aim for this paper to serve as an accessible introduction to umbrella sampling for astronomers unfamiliar with this fruitful technique. Our present work was primarily inspired by Matthews et al. [2018], which also stemmed from the twin motivations of applying umbrella sampling to an astrophysical problem (in their case, sampling the low-probability tails of distributions in order to compare cosmological models) and introducing umbrella sampling to astronomers in a pedagogically accessible manner.

CHAPTER 2

THE GEOMETRY OF GRAZING TRANSITS

In this section, we review a few salient aspects of transit lightcurve modeling relevant to the analysis of grazing transits. A full pedagogical introduction to the geometry of transit lightcurves is presented in Winn [2010].

2.1 The transit model

A transit lightcurve can be fully specified by just five parameters when the following assumptions are met: (1) the planet's orbit is circular, (2) the stellar disk has uniform surface brightness, (3) the planet is opaque, (4) $m_p << M_\star$, so that the small planet approximation holds, (5) the starlight is from a single, unblended star, and (6) at least two transits are observed such that the planet's orbital period can be determined [Seager and Mallén-Ornelas, 2003]. In practice, conditions 3-6 are met under most ordinary circumstances. Conditions 1 and 2 in general are not met - real planets often have eccentric orbits and real stars typically have non-negligible limb darkening - but techniques for modifying the simple five parameter model to account for these complications are well developed [Mandel and Agol, 2002, Ford et al., 2008, Dawson and Johnson, 2012, Kipping, 2013, Luger et al., 2019]. Regardless, most real exoplanet transit lightcurves possess a shape similar enough to the idealized case that an examination of the simple five parameter model is instructive.

The observables which can be directly recovered from a single transit light curve are the mid-transit time, t_0, the transit depth, δ, the transit duration, T, and the ingress/egress timescale, τ (see Figure 2.1). When multiple transits are observed, the orbital period, P, can be inferred as well. From these five observables, one can derive four physical quantities: the planet-to-star radius ratio, $r \equiv r_p/R_\star$, the scaled separation, $a/R_\star$, the mean stellar density, $\rho_\star$, and the normalized impact parameter, b, defined as

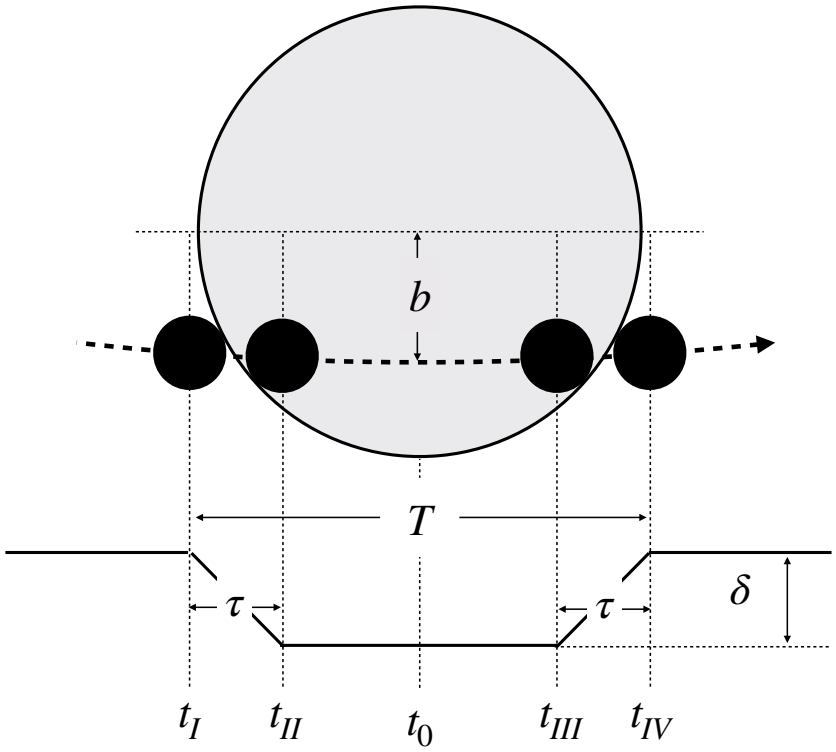

Figure 2.1: Schematic illustration of a non-grazing transit geometry with the corresponding lightcurve approximated as a trapezoid. The transit depth, δ; transit duration, T; ingress/egress duration τ; mid-transit time, t_0; impact parameter, b; and 1$^{\rm st}$ through 4$^{\rm th}$ contact points, t_I - t_{IV}, are indicated. Note that for this study, the transit duration, T, will always refer to the full first-to-fourth contact duration, T_{14}, unless otherwise specified because this is the only duration which is defined for all grazing and non-grazing geometries. The approximation $\tau_{12} = \tau_{34}$ (i.e. ingress and egress timescales are equal) is valid as long as eccentricities are not very large. Figure adapted from Winn [2010].

$$b = \frac{a\cos i}{R_\star}\left(\frac{1-e^2}{1+e\sin\omega}\right) \tag{2.1}$$

where i is the sky projected inclination, e is the orbital eccentricity and ω is the argument of periastron [Seager and Mallén-Ornelas, 2003, Winn, 2010].

When fitting a model to photometric data, the lightcurve can be described by any non-degenerate combination of the five observables $\{P, t_0, \delta, T, \tau\}$ and the four physical quantities $\{r, a/R_\star, b, \rho_\star\}$ or any derivable quantity thereof (e.g. $\cos i$, the sky-projected inclination). These five parameters constitute the model basis set, and choosing which parameters are

best for a given problem presents a surprisingly difficult challenge. In practice, P, t_0, and δ, are usually well constrained by the data, so most reasonable basis sets will include P, t_0, and either δ or r (or slight modifications of these quantities) as the first three free parameters. The transit duration, T, is usually well constrained by the data as well, but for even modestly noisy lightcurves, τ is difficult to resolve, making the selection of the final basis parameter a non-trivial task. Moreover, because $\tau \approx 5$ min is shorter than the observing cadence in many cases ($\Delta t \approx 30$ min for Kepler long cadence data and TESS primary mission full frame images), data binning often precludes a precise ingress/egress characterization even for high signal-to-noise transits [Kipping, 2010, Price and Rogers, 2014]. The ratio T/τ is as important as the values of T and τ in isolation, so the choices of these final two basis parameters must be considered in tandem. A straightforward approach is to use the basis pair $\{T, \tau\}$ directly, although one might just as easily choose $\{b, \rho_\star\}$, $\{T_{14}, T_{23}\}$ or even more exotic pairs such as $\{b^2, 1/T\}$. Numerous basis sets have been proposed in the literature [e.g. Bakos et al., 2007, Carter et al., 2008, Pál, 2008], although none has yet been adopted as the standard set applied in all cases. This lack of a standard basis set and the wide variety of different parameterizations in use speaks to the subtle challenge of transit model fitting. Often the motivating factor for choosing one basis set over another is the degree to which model parameters are mutually orthogonal. While state-of-the-art computational tools can usually handle moderate - or even strong - covariances between parameters [Neal, 2011, Foreman-Mackey et al., 2013], in many circumstances the best tactic for improving performance is still to reparameterize a model into an orthogonal basis. This is no easy task, as the optimal choice of basis set is often context dependent. To make matters worse, it is often not obvious when the "right" basis set has been found, and selecting the "wrong" basis set can lead to inefficient exploration of the posterior distribution that nevertheless appears to appears to produce well mixed posterior chains. As we will demonstrate below, these complications can even propagate through to quantities such as r which are usually

assumed to be robust against changes in model parameterization.

Throughout this study, T will refer to the full first-to-fourth contact transit duration T_{14} unless otherwise noted, and this will be the transit duration we adopt throughout. In practice, however, the center-to-center contact transit duration, T_{c-c}, sometimes called the 1.5 to 3.5 contact is usually better constrained by the data and is preferred as a model parameter over T_{14}. We have chosen to adopt T_{14} here because it is the only transit duration which is readily defined for all grazing and non-grazing geometries. One could alternatively use the full width half max transit duration, $T_{\rm FWHM}$, which can be defined in relation to the transit depth, δ, for strongly for strongly grazing geometries ($b > 1$) where T_{c-c} is undefined. However, this parameterization would necessitate numerical determination of the lightcurve and introduce additional covariances, negating a primary motivation for our parameterization efforts. In §4.2 will discuss how to modify our proposed transit modeling technique in order to incorporate different duration parameters. For the time being, we limit consideration to T_{14}, which simplifies our discussion considerably and allows us to focus on the ideas which are unique to this paper.

In addition to adopting five of the geometric and physical basis parameters discussed above, most transit lightcurve models will also need to include some way to handle both nonzero limb darkening and nonzero eccentricity. The former necessitates the introduction of at least one new model parameter, whereas the latter can be accomplished without any new parameters at all.

The standard approach for handling nonzero limb darkening is to employ a quadratic limb darkening law [Mandel and Agol, 2002, Kipping, 2013]. Although other limb darkening laws exist in the literature [Claret, 2000, Claret and Bloemen, 2011], the quality of available photometric data rarely justifies using more than two limb darkening parameters. For some applications one may wish to impose priors on the limb darkening coefficients with constraints derived from physical laws, but the more general approach is to marginalize over

uncertain limb darkening using uninformative priors [Kipping, 2013]. The effect of nonzero limb darkening on the transit shape is to "round out" the bottom of the transit and add curvature to the ingress and egress. While the inclusion of limb darkening is often necessary, the added flexibility in an already under-constrained problem can make inferring the ingress/egress duration - and thus the impact parameter and other related quantities - even more difficult. Changes in stellar surface brightness becomes most extreme near the limb of the star, making inferences of transit parameters for grazing or near-grazing transits highly sensitive to changes in the stellar limb darkening profile. Throughout this paper, we adopt the uninformative quadratic limb darkening treatment advocated for by Kipping [2013].

The standard approach for handling nonzero eccentricity is to employ the so-called photo-eccentric effect [Ford et al., 2008, Dawson and Johnson, 2012]. A planet on an eccentric orbit will travel with a variable velocity, and depending on the orientation of the orbit's periastron, the planet's velocity during transit may be either sped up or slowed relative to the expectation for a circular orbit with the same semimajor axis. An increase (decrease) in velocity translates into a shortened (lengthened) transit duration, so by comparing the ratio of the observed transit duration to the predicted circular transit duration, the orbital eccentricity can be recovered. The situation is complicated somewhat by the interaction of the periastron and the orbital inclination, and depending on the orientation of the orbit, an eccentric orbit may produce a higher or lower impact parameter relative to that of a circular orbit. Once again, impact parameter plays a central role in transit modeling. Predicting the circular transit duration requires an independent estimate of the stellar density, and so in practice we usually estimate orbital eccentricity by comparing the transit constrained stellar density obtained from fitting a circular transit model, $\rho_{\rm circ}$, to our reference density derived from independent observations, $\rho_{\rm obs}$. The pertinent equation [Ford et al., 2008, Dawson and Johnson, 2012] is

$$\frac{\rho_{\rm circ}}{\rho_{\rm obs}} = \Big(\frac{1+e\sin\omega}{\sqrt{1-e^2}}\Big)^3 \tag{2.2}$$

where e is the eccentricity and ω is the argument of periastron. Although there are confounding factors which may alter the measured transit duration [Kipping, 2014], nonzero eccentricity is usually the dominant factor. Tracking eccentricity via the photoeccentric effect is in fact often a better approach than directly incorporating the variables $\{e, \omega\}$ into the transit model because only one of the two eccentricity vector components ($e\sin\omega$) is constrained by primary transit observations, while the other ($e\cos\omega$) is not, at least to low order in e. Eccentricity can thus be seamlessly integrated into the transit model without introducing any new model parameters.

With all of this in mind, we will take as our fiducial basis set the parameters $\{P, t_0, \log r, b, \log T, q_1, q_2\}$. The first five variables are drawn from the geometric and physical transit parameters described above, where the logarithm on r and T enforces positivity of the two scale parameters and facilitates sampling over multiple orders of magnitude; the variables q_1 and q_2 are the quadratic limb darkening coefficients from Kipping [2013], defined as

$$q_1 = (u_1 + u_2)^2 \tag{2.3}$$

$$q_2 = \frac{u_1}{2(u_1+u_2)} \tag{2.4}$$

where u_1 and u_2 are the standard quadratic limb darkening coefficients [Mandel and Agol, 2002, Claret, 2000]. As we will see shortly, this basis set performs quite well for non-grazing transits but performs poorly for grazing transits due to an emergent degeneracy between r and b in the grazing regime.

2.2 Model degeneracy in the grazing regime

We will now consider the specific challenges that arise when modeling the transit lightcurves of exoplanets on grazing trajectories. In the discussion that follows, the term "grazing" refers to any transit geometry for which the planetary disk does not fully overlap the stellar disk at the mid-transit point. In other words, we consider a transit to be non-grazing if $b \leq 1 - r$ and grazing if $b > 1 - r$.

Before examining the mathematics of grazing transits, let us first examine the qualitative features of grazing vs non-grazing transits. Figure 2.2 illustrates how the transit shape changes as the planet's trajectory moves from the non-grazing regime into the grazing regime. At higher impact parameters, the transit chord is shortened, reducing the transit duration. Furthermore, at higher impact parameters, the planet crosses a dimmer region of the limb-darkened stellar disk, reducing the transit depth as well. The change in transit shape is gradual for $0 \leq b \lesssim 1 - 2r$, then changes rapidly as the planet enters the near-grazing transition region. When the planet crosses the grazing boundary at $b = 1 - r$, the transit switches from U-shaped to V-shaped. For high signal-to-noise cases, this change in morphology can be used to distinguish between grazing and non-grazing transits. For lower signal-to-noise cases, however, there is enough flexibility in the model that the shape of the transit is ambiguous and so it is unclear whether a planet is on a grazing or non-grazing trajectory. In particularly troublesome cases, this ambiguity can lead to a strong degeneracy between r and b.

If the shape of ingress/egress can be resolved or the mean stellar density is known to high precision, then the transit lightcurve of a small planet on a center-crossing orbit ($b = 0$) can be easily distinguished from that of a larger planet on a grazing or near-grazing orbit ($b \approx 1$). Resolving ingress/egress from a single transit requires a high cadence to adequately sample the ingress/egress shape. However, many targets have been observed only in Kepler long cadence mode or TESS full frame images, with $\tau_{\rm obs} \approx 30$ min considerably longer

than the typical ingress/egress timescale, $\tau \approx 5$ min. One can in principle sidestep the binning issue and resolve the ingress/egress shape by observing multiple transits, but this approach requires a precise estimate of the orbital ephemeris which can be obscured by the presence of uncertain transit timing variations. So, data binning can still inhibit accurate characterization of the transit shape, even for high signal-to-noise cases [Kipping, 2010, Price and Rogers, 2014].

For a typical stellar density precision and a typical photometric noise level, the transit shape is ambiguous and the basis set $\{r, b, T\}$ is flexible enough to accommodate a wide range of transit shapes. These various complications make impact parameters extremely difficult to measure, and inconsistencies become apparent upon comparing results between different Kepler data releases. Figure 2.3 plots the impact parameters reported by Kepler DR22 [Mullally et al., 2015] against those reported by Kepler DR25 [Thompson et al., 2018]. Each point on the figure marks an individual Kepler Object of Interest (KOI) reported in both catalogs and so points should cluster around the line $b_{22} = b_{25}$. The high degree of scatter indicates that the actual results are inconsistent, despite the fact that they were obtained from nearly identical input observations and similar data processing pipelines. Although there is some evident correspondence between the catalogs for high impact parameters ($b \gtrsim 0.7$), there is nonetheless a substantial fraction of points which report $b \approx 0$ in one catalog and $b \approx 1$ in the other. Even the 1D single-catalog distributions exhibit inhomogeneity, with a pile-up of reported impact parameters near $b = 0$ seen in both catalogs. In fact, the pile-up is most severe in the more recent catalog. Because constraints on r and b inferred from transit lightcurves are correlated for stars with non-negligible limb darkening, any mismeasurement of impact parameter will propagate through to a mismeasurement of planetary radius as well.

A full derivation of the analytic equations governing transit lightcurves for both grazing and non-grazing geometries is presented in Mandel and Agol [2002], which we summarize

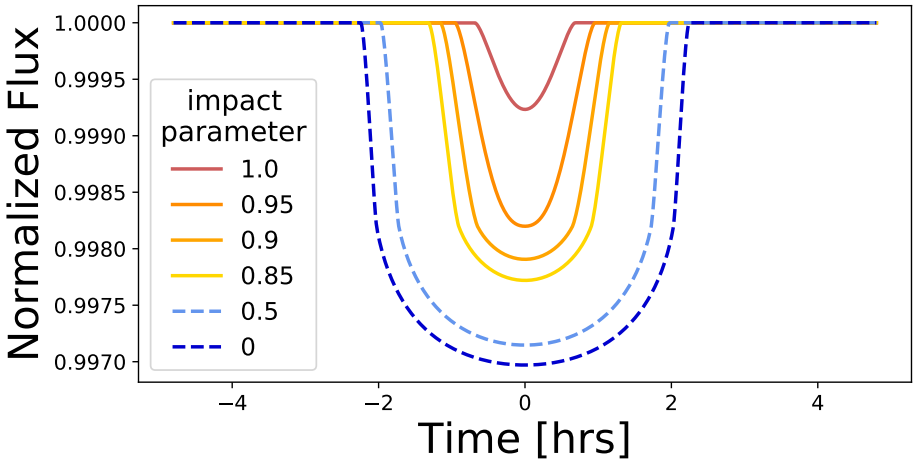

Figure 2.2: Lightcurve model illustrating how the transit shape changes as a function on impact parameter, b. This model is for a warm Neptune orbiting a Sun-like star with model parameters $r = 0.05$, $P = 13.0$ days, $R_\star = R_\odot$, $M_\star = M_\odot$, $u = (0.4, 0.25)$. Warm colored, solid lines indicate grazing or near-grazing transits (informally defined here as $b \gtrsim 0.8$), while cool-colored, dashed lines indicate non-grazing transits. As b increases, transit duration decreases (due to the shorter transit chord) and transit depth decreases (due to stellar limb darkening). There is little change in the transit shape between $0 \leq b \lesssim 0.5$, making differences between low impact parameters difficult to resolve. The transit depth and duration both change more rapidly above $b \gtrsim 0.5$. At $b = 1 - r = 0.95$, the transit morphology switches from "U-shaped" to "V-shaped", providing a diagnostic avenue for distinguishing grazing from non-grazing transits, although the effects of limb darkening blur this transition somewhat.

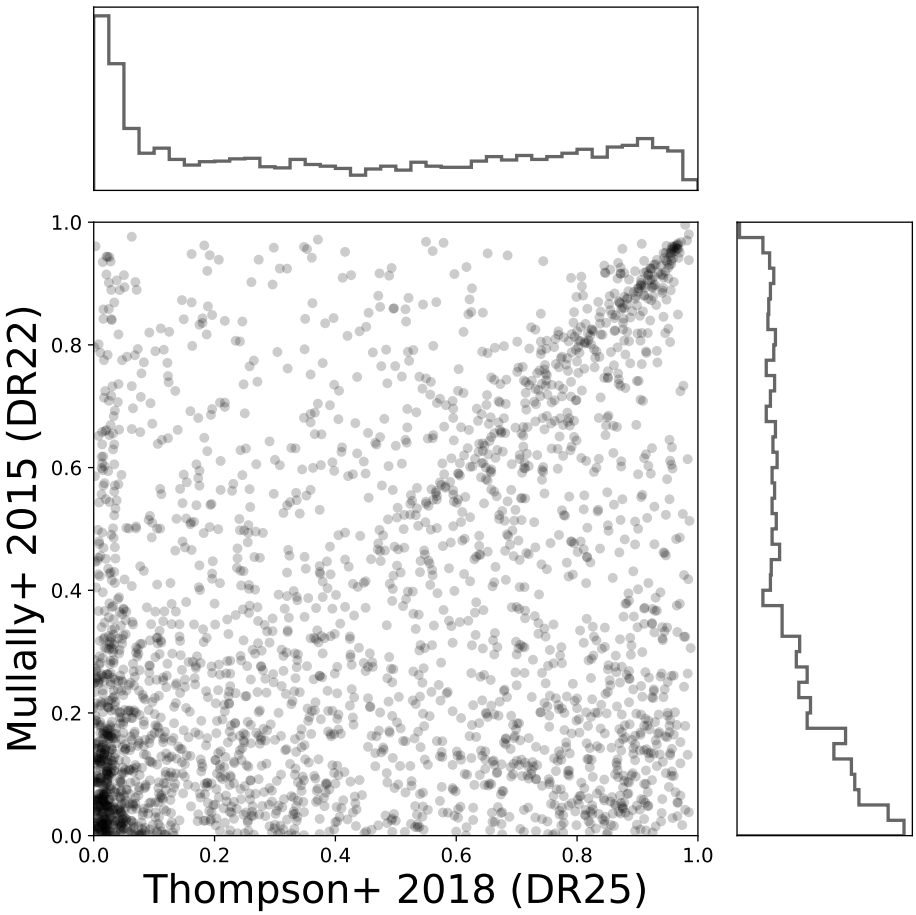

Figure 2.3: Impact parameters reported by Kepler DR22 [Mullally et al., 2015] vs DR25 [Thompson et al., 2018]. Plotted values correspond to best-fit point estimates. Each point indicates a single Kepler Object of Interest reported in both catalogs, so points should cluster around the line $b_{22} = b_{25}$. The high degree of scatter in the actual data indicates that results are inconsistent and therefore unreliable. While there is some correspondence of values where $b \gtrsim 0.7$, in a substantial fraction of cases one catalog reports $b \approx 0$ while the other reports $b \approx 1$. A pile-up of reported values near $b = 0$ can be seen in both catalogs, indicating that results are inaccurate. Because r and b are correlated for stars with non-negligible limb darkening, any mismeasurement of b will propagate through to a mismeasurement of r.

here. For a uniform (i.e. zero limb darkening) stellar source, the fractional change in flux during transit is

$$\Lambda(r,z) = \begin{cases} \frac{1}{\pi}\left[r^2\kappa_0 + \kappa_1 - \frac{1}{2}\zeta\right] & 1-r < z < 1+r \\ r^2 & 0 \leq z \leq 1-r \end{cases} \tag{2.5}$$

where $z = z(t)$ is the is the normalized separation of centers and

$$\begin{aligned} \kappa_0 &= \cos^{-1}[(r^2+z^2-1)/2rz] \\ \kappa_1 &= \cos^{-1}[(1-r^2+z^2)/2z] \\ \zeta &= \sqrt{z^2-(1+z^2-r^2)^2} \end{aligned} \tag{2.6}$$

Making the substitution $z(t_0) \to b$ at mid-transit time t_0 gives the transit depth, $\delta = \Lambda(r,b)$. For non-grazing trajectories, Λ is independent of b, whereas for grazing ones Λ has a complicated dependence on b and r. It is this complicated geometry that leads to parameter degeneracies and considerable difficulties when modeling grazing transits. The key problem is that the covariance properties of the transit parameters change drastically in the grazing regime compared to the not-grazing regime. More specifically, whereas b and r are nearly uncorrelated for non-grazing geometries, they become highly correlated for grazing ones. The upshot is that a basis set which performs well in the non-grazing regime may perform quite poorly in the grazing regime, and vice versa.

In the grazing regime, as a planet moves to higher b with fixed r, the overlap area between the stellar and planetary disk will decrease, thereby reducing the transit depth. However, if r is allowed to float as a free parameter, a large b can be compensated for by a commensurate increase in r. Thus r and b become almost perfectly correlated and sometimes a sampler will find an extremely large radius ($r_p \gg R_\star$) and extremely high impact parameter ($b \gg 1$), which is obviously unphysical. This is a well known problem in transit fitting [Rowe et al., 2014, 2015] and is a clear case where common sense is in conflict with the analysis.

Unfortunately, the degeneracy problem is not isolated to a few easily identified cases, but instead affects any planet for which a grazing geometry can not be conclusively ruled out.

The $r - b$ degeneracy can be readily seen upon inspection of real Kepler data. Figure 2.4 demonstrates that compared to isotropic expectations for the cumulative KOI catalog,[1] there is an overabundance of supergiant planets ($r_p \gtrsim 2R_J$) found on grazing trajectories. For objects with astrophysically sensible radii, the fraction of planets inferred to be on grazing trajectories is roughly in line with expectations, but for implausibly large super-giant planets, nearly every KOI has a quoted impact parameter consistent with a grazing transit. Furthermore, a number of these grazing super-giants cluster at the $b = 1 + r$ boundary that marks where planets are not only on grazing orbits, but on *extremely* grazing orbits for which the planetary and stellar disks barely overlap at all. While it is possible there is some complicated selection effect at play wherein only super-giants on grazing orbits pass all vetting thresholds necessary to be included in the cumulative KOI catalog, the simpler explanation is that the majority of these supposed super-giants are actually super-Earths or mini-Neptunes on non-grazing orbits, with inferred r and b values that are artifacts of a transit fitting procedure gone awry.

2.3 Experiments using simulated data

In order to illuminate the origin of the skewed $r-b$ distribution seen in real data, we perform a straightforward experiment which applies a Markov Chain Monte Carlo (MCMC) model fitting routine to synthetic data. In the next several paragraphs, we describe our method for simulating data and for subsequently fitting a transit model to that data using Hamiltonian Monte Carlo [HMC; Neal, 2011]. The casual reader may wish to skim these paragraphs so as not to become bogged down in the details. The important point is that we simulate an ordinary transit of an unremarkable star-planet system and then model that transit using

1. `https://exoplanetarchive.ipac.caltech.edu`

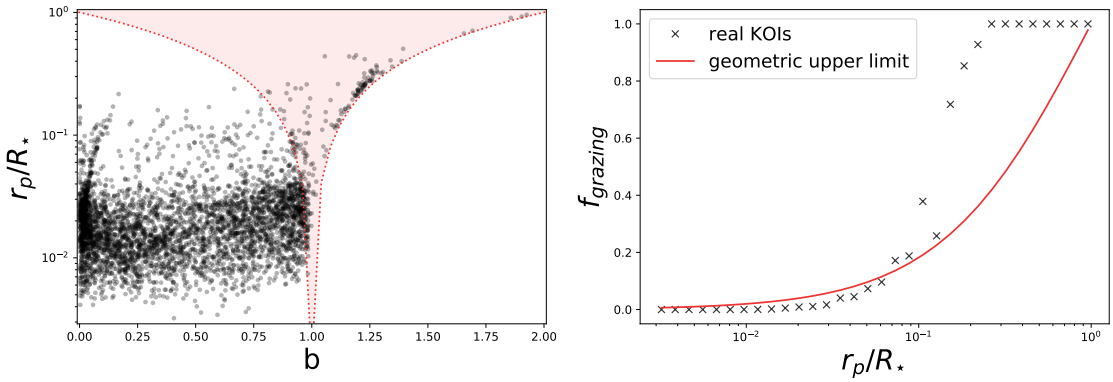

Figure 2.4: Distribution of $r_p/R_\star$ and b for cumulative Kepler Object of Interest (KOI) planet candidates. Data were downloaded from the NASA Exoplanet Archive (https://exoplanetarchive.ipac.caltech.edu) on 18 July 2021. *Left panel*: joint 2D distribution of $r_p/R_\star$ and b. Each point represents an individual validated or candidate planet. The red shaded region highlights planets inferred to orbit on a grazing trajectory. There is a suspicious pile-up of planets near $b \approx 1.25$ at the $b = 1 - r$ boundary, hinting that the radius and impact parameter measurements derived for these planets may not be reliable. Non-isotropic structure in the distribution among planets on non-grazing trajectories - particularly near $b = 0$ - suggests that measurements for these planets should be approached with some skepticism as well. *Right panel*: fraction of planets inferred to be on grazing trajectories as a function of radius ratio. The red line plots the relation $f = 2r/(1+r)$, which is the geometric upper limit on how many planets are expected to be on grazing orbits, ignoring any reduced detection efficiencies for grazing transits. For planets with physically plausible radii ($r_p \lesssim 2R_J$), the observed fraction of grazing transits is in line with expectations, but for super-giant planets an overabundance of KOIs are found on grazing trajectories, again suggesting that their radius and impact parameter measurements may be unreliable.

our fiducial basis set and standard Monte Carlo sampling techniques.

The details of our data simulation procedure are as follows. First, we simulate a low signal-to-noise transit of a warm Jupiter on a near-grazing, $P = 13$ day circular orbit around a Sun-like star. The period was calculated in order to create a 3 hr transit duration for an impact parameter $b = 0.85$. We generated 500 data points between $t_0 \pm T$, each with an integrated exposure time of 14.4 minutes (0.01 days). The finite data points were spaced randomly over the interval in order to minimize aliasing artifacts that might arise from a uniform observing cadence. We then added $\sigma_F/F = 10^4$ ppm Gaussian noise to the data. We did not include any long term trends or correlated noise in our simulation. Ground-truth parameter values are collected as simulation J-85 in Table 2.1, and the simulated photometry is shown in the middle panel of Figure 2.5.

We parameterized the model using our fiducial basis set - $\{P, t_0, \log r, b, \log T, q_1, q_2\}$ - plus a baseline flux offset, F_0, and a jitter term, $\log \sigma_F$. We held P fixed at the true value and placed uninformative priors on all other variables, the mathematical details of which are collected in Table 2.2. Fixing P is equivalent to assuming that the planet's ephemeris is tightly constrained, which is often the case even for noisy transits. Placing uninformative priors on $\log T$ is roughly equivalent to placing uninformative priors on both $\rho_\star$ and e. Although in practice most applications will use the best available stellar characterization to place at least a modestly informative prior on $\rho_\star$ (and, indirectly, on T and e via the photoeccentric effect), for our present experiment we are more concerned with the sampler behavior (i.e. whether MCMC chains are well behaved) rather than the posterior inferences. Our philosophy is that the model should converge regardless of any particular choice of prior, so we adopt minimally restrictive priors wherever possible.

We sampled from the posterior distribution using HMC as implemented by `PyMC3` [Salvatier et al., 2016] and the No U-Turn Sampler [NUTS; Hoffman and Gelman, 2011]. Each sampling run consisted of two independent chains tuned for 5000 steps and sampled for 1000

Parameter	Unit	J-22	J-85	J-100	SE	MN	MHZ
Star							
$R_\star$	$R_\odot$	1.0	1.0	1.0	0.92	0.92	0.37
$M_\star$	$M_\odot$	1.0	1.0	1.0	0.86	0.86	0.38
u_1	-	0.40	0.40	0.40	0.48	0.48	0.46
u_2	-	0.25	0.25	0.25	0.22	0.22	0.28
σ_F	ppm	1×10^4	1×10^4	5×10^3	300	300	200
Planet							
P	days	3.6	13.0	44.9	21.0	21.0	37.0
r_p	$R_\oplus$	11.2	11.2	11.2	1.3	2.2	0.38
b	-	0.22	0.85	1.00	0.70	0.98	0.70
T	hrs	3.0	3.0	3.0	3.24	1.26	2.33
Derived							
r	-	0.103	0.103	0.103	0.012	0.020	0.009
γ	-	7.57	1.36	0.0	25.2	2.48	32.7
λ	-	0.091	0.026	0.011	0.004	0.001	0.003

Table 2.1: Ground-truth parameter values for simulated lightcurves used to compare a standard sampling approach to our new method. The quantities λ and γ are defined in Equation 3.1. All simulated planets were placed on circular orbits. The first set of simulations (J-22, J-85, & J-100) placed a Jupiter-sized planet around a Sun-like star at three different impact parameters in order to produce a non-grazing ($b = 0.22$), nearly grazing ($b = 0.85$), and grazing ($b = 1.00$) geometry; the orbital period was scaled to preserve a circular orbit for a consistent transit duration $T = 3$ hrs. The second pair of simulations (SE, MN), placed a super-Earth ($r_p = 1.6R_\oplus$) and a mini-Neptune ($r_p = 2.2R_\oplus$ on a 21 day orbit around a K star; this experiment was designed to mimic the detection of radius valley planet that will require a precise impact parameter measurement in order to determine the its composition. The final simulation (MHZ) placed a small rocky planet in the habitable zone of an M-dwarf.

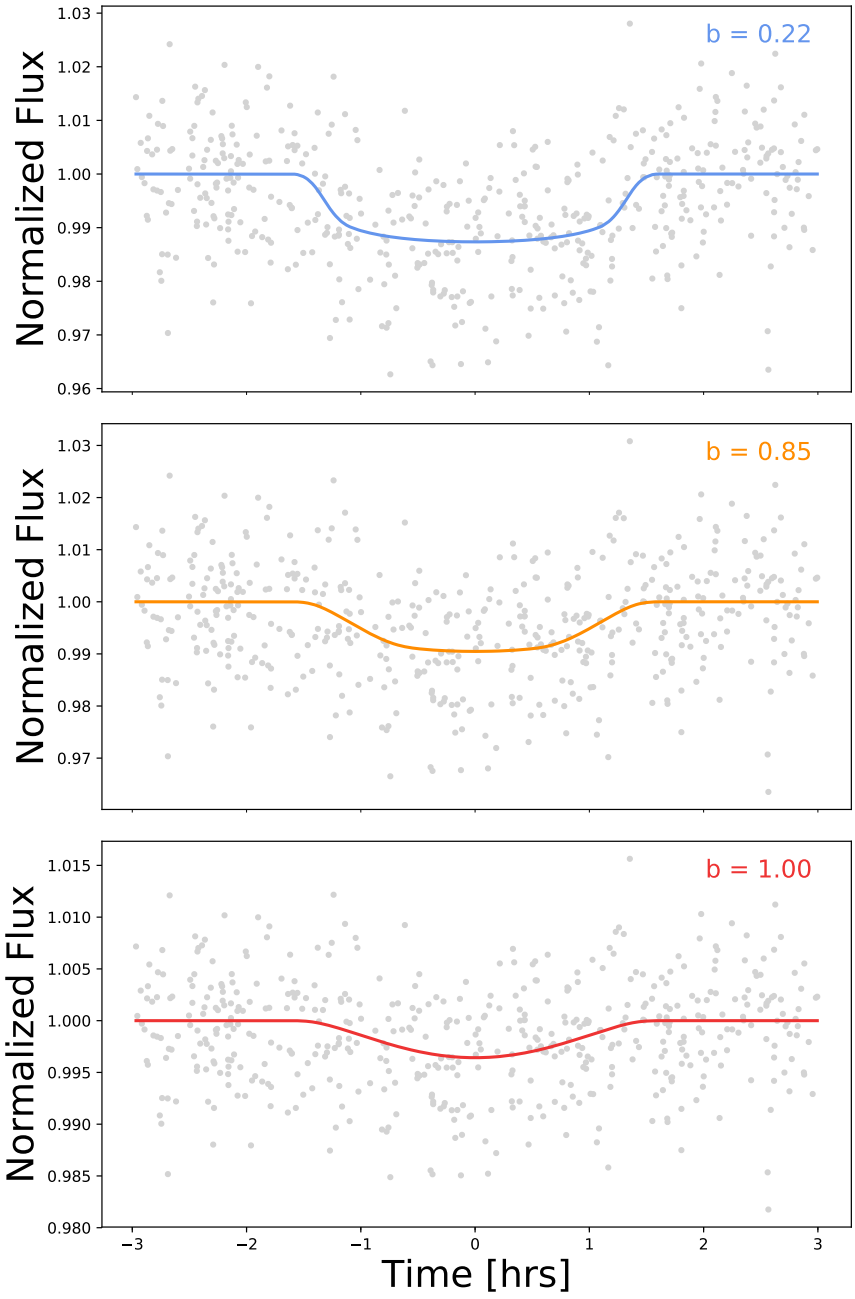

Figure 2.5: Simulated lightcurve photometry for a Jupiter-size planet on a circular orbit around a Sun-like star with various impact parameters. The orbital period was adjusted to maintain a consistent transit duration of $T = 3$ hrs. The solid colored lines show the true underlying model while the grey points have additive Gaussian noise. The low signal-to-noise ratio of the transit makes the orbital trajectory (grazing vs. non-grazing) ambiguous. *Top panel*: $b = 0.22$, placing the planet on a non-grazing trajectory, corresponding to model J-22 in Table 2.1. *Middle panel*: $b = 0.85$, a near-grazing trajectory (model J-85). *Bottom panel*: $b = 1.0$, a grazing trajectory (model J-100).

Parameter	Prior
P	*fixed*
t_0	$\mathcal{N}(0.0, 2.4)$
$\log r$	$\mathcal{U}(-4, -0.02)$
b	$\mathcal{U}(1-r, 1+r)$
$\log T$	$\mathcal{U}(-0.6, 0.8)$
q_1, q_2	$\mathcal{U}(0, 1)$
F_0	$\mathcal{N}(0, 1)$
$\log \sigma_F$	$\mathcal{N}(0, 1)$
$\rho_\star$	*see text*
e	*see text*

Table 2.2: Model priors. Normal distributions $\mathcal{N} = \mathcal{N}(\mu, \sigma)$ and uniform distributions $\mathcal{U} = \mathcal{U}(min, max)$. Note that b is defined as a conditional distribution predicated on r, i.e. $f(b) \equiv f(b|r)$. The limb darkening treatment of Kipping [2013] places uninformative priors on the two quadratic limb darkening coefficients.

draws, for a total of 2000 samples per run. We deliberately left the independent chains short in order to highlight the stochastic nature of the problem, but note that with HMC the autocorrelation length is typically much shorter than for standard random walk Metropolis-Hastings algorithms [Metropolis et al., 1953, Hastings, 1970], so that the number of effective samples is usually $\gtrsim 25\%$ and under ideal circumstances can approach 100%. This high effective sample rate is achievable with HMC because the algorithm adds a "momentum" term to the proposal generation which enables much larger steps sizes compared to a random walk. During the tuning phase (analagous to the burn-in phase of other MCMC routines), the sampler "learns" the posterior geometry and adaptively selects an optimal steps size for efficient exploration of the posterior. While the computational cost per step is higher for HMC compared to random walk Metropolis-Hastings, the cost per effective sample is usually considerably lower, especially for high dimensional problems. HMC has only recently begun to gain popularity among astrophysicists, so we direct the interested reader to the excellent review by Betancourt [2017], as well as tutorials for the Python software packages `PyMC3`[2]

2. `https://docs.pymc.io`

[Salvatier et al., 2016] and `exoplanet`[3] [Foreman-Mackey et al., 2021b,a].

Figure 2.6 illustrates results of four independent attempts to model the simulated transit data using HMC. The only difference from run-to-run was the random number seed for the sampler. Despite identical setups, each run produced a remarkably different posterior distribution, sometimes getting stuck in the grazing regime and sometimes failing to explore that regime altogether. The issue is not merely that the chains had not converged, and even increasing the length of the sampling and/or tuning phase by orders of magnitude did not reliably produce consistent results. The root of the problem is the complicated geometry, with two "sticking points" for the sampler. One sticking point is at the grazing to non-grazing transition, $b = 1 - r$, and the other is in the tails of the distribution at high b and large r. Because standard sampling methods cannot be counted on to adequately sample both the grazing and non-grazing portions of the distribution, our inferences are unreliable, and we must find a new method for modeling exoplanet transit lightcurves.

The problem is two-fold. First, we need to use a different basis set for grazing vs. non-grazing geometries because the covariance properties of r and b are quite different between the two regimes. Second, we need to find a way to efficiently explore the full posterior space without getting stuck at the grazing transition boundary. The solutions to these problems are interrelated and are discussed in the next two sections of this paper.

3. `https://docs.exoplanet.codes`

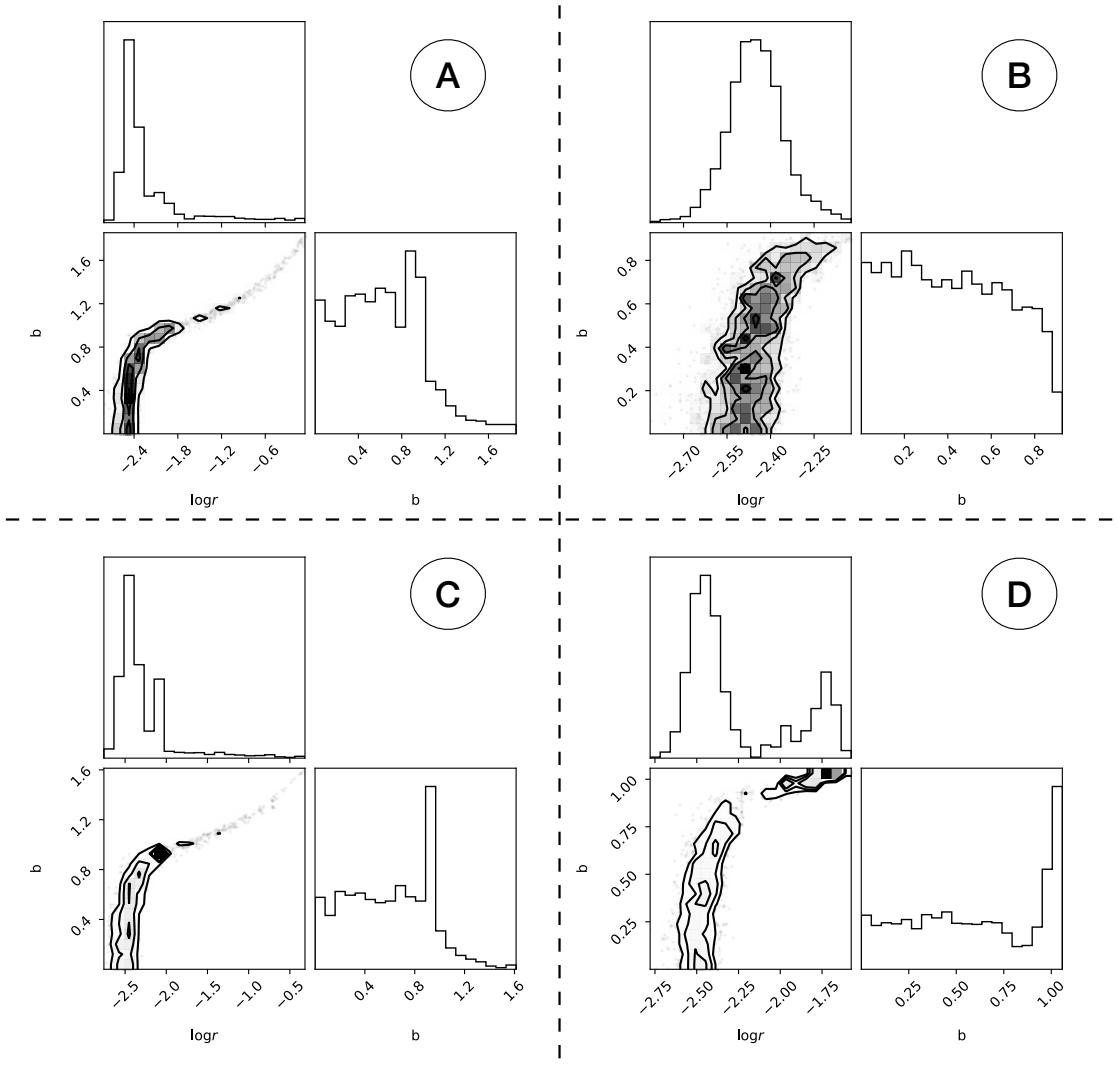

Figure 2.6: Corner plots of the posteriors from four identical MCMC runs of model J-85 using our fiducial $\{\log r, b\}$ basis. See §3.2 and Table 2.1 for details of the model setup. The only difference between the runs was the random seed for the Markov chains. Despite their identical setups, each run produces a remarkably different posterior geometry. *Panel A*: The sampler appears to fully explore the posterior region, with most samples consistent with a non-grazing geometry and a smaller fraction extending into the grazing regime. There is a "dog leg" feature at $b \approx 1$ where the geometry transitions from non-grazing to grazing, and there is a strong degeneracy between $\log r$ and b for grazing transits. *Panel B*: The sampler fails to explore the grazing regime entirely, giving the illusion of a well-behaved posterior. *Panel C*: The sampler extends to high impact parameters, but catches at the boundary between grazing and non-grazing geometries, producing a sharp spike at $b \approx 1$. *Panel D*: The samples pile up at $b \approx 1$, leading to a bimodal posterior distribution that barely explores the grazing regime at all. Increasing the length the tuning phase and/or the sampling phase does not reliably fix these issues.

CHAPTER 3

A NEW BASIS FOR GRAZING TRANSITS

Our solution to the grazing transit problem is to split the Monte Carlo sampling into separate runs for the grazing ($b > 1 - r$) and non-grazing ($b < 1 - r$) regimes. We then combine these independent runs into a single posterior distribution using umbrella sampling (see §4). Before we describe our full umbrella sampling routine, we first present a new basis set which is designed for optimal performance in the grazing regime.

3.1 Specification of the model parameters

Of the seven parameters in our fiducial basis set - $\{P, t_0, \log r, b, \log T, q_1, q_2\}$ - four can be carried over to our new grazing basis without modification: P, t_0, q_1, and q_2. Both P and t_0 are generally tightly constrained by the data and are minimally covariant with the other parameters, and the two uninformative limb darkening coefficients, q_1 and q_2 [Kipping, 2013], perform well for both grazing and non-grazing orbits. Only r, b, and T now remain. The transit duration, T, is usually well constrained by the data (albeit somewhat less so than P and t_0) and is closely related to the eccentricity via the photoeccentric effect; we therefore maintain $\log T$ as one of our seven basis parameters. With five parameters in common between the fiducial non-grazing basis and our new grazing basis, our reparameterization effort now hinges on a transformation of r and b (which are highly covariant for grazing transits) into a new parameter pair which is more nearly orthogonal for grazing geometries. Rather than producing new parameters wholesale, our strategy is to find some mapping of $\{r, b\} \rightarrow \{x_1, x_2\}$ with the desired orthogonality when $b > 1 - r$.

After some experimentation, we identified a suitable pair of quantities, which we define according to the non-linear combination

$$\begin{aligned} \lambda &= r^2 + \beta r \\ \gamma &= \frac{\beta}{r} \end{aligned} \tag{3.1}$$

where $\beta \equiv 1 - b$ is a convenience variable. Because both r and b are unitless, λ and γ are unitless as well.

The first quantity, λ, is derived from a linear approximation to the area of overlap between two spheres (Equation 2.5). Thus, λ is closely related to the transit depth in the grazing regime. But, note that because λ ranges over $(0,\, 2r^2)$ for grazing transits, the relation is closer to $\lambda \approx 2\delta$ than to $\lambda \approx \delta$. We caution the reader <u>not</u> to use λ as a basis parameter outside of the grazing regime because it is explicitly tied to the geometry of grazing transits. Figure 3.1 demonstrates that the exact $\Lambda(r, b)$ curve is well matched by a simple linear function $\lambda(b)$ at fixed r as long as $r < 1$, which will virtually always be the case for exoplanets orbiting main sequence or giant branch stars. We have not rigorously checked how the validity of our assumptions break down when $r \geq 1$, and so the results in this paper will likely need to be adjusted if they are to be applied to substellar companions of brown dwarfs [Jung et al., 2018] or white dwarfs [Vanderburg et al., 2020]. A full derivation of $\lambda(r, b)$ from $\Lambda(r, b)$ is given in the appendix.

The second quantity, γ, indicates the extent to which a transit is grazing or non-grazing, with transition occurring at $\gamma = 1$. When $\gamma \geq 1$, the transit is non-grazing; when $-1 < \gamma < 1$, the transit is grazing; when $\gamma \leq -1$ the planet does not transit at all. We refer to γ as the ***grazing coordinate***, and in §4 we will see that it plays a special role in our umbrella sampling routine.

When converting from one basis to another, care must taken in order to avoid inadvertently introducing unwanted priors. For a thorough discussion of the implicit priors introduced by our reparameterization and for a recipe to establish sensible prior distributions for λ and γ, see the appendix. The important point is that in addition to mapping

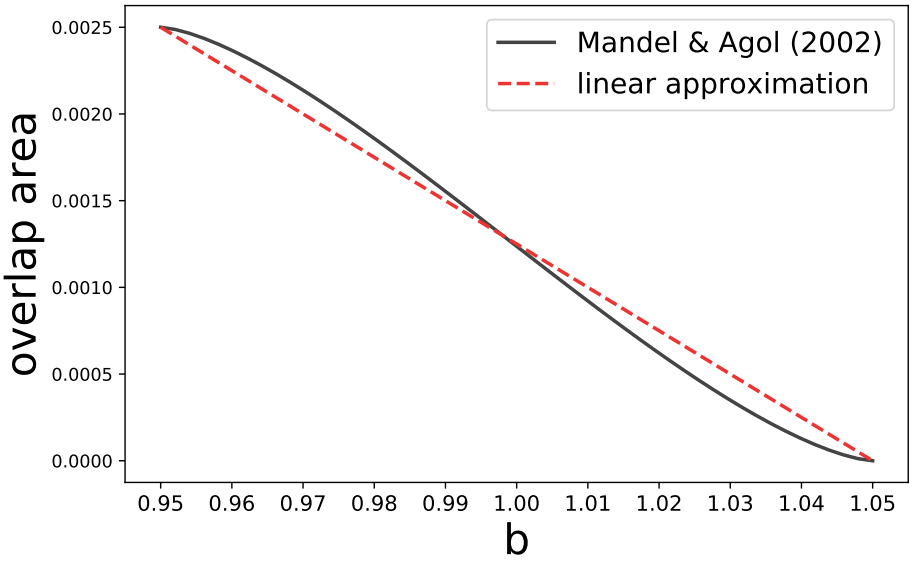

Figure 3.1: Planet-star overlap area at mid-transit as a function of impact parameter, b, for a star with zero limb darkening and a planet-to-star radius ratio $r = 0.05$. The plot is restricted to show only the grazing regime, i.e. $1 - r < b < 1 + r$. The solid black line shows the exact geometric solution derived by Mandel and Agol [2002] and presented in this paper as Equation 2.5. Even though the full geometry is quite complicated, the resultant curve is well approximated by a simple linear function (dashed red line).

$\{r, b\} \rightarrow \{\lambda, \gamma\}$, we add additional terms to the log-likelihood function as needed to ensure that our priors remain consistent between parameterizations.

By construction, our new $\{\lambda, \gamma\}$ basis is far more orthogonal than the fiducial $\{r, b\}$ basis is for grazing transits. Conversely, $\{\lambda, \gamma\}$ is far *less* orthogonal than $\{r, b\}$ is for non-grazing transits. To achieve good sampler performance, we must therefore make the restriction $\gamma < 1$ when using our new basis and the restriction $\gamma \geq 1$ when using the old basis. We stress that our new $\{\lambda, \gamma\}$ parameterization is specifically designed with grazing transits in mind and should not be applied to non-grazing geometries.

For many transits, some fraction of the posterior distribution will be consistent with both grazing and non-grazing trajectories, so fitting a transit will require at least two independent sampling runs, one to sample the grazing regime using $\{\lambda, \gamma\}$ and the other to sample the non-grazing regime using $\{r, b\}$. Recombining independent posterior chains into a single posterior distribution can be performed using the statistical technique of umbrella sampling, which will be introduced in §4. For now, we will restrict our analysis to consideration of the grazing regime in order to compare the relative performance of the two basis sets.

3.2 Performance of the $\{r, b\}$ vs $\{\lambda, \gamma\}$ basis

The fiducial basis set we have used thusfar is $\{P, t_0, \log r, b, \log T, q_1, q_2\}$, which we now compare to our new parameterization $\{P, t_0, \log \lambda, \gamma, \log T, q_1, q_2\}$. As a shorthand, we will continue to refer to these as the $\{r, b\}$ and $\{\lambda, \gamma\}$ bases, respectively, although any actual sampling will always be performed using $\log r$ and $\log \lambda$ in place of r or λ.

To compare the two basis sets, we simulate a low signal-to-noise transit of a warm-Jupiter orbiting a Sun-like star and sample from the posteriors using Hamiltonian Monte Carlo. Simulated photometry is shown in Figure 2.5 and ground truth parameter values are presented in Table 2.1. Our model setup and sampling routine both follow the procedure described in §2.2, modified to restrict samples to grazing geometries. For each basis set, we

perform 100 independent MCMC runs using two chains run for 5000 tuning steps and 1000 draws, generating 2000 samples per run. Rather than merely setting a hard boundary at the grazing transition, we added a biasing potential, ψ, to the likelihood such that

$$\psi(\gamma) = \begin{cases} 1+\gamma & \gamma \leq 0 \\ 1-\gamma & 0 < \gamma < 1 \end{cases} \tag{3.2}$$

which has the effect of preferentially biasing posterior samples toward the middle of the grazing regime. The term is related to umbrella sampling, and the motivation behind its inclusion will become apparent in §4.

Our new $\{\lambda, \gamma\}$ basis performs more efficiently and more than the standard $\{r, b\}$ basis and produces consistent posterior distributions. For a simulated near-grazing transit (model J-85), the total runtime for a given run using $\{r, b\}$ was 389±34 seconds, compared to 386±26 seconds using $\{\lambda, \gamma\}$, a nearly identical wall clock time. On the balance, the autocorrelation length of the chains was a little longer when using the new basis compared to the standard basis, resulting in a larger number of effective samples obtained using our the $\{\lambda, \gamma\}$ basis compared to the standard $\{r, b\}$ basis. Evaluated using the autocorrelation length for r, the time per effect sample was 1.6 seconds using $\{\lambda, \gamma\}$ vs 2.3 seconds using $\{r, b\}$, a 29% gain in efficiency. We repeated this autocorrelation analysis using posterior chains for b and T finding a gains in efficiency of 28% and 4%, respectively. when using our new basis. The relative performances of the two bases was comparable for various other simulated transit geometries (see Table 2.1), typically producing a $\sim$ 20% gain in efficiency for generating effective samples of r and b and a roughly equivalent efficiencies for generating effective samples of T. The two bases produce consistent posterior distributions (Figure 3.2). We conclude that our new $\{\lambda, \gamma\}$ basis will be preferred under most circumstances.

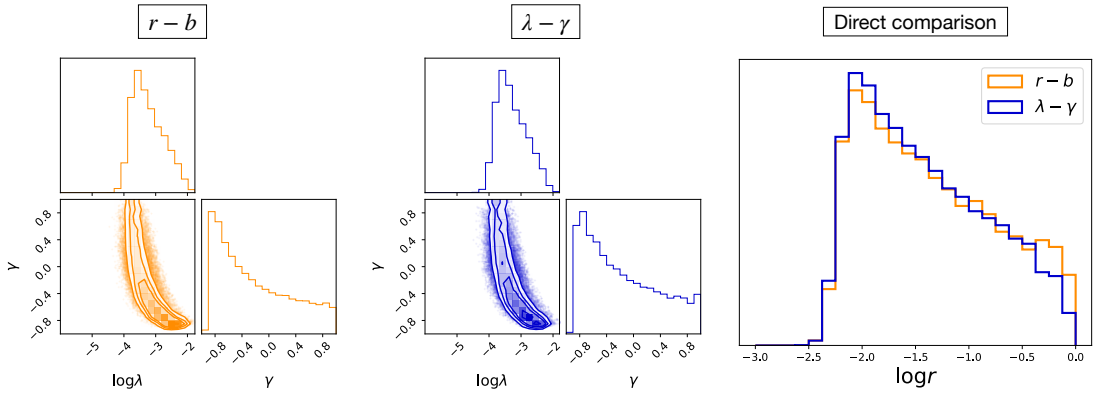

Figure 3.2: Posterior distributions of $\log \lambda$, γ, and $\log r$ for transit geometries restricted to the grazing regime for a near-grazing transit of a warm Jupiter orbiting a Sun-like star (model J-85). The effects of the biasing potential have been removed. See Table 2.1 for ground truth simulation parameters and Figure 2.5 for the simulated photometry. The two parameterizations produce comparable posterior distributions, although the new $\lambda - \gamma$ basis performs $\sim 20\%$ more efficiently. A detailed discussion of the model parameterization is presented in §3.

CHAPTER 4

UMBRELLA SAMPLING

Umbrella sampling [Torrie and Valleau, 1977] is a statistical tool designed for estimating complicated target distributions - e.g. multimodal distributions or degeneracy ridges - for which standard sampling techniques fail. Umbrella sampling does not replace existing sampling methods, but rather works in tandem with these methods to produce more robust posterior estimates. The basic idea is to split a complicated sampling problem into multiple smaller, more manageable problems, each restricted to a narrow region (or window, in the standard nomenclature) of parameter space. Samples are obtained separately from each window using whatever sampling technique the user prefers - e.g. Hamlitonian Monte Carlo [Neal, 2011, Salvatier et al., 2016] or ensemble sampling [Goodman and Weare, 2010, Foreman-Mackey et al., 2013] - after which the samples are recombined into a single joint posterior distribution.

Although umbrella sampling has rarely been applied to astrophysical problems, the technique is widely used in the field of molecular dynamics where it originated [Torrie and Valleau, 1977]. The literature on umbrella sampling is extensive, but because most examples of its use are presented in the context of highly technical chemical analyses, there exists a precipitous barrier to entry for many astronomers (this paper's author included) who lack the domain expertise to easily comprehend the specialized scientific content surrounding the general statistical tool we wish to adopt. One goal of this paper is therefore to present an accessible, high-level introduction to umbrella sampling tailored toward the needs of astronomers in order to establish a gateway into the wider umbrella sampling literature. For a more rigorous introduction, we direct the interested reader to a recent review of umbrella sampling by Kästner [2011], as well as to the first astrophysical application of by Matthews et al. [2018]. Indeed, much of the pedagogy in this section was borrowed from Matthews et al. [2018] - particularly their §2.1 - and any astronomer wishing to implement umbrella sampling them-

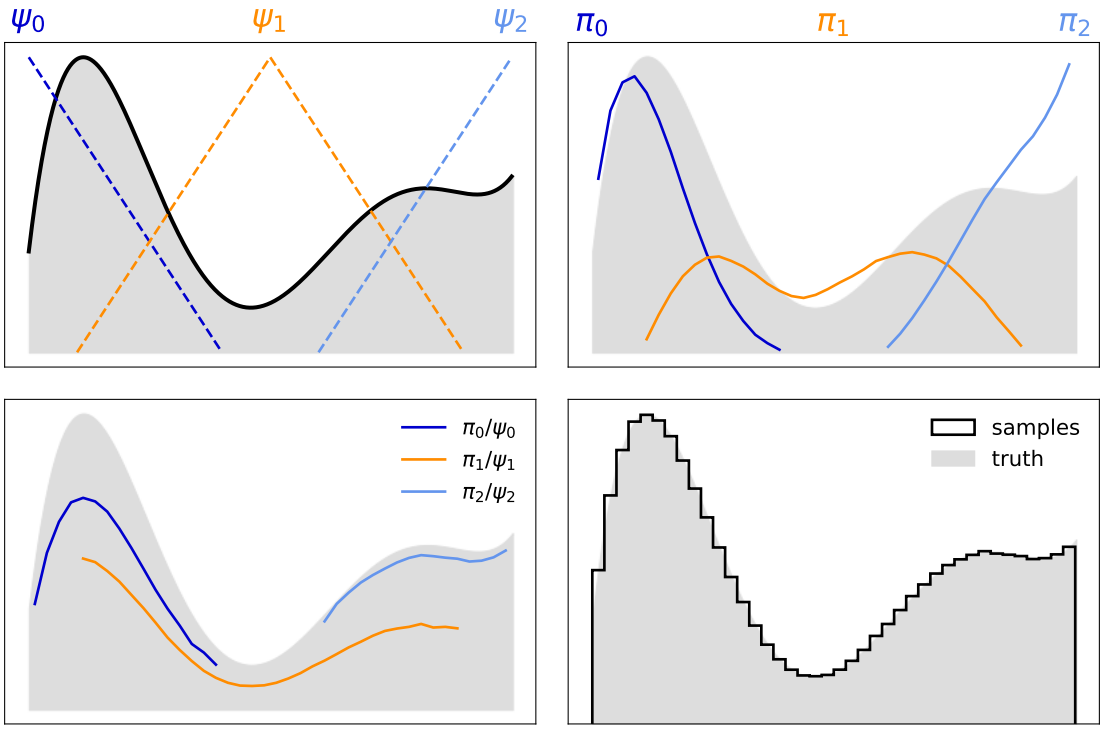

Figure 4.1: Schematic illustration of the umbrella sampling method. The method is designed to facilitate sampling from multi-modal target distributions (shaded grey region in all panels). *Top left*: The target distribution (emphasized with a thick black line) has low-probability valley which will create a bottleneck for standard sampling techniques. In order to ensure sampling from the full posterior space, we split the problem into three windows, each assigned a bias function, ψ_i. *Top right*: After sampling independently from each window, we have three biased sub-distributions π_i. *Bottom left*: Removing the effect of the bias functions, ψ_i produces three unbiased sub-distributions with unknown offsets between one another. *Bottom right*: calculating the window weights, z_i, and recombining all sub-distributions π_i into a single joint posterior recovers the true target distribution. See §4.1 for a detailed discussion. A tutorial for reproducing this plot by implementing umbrella sampling can be found at `https://gjgilbert.github.io/tutorials/umbrella_sampling/`.

selves will likely benefit from reviewing that paper in tandem with the present manuscript. Because learning to use new mathematical tools is often best accomplished through a "hands-on" approach, we have also developed a Python tutorial for implementing umbrella sampling, available at `https://gjgilbert.github.io/tutorials/umbrella_sampling/`.

4.1 A brief overview of umbrella sampling

Let us begin by assuming that we wish to sample from some arbitrary target distribution which possesses a complicated geometry (Figure 4.1). Standard sampling techniques will do a poor job at traversing the low probability "valleys" between high probability "peaks," resulting in poorly mixed posterior chains and incomplete sampling of the target distribution. One way around this issue is to add an additional bias term to the likelihood in order to "level out" the peaks and valleys, thereby simplifying the geometry. If the target distribution were known *a priori* (which of course it is not), we could add a single bias term to the entire distribution to make it flat throughout. In practice, however, the more viable approach is to break the complicated target distribution into several overlapping ***windows***, sample separately from each window, and then recombine the sub-samples into a joint posterior distribution. Each window will be assigned its own ***bias function***, sometimes called simply a ***bias*** or ***umbrella***. The bias functions serve to restrict the sampler to a given window and ensure that a significant fraction of samples are drawn from the low probability valleys.

Before defining our windows and biases, we must first identify a suitable variable, x, which we will use to construct a sampling framework. In the molecular dynamics literature, x is usually called the *reaction coordinate* because it corresponds to a real physical quantity related to chemical reactions such as free energy or molecular bond strength; in this manuscript we will refer to x as the ***umbrella coordinate*** (hence our earlier terminology for the grazing coordinate). The optimal choice of x will be dictated by the geometry of the target distribution. For example, if the target distribution consists of several isolated peaks,

x could be defined along the line connecting those peaks. Such detailed advance knowledge is not strictly necessary, however, and in many cases it is possible to select a good (though perhaps sub-optimal) umbrella coordinate even for a blind search. In any case, the prior information needed to identify a suitable umbrella coordinate is comparable to the prior information needed to properly specify the model in the first place, and the choice of umbrella coordinate should follow from the structure of the problem. For a more in-depth discussion of strategies for choosing umbrella coordinates, particularly under information-limited circumstances, see Matthews et al. [2018].

Once we have selected our umbrella coordinate, x, our next task is to define our window bounds and a set of N corresponding bias functions, $\psi_i(x)$ (Figure 4.1, top left panel). Once again, the optimal choice of windows and biases depends on the geometry, so the more that can be learned via exploratory analysis, the better. Fortunately, however, the results of umbrella sampling are insensitive to the particular choice of window bounds and bias functions provided that two conditions are met: (1) each window is adequately sampled, and (2) there is sufficient overlap between windows in order to allow accurate determination of relative window weights. We are thus free to define ψ_i in whatever manner is most convenient for the problem at hand.

With windows and biases defined, we now sample from the target distribution, $\pi(x)$ separately from each of the N windows, thereby producing N biased posterior sub-distributions $\pi_i(x)$ (Figure 4.1, top right panel). The sub-distributions relate to the (known) bias functions and to the (unknown) target distribution, via the equation

$$\pi_i(x) = \frac{1}{z_i}\psi_i(x)\pi(x) \tag{4.1}$$

where z_i are the window weights quantifying the relative contribution of each π_i to the combined target distribution, π. Because each π_i is a probability distribution, $\int \pi_i(x)dx = 1$, and the window weights z_i can be calculated via integration of Equation 4.1 as

$$z_i = \int \psi_i(x)\pi(x)dx = \langle \psi_i \rangle_\pi \tag{4.2}$$

where $\langle \psi_i \rangle_\pi$ denotes the average of some function f with respect to π. In other words, to determine z_i, we take the average of each ψ_i weighted by the empirically sampled target distribution, π. If the full target distribution π were known, calculating the window weights z_i would be trivial. But of course π is not known - it is precisely the quantity we are trying to determine! Furthermore, we don't actually have samples of π yet. Rather, we have N sets of biased sub-samples, π_i, meaning we will need to compute $\langle \psi_i \rangle_{\pi_j}$ for each (i, j) and then combine these to estimate $\langle \psi_i \rangle_\pi$. The challenge is that this final combination step depends on z, making the whole process a bit circular. Once the z_i are known, however, the biased sub-distributions, π_i, can be easily combined into a single joint posterior distribution, π (Figure 4.1, bottom panels).

Different methods for implementing umbrella sampling more or less come down to different strategies for solving the integral in Equation 4.2. The most popular method is the Weighted Histogram Analysis Method [WHAM; Kumar et al., 1992], which works by binning the data and computing a histogram in the overlap region. Another popular method is the Multistate Bennet Acceptance Ratio [MBAR; Shirts and Chodera, 2008], which does not require discretization of the data. Both WHAM and MBAR can be derived from maximum likelihood or minimum asymptotic variance principles (see the references above for proofs). Recently, Thiede et al. [2016] and Dinner et al. [2017] demonstrated that the determination of umbrella weights z_i can be recast as an eigenvector problem, a method which they term the Eigenvector Method for Umbrella Sampling (EMUS). Establishing umbrella sampling as an eigenvector problem has the twin advantages of being computationally efficient and facilitating accurate error analysis, and so we adopt EMUS as our method of choice here.

Following Matthews et al. [2018], we restate Equation 4.2 as an explicit sum

$$z_j = \sum_{i=1}^{N} \left\langle \frac{\psi_j(x)}{\sum_{k=1}^{N} \psi_k(x)/z_k} \right\rangle_{\pi_i} \tag{4.3}$$

where $\langle\rangle_{\pi_i}$ denotes an average with respect to π_i. Because the umbrella weights z_i enter the equation both on the left-hand side of the equation and in the denominator sum on the right-hand side, Equation 4.3 must be solved iteratively. To do so using EMUS, we first define a square *overlap matrix*, F, with each element (i, j) defined as

$$F_{ij} = \left\langle \frac{\psi_j/z_i}{\sum_{k=1}^{N} \psi_k/z_k} \right\rangle_{\pi_i} \tag{4.4}$$

As its name implies, F tracks the extent to which samples drawn within one window fall under the umbrella of any other window. On diagonal terms will usually have larger values (because all samples drawn from window i by construction fall under umbrella ψ_i), and when windows (i, j) do not overlap, $F_{ij} = F_{ji} = 0$.

In order to calculate z_i using linear algebra, we first define $z \equiv [z_1, z_2, ..., z_N]$ as a vector. Taking the product of z and the j^{th} column of F yields

$$\sum_{i=1}^{N} z_i F_{ij} = \sum_{i=1}^{N} \left\langle \frac{\psi_j}{\sum_{k=1}^{N} \psi_k/z_k} \right\rangle_{\pi_i} = \langle \psi_j \rangle_{\pi} \tag{4.5}$$

Recall from Equation 4.2 that $z_j = \langle \psi_j \rangle_{\pi}$, so $\sum_i z_i F_{ij} = z_j$. Considering all columns in F simultaneously yields the left eigenvalue problem

$$zF = z \tag{4.6}$$

which when solved provides an estimate of the window weights.

If we knew F *a priori*, finding the eigenvalues and eigenvectors of Equation 4.6 would be a straightforward application of linear algebra. But, in practice, we need to estimate both z and F from our empirical samples, π_i. As we noted earlier, this must be done iteratively.

Our strategy will be to pick a starting guess for z and calculate a first estimate of F from Equation 4.4. We'll then use our estimate of F to calculate an updated value for z using Equation 4.6, and then iterate between Equations 4.4 and 4.6 until the result converges. In practice the problem is often nearly converged after just one or two iterations, and both the final result and convergence rate are insensitive to the particular starting estimate of z.

In summary, the steps of umbrella sampling are: (1) choose a suitable umbrella coordinate x, (2) define windows and biases ψ_i, (3) sample from each window to produce biased sub-distributions π_i, (4) calculate window weights z_i by iteratively solving Equations 4.4 and 4.6, and finally (5) recombine all sub-samples into the joint posterior estimate π by inverting Equation 4.1. Note that unlike standard direct sampling methods which produce a single set a unweighted samples, umbrella sampling produces multiple sets of weighted samples (with weights given by z_i), and these weights must be taken into account when estimating posterior distributions or summary statistics.

A Python tutorial for implementing EMUS can be found at `https://gjgilbert.github.io/tutorials/umbrella_sampling/`.

4.2 Applying umbrella sampling to the transit model

We now introduce our full umbrella sampling routine as applied to the transit fitting problem. Properly implemented, our new method produces posterior estimates which are more accurate than estimates obtained using standard direct sampling techniques. The key components of our method are (1) splitting the transit fitting problem into separate windows for grazing vs non-grazing geometries and (2) adopting a unique parameter basis within each window tailored to the specific geometry at hand.

For our umbrella coordinate we adopt the grazing coordinate, $\gamma \equiv (1-b)/r$, which was introduced in §3.1. Defining our windows in terms of γ allows us to easily separate posterior sampling into grazing and non-grazing runs, with the cutoff occurring at $\gamma = 1$. While

developing our method, we first attempted to implement a simple two-umbrella scheme wherein the non-grazing window extended slightly into the grazing regime and, conversely, the grazing window extended slightly into non-grazing regime. However, we found that the sampler still often became stuck at the grazing to non-grazing transition, leading to poorly mixed chains and inaccurate results. We therefore found it necessary to restrict the grazing umbrella to strictly grazing geometries ($\gamma < 1$) and the non-grazing umbrella to strictly non-grazing geometries ($\gamma > 1$). Windows must have at least some overlap with their neighbors, so we introduced a third "transition" umbrella centered on the grazing to non-grazing boundary at $\gamma = 1$ and extending a little way into both the grazing and non-grazing regimes in order to bridge the gap. We find that this simple three-umbrella scheme performed well under a wide range of circumstances.

We define our bias functions over the non-grazing (N), transition (T), and grazing (G) windows as

$$\psi_N \simeq \begin{cases} \gamma - 1 & 1 < \gamma < 2 \\ 1 & \gamma \geq 2 \end{cases} \tag{4.7}$$

$$\psi_T \simeq \begin{cases} \gamma & 0 \leq \gamma < 1 \\ 2 - \gamma & 1 \leq \gamma < 2 \end{cases} \tag{4.8}$$

$$\psi_G \simeq \begin{cases} 1 + \gamma & \gamma \leq 0 \\ 1 - \gamma & 0 < \gamma < 1 \end{cases} \tag{4.9}$$

where the symbol "$\simeq$" denotes an un-normalized distribution. These biases are shown graphically in Figure 4.2. We have opted to use tent biases out of mathematical convenience, but as noted above, the results of umbrella sampling are in general insensitive to any particular choice of bias function. The reader is thus free to chose any other bias function if

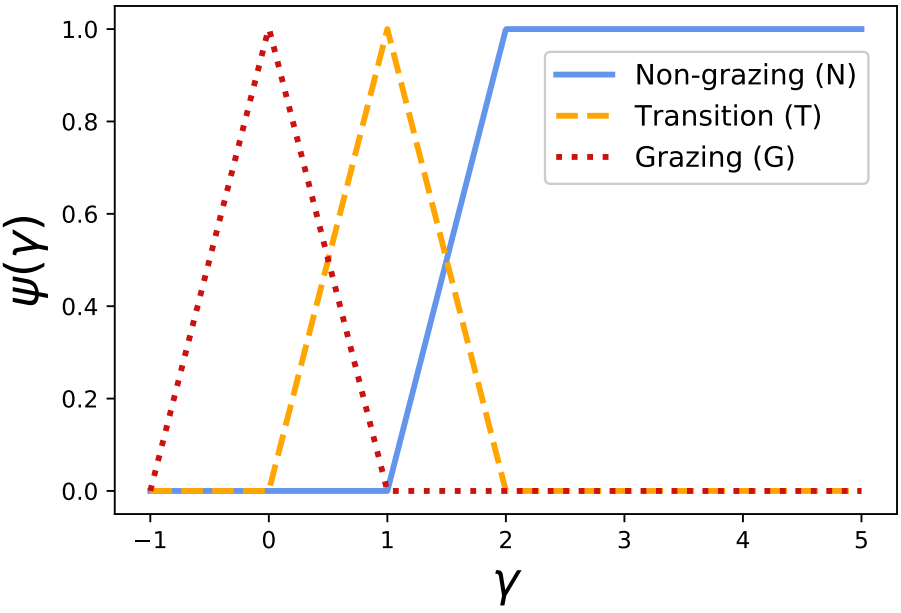

Figure 4.2: Our umbrella bias functions, $\psi(\gamma)$. The solid blue line is the non-grazing umbrella (ψ_N). The dashed orange line is the transition umbrella (ψ_T). The dotted red line is the grazing umbrella (ψ_G). We have opted to use tent biases because these are simple to perform calculations with, but because umbrella sampling is insensitive to the particular choice of bias functions —provided that windows overlap —many other functional forms would perform just as well.

they so desire. However, we do caution that while the shape of the bias within each window is mostly unimportant, altering the window widths (i.e. the range of γ spanned by each ψ) can have a significant effect. Indeed, while developing this method we undertook considerable effort to ensure that windows overlapped enough to facilitate calculation of the window weights without being so wide as to lead to geometric degeneracies. We therefore advise that anyone attempting to apply our method should only adjust the window bounds after careful consideration of the consequences. Unless one has a strongly motivated reason to alter the windows, the safest approach is to stick with the limits presented in Equations 4.7 - 4.9.

Because each window will be sampled independently of the others, we are free to use a different parameter basis within each window. Motivated by the results of §3.2, we adopt the standard $\{\log r, b\}$ basis for the non-grazing umbrella, while for the grazing umbrella we

adopt our new $\{\log\lambda, \gamma\}$ basis. For sampling runs under the transition umbrella, we adopted the hybrid parameter basis $\{\log r, \gamma\}$, which we found worked well over a range of conditions. Each basis set is tailored to the specific geometry within its corresponding window, and thus performs well under its own umbrella. Umbrellas are always defined in terms of γ, but γ is not always a basis parameter. In practice, in order to define our window bounds and bias functions, we must first calculate *gamma* from any two out of three basis parameters $\{r, b, \lambda\}$ following Equation 3.1.

We have so far always used the full first-to-fourth contact transit duration T_{14} as our transit duration because it is defined regardless of transit geometry. However, the center-to-center duration (1.5 to 3.5 contact), T_{c-c} is often better constrained by the data and is therefore often preferred as a basis parameter as long as it is defined, which it will be as long as $\gamma > 0$. Given our window bounds (Equations 4.7-4.9), this means that we are free to use T_{c-c} in place of T_{14} for the N and T umbrellas. Swapping one T for another adds an additional step to the procedure, in that a consistent T must be used to produce the final joint posterior distribution. Fortunately, once samples have been obtained for all parameters, determining T_{14} from T_{c-c} is a matter of straightforward arithmetic, and vice versa. For this study, we will continue to use T_{14} as our basis parameter in order to avoid introducing another complicating factor into an already complicated routine, but keep in mind that when applying umbrella sampling in the future, T_{c-c} should probably be adopted instead of T_{14} whenever possible.

The stage is now set, and at this point one could in principle draw samples from each window and then recombine them into a final joint posterior following the weighting prescription described in §4.1. However, there is one final complication that must be addressed first, namely that we do not know the transit geometry ahead of time and so we cannot be sure whether samples π_T obtained under the T umbrella will overlap with both ψ_G and ψ_N. Recall that umbrella sampling does not merely require that windows overlap, but instead

imposes the more stringent requirement that at least some samples obtained within each window fall into the overlap region with their neighboring windows. This subtle condition demands careful attention, but turns out to be a blessing in disguise.

In order to ensure that our window weights will be properly determined, we first sample from the transition umbrella, producing a (biased) posterior distribution $\pi_T(\gamma)$ with samples restricted by ψ_T to lie between $0 < \gamma < 2$. Because γ tells us how strongly grazing the transit is and because all samples $\pi_T(\gamma)$ must by construction fall near the grazing/non-grazing transition boundary, we can use $\pi_T(\gamma)$ to infer whether the transit geometry is grazing or not. If all samples $\pi_T(\gamma)$ have $\gamma < 1$, we can be confident that the planet is on a grazing trajectory; conversely if all samples $\pi_T(\gamma)$ have $\gamma > 1$, we can be confident that the planet is on a non-grazing trajectory. In the former case (all $\gamma < 1$) we then need only draw samples from the grazing window, whereas in the latter case (all $\gamma > 1$) we need only draw samples from the non-grazing window. In fact, under these circumstances umbrella sampling may no longer be needed, as the N or G windows will by themselves cover the full span of the relevant parameter space. However, at this point samples from the T umbrella have already been obtained, so one may as well proceed with a two-umbrella scheme.

We recommend that all future transit modeling efforts — even those which do not intend use umbrella sampling for their final analysis — first conduct an exploration of the grazing/non-grazing transition boundary, aided by ψ_T to ensure adequate sampling of the region immediately surrounding $\gamma = 1$. Depending on the circumstances, one may wish to set a more or less lenient condition for categorizing a transit as grazing/non-grazing than we have proposed here (i.e. all $\gamma < 1$ vs all $\gamma > 1$), but the core strategy would remain the same. Conclusively ruling in/out grazing geometries will afford us greater confidence in results derived from transiting modeling, and if widely adopted we anticipate our "check the transition region first" approach will reveal previously unnoticed inaccuracies or systematic offsets in transiting exoplanet catalogs.

Samples may be drawn using any suitable sampling method, and provided that all posterior chains are well mixed and pass the necessary convergence checks, the choice of sampler will be inconsequential to the final results, save perhaps a difference in computational efficiency. Once we have drawn samples from all three windows (or perhaps only two, if $\pi_T(\gamma)$ rules out one geometry or another), calculation of the window weights, z_i, is a straightforward application of the EMUS algorithm presented in §4.1; once z_i have been calculated, we can then immediately estimate the posterior distributions and summary statistics.

CHAPTER 5

COMPARISON OF RESULTS FROM STANDARD SAMPLING TECHNIQUES TO UMBRELLA SAMPLING

We now test our proposed method by simulating transit lightcurve photometry for several prototypical star-planet configurations and then comparing posterior inferences obtained via umbrella sampling to inferences obtained using a standard direct sampling approach. Throughout these tests, we follow the same data simulation procedure and Hamiltonian Monte Carlo sampling routine described in §2.3, modified to incorporate a moderately informative prior on eccentricity. Rather than incorporating e and w as free parameters in our model, we instead inferred these quantities using the photoeccentric effect [Ford et al., 2008, Dawson and Johnson, 2012], thus necessitating an priors on both e and $\rho_\star$. For e we assumed a Raleigh distribution with scale parameter $\sigma_e = 0.21$, corresponding to the single-planet value found by Mills et al. [2019]; for $\rho_\star$, we assumed a 10% Gaussian measurement uncertainty. In practice, placing priors on e and ρ serves to place indirect priors on T and b. We will address the role of eccentricity priors and describe the effects of several alternative prior distributions in greater detail is §5.1.4 below.

We perform three tests of our method, each focused on a different star-planet architecture. In the first test (the "J" models; see Table 2.1), we place a warm Jupiter in orbit around a Sun-like star at various impact parameters in order to simulate grazing, near-grazing, and non-grazing trajectories. In the second test (models "SE" & "MN") we place a super-Earth and mini-Neptune on 21 day orbits around a star typical of the Kepler field, with inclinations scaled to produce comparable transit depths. In the third (model "MHZ"), we place a rocky planet in the habitable zone of an M dwarf. Simulated photometry is shown in Figures 2.5, 5.5, and 5.8, and ground truth parameter values for each simulated lightcurve are collected in Table 2.1. As before, the important point throughout is that we have endeavored to simulate unremarkable transits, which we then model using techniques which are intended to be as

uncontroversial as possible.

5.1 A giant planet orbiting a solar twin

For our first test, we placed a warm Jupiter ($r = 0.1$) on a circular orbit around a Sun-like star at three different impact parameters in order to create a grazing ($b = 1.00$, model J-100), near-grazing ($b = 0.85$, model J-85), and non-grazing ($b = 0.22$, model J-22) trajectory. The transit duration for all three cases was set to $T = 3.0$ hrs and then the orbital period was calculated in order to preserve $e = 0$, resulting in orbital periods of 44.9, 13.0, and 3.6 days, respectively. In order to produce a comparable signal-to-noise, the simulated Gaussian noise for the grazing transit (J-100) was reduced by a factor of two relative to the non-grazing and near-grazing transits. The simulated photometry for all three configurations is shown in Figure 2.5.

5.1.1 Simulation J-85: a near-grazing transit

We being by placing our warm Jupiter on a $P = 13$ day orbit around its host star with $b = 0.85$, thereby producing a transit chord that is non-grazing yet close enough to the stellar edge that limb darkening becomes significant. In this near-grazing regime, the transit morphology begins to shift from U-shaped to V-shaped, so we expect that some fraction of the posterior distribution will be consistent with both a grazing and non-grazing trajectory. Examination of γ samples obtained under the transition umbrella, ψ_T, confirm that this is indeed the case (Figure 5.1), validating our assertion that umbrella sampling is warranted.

Both direct sampling and umbrella sampling produce comparable distributions for T and broadly similar estimates of r and b (Figure 5.2). However, direct sampling does not fully explore the high-b, high-r tail of the distribution. By-eye the differences appear slight, but the consequences of these skewed distributions become apparent when one calculates the marginalized 1σ uncertainties for r and b. From direct sampling, we estimate $r = 0.098^{+0.042}_{-0.013}$,

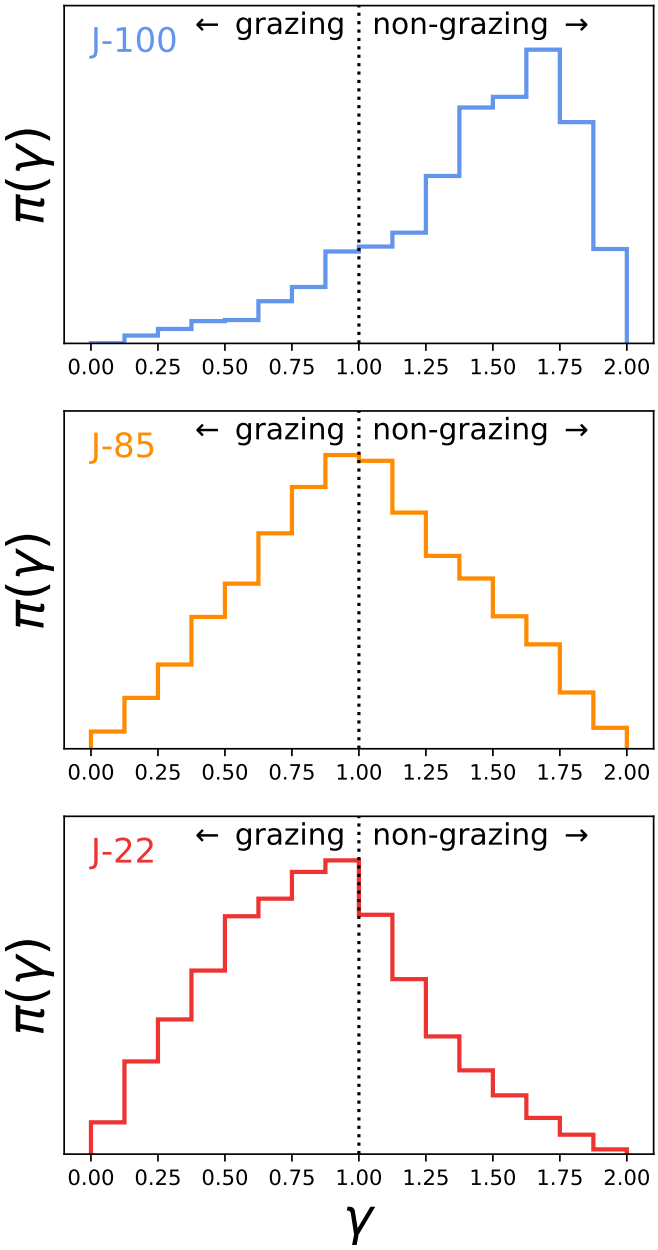

Figure 5.1: Distribution of the grazing coordinate, γ, for posterior MCMC samples obtained under the transition umbrella, ψ_T, for three simulated transits of a warm Jupiter orbiting a Sun-like star at various impact parameters. Simulated parameter values are collected in Table 2.1 and corresponding simulated lightcurves are shown in Figure 2.5. Unsurprisingly, the fraction of posterior samples consistent with a non-grazing geometry is highest for the simulated non-grazing transit (top), and vice-versa for a grazing geometry (bottom). The near-grazing transit (middle) reflects an intermediate state. In all three cases, at least some fraction of the posteriors are consistent with both a grazing and a non-grazing trajectory, indicating the transit geometry is ambiguous and the application of umbrella sampling is warranted.

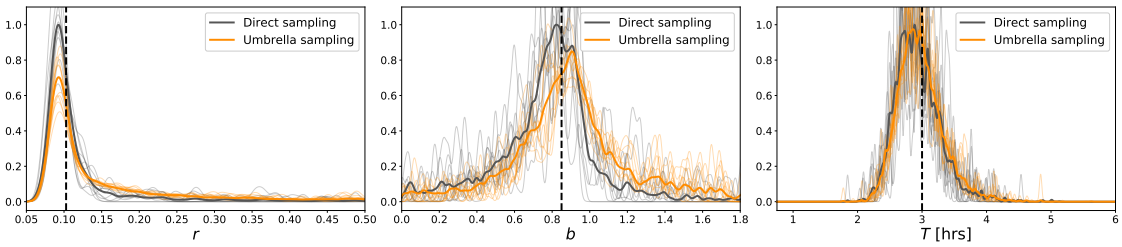

Figure 5.2: Posterior distributions of r, b, and T for a simulated near-grazing transit of a Jupiter-size planet on a 13 day orbit around a Sun-like star (simulation J-85). See Table 2.1 for simulated model parameters and Figure 2.5 for the simulated photometry. Each thin line represents a 2000 sample chain from a single independent Monte Carlo run, while the thick lines give the combined results of 20 such runs. Vertical dashed lines represent ground-truth parameter values. Both methods produce posterior distributions consistent with the true value, but only umbrella sampling is able to fully explore the high-b, high-r tail of the distribution.

whereas from umbrella sampling, we estimate $r = 0.108^{+0.187}_{-0.021}$ (based on the 16$^{\text{th}}$, 50$^{\text{th}}$, and 84$^{\text{th}}$ percentiles). Although one might naively prefer the narrower posterior obtained via direct sampling, this result is inaccurate, and the misleadingly tight constraint on r is predicated on the false assumption that the high-b high-r tail has been rule out, when in fact it has simply not been explored. Umbrella sampling, on the other hand, ensures that the difficult to explore regions of the posterior have indeed been adequately sampled.

5.1.2 Simulation J-22: a non-grazing transit

We next modify our simulated transit by changing the impact parameter to $b = 0.22$ in order to place the planet on a non-grazing trajectory. In order to keep the transit duration consistent at $T = 3$ hrs, we shifted the orbital period to $P = 3.6$ days. In this case, the results of the two methods are entirely consistent with one another (Figure 5.3), as expected for a planet with negligible posterior mass consistent with a grazing geometry. Because there is a small but non-zero fraction fraction of samples with $b > 1 - r$ (Figure 5.1), trusting the results from direct sampling hinges on the implicit assumption that the sampler did not explore the grazing regime because the model and data are poorly matched there, rather

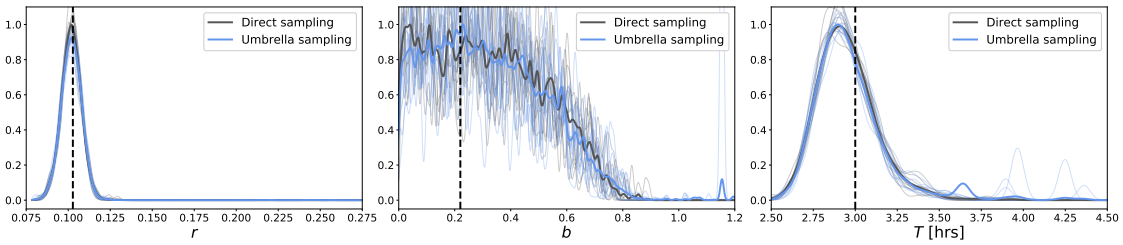

Figure 5.3: Posterior distributions of r, b, and T for a simulated non-grazing transit of a Jupiter-size planet on a 3.6 day orbit around a Sun-like star. See Table 2.1 for simulated model parameters and Figure 2.5 for the simulated photometry. Each thin line represents a 2000 sample chain from a single independent Monte Carlo run, while the thick lines give the combined results of 100 such runs. Vertical dashed lines represent ground-truth parameter values. Both methods produce comparable results, however only umbrella sampling is able to smoothly explore deep into the grazing regime. This augmented exploration allows us to confidently rule out a grazing geometry by placing reliable upper limits on r and b.

than because the sampler encountered a bottleneck at the grazing/non-grazing boundary. The advantage of using umbrella sampling is that we can be more confident is our inferences because the sampler explores smoothly deep into the grazing regime, allowing us to be sure that the posterior likelihood there is indeed small.

5.1.3 Simulation J-100: A grazing transit

For our last test we shift the transit to $b = 1.0$ in order to create a grazing trajectory (model J-100). Once again, we preserve the transit duration at $T = 3$ hrs by adjusting the orbital period, in this case to to $P = 45$ days. In order to compensate for the reduced transit depth of the grazing geometry, we reduce the photometric noise level by a factor of two, which gives this simulated transit (J-100) a similar signal-to-noise ratio compared to the first two simulations (J-85 & J-22).

The performance of the two methods for fitting a grazing transit is quite similar (Figure 5.4). From direct sampling, we estimate $r = 0.148^{+0.262}_{-0.077}$, $b = 1.05^{+0.28}_{-0.11}$; whereas from umbrella sampling, we estimate $r = 0.130^{+0.265}_{-0.070}$, $b = 1.03^{+0.29}_{-0.13}$. As with the non-grazing case (simulation J-22), the main advantage of umbrella sampling is that we can be sure we

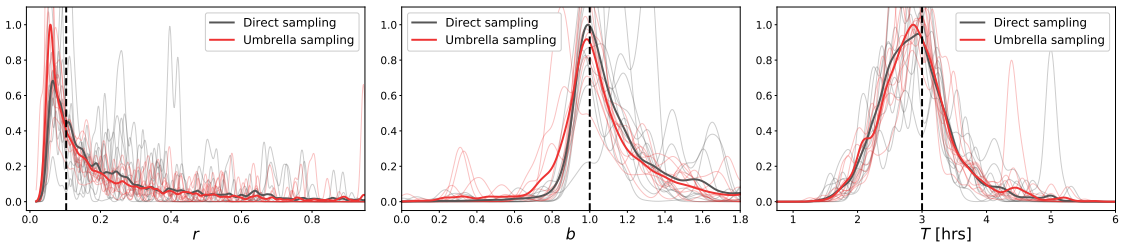

Figure 5.4: Posterior distributions of r, b, and T for a simulated grazing transit of a Jupiter-size planet on a 45 day orbit around a Sun-like star (simulation J-100). See Table 2.1 for simulated model parameters and Figure 2.5 for the simulated photometry. Each thin line represents a 2000 sample chain from a single independent Monte Carlo run, while the thick lines give the combined results of 100 such runs. Vertical dashed lines represent ground-truth parameter values. In this case, both mehods produce comparable results.

have explored the full posterior geometry, lending us greater confidence in our results.

5.1.4 The role of eccentricity priors

In order to investigate the effect of eccentricity priors, we repeated the experiment for the near-grazing transit (simulation J-85) using three additional eccentricity priors distributions. For the first two, we again used a Rayleigh prior, but now with scale parameter $\sigma_e = 0.0355$ or $\sigma_e = 0.008$. The former corresponds to the value found by Mills et al. [2019] for multiplanet systems, while the latter corresponds to the value by Lithwick et al. [2012] for systems exhibiting large-amplitude transit timing variations. Recall that our original test used $\sigma_e = 0.21$, the Mills et al. [2019] single planet value. Our fourth and final test placed uniform (i.e. uninformative) priors on e. In all cases, we assumed a 10% Gaussian measurement uncertainty on $\rho_\star$.

Both methods show a similar sensitivity to choice of eccentricity prior (see Table 5.1).

5.2 A pair of planets straddling the radius valley

A primary motivation for developing our umbrella sampling method is to accurately determine the radii of exoplanets in or near the radius valley [Fulton et al., 2017]. More specifically,

Prior distribution	Direct Sampling	Umbrella Sampling
Rayleigh, $\sigma_e = 0.008$	$0.108^{+0.037}_{-0.014}$	$0.118^{+0.198}_{-0.021}$
Rayleigh, $\sigma_e = 0.0355$	$0.108^{+0.038}_{-0.014}$	$0.120^{+0.208}_{-0.022}$
Rayleigh, $\sigma_e = 0.21$	$0.098^{+0.042}_{-0.013}$	$0.108^{+0.187}_{-0.021}$
Uniform, $e \sim (0, 1)$	$0.091^{+0.033}_{-0.010}$	$0.097^{+0.136}_{-0.015}$

Table 5.1: Marginalized MCMC posterior values for the planet-to-star radius ratio, r, of a simulated transit (simulation J-85), assuming four different eccentricity prior distributions. The true value is $r = 0.103$. Posterior values quoted in this table correspond to the retrieved 16th, 50th, and 84th percentiles of r, with results arranged from most informative prior (top) to least informative (bottom). See text of §5.1.4 for discussion.

we would like to be able to measure the size of planets with $r_p \approx 1.6R_\oplus$ and periods $P \lesssim 100$ days orbiting FGK stars, i.e. planets typical of the Kepler and *K2* samples. For this case study, we simulate the transits of a pair of planets, each on a circular 21 day orbit around a K dwarf ($R_\star = 0.92R_\odot$). The first planet (simulation SE; a super-Earth) has $r_p = 1.3R_\oplus$ and a non-grazing trajectory ($b = 0.70$). The second planet (simulation MN; a mini-Neptune) has $r_p = 2.2R_\oplus$ and a barely grazing trajectory ($b = 0.98$). These setups produce a pair of transits with comparable transit depths, albeit distinct transit durations (Figure 5.5). For these test cases, an accurate estimate of r thus hinges on accurate estimates of both T and b. Our goal then is to investigate whether our competing sampling methods can constrain these three parameters with sufficient reliability to determine whether each planet exists on the rocky or gaseous edge of the radius valley.

5.2.1 Simulation SE: A non-grazing super-Earth

Posterior distributions for the super-Earth simulation produce consistent results regardless of which method is used (Figure 5.6). Because the transit trajectory is far from grazing ($b = 0.7, r = 0.012$) this agreement is to be expected. The marginalized constraints for this

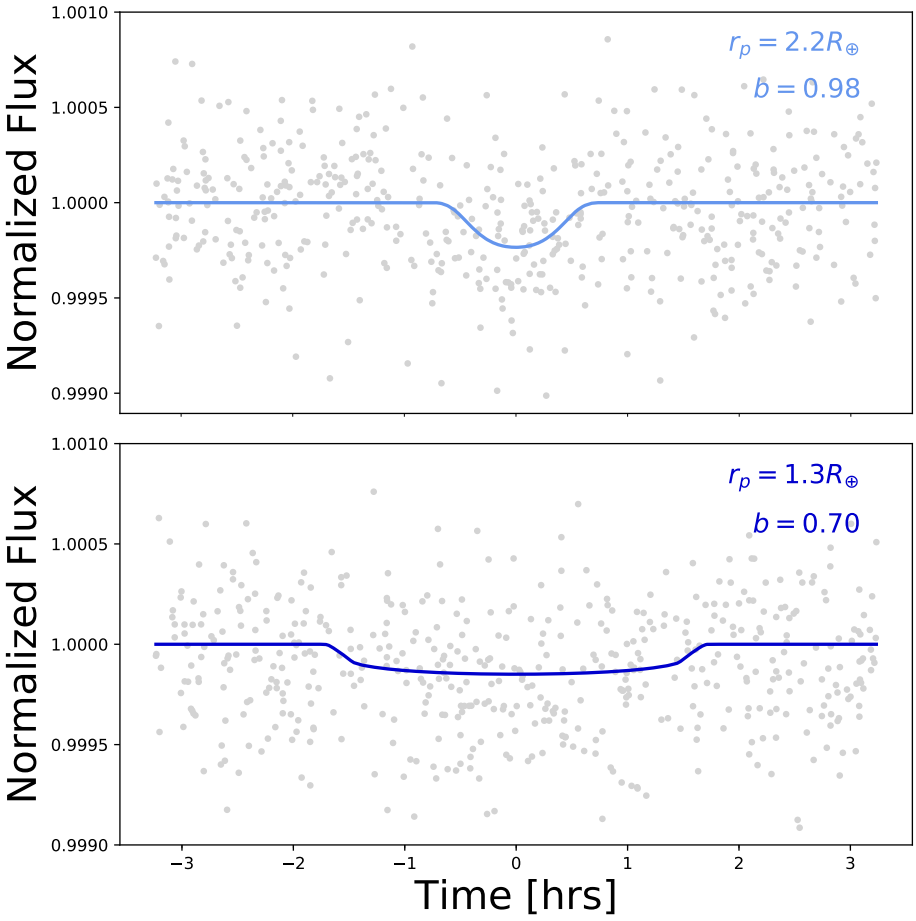

Figure 5.5: *Top panel*: Simulated photometry for a mini-Neptune (model MN) orbiting a K-dwarf star on a barely grazing orbit. *Bottom panel*: Simulated photometry for a super-Earth (model SE) orbiting the same star on a non-grazing orbit. Ground truth simulation parameters are collected in Table 2.1. See §5.2 for discussion.

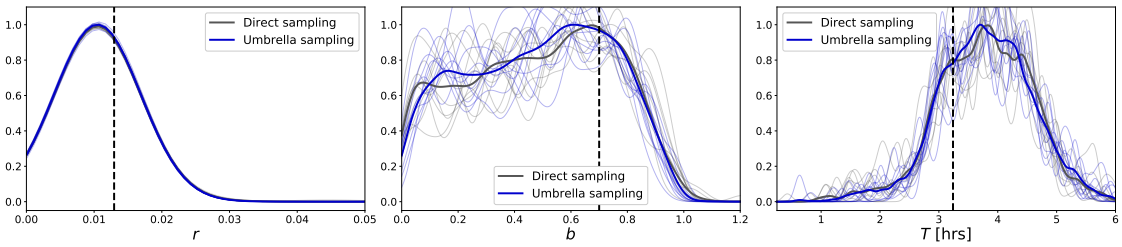

Figure 5.6: Posterior distributions of r, b, and T for a simulated non-grazing transit of a super-Earth orbiting a K-dwarf star (simulation SE). See Table 2.1 for simulated model parameters and Figure 5.5 for the simulated photometry. Each thin line represents a 2000 sample chain from a single independent Monte Carlo run, while the thick lines give the combined results of 20 such runs. Vertical dashed lines represent ground-truth parameter values. The transit trajectory is far from grazing ($b = 0.7, r = 0.012$), and the two methods produce comparable results ,as expected.

case are $r_p = 1.06 \pm 0.17 R_\oplus$, $b = 0.51 \pm 0.30$, correctly identifying the planet as a rocky object with a non-zero, non-grazing impact parameter.

Once again, the primary advantage of umbrella sampling is that it affords us confidence in our results. A small fraction of the posterior samples are consistent with $b > 1$, and by employing umbrella sampling we can be sure that we have correctly weighted the high-b tail of the distribution, whereas with direct sampling alone there would be ambiguity as to whether the tail has been properly explored. In this case, direct sampling does manage to produce the correct result, but we only know this because we have also fit the transit using umbrella sampling. In this specific case, a larger fraction of samples consistent with a grazing trajectory would have made the radius uncertainty larger, which in turn would make the composition of the planet ambiguous, a major detriment for studies of planets near the radius valley.

5.2.2 Simulation MN: A barely grazing mini-Neptune

For our mini-Neptune simulation, posterior inferences made via umbrella sampling are significantly better than those made via direct sampling (Figure 5.7). Whereas umbrella sampling returns $r_p = 2.17^{+6.16}_{-0.55}$, $b = 0.96^{+0.10}_{-0.02}$, direct sampling returns $r_p = 5.3^{+19.4}_{-2.95}$, $b = 1.03^{+0.19}_{-0.04}$.

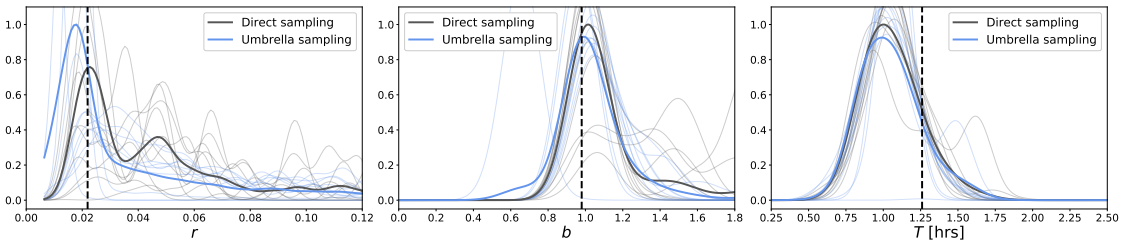

Figure 5.7: Posterior distributions of r, b, and T for a simulated barely grazing transit of a mini-Neptune orbiting a K-dwarf star (simulation MN). See Table 2.1 for simulated model parameters and Figure 5.5 for the simulated photometry. Each thin line represents a 2000 sample chain from a single independent Monte Carlo run, while the thick lines give the combined results of 20 such runs. Vertical dashed lines represent ground-truth parameter values. In this case, umbrella sampling produces obviously improved results, as direct sampling struggles to smoothly explore the grazing regime.

The reduced precision in r_p from direct sampling will have dramatic consequences for understanding the composition of the individual planet. Even though there is indeed a fairly large uncertainty on the planet radius no matter what method is used - which is to be expected for grazing transits - the implied planet composition is far more ambiguous using direct sampling.

5.3 A rocky planet in the M-dwarf habitable zone

For our final test (simulation MHZ), we place a Mercury-sized planet ($r_p = 0.38 R_\oplus$) on a $P = 37$ day orbit around a $R_\star = 0.38 R_\odot$ M dwarf, which puts the planet squarely in that star's habitable zone. See Figure 5.8 for the simulated photometry and Table 2.1 for the ground truth simulation parameters.

We find that direct sampling and umbrella sampling perform equally well for this test case, with both methods recovering the true values for r, b, and T with nearly identical accuracy (Figure 5.9). Specifically, both methods find $r_p = 0.32 \pm 0.20 R_\oplus$ and a broad, predominantly non-grazing distribution for b. Yet even in this case where marginalized statistics are nearly identical, umbrella sampling still confers an advantage over direct sam-

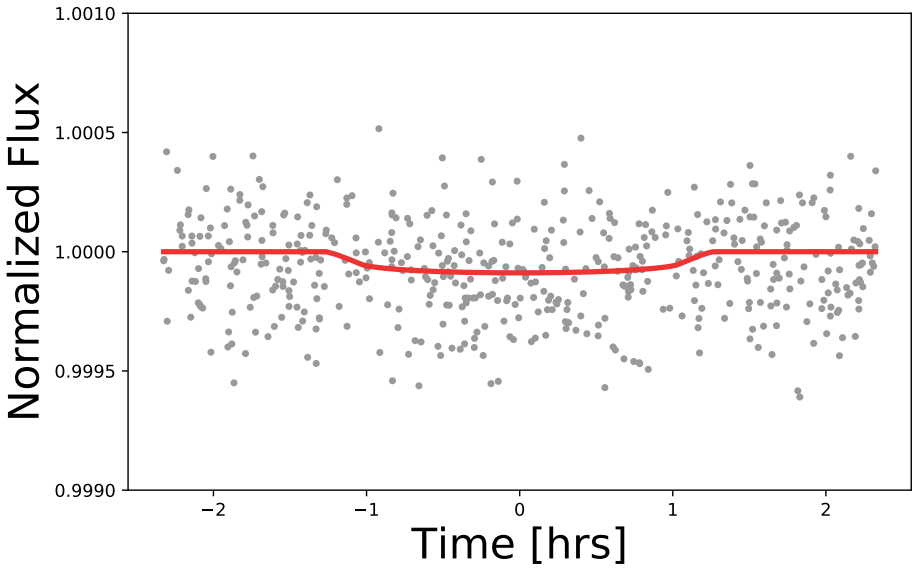

Figure 5.8: Simulated photometry for a Mercury radius planet orbiting in the habitable zone of an M-dwarf host star (model MHZ). Ground truth simulation parameters are collected in Table 2.1. See §5.3 for discussion.

pling. Because the posterior distribution for impact parameter extends above $b = 1$ for both methods, with direct sampling we cannot be certain that the entire full posterior space has been adequately explored. Rather, it is possible we encountered the usual bottleneck at the grazing/non-grazing boundary, leaving the grazing regime undersampled. With umbrella sampling, however, we can be confident —without the need for follow-up observations —that the posterior geometry has been fully explored, meaning that the planet is indeed on a non-grazing orbit and therefore has an accurately measured radius.

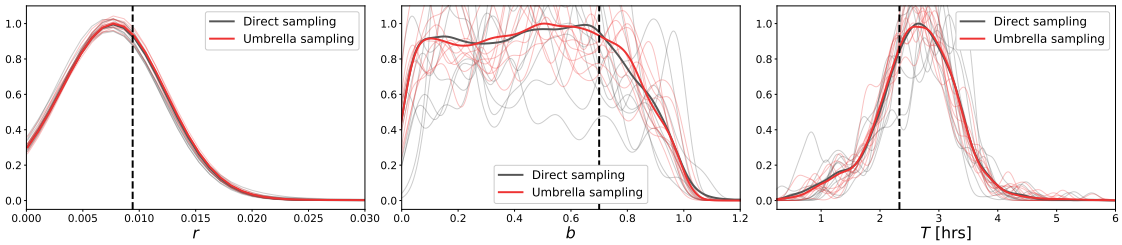

Figure 5.9: Posterior distributions of r, b, and T for a simulated barely grazing transit of a Mercury-sized planet orbiting in the habitable zone of an M-dwarf star (simulation MHZ). See Table 2.1 for simulated model parameters and Figure 5.8 for the simulated photometry. Each thin line represents a 2000 sample chain from a single independent Monte Carlo run, while the thick lines give the combined results of 100 such runs. Vertical dashed lines represent ground-truth parameter values. For this test case, both direct sampling and umbrella sampling perform equally well.

CHAPTER 6

ANALYSIS OF REAL TARGETS

We will now use umbrella sampling to estimate the impact parameters and radii of several KOI planet candidates with $b > 1$ reported by the NASA Exoplanet Archive cumulative KOI table. For the KOIs in multiplanet systems, we also fit the sibling planets (which are not on grazing trajectories) in order to verify that our results are consistent with previous measurements. For the KOIs in single planet systems, we select an unrelated planet with with a non-grazing b but otherwise similar properties to serve as a basis of comparison.

Our data reduction and transit fitting pipeline is described below. Similarly to our experiments with simulated data, we have endeavored to use standard techniques wherever possible, except of course for the steps of the procedure which directly implement umbrella sampling.

We begin by downloading the Pre-search Data Conditioning Simple Aperature Photometry (PDCSAP) flux from the Mikulski Archive for Space Telescopes (MAST). We then flag bad cadences and remove any large outliers with iterative sigma clipping at the 5σ level. We next remove long-term trends using a Gaussian Process (GP) implemented by `celerite` [Foreman-Mackey et al., 2017]. For the GP kernel, we adopted a stochastically driven simple harmonic oscillation SHOTerm[1], which has been shown to produce good results for astronomical time series [Foreman-Mackey et al., 2017]. In order to protect the transit shape during detrending, we mask all cadences within 1.5 transit durations of each expected mid-transit time and project our GP trend across the masked transit region.

To account for possible transit timing variations (TTVs), we read in the transit time measurements of Holczer et al. [2016] and fit a smooth model to these using a GP regression and a Matern-3/2 kernel. We obtain a self-consistent starting estimate for the transit shape and transit times by first fitting $\{P, t_0, \log r, b, T\}$ and holding transit times fixed, then

1. `https://celerite.readthedocs.io/`

reversing the procedure to hold transit shape parameters fixed and fitting independent transit times. Finally, we model the independent transit times using a 1st-3rd order polynomial and either zero or one single frequency sinusoids. The complexity of the TTV model was selected based on the Akaike Information Criterion [AIC; Akaike, 1974]. For all steps of this TTVs initialization procedure we hold limb darkening coefficients fixed to the theoretical values obtained from the NASA Exoplanet Archive.

While sampling from the posterior, we hold transit times fixed at our low order polynomial + sinusoid model and sample each of the umbrellas independently following the prescription in §4 and §5. This means that the free parameters in the model are $\{\log T, q_1, q_2, F, \log \sigma_F\}$, plus either $\{r, b\}$ for the N umbrella, $\{\log r, \gamma\}$ for the T umbrella, or $\{\log \lambda, \gamma\}$ for the G umbrella. For simplicity, we only consider non-overlapping transits and fit planets one at a time for multiplanet systems. Overlapping transits were defined as any transit pair for which $|t_{0,b} - t_{0,c}| < (T_b + T_c)$ for any two planets b and c. Each HMC run consisted of two independent chains, with each chain run for a default length of 10,000 tuning steps and 5,000 draws, generating 10,000 samples total per run. In a few cases, the chains did not converge on our first attempt to fit the data, in which case extending the length of the tuning phase remedied the issue.

After drawing samples from all three windows - N, T, and G - we check that the posterior samples of r, b, and T are consistent between the sub-distributions π_N, π_T, and π_G for each planet. This does not mean that the distributions must overlap completely (indeed, they are expected not to), but rather that they have at least some overlap, with perhaps some modest tension between umbrellas. In practice, we found that posterior sub-distribution were nearly always either obviously consistent or obviously inconsistent, with the later case indicating that the algorithm had not been properly tuned prior to sampling. In some cases, even though sub-distributions were clearly inconsistent when considered simultaneously, results initially appeared reasonable when each umbrella was considered in isolation. Thus, our

method provides a new avenue for verifying that the results of a transit fit are trustworthy: if Markov chains do not properly behave within all three windows and produce self-consistent results, we know to investigate further. Thus, the sub-distribution comparison step of our algorithm builds in an extra redundancy for checking convergence.

6.1 KOI-2068

KOI-2068 is a $0.91R_{\odot}$ star hosting a single planet candidate at $P = 42$ days with 1σ upper limits $b \leq 58$ and $r_p \leq 42R_{\oplus}$. With signal-to-noise $S/N = 21$ and a disposition score of 0.89, the object is unlikely to be a false positive. This combination of degenerate, poorly constrained r and b values plus a low false positive probability makes this object an ideal test case for our umbrella sampling scheme. As a comparison target, we select KOI-2285, a $0.87R_{\odot}$ star hosting a single confirmed planet at $P = 38$ days, with $b = 0.26 \pm 0.23$, $r_p = 2.79 \pm 0.30$, and $S/N = 24$.

After sampling, for KOI-2068.01 we recover $r_p = 5.81 \pm 2.38R_{\oplus}$, $b = 0.99 \pm 0.03$, $T = 1.39 \pm 0.08$ hrs, a marked improvement in precision over the literature values. While the radius constraint remains somewhat broad, we are nonetheless now able to confidently say that the object is indeed a planet-sized object on a grazing orbit. For the comparison target, we recover $r_p = 2.88 \pm 0.15R_{\oplus}$, $b = 0.55 \pm 0.27$, $T = 4.17 \pm 0.16$ hrs, consistent with the literature values. We will not comment on the relative precision or accuracy of our results vs reliable literature results, as any differences in measured values are more likely to be driven by differences in data reduction techniques than by which sampling method was used. The important point is that our analysis was able to reproduce known reliable results, validating our pipeline and affording us confidence in any new measurements which improve upon the state of the art.

6.2 KOI-2150

KOI-2150 is a $0.94R_\odot$ star hosting two planet candidates, both with impact parameters greater than unity reported on the NASA Exoplanet Archive. The inner planet ($P = 19$ days) has 1σ upper limits $b \leq 73$ and $r_p \leq 38R_\oplus$, while the outer planet ($P = 45$ days) has $b \leq 72$ and $r_p \leq 94R_\oplus$. Both candidates have a disposition score > 0.99, indicating that neither is likely to be a false positive.

After umbrella sampling, for the inner planet we recover $b = 0.37 \pm 0.29$, $r_p = 2.41 \pm 0.19R_\oplus$ (8% radius uncertainty), and for the outer planet we recover $b = 0.47 \pm 0.39$, $r_p = 1.86 \pm 0.39R_\oplus$ (21% radius uncertainty). Thus, umbrella sampling places both objects on non-grazing trajectories - albeit with poorly constrained impact parameters - and finds a plausible radius for each. These candidates are probably mini-Neptunes, both possessing individual properties consistent with a depleted radius valley [Fulton and Petigura, 2018, Van Eylen et al., 2018] and relative sizes consistent with the "peas in a pod" hypothesis [Weiss et al., 2018], adding further credulity to our results.

6.3 KOI-1426

KOI-1426 is a $0.90R_\odot$ star hosting two confirmed planets and one planet candidate. The two confirmed planets have well-constrained properties reported on the NASA Exopplanet Archive (KOI-1426.01: $P = 39$ days, $r_p = 2.81 \pm 0.04\ R_\oplus$, $b = 0.03^{+0.33}_{-0.03}$, KOI-1426.02: $P = 75$ days, $r_p = 6.39 \pm 0.10R_\oplus$, $b = 0.80^{+0.01}_{-0.06}$), but the candidate planet (KOI-1426.03: $P = 150$ days, $r_p \leq 36R_\oplus$, $b \leq 68$) exhibits the $r - b$ degeneracy. Unlike either candidate in the KOI-2150 system, KOI-1426.03 possesses an impact parameter constraint $b = 1.25^{+67}_{-0.17}$ that marks its orbit (if real) as unambiguously grazing.

Our umbrella sampling analysis confirms the grazing transit hypothesis for KOI-1426.03, finding $b = 1.03 \pm 0.08, r = 16.6 \pm 6.9R_\oplus$ and $< 4\%$ of posterior samples drawn under

the transition umbrella consistent with a non-grazing geometry. In contrast, none of the samples drawn from the transition window for either of the two confirmed planets were consistent with a grazing geometry (Figure 6.1), highlighting the utility of our approach for distinguishing grazing from non-grazing transits. Although the uncertainty on the radius of the grazing candidate planet is somewhat high at 41%, the more physically sensible value lends us greater confidence that the planet is real. As expected, our posterior results for the two confirmed planets (KOI-1426.01: $r_p = 2.72 \pm 0.07\ R_\oplus$, $b = 0.26 \pm 0.15$, KOI-1426.02: $r_p = 6.52 \pm 0.16 R_\oplus$, $b = 0.84 \pm 0.02$) are consistent with the literature values. Our constraints are somewhat less precise than those reported on the archive, but recall that experiments with synthetic data (§5) demonstrated that standard sampling methods are prone to producing misleadingly narrow posterior distributions. This situation is not necessarily the case for KOI-1426, and a deeper investigation (which is beyond the scope of this work) will be needed to resolve the issue.

Revealing the true properties of the KOI-1426 system will likely require a full photodynamical analysis. In retrospect, this is unsurprising for two reasons. First, all three planets exhibit large transit timing variations [Diamond-Lowe et al., 2015, Holczer et al., 2016] which may be insufficiently characterized by our parametric model. Second, grazing transits often have a time-dependent transit shape [Hamann et al., 2019, Dawson, 2020], and thus our approximation of an invariant transit shape may yield biased inferences. These complications do not mean that our present efforts to model the system were a waste of time. On the contrary, the results obtained with umbrella sampling will serve as useful priors for setting up the computationally expensive photodynamical model. Informed priors (such as the fact that KOI-1426.02 is both real and on a grazing trajectory) can place meaningful limits on the system architecture and thereby greatly improve both the accuracy and efficiency of the full photodynamical treatment. Furthermore, the techniques of photodynamics and umbrella sampling are not mutually exclusive, and it may ultimately prove necessary to combine

the two methods in order to achieve a definitive result for this or other dynamically active systems.

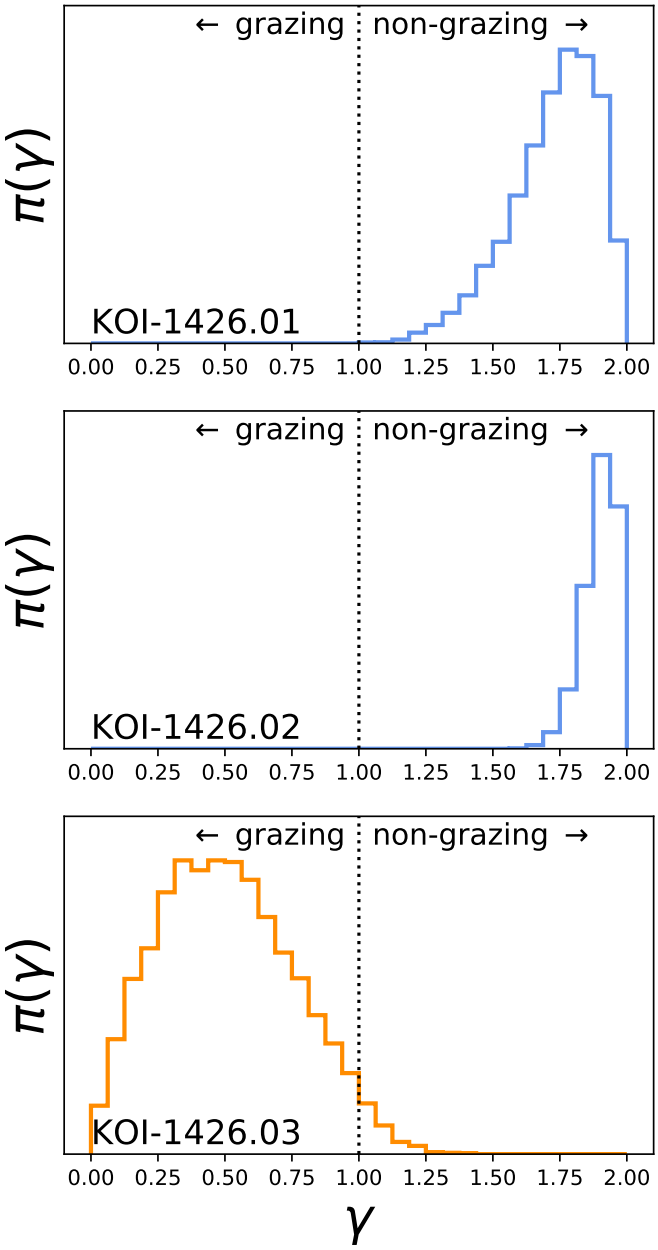

Figure 6.1: Samples of γ obtained under the transition umbrella, ψ_T, for the three planets in the KOI-1426 system. Neither of the two confirmed objects (top two panels, blue) have any posterior samples $\gamma < 1$, indicating that both planets are almost certainly on non-grazing trajectories. For these two objects, we can comfortably ignore the grazing umbrella, ψ_G, and perform a two-window analysis. The candidate object (bottom panel, orange) has 97% of samples with $\gamma < 1$, indicating that this planet is probably on a grazing trajectory. For this object, we might choose to ignore the non-grazing umbrella, ψ_N, but the more conservative approach would be to perform a three-window analysis as usual. The distributions seen here for KOI-1426 are typical of Kepler targets, in that samples of γ from the transition umbrella can often be used to rule out/in certain transit geometries.

CHAPTER 7

SUMMARY AND RECOMMENDATIONS

We have introduced a new method for modeling exoplanet transit lightcurves which explicitly accounts for the differences in transit geometry between grazing and non-grazing trajectories. Our technique employs the well-established framework of umbrella sampling [Torrie and Valleau, 1977] by splitting the transit fitting problem into three sub-problems, each restricted to either the grazing, non-grazing, or transition regions of the parameter space. We draw samples independently from each window using an MCMC sampler to produce three posterior sub-distributions which we then recombine into a single joint posterior distribution using the Eigenvecor Method of Umbrella Sampling [Thiede et al., 2016, Dinner et al., 2017].

Although umbrella sampling is widely used by molecular dynamicists and biochemists, it has only recently begun to gain the attention of astronomers [Matthews et al., 2018]. Yet umbrella sampling is itself a general statistical tool not tied to any particular content domain. At heart, umbrella sampling is designed to estimate complicated posterior geometries (e.g. isolated modes or degeneracy ridges) - geometries of the sort that arise frequently in astrophysical studies. By applying umbrella sampling to a familiar astronomical problem and illustrating its efficacy, we hope to raise awareness of this powerful statistical technique which is well suited to astronomical data analysis. To aide astronomers first learning to use umbrella sampling, we have provided an introductory Python tutorial at `https://gjgilbert.github.io/tutorials/umbrella_sampling/`.

Our umbrella sampling routine reliably produces posterior estimates of planetary radii and impact parameters which are both more accurate than estimates obtained using a standard approach. We tested our method under a wide range of conditions using both real and synthetic data, finding that umbrella sampling performed at least as well as —and usually better than —the standard direct sampling approach for every star-planet configuration we considered. Moreover, even in cases where umbrella sampling did not provide higher preci-

sion estimates than direct sampling, we were able to have greater confidence in the results of umbrella sampling because only this method is able to efficiently explore deep in the the grazing regime.

Throughout this paper, we have offered numerous suggestions for how to modify existing transit modeling procedures in order to produce more robust posterior estimates. We now summarize these recommendation here.

1. Before fitting any transit model, perform an exploratory analysis restricted to the region of parameter space immediately surrounding the grazing/non-grazing transition at $r = 1 - b$. This exploration can be efficiently executed using our $\{r, \gamma\}$ basis and and the transition umbrella, ψ_T.

2. If all samples $\pi_T(\gamma)$ are consistent with a non-grazing geometry (i.e. all $\gamma > 1$), one may proceed with a standard analysis, restricting the model to non-grazing geometries. Conversely, if all samples $\pi_T(\gamma)$ are consistent solely with a grazing geometry (all $\gamma < 1$), one may instead restrict the model to grazing geometries and sample using our new $\{\lambda, \gamma\}$ basis.

3. If, however, samples $\pi_T(\gamma)$ are mixed between grazing and non-grazing geometries, the transit should be modeled using the scheme we have described in detail in §3.1 and §4.2.

4. After sampling from under the various umbrellas ψ_i, compare posterior sub-distributions π_i for each transit parameter to ensure that inferences are consistent between samples drawn from different windows. If samples are in disagreement, closer investigation is needed. This comparison step provides an additional convergence check for the user.

5. For planets inferred to orbit on a grazing trajectory, consider whether a fully photodynamical analysis is needed. If so, the results obtained via umbrella sampling will serve

as useful priors for initializing the more computationally expensive photodynamical model, thereby improving efficiency.

For most of the history of exoplanet science, uncertainties on planetary radii and orbital parameters have been dominated by uncertainties on stellar parameters. Now, however, with improved stellar radius estimates from Gaia [Gaia Collaboration et al., 2018], and high resolution spectroscopy [Johnson et al., 2017, Petigura et al., 2017], the details of the transit fitting problem have once again become relevant for obtaining state-of-the-art estimates of planet properties. Statistical studies of exoplanets will remain dominated by the population of transiting planets for at least the next decade, and so transit modeling will remain at the foundation of many astrophysical analyses. By adopting umbrella sampling as a new tool, we will ensure that our understanding of exoplanet demographics, architectures, and formation histories will reach as far as the data allow.

REFERENCES

H. Akaike. A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, 19:716–723, January 1974.

Astropy Collaboration, T. P. Robitaille, E. J. Tollerud, P. Greenfield, M. Droettboom, E. Bray, T. Aldcroft, M. Davis, A. Ginsburg, A. M. Price-Whelan, W. E. Kerzendorf, A. Conley, N. Crighton, K. Barbary, D. Muna, H. Ferguson, F. Grollier, M. M. Parikh, P. H. Nair, H. M. Unther, C. Deil, J. Woillez, S. Conseil, R. Kramer, J. E. H. Turner, L. Singer, R. Fox, B. A. Weaver, V. Zabalza, Z. I. Edwards, K. Azalee Bostroem, D. J. Burke, A. R. Casey, S. M. Crawford, N. Dencheva, J. Ely, T. Jenness, K. Labrie, P. L. Lim, F. Pierfederici, A. Pontzen, A. Ptak, B. Refsdal, M. Servillat, and O. Streicher. Astropy: A community Python package for astronomy. , 558:A33, October 2013. doi: 10.1051/0004-6361/201322068.

Astropy Collaboration, A. M. Price-Whelan, B. M. Sipőcz, H. M. Günther, P. L. Lim, S. M. Crawford, S. Conseil, D. L. Shupe, M. W. Craig, N. Dencheva, A. Ginsburg, J. T. Vand erPlas, L. D. Bradley, D. Pérez-Suárez, M. de Val-Borro, T. L. Aldcroft, K. L. Cruz, T. P. Robitaille, E. J. Tollerud, C. Ardelean, T. Babej, Y. P. Bach, M. Bachetti, A. V. Bakanov, S. P. Bamford, G. Barentsen, P. Barmby, A. Baumbach, K. L. Berry, F. Biscani, M. Boquien, K. A. Bostroem, L. G. Bouma, G. B. Brammer, E. M. Bray, H. Breytenbach, H. Buddelmeijer, D. J. Burke, G. Calderone, J. L. Cano Rodríguez, M. Cara, J. V. M. Cardoso, S. Cheedella, Y. Copin, L. Corrales, D. Crichton, D. D'Avella, C. Deil, É. Depagne, J. P. Dietrich, A. Donath, M. Droettboom, N. Earl, T. Erben, S. Fabbro, L. A. Ferreira, T. Finethy, R. T. Fox, L. H. Garrison, S. L. J. Gibbons, D. A. Goldstein, R. Gommers, J. P. Greco, P. Greenfield, A. M. Groener, F. Grollier, A. Hagen, P. Hirst, D. Homeier, A. J. Horton, G. Hosseinzadeh, L. Hu, J. S. Hunkeler, Ž. Ivezić, A. Jain, T. Jenness, G. Kanarek, S. Kendrew, N. S. Kern, W. E. Kerzendorf, A. Khvalko, J. King, D. Kirkby, A. M. Kulkarni, A. Kumar, A. Lee, D. Lenz, S. P. Littlefair, Z. Ma, D. M. Macleod, M. Mastropietro, C. McCully, S. Montagnac, B. M. Morris, M. Mueller, S. J. Mumford, D. Muna, N. A. Murphy, S. Nelson, G. H. Nguyen, J. P. Ninan, M. Nöthe, S. Ogaz, S. Oh, J. K. Parejko, N. Parley, S. Pascual, R. Patil, A. A. Patil, A. L. Plunkett, J. X. Prochaska, T. Rastogi, V. Reddy Janga, J. Sabater, P. Sakurikar, M. Seifert, L. E. Sherbert, H. Sherwood-Taylor, A. Y. Shih, J. Sick, M. T. Silbiger, S. Singanamalla, L. P. Singer, P. H. Sladen, K. A. Sooley, S. Sornarajah, O. Streicher, P. Teuben, S. W. Thomas, G. R. Tremblay, J. E. H. Turner, V. Terrón, M. H. van Kerkwijk, A. de la Vega, L. L. Watkins, B. A. Weaver, J. B. Whitmore, J. Woillez, V. Zabalza, and Astropy Contributors. The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package. , 156(3):123, September 2018. doi: 10.3847/1538-3881/aabc4f.

G. Á. Bakos, A. Shporer, A. Pál, G. Torres, Géza Kovács, D. W. Latham, T. Mazeh, A. Ofir, R. W. Noyes, D. D. Sasselov, F. Bouchy, F. Pont, D. Queloz, S. Udry, G. Esquerdo, B. Sipőcz, Gábor Kovács, R. Stefanik, J. Lázár, I. Papp, and P. Sári. HAT-P-5b: A Jupiter-like Hot Jupiter Transiting a Bright Star. , 671(2):L173–L176, December 2007. doi: 10.1086/525022.

Travis A. Berger, Daniel Huber, Eric Gaidos, and Jennifer L. van Saders. Revised Radii of Kepler Stars and Planets Using Gaia Data Release 2. , 866(2):99, October 2018. doi: 10.3847/1538-4357/aada83.

Michael Betancourt. A Conceptual Introduction to Hamiltonian Monte Carlo. *arXiv e-prints*, art. arXiv:1701.02434, January 2017.

Joshua A. Carter, Jennifer C. Yee, Jason Eastman, B. Scott Gaudi, and Joshua N. Winn. Analytic Approximations for Transit Light-Curve Observables, Uncertainties, and Covariances. , 689(1):499–512, December 2008. doi: 10.1086/592321.

A. Claret. A new non-linear limb-darkening law for LTE stellar atmosphere models. Calculations for -5.0 ¡= log[M/H] ¡= +1, 2000 K ¡= T_{eff} ¡= 50000 K at several surface gravities. , 363:1081–1190, November 2000.

A. Claret and S. Bloemen. Gravity and limb-darkening coefficients for the Kepler, CoRoT, Spitzer, uvby, UBVRIJHK, and Sloan photometric systems. , 529:A75, May 2011. doi: 10.1051/0004-6361/201116451.

Rebekah I. Dawson. Robustly Detecting Changes in Warm Jupiters' Transit Impact Parameters. , 159(5):223, May 2020. doi: 10.3847/1538-3881/ab7fa5.

Rebekah I. Dawson and John Asher Johnson. The Photoeccentric Effect and Proto-hot Jupiters. I. Measuring Photometric Eccentricities of Individual Transiting Planets. , 756 (2):122, September 2012. doi: 10.1088/0004-637X/756/2/122.

Hannah Diamond-Lowe, Kevin B. Stevenson, Daniel Fabrycky, Sarah Ballard, Eric Agol, Jacob Bean, Matthew J. Holman, and Darin Ragozzine. Determining the architecture of the Kepler-297 system using transit timing variations. In *American Astronomical Society Meeting Abstracts #225*, volume 225 of *American Astronomical Society Meeting Abstracts*, page 438.01, January 2015.

Aaron R. Dinner, Erik Thiede, Brian Van Koten, and Jonathan Weare. Stratification as a general variance reduction method for Markov chain Monte Carlo. *arXiv e-prints*, art. arXiv:1705.08445, May 2017.

Néstor Espinoza. Efficient Joint Sampling of Impact Parameters and Transit Depths in Transiting Exoplanet Light Curves. *Research Notes of the American Astronomical Society*, 2(4):209, November 2018. doi: 10.3847/2515-5172/aaef38.

Daniel C. Fabrycky, Jack J. Lissauer, Darin Ragozzine, Jason F. Rowe, Jason H. Steffen, Eric Agol, Thomas Barclay, Natalie Batalha, William Borucki, David R. Ciardi, Eric B. Ford, Thomas N. Gautier, John C. Geary, Matthew J. Holman, Jon M. Jenkins, Jie Li, Robert C. Morehead, Robert L. Morris, Avi Shporer, Jeffrey C. Smith, Martin Still, and Jeffrey Van Cleve. Architecture of Kepler's Multi-transiting Systems. II. New Investigations with Twice as Many Candidates. , 790(2):146, August 2014. doi: 10.1088/0004-637X/790/2/146.

Eric B. Ford, Samuel N. Quinn, and Dimitri Veras. Characterizing the Orbital Eccentricities of Transiting Extrasolar Planets with Photometric Observations. , 678(2):1407–1418, May 2008. doi: 10.1086/587046.

Daniel Foreman-Mackey, David W. Hogg, Dustin Lang, and Jonathan Goodman. emcee: The MCMC Hammer. , 125(925):306, March 2013. doi: 10.1086/670067.

Daniel Foreman-Mackey, Eric Agol, Sivaram Ambikasaran, and Ruth Angus. Fast and Scalable Gaussian Process Modeling with Applications to Astronomical Time Series. , 154 (6):220, December 2017. doi: 10.3847/1538-3881/aa9332.

Daniel Foreman-Mackey, Rodrigo Luger, Eric Agol, Thomas Barclay, Luke Bouma, Timothy Brandt, Ian Czekala, Trevor David, Jiayin Dong, Emily Gilbert, Tyler Gordon, Christina Hedges, Daniel Hey, Brett Morris, Adrian Price-Whelan, and Arjun Savel. exoplanet: Gradient-based probabilistic inference for exoplanet data & other astronomical time series. *The Journal of Open Source Software*, 6(62):3285, June 2021a. doi: 10.21105/joss.03285.

Daniel Foreman-Mackey, Rodrigo Luger, Eric Agol, Thomas Barclay, Luke G. Bouma, Timothy D. Brandt, Ian Czekala, Trevor J. David, Jiayin Dong, Emily A. Gilbert, Tyler A. Gordon, Christina Hedges, Daniel R. Hey, Brett M. Morris, Adrian M. Price-Whelan, and Arjun B. Savel. exoplanet: Gradient-based probabilistic inference for exoplanet data & other astronomical time series, June 2021b.

Benjamin J. Fulton and Erik A. Petigura. The California-Kepler Survey. VII. Precise Planet Radii Leveraging Gaia DR2 Reveal the Stellar Mass Dependence of the Planet Radius Gap. , 156(6):264, December 2018. doi: 10.3847/1538-3881/aae828.

Benjamin J. Fulton, Erik A. Petigura, Andrew W. Howard, Howard Isaacson, Geoffrey W. Marcy, Phillip A. Cargile, Leslie Hebb, Lauren M. Weiss, John Asher Johnson, Timothy D. Morton, Evan Sinukoff, Ian J. M. Crossfield, and Lea A. Hirsch. The California-Kepler Survey. III. A Gap in the Radius Distribution of Small Planets. , 154(3):109, September 2017. doi: 10.3847/1538-3881/aa80eb.

Gaia Collaboration, A. G. A. Brown, A. Vallenari, T. Prusti, J. H. J. de Bruijne, C. Babusiaux, C. A. L. Bailer-Jones, M. Biermann, D. W. Evans, L. Eyer, F. Jansen, C. Jordi, S. A. Klioner, U. Lammers, L. Lindegren, X. Luri, F. Mignard, C. Panem, D. Pourbaix, S. Randich, P. Sartoretti, H. I. Siddiqui, C. Soubiran, F. van Leeuwen, N. A. Walton, F. Arenou, U. Bastian, M. Cropper, R. Drimmel, D. Katz, M. G. Lattanzi, J. Bakker, C. Cacciari, J. Castañeda, L. Chaoul, N. Cheek, F. De Angeli, C. Fabricius, R. Guerra, B. Holl, E. Masana, R. Messineo, N. Mowlavi, K. Nienartowicz, P. Panuzzo, J. Portell, M. Riello, G. M. Seabroke, P. Tanga, F. Thévenin, G. Gracia-Abril, G. Comoretto, M. Garcia-Reinaldos, D. Teyssier, M. Altmann, R. Andrae, M. Audard, I. Bellas-Velidis, K. Benson, J. Berthier, R. Blomme, P. Burgess, G. Busso, B. Carry, A. Cellino, G. Clementini, M. Clotet, O. Creevey, M. Davidson, J. De Ridder, L. Delchambre, A. Dell'Oro, C. Ducourant, J. Fernández-Hernández, M. Fouesneau, Y. Frémat, L. Galluccio, M. García-Torres, J. González-Núñez, J. J. González-Vidal, E. Gosset, L. P. Guy,

J. L. Halbwachs, N. C. Hambly, D. L. Harrison, J. Hernández, D. Hestroffer, S. T. Hodgkin, A. Hutton, G. Jasniewicz, A. Jean-Antoine-Piccolo, S. Jordan, A. J. Korn, A. Krone-Martins, A. C. Lanzafame, T. Lebzelter, W. Löffler, M. Manteiga, P. M. Marrese, J. M. Martín-Fleitas, A. Moitinho, A. Mora, K. Muinonen, J. Osinde, E. Pancino, T. Pauwels, J. M. Petit, A. Recio-Blanco, P. J. Richards, L. Rimoldini, A. C. Robin, L. M. Sarro, C. Siopis, M. Smith, A. Sozzetti, M. Süveges, J. Torra, W. van Reeven, U. Abbas, A. Abreu Aramburu, S. Accart, C. Aerts, G. Altavilla, M. A. Álvarez, R. Alvarez, J. Alves, R. I. Anderson, A. H. Andrei, E. Anglada Varela, E. Antiche, T. Antoja, B. Arcay, T. L. Astraatmadja, N. Bach, S. G. Baker, L. Balaguer-Núñez, P. Balm, C. Barache, C. Barata, D. Barbato, F. Barblan, P. S. Barklem, D. Barrado, M. Barros, M. A. Barstow, S. Bartholomé Muñoz, J. L. Bassilana, U. Becciani, M. Bellazzini, A. Berihuete, S. Bertone, L. Bianchi, O. Bienaymé, S. Blanco-Cuaresma, T. Boch, C. Boeche, A. Bombrun, R. Borrachero, D. Bossini, S. Bouquillon, G. Bourda, A. Bragaglia, L. Bramante, M. A. Breddels, A. Bressan, N. Brouillet, T. Brüsemeister, E. Brugaletta, B. Bucciarelli, A. Burlacu, D. Busonero, A. G. Butkevich, R. Buzzi, E. Caffau, R. Cancelliere, G. Cannizzaro, T. Cantat-Gaudin, R. Carballo, T. Carlucci, J. M. Carrasco, L. Casamiquela, M. Castellani, A. Castro-Ginard, P. Charlot, L. Chemin, A. Chiavassa, G. Cocozza, G. Costigan, S. Cowell, F. Crifo, M. Crosta, C. Crowley, J. Cuypers, C. Dafonte, Y. Damerdji, A. Dapergolas, P. David, M. David, P. de Laverny, F. De Luise, R. De March, D. de Martino, R. de Souza, A. de Torres, J. Debosscher, E. del Pozo, M. Delbo, A. Delgado, H. E. Delgado, P. Di Matteo, S. Diakite, C. Diener, E. Distefano, C. Dolding, P. Drazinos, J. Durán, B. Edvardsson, H. Enke, K. Eriksson, P. Esquej, G. Eynard Bontemps, C. Fabre, M. Fabrizio, S. Faigler, A. J. Falcão, M. Farràs Casas, L. Federici, G. Fedorets, P. Fernique, F. Figueras, F. Filippi, K. Findeisen, A. Fonti, E. Fraile, M. Fraser, B. Frézouls, M. Gai, S. Galleti, D. Garabato, F. García-Sedano, A. Garofalo, N. Garralda, A. Gavel, P. Gavras, J. Gerssen, R. Geyer, P. Giacobbe, G. Gilmore, S. Girona, G. Giuffrida, F. Glass, M. Gomes, M. Granvik, A. Gueguen, A. Guerrier, J. Guiraud, R. Gutiérrez-Sánchez, R. Haigron, D. Hatzidimitriou, M. Hauser, M. Haywood, U. Heiter, A. Helmi, J. Heu, T. Hilger, D. Hobbs, W. Hofmann, G. Holland, H. E. Huckle, A. Hypki, V. Icardi, K. Janßen, G. Jevardat de Fombelle, P. G. Jonker, Á. L. Juhász, F. Julbe, A. Karampelas, A. Kewley, J. Klar, A. Kochoska, R. Kohley, K. Kolenberg, M. Kontizas, E. Kontizas, S. E. Koposov, G. Kordopatis, Z. Kostrzewa-Rutkowska, P. Koubsky, S. Lambert, A. F. Lanza, Y. Lasne, J. B. Lavigne, Y. Le Fustec, C. Le Poncin-Lafitte, Y. Lebreton, S. Leccia, N. Leclerc, I. Lecoeur-Taibi, H. Lenhardt, F. Leroux, S. Liao, E. Licata, H. E. P. Lindstrøm, T. A. Lister, E. Livanou, A. Lobel, M. López, S. Managau, R. G. Mann, G. Mantelet, O. Marchal, J. M. Marchant, M. Marconi, S. Marinoni, G. Marschalkó, D. J. Marshall, M. Martino, G. Marton, N. Mary, D. Massari, G. Matijevič, T. Mazeh, P. J. McMillan, S. Messina, D. Michalik, N. R. Millar, D. Molina, R. Molinaro, L. Molnár, P. Montegriffo, R. Mor, R. Morbidelli, T. Morel, D. Morris, A. F. Mulone, T. Muraveva, I. Musella, G. Nelemans, L. Nicastro, L. Noval, W. O'Mullane, C. Ordénovic, D. Ordóñez-Blanco, P. Osborne, C. Pagani, I. Pagano, F. Pailler, H. Palacin, L. Palaversa, A. Panahi, M. Pawlak, A. M. Piersimoni, F. X. Pineau, E. Plachy, G. Plum, E. Poggio, E. Poujoulet, A. Prša, L. Pulone, E. Racero, S. Ragaini, N. Rambaux, M. Ramos-Lerate, S. Regibo,

C. Reylé, F. Riclet, V. Ripepi, A. Riva, A. Rivard, G. Rixon, T. Roegiers, M. Roelens, M. Romero-Gómez, N. Rowell, F. Royer, L. Ruiz-Dern, G. Sadowski, T. Sagristà Sellés, J. Sahlmann, J. Salgado, E. Salguero, N. Sanna, T. Santana-Ros, M. Sarasso, H. Savietto, M. Schultheis, E. Sciacca, M. Segol, J. C. Segovia, D. Ségransan, I. C. Shih, L. Siltala, A. F. Silva, R. L. Smart, K. W. Smith, E. Solano, F. Solitro, R. Sordo, S. Soria Nieto, J. Souchay, A. Spagna, F. Spoto, U. Stampa, I. A. Steele, H. Steidelmüller, C. A. Stephenson, H. Stoev, F. F. Suess, J. Surdej, L. Szabados, E. Szegedi-Elek, D. Tapiador, F. Taris, G. Tauran, M. B. Taylor, R. Teixeira, D. Terrett, P. Teyssandier, W. Thuillot, A. Titarenko, F. Torra Clotet, C. Turon, A. Ulla, E. Utrilla, S. Uzzi, M. Vaillant, G. Valentini, V. Valette, A. van Elteren, E. Van Hemelryck, M. van Leeuwen, M. Vaschetto, A. Vecchiato, J. Veljanoski, Y. Viala, D. Vicente, S. Vogt, C. von Essen, H. Voss, V. Votruba, S. Voutsinas, G. Walmsley, M. Weiler, O. Wertz, T. Wevers, Ł. Wyrzykowski, A. Yoldas, M. Žerjal, H. Ziaeepour, J. Zorec, S. Zschocke, S. Zucker, C. Zurbach, and T. Zwitter. Gaia Data Release 2. Summary of the contents and survey properties. , 616:A1, August 2018. doi: 10.1051/0004-6361/201833051.

Gregory J. Gilbert and Daniel C. Fabrycky. An Information Theoretic Framework for Classifying Exoplanetary System Architectures. , 159(6):281, June 2020. doi: 10.3847/1538-3881/ab8e3c.

Sivan Ginzburg, Hilke E. Schlichting, and Re'em Sari. Core-powered mass-loss and the radius distribution of small exoplanets. , 476(1):759–765, May 2018. doi: 10.1093/mnras/sty290.

Jonathan Goodman and Jonathan Weare. Ensemble samplers with affine invariance. *Communications in Applied Mathematics and Computational Science*, 5(1):65–80, January 2010. doi: 10.2140/camcos.2010.5.65.

Akash Gupta and Hilke E. Schlichting. Sculpting the valley in the radius distribution of small exoplanets as a by-product of planet formation: the core-powered mass-loss mechanism. , 487(1):24–33, July 2019. doi: 10.1093/mnras/stz1230.

Aaron Hamann, Benjamin T. Montet, Daniel C. Fabrycky, Eric Agol, and Ethan Kruse. K2-146: Discovery of Planet c, Precise Masses from Transit Timing, and Observed Precession. , 158(3):133, September 2019. doi: 10.3847/1538-3881/ab32e3.

Kevin K. Hardegree-Ullman, Jon K. Zink, Jessie L. Christiansen, Courtney D. Dressing, David R. Ciardi, and Joshua E. Schlieder. Scaling K2. I. Revised Parameters for 222,088 K2 Stars and a K2 Planet Radius Valley at 1.9 $R_\oplus$. , 247(1):28, March 2020. doi: 10.3847/1538-4365/ab7230.

Charles R. Harris, K. Jarrod Millman, Stéfan J van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and

Travis E. Oliphant. Array programming with NumPy. *Nature*, 585:357–362, 2020. doi: 10.1038/s41586-020-2649-2.

W. K. Hastings. Monte Carlo Sampling Methods using Markov Chains and their Applications. *Biometrika*, 57(1):97–109, April 1970. doi: 10.1093/biomet/57.1.97.

Matthias Y. He, Eric B. Ford, and Darin Ragozzine. Architectures of exoplanetary systems - I. A clustered forward model for exoplanetary systems around Kepler's FGK stars. , 490 (4):4575–4605, December 2019. doi: 10.1093/mnras/stz2869.

Matthew D. Hoffman and Andrew Gelman. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *arXiv e-prints*, art. arXiv:1111.4246, November 2011.

Tomer Holczer, Tsevi Mazeh, Gil Nachmani, Daniel Jontof-Hutter, Eric B. Ford, Daniel Fabrycky, Darin Ragozzine, Mackenzie Kane, and Jason H. Steffen. Transit Timing Observations from Kepler. IX. Catalog of the Full Long-cadence Data Set. , 225(1):9, July 2016. doi: 10.3847/0067-0049/225/1/9.

Andrew W. Howard, Geoffrey W. Marcy, Stephen T. Bryson, Jon M. Jenkins, Jason F. Rowe, Natalie M. Batalha, William J. Borucki, David G. Koch, Edward W. Dunham, III Gautier, Thomas N., Jeffrey Van Cleve, William D. Cochran, David W. Latham, Jack J. Lissauer, Guillermo Torres, Timothy M. Brown, Ronald L. Gilliland, Lars A. Buchhave, Douglas A. Caldwell, Jørgen Christensen-Dalsgaard, David Ciardi, Francois Fressin, Michael R. Haas, Steve B. Howell, Hans Kjeldsen, Sara Seager, Leslie Rogers, Dimitar D. Sasselov, Jason H. Steffen, Gibor S. Basri, David Charbonneau, Jessie Christiansen, Bruce Clarke, Andrea Dupree, Daniel C. Fabrycky, Debra A. Fischer, Eric B. Ford, Jonathan J. Fortney, Jill Tarter, Forrest R. Girouard, Matthew J. Holman, John Asher Johnson, Todd C. Klaus, Pavel Machalek, Althea V. Moorhead, Robert C. Morehead, Darin Ragozzine, Peter Tenenbaum, Joseph D. Twicken, Samuel N. Quinn, Howard Isaacson, Avi Shporer, Philip W. Lucas, Lucianne M. Walkowicz, William F. Welsh, Alan Boss, Edna Devore, Alan Gould, Jeffrey C. Smith, Robert L. Morris, Andrej Prsa, Timothy D. Morton, Martin Still, Susan E. Thompson, Fergal Mullally, Michael Endl, and Phillip J. MacQueen. Planet Occurrence within 0.25 AU of Solar-type Stars from Kepler. , 201(2):15, August 2012. doi: 10.1088/0067-0049/201/2/15.

John Asher Johnson, Erik A. Petigura, Benjamin J. Fulton, Geoffrey W. Marcy, Andrew W. Howard, Howard Isaacson, Leslie Hebb, Phillip A. Cargile, Timothy D. Morton, Lauren M. Weiss, Joshua N. Winn, Leslie A. Rogers, Evan Sinukoff, and Lea A. Hirsch. The California-Kepler Survey. II. Precise Physical Properties of 2025 Kepler Planets and Their Host Stars. , 154(3):108, September 2017. doi: 10.3847/1538-3881/aa80e7.

Y. K. Jung, A. Udalski, A. Gould, Y. H. Ryu, J. C. Yee, and, C. Han, M. D. Albrow, C. U. Lee, S. L. Kim, K. H. Hwang, S. J. Chung, I. G. Shin, W. Zhu, S. M. Cha, D. J. Kim, Y. Lee, B. G. Park, D. J. Lee, H. W. Kim, R. W. Pogge, KMTNet Collaboration, M. K. Szymański, P. Mróz, R. Poleski, J. Skowron, P. Pietrukowicz, I. Soszyński, S. Kozłowski,

K. Ulaczyk, M. Pawlak, K. Rybicki, and OGLE Collaboration. OGLE-2017-BLG-1522: A Giant Planet around a Brown Dwarf Located in the Galactic Bulge. , 155(5):219, May 2018. doi: 10.3847/1538-3881/aabb51.

David M. Kipping. Binning is sinning: morphological light-curve distortions due to finite integration time. , 408(3):1758–1769, November 2010. doi: 10.1111/j.1365-2966.2010.17242.x.

David M. Kipping. Efficient, uninformative sampling of limb darkening coefficients for two-parameter laws. , 435(3):2152–2160, November 2013. doi: 10.1093/mnras/stt1435.

David M. Kipping. Characterizing distant worlds with asterodensity profiling. , 440(3):2164–2184, May 2014. doi: 10.1093/mnras/stu318.

David M. Kipping and Emily Sandford. Observational biases for transiting planets. , 463(2):1323–1331, December 2016. doi: 10.1093/mnras/stw1926.

Edwin S. Kite, Jr. Fegley, Bruce, Laura Schaefer, and Eric B. Ford. Superabundance of Exoplanet Sub-Neptunes Explained by Fugacity Crisis. , 887(2):L33, December 2019. doi: 10.3847/2041-8213/ab59d9.

Shankar Kumar, John M. Rosenberg, Djamal Bouzida, Robert H. Swendsen, and Peter A. Kollman. The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method. *Journal of Computational Chemistry*, 13(8):1011–1021, 1992. doi: https://doi.org/10.1002/jcc.540130812. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.540130812`.

Johannes Kästner. Umbrella sampling. *WIREs Computational Molecular Science*, 1(6):932–942, 2011. doi: https://doi.org/10.1002/wcms.66. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/wcms.66`.

Yoram Lithwick, Jiwei Xie, and Yanqin Wu. Extracting Planet Mass and Eccentricity from TTV Data. , 761(2):122, December 2012. doi: 10.1088/0004-637X/761/2/122.

Rodrigo Luger, Eric Agol, Daniel Foreman-Mackey, David P. Fleming, Jacob Lustig-Yaeger, and Russell Deitrick. starry: Analytic Occultation Light Curves. , 157(2):64, February 2019. doi: 10.3847/1538-3881/aae8e5.

Kaisey Mandel and Eric Agol. Analytic Light Curves for Planetary Transit Searches. , 580(2):L171–L175, December 2002. doi: 10.1086/345520.

Charles Matthews, Jonathan Weare, Andrey Kravtsov, and Elise Jennings. Umbrella sampling: a powerful method to sample tails of distributions. , 480(3):4069–4079, November 2018. doi: 10.1093/mnras/sty2140.

Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of State Calculations by Fast Computing Machines. , 21(6):1087–1092, June 1953. doi: 10.1063/1.1699114.

Sarah Millholland, Songhu Wang, and Gregory Laughlin. Kepler Multi-planet Systems Exhibit Unexpected Intra-system Uniformity in Mass and Radius. , 849(2):L33, November 2017. doi: 10.3847/2041-8213/aa9714.

Sarah C. Millholland, Matthias Y. He, Eric B. Ford, Darin Ragozzine, Daniel Fabrycky, and Joshua N. Winn. Evidence for a Non-Dichotomous Solution to the Kepler Dichotomy: Mutual Inclinations of Kepler Planetary Systems from Transit Duration Variations. *arXiv e-prints*, art. arXiv:2106.15589, June 2021.

Sean M. Mills, Andrew W. Howard, Erik A. Petigura, Benjamin J. Fulton, Howard Isaacson, and Lauren M. Weiss. The California-Kepler Survey. VIII. Eccentricities of Kepler Planets and Tentative Evidence of a High-metallicity Preference for Small Eccentric Planets. , 157 (5):198, May 2019. doi: 10.3847/1538-3881/ab1009.

F. Mullally, Jeffrey L. Coughlin, Susan E. Thompson, Jason Rowe, Christopher Burke, David W. Latham, Natalie M. Batalha, Stephen T. Bryson, Jessie Christiansen, Christopher E. Henze, Aviv Ofir, Billy Quarles, Avi Shporer, Vincent Van Eylen, Christa Van Laerhoven, Yash Shah, Angie Wolfgang, W. J. Chaplin, Ji-Wei Xie, Rachel Akeson, Vic Argabright, Eric Bachtell, Thomas Barclay, William J. Borucki, Douglas A. Caldwell, Jennifer R. Campbell, Joseph H. Catanzarite, William D. Cochran, Riley M. Duren, Scott W. Fleming, Dorothy Fraquelli, Forrest R. Girouard, Michael R. Haas, Krzysztof G. Hełminiak, Steve B. Howell, Daniel Huber, Kipp Larson, III Gautier, Thomas N., Jon M. Jenkins, Jie Li, Jack J. Lissauer, Scot McArthur, Chris Miller, Robert L. Morris, Anima Patil-Sabale, Peter Plavchan, Dustin Putnam, Elisa V. Quintana, Solange Ramirez, V. Silva Aguirre, Shawn Seader, Jeffrey C. Smith, Jason H. Steffen, Chris Stewart, Jeremy Stober, Martin Still, Peter Tenenbaum, John Troeltzsch, Joseph D. Twicken, and Khadeejah A. Zamudio. Planetary Candidates Observed by Kepler. VI. Planet Sample from Q1–Q16 (47 Months). , 217(2):31, April 2015. doi: 10.1088/0067-0049/217/2/31.

Radford Neal. *MCMC Using Hamiltonian Dynamics*, pages 113–162. 2011. doi: 10.1201/b10905.

Andrew R. Neil and Leslie A. Rogers. A Joint Mass-Radius-Period Distribution of Exoplanets. , 891(1):12, March 2020. doi: 10.3847/1538-4357/ab6a92.

James E. Owen and Yanqin Wu. The Evaporation Valley in the Kepler Planets. , 847(1): 29, September 2017. doi: 10.3847/1538-4357/aa890a.

András Pál. Properties of analytic transit light-curve models. , 390(1):281–288, October 2008. doi: 10.1111/j.1365-2966.2008.13723.x.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12(85):2825–2830, 2011. URL `http://jmlr.org/papers/v12/pedregosa11a.html`.

Erik A. Petigura. Two Views of the Radius Gap and the Role of Light Curve Fitting. , 160 (2):89, August 2020. doi: 10.3847/1538-3881/ab9fff.

Erik A. Petigura, Andrew W. Howard, Geoffrey W. Marcy, John Asher Johnson, Howard Isaacson, Phillip A. Cargile, Leslie Hebb, Benjamin J. Fulton, Lauren M. Weiss, Timothy D. Morton, Joshua N. Winn, Leslie A. Rogers, Evan Sinukoff, Lea A. Hirsch, and Ian J. M. Crossfield. The California-Kepler Survey. I. High-resolution Spectroscopy of 1305 Stars Hosting Kepler Transiting Planets. , 154(3):107, September 2017. doi: 10.3847/1538-3881/aa80de.

Ellen M. Price and Leslie A. Rogers. Transit Light Curves with Finite Integration Time: Fisher Information Analysis. , 794(1):92, October 2014. doi: 10.1088/0004-637X/794/1/92.

Leslie A. Rogers. Most 1.6 Earth-radius Planets are Not Rocky. , 801(1):41, March 2015. doi: 10.1088/0004-637X/801/1/41.

Jason F. Rowe, Stephen T. Bryson, Geoffrey W. Marcy, Jack J. Lissauer, Daniel Jontof-Hutter, Fergal Mullally, Ronald L. Gilliland, Howard Issacson, Eric Ford, Steve B. Howell, William J. Borucki, Michael Haas, Daniel Huber, Jason H. Steffen, Susan E. Thompson, Elisa Quintana, Thomas Barclay, Martin Still, Jonathan Fortney, III Gautier, T. N., Roger Hunter, Douglas A. Caldwell, David R. Ciardi, Edna Devore, William Cochran, Jon Jenkins, Eric Agol, Joshua A. Carter, and John Geary. Validation of Kepler's Multiple Planet Candidates. III. Light Curve Analysis and Announcement of Hundreds of New Multi-planet Systems. , 784(1):45, March 2014. doi: 10.1088/0004-637X/784/1/45.

Jason F. Rowe, Jeffrey L. Coughlin, Victoria Antoci, Thomas Barclay, Natalie M. Batalha, William J. Borucki, Christopher J. Burke, Steven T. Bryson, Douglas A. Caldwell, Jennifer R. Campbell, Joseph H. Catanzarite, Jessie L. Christiansen, William Cochran, Ronald L. Gilliland, Forrest R. Girouard, Michael R. Haas, Krzysztof G. Hełminiak, Christopher E. Henze, Kelsey L. Hoffman, Steve B. Howell, Daniel Huber, Roger C. Hunter, Hannah Jang-Condell, Jon M. Jenkins, Todd C. Klaus, David W. Latham, Jie Li, Jack J. Lissauer, Sean D. McCauliff, Robert L. Morris, F. Mullally, Aviv Ofir, Billy Quarles, Elisa Quintana, Anima Sabale, Shawn Seader, Avi Shporer, Jeffrey C. Smith, Jason H. Steffen, Martin Still, Peter Tenenbaum, Susan E. Thompson, Joseph D. Twicken, Christa Van Laerhoven, Angie Wolfgang, and Khadeejah A. Zamudio. Planetary Candidates Observed by Kepler. V. Planet Sample from Q1-Q12 (36 Months). , 217(1):16, March 2015. doi: 10.1088/0067-0049/217/1/16.

John Salvatier, Thomas V. Wiecki, and Christopher Fonnesbeck. Probabilistic programming in python using pymc3. *PeerJ Comput. Sci.*, 2:e55, 2016. URL `http://dblp.uni-trier.de/db/journals/peerj-cs/peerj-cs2.html#SalvatierWF16`.

S. Seager and G. Mallén-Ornelas. A Unique Solution of Planet and Star Parameters from an Extrasolar Planet Transit Light Curve. , 585(2):1038–1055, March 2003. doi: 10.1086/346105.

Michael R. Shirts and John D. Chodera. Statistically optimal analysis of samples from multiple equilibrium states. , 129(12):124105–124105, September 2008. doi: 10.1063/1.2978177.

Erik H. Thiede, Brian Van Koten, Jonathan Weare, and Aaron R. Dinner. Eigenvector method for umbrella sampling enables error analysis. , 145(8):084115, August 2016. doi: 10.1063/1.4960649.

Susan E. Thompson, Jeffrey L. Coughlin, Kelsey Hoffman, Fergal Mullally, Jessie L. Christiansen, Christopher J. Burke, Steve Bryson, Natalie Batalha, Michael R. Haas, Joseph Catanzarite, Jason F. Rowe, Geert Barentsen, Douglas A. Caldwell, Bruce D. Clarke, Jon M. Jenkins, Jie Li, David W. Latham, Jack J. Lissauer, Savita Mathur, Robert L. Morris, Shawn E. Seader, Jeffrey C. Smith, Todd C. Klaus, Joseph D. Twicken, Jeffrey E. Van Cleve, Bill Wohler, Rachel Akeson, David R. Ciardi, William D. Cochran, Christopher E. Henze, Steve B. Howell, Daniel Huber, Andrej Prša, Solange V. Ramírez, Timothy D. Morton, Thomas Barclay, Jennifer R. Campbell, William J. Chaplin, David Charbonneau, Jørgen Christensen-Dalsgaard, Jessie L. Dotson, Laurance Doyle, Edward W. Dunham, Andrea K. Dupree, Eric B. Ford, John C. Geary, Forrest R. Girouard, Howard Isaacson, Hans Kjeldsen, Elisa V. Quintana, Darin Ragozzine, Megan Shabram, Avi Shporer, Victor Silva Aguirre, Jason H. Steffen, Martin Still, Peter Tenenbaum, William F. Welsh, Angie Wolfgang, Khadeejah A. Zamudio, David G. Koch, and William J. Borucki. Planetary Candidates Observed by Kepler. VIII. A Fully Automated Catalog with Measured Completeness and Reliability Based on Data Release 25. , 235(2):38, April 2018. doi: 10.3847/1538-4365/aab4f9.

G. M. Torrie and J. P. Valleau. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *Journal of Computational Physics*, 23(2):187–199, February 1977. doi: 10.1016/0021-9991(77)90121-8.

V. Van Eylen, Camilla Agentoft, M. S. Lundkvist, H. Kjeldsen, J. E. Owen, B. J. Fulton, E. Petigura, and I. Snellen. An asteroseismic view of the radius valley: stripped cores, not born rocky. , 479(4):4786–4795, October 2018. doi: 10.1093/mnras/sty1783.

Andrew Vanderburg, Saul A. Rappaport, Siyi Xu, Ian J. M. Crossfield, Juliette C. Becker, Bruce Gary, Felipe Murgas, Simon Blouin, Thomas G. Kaye, Enric Palle, Carl Melis, Brett M. Morris, Laura Kreidberg, Varoujan Gorjian, Caroline V. Morley, Andrew W. Mann, Hannu Parviainen, Logan A. Pearce, Elisabeth R. Newton, Andreia Carrillo, Ben Zuckerman, Lorne Nelson, Greg Zeimann, Warren R. Brown, René Tronsgaard, Beth Klein, George R. Ricker, Roland K. Vanderspek, David W. Latham, Sara Seager, Joshua N. Winn, Jon M. Jenkins, Fred C. Adams, Björn Benneke, David Berardo, Lars A. Buchhave, Douglas A. Caldwell, Jessie L. Christiansen, Karen A. Collins, Knicole D. Colón, Tansu Daylan, John Doty, Alexandra E. Doyle, Diana Dragomir, Courtney Dressing, Patrick Dufour, Akihiko Fukui, Ana Glidden, Natalia M. Guerrero, Xueying Guo, Kevin Heng, Andreea I. Henriksen, Chelsea X. Huang, Lisa Kaltenegger, Stephen R. Kane, John A. Lewis, Jack J. Lissauer, Farisa Morales, Norio Narita, Joshua Pepper, Mark E. Rose,

Jeffrey C. Smith, Keivan G. Stassun, and Liang Yu. A giant planet candidate transiting a white dwarf. , 585(7825):363–367, September 2020. doi: 10.1038/s41586-020-2713-y.

Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2.

Lauren M. Weiss, Geoffrey W. Marcy, Erik A. Petigura, Benjamin J. Fulton, Andrew W. Howard, Joshua N. Winn, Howard T. Isaacson, Timothy D. Morton, Lea A. Hirsch, Evan J. Sinukoff, Andrew Cumming, Leslie Hebb, and Phillip A. Cargile. The California-Kepler Survey. V. Peas in a Pod: Planets in a Kepler Multi-planet System Are Similar in Size and Regularly Spaced. , 155(1):48, January 2018. doi: 10.3847/1538-3881/aa9ff6.

J. N. Winn. *Exoplanet Transits and Occultations*, pages 55–77. 2010.

Joshua N. Winn and Daniel C. Fabrycky. The Occurrence and Architecture of Exoplanetary Systems. , 53:409–447, August 2015. doi: 10.1146/annurev-astro-082214-122246.

APPENDIX

A Derivation of lambda

Let $\Lambda(r, z)$ be the area of overlap area between two spheres (Equation 2.5) for a radius ratio, r and projected separation of centers, $z = z(t)$. Making the substitution $z(t_0) \to b$ gives the area of overlap at midtransit, Λ_0. At the grazing/non-grazing transition boundary, $b = 1 - r$ and $\Lambda_0 = r^2$; at the grazing limit where the planet just barely transits, $b = 1 + r$ and $\Lambda_0 = 0$. If we define the convenience variable $\beta \equiv 1 - b$, then the two endpoints of grazing geometries in (β, Λ_0) space for a given r are (r, r^2) and $(-r, 0)$. Connecting a straight line between these two points yields

$$\Lambda_0 = \Big(\frac{0 - r^2}{-r - r}\Big)(\beta + r) = \frac{r}{2}(\beta + r) = \frac{r}{2}\beta + \frac{r^2}{2} \tag{1}$$

Eliminating numerical factors and defining $\lambda \equiv 2\Lambda_0$ gives $\lambda = \beta r + r^2$, which is the definition of λ we presented in Equation 3.1 in the main text.

B Accounting for implicit priors

In general, a transformation from one basis set to another can induce spurious priors unless we properly track the effect of the coordinate transformation on the likelihood. For our specific case, we have defined the transformation $\{r, b\} \to \{\lambda, \gamma\}$ as

$$\begin{aligned} \lambda &= r^2 + \beta r \\ \gamma &= \frac{\beta}{r} \end{aligned} \tag{2}$$

where $\beta \equiv 1 - b$. In this case, the Jacobian will be

$$J = \begin{vmatrix} \frac{\partial \lambda}{\partial r} & \frac{\partial \lambda}{\partial \beta} \\ \frac{\partial \gamma}{\partial r} & \frac{\partial \gamma}{\partial \beta} \end{vmatrix} = \begin{vmatrix} 2r + \beta & r \\ -\beta / r^2 & 1/r \end{vmatrix} \tag{3}$$

and adding its log-determinant to the log-likelihood will account for implicit priors induced by the coordinate transformation.

One common approach for setting priors on r and b is to assume that these variables are uncorrelated and then draw (r, b) pairs uniformly from the physically permissible region of the $r - b$ plane [Espinoza, 2018]. However, this approach introduces a marginal prior on r, more heavily weighting larger values of r. We instead opt to use a proposal distribution which produces an uninformative marginal prior on both r and b, replicating the default `ImpactParameter` distribution implemented in the popular Python package `exoplanet` [Foreman-Mackey et al., 2021a].

Because our variables are correlated in non-trivial ways due to the complicated underlying geometry of the problem, in practice we find it convenient to draw our variables from conditional distributions. For non-grazing transits (the "N" umbrella), we draw r and b sequentially as

$$\begin{aligned} f(r) &\sim \mathcal{U}(r_{\min}, r_{\max}) \\ f(b|r) &\sim \mathcal{U}(0, 1 - r) \end{aligned} \tag{4}$$

where $\mathcal{U}$ is a uniform distribution and $f(b|r)$ is a conditional distribution of b given r. These proposal distributions do not automatically produce the desired prior. The Espinoza [2018] prior can be recovered by adding $\log(1 - r)$ to the log-likelihood, and the `exoplanet` prior can be produced from the Espinoza [2018] prior by subsequently subtracting $\log(1 + r)$.

For grazing transits (the "G" umbrella), in order to produce our desired priors, we can draw λ and γ from the proposal distributions

$$
\begin{aligned}
f(\gamma) &\sim \mathcal{U}(-1, 1) \\
f(\lambda|\gamma) &\sim \mathcal{U}\Big((\gamma+1)r_{\min}^2, (\gamma+1)r_{\max}^2\Big)
\end{aligned}
\tag{5}
$$

Subtracting $\log(2+2\gamma)$ from the log-likelihood reproduces the Espinoza [2018] prior, and subsequently subtracting $\log(1+r)$ reproduces the `exoplanet` prior. Coincidentally, the term required to produce the Espinoza [2018] prior exactly cancels the log-determinant of the Jacobian.

For the transition window (the "T" umbrella), we sample using the hybrid basis $\{r, \gamma\}$. Here, we can draw r and γ from the proposal distributions

$$
\begin{aligned}
f(r) &\sim \mathcal{U}(r_{\min}, r_{\max}) \\
f(\gamma) &\sim \mathcal{U}(0, \gamma_{\max})
\end{aligned}
\tag{6}
$$

where $\gamma_{\max} = 2$ if $r < 0.5$ and $\gamma_{\max} = 1/r^2$ if $r \geq 0.5$. This variable upper bound on γ prevents negative impact parameters from being drawn. The Jacobian determinant for the transformation $\{r, b\} \to \{r, \gamma\}$ is $|J| = 1/r$. When $r < 0.5$, adding $2\log 2r$ to the log-likelihood reproduces the Espinoza [2018] prior, and when $r \geq 0.5$ adding $\log r$ does so. Subsequently subtracting $\log(1+r)$ once again reproduces the `exoplanet` prior.