On the Detection and Characterization of Exomoons Through Survey and Targeted Observations

Alexander Macaulay Teachey

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2020

Abstract

On the Detection and Characterization of Exomoons

Through Survey and Targeted Observations

Alexander Macaulay Teachey

Exomoons remain amongst the most elusive targets in observational astronomy. Nevertheless, these worlds stand to provide an unprecedented window into the formation and evolution of planetary systems. If the Solar System is any guide, we can expect exomoons will be geologically active and diverse, with the potential for hosting volatiles, atmospheres, and even life. Moreover, a thorough understanding of the population and occurrence rates of exomoons will help to place our own Solar System in a galactic context, speaking to the commonality of our own history. And though there are a variety of known pathways for moon formation, the discovery of exomoons may yet reveal heretofore unanticipated system architectures and defy easy explanation, thereby enriching our theoretical understanding of system formation. In this Dissertation I present a population study of exomoons in the *Kepler* data, finding an apparent dearth of Galilean-analog satellites orbiting planets between 0.1 and 1 AU. I then present evidence for a large exomoon orbiting Kepler-1625b – potentially the first ever discovery of a transiting exomoon – as suggested by a joint analysis of *Kepler* and Hubble Space Telescope data. The following chapter further investigates a number of alternative hypotheses relating to the candidate moon, though the conclusion that an exomoon best explains the data in hand remains unchanged. Finally, I present the results of an effort to identify candidate exomoon signals in the *Kepler* data by developing a convolutional neural network trained on $\mathcal{O}(10^5)$ *Kepler* light curves injected with simulated planet and moon transit signals. The most promising exomoon candidates identified by the neural network are examined in detail, undergoing a full photodynamical model fit and Bayesian model selection. I conclude by discussing the outlook for the moon search, highlighting strategies for future work and myriad unanswered questions that should be pursued in the coming years.

Table of Contents

List of Tables

List of Figures

Acknowledgements

"You know, you don't act like a scientist... you're more like a game show host."
– Dana Barrett

How on Earth did I get here?

It's been quite a peculiar journey, and whenever I think back on the last two decades or so I'm freshly amazed. You see, unlike so many of my peers here at Columbia, I really had no intention or expectation of being a scientist when I was growing up. Instead, it was quite obvious to me as a kid that I was going to be a performer. I just loved movies, and the theater, making people laugh, playing roles, doing voices and impersonations, singing and dancing... and people seemed to think I was reasonably good at all that. I thought so too, I guess... and not only that, I really thought I had what it takes to "make it". I'm not talking about becoming a famous actor, necessarily — though every actor secretly dreams of that (I don't think I have the looks for it, quite frankly) – but you know, just a working actor, teaching maybe, doing small shows... that was going to suit me just fine. It was hard for me to imagine anything else that could possibly be so fulfilling, honestly, as the life of an actor. So eventually I went off to school to study the theater, and that seemed to go pretty well, too. For a long time I was on that path, and I loved it. It was a great privilege, and I met so many fabulous artists along the way.

Eventually, though – after graduation, getting an office job, knowing what it's like to have money in my pocket, and finding out I'm actually fairly allergic to the life of an actor (which, it turns out, is at least as much about selling yourself, *looking* the part and *getting* the part as it is about being able to *act* the part... to say nothing of the financial insecurity and nomadic lifestyle) – I came around slowly and painfully to the conclusion that that life really just wasn't for me after all. I felt like a failure, to be honest... at a dead end, and only 25 years old! What could I possibly do that would be as rewarding as life in the theater? What else can I do that contributes something to the world, and that also never gets boring?

What else am I good at? Why must I insist on having a career that's *interesting*?

Of course if you're reading this you know how it all turned out, so I'll spare you all the gory details.

The jump from actor to scientist is perhaps not quite so huge or bizarre, though, as some people might think; after all, theatre and science are both creative endeavors, – though the latter isn't always thought of that way – and both are jobs one might pursue if you tend to think, as I do, that we've all only got one shot at this thing called life, so we better well make it count. And, like actors, astronomers must also somehow embrace this strange duality of our particular field, which may be considered at once both profoundly important and highly impractical. You see, an actor can be keenly aware of the importance of their work – "holding up a mirror to humanity", and all that – and also plainly acknowledge that their job is to put on a costume and maybe a funny voice and play pretend for two hours. Similarly, we astronomers are the keepers of an ancient tradition, explorers of the vast cosmic wilderness, adherents to the highest standards of scientific rigor and scrutiny – and yet, it is not so easy to point towards the practical *necessity* of our work in many cases... especially when we're met with hostile questions along these lines by people who are decidedly skeptical of our value. Sure there are some branches of astronomy that are indeed of significant concern to all of us down here on the Earth – space weather, identifying and monitoring near-Earth objects, and climate studies of other planets come to mind – and of course historically so-called pure scientific efforts have paid innumerable practical dividends. But these days most of us puzzle over rather esoteric points of interest. I don't mean to diminish these efforts at all; they are thrilling, and inspire awe amongst astronomers and non-astronomers alike. But do we *need* these discoveries? Is this even a *job? How am I getting paid for this?!* I feel I've bitten off more than I can chew with this particular line of questioning, but maybe you catch my meaning. At least insomuch as we need the theatre, and film, and music, and paintings, and sculpture – yes, we absolutely need astronomy. And I trust that our efforts will pay some unimagined dividends in an unimaginable future.

I've accumulated a lot of letters along the way: BFA, BA, MA, MPhil, and soon PhD... I think it's time to give it a rest. From New York University, to Hunter College, to Columbia, I have been blessed by a life lived (so far) in the academy. It's by no means assured I'll stay here forever, but whatever I end up doing, I will always remember fondly the time I have spent here, the people I've come to know and love, and the incredible projects to which I've had the good fortunate of contributing.

For now though it looks like I'll get to stick around a bit longer! I am beyond excited to start the next chapter of my academic life as a postdoctoral fellow at the Academia Sinica Institute of Astronomy & Astrophysics in Taipei. I am honored and thrilled that they have taken a chance on me to continue with the work I have started here at Columbia.

Before I get on to thanking everyone who has helped me get where I am today, I'd first like to acknowledge some folks I have never actually met. First let me express my profound thanks to the innumerable scientists, engineers, administrators, and astronauts for their role in making the observations and analysis in this work possible. I am thinking in particular about those who have built and operated our ground-based telescopes, those who have constructed and launched our incredible space-based observatories, the crews of six Space Shuttle missions who put their lives on the line to launch and service the great Hubble Space Telescope, the designers of the incredibly powerful supercomputers on which our analysis has been performed, and those who have been responsible, in large ways and small, for championing science through political and administrative means to enable us to make these amazing discoveries about our Universe. We are forever indebted to you for making this work possible.

I also wish to acknowledge the indigenous peoples of this nation, whose lands were taken from them, and defiled, by European colonizers during an age of hideous oppression and subjugation. I am an inheritor of that obscene legacy, so I believe it is my responsibility to meditate on that history, remember the rich culture and heritage of this land's true natives, now nearly lost, and do what is in my power to ensure that we never repeat the terrible

deeds of our ancestors. May we learn to live in harmony with the natural world and treat respectfully our fellow human beings.

Along these same lines, let me also say something about the privileges I have had as a middle class white male, whose ancestry in this country stretches back to the 1700s. A great deal of this country was built with slave labor, and indeed some of my own ancestors enslaved other human beings. The plight of African Americans of course did not end with the Civil War, but remains with us even today. From the education and relative wealth of my family, to the schools I attended, the opportunities I had, and the countless other privileges afforded to whites in this country by virtue of our skin color alone – I did not earn these privileges, but there is no doubt in my mind that my accomplishments in this life could have easily been foreclosed without them. We owe a debt of gratitude to the people of color who built and continue to contribute to the greatness of this country despite their historic mistreatment. And more than that, we are obliged to right these historic wrongs.

I feel compelled also to address a new tragedy that is happening even as we speak. As I write this, we are witnessing an unprecedented attack on immigration into this country which, as we all know, is a nation of immigrants. Foreign students and professionals, who work incredibly hard to come here, to better themselves and to contribute to our society, are being systematically maligned, locked out and forced out of this country under the guise of protecting American workers. Of course, these new policies will do no such thing, but they will inflict significant harm to our economy, our great universities, and most devastatingly, to the lives of everyone involved. I am appalled by the short-sightedness and cruelty of the government's efforts in this regard, especially in the midst of the global COVID-19 crisis, and I stand in solidarity with everyone suffering under these disgraceful policies – including many of my friends and colleagues, who have enriched my life enormously, and who are the very embodiment of that spirit we have sometimes called "the American dream."

Finally, on a more cheerful note, I want to express my tremendous gratitude and love for the LGBTQ+ community. We have come so far, thanks in large part to the activism

of folks who put themselves at great risk to come out, to advocate for equal rights, and to raise awareness especially during the AIDS epidemic. Long-entrenched attitudes of moral superiority, bigotry and hatred – these are steadily eroding, and today I have the freedom to be open about who I am, love who I love, marry who I want to marry, and not fear termination from my employer on the basis of my orientation. I remember all those who have been rejected, abused, imprisoned, tortured, or even killed, simply for being themselves, and I am mindful, of course, of the continuing struggles of sexual minorities around the world, and indeed, even still in this country. There remains a great deal of work we must do to expand equal protection under the law, and guarantee access to critical healthcare. I am proud to stand alongside my community and our allies in continuing to pursue a more just world.

* * *

On a personal level, I first wish to express my profound thanks to my thesis advisor, David Kipping, who graciously took me in as a first year student to work on exomoons – a problem which he quite rightly warned me was exceptionally challenging, and which I quite characteristically underestimated. My fascination with the subject predated that first meeting, so I jumped right in and never looked back. After all we've been through, the remarkable ups and downs over these last five years, I count myself incredibly lucky to have found this work, and to have had David there every step of the way to guide me. My instincts and skills as an astronomer are far greater than they were when I arrived at Columbia, thanks in large part to his tutelage, and I feel extraordinarily privileged to work in this small but incredibly exciting field.

Of course I cannot forget the wonderful education, guidance, and support I have received from my other mentors along the way: Tim Paglione, Kelle Cruz, Marcel Agüeros, Betsy Mills, Jürgen Ott, and Dave Meier. It has been my privilege learning from them all, and of course, I would not be here today without them. I also wish to thank the folks who have been involved with my thesis committee at some point or other: Marcel (again), Jacqueline

Van Gorkom, Caleb Scharf, Melissa Ness, Rebecca Oppenheimer, and Andrew Howard. I'm honored to have worked with these scientists, and my sincere thanks to everyone who has signed on to read this monstrosity.

Indeed I also wish to thank the rest of the faculty in the Department of Astronomy. It is no exaggeration to say I have learned so much from them all, and they all have my deep admiration. I must confess that their various idiosyncrasies have often been fodder for my irrepressible urge to impersonate, to the occasional delight of my fellow students, and my delight in amusing them. But if any of the faculty are reading – a few certainly are – they may rest assured, every impersonation has been rendered with love, born from my own idiosyncratic zeal for identifying and reproducing idiomatic behaviors and speech patterns. I tend to think I'm quite fortunate that I will be able to recall you all in the most vivid detail long after I've left these halls.

Thanks is also due to the collaborators with whom I've worked on various papers and proposals over the years. Some I have already mentioned, so to them and to everyone I'm about to name, my sincere thanks for their help and their contributions: Ryan Abrahams, Ruth Angus, Al Schmitt, Paul Dalba, Stephen Kane, and Chris Burke. Thanks also to folks in the astronomy community who have been particularly supportive and encouraging towards me and this work over the years, including Emily Rice, Eric Agol, Matthew Kenworthy, Judit Szulágyi, Mercedes López-Morales, and Giovanna Tinetti. Not everyone is an exomoon enthusiast, so it has been great having these folks in our corner. I'm sure I'm leaving some people out of this list, my apologies.

I must also extend my deep gratitude to our two fabulous administrators in Columbia Astronomy, Ayoune Payne and Millie Garcia, who tirelessly support the operations of the department, cheerfully and meticulously. Having been a department administrator myself, I know better than most the wide variety of their seen and unseen responsibilities, and have had my fair share of interactions with administrators over the years. It is often a thankless job, and I have met more than a few who are, shall we say, visibly less than enthusiastic

about their work. All that's to say, having good administrators is a blessing, and I can't thank them enough for the countless times they've helped me over the years.

My life has been enriched enormously by the friendship and good cheer of the other graduate students and postdocs at Columbia Astronomy, past and present. I have learned so much from you all, too, and will cherish the memories of our silly exploits forever. As tempted as I am to start naming everyone here, it really would be unwieldy, and I would be embarrassed to leave someone out. You all have meant so much to me. Even so, I hope it will not be too crass to say that I am particularly grateful for my friends Nicole Melso and Emily Sandford, two truly inspiring women who have been my best buds since the very beginning. I'm afraid they are both too modest to realize how remarkable they are, so I do my best to remind them periodically. I will miss everyone here at Columbia terribly, but I know our paths will cross many times in the years to come.

While I'm at it, let me also acknowledge the fabulous astronomers I met and worked along side at CUNY, the American Museum of Natural History, and the National Radio Astronomy Observatory. I'm also grateful for the professional outreach collaborations in which I've had the good fortune of participating, including Astronomy on Tap, the Weekly Space Hangout, and Out In Space. My science outreach opportunities have been so much fun. Oh, and of course, thanks to the remarkably energetic and hardworking undergraduates here at Columbia and Barnard: the astronomy majors, and my students!

Let me also thank (some of) the other teachers I have not yet mentioned who have had a powerful impact on my life, from middle school up through college: Jim Frizzell, Mary Spears, Jim Pennington, Homer "Butch" Alberti, Laura McGlinchey Akesson, Rita Bsat, Andy Koebler, Sally Maynard, Susan Finch, Paul Urcioli, Patrick Huggins, Lim Hyungsik, and Chen Ying-Chih. Teachers rarely know how much they mean to their students, and how long after their lessons they're still remembered. Not all of these folks are still with us, unfortunately, but I hope the ones that are know their impact is immeasurable.

I must also thank my broader group of friends in New York and elsewhere for their many

years of support and encouragement, particularly during the years when I was transitioning careers. It's hard to believe that I have known some of them for more than 20 years! I have watched so many of them develop their own fabulous careers, and start families, and I have been exceedingly proud to be a part of their lives and have them in mine. There really are too many to thank, but please forgive my listing a sizeable number of them here (in no particular order!): Brad Mielke, Jolenta Greenberg, Josh Young, Samir Ashraf, Juan Rosales, Sam Litton, Arwa Gunja, Oz Hazel, Mimi Wong, David DeKeyser, Pricilla Cheng, Takuma Mizuno, Leah Flax, Jaeah Lee, Emory Al-Imam, Seth Flaxman, Jim McCormick, Dan Abeles, Jamie Kelley, Sean O'Donnell, Nathan Wilson, Beau Cribbs, Brawner Alcorn, Sanders Pearsall, Ashley Bell, Matt Schock, Rachel Walker, Justin Lauro, Natalia De Sosa, Dan Halden, Mike Sidman, Jason Desmarais, Matt Hadley, Mike Piazza, Brett Underhill, Ellie Vance, John Moran, Kenneth French, Sandi Adams, Munazza Alam, JP Ventura, Jake Taylor, Malena Rice, Mark Popinchalk, Eileen Gonzales, Paige Godfrey, Pat Huang, Katie McNish, Ke Ma... the list goes on and on. My apologies to those I've left off.

I am obliged to thank my longtime partner, Johnson, whose support during my years at Hunter College especially made this all possible. I truly would not be here today without him, and though we have since gone our separate ways, I will be forever grateful for his love, encouragement, and generosity during our time together. In all honesty, I can never repay him. Thanks also to his wonderful family for their warmth and embrace during our time together.

Finally, I wish to thank my own loving family. The Macaulays and the Teacheys have always meant the world to me. I feel incredibly fortunate to have such a large, warm and fun-loving family, and they have given me so much laughter over the years. I'm quite proud of these roots. One thing I realized a while back was just how many educators we have! All four of my grandparents, both my parents, three aunts, one uncle, and two cousins – all teachers. It must run in our family. Thanks to everyone who has enriched my life from my earliest memories to the present.

And now to my Mom, Dad, and brother. They each get their own paragraphs.

In some ways my brother and I are quite different: most obviously, I relish attention, while he tends to avoid it. And as kids we weren't exactly the best of friends. But as we grow older, I've learned that we have so much in common, and there is much I admire in him. He's dependable, forthright, practical, methodical, and honorable. I am, well, only some of those things. It's always a blast joking around with him – making obscure references to *Seinfeld*, or some movie only he and I know because it was taped off TV in the early 90s – and I just love playing music together. We don't get enough opportunities. Burt, I'm very fortunate to have you as my big brother.

My father, Philip Teachey, is an exceedingly interesting man. He is most notably an exceptionally talented musician, particularly well regarded during his tenure as principal oboist of the Richmond Symphony. I have also had the great fortune of witnessing his musicianship at the piano. He has the keenest ear, which he has employed also as a violinist, as a self-taught piano technician and refurbisher, and as a sometime critic of my own playing, singing and composing! It has been a privilege to inherit some of my musical talents from him, to learn from him, to try to emulate him. Dad is also a gifted woodworker and carpenter, having crafted many beautiful pieces of furniture, and even designing and building his own house from the ground up... before going on to single-handedly restore the farmhouse in which my grandmother was born. And these days, he enjoys tending to that same farm, providing the music in local churches, and helping folks dig into the local history at the public library in Sampson County, North Carolina. Not only have I inherited some musicality from him, and some of my intellect; I also seem to have gotten my speaking voice, my dry wit, and evidently, my impending baldness from him. Thanks Dad.

My mother, Pat Macaulay, is a most remarkable woman, who worked incredibly hard to raise her two sons, and did it with grace and positivity. From piano lessons to choir practice, boy scouts, play rehearsals, cotillion, youth trips, band trips... she was there through it all. She made costumes, and trombone slide covers! Whatever needed to be done, you could count

on Mom. By the time she retired from the University of Richmond she was widely regarded there as an institution, and she particularly enjoyed recounting the history of the University for new arrivals. She is most known by her own enormous group of friends for her tireless devotion to supporting the church and its congregation, her immutable sense of humor, and her inimitable storytelling. On top of that, I have known her as the person who has always been there for me, helped me through tough times, thrilled in my accomplishments, and encouraged my curiosity. My fascination with the world, with science, with history, with language, with people – these were both inherited from and fostered by her. I can't imagine a more loving mother. Thank you Mom.

I love you all more than I can say. Thank you for being a part of my life.

July 2020

Dedication

To the makers of music – all worlds, all times.

Preface

I tend to think that the search for exomoons is among the most exciting objectives in observational astronomy today. Clearly not everyone agrees, as evidenced by the fairly small number of other astronomers actively working on this problem. Sure, there is a growing pile of papers out there on the subject, both observationally and theoretically oriented – many of which I will cite in the coming pages. But to my knowledge, this is only the third *dissertation* on the subject (Lewis, 2011; Kipping, 2011; Noyola, 2015). In my conversations with exoplanet colleagues around the world, I sense broad interest in exomoons. So why hasn't the subject caught on, like so many others? I believe there are a few reasons.

To start, the barrier to entry is quite high. I have been incredibly fortunate to study under one of the pioneers of this field. It is my hope to continue working on exomoons in the future, because hard-won expertise needs to be propagated somehow, not just through papers, but also through mentorship. Most researchers interested in exomoons, however, are necessarily self-taught, and that tends to be a bit more difficult. I confess that, were I in their shoes these last five years, without an experienced mentor to guide me, it's quite likely that I would have been rather intimidated and overwhelmed by the array of challenges, despite my sincere enthusiasm. As I grow into my next and future roles, I hope and expect I will be less afraid of diving into such uncharted waters on my own. But as a graduate student especially, it can be quite daunting. All the more reason to be impressed with my advisor's early work on the subject, as an intrepid graduate student himself.

The other major reason why I think we still don't see much buy-in from the community

is broad skepticism that these worlds can even be detected in the first place. I must say, this seems to me like a lack of vision, and frankly depresses me. I confess I myself am frequently pessimistic about what we can accomplish in observational astronomy in the next, say, twenty or thirty years (that is, the likely duration of my career as an astronomer). There are enormous obstacles to achieving our lofty goals. Even with huge new telescopes on the ground, and incredibly sophisticated next-generation space-based observatories, we have a long way to go before we can characterize exoplanets the way we'd like, with detailed understanding of atmospheric compositions, imaging, and identifying biosignatures. On the other hand, we see remarkable innovations in exoplanetary science on almost a daily basis! With this in mind, while we can all individually only dream of what we might be able to do down the line, we have to have some faith in the community's ingenuity. There is good reason to think we really can achieve incredible things, through a joint effort.

Coming back to exomoons, I fully expect we can make these same great strides, *if* we work together, *if* we get people working on the problem. We simply cannot expect to make great strides if we only have ten or twenty astronomers in the world making any effort along these lines. I sure can't do it alone. To quote Virgil Sollozzo, "you think too much of me, kid. I'm not that clever!"

A useful point of reference is the recent, astonishing history of exoplanets. Exoplanet discoveries began only in the early 1990s. Here, some thirty years later, we now sit atop a mountain of exoplanet discoveries, and indeed, they are now so common that writing a paper about a single new planet discovery seems like overkill, unless it is some kind of superlative. As such, the community has now transitioned to other objectives beyond detection – atmospheric characterization, and teasing out formational histories through studies of the exoplanet *population*. These studies might have been unfathomable at the dawn of the exoplanet era. One can easily imagine a skeptical, conventionally-minded astronomer of the time saying, "look, it's really hard to find an exoplanet in the first place, and you want to *look at its atmosphere?* You want to *understand the population?*" Yet, here we are.

The study of exomoons now feels very much like where the field of exoplanet research was decades ago: a few tentative detections, hints of what might be out there, and maybe some wholly unexpected objects. There is, rightly, a good deal of skepticism about these possible detections in the literature. I don't expect many astronomers think exomoons are uncommon, or unlikely. But there seems to be a broad lack of imagination about what we can learn by actually finding these moons, not just a few of them, but a whole host of them. It is amazing now to think about the degree of our ignorance a few decades ago when it comes to exoplanets. We are at this stage now with exomoons, but we have to forge ahead. Just as detecting exoplanets once seemed like a fool's errand, detecting exomoons evidently seems that way to many folks in the community. Is it really? I don't think so. Perhaps we have been too good at identifying and communicating the ongoing challenges to making these discoveries. But highlighting these difficulties is the best way to get someone on the job of trying to actually solve them.

This dissertation is focused on the work I have done, along with some esteemed colleagues, to push the field ever so slightly forward. We have utilized both survey and targeted observations to carry out the search for exomoons and, to the extent possible, tease out some understanding of what is going on in these distant exoplanetary systems that we've studied. This work is detailed in Chapters 2 through 5.

But before we dive in to the work itself, it is useful to survey the state of the field more broadly. In the following introductory chapter I will therefore begin by tracing a brief history of our understanding of moons in our Solar System and beyond (section 1.1). I will then spend some time highlighting the myriad motivations for finding and characterizing exomoons (section 1.2). From there we will examine the various detection methods that have been proposed to find these worlds (section 1.3), and I will conclude the chapter by discussing some of the lingering challenges for this work (section 1.4).

Chapter 1: An Introduction to the History, Motivations, Approaches, and Challenges of Finding and Characterizing Moons

"We choose to go to the Moon, not because it is easy, but because it is hard."
- President John F. Kennedy

1.1 A Very Brief History

1.1.1 Our Moon

Every human being who has ever lived has lived under the influence of the Moon. It has been ever present in our lives; chasing the setting Sun or announcing the dawn, sailing serenely overhead and casting long shadows in the darkness... throughout the world the Moon has been a source of wonder, and terror. We have imagined it as the doleful companion of our mighty Sun, bewitching and beguiling.

The deification of the Moon in cultures throughout the world suggests its central place in our collective consciousness, and its importance in ancient society. Little wonder; it is not merely a ghostly presence in the night. It is also nature's most obvious timepiece for timescales longer than a day. For the earliest farmers, hunters and foragers living in seasonal lands, tracking the progress of the year might have been a matter of life and death, and there are reports of lunar calendars that are tens of thousands of years old – far older than any solar calendar (Rappenglück, 2014). The cycles of the Moon are easily monitored, without sophisticated observations, and it should hardly come as a surprise that lunar calendars remain relevant today among cultures who have retained a connection to their ancient traditions and, in particular, continue their traditional festivals.

For millennia the Moon remained thus – a mysterious presence in the sky, a natural clock, maybe a god or goddess. Starting around the 5th century BCE, however, the first intimations

of what the Moon *really* is began to take shape, through the observations and philosophical speculations of Babylonian, Greek, Indian, and Chinese astronomers. The illumination and phases of the Moon, its true size and distance, the prediction of eclipses, even the cause of ocean tides... all these were first supposed by these early observers. Even so, a deeper understanding of these phenomena would have to wait until the advent of modern scientific inquiry during the European Renaissance.

It was Galileo Galilei who first reported the dazzling array of surface features he saw on the Moon, utilizing (according to his own account) a telescope of his own manufacture based on "rumors" of such a device invented in Holland. He recorded his observations in his *Sidereus Nuncius* (1610) as follows:

> The brighter part appears to surround and pervade the entire hemisphere, but the darker part, like some cloud, stains its very face and renders it spotted. Indeed, these darkish and rather large spots are obvious to everyone, and every age has seen them. For this reason we shall call them the large or ancient spots, in contrast with other spots, smaller in size and occurring with such frequency that they besprinkle the entire lunar surface, but especially the brighter part. These were, in fact, observed by no one before us. By oft-repeated observations of them we have been led to the conclusion that we certainly see the surface of the Moon to be not smooth, even, and perfectly spherical, as the great crowd of philosophers have believed about this and other heavenly bodies, but, on the contrary, to be uneven, rough, and crowded with depressions and bulges. And it is like the face of the Earth itself, which is marked here and there with chains of mountains and depths of valleys. *(Translation by Albert Van Helden)*

Galileo went on to describe and illustrate, in breathtaking detail, the evidence for these conclusions, and correctly interpreted the phenomenon we know today as Earthshine. From then on, the Moon would be in the collective consciousness a world something like the Earth, waiting to be explored – that is, if it could be reached somehow.

Stories of journeys to the Moon and its imagined inhabitants abound in literature beyond mythological treatments, with several examples predating Galileo's revelations. The earliest extant tale about an imagined visit to the Moon, slyly entitled *A True Story*, is attributed to the Assyrian satirist Lucian of Samosata (c. 125 - 180 CE). In Japan, *The Tale of the Bamboo Cutter* dating from the 10th century CE tells the story of Kaguya-hime, a princess of the people living on the Moon. In time, with the dawning of modern astronomy, there

came to be ever more works involving journeys to the Moon, including Johannes Kepler's own *Somnium*, written just two years before *Sidereus Nuncius* but published posthumously in 1634.

Following Galileo's pioneering work, imaginative stories proliferated, and increasingly, these stories began to take on more realistic contours. By the time of Jules Verne's *From the Earth to the Moon* (1865), the notion that humans might actually be able to make such a journey was becoming ever more plausible. In Verne's novel, he imagined employing a giant cannon to launch his astronauts to the Moon. In reality, the enormous acceleration force required to launch astronauts to the Moon this way would probably be lethal, but it is arguably closer to the approach that finally prevailed than the flying machines and balloons that were common to earlier stories.

The early twentieth century saw a period of rapid technological advancement – including the advent of modern rocketry, pioneered by the likes of Robert Goddard, Hermann Oberth, Robert Esnault-Pelterie, Konstantin Tsiolkovsky, and later, Werner von Braun and Sergei Korolev. Government support for these efforts had a decidedly belligerent component; it was Nazi Germany who first crossed the boundary into space with their V-2 rocket in 1944, demonstrating the far-reaching military applications of such a technology. Following the end of the second World War, The United States and the Soviet Union scaled up their own rocketry efforts dramatically, towards a capability of delivering nuclear warheads to locations around the world on intercontinental ballistic missiles. At the same time, peaceful and scientific applications of these new technologies were also pursued – the lofting of the first artificial satellites for communication and weather forecasting, crewed exploration of low-Earth orbit and cislunar space, and even interplanetary robotic missions. These advancements, in slow but deliberate succession, would eventually culminate in the Apollo program's placing human beings on the surface of the Moon in 1969 – arguably among the crowning achievements of the human species.

A tiny fraction of the Moon's surface was explored by a mere twelve American astronauts,

from 1969 to 1972. A few hundred kilograms of Moon rocks were collected and returned to the Earth for study, and various experiments were carried out and left on the surface. At the time, settlements on the Moon may have seemed inevitable, perhaps even imminent. Stanley Kubrick's 1968 film *2001: A Space Odyssey* (written with Arthur C. Clarke) imagined what then seemed quite plausible: multinational Moon bases, an enormous rotating space station with artificial gravity, and routine business space travel by – you guessed it – the year 2001. But in reality, the United States turned its attention away from continued lunar exploration, shifting focus towards crewed missions to low-Earth orbit. Even so, the Moon has remained an important target for uncrewed exploration, and has been visited by robotic spacecraft from the U.S., Russia, China, Japan, India, and Israel. The surface has been mapped in exquisite detail by NASA's Lunar Reconnaissance Orbiter, and the Chinese Cháng'é mission series has placed a rover on the far side of the Moon. NASA's current plans call for returning humans to the surface of the Moon by 2024 through the Artemis program, though whether this is an achievable timeframe is a matter of considerable debate.

Regardless, renewed interest in crewed missions lays the groundwork for still bolder visions of a future on the Moon: mining for rare materials, building telescopes, using the Moon as a launch platform, even the construction of a space elevator (Penoyre & Sandford, 2019). In light of the enormous difficulties in securing and retaining funding for ambitious space programs, realization of these plans may seem as remote as ever. But there is reason for some optimism; the burgeoning commercial space economy has already reduced launch costs dramatically, and there is good reason to expect this new paradigm will continue to result in further cost savings while also better identifying real economic benefits of ambitious exploration and utilization of space resources. If these visions do in fact come to fruition, however, it will be crucial that we revisit and expand upon the legal frameworks for commercial use and militarization of space.

It has been an incredible journey, thousands of years in the making. And yet, the Moon is just our nearest neighbor. It calls out to us for exploration, and teases the great wonders

that await us across the vast interplanetary and interstellar oceans of space. Moons abound.

1.1.2 Moons of our Solar System

It was Galileo also who discovered in 1610 the four largest moons of Jupiter, worlds which today still bear his name. This discovery was of course monumental in its own right – the first moons discovered besides our own! – but it also played an important role in supporting the Copernican heliocentric model; by analogy of Jupiter being its own little planetary system, these revelations in turn gave rise to a Keplerian and subsequently Newtonian understanding of planetary motion. On top of this, observations of Jupiter's moon Io by the Danish astronomer Ole Rømer led to the first calculation of the speed of light, by measuring the changing duration of eclipses, correlated with the changing distances of Jupiter from Earth. It is no exaggeration then to say then that the discovery of the Galilean moons is one of the key milestones in the development of modern physics. Apart from our Moon, the Galileans were the only other moons known for another 45 years, until the discovery of Saturn's largest moon Titan in 1655 by Christiaan Huygens. Four more Saturnian moons were discovered over the years by Huygens and Giovanni Cassini before the close of the 17th century, and then it would be a full century before two more moons, this time orbiting Uranus, were discovered by William Herschel. The early history of moon discovery reads as a who's who of astronomy pioneers.

Until the advent of interplanetary spacecraft, we could of course only see the Galilean moons through a telescope. The largest telescope of the pre-spaceflight era was the Hale 200-inch at Palomar, which based on its aperture has a theoretical resolution of $\sim 0.02''$ at 0.5 μm but was of course seeing-limited at the time. Historical seeing records are not easy to find, but we know that at today's most pristine astronomical sites we can achieve no better than about $0.4''$; that is, we have no hope of resolving features below this threshold without adaptive optics technology. Jupiter's largest moon Ganymede never exceeds an apparent diameter of about $\sim 1.8''$, so making out surface features on the Galilean moons through a ground-

based telescope at the time would be exceptionally challenging. Nevertheless, through the ingenuity of astronomers, we could infer some important features of these worlds, such as their albedo and color, and even draw up crude maps of the surface (e.g. Reese, 1951). The true nature and extent of the marvels that lay in store for us in the outer Solar System could only be guessed.

Our rapidly evolving understanding of these worlds then was set in motion by the launch of the robotic interplanetary spacecraft *Pioneer 10* in March 1972, which would become the first mission to visit the Jovian system. Subsequent visits by *Pioneer 11*, *Voyagers 1* and *2*, *Galileo*, and *New Horizons* have revealed the astonishing beauty and rich geological features of these worlds. Meanwhile, the *Cassini-Huygens* mission to Saturn revealed its host of satellites and its ring system in exquisite detail, and even placed a lander on the surface of Titan. Moons of Uranus, Neptune, and now Pluto have all been observed close-up, and the scientific yield from these missions is immeasurable. Among the most notable examples, we have found active volcanoes on Io (Smith et al., 1979), and inferred the presence of a liquid ocean underneath the icy surface of Europa (Khurana et al., 1998). Another liquid ocean is believed to exist under the surface of Saturn's moon Enceladus (Postberg et al., 2009), based on compositional analysis of geysers at its south pole. On Titan we have observed lakes of liquid hydrocarbons (Stofan et al., 2007).

These results are not *just* scientific curiosities. They help to reveal the complexity of processes that were evidently at work during the formation of these systems, and indeed, the gravitational interactions they still feel to this day, which have a powerful geological influence on these moons. Moreover, the presence of liquid water carries with it the real possibility that life could have arisen, or could even now be present, under the surfaces of these moons. The potential for finding evidence of life past or present in the Solar System is quite real, and extends beyond Mars. Great care must be taken, then, to ensure that we do not pollute these worlds with microbes from the Earth.

Planned missions to Europa and Titan stand to extend our understanding of these worlds

enormously. Both of these moons are generally considered attractive targets from an astrobiological perspective. With an anticipated launch date of 2024, the *Europa Clipper* (Phillips & Pappalardo, 2014) will perform a planned 44 flybys of Europa over a 3.5 year residence in orbit around Jupiter, mapping Europa's surface and preparing for an eventual lander mission. The spacecraft may also fly through a geyser plume, thereby providing a direct sample of the water beneath the icy surface.

Expected to launch in 2026, *Dragonfly* (Turtle et al., 2017) will actually land on the surface of Titan, and explore as a rotorcraft drone. Recent advances in artificial intelligence will no doubt be a critical component of *Dragonfly*'s capabilities, since real-time control of the vehicle is out of the question, and its ability to fly should allow it to cover and survey large distances, far more efficiently than the wheeled rovers on Mars, and even allow for inspection of the great hydrocarbon lakes on Titan. We can expect enormous scientific returns from both of these missions, sometime in the 2030s.

In any case, it should be clear that we are still just scratching the surface when it comes to exploring the gorgeous array of moons in our Solar System. Consider, as a point of comparison, the rich complexity of small scale geological features we find on the Earth; peculiar geometric formations, enormous deposits of petrified organic material buried in the mountains, exquisitely rare minerals and gems embedded deep in the caves... how much of this can we see or infer from orbit? There is no doubt we have learned a great deal about the moons in our Solar System from our robotic spacecraft inspections, but these are entire worlds waiting for us out there, and it will take centuries to explore them fully.

1.1.3 Exomoons

Serious consideration of detecting exomoons appears to be only a few decades old, and grew naturally out of the emerging possibility of detecting exoplanets. But dreams of exomoons are certainly far older. As just one example, Christiaan Huygens in his *Cosmotheoros* (1698) flatly assumed that other stars must have their own planets, and that these planets

should have all the attendant features of the planets in our Solar System:

> For then why may not every one of these Stars or Suns have as great a Retinue as our Sun, of Planets, with their Moons, to wait upon them? Nay there's a manifest reason why they should. For let us fancy our selves placed at an equal distance from the Sun and fix'd Stars; we should then perceive no difference between them. For, as for all the Planets that we now see attend the Sun, we should not have the least glimpse of them, either that their Light would be too weak to affect us, or that all the Orbs in which they move would make up one lucid point with the Sun. In this station we should have no occasion to imagine any difference between the Stars, and should make no doubt if we had but the sight, and knew the nature of one of them, to make that the Standard of all the rest. We are then plac'd near one of them, namely, our Sun, and so near as to discover six other Globes moving round him, some of them having others performing them the same Office. Why then shall not we make use of the same Judgment that we would in that case; and conclude, that our Star has no better attendance than the others? So that what we allow'd the Planets, upon the account of our enjoying it, we must likewise grant to all those Planets that surround that prodigious number of Suns. They must have their Plants and Animals, nay and their rational ones too, and those as great Admirers, and as diligent Observers of the Heavens as our selves; and must consequently enjoy whatsoever is subservient to, and requisit for such Knowlege. *(contemporaneous English translation)*

As exoplanets became commonplace in fiction, so too did exomoons; they are featured prominently in films like *Star Wars*, *Return of the Jedi* and *Avatar*, for example. In these movies, the moons are depicted as inhabited and rather Earth-like. The immense popularity of these blockbusters has no doubt shaped perceptions that moons can indeed be rather hospitable, markedly different from the barren austerity of moons here in the Solar System.

Serious efforts have been made in the field of exomoon research along three lines: 1) development of methods for detection and characterization of exomoons, 2) observational application of these methods, and 3) theoretical predictions for the architectures, compositions, and habitability of the moon systems we will eventually find. In section 1.3 we will discuss at length the various methods that have been put forward, so let us postpone tracing that history for now. Likewise, to the extent these approaches have been implemented observationally, they too will be discussed in section 1.3 – and of course in the following chapters as well, which represent a significant fraction of the observational work that has been attempted to date. But let us comment briefly on the theoretical predictions that have been made. What should we expect to find out there?

Composition

As we will discuss in section 1.2, there are a great many questions we seek to answer in relation to exomoons. Theoretical studies are a natural place to start, especially while observation remains challenging. With regard to the formation and composition of moons, the vast majority of the literature has understandably focused on the variety of mechanisms needed to explain the moons we find here in our Solar System, since these worlds already present a number of puzzles to be solved. We will discuss theses various mechanisms for moon formation further in section 1.2. But following our work on the exomoon candidate in the Kepler-1625 system (chapters 2, 3, and 4), several studies have now also examined how such a massive (ostensibly gaseous) moon might be created (Heller, 2018; Hamers & Portegies Zwart, 2018; Hansen, 2019; Moraes & Vieira Neto, 2020). Thus, the burgeoning field of exomoon observation is inspiring the examination of new processes which until now never needed to be invoked. Other studies have simply extended the known mechanisms for moon formation here in the Solar System and asked how these processes might play out elsewhere, for example, through collisional formation (Malamud et al., 2020), or the formation of a single massive moon in the circumplanetary disk of a gas giant (Fujii & Ogihara, 2020)

Dynamics

A recurring theme in theoretical exomoon papers is the question of orbital stability and longevity. The last few decades of exoplanetary observations have made clear that other planetary systems can look quite different from our own, so naturally we want to know what might happen to the moons in these contexts. A moon's stability will be a function of (for example) the planet's Hill radius (Barnes & O'Brien, 2002; Namouni, 2010; Sucerquia et al., 2020), tidal dissipation quality factor Q (Sasaki et al., 2012; Sasaki & Barnes, 2014), the proximity and relative inclination of other planets (Hong et al., 2015), its migrational history (Namouni, 2010; Spalding et al., 2016; Hong et al., 2018), the moon's semimajor axis (Domingos et al., 2006), and the moon's inclination and eccentricity (Donnison, 2014).

These works help guide our thinking in terms of the plausibility of any putative moon signal reported in the future, and to some extent, point us towards more promising approaches to the search. A few studies have also examined what happens to moons that are "liberated" by going unstable, and have proposed that we might even be able to observe spectroscopic or photometric signatures of these objects, particularly if they are in the process of disintegration (Sucerquia et al., 2019; Martinez et al., 2019).

Habitability

The final major category into which theoretical moon work can be classified relates to the question of habitability. As we will discuss in section 1.2, the search for life is a powerful motivator for investigating moons, and we have a handful of interesting astrobiological targets in our own Solar System already. It is therefore not a stretch to expect there will be at least some exomoons out there with comparable environments that are equally habitable. But could exomoons be more attractive for life than just dark ocean worlds with icy shells? Could some in fact look like the moons we've seen in science fiction? Could they have thick atmospheres, stable liquid water on the surface, the necessary chemical ingredients for life, and adequate energy flux? In other words, could we find life *as we know it* on these worlds? And might we be able to characterize them?

We are accustomed to thinking about habitable planets. But of course moons may have some physical features that planets necessarily lack, and it is reasonable to question how these differences may affect habitability. For example, most every location on our Moon experiences almost 15 Earth days of sunlight at a time, followed by 15 days of darkness. The Moon experiences dramatic swings in surface temperatures as a result, but if we are imagining a moon with an atmosphere we may expect some re-distribution of heat from the dayside to the nightside. But what does redistribution look like for such a slow rotation period coincident with tidal locking to the planet? Consider also the moons of a planet like Jupiter, which will experience both regular stellar eclipses, and one hemisphere may

also see significant illumination from its host planet; what does this look like? Then we consider the significant tidal forces felt by these moons, which drive energetic seismic and geological phenomena. And then consider the intense radiation belts around Jupiter, its magnetosphere funneling charged particles towards the moons and presenting a significant hazard for spacecraft and, eventually, whatever Earthlings come to visit. Is such ionizing radiation necessarily deleterious for the emergence of life, or could it actually be a benefit?

An early study from Scharf 2006 examined both the dynamical stability and habitability of exomoons, concluding some 15 - 27% of exoplanetary systems known at the time could be capable of hosting small icy moons, while 22 - 28% could harbor a moon in a temperate "sublimation zone," based on both stellar insolation and tidal heating. Heller & Barnes 2013 found that combinations of tidal heating and illumination from the host planet (both reflected light and thermal emission) could in fact result in a runaway greenhouse effect on some exomoons, suggesting a minimum moon semimajor axis for habitability. Heller & Zuluaga 2013 considered another important feature for (Earth-like) habitability, namely, the presence of a magnetosphere as a shield for blocking high-energy particles from the host star and preventing atmospheric ablation. Because moons are expected to be small, they may lack their own magnetospheres, and therefore will rely on their host planet's for protection, but this will extend only so far away from the planet. There may therefore be both inner and outer boundaries for moon habitability. Lammer et al. 2014 focused on exomoon atmospheric retention also, with a particular emphasis on volatile outgassing and the ionizing radiation flux of early stars, finding (perhaps not surprisingly) that low mass moons would be unlikely to retain their atmospheres while moons larger than Mars could conceivably hold onto them. Lingam & Loeb 2019 went on to consider the potential for (possibly exotic) photosynthesis on moons driven by reflected light of their host giant planet, concluding that it could be possible for systems other than planets orbiting M-dwarfs, as illumination from these stars would require the moon to be too close for habitability. Night side illumination of tidally-locked exoplanets by exomoons could also be sufficient to drive photosynthesis in some cases.

These studies already raise a number of intriguing questions and possibilities to be examined. And there are still more. Let us turn now to examining the motivations for carrying out a systematic search for exomoons.

1.2 Motivations for Finding Exomoons

In science we are well advised to keep sight of the reasons for pursuing our objectives. This is not to impose some strict sense of scientific utility, or embrace a philosophy that dictates that every scientific endeavor must pay immediate, concrete dividends. History is replete with examples of so-called "pure" scientific discoveries that may have had little contemporaneous practical application but that nevertheless ultimately revolutionized our world. Einstein may not have ever dreamt of GPS, for example, but his work on relativity is nevertheless a critical component of making that technology possible.

As we extend our reach ever deeper into the cosmic unknown, accelerated by the great technological leaps of the last 150 years or so, the gap between "pure" scientific discovery and its "practical" applications may seem ever wider. Where is the (what might be called in other professions) "actionable intelligence" in understanding the origins of the Universe, the formation of the galaxies, the collisions of black holes, the birth of stars, the evolution of planetary systems? Certainly we can find some potentially practical physical insights if we look hard enough, perhaps in, say, quantum computing or fusion reactors. That is, to the extent our understanding of the relevant physics in these cases is incomplete, astrophysics may yield some important results. The whole Universe is our laboratory. Or, we can imagine the practical payoffs that are several centuries into our future, technologies that are as far removed from us as the Concorde is from Leonardo da Vinci's flying machines. What might we build if we could harness dark matter, antimatter, or dark energy? There is no shortage of dreamers who have authored either serious scientific investigations or imaginative science fiction on these subjects. But if history is any guide, we may be amusingly far off track in picturing the true realities of our deep future.

Let us set aside, then, our need to imagine all the practical applications of our work, while not entirely casting off the great value of pure scientific research to the practical. We can content ourselves with the knowledge that payoffs are inevitable, but because we cannot perfectly imagine the future, we simply cannot and should not impose our present day metrics of what constitutes practical science onto our research efforts. Suffice it to say, we should want to understanding everything we possibly can about the Universe in which we find ourselves, and rest assured that there is some inherent value to these discoveries.

Still, it is important to have goals in mind, particularly in the context of modern science, because we are all working individually at revealing a small piece of the big picture. How then, does our work connect to that broader picture? In what ways might it help reveal other parts of the picture? With apologies for appealing to a rather cliché metaphor, each result is a piece of the puzzle; one piece by itself is orphaned and possibly meaningless, it must be anchored to the broader picture to give it some context, and to contextualize other pieces.

Before diving in to the specific motivations for exomoon research, there is one last thing to say about our motivations in astronomy more broadly. In our line of work we are truly plumbing the depths of time and space to learn who we are, where we came from, and what our fate may be. It is difficult to imagine a more profound endeavor. This may seem about as far afield from those practical considerations we just discussed, but it is worth remembering that investigating or providing answers to these very same enormous questions has been central to religions and cultures around the world for millennia. The fact that these philosophical questions have arisen repeatedly and independently around the world might be considered a form of convergent evolution – just as birds, insects, and mammals all independently developed wings, or fish and aquatic mammals independently developed fins. That is, the repeated emergence of these questions and efforts to answer them in isolation speaks to their central importance in the human psyche. They should not be trivialized, and we have some reason to rejoice in the fact that we live in an age where humans are finally

able to answer some of these questions, not with mysticism, but through science. Astronomy does not purport to provide spiritual guidance, but it can nevertheless reveal the awesome scope of nature and our vanishingly small part in it.

Let us now examine the specific motivations for finding and characterizing exomoons.

1.2.1 Contextualizing our Solar System

The history of astronomy has been that of gradually expanding our horizons, and for most of that history, the Solar System was essentially the extent of our arena. Its true architecture was not fully grasped until the European Renaissance, through the work of Nicolaus Copernicus, Galileo Galilei, and Johannes Kepler. It was about this same time that Giordano Bruno was speculating that the stars were other suns, accompanied themselves by their own retinue of planets. Since these auspicious discoveries and ruminations there has been a series of "demotions", as Carl Sagan put it in his *Pale Blue Dot*, a cascade of discoveries that have placed humanity ever farther from our perceived centrality: ours is not a particularly extraordinary star, nor do we live in a conspicuously special region of the galaxy, nor is our galaxy anything especially remarkable amongst the hundreds of billions of other galaxies in the observable Universe. Still, even as our discoveries about the wider Universe have proceeded at an incredible rate these last few centuries, we remained, as recently as the 1980s, in complete ignorance about the nature and abundance of other planetary systems. Fortunately, the veil of mystery is finally being lifted.

Before the exoplanet revolution, we were of course interested in learning how the Solar System came to be, and we are still today working to answer any number of open questions about our history. Prior to the first exoplanet discoveries it may have been quite tempting to apply the mediocrity principle to our Solar System, concluding that what we would eventually find elsewhere should be, broadly speaking, more or less like what we see here at home. With the benefit of hindsight we can see now that such thinking would have been woefully simplistic, and could lead us astray. Indeed it did at times, and to the extent

that such assumptions were built into theoretical models of planetary system formation, real observations of exoplanetary systems were crucial for painting a fuller picture. Now we know that there are systems that look astonishingly different from our own, with massive planets orbiting extremely close to their host stars (e.g. Mayor & Queloz, 1995), planets on highly eccentric orbits (e.g. Jones et al., 2006), spin-orbit misalignments (e.g. Fabrycky & Winn, 2009; Huber et al., 2013; Winn et al., 2005), compact multi-planet systems locked in Laplace resonances (e.g. Luger et al., 2017), and a whole class of planets – the super-Earths or mini-Neptunes – that are completely absent from our own Solar System. We have uncovered new relationships and mechanisms at work through a study of the exoplanet population, for example, the so-called "peas in a pod" ordering of planets of similar size (Weiss et al., 2018), the inflation and obliquity of planets just beyond orbital resonance (Millholland, 2019; Millholland & Laughlin, 2019), and the radius gap (Fulton et al., 2017; Fulton & Petigura, 2018).

It should be clear, then, that observing exoplanets and by extension exomoons holds the key to helping us contextualize our Solar System. In what ways are we ordinary, and in what ways are we peculiar? At present we lack examples of systems that look very much like our own, with a terrestrials inside the snow line and several giants at large semimajor axes. Statistical analyses of the population of planetary systems are suggesting that our architecture may actually be somewhat uncommon (Gould et al., 2010). On the other hand, it is worth keeping in mind that in general, few of our current planet detection techniques are particularly well suited for finding planets like Saturn, Uranus and Neptune. They are unlikely to show up through direct imaging (too cold and dark), radial velocities or astrometry (very long time baselines are needed), or through transits (low transit probabilities). Microlensing may hold some promise for revealing such systems (e.g. Gould et al., 2010), but there are significant challenges for this approach, as well (see section 1.3).

While we now have many points of comparison between exoplanets and our own Solar System, we completely lack an equivalent point of comparison when it comes to moons.

Our Solar System is home to hundreds of moons, most rather small, a handful large enough that we could potentially detect their analogs in other systems with our current technology. We see most of these large moons in the outer Solar System. The question is, is this the way it must be? Will moons be rare in the regions close to their host stars, and if so, why is this the case? By the same token, should we always expect an abundance of moons out beyond the snow line, or do we just have a peculiar build-up of these moons in our Solar System, commensurate with our potentially peculiar build-up of gas giants at these distances? In order to answer these questions, we will need to make real efforts towards systematic detection of exomoons. Our degree of success or failure along these lines will ultimately help shape our understanding of these planetary systems in relation to our own.

1.2.2 Planet and Moon Formation

There are a variety of mechanisms through which we expect to form moons. The first is formation through accretion of material in a circumplanetary disk, something very much akin to the formation of planets in a circumstellar disk (e.g. Lunine & Stevenson, 1982). It is generally thought that the regular moons of Jupiter, Saturn, Uranus, and Neptune formed in this way (e.g. Canup & Ward, 2006; Ogihara & Ida, 2012; Heller et al., 2015; Cilibrasi et al., 2018; Szulágyi et al., 2018; Batygin & Morbidelli, 2020).

Then there is the capture scenario, in which an object that was once gravitationally unbound is through some mechanism pulled into orbit around the primary. This could occur, for example, through an exchange of energy and angular momentum in a 3-body encounter (Agnor & Hamilton, 2006; Nesvorný et al., 2007), tidal dissipation (Hamers & Portegies Zwart, 2018), pull-down capture (Hansen, 2019), or through gas drag in a protoplanetary disk (Pollack et al., 1979). The best example of a large captured moon is probably Neptune's large moon Triton, which appears to be a captured Kuiper Belt object based on its composition and orbital properties (Agnor & Hamilton, 2006). It has been suggested that the capture event may have disrupted the original regular moons of Neptune, and that the debris from

the assoociated collisions ultimately provided the material for the regular moons we see today (Banfield & Murray, 1992).

Then we have the giant impact scenario, which is thought to be the explanation for Earth's large moon (e.g. Hartmann & Davis, 1975; Lucey et al., 1995; Canup & Asphaug, 2001; Canup, 2004, 2012) as well as Pluto's giant moon Charon (McKinnon, 1989; Canup, 2005). In these cases, two large protoplanets collided, spinning off an enormous amount of material that would ultimately become the large moon. Smaller impact events may have resulted in the creation of Mars' two small moons (Craddock, 2011; Citron et al., 2015) as well as Pluto's four smaller moons (Stern et al., 2006; Canup, 2011). In all these cases, the moons are thought to have formed out of a disk of material created by the associated impacts.

Finally, we have systems that blur the line between primary and satellite. These are effectively binaries, having formed near one another and due to their low relative velocities simply remain a pair. In some cases they may even become contact binaries. The Kuiper Belt object MU69 is thought to be one such object (Stern et al., 2019), and indeed, there may be many of these systems particularly in the depths of the outer Solar System (Nesvorný & Vokrouhlický, 2019). These objects blur the line between the circumplanetary disk pathway and the capture pathway.

How often do these various architectures manifest elsewhere? And to what extent might they impact the habitability of these systems? Are these outcomes common? What might have happened if, say Venus or Mars, had experienced a similar impact? It has been suggested, for example, that the Moon has played an important role with respect to life on Earth by maintaining a stable axial tilt (Laskar et al., 1993), though the importance of having a moon for habitability is disputed (Lissauer et al., 2012).

Let's turn our attention to the regular moons, a few of which here in the Solar System we have established are potentially attractive places to look for life. What might these worlds look like in orbit around giant planets closer to their host stars? Will they even be found

in these regions in any significant number? Might moon capture be a bigger player in other planetary systems? And what if we find systems that defy the tidy formation explanations offered here? What insights might be gleaned with respect to the common features of our Solar System?

In the preceding two paragraphs alone there are a number of fascinating questions, the answers to which could open even more lines of inquiry. There are two recurring themes emerging here: 1) theory has great explanatory power, but needs to be considered in tandem with observation, and 2) there is any number of exciting opportunities to further our understanding, but that a statistical accounting of these worlds is needed, and this can only be accomplished through broader, sustained effort.

1.2.3 Planetary Migration

Solar System and exoplanetary scientists alike continue to puzzle over the question of planetary migration. Once again we come at this problem from a theoretical standpoint to specifically explain the features we see in the Solar System, and through observations of other planetary systems. The "Grand Tack" hypothesis provides explanations for a variety of features of our Solar System, including Mars' small size, the distribution and composition of asteroids and comets, and volatile delivery to the terrestrial planets via impactors (Walsh et al., 2012; O'Brien et al., 2014; Dones et al., 2015). At the same time, some lingering questions remain (c.f. Raymond & Morbidelli, 2014).

Now our understanding is also aided by the observation of exoplanetary systems, which display features we wish to explain from a theoretical standpoint, and high resolution observations of protoplanetary disks with the Atacama Large Millimeter Array (ALMA). These observations have provided new puzzles, for example, the existence of Hot Jupiters, and what appear to be gaps in the disks, which are suggestive of planetary formation but that occur in some cases at hundreds of astronomical units (AU) from the star; in the case of CL Tau there is evidence for a Hot Jupiter *and* disk gaps at large semimajor axes (Clarke

et al., 2018). Meanwhile, the formation mechanism of Hot Jupiters remains fiercely debated even decades after the first discovery (Mayor & Queloz, 1995), falling into two main camps: in-situ formation and disk migration. In any case, these observations, while revealing, all have one thing in common: they are but one snapshot in time, allowing us to see only a few frames of a movie that plays out over the course of millions of years. And in fact, what we're seeing in each case is more like capturing a single frame of *different* movie, something like, say, different adaptations of the same Shakespeare play. We can piece together the story, but with lots of caveats and incomplete knowledge of the unique story elements.

Where do moons fit into this picture? Regular moon formation will occur primarily during the early years of the planetary system, as gas from the circumstellar disk feeds the circumplanetary nebula and "subdisk". These disks, in fact, are not only the birthplace of satellites, but they are also responsible for regulating the late stage accretion of the planets (Szulágyi, 2017). As the leftover material of giant planet formation, after earlier generations of satellites have fallen into the planet (Canup & Ward, 2002), the moons we see today therefore retain some traces of that early history. Likewise, moons created through giant impacts are also relics of an early, far more violent epoch in the history of the planetary system. In our studies of our Solar System, we have drawn conclusions about the moons' formation histories through both compositional and dynamical analysis.

Of course, orbits evolve, dynamics are complicated. But we know, for example, that moons can be lost during planetary migration as the Hill sphere shrinks (Namouni, 2010). Evection resonance can also destroy moons during inward migration (Spalding et al., 2016). Other mechanisms, not yet imagined or investigated, could conceivably play a role in shaping the moon population, particularly at this critical time of planetary system formation. By developing a statistical picture of the moon population – their distribution, abundance, sizes, and ultimately their compositions – we will have a new window in understanding the dynamical histories of the planetary systems in which we find them, how they formed and how they evolved in time.

1.2.4 Future Characterization

In observational astronomy we have to be thinking about our objectives far into the future. The flagship instruments, on which groundbreaking discoveries are often made, can often take decades to develop, from initial concept to construction to first light. The idea for the *Kepler* mission, for example, was first advanced in Borucki & Summers 1984; the spacecraft launched some 25 years later, and here more than a decade after that, we are still making new discoveries with the data. The James Webb Space Telescope (JWST) has famously seen its share of major delays; with the concept first advanced as far back as 1994, it had an original launch date between 2007 and 2011; we have still yet to see its launch. Meanwhile, the Thirty Meter Telescope (TMT) concept was studied as far back as 2002 (CELT collaboration, 2002), and though the collaboration broke ground in 2014, the telescope has still yet to be built due to ongoing political turmoil in Hawai'i stemming from opposition to new construction on Mauna Kea. Space-based telescopes, fortunately, tend to face merely financial and technical, but not territorial, obstacles.

As of this writing NASA is now exploring four new mission concepts for the next flagship space telescope: The Large Ultraviolet Optical Infrared Surveyor (LUVOIR, The LUVOIR Team, 2019), The Habitable Exoplanet Imaging Mission (HabEx, Gaudi et al., 2020), the Origins Space Telescope (OST, Meixner et al., 2019), and the Lynx X-ray Observatory (The Lynx Team, 2018). Whichever mission is selected, the launch date is expected to be at least 15 years away. If history is any guide, it may be considerably later.

With these realities in mind, it is far from outrageous to be thinking seriously now about the kinds of observations we wish to make in the next several decades. The present challenges for finding exomoons, though significant as we will see in section 1.4, should not deter us from keeping exomoons on the wishlist, not just for discovery, but eventually for characterization as well. If the moons of our Solar System are any guide, we can expect exomoons will be geologically active and diverse, with an abundance of water – in some cases even liquid water, and (possibly thick) atmospheres. There is good reason to expect that, just as exoplanetary

systems have broadened our understanding and expectations for what we will find out there, so too will exomoons. A rich diversity of terrestrial worlds, we have good reason to expect, are waiting out there to be discovered, and known.

As was the case in our Solar System, and has already been argued with regard to exomoon observations, dynamical inferences can tell us a great deal about the system and its history. But close-up examination of the Solar System moons has revolutionized our understanding of these worlds. For example, one key piece of evidence for a giant impact explanation of our Moon is the geological similarity of the Moon and the Earth's crust. The potential habitability of moons in the vicinity of Jupiter and Saturn might have seemed exceptionally remote had we not also detected evidence of liquid water under the surface through close-up inspection. Of course, when it comes to exomoons, we will not be able to send robotic probes to photograph these surfaces. But to the extent that we are optimistic about characterizing terrestrial exoplanets, we can and should at least set the goal of achieving the same with respect to exomoons. We will discuss observational approaches in greater detail in the next section.

1.2.5 Life

Finally we come to what is perhaps one of the most powerful motivators for this work: the search for life. In our Solar System we have several terrestrial worlds beyond the Earth that are thought to be attractive places to carry out a serious investigation: most notably, Mars, Europa, and Enceladus. Mars of course is much closer to the Sun than these icy moons, and though today it is itself a frozen desert world, it is believed to have once supported liquid water on the surface (e.g. Bibring et al., 2006). It is at least conceivable that with extensive, human exploration of Mars, we could one day uncover fossilized evidence of life that once lived there. It is even possible that as the years of prime Martian habitability waned, organisms evolved to cope with the increasingly hostile conditions and migrated underground. Supporting this possibility is the fact that here on the Earth we have extremophilic organisms

that have occupied virtually every niche on the Earth, from the frigid, crushing depths of the deep ocean to scalding hot springs on the surface. With this in mind, it is just possible that we could find life living today on Mars.

Meanwhile, there is now convincing evidence of subsurface oceans of water on both Europa and Enceladus (Khurana et al., 1998; Postberg et al., 2009). Because we have not yet established how abiogenesis occurred here on the Earth – and it would be a mistake to assume there is only one possible pathway to abiogenesis anyway – we must also admit the possibility of life having been able to gain a foothold in other, rather different and ostensibly more hostile environments, and thriving right now in these subsurface oceans. What Europa and Enceladus have always lacked, of course, is the incredible power source that is Solar radiation. Photosynthesis *as we know it* is clearly out of the question. Even so, here on the Earth we know of entire ecosystems that exist entirely independent of the Sun as an energy source; at hydrothermal vent oases, chemolithoautotrophic bacteria form the base of the food chain, which might also include worms, corals, mollusks, and fish. It is also clear that these ocean moons have a source of internal energy, through the gravitational tidal heating, and indeed, we see evidence of volcanism or cryovolcanism on Io (Smith et al., 1979), Enceladus (Hansen et al., 2006; Waite et al., 2006), and Europa (e.g. Roth et al., 2014; Sparks et al., 2016). hydrothermal vents are therefore not out of the question on these icy moons.

The Solar System has one large moon in the so-called habitable zone: our Moon. Apart from some water ice at the poles (Colaprete et al., 2010), it appears to be bone dry, and rather inhospitable to life as we know it. Even so, the great diversity of terrestrial worlds in our Solar System, and the variety of ecosystems that exist right here on Earth, suggest that a habitable exomoon is entirely plausible. Indeed, given the fact that we currently only know of one world on which life has developed, we should be cautious about a tendency towards planet chauvinism. It could turn out that moons, not planets, are more clement environments for abiogenesis. We cannot know for sure until we actually probe these worlds, understand them, and characterize them.

1.3 Detection

We now turn to examining the variety of techniques that have been utilized for exoplanet discovery and characterization, and comment in turn on their applicability to the exomoon search. We will categorize these various methodologies by the primary observational tools and techniques that will be employed for each one, keeping in mind that in some cases these methods will blur the line between observational techniques.

1.3.1 Spectroscopy

Spectroscopic observations have long been the bread and butter of observational astronomy. By analyzing the spectrum of light we collect from a distant object, we can determine not only its composition, but also its temperature, its speed, and (in a cosmological context) its distance, as well. Carrying out multiple spectroscopic observations over time further allows us to study dynamic events. Let us examine now how spectroscopic observations might reveal the presence of exomoons.

Radial velocity

As one of the earliest techniques used for planet discovery (e.g. Mayor & Queloz, 1995), radial velocity (RV) measurements have been responsible for approximately 20% of confirmed planet discoveries to date. Most powerfully, RVs are capable of revealing planets in a wide variety of orbits; that is, they are not restricted to edge-on orbits, and they are sensitive to planets on extremely wide orbits. At the time of this writing, the longest period planet for which the period uncertainty has been constrained to less than 100 days is HD 181234, with $P = 7462^{+80}_{-76}$ days ($a = 7.52 \pm 0.16$, $e = 0.73 \pm 0.01$). RVs have thus yielded a sizeable number of planets that are, at least in some respects, true analogs of our outer Solar System giants. RVs necessarily also allow us to constrain the masses of these worlds. The promise of RVs to reveal ever smaller planets, even Earth-mass worlds, has become clear, and today we

are achieving RV precision on the order of 10 cm s^{-1} through instruments like the EXtreme PREcision Spectrograph (EXPRES, Jurgenson et al., 2016). While a true Earth-analog orbiting a Sun-like star is currently out of reach, it is nevertheless possible to find Earth-mass planets around low mass stars in tight orbits (e.g. Proxima Centauri b Anglada-Escudé et al., 2016). Improvements in spectrographic technology in the years to come will no doubt continue to lower the mass floor, eventually allowing us to home in on ever more Earth-like systems.

Given the power of these techniques, the question naturally arises, can we use these tools to find exomoons? The short answer is, unfortunately no (see however 1.3.2). It is the barycenter of the planet-moon system that follows the Keplerian orbit around the star, and it is the combined system mass that we are measuring when we fit an orbit to the RV data. One might wonder whether the changing geometry of the planet-moon system might still induce some additional effect on the RV signal, and in fact it does, though the effect is extremely small. For a Jupiter mass planet with an Earth-mass satellite, Morais & Correia 2008 calculated a residual RV amplitude on the order of 10^{-7} m s^{-1}, far below our current detection limits. Nevertheless, as discussed at length in later chapters, RVs do have a key role to play in the exomoon search. We may obtain planet mass predictions from a photodynamical model fit, and RVs are capable of either supporting such predictions or placing tension on them. Moreover, RVs will be critical in testing the hypothesis of a perturbing planet in the system, a hypothesis which will be often be the primary competing explanation for the presence of transit timing variations (TTVs). We will discuss these TTVs further below.

Rossiter-McLaughlin Effect

At any given moment when we observe a distant star, if we are not looking directly along its spin axis, one hemisphere of that star is rotating away from us while the other hemisphere is rotating towards us. The photons we receive from these two hemispheres are

thus redshifted and blueshifted, respectively. This effect will be slight, and of course for an unresolved source we see the light from both hemispheres blended together. The overall effect of this blending is spectral line broadening, and provides an effective method for measuring the rotational speed of a star.

This effect can be used to our great advantage when a planet transits, because the planet will block out light from either blue- or redshifted regions as it passes in front of the face of the star. Monitoring the changing shape and position of the spectral lines during the transit provides a clue as to the direction of the planet's orbit (either prograde or retrograde), and even reveals its inclination with respect to the star's equator. This is easy enough to visualize: for a prograde planet orbiting co-planar to the star's equator, it will pass first in front of the blueshifted hemisphere, and then the redshifted hemisphere. We will therefore observe a shift towards the red, followed by a shift towards the blue. At the other extreme, in the limit of a planet inclined such that its orbital plane is aligned with the star's spin axis, it will pass entirely over the blueshifted hemisphere, or the redshifted hemisphere, and we will observe a single, constant shift in the spectral lines that last the duration of the planet's transit. It is through this technique that we have identified several systems with significant spin-orbit misalignments (e.g. Winn et al., 2005).

Clearly, for a transiting planet-moon system, we will now have two objects occulting the red- and blueshifted hemispheres of the star, different ingress and egress times, and potentially different impact parameters, as well. The Rossiter-McLaughlin effect may therefore be used, perhaps in combination with transit photometry, to constrain the geometry of the system (Simon et al., 2010; Zhuang et al., 2012) to an extent not fully possible with transit photometry alone. It may therefore be a powerful characterization tool in the future, though because it is rather expensive in terms of labor and telescope resources, it is probably unlikely to be a first detection tool.

Plasma tori and outgassing

By analogy with the Solar System, it has been suggested that we may see evidence of exomoons, not directly, but through absorption features in the region of space around the transiting planet that can be attributable to volcanism on a close-in moon. Observations in the early 1970s (e.g. Kupo et al., 1976) suggested the presence of a plasma torus around Jupiter, which was attributed to ionization of sulfur ejecta from Io (Brown, 1976). It is argued, then, that spectral features during a planet transit might therefore be an observable signature of exomoons (Ben-Jaffel & Ballester, 2014; Oza et al., 2019). These would manifest, for example, as spectral transmission or absorption features that are far too extended to be attributable to the planet's atmosphere.

It is expected that these signatures will be strongest around planets with high irradiation, that is, Hot Jupiters. Of course, for very close-in planets, their Hill spheres will be quite small, restricting the region of stability for a moon; but this is not necessarily a problem, since the required volcanism is expected from moons that are orbiting very close to their host planet anyway. In the case of Io, the tidal heating is due not just to its proximity to Jupiter, but also the presence of the outer Galilean satellites, which work to preserve its non-zero eccentricity. Outer satellites for a Hot Jupiter may be out of the question given the restricted region of stability, but in these cases the host star's proximity could conceivably play the equivalent role of maintaining the moon's eccentricity.

It has been suggested that the planet WASP-49b may be displaying such a signature (Oza et al., 2019). The key challenge will be whether such a signal is an unambiguous indicator for the presence of an exomoon. As always, we will want to pursue as many avenues for independent confirmation as possible.

1.3.2 Direct Imaging

Among the more dazzling achievements in exoplanetary science has been the advent of direct imaging. For relatively nearby systems, with (typically) young, large planets orbiting

at large semimajor axes, it is possible to directly image these planets. Achieving this has required the employment of large aperture telescopes such as the Very Large Telescope (VLT) (SPHERE, Beuzit et al., 2019) and Gemini South (GPI, Macintosh et al., 2006) in Chile, along with the development of fast and exquisitely precise adaptive optics, high contrast imaging, and integral field spectroscopy. In the future, proposed next-generation space telescopes such as HabEx (Gaudi et al., 2020) and LUVOIR (The LUVOIR Team, 2019) are expected to improve considerably our ability to directly image smaller worlds with shorter orbital periods, decreasing the inner working angle through the use of a coronagraph or possibly even a free-flying starshade. These missions hold great promise for identifying Solar System analogs in the coming decades.

The question is, can we resolve a planet-moon system through direct imaging? This is currently out of reach, and will likely remain so for the foreseeable future. There have been recent suggestions of a possible exomoon detection with SPHERE, through residual analysis of a directly imaged planet (Lazzoni, 2019), but this is unpublished work and as yet unconfirmed. In any case, there are some other exciting possible approaches in the literature that may be employed in the near term.

Planet reflex motion

One very promising idea comes from Vanderburg et al. 2018, who found that it should be possible with near-future instrumentation to measure the radial velocity of the *planet* (not the star), and identify a moon in orbit around that planet through a short period oscillation superposed on the larger amplitude velocity signal of the planet. To achieve this, we will need sufficient sensitivity to measure chemical emission or absorption lines in the planet's atmosphere accurately enough to detect a Doppler shift. A sizeable planet-to-moon mass ratio is also a prerequisite; Galilean analogs are likely out of reach in the near term.

Center of light

Another exciting possibility for detecting exomoons in the near future through direct imaging is spectroastrometric detection (Agol et al., 2015). In this approach, we leverage the fact that a moon may have a sufficiently different spectral reflection profile from its host planet that the brightness of the two objects will vary differently across the detector's wavelength range. As a result, when observing an unresolved planet-moon system, the light centroid will shift, as the planet is brighter in some wavelengths and the moon is brighter in others. If we can measure this centroid shift with sufficient accuracy, we could infer the presence of a moon. A planet, of course, would show no such spatial shift if it is the only light source. It is at least conceivable that this method could be applied with LUVOIR or HabEx, or even near-future thirty-meter class telescopes.

In any case, we must keep in mind that direct imaging may still be limited to a relatively small fraction of nearby systems. Even so, these few systems stand to provide a disproportionate science yield.

1.3.3 Radio

We tend to think of radio observations as being useful primarily for studying gas in the local Universe, and for probing extragalactic sources which may display a range of phenomena at various energies redshifted into the radio. There are however at least two approaches to detecting exomoons with radio that have been suggested to date.

Pulsar timing

One particularly exotic approach to exomoon detection is through precise pulsar timing (Lewis et al., 2008; Pasqua & Assaf, 2014). The first pulsar exoplanets were discovered by Wolszczan & Frail 1992, and indeed, these were the first exoplanet discoveries ever, though they are sometimes overlooked. Despite the power of this technique, there are several roadblocks to undertaking a large-scale search for exomoons in this fashion. Nearly thirty

years after the first discovery, there are today still only seven confirmed exoplanets discovered by pulsar timing. Thus, the hunting ground for exomoons is quite limited.

But why are there so few? To date there are only a few thousand pulsars known (Manchester et al., 2005), already limiting significantly the number of targets that might be examined. More importantly, it has been suggested that planets around pulsars are intrinsically rare (Miller & Hamilton, 2001). When it comes to finding moons around these planets, Lewis et al. 2008 finds that any detectable moon would need to have a reasonably large semimajor axis, and a mass ratio of at least 5%. In our Solar System, only Charon and Pluto display such a large ratio, but of course, we do not yet know how often we might see such objects in extrasolar systems.

Habitability of these pulsar planets and any moons is highly questionable, but the discovery of an exomoon orbiting a pulsar planet could potentially yield insights into how the planet formed in the first place, whether it is a survivor of the supernova event or formed as a second generation planet out of the debris of the explosion. This remains very much an open question.

Circumplanetary radio emission

Planets in our Solar System are also sources of radio emission resulting from charged particle flow within the planets' magnetospheres. It has been suggested, then, that exoplanets might be detected through the radio, as well. More than this, just as we see the magnetic interaction with Jupiter's moons, and in particular the plasma torus mentioned previously, exomoons may also be detectable through a characteristic radio signal (Bastian et al., 2000; Noyola et al., 2016). This emission is conceivable detectable, at least for nearby sources, but it remains unclear whether observations of this sort could yield an unambiguous exomoon detection.

1.3.4 Time-Domain Photometry

We turn at last to the approach on which the work in this dissertation is largely based. Time-domain photometry, particularly through missions like *Kepler* and TESS, has been astonishingly successful in detecting the majority of known exoplanets, through relatively inexpensive methods. Large scale monitoring efforts in the time-domain means that we can (broadly speaking) avoid the need for targeted observations and in the process arguably democratizes the exoplanet search. This is because, unlike most of the other approaches highlighted above, exoplanet transit searches can be carried out without the need for individual astronomers or teams to propose for telescope time, or to be associated with a particular observing campaign, telescope, or instrument. Instead, virtually anyone, including amateurs, may download data from a massive database and carry out a search project, potentially with no funding for the work at all. We are, of course, indebted to the project scientists who build the instruments, launch the spacecraft, and design the front-end analysis pipelines, but after a (typically short) proprietary period, there is no monopoly on the data's usage, and there is far more data to analyze than can possibly be handled by one research group. Time-domain surveys have also yielded any number of other major astrophysical insights beyond exoplanet discoveries, through (for example) stellar aging (e.g. McQuillan et al., 2014; Angus et al., 2015; Douglas et al., 2016), stellar activity (e.g. Davenport, 2016; Morris et al., 2018), and asteroseismology (e.g. Gilliland et al., 2010; Angus et al., 2015).

Time-domain photometry, I argue, is also the most promising avenue for detecting exomoons in the near future. This is because, when a moon is orbiting a transiting planet, there will be at least three signals imprinted in the photometry: planetary transit timing variations (TTVs), transit duration variations (TDVs), and transits of the moon itself. These signals can be independently measured, but of course they must all be consistent with one another, and this provides useful opportunities for cross-validation and model selection.

A few other approaches to leveraging time-domain photometry have been proposed, so let us examine these first before diving into the methods we have primarily utilized in the

subsequent chapters.

Astrometry

The advent of extremely precise astrometric measurements through the *Gaia* spacecraft (Perryman et al., 2001; Gaia Collaboration et al., 2016, 2018) has raised the intriguing possibility that planets, particularly long period planets, may be identified in significant numbers through precise measurement of stellar reflex motion. For sufficiently nearby stars hosting a massive planet, it should be possible to identify exoplanets and obtain an orbital solution through precision measurements of the host stars' sky position, distance, and proper motion. For stars brighter than $G = 15$, the typical positional uncertainty is 0.02 - 0.04 milliarcseconds (mas), and proper motion uncertainty is 0.07 mas yr^{-1} (Gaia Collaboration et al., 2018).

For a point of comparison, a Solar-type star ($M_V = 4.83$) at 10 parsecs away with a Jupiter-analog companion would display an edge-on reflex motion of amplitude $\approx$ 2 Solar radii. As small as this may sound, this translates to a positional change of 0.93 mas, or a proper motion of 0.157 mas yr^{-1} – comfortably detectable considering the typical uncertainties. With a sufficiently long observational baseline, this motion can conceivably be disentangled from the star's motion through the galaxy. A Jovian-equivalent at a smaller semimajor axis would have a correspondingly smaller amplitude due to its shorter lever arm, but would also have a correspondingly shorter orbital period. And of course, most stars will not be seen perfectly edge-on, so the star will trace out a small ellipse, further aiding our distinguishing the star's reflex motion from its galactic orbit. Through this simple example we can begin to appreciate the sheer number of worlds, at a variety of semimajor axes and orientations, that may be found through astrometry in the years to come.

When it comes to finding moons, we are faced with the same challenges that are present with stellar RVs: the reflex motion we measure will be due to the planet-moon barycenter, and so there will be no evidence of a moon in the astrometric measurements. Nevertheless,

because astrometry stands to expand enormously the number of known exoplanets, and in at least some subset of cases their orbits will be well-constrained, some fraction of these planets may be new candidates for direct imaging, and these worlds may be ripe for the moon search through the RV technique put forward in (Vanderburg et al., 2018). Through astrometry we may also have a new, lower cost method for corroborating a planet mass prediction made by photodynamical moon modeling.

Microlensing

Microlensing holds a great deal of promise for detecting low mass objects like moons (Han, 2008). The anticipated Nancy Grace Roman Space Telescope (formerly the Wide-Field Infrared Survey Telescope, WFIRST), is expected to discover ~ 2600 bound exoplanets through microlensing, with a sensitivity down as low as 0.1 Earth mass, and will be particularly sensitive to planets at larger semimajor axes (Spergel et al., 2015). This is exciting in part because low mass objects with long orbital periods are mostly out of reach for detection through the other methods described here, so this will fill in a critical gap in our statistical sample of exoplanets. There have even been some tentative exomoon detections through microlensing already Bennett et al. (2014); Hwang et al. (2018), though both of these results have lingering uncertainties that point to ongoing challenges with the microlensing approach. Nevertheless, it is reasonable to expect the ambiguities associated with these challenges can be overcome in future work, and we may see an uptick in the number of researchers employing microlensing once the NGRST comes online.

The challenge with utilizing microlensing for the moon search is that observational follow-up opportunities are not guaranteed, and will be quite difficult to obtain and perform. Microlensing events are recorded as the low-mass objects pass between the observer and a background light source, which is lensed by the foreground objects. The system may be targeted for follow-up observations with high angular resolution telescopes (large aperture ground-based telescopes with adaptive optics, or space based observatories like HST), but

only once the lensing (target) star is sufficiently separated from the background light source on the sky, and then the lensing star would also need to be bright enough for adequate observations. Even then, follow-up observation is not straightforward; there is only a very slim chance that the planet transits, and ephemerides for the planet will likely be poorly established in any case. Radial velocity measurements could be feasible, but substantial observational resources over several years would have to be invested if the planet has a long period, and as we've already seen, RVs by themselves will not reveal a moon in any case. The best remaining hope is direct imaging, but this too carries the significant limitations discussed above. Meanwhile, rogue planets and their potential moons, while also expected to be detected in small numbers by NGRST, will not be visible at all after the microlensing event, and therefore cannot be observationally confirmed after the fact. All this suggests that that microlensing will be of limited value in identifying and characterizing individual exomoons for future study.

What microlensing has to offer, then, is a broader statistical picture. As discussed later in this chapter and in Chapter 2, there is a great deal we need to understand about the population of exomoons, and we find ourselves in something of a Catch-22: to better find individual systems, we need to understand what we're seeing across the population, but our understanding of the population is limited in part by our difficulties identifying individual systems. Microlensing then suggests a path forward, where we can begin to build a statistical picture of the population of exomoons out there that will inform our search through other approaches.

Transit timing

A planet with a moon will orbit the two bodies' common center of mass, and as previously noted, it is this center of mass that travels on the Keplerian orbit. Under the reasonable assumption that the moon's orbital period is not a perfect integer multiple of the planet's period, for any given epoch we will observe the moon at a different phase in its orbit. From the

perspective of an observer viewing this event edge-on, the planet will have some sky-projected displacement from the barycenter; it may lead the barycenter, or trail it. In the case of a single planet system, where planet-planet interactions are absent, the barycenter's transit time will be effectively fixed, following strict linear ephemeris, but because the planet appears to lead or trail the barycenter, the timings of the planet's transit will be correspondingly early or late. These are the moon-induced transit timing variations, first predicted to arise in Szabó et al. 2006.

Subsequent work identified another effect, the TDV (Kipping, 2009a,b), which involves an oscillating planet transit duration over the course of several epochs. These duration variations can arise in two ways. First, because the planet is orbiting the planet-moon barycenter, it now has a non-zero tangential velocity with respect to the Keplerian orbit. From the perspective of the observer, the planet may appear to be moving faster or slower than the barycenter, depending on the phase of the moon's orbit. This results in transit durations that are shorter or longer, respectively, and are referred to as TDV-Vs (V for 'velocity'- induced). A second effect arises for any system where the moon's orbit is not seen perfectly edge on by the observer. As such, the barycenter will maintain a fixed impact parameter for every transit, but the planet will now necessarily pass above or below the barycenter from our perspective, thereby changing the impact parameter of the planet for that epoch. As a result, the transit chord of the planet becomes shorter or longer, and this too produces variations in the duration of the transit. This effect is typically referred to as the TDV-TIP effect, TIP for 'transit impact parameter'. These two TDV effects will virtually always be present, but their amplitude may often be quite small.

Fortuitously, Kipping 2009a found that while TTVs and TDVs are not unique signatures of exomoons, when these signals are produced by an exomoon they should be out of phase by precisely $\pi/2$, and this is in fact a unique signature of a moon. The reason for this is easy enough to visualize if we imagine a planet with its moon orbiting edge on from our perspective. If we see the planet and moon in conjunction, the barycenter will also be aligned

with the planet and we therefore see no displacement of the transit, that is, it is neither early nor late based on linear ephemeris[1]. At the same time, we will be seeing the planet at the instant of its maximum tangential velocity with respect to the barycenter, and therefore we will observe it at one of the transit duration extrema. Conversely, if we observe the moon at maximum elongation, the planet will be at its most extreme projected displacement from barycenter, corresponding to a maximum displacement from linear ephemeris, and we will see it with essentially zero tangential velocity, corresponding to its average transit duration. In the case where we can measure a transit duration with adequate precision, this powerful relationship between projected planet velocity and position can be a strong indicator of a moon's presence.

Moon transits

Apart from the TTVs and TDVs, for a transiting planet with an attendant moon we can expect to observe two transits: one of course from the planet, and a second, smaller transit from the moon. So long as the moon's orbital period is not an integer multiple of the planet's period, the moon's transit will appear variously before, during, or after the planet's transit. Clearly, the relative positions of these two transit events reflect the sky-projected positions of the planet and moon, and these positions will also be reflected in the corresponding TTVs and TDVs. For an epoch displaying a moon transit preceding the planet's transit, for example, we know that the moon is leading the planet from our point of view, and therefore the planet's transit should be late. Thus, in the time-domain photometry we have three signals that must all be self-consistent for the moon to hold water. Not only this, but the moon model must adequately explain booth the morphology and timing of every transit in the time series.

The search for and analysis of putative exomoon transit signals, and their associated TTVs and TDVs, makes up the bulk of this dissertation, so we won't belabor the discussion

[1]It is like a wizard in this regard: it arrives precisely when it means to.

now.

1.4 Challenges of The Exomoon Search

The question naturally arises, "if exomoons are so interesting, why haven't we found them yet?". The answer, of course, is that there are a number of critical challenges hindering our ability to detect them, challenges that we are still only beginning to solve. Great strides have been made in the last two decades in terms of theory and mitigating technical roadblocks, as we've seen, but there is still much to be done. In this section I will highlight some of these enduring challenges, and suggest ways forward where appropriate.

1.4.1 Instrumental limitations

Let us first examine the challenges present in our current toolkit. For an observer, the primary tool is of course the telescope. Today observers have a great variety of instruments at our disposal: ground- and space-based observatories, some providing public datasets of unprecedented size and scope, others enabling highly sensitive observations of select targets (generally through priority access to privately funded instruments and/or through a proposal process). A number of technical developments in the last twenty years have provided astronomers new techniques for planet- and moon-hunting, as we've previously discussed. From an instrumentation perspective, there have been extraordinary advances in adaptive optics, astrometric precision, and spectroscopic resolution. As a result, we have achieved incredibly stable imaging from the ground, exceptionally precise distances and proper motions of the stars under consideration, and stellar radial velocities down to as low as 10 cm s^{-1} (Jurgenson et al., 2016). However, despite all these advances, unambiguous exomoon detections remain elusive. Why?

Aperture

As pointed out in Section 1.3, time-domain photometry remains to date the best bet for detecting exomoons in the near term. This poses a key challenge in the detection approach, however. Since the invention of photography, long exposures have been used for imaging faint objects in the sky. For any object that is static on the relevant time scale of the observation, we can typically just integrate for as long as we want to achieve the desired signal-to-noise ratio (SNR). A larger aperture in this case is clearly handy, because a greater SNR can be achieved with shorter integrations, improving the efficiency of the telescope. A larger aperture will also improve the angular resolution of the image, which will be important for resolving extended (i.e. not point-like) sources and distinguishing point sources with small angular separations. And, a larger aperture may also help mitigate the interference or distortion of an image due to atmospheric turbulence. Still, we have been able to see to the edges of the observable universe with apertures that are considered modest by today's standards. The Hubble Ultra Deep Field (Beckwith et al., 2006), for example, revealed the highest redshift galaxies then known, despite its relatively small aperture of 2.4 meters (more on that later).

However, when it comes to time-domain photometry, aperture size is paramount. We do not have the luxury of integrating over minutes or hours for the sake of building up our SNR. Why then, it might be asked, do we routinely see long integrations in time-domain photometry? For example, the standard long-cadence *Kepler* data provided one data point every half hour, while the short-cadence provided an observation every two minutes. TESS, likewise, has provided 30-minute long cadence full frame images (which will become a 10-minute cadence for the extended mission) and a two minute cadence for select targets. In these cases, the answer has chiefly to do with limited download bandwidth, and to a lesser extent, finite detector readout times. But for most telescopes, the readout time will be our limiting factor.

When performing an observation in the time domain, we wish to find an optimum that

maximizes the observing cadence (as short as possible) and minimizes the amount of time lost to overheads. For many instruments, readout cannot be done simultaneous with data collection. Rather, an integration is performed, received photons are converted to electrons by the detector, and those electrons are then counted in a destructive readout process. Then the cycle is repeated. It is apparent, then, that for any finite readout time, increasing the number of readouts will reduce the efficiency of time on target. On the other hand, we have to keep the *time* in the time-domain; the longer we integrate, the less timing precision we achieve. It is also worth pointing out that we do not in fact improve our SNR by integrating for longer. We may shrink the uncertainties on each data point, but this is no different than binning a number of shorter cadence observations, and comes at the expense of reduced timing precision and the blurring of transit features (Kipping, 2010). Therefore, the shortest reasonable cadence is most desirable.

Ground-based limitations

The first exoplanet discoveries were of course made with ground-based telescopes, and though space-based observatories clearly offer some major advantages, there remains a number of key contributions that can and in some cases must be provided by observatories on Earth. For example, telescopes on the ground are not constrained by size and mass the way anything lofted into space will be; with the advent of thirty meter-class telescopes in the near future, there are clearly new technical challenges to be overcome, but there are no restrictions placed on these instruments by the size or power of a launch vehicle.

There are, however, some key difficulties with ground based observations. They are 1) lack of high cadence time-domain surveys, 2) restricted access on major instruments, 3) weather, and 4) short continuous observing windows.

There have been a number of ground-based, time-domain surveys conducted over the past several years – for example, the Palomar Transient Factory (PTF, 1.2 meter Law et al., 2009; Rau et al., 2009), its successor, the Zwicky Transient Facility (Bellm, 2014; Bellm et al.,

2019; Graham et al., 2019; Masci et al., 2019), Pan-STARRS (2 × 1.8 meter) (2 × 1.8 meter Kaiser et al., 2002, 2010), The MEarth Project (8 × 40 cm, Nutzman & Charbonneau, 2008; Irwin et al., 2009), the Kilodegree Extremely Little Telescope (KELT, 2 × 4.2 cm telephoto lens, Pepper et al., 2007), the Hungarian Automated Telescope (HAT, 6 × 11 cm, Bakos et al., 2004), and the Wide Angle Search for Planets (WASP, Pollacco et al., 2006). Also noteworthy is the upcoming Legacy Survey of Space and Time (LSST) at the Vera C. Rubin Observatory (LSST Science Collaboration et al., 2009; Ivezić et al., 2019).

Of these, PTF, ZTF, Pan-STARRS, and LSST are all-sky surveys with coarse time sampling in variety of different bandpasses, with a variety of science objectives, including probing dark energy and dark matter, solar system science, mapping the Milky Way and nearby galaxies, and identifying transient events. They stand to make some contribution in stellar astrophysics, though they are not designed for this purpose, and they are unlikely to provide much leverage on the planet hunting front. LSST, however, may be capable of detecting transiting planets around white dwarfs, however (Cortés & Kipping, 2019; Lund et al., 2018).

On the other hand, the MEarth, KELT, HAT, and WASP observatories have been built for the purpose of detecting planets. These projects have all been modest, utilizing inexpensive, over-the-counter telescopes and optics in concert, and going after the low hanging fruit... hot Jupiters and M-dwarf planets, both of which present deep transit signatures. However, in every case, these projects have focused on a small sample of stars pre-identified as being of interest (of course, *Kepler* also only monitored a fraction of the stars in its field of view with sufficient cadence to detect planets). A good number of planets have been discovered through these projects, but they are clearly incapable of scanning the entire sky.

A particularly exciting project that marries the all-sky approach of ZTF / LSST-type instruments and the high cadence needed for planet hunting is the Evryscope project (Law et al., 2014). This unique design incorporates an array of 23, 7cm telescopes built into a dome, something like the eye of a housefly, but with a fast, 2-minute cadence, on par

with the short cadence of TESS, and can image an unprecedented 3060 square degrees at a time. Of course it cannot go as deep as LSST (Evryscope can achieve down to V=19 for one hour integrations), but it will undoubtedly be a valuable resource for planet hunting, and especially, ephemeris monitoring, when the data become publicly available. Once again, though, the precision and resolving power is limited by the modest size of each telescope.

There are, of course, ground-based telescopes with apertures one to two orders of magnitude larger than the planet-hunting instruments mentioned above. But usage of these instruments is often limited to scientists at member institutions. At the same time, nightly operating costs for these instruments can be enormous. These large telescopes will be oversubscribed for a variety of scientific objectives, but for exoplanets, they are typically devoted to spectroscopic (RV) analysis and in some cases direct imaging. Devoting an entire night to the observation of a single transit is likely unfeasible, or could be rejected as an inefficient use of resources.

For the sake of argument, let us suppose we have a planet worthy of acquiring high cadence transit photometry on a large aperture telescope. Our difficulties do not end here. Utilizing a large telescope would primarily be of use only for longer period planets; after all, for any short period planet, we will have ample opportunity to observe it transit, and can thereby build up the desired SNR for the planet's transit by observing multiple transits (moon transits will be more complicated). For longer period planets, however, we will require larger apertures to acquire comparable SNR on reasonable time scales. For a temperate world, even those interior to 1 AU, we may have only two or three transit observation opportunities in a given year. But if there is bad weather at the observatory on the date of observation, this precious opportunity will be lost, and the scientific result will be deferred if not lost altogether.

As we push to longer period planets – those that are of interest because they may be in the so-called habitable zone, or even colder – we must also contend with longer transit durations. It is quickly apparent that a single telescope on the ground cannot adequately

monitor a full transit, including the acquisition of adequate out-of-transit baseline, once the transit duration becomes a significant fraction of the length of the night. The length of the night, of course, will depend on both the latitude of the observation and the time of year. The only locations on Earth that are sufficient to observe very long transits are at the poles, which can experience darkness for months at a time, but only in the Southern Hemisphere is there land on which to build a permanent structure. In the years to come, we may wish to think seriously about performing transit observations of long-period planets in the southern sky, but before that we will need a critical mass of planet discoveries with very long orbital periods.

The long-duration observation problem is even worse when it comes to searching for exomoons. Now we need extensive out-of-transit baseline, not just to constrain transit depths or characterize and correct for stellar activity. The moon transit itself may be found preceding or following the planet's transit, and in the case where we do not know where to expect the moon transit, either because it is a blind search or because our orbital parameters for a candidate are poorly constrained, we have to observe long enough pre-ingress and post-egress to search for the moon, *and* have enough out-of-transit data to differentiate the moon's transit, as well. This is clearly intractable for virtually all existing optical / NIR instruments on the ground. Only through coordinated observations around the world could we achieve the temporal coverage required. But this approach requires good weather at multiple sites, multiple coordinated proposals, handling of unique systematics for each instrument, and non-trivial stitching together of these datasets.

Space-based limitations

With some of the challenges of ground-based observations now enumerated, let us turn our attention to the challenges posed by moving our observations to space. Naturally, some of the difficulties highlighted above are mitigated by employing telescopes in space, but other issues arise from this choice. Briefly, they are 1) launch vehicle limitations, 2) limited access

of targeted observations, 3) limitations of survey instruments, and 4) limited bandwidth.

Perhaps the most serious limitation of space-based telescopes is their aperture. To date, the largest and arguably most successful telescope in space is the Hubble Space Telescope (HST), located in low-Earth orbit (LEO) with a 2.4 meter mirror. This phenomenal telescope has been the workhorse of space-based astronomy for more than 26 years, and time on the telescope remains highly prized and oversubscribed (in proposal Cycle 27, for example, there were 24,454 orbits requested for the 2,700 available – a factor of ~ 9 oversubscription). The enduring popularity of HST is thanks in large part to its remarkable scientific flexibility and its unmatched precision.

As we pointed out before, a 2.4 meter aperture is by itself a modest scientific grade instrument. There are a great deal of ground-based telescopes of comparable size, some of which are undersubscribed. The key difference is of course Hubble's perch, situated above the atmosphere and thereby avoiding the challenges of weather, atmospheric turbulence, telluric absorption lines, and ultraviolet shielding. HST can achieve 1-2 orders of magnitude better spatial resolution than a telescope of the same size on the ground operating without adaptive optics. In addition, for at least some subset of targets, HST can observe for an essentially unlimited amount of time, uninterrupted (that is, limited to the amount of time an allocation committee is willing to award). For many targets, on the other hand, targets can be observed for long periods but are subject to periodic occultation by the Earth.

HST was famously deployed, and serviced, by NASA's Space Shuttle fleet, so the size of the aperture was restricted to that which could fit inside the payload bay. Launching a future telescope onboard an uncrewed rocket may allow for a somewhat larger payload, but probably not significantly so. To achieve substantially larger apertures, we will have to design more complex, unfolding spacecraft with segmented mirrors, as utilized for JWST and envisioned for the proposed LUVOIR mission.

So far our discussion of space-based instruments has implicitly focused on targeted observations. But as the last decade or so has made clear, the scientific payoff from *survey*

missions can be enormous. With the advent of *Kepler* and TESS, exoplanet science has moved into an era of large data volume, automation, and statistics.

Kepler famously stared at one patch of sky in the Cygnus constellation for ~ 4.5 years, monitoring a few hundred thousands stars for transit signals. Its successor TESS takes a different approach, imaging the entire sky, but only providing ~ 27.5 days baseline for most of the sky; overlapping regions as the telescope scans the sky get longer baselines, and a small fraction of the sky (the "continuous viewing zone" at the ecliptic poles) has a full year of data. It is clear that we have a trade-off here: either we get extensive baselines for a small patch of the sky, or short baselines for the entire sky. Which is better? It depends entirely on your science. TESS will discover any enormous number of close-in planets, but in most cases the period will be precisely known only for planets with an orbital period around 13 days or less. These are therefore inhospitably hot worlds unless they are orbiting very cool stars. Over the mission lifetime there will also be a sizeable sample of longer period planets whose transits just happened to be captured in the time window, but in these cases, period estimates will be challenging and thus, follow-up observations will be also.

It is worth remembering that the hardware restrictions for space telescopes are not limited to mirror size. Another challenge is data transfer rates. Spacecraft in LEO or cislunar space (HST and TESS) can communicate their observations reasonably quickly due to their proximity to home. Objects in orbit around the Sun (*Spitzer* and *Kepler*) are a great deal farther from the Earth, dramatically reducing the bit rate. This is why *Kepler* provided one data point every half hour for most of its target stars, despite the fact that the combined exposure and readout time was 6.54 seconds; the photometry was binned on board.

1.4.2 Surveys vs Targeted Observations

For time-domain photometry, as we've discussed, we have both survey and targeted observations at our disposal. As the results of this dissertation make clear, there is and will remain a role for both of these approaches to finding exomoons. Naturally we would like to

utilize the most powerful telescopes available when performing our work, to achieve the best possible SNR and the most conclusive results. However, it is also clear that these resources are limited, and available only to the extent that an observation is deemed sufficiently worthy when weighed against an enormous number of other scientific objectives. Moreover, targeted observations in the time-domain will likely only yield results for a single target, requiring this target to have exceptional scientific value, at least in the eyes of a time allocation committee. That is, a single target is expected to revolutionize, to some degree, our understanding of these systems.

Given this heavy lift, we must continue to utilize the surveys, which provide the community with a tremendous amount of data at virtually no cost to the researcher. It is essential that these surveys be fully utilized to identify promising targets, particularly those worthy of targeted follow-up. For this, however, we must develop appropriate strategies. How can we best identify promising targets? Such a large dataset presents any number of challenges, including unique astrophysical nuisances for each target, data volume management, and some trade-off between efficiency and depth. One approach is to choose, based on observable metrics, a subset of targets which might be particularly attractive. We might choose, for example, small stars or bright stars, due to their observational attributes, or large planets, due to their physical attributes. Another approach is to attempt a simplification of the process by which candidate moon signals are identified. Fast algorithms that can identify interesting systems through transit timing analysis, periodograms, or supervised learning, appear to be particularly promising avenues.

1.4.3 Stellar activity

In interpreting signals that might be due to an exomoon, we are often faced with the question, "how do we know we are not observing stellar activity?" After all, we know that starspots darken the stellar disk, and these might therefore be interpreted as a transit event. How might we distinguish these two effects?

First, it will be important to determine all that we can about the star itself. In the case of *Kepler*, we have roughly 4.5 years' worth of data with which we can analyze the stellar activity, most of which will not be coincident with a planetary transit. What amplitude of modulation / variability do we see over the course of that observation? Do we see periodicities? Are we able to measure a rotation period for the star? How old is this star, and what degree of activity should we expect from a star of this age and spectral type? We attempt to address some of these questions in chapters 3 and 4.

We know that star spots grow and decay over time, and it is generally believed that spot lifetimes tend to be much longer than the rotation period of the star (e.g. Henry et al., 1995; Namekata et al., 2020). Moreover, the spot rotate in and out of view at the rotational speed of the star. Thus, if we have a star with rotational period of, say, 20 days, the spot will be visible (that is, darkening the disk) for a full 10 days – much longer than the duration of any putative transit signal. This darkening can then reasonably be expected to be removed from the light curve through a detrending approach that removes low frequency signals and preserves those at higher frequencies. By this logic, faster rotating stars will be comparatively trickier, and these stars will tend to be younger and more active (e.g. Skumanich, 1972; Baliunas et al., 1995)

Supposing we have a putative moon transit signal in hand, we can interrogate this further. Coming at it from the other direction, if we have a moon transit that lasts 10 hours, what are the chances that this signal is due to a starspot? For this to be the case, we would have to assume that either 1) the star does a full rotation in twice this time, which would indicate a very (maybe impossibly) fast rotator, or 2) the star spot has grown and decayed in this amount of time, which also seems unlikely. Morever, to the extent that we have multiple dips which we are attributing to the moon, one would need to appeal to multiple rapidly emerging and decaying star spots, all roughly coincident with the planet's transit. For a very active star this may not be wholly impossible, but we would want to weigh these two hypotheses in part on their relative probabilities.

We can perform additional checks on the possibility that the purported moon signal is due instead to stellar activity. To examine whether such a signal could simply be stellar activity, we can perform a moving box test on the full light curve and ask the question, how often do we see a random portion of the light curve displaying a dip of this depth and duration? We perform this test, for example, in Chapter 4, where we ask the question regarding the likelihoood of seeing a hypothesized moon-like dip in a finite time window due purely to the stellar activity. Finally, in the case where we happen to have some spectral information during the transit, we should ask whether the supposed moon transit displays any chromaticity. While we expect stellar activity to be chromatic, a moon transit's computed depth should be the same at all wavelengths; it will not matter if the star is brighter in the red or in the blue, the transit depth will simply be the ratio of radii R_S/R_*, and therefore we should measure the same transit depth in all bands.

So far this treatment has not addressed the potentially thornier issue of whether a planet - star spot *crossing* could mimic a moon signal. When a planet transits directly in front of a star spot, this will cause a temporary brightening of the star, because the combined area of reduced flux due to the planet's shadow and the starspot has now decreased from our point of view. As a result, we will see a peculiar in-transit morphology; instead of a clean, trapezoid-like shape, there will may be two or more dips. At first glance, this can appear very similar to moon signals, which can also be coincident with the planet's transit. However, it is possible to disentangle these two scenarios. To start, the inferred size of the moon required to explain the transit morphology might be difficult to explain in the case of a star spot crossing, and particularly when viewed in conjunction with the presence or absence of timing and duration variations. That is, an inferred moon size would need to be consistent with an inferred moon mass, and if there is tension between these two, the moon hypothesis is suspect. Secondly, we would want to interrogate all the transits of this planet to see whether a full moon model could adequately explain the morphology seen across the various epochs. A large moon might reasonably explain one or two, but might be more difficult to

reconcile if suddenly the planet stopped displaying moon-like signals altogether. We might also attempt to observe the Rossiter-McLaughlin effect, thereby establishing whether the planet's orbit and the star's spin axis are aligned. If they are not, it may be less likely that the planet and starspot intersect repeatedly.

Furthermore, the required transit durations required for a moon to adequately explain the starspot crossings may be too long or too short; in general, moon transit durations can only be between one-half and twice the duration of the planet's transit, depending on its phase and semimajor axis. Finally, it is noteworthy that starspot crossing can *only* mimic planet and moon transits that are temporally coincident. That is, they cannot mimic moon dips that are outside the planet's transit. In such a case, requiring the moon's transit to always coincide with the planet's would mean the moon must be orbiting very close to the planet. Such a hypothesis would again have to be considered in relationship to 1) the physical plausibility of such a system, from the perspective of formation, dynamical stability, and mass-radius agreement, and 2) corroborating dynamical evidence. Therefore, even though starspots are potentially confounding, we have a variety of tools for investigating this hypothesis.

1.4.4 Intrinsic Moon Population

We turn now to a problem of expectations. At present there remain many open questions about exomoons, but to some extent they are are all driven by this central question: what does the underlying population look like? That is, how large will they be, how abundant are they, and where will we find them? These questions will be raised repeatedly throughout this dissertation, and addressed most directly in Chapter 2. But it is worth highlighting from the start some of these questions.

Until we have established a solid observational point of reference for exoplanetary systems, our expectations will be shaped, inevitably, by what we see here in the Solar System. Through theoretical studies we can hope to extrapolate beyond the Solar System, but they are at present fundamentally anchored in explaining and agreeing with what we see in our

immediate neighborhood. To the extent that protoplanetary disk studies are able to explain the formation of systems radically different than our own, these conditions may then point to what the underlying exomoon population may have in store for us, but at present it is challenging to model both planet and moon formation simultaneously.

Let us briefly survey the Solar System. Mercury and Venus have no moons at all, but Earth has quite a massive moon, at $M_S/M_P \sim 1.2\%$. Then we see that Mars has but two tiny moons. It is only when reach the outer Solar System that we see an abundance of very massive satellites. Jupiter, Saturn, Uranus and Neptune all host major moons that are potentially detectable by an outside observer using our present or near-future methods. Of these, Jupiter displays its tidy Galilean moons in Laplace resonance; Saturn has a host of smaller regular moons and an exceptionally massive one, Titan; Uranus displays its peculiar axial tilt, with its moons orbiting co-planar with the equator; and Neptune evidently hosts a captured moon, Triton, evidently originating from the Kuiper Belt (Agnor & Hamilton, 2006). Finally, observations of the Kuiper Belt itself suggests satellites will be common (Stern et al., 2006; Nesvorný & Vokrouhlický, 2019). Pluto, we know now, hosts at least five moons, including the massive Charon, at $M_S/M_P = 12.4\%$. At least three mechanisms – capture, disk accretion, and impacts – are thought to be responsible for the variety of systems we observe here, and it is clear that these various satellite architectures are relics of a complicated Solar System formation history.

Can we extrapolate these observations to make adequate predictions about exoplanetary systems? This is challenging, because we know now that exoplanetary systems display many features that are not present in our own Solar System. We see planets at very small semimajor axes, Hot Jupiters, high eccentricity planets, spin-orbit misalignments, super-Earths, and mini-Neptunes. Confident pronouncements on what we will find out there with respect to moons are ill-advised.

We wish to establish some answers to these questions of moon size, occurrence rate, and planet semimajor distribution, because they bear directly on this question of where and

how often we should expect to find exomoons in existing datasets. Does our present lack of exomoon detections indicate a real dearth of these worlds, or have we simply not been sufficiently clever to identify them? Chapter 2 takes a systematic look at these questions to establish a moon occurrence rate in the *Kepler* data, bearing in mind the degeneracy between moon size and inferred abundance. That is, through the methods utilized in Chapter 2, there may be many small moons, or a few large moons, but telling these two scenarios apart is not trivial.

One more issue affecting our inferences regarding the abundance of transiting moons is the question of moon inclination. To what extent are moons in the population inclined such that they do not transit in every epoch? As highlighted in Martin et al. 2019, inclined moons will not transit every time the planet does, unless their semimajor axis is sufficiently small. We will see this issue arise repeatedly in the coming chapters. Is it reasonable to expect moon orbits to be roughly co-planar with the orbit of the planet? Our own Solar System provides some insight, though we should proceed with caution. Titan and Triton are significantly inclined with respect to their host planets' orbital planes, and of course, all the regular moons of Uranus are aligned with the planet's own "tipped over" orientation. But we are seeing in these cases two different scenarios: in the cases of Saturn and Uranus, the spin axes of the planets themselves are inclined with respect to their orbital plane, whereas with Neptune, Triton's inclination appears to be an artifact of its capture history.

What then do we make of inclined moons in exoplanetary systems? We cannot readily appeal to one scenario over the other, nor can we (at least right now) extrapolate which of these situations is more likely. To constrain these, we would want to establish whether the planet itself displays some axial tilt, but this is not easily observable at present. For a capture scenario, a retrograde solution (like Triton) might be indicative. A retrograde orbit may or may not be distinguishable from a prograde orbit (Heller & Albrecht, 2014), but we can safely infer a retrograde orbit if the moon solution places it beyond the prograde stability limit of $\sim 0.48 R_{Hill}$ (Domingos et al., 2006). Here again, though, we ask: how plausible is a

moon solution that requires a retrograde orbit? It is difficult to answer this given our present lack of observational and theoretical predictions along these lines.

1.4.5 Computational limitations

Compounding the difficulty in detecting exomoons are the demands on computational resources. As of this writing there are more than 4700 confirmed or candidate transiting planets identified by the *Kepler* mission. Naturally we would like to examine all of these worlds. Of course, many of them may not be suitable targets for follow-up, due to (for example) excessive photometric noise. Even so, it is likely there are a few thousand planets in this sample that deserve at least some investigation. How might we go about this? The data volume is simply too large to examine by eye, so we will require a variety of automated tools for processing them, which includes both a reliable, robust detrending approach, and some means for further distilling the sample.

How much does the sample have to be pared down? Kipping et al. 2015 found that the median CPU time for a single KOI moon fit was $\sim 33,000$ hours, and this figure does not even include the various other models we must examine in testing a moon hypothesis. As demonstrated in subsequent chapters, we typically test at least four models: a planet-only model, a TTV model, a zero-radius moon model, and a fully-physical moon model, and we quantify our belief in these respective models by using the model evidences to compute a Bayes factor. It is easy to see though that examining a single KOI may very well exceed 100,000 CPU hours. In this work, and previous exomoon studies carried out through the Hunt for Exomoons with Kepler (HEK) project, these models were run using `MultiNest` (Feroz et al., 2009), a multi-modal nested-sampling algorithm that computes the evidence for each model fit and is particularly well suited to curved degeneracies in high dimensional parameter space. The moon problem can display multimodality in, for example, TTV and TDV harmonics (Kipping et al., 2012), and therefore exploration of this parameter space requires an approach that adequately handles these contours. Still, this extremely compu-

tationally expensive approach severely limits the number of systems that may be examined, and indirectly limits broader participation in the moon search, because such an approach necessarily requires access to high performance computing clusters which not every researcher enjoys.

Our difficulties do not end here, because `MultiNest` is typically limited to fitting no more than ~ 20 parameters. Beyond this and the problem is computationally intractable. This presents a challenge, for example, for running TTV models on short period planets, since each transit time becomes a free parameter and this rapidly explodes the number of model inputs. As shown in Chapter 2, this problem can be circumvented by segmenting the light curves and running multiple TTV models. More difficult is the problem of fitting more than one moon in the system. For a single moon fit, we have a total of 14 parameters: 7 for the planet, and 7 for the moon (period of the satellite P_S, semimajor axis of the satellite a_S, a reference phase ϕ_S, an inclination i_S, longitude of the ascending node Ω_S, a mass ratio M_S/M_P, and a radius ratio R_S/R_P). For each additional moon modeled we must add an additional seven parameters, so practically speaking it is impossible for `MultiNest` to model more than two moons simultaneously. Yet we know that Jupiter, for example, has four large moons. What would modeling of such a system with only one moon look like?

A detailed investigation of this question is needed, and planned for future work. But for now we can tease out some possible results. In modeling the Jovian system with a single moon, we might see evidence for multiple moon dips in a single epoch, and modeling might therefore suggest the existence of four modes with comparable evidences, one for each of the Galilean satellites. All these models could be consistent with transiting moons present across many subsequent epochs. However, the presence of multiple moons may work to cancel out or dilute the transit timing effects induced by the moons, since there are now competing gravitational influences present. Or perhaps the observed transit timing effects, if still detectable, might be quite difficult to reconcile with something that could be produced by a single moon. On the other hand, if the combined gravitational tug of multiple moons

is attributed to a single moon, the previously mentioned attenuation of the resulting transit timing variation signal could lead us to infer an unphysically low-mass moon, which we might reject on those grounds.

It is clear, then, that modeling systems under the single-moon system is problematic, but that we also have little recourse at present. Going forward we will want to make every effort to reduce the computational demands of this work, while maintaining the same rigorous standards required to test the moon hypothesis in a Bayesian framework.

1.5 Summary

In this introductory chapter we started by tracing the history of our scientific understanding of moons in our Solar System and beyond. We then examined the various motivations for finding and studying exomoons. From there we highlighted a wide variety of observational approaches to the problem, and concluded with a look at some of the challenges facing us as we continue this work.

Let us now turn to the body of research that comprises the bulk of this dissertation. In Chapter 2 we characterize the exomoon population in the *Kepler* data, and report our initial results on the candidate exomoon host Kepler-1625b. In Chapter 3 we report on the results of our Hubble observation of this target and announce promising if still somewhat ambiguous results. In Chapter 4 we apply a number of additional tests to the candidate. Chapter 5 explores a new approach to the detection of exomoons, leveraging powerful machine learning techniques. We conclude in Chapter 6 with some final words about the future of the exomoon search.

Chapter 1: Bibliography

Agnor, C. B., & Hamilton, D. P. 2006, Nature, 441, 192

Agol, E., Jansen, T., Lacy, B., et al. 2015, ApJ, 812, 5

Anglada-Escudé, G., Amado, P. J., Barnes, J., et al. 2016, Nature, 536, 437

Angus, R., Aigrain, S., Foreman-Mackey, D., et al. 2015, MNRAS, 450, 1787

Bakos, G., Noyes, R. W., Kovács, G., et al. 2004, PASP, 116, 266

Baliunas, S. L., Donahue, R. A., Soon, W. H., et al. 1995, ApJ, 438, 269

Banfield, D., & Murray, N. 1992, Icarus, 99, 390

Barnes, J.W., & O'Brien, D.P. 2002, ApJ575, 1087

Bastian, T. S., Dulk, G. A., & Leblanc, Y. 2000, ApJ, 545, 1058

Batygin, K., & Morbidelli, A. 2020, ApJ, 894, 143

Beckwith, S. V. W., Stiavelli, M., Koekemoer, A. M., et al. 2006, AJ, 132, 1729

Bellm, E. 2014, The Third Hot-wiring the Transient Universe Workshop, 27

Bellm, E. C., Kulkarni, S. R., Graham, M. J., et al. 2019, PASP, 131, 018002

Ben-Jaffel, L., & Ballester, G. E. 2014, ApJ, 785, L30

Bennett, D. P., Batista, V., Bond, I. A., et al. 2014, ApJ, 785, 155

Beuzit, J.-L., Vigan, A., Mouillet, D., et al. 2019, A&A, 631, A155

Bibring, J.-P., Langevin, Y., Mustard, J. F., et al. 2006, Science, 312, 400

Borucki, W. J., & Summers, A. L. 1984, Icarus, 58, 121

Brown, R. A. 1976, ApJ, 206, L179

Canup, R. M., & Asphaug, E. 2001, Nature, 412, 708

Canup, R. M., & Ward, W. R. 2002, AJ, 124, 3404

Canup, R. M. 2004, Icarus, 168, 433

Canup, R. M. 2005, Science, 307, 546

Canup, R. M., & Ward, W. R. 2006, Nature, 441, 834

Canup, R. M. 2011, AJ, 141, 35

Canup, R. M. 2012, Science, 338, 1052

California Extremely Large Telescope : conceptual design for a thirty-meter telescope, 2002. https://authors.library.caltech.edu/46821/1/CELTgreenbook2002.pdf

Cilibrasi, M., Szulágyi, J., Mayer, L., et al. 2018, MNRAS, 480, 4355

Citron, R. I., Genda, H., & Ida, S. 2015, Icarus, 252, 334

Clarke, C. J., Tazzari, M., Juhasz, A., et al. 2018, ApJ, 866, L6

Colaprete, A., Schultz, P., Heldmann, J., et al. 2010, Science, 330, 463

Cortés, J., & Kipping, D. 2019, MNRAS, 488, 1695

Craddock, R. A. 2011, Icarus, 211, 1150

Davenport, J. R. A. 2016, ApJ, 829, 23

Dones, L., Brasser, R., Kaib, N., et al. 2015, Space Sci. Rev., 197, 191

Domingos, R. C., Winter, O. C., & Yokoyama, T. 2006, MNRAS, 373, 1227

Donnison, J. R. 2014, Earth Moon and Planets, 113, 73

Douglas, S. T., Agüeros, M. A., Covey, K. R., et al. 2016, ApJ, 822, 47

Fabrycky, D. C., & Winn, J. N. 2009, ApJ, 696, 1230

Feroz, F., Hobson, M. P., & Bridges, M. 2009, MNRAS, 398, 1601

Fujii, Y. I., & Ogihara, M. 2020, A&A, 635, L4

Fulton, B. J., Petigura, E. A., Howard, A. W., et al. 2017, AJ, 154, 109

Fulton, B. J., & Petigura, E. A. 2018, AJ, 156, 264

Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, A&A, 595, A1

Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, A&A, 616, A1

Gaudi, B. S., Seager, S., Mennesson, B., et al. 2020, arXiv e-prints, arXiv:2001.06683

Gilliland, R. L., Brown, T. M., Christensen-Dalsgaard, J., et al. 2010, PASP, 122, 131

Gould, A., Dong, S., Gaudi, B. S., et al. 2010, ApJ, 720, 1073

Graham, M. J., Kulkarni, S. R., Bellm, E. C., et al. 2019, PASP, 131, 078001

Hamers, A. S., & Portegies Zwart, S. F. 2018, ApJ, 869, L27

Han, C. 2008, ApJ, 684, 684

Hansen, C. J., Esposito, L., Stewart, A. I. F., et al. 2006, Science, 311, 1422

Hansen, B. M. S. 2019, Science Advances, 5, eaaw8665

Hartmann, W. K., & Davis, D. R. 1975, Icarus, 24, 504

Heller, R., & Barnes, R. 2013, Astrobiology, 13, 18

Heller, R., & Zuluaga, J. I. 2013, ApJ, 776, L33

Heller, R., & Albrecht, S. 2014, ApJ, 796, L1

Heller, R., Marleau, G.-D., & Pudritz, R. E. 2015, A&A, 579, L4

Heller, R. 2018, A&A, 610, A39

Henry, G. W., Eaton, J. A., Hamer, J., et al. 1995, ApJS, 97, 513

Hong, Y.-C., Tiscareno, M. S., Nicholson, P. D., et al. 2015, MNRAS, 449, 828

Hong, Y.-C., Raymond, S. N., Nicholson, P. D., et al. 2018, ApJ, 852, 85

Huber, D., Carter, J. A., Barbieri, M., et al. 2013, Science, 342, 331

Hwang, K.-H., Udalski, A., Bond, I. A., et al. 2018, AJ, 155, 259

Irwin, J., Charbonneau, D., Nutzman, P., et al. 2009, 15th Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun, 445

Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, ApJ, 873, 111

Jones, H. R. A., Butler, R. P., Tinney, C. G., et al. 2006, MNRAS, 369, 249

Jurgenson, C., Fischer, D., McCracken, T., et al. 2016, Proc. SPIE, 99086T

Kaiser, N., Aussel, H., Burke, B. E., et al. 2002, Proc. SPIE, 154

Kaiser, N., Burgett, W., Chambers, K., et al. 2010, Proc. SPIE, 77330E

Khurana, K. K., Kivelson, M. G., Stevenson, D. J., et al. 1998, Nature, 395, 777

Kipping, D. M. 2009, MNRAS, 392, 181

Kipping, D. M. 2009, MNRAS, 396, 1797

Kipping, D. M. 2010, MNRAS, 408, 1758

Kipping, D. M. 2011, Ph.D. Thesis

Kipping, D. M., Bakos, G. Á., Buchhave, L., et al. 2012, ApJ, 750, 115

Kipping, D. M., Schmitt, A. R., Huang, X., et al. 2015, ApJ, 813, 14

Kupo, I., Mekler, Y., Mekler, Y., et al. 1976, ApJ, 205, L51

Law, N. M., Kulkarni, S. R., Dekany, R. G., et al. 2009, PASP, 121, 1395

Law, N. M., Fors, O., Wulfken, P., et al. 2014, Proc. SPIE, 91450Z

Lammer, H., Schiefer, S.-C., Juvan, I., et al. 2014, Origins of Life and Evolution of the Biosphere, 44, 239

Laskar, J., Joutel, F., & Robutel, P. 1993, Nature, 361, 615

Lazzoni, C. 2019, AAS/Division for Extreme Solar Systems Abstracts 51, 101.03

Lewis, K. M., Sackett, P. D., & Mardling, R. A. 2008, ApJ, 685, L153

Lewis, K. M. 2011, Ph.D. Thesis

Lingam, M., & Loeb, A. 2019, arXiv e-prints, arXiv:1907.12576

Lissauer, J. J., Barnes, J. W., & Chambers, J. E. 2012, Icarus, 217, 77

LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, arXiv e-prints, arXiv:0912.0201

Lucey, P. G., Taylor, G. J., & Malaret, E. 1995, Science, 268, 1150

Luger, R., Sestovic, M., Kruse, E., et al. 2017, Nature Astronomy, 1, 0129

Lund, M. B., Pepper, J. A., Shporer, A., et al. 2018, arXiv e-prints, arXiv:1809.10900

Lunine, J. I., & Stevenson, D. J. 1982, Icarus, 52, 14

The LUVOIR Team 2019, arXiv e-prints, arXiv:1912.06219

The Lynx Team 2018, arXiv e-prints, arXiv:1809.09642

Macintosh, B., Graham, J., Palmer, D., et al. 2006, Proc. SPIE, 62720L

Malamud, U., Perets, H. B., Schäfer, C., et al. 2020, MNRAS, 492, 5089

Manchester, R. N., Hobbs, G. B., Teoh, A., et al. 2005, AJ, 129, 1993

Martin, D. V., Fabrycky, D. C., & Montet, B. T. 2019, ApJ, 875, L25

Martinez, M. A. S., Stone, N. C., & Metzger, B. D. 2019, MNRAS, 489, 5119

Masci, F. J., Laher, R. R., Rusholme, B., et al. 2019, PASP, 131, 018003

Mayor, M., & Queloz, D. 1995, Nature, 378, 355

McKinnon, W. B. 1989, ApJ, 344, L41

McQuillan, A., Mazeh, T., & Aigrain, S. 2014, ApJS, 211, 24

Meixner, M., Cooray, A., Leisawitz, D., et al. 2019, arXiv e-prints, arXiv:1912.06213

Miller, M. C., & Hamilton, D. P. 2001, ApJ, 550, 863

Millholland, S. 2019, ApJ, 886, 72

Millholland, S., & Laughlin, G. 2019, Nature Astronomy, 3, 424

Moraes, R. A., & Vieira Neto, E. 2020, MNRAS, 495, 3763

Morais, M. H. M., & Correia, A. C. M. 2008, A&A, 491, 899

Morris, B. M., Curtis, J. L., Douglas, S. T., et al. 2018, AJ, 156, 203

Namekata, K., Davenport, J. R. A., Morris, B. M., et al. 2020, ApJ, 891, 103

Namouni, F. 2010, ApJ, 719, L145

Nesvorný, D., Vokrouhlický, D., & Morbidelli, A. 2007, AJ, 133, 1962

Nesvorný, D., & Vokrouhlický, D. 2019, Icarus, 331, 49

Noyola, J. P. 2015, Ph.D. Thesis

Noyola, J. P., Satyal, S., & Musielak, Z. E. 2016, ApJ, 821, 97

Nutzman, P., & Charbonneau, D. 2008, PASP, 120, 317

O'Brien, D. P., Walsh, K. J., Morbidelli, A., et al. 2014, Icarus, 239, 74

Ogihara, M., & Ida, S. 2012, ApJ, 753, 60

Oza, A. V., Johnson, R. E., Lellouch, E., et al. 2019, ApJ, 885, 168

Pasqua, A., & Assaf, K. A. 2014, Advances in Astronomy, 2014, 450864

Penoyre, Z., & Sandford, E. 2019, arXiv e-prints, arXiv:1908.09339

Pepper, J., Pogge, R. W., DePoy, D. L., et al. 2007, PASP, 119, 923

Perryman, M. A. C., de Boer, K. S., Gilmore, G., et al. 2001, A&A, 369, 339

Phillips, C. B., & Pappalardo, R. T. 2014, EOS Transactions, 95, 165

Pollacco, D. L., Skillen, I., Collier Cameron, A., et al. 2006, PASP, 118, 1407

Pollack, J. B., Burns, J. A., & Tauber, M. E. 1979, Icarus, 37, 587

Postberg, F., Kempf, S., Schmidt, J., et al. 2009, Nature, 459, 1098

Rappenglück, M.A. 2015. Handbook of Archaeoastronomy and Ethnoastronomy. 1197-1204. https://doi.org/10.1007/978-1-4614-6141-8$_1$15

Rau, A., Kulkarni, S. R., Law, N. M., et al. 2009, PASP, 121, 1334

Raymond, S. N., & Morbidelli, A. 2014, Complex Planetary Systems, Proceedings of the International Astronomical Union, 194

Reese, E. J. 1951, Strolling Astronomer, Vol. 5, No. 7, pg. 1 http://articles.adsabs.harvard.edu/full/1952StAst...6...78B

Roth, L., Saur, J., Retherford, K. D., et al. 2014, Science, 343, 171

Sartoretti, P., & Schneider, J. 1999, A&AS, 134, 553

Sasaki, T., Barnes, J. W., & O'Brien, D. P. 2012, ApJ, 754, 51

Sasaki, T., & Barnes, J. W. 2014, International Journal of Astrobiology, 13, 324

Scharf, C. A. 2006, ApJ, 648, 1196

Simon, A. E., Szabó, G. M., Szatmáry, K., et al. 2010, MNRAS, 406, 2038

Skumanich, A. 1972, ApJ, 171, 565

Smith, B. A., Soderblom, L. A., Johnson, T. V., et al. 1979, Science, 204, 951

Spalding, C., Batygin, K., & Adams, F. C. 2016, ApJ, 817, 18

Sparks, W. B., Hand, K. P., McGrath, M. A., et al. 2016, ApJ, 829, 121

Spergel, D., Gehrels, N., Baltay, C., et al. 2015, arXiv e-prints, arXiv:1503.03757

Stern, S. A., Weaver, H. A., Steffl, A. J., et al. 2006, Nature, 439, 946

Stern, S. A., Weaver, H. A., Spencer, J. R., et al. 2019, Science, 364, aaw9771

Stofan, E. R., Elachi, C., Lunine, J. I., et al. 2007, Nature, 445, 61

Sucerquia, M., Alvarado-Montes, J. A., Zuluaga, J. I., et al. 2019, MNRAS, 489, 2313

Sucerquia, M., Ramírez, V., Alvarado-Montes, J. A., et al. 2020, MNRAS, 492, 3499

Szabó, G. M., Szatmáry, K., Divéki, Z., et al. 2006, A&A, 450, 395

Szulágyi, J. 2017, ApJ, 842, 103

Szulágyi, J., Cilibrasi, M., & Mayer, L. 2018, ApJ, 868, L13

Turtle, E. P., Barnes, J. W., Trainer, M. G., et al. 2017, AGU Fall Meeting Abstracts 2017,

P53D-2667

Vanderburg, A., Rappaport, S. A., & Mayo, A. W. 2018, AJ, 156, 184

Waite, J. H., Combi, M. R., Ip, W.-H., et al. 2006, Science, 311, 1419

Walsh, K. J., Morbidelli, A., Raymond, S. N., et al. 2012, Meteoritics and Planetary Science, 47, 1941

Weiss, L. M., Marcy, G. W., Petigura, E. A., et al. 2018, AJ, 155, 48

Winn, J. N., Noyes, R. W., Holman, M. J., et al. 2005, ApJ, 631, 1215

Wolszczan, A., & Frail, D. A. 1992, Nature, 355, 145

Zhuang, Q., Gao, X., & Yu, Q. 2012, ApJ, 758, 111

Chapter 2: On The Dearth of Galilean Analogs in *Kepler*, and the Exomoon Candidate Kepler-1625b I

> *"We wish to pursue the truth no matter where it leads. But to find the truth, we will need imagination and skepticism both. We will not be afraid to speculate, but we will be careful to distinguish speculation from fact."*
> – Carl Sagan, *Cosmos*

2.1 Introduction

Moons present unique scientific opportunities. In our Solar System, they offer clues to the mechanisms driving early and late planet formation, and several of them are thought to be promising targets in the search for life, as several are rich in volatiles (e.g. Squyres et al., 1983; Hansen et al., 2006) and possess internal heating mechanisms (e.g. Morabito et al., 1979; Hansen et al., 2005; Sparks et al., 2016). The moons of our Solar System also demonstrate the great variety of geological features that may be found on other terrestrial worlds.

In this new era of exoplanetary science it stands to reason that moons in extrasolar systems, so-called exomoons, should tell us a great deal about the commonality of the processes that shaped our Solar System and may yield just as many surprises as their host planets before them. Just as the study of exoplanets has complicated our picture of planetary formation by revealing (for example) the existence of Hot Jupiters (Mayor & Queloz, 1995) – worlds without Solar System analogs – so too might moons show us what else is possible and uproot conventional thinking about satellite formation mechanisms.

This chapter was first published in The Astronomical Journal, Volume 155, Issue 1, article id. 36, 20 pp. (January 2018), co-authored with David Kipping and Al Schmitt.
https://ui.adsabs.harvard.edu/abs/2018AJ....155...36T/abstract

Galilean-sized moons ($\sim$ 0.2-0.4 $R_\oplus$) are generally thought to be able to form in a variety of ways. For the regular satellites of Jupiter, the Galilean moons are thought to have condensed out of a circumplanetary disk, akin to planet formation within a protoplanetary disk (Canup & Ward, 2002). This process is expected to limit regular satellites to a cumulative mass of $\mathcal{O}[10^{-4}]$ that of the primary (Canup & Ward, 2006). Higher mass-ratio moons, such as the Earth's Moon, are evidently viable too and may form from catastrophic collisions in the first few hundred million years of the solar system, coalescing from that collision's debris (e.g. Ida et al., 1997). Finally, retrograde Triton is hypothesized to have originated from a capture event via a binary exchange mechanism (Agnor & Hamilton, 2006). Put together, Galilean-sized satellites appear to have formed via at least three independent pathways within the Solar System, and their existence around exoplanets can therefore be reasonably hypothesized.

Galilean-sized exomoons are challenging to detect using the transit method[1] for a number of reasons. First, the transit of a 0.2-0.4 $R_\oplus$ moon across a Sun-like star results in a depth of 3-13 ppm, below the typical sensitivity achievable with *Kepler* (Christiansen et al., 2012). Second, the moon signal will almost certainly be found at each epoch in a different location with respect to the host planet, sometimes occurring before the transit, sometimes after, and at a different projected distance from the planet (Kipping, 2011). Third, multiple moons around a single planet may wash out any transit timing (Sartoretti & Schneider, 1999) or duration variations (Kipping, 2009a,b). The three-body motion combined with the potential for overlapping disks (syzygies), makes accurate modeling of exomoon transits computationally demanding (Kipping, 2011). Fourth, the long-period nature of plausible moon hosts means that relatively few transits are usually available. In this regime, a planet+moon light curve model has sufficient flexibility to almost always provide a superior fit to the limited data (Kipping, 2013b), thereby necessitating rigorous Bayesian approaches to model selection. Despite these challenges, the "Hunt for Exomooons with Kepler" (HEK) project has

[1]Note that the transit method is by no means the only method sensitive to exomoons; microlensing, for example, is another promising avenue (Bennett et al., 2014).

performed Bayesian photodynamical fits of ~60 exoplanets to date (Kipping et al., 2012, 2013,b, 2014, 2015), with no unambiguous detections and limits typically hovering around an Earth-mass.

In this work, our project pursues a different approach to searching for evidence of moons which focuses on seeking a population of moons around a population of planets. Rather than pursuing individual limits which are then combined to constrain the population, we here approach the problem from a broad statistical perspective in order to directly measure the occurrence rate of moons, $\eta_{☾}$. Resulting from this analysis, we identify a single exomoon candidate, Kepler-1625b I. We briefly describe its detection and vetting ahead of scheduled observations of the planet with the Hubble Space Telescope in October 2017.

2.2 Stacking Exomoons

2.2.1 Phase-Folding

The work presented here aims to exploit the power of stacking in order to search for exomoons. Stacking is a familiar technique to those studying exoplanet transits, who typically phase-fold a light curve upon the period. For a linear ephemeris, the transits align leading to a coherent signal. It is important to stress that this act does not improve the signal-to-noise ratio (SNR). The amount of data before and after stacking is the same, with the only difference being that stacking assumes the ephemeris of the planet is known to infinite precision. Modeling the full unstacked light curve with a model conditioned upon the same ephemeris assumption would result in identical posteriors and thus no improvement is achieved for the inference itself. Despite this, stacking is attractive because the signal's coherence means that full light curve modeling is unnecessary in the context of signal detection. Specifically, one may simply evaluate the weighted mean centered around the pivot point of the fold to achieve a detection, which is why the popular Box Least Squares (BLS) algorithm (Kovács et al., 2002) is a computationally efficient yet sensitive tool in conventional transit detection.

Stacking light curves in pursuit of exomoons is complicated by the fact that simply

phase-folding light curves upon a linear ephemeris will lead to the moon appearing at different phases in each epoch. Despite the fact that the moon is not perfectly coherent, it is constrained to lie within a fraction of the Hill sphere radius (Barnes & O'Brien, 2002) and this imparts some quasi-coherent properties into the phase-folded light curve. Simon et al. 2012 were the first to describe this possibility, where they argued that this quasi-coherence will lead to an increase in the photometric scatter in the temporal region surrounding the planetary event - an effect they dubbed "scatter peak". A similar idea is discussed in Heller (2014), who instead considered looking for a slight photometric decrease in this temporal region. By considering the probability density of the moon's sky-projected position, an effect dubbed the "orbital sampling effect" (OSE), Heller 2014 derives a simple formula for predicting phase-folded light curve shape in the presence of moons, enabling a simple approach to seeking exomoon shadows.

As with the case of a planetary transit, or indeed any kind of stacking, this approach does not boost SNR in any way. The data volume and quality are the same before and after the stacking. However, unlike the planet case, the shape of the phase-folded moon signal represents a washed-out depiction of the individual signals. Accordingly, the subtle individual variations in durations, positions and shapes are lost, meaning that stacking imposes a fundamental loss of information content, and therefore sensitivity.

A similar situation occurs when observing planetary transits with long exposures, such as the 30 minute cadence (LC) mode of *Kepler*, causing the shape of the transit to be slightly washed-out, thereby degrading the information content. It is for this reason that short-cadence (SC) *Kepler* data provides tighter constraints on transit times, despite the fact that both see the same SNR transit depth (e.g. see Kipping et al. 2013 and Kipping et al. 2014).

Accordingly, searching for exomoons in phase-folded light curves will always be less sensitive than full photodynamic fits – although precisely how much has not been formally evaluated and would be sensitive to the specific planet-moon parameters. Despite this, phase-folding is attractive for its simplicity and as a pure detection tool – analogous to BLS

for planet hunting.

One particularly attractive feature of phase-folding exomoons is that a sizable fraction of the quasi-coherent signal appears exterior to the planetary transit. Assuming the planetary transits are well-aligned and the duration is well-known, one may simply crop the planetary transit leaving behind a pure exomoon signal. This greatly simplifies the analysis, since the planet properties are not covariant with this signal[2]. For these reasons, we identify this out-of-transit phase-folded moon signal as the target signature in this work.

2.2.2 Planet-Stacking

Unlike a simple planetary phase-fold, the quasi-coherent nature of the phase-folded moon light curve means that a large number of transits are needed to produce a predictable signal. At its core, the phase-folded moon signal depends upon the law of averages and thus co-adding relatively few transits can lead to a phase-folded moon signal which is highly irregular and erratic. Without a characteristic and predictable shape, it is very difficult to convincingly argue the signal is genuinely a moon, rather than some peculiarity of the data in those limited co-added events. Indeed, Heller (2014) argue that at least a dozen events are generally needed, such that $N \gg 1$ and the averaging effect can become noticeable.

Unfortunately, *Kepler*'s primary mission lifetime of 4.35 years means that the long-period planets, where moons are most *a priori* expected to be viable, were only observed to transit a few times. Only in a small number of cases are there *Kepler* planets for which their period is long enough such that moons are dynamically stable for Gyr *and* we possess $N \gg 1$ transit events within the 4.35 years of *Kepler* observations. This point seemingly excludes phase-folding as a viable exomoon approach, except for a few rare cases.

We devised an approach to solve this problem, inspired by the work of Sheets & Deming (2014). In that work, the authors not only phase-folded each planetary light curve but also co-added different planets together. This allowed them to greatly increase the number of

[2]Phase variations are not included in our model and thus cannot be covariant, nor should such phase variations persist given our detrending algorithm will have largely removed them.

ensemble phase folded signals, which in their case was used to boost the overall SNR. In many ways, this approach is reminiscent of a hierarchical Bayesian model (HBM; Hogg et al. 2010) but by stacking the objects the identities of each object are sacrificed. While an HBM approach would be better suited, in general, direct planet stacking is attractive for its simplicity, particularly if the objective is purely to test whether an ensemble signal even exists rather than attempting to perform detailed characterization of said signal. Therefore, in the same vein, we decided to try stacking different phase-folded planet signals together, to solve the $N \gg 1$ problem. We highlight that Hippke (2015) independently arrived at the same idea and published before our effort, although many differences exist in our actual implementations, as will become clear throughout this paper.

Co-adding different planets decreases the overall noise, since we are extending the data volume upon which our inferences are conditioned. However, this approach is not guaranteed to increase the SNR, since many of the objects co-added may not even possess moons and thus their inclusion only dilutes the overall signal, rather than co-adding to it. Nevertheless, we can quantify the overall signal amplitude as being a combination of the occurrence rate and the moon radii. Even so, selecting a sample of planets which are expected, *a priori*, to be plausible hosts for large moons will be crucial for maximizing our chances of a successful detection and correspondingly deriving meaningful, physically-constraining upper limits.

2.3 Target Selection

2.3.1 Automated Target Selection

Not all exoplanets are equally likely, *a priori*, to yield an exomoon detection. At the most basic level, two questions guide our target selection process: 1) what is the largest stable moon plausible around a given planet 2) would this moon be detectable, given the current data in hand?

In this work, we estimate whether a detectable moon is plausible following a similar approach to that adopted in previous HEK papers. Specifically, we employ the *Target*

Selection Automatic (TSA) algorithm described in Kipping et al. (2013).

To summarize, the algorithm first estimates a mass for each Kepler Object of Interest (KOI) using the maximum likelihood radius reported on the NASA Exoplanet Archive (Akeson et al. 2013) and the mass-radius relation defined in Kipping et al. (2013). This is used to further estimate the extent of the planet's Hill sphere. Moons are expected to have their lifetime limited by the time it takes to tidally spin out from the Roche limit to some critical fraction of the Hill sphere, fR_H. Using the expressions of Barnes & O'Brien (2002), this logic may be inverted to compute the maximum allowed moon mass, $M_{S,\max}$, which can survive for a fiducial age of $t_\star = 5\,\mathrm{Gyr}$. For TSA, we set $f = 0.9309$ for the optimistic case of a retrograde moon (Domingos et al., 2006). Note that we assume a single moon for stability estimates and tidal evolution timescales; the presence of multiple moons is expected to modify these values. However, since we do not know *a priori* how many moons may be present in a given system, it is impossible and impractical to apply more sophisticated stability estimates at this stage.

In order to compute $M_{S,\max}$, we must adopt a value for the ratio (k_2/Q), which represents the efficiency of tidal dissipation. Whereas in previous HEK papers we simply adopted $(k_2/Q) = 0.5/10^5$, here we use an empirical relation based on the Solar System. As shown in Figure 2.1, empirical estimates of (k_2/Q) for the Moon (Dickey et al., 1994), the Earth (Kozai, 1968), Neptune (Trafton, 1974), Jupiter (Lainey et al., 2009) and Saturn (Lainey et al., 2012) follow a power-law distribution versus planetary radii, except for Saturn. In this work, we invert this empirical relation (ignoring Saturn) to estimate (k_2/Q) for a given R_P. While we don't claim this to be a fundamentally general law, it at least provides a somewhat more reasonable estimate than the blanket fixed value assumed previously.

Once the maximum moon mass has been computed, it is then converted into a moon radius using the same mass-radius relation as before. Since target selection was conducted early on in the two-year duration of this research comprising this paper, it predated the more robust probabilistic mass-radius relation of Chen & Kipping (2017a), which is why

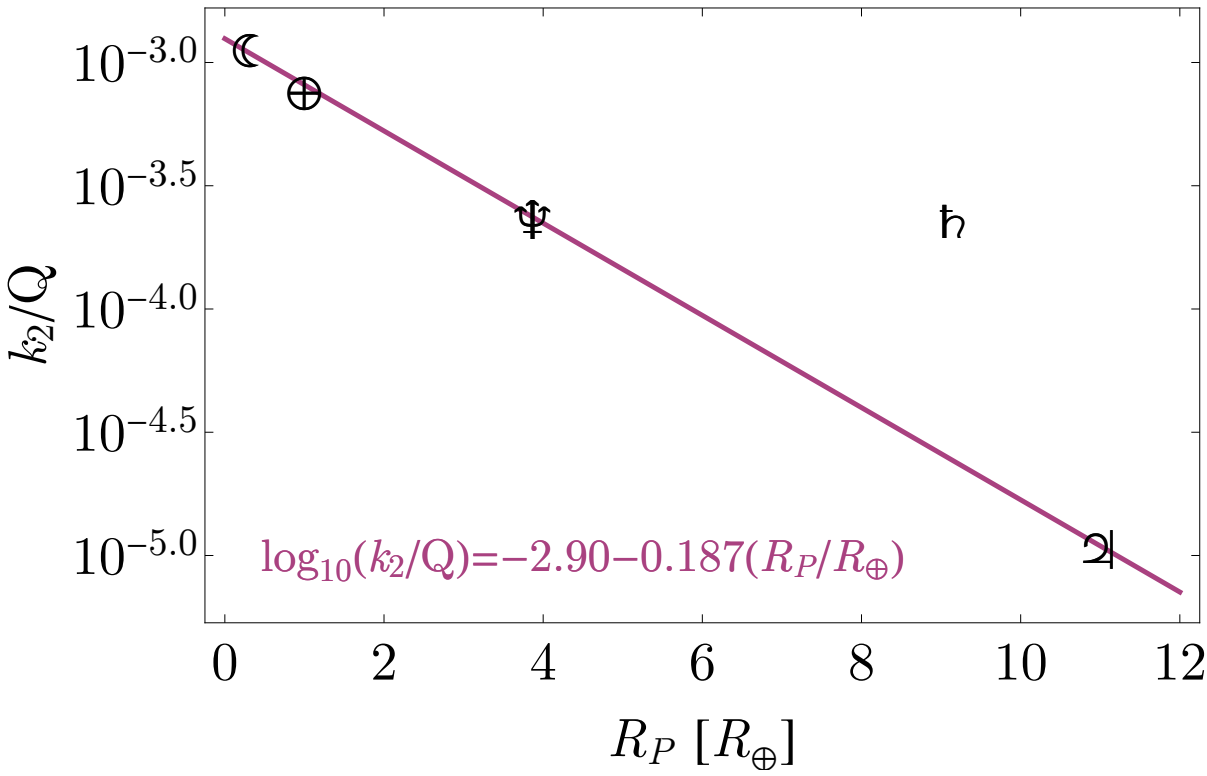

Figure 2.1: Relationship between the tidal property k_2/Q and planet size for the Solar System bodies. When excluding Saturn, the remaining four points closely follow a power-law used in this work.

that latter relation was not used for these calculations. We query the combined differential photometric precision (CDPP) of each host star (Christiansen et al., 2012), which along with the maximum moon radius, allows us to estimate the signal-to-noise ratio (SNR) expected due to the moon.

For the SNR calculation, we estimate the phase-averaged signal amplitude using the so-called "orbital sampling effect" (OSE) described in Heller (2014). While these expressions formally assume a large number of transits, which is rarely true, they work well as an approximation for the signal strength marginalized over the moon's phase, which is of course unknown to us *a priori*. The expected OSE flux decrease for the out-of-transit data is given in Heller (2014), from which we may integrate over the signal to compute the signal strength of the out-of-transit portion to be

$$S = \left(\frac{R_S}{R_\star}\right)^2 \frac{\sqrt{(a_{SP}/R_P)^2 - 4} - 2\cos^{-1}(2R_P/a_{SP})}{\pi((a_{SP}/R_P) - 2)}. \tag{2.1}$$

The fraction on the right-hand side varies from about a quarter to a third for (a_{SP}/R_P) in the range of 5 to 100 i.e. it is a relatively gentle function. We therefore take the limit of large (a_{SP}/R_P), giving

$$S = \frac{1}{\pi}\left(\frac{R_S}{R_\star}\right)^2. \tag{2.2}$$

The SNR may now be calculated by dividing this by the noise expected

$$\mathrm{SNR} = \frac{1}{\pi}\left(\frac{R_S}{R_\star}\right)^2 \frac{\sqrt{T_{\mathrm{Hill}}/0.25}}{\mathrm{CDPP}_6}\sqrt{\frac{B}{P}} \tag{2.3}$$

where B is the **time** baseline of observations, optimistically assumed to be the full Q1-17

baseline for these calculations and CDPP_6 is the combined differential photometric precision on a 6 hour timescale. We may express the Hill time, assuming a simple circular orbit approximation, as

$$T_{\mathrm{Hill}} = f\left(\frac{P}{2\pi}\right)\left(\frac{M_P}{3M_\star}\right)^{1/3}, \tag{2.4}$$

which when substituted in leads to the P terms cancelling out, such that

$$\mathrm{SNR} = \frac{(R_S/R_P)^2}{\pi}\left[\frac{\sqrt{2Bf}}{3^{1/6}\sqrt{\pi}\mathrm{CDPP}_6}\left(\frac{M_P}{3M_\star}\right)^{1/6}\right]. \tag{2.5}$$

As discussed later in Section 2.6, we find that the OSE model overestimates the signal-to-noise for large a_{SP}, with numerical experiments showing it is around a factor of three too high by the time we hit $(a_{SP}/R_P) = 100$. We therefore correct the SNR quoted above by dividing by a factor of ~ 3, which together with by the π denominator we simply approximate to a factor of ~ 10 denominator, yielding

$$\mathrm{SNR} \simeq \frac{(R_S/R_P)^2}{10}\left[\frac{\sqrt{2Bf}}{3^{1/6}\sqrt{\pi}\mathrm{CDPP}_6}\left(\frac{M_P}{3M_\star}\right)^{1/6}\right]. \tag{2.6}$$

2.3.2 Applying to the Kepler Planetary Candidates

TSA was first run for this project in November 2014, at which time 7305 KOIs were listed on the NASA Exoplanet Archive. However, 27 were removed due to having some incomplete column entries. Of these, 4109 were not classified as a "false-positive" by the NASA Exoplanet Archive and thus were considered further. In order to calculate SNR, basic stellar properties are required and so we cross-referenced this list with the Huber et

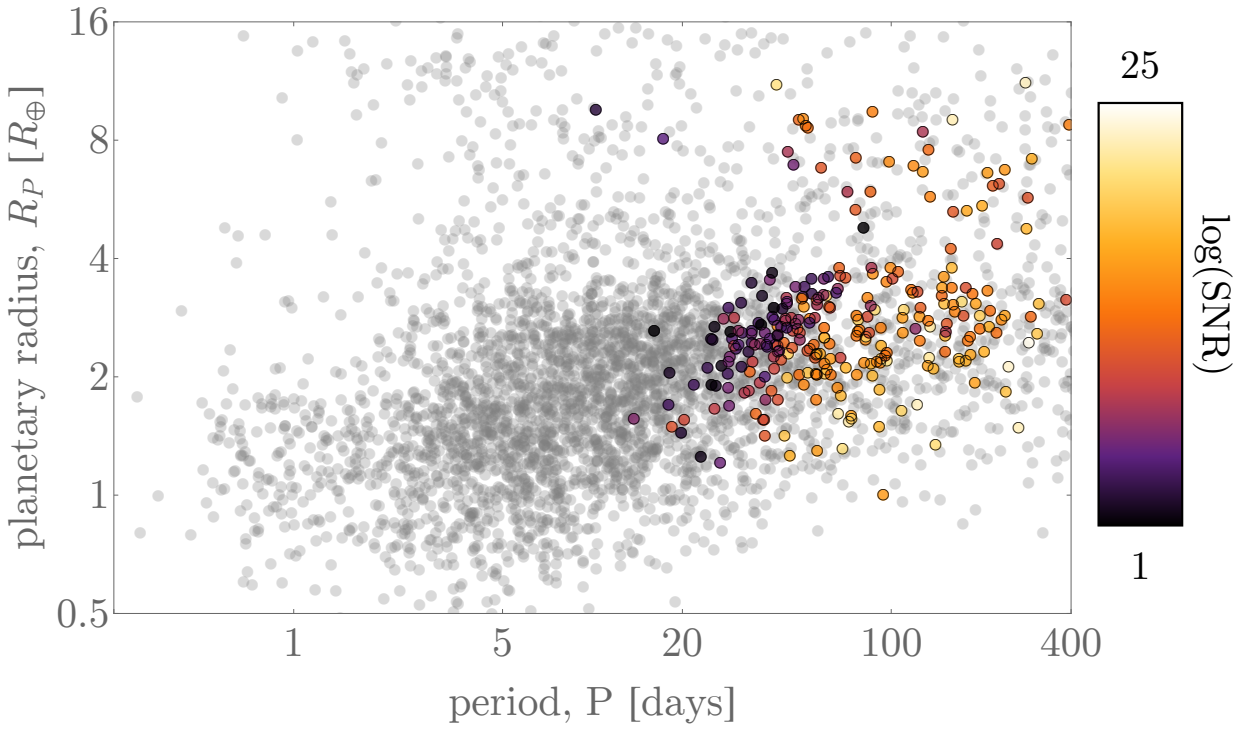

Figure 2.2: Location of the KOIs selected to search for exomoons around. The 966 colored points, color-coded by the SNR given in Equation 2.6, tend to be at relatively long orbital periods, where Hill spheres are larger. Points with a solid, black circle around it are the 347 KOIs found to pass our data quality vetting.

al. (2014) catalog, in which we were unable to find a match for 11 KOIs. They were thus removed giving us a total of 4098 KOIs which were then put through the TSA algorithm.

Due to the ensemble nature of our analysis, the total SNR is expected to be much greater than that of individual objects and thus we can afford to use a relatively generous SNR cut. Accordingly, we elected to use SNR>0.1 and apply the criteria that $P < B/4 = 397.39$ days (to give three transits), leading to a sample of 966 KOIs outputted from the TSA procedure. These targets are visualized in Figure 2.2.

Since this work has taken several years to complete, some of the objects that were considered exoplanet candidates when we began are now considered false-positives. In December 2015, we elected to remove all KOIs if they were either classified as "false-positives" at NEA or the probability of any false-positive scenario was in excess of 1%, as given by the values

listed on NEA. This filter removed 292 ($\sim 30\%$) of the objects originally considered, reducing the number of usable KOIs from 966 to 674.

2.4 Data Processing Requirements

2.4.1 Overview

The objective of this work is to create a phase-folded planet-stacked out-of-transit light curve, which may be used to search for evidence of exomoons. For the sake of brevity, we will refer to this light curve as the grand light curve, or simply GLC, in what follows.

We identify four unique and critical requirements for realizing this objective, specifically:

1. removal of TTVs,
2. very high quality light curves,
3. temporal rescaling,
4. two-pass data processing.

We explain and discuss these requirements in what follows.

2.4.2 TTVs

In order to create an accurate phase-folded light curve of a sequence of planetary transits, it is necessary to ensure the transits accurately phase up. In the absence of transit timing variations (TTVs), this is straightforward and is a simple linear ephemeris fold. However, the signal we seek, an exomoon, will always introduce a small TTV signal into the data (Sartoretti & Schneider, 1999). Moreover, TTVs can be caused by other effects, notably perturbing planets (Holman & Murray, 2005; Agol et al., 2005) and thus TTVs are observed to be fairly common ($\gtrsim$10%) in *Kepler* planetary systems (Holczer et al., 2016). Carefully removing these TTVs is crucial in creating an accurate phase-folded transit signal.

One approach might be to take the catalog of known TTVs from Holczer et al. (2016) and use these for corrections. There are several reasons why this is unsatisfactory for this work though. First, in order to assess robust confidence limits, we require covariant, joint posteriors distributions of the transit times and basic planet parameters, which were not derived in Holczer et al. (2016). Second, whenever possible, accurate phase stacking is aided by first conducting model selection between the TTV and linear ephemeris models, which itself formally requires computation of the Bayesian evidence - again something not derived in Holczer et al. (2016). Third, the TTVs derived in Holczer et al. (2016) were conditioned on a different data set to that used in this work. More specifically, although both Holczer et al. (2016) and this work are based on *Kepler* photometry, our data detrending methods are distinct meaning that these differences should be expected to affect the TTV measurements to some degree. When one is ultimately seeking the discovery of a few parts-per-million signal, these conventionally minor issues cannot be ignored and should be expected to influence the results.

For these reasons, we concluded that creating an accurate grand light curve was not possible without first deriving TTV posteriors ourselves for every system considered.

2.4.3 Data Quality

There is a unique property of the phase-folded moon signal that has strong implications for the data quality requirements, which is not conventionally an issue for planet analyses. The GLC signal is a phase-fold of the planetary transit, after removing TTVs, and thus at any given instant in phase, the moon actually only induces a transit-dip for some fraction, F, of the co-added light curves. Geometry demands that this fraction must always be less than one-half (i.e. $F < 0.5$) for all phase points occurring outside of the planetary transit signal. This is a key point which has a major implication: median binning kills the GLC signal.

This is extremely important, because median binning is a robust point estimate. The

forgiving nature of median binning means that one can actually do a bad job of detrending some small fraction of your light curves (which represent outliers) yet still recover an accurate phase-folded signal. However, if one cannot use median binning, then one is forced to use mean-based estimates which are sensitive to each and every transit co-added. In this case, even a single inaccurately detrended transit light curve will contribute to the phase-stacked signal. Once again, since we seek the detection of a signal with an amplitude of a few parts per million, this cannot be ignored and demands the highest levels of scrutiny and data quality.

We therefore establish that each and every transit used in our grand light curve must be verified to be of very high data quality, which of course greatly increases the time demands needed to complete such an analysis.

2.4.4 Temporal Rescaling

When we finally arrive at an accurate phase-folded light curve for each planet, they must be combined into a single grand light curve. This is similar to the co-addition performed for occultations by Sheets & Deming (2014). In their case, each occultation has a distinct duration and thus simply co-adding the occultations would cause the signal to smear out and produce an averaged signal distorted from the true morphology. To overcome this, Sheets & Deming (2014) re-scaled each event by the known duration and then co-added, producing a more coherent signal. Just as with the occultations, each GLC signal will have a different velocity and impact parameter and thus cause a different duration. However, the problem is actually worse since unlike Sheets & Deming (2014), we don't know what the true duration of each event should be, since the duration is highly sensitive to the semi-major axis of the moon(s), which are of course not yet discovered.

Ultimately, re-scaling will always be flawed since we can't know the semi-major axis of the moon prior to discovering it. A full hierarchical Bayesian model (HBM) would be an appealing way of approaching this problem, allowing each object to have a unique semi-major

axis. However, since each planet would not satisfy $N \gg 1$ transits, the OSE approximation would break down and thus each system would require modeling with a full photodynamic simulation, such as that from `LUNA` (Kipping, 2011). For five years, we in the HEK project have been conducting Bayesian regression of individual systems with `LUNA`, and the computational demands for even individual systems are formidable ($\sim 30,000$ CPU hours per planet). Linking this into a full HBM would be computationally extremely challenging and was not a strategy we elected to pursue here.

Moreover, in this work, we ultimately hoped to find a signal which was visually evident in the final grand light curve and thus not conditional upon the inferences of an HBM. While this does not maximize the information content of the final data product, we are motivated to follow this philosophy on the basis that the discovery of any novel phenomenon, which exomoons would represent, requires a much higher confidence than routine discoveries (Gould et al., 2004).

Accordingly, we seek a method of re-scaling which is "least-bad" and will maximize the expected signal coherence even when marginalizing out our ignorance of the moon's semi-major axis. One approach is that of Hippke (2015), who re-scaled by the duration of the planetary transit. The advantage of this is mostly simplicity, the duration is well-constrained and easy to understand. One downside of this is that even if all the moons had the same semi-major axes, they would still lead to the grand light curve having a smeared out OSE signal, since each system has different barycentric velocity and impact parameter across the star. Another approach, re-scaling by the Hill radii, is not possible since the exoplanet masses are unknown.

Instead, in this work, we argue a better approach is to re-scale the time axis into distance from the planet, in units of planetary radii. This can be accomplished by considering the original Seager & Mallén-Ornelas (2003) equation for the duration of a planet, under the assumption of circular orbits:

$$T_{23}^{14} = \frac{P}{\pi} \sin^{-1} \left(\sqrt{\frac{(1 \pm p)^2 - b^2}{a_R^2 - b^2}} \right). \tag{2.7}$$

If we let $(1 \pm p) \rightarrow 1$ in the above, we recover the transit duration as defined when the center of the planet overlaps with the stellar limb, $\tilde{T}$ (Kipping, 2010). Thus, at contact point 1 & 4, we can think of this instant in time as when a shell of radius p centered on the planet first starts to induce transit-dip features. By extension, we could adapt $(1 \pm p) \rightarrow (1 + t'p)$ in the above, which would equal the duration of a shell of radius $t'p$, centered on the planet, to start/end creating transit-dip features. In this way, we can think $t'p$ as being the orbital distance of the moon at the instant in time when the transit begins/ends. Accordingly, t' represents the planet-moon distance in units of the planetary radius. This convenient form allows us to use the transit observables directly to convert from time into a physically motivated dependent variable via:

$$t' = \left[\sqrt{b^2 + (a_R^2 - b^2) \sin^2 \left(\frac{2\pi}{P} (t - \tau) \right)} - 1 \right] / p. \tag{2.8}$$

If all of the moons shared the same (a_{SP}/R_P), this would produce a coherent signal. In reality, we do not expect this statement to be true, but moons do appear in the Solar System to be distributed log-uniformly with respect to this term (Kane et al., 2013). This approach means that we could model the resulting grand light curve assuming exomoons followed a formulaic distribution for (a_{SP}/R_P), such as a log-uniform.

In order to convert from $t \rightarrow t'$, we need estimates for the impact parameter and scaled semi-major axis. Since our data is not strictly the same as that used for the inferences quoted elsewhere, a self-consistent analysis demands we derive these estimates ourselves, which forms another requirement for our work.

While our conversion equation assumes a circular orbit, if we fit the data under the same

assumption, the relative estimate is at least self-consistent. Further, eccentric planets have smaller regions of stability for exomoons (Domingos et al., 2006) and have likely experienced scattering which decreases the chances for moons further (Gong et al., 2013). Thus, if the planet is eccentric, the incorrect conversion is likely irrelevant since such planets likely do not contribute OSE-signals into the grand light curve anyway.

2.4.5 Two-Pass Detrending

In this work, we use the `CoFiAM` algorithm to detrend the *Kepler* light curves. We direct the reader to Kipping et al. (2013) for details on the algorithm, but essentially its goal is to remove long-term trends without introducing any power, in a Fourier sense, at periodicities less than the transit duration. This requirement ensures that any signals with a time-scale approximately equal to or less than this duration will not be distorted by the detrending process itself, since a transit can be thought of as a Fourier series with the lowest frequency being that of the duration (Waldmann, 2012). Accordingly, both the planet and moon transits are preserved, in contrast to polynomial-based methods which introduce power at all frequencies.

High frequency noise is not even attempted to be removed, but is monitored by measuring the autocorrelation at the cadence-lag and used to optimize both the harmonic filtering and subsequent identification of "bad transits", which are typically rejected.

A disadvantage of `CoFiAM` is the requirement for a precise estimate of the time and duration of all transits in the time series. For this reason, it generally is not useful for blind searches for exoplanet transits. However, when seeking exomoons this requirement is generally true and indeed `CoFiAM` should be really thought of as an exomoon optimized detrending method more than anything else.

As before, transit times and durations are often available for the planets under consideration but those times and durations were conditioned on a different data detrending. In order to make our analysis self-consistent, we must derive these times and durations ourselves.

Here-in lies a chicken-and-egg problem though, since to derive these times we first need detrended data, which itself first requires the times. To tackle this, we use two passes to iterate onto the best solution. The first pass uses the literature values for the times and duration of the transits and then performs `CoFiAM`. The second pass takes the times and durations inferred using the first-pass data product, and then performs a new round of detrending with `CoFiAM`. This approach ensures both self-consistency and reliability in our estimates, and provides several opportunities to vet the data quality ensuring only the highest quality light curves are used in the final analysis.

2.5 Data Processing Pipeline

2.5.1 Overview

As motivated in Section 2.4, we require a joint posterior distribution for the times and basic transit parameters of all planets used for the final analysis. Using a two-pass approach to detrending-fitting ensures that our inferences are self-consistent and conditioned upon the actual data product used in this work. In this section, we provide a detailed explanation of the data processing steps comprising each pass, which are tailored to the specific and unique goals of the two. An illustrative overview is provided in Figure 2.3 for reference.

2.5.2 Pass A

PASS A GOALS

We first re-assert that the primary goals of pass A are to: 1) derive transit times for each transit 2) derive the transit duration for each planetary candidate. These products may then be used to conduct a second detrending later in pass B, since our detrending procedure (`CoFiAM`) requires the times and durations for execution.

These two objectives necessitate detrending of the data first, since we do a covariant detrending plus inference model which require an HBM, which is beyond the scope of this

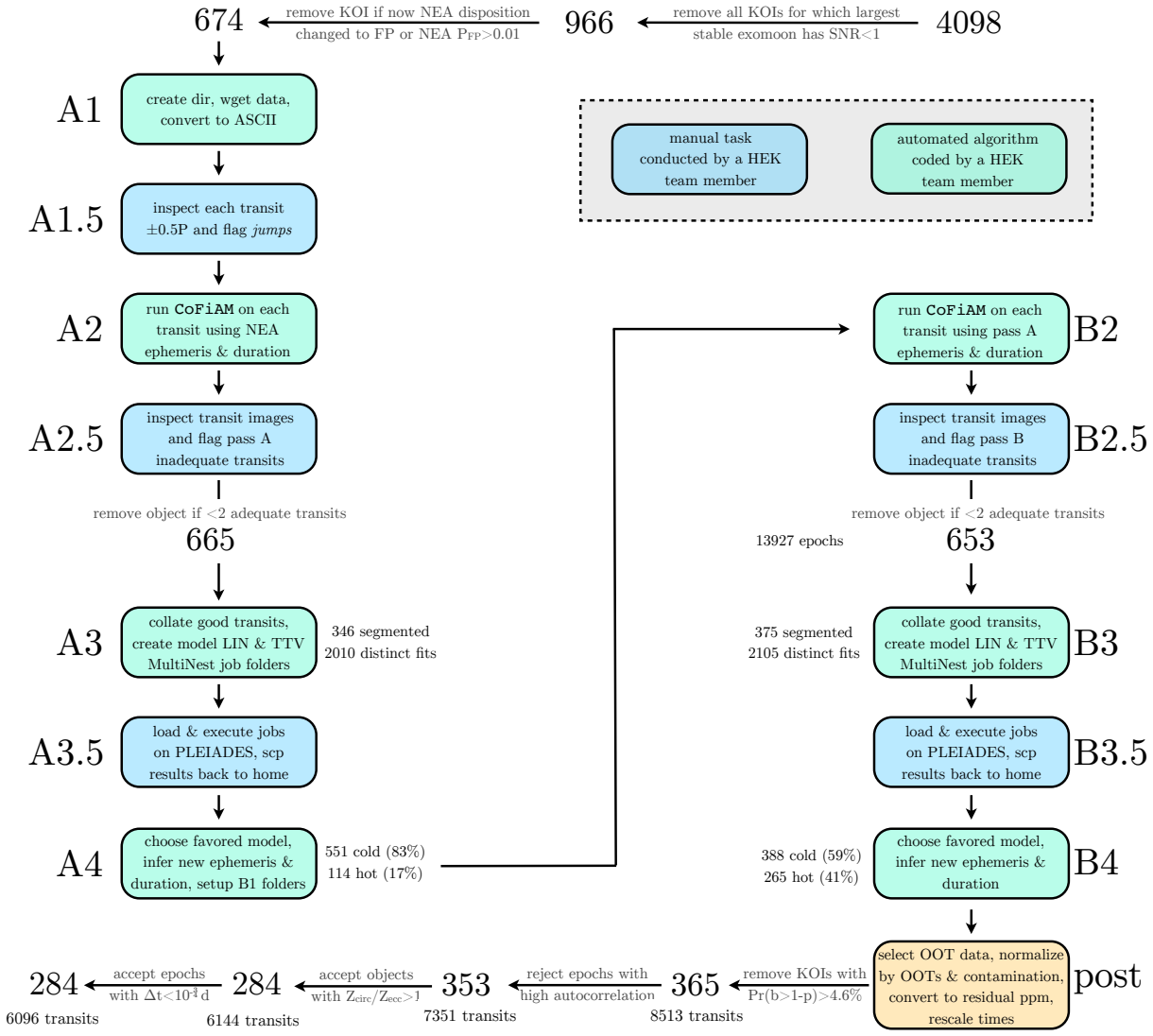

Figure 2.3: Schematic diagram of the pipeline used to process the *Kepler* SAP photometry to ultimately construct the grand light curve. We color code the manual steps in blue and the automated steps in green. The left column broadly describes pass A, and the right column pass B. For details on each step, we direct the reader to the relevant subsection in Section 2.5.

work for reasons described earlier.

For pass A, we make all decisions regarding data quality based upon whether we conclude that these two goals can be achieved. For example, in cases where there is plenty of out-of-transit data, but no in-transit, these would be rejected in pass A (but can be picked up later in pass B since such signals may still contain exomoon transits). Ultimately, defining a clear and independent objective for each pass allows us to optimize required steps.

In what follows, we describe the different stages of data analysis which are performed in pass A.

STAGE A1

The first step is simply to download the *Kepler* Simple Aperture Photometry (SAP) for each target. We wrote a shell script to step through the target list (see Section 2.3), creating a local directory for each KOI. Using a WGET script, any and all of the target's LC data is then downloaded from MAST (from archive.stsci.edu/pub/kepler/lightcurves/) and saved in the appropriate directory. As the data were downloaded in batches, the *Kepler* Science Operations Center data processing pipeline used for these files varied from 9.0.3 (2013-04-18 creation date) to 9.2.23 (2014-11-18 creation date).

Pre-prepared template detrending scripts are copied into each target directory. As with previous HEK papers, we use `CoFiAM` to perform our detrending. We direct the reader to Kipping et al. (2013) and Kipping et al. (2013b) for details on the algorithm. We stress here that the algorithm is designed specifically for the exomoon problem and requires detailed initial information such as planetary periods, transit times and durations in order to work, which ultimately again explains why the two-pass data processing strategy is used.

We use the *Kepler* SAP time series throughout, since this time series has less chance of having unintended artificial signals present than the Pre-search Data Condition (PDC) time series (Stumpe et al., 2012; Smith et al., 2012), which has been subject to data massaging techniques already.

STAGE A1.5

We manually inspect an image of each and every transit light curve epoch, centred upon the time of transit minimum expected from a linear ephemeris and including ±0.5 orbital periods of data either side. We never attempt to stitch different quarters together and instead simply reject any data which occurs in a different quarter to that of the transit epoch under consideration. As before, ephemeris parameters are taken from the NASA Exoplanet Archive for this task.

At each epoch, we identify if any sharp jumps, exponential flux variations or any other anomalous light curve feature exists in the data. This process is performed by one of us with the perspective of whether `CoFiAM` would be able to fit the light curve variations or not. The aim is to keep a sufficiently long series of data to perform a robust detrending, but clip out patterns which may degrade the performance of `CoFiAM`, with an appreciation for the basis set which `CoFiAM` employs.

We initially pursued a variety of automated metrics for this purpose, such as standard deviation, autocorrelation and linear trends. However, we found that a wide variety of anomalous features survived and thus deceived these simple metrics. Rather than creating an ever larger battery of metrics, for which still no guarantee of completeness could be assured, we instead acknowledged that the human eye remains an unparalleled tool in quickly identifying anomalies in time series data.

Anomalous features are flagged by saving the instant in time just before/after the feature, depending on whether the feature occurs after/before the time of transit minimum. This process required approximately 60 hours of human labor in total.

We note that provided cotrending basis vectors are derived from the study of common trends between stars and, in an ideal world, would provide a perfect removal of instrumental effects. However, they do not remove stellar variations, which must also be removed to apply our method. We therefore opt to use `CoFiAM`, not only because it is optimized for the moon problem, but also because it accounts for both instrumental and astrophysical trends in a

single step, which reduces the chances of artificially injecting or removing a moon signal. Fewer manipulations of the data are preferable, and by setting a strict frequency limit to protect the transit Fourier decomposition signal, we ensure that our method does not overfit out small signals of potential interest.

STAGE A2

Stage A2 involves the first detrending of the light curves. As mentioned earlier, this is performed automatically using `CoFiAM` and the list of anomalous features to ignore (found manually in stage A1.5). While details of `CoFiAM` can be found in Kipping et al. (2013,b), we point out some general options selected for the execution in this paper.

The outlier threshold was set to 3-σ from a 20-point moving median. Before detrending, all planetary transits are removed with an exclusion window of $\pm 0.6 T_{14}$ for all events (half a duration either side plus 20% buffer), including the object of interest.

Each transit is detrended separately using $\pm 0.5P$ of the data surrounding each event. If the transit epoch has any associated anomalous flux changes, as discussed in stage A1.5, then points beyond these times are also cropped.

Periodic functions described by a sum of harmonic cosines are explored from twice the data baseline down to twice the transit duration, with a cap of 30 harmonics (beyond which we tend to encounter numerical instabilities). This choice ensures that `CoFiAM` does not disturb the shape of the planetary transit in a Fourier sense (Waldmann, 2012).

Each model is regressed to the data then ranked by the local autocorrelation, as computed using the Durbin & Watson (1950) statistic (DW) on the timescale of the LC cadence, with the lowest autocorrelation being favored. The favoured model is then applied to original time series, re-including the planetary transit of interest. The final light curve is saved with a $\pm 6.66 T_{14}$ gap either side (which is the local timescale used for the DW calculation).

Finally, an image of every detrended transit is stored along with the best DW statistic.

STAGE A2.5

This is the second manual stage in pass A, where we manually identify "bad" transits. In some rare cases, `CoFiAM` fails and produces a light curve which cannot be used for fitting, due to visually evident trends remaining in the data. For example, if we missed a location of an anomalous flux variation in stage A1.5, `CoFiAM` may be trying to detrend sharp features with a smooth cosine function, producing a poorly detrended light curve. By manually going through the light curves in this way, it is essentially a second-check of the data quality, catching any missed anomalies from earlier.

In general, these assessments are made by one of us searching for any visually evident trends which would significantly impede our ability to fit the light curves to determine T_{14}. A bad transit does not necessarily have a poor DW statistic, although that tends to be a common scenario. Because we anticipate a second pass, we can be generous in considering acceptable data qualities at this stage. This acceptance level is non-constant, since we try to allow KOIs displaying very frequent bad transits to have at least a few transits which can be used for fitting in stage A3. Vice versa, if a KOI has many clean transits, we apply more stringent conditions in assessing data quality. Finally, we note that assessments are generally based on SNR of the transit, not the raw wiggles in the data, but the relative size of those wiggles compared to the transit. In cases of very low SNR, where the transit is not visible in a single epoch, we work under the assumption that sometimes data will wiggle up and sometimes down, but we must trust that on the average there is power (thus we try to allow for almost anti-transit like features in the interests of being unbiased and balanced). In such cases, our criteria switches from trying to make a good measure of the transit duration to simply avoiding "catastrophic" detrendings.

In addition to a "bad-transit flag", we also use a "sparse flag" for transits where there is insufficient in-transit data (or none at all). In some rare cases, a third type of flag was used, "missing flag", where the data are well-detrended, we have good temporal coverage, but a transit which should be visually obvious (given the transit depth) is missing in the data.

We consider these cases to be most likely due to an erroneous transit ephemeris on the the NASA Exoplanet Archive.

If fewer than two good transits remain for a KOI, the object is removed from our sample as being a useful object. In total, this removes 9 KOIs dropping our sample down from 674 to 665 KOIs. It should be noted that any rejection of bad data, manual or automated, injects additional uncertainty into the occurrence rate of exomoons calculated in this work which is, strictly speaking, of unquantifiable magnitude. If there is any correlation between unusable transits and the presence of moons the calculation could be particularly affected, but this is impossible to measure since the bad transits are by definition unusable for exomoon characterization. We proceed under the assumption that bad data result from instrumental effects that are distributed randomly across the data set.

STAGE A3

The third-stage is an automated shell script which begins by stitching the good transits together for each KOI into a single file. Our script then creates two directories for a linear ephemeris model fit, $\mathcal{H}_{\rm LIN}$, and a transit timing variation model fit, $\mathcal{H}_{\rm TTV}$, to be fed into MultiNest (Feroz & Hobson, 2008; Feroz et al., 2009). The script then queries the orbital period, P, and time of transit minimum, τ, from the NASA Exoplanet Archive database to construct priors for these terms. In the case of model $\mathcal{H}_{\rm LIN}$, the prior on both terms is uniform centered on reported the NASA Exoplanet Archive value with ± 1.0 days window. For model $\mathcal{H}_{\rm TTV}$, the period is treated as a fixed parameter with the individual transit times following a uniform prior centered on the expected time of transit for a linear ephemeris with a ± 1.0 day window again. The choice of the 1.0 day window is essentially arbitrary, but assumes that transit timing variations larger than one day are highly unlikely in the region between 0.1 and 1.0 AU (c.f. Holczer et al. (2016) who find $< 0.3\%$ of their sample have TTV amplitudes larger than one day). A larger window will unnecessarily increase the time it takes to explore the parameter space, while a smaller window tailored to the linear

ephemerides of each planet could result in a bias against finding large TTVs.

For model $\mathcal{H}_{\rm LIN}$, the basic free parameters are P & τ, two quadratic limb darkening terms (we use the q_1 & q_2 prescription from Kipping 2013a), the ratio-of-radii, p, the impact parameter, b, and the stellar density, $\rho_\star$. Uniform priors are adopted for all except $\rho_\star$ which follows a log-uniform prior. This gives a total of $d = 7$ free parameters, which is easily handled by MultiNest. For model $\mathcal{H}_{\rm TTV}$, the period is treated as fixed, removing one degree of freedom, but then each transit epoch has a unique τ parameter (same prior as $\mathcal{H}_{\rm LIN}$), giving us $d = 6 + N$ free parameters. For $d \gtrsim 20$, the performance of MultiNest is severely impeded and global fits are not possible.

A common approach is "templating", where one folds the transits, creates a template which is then regressed to the individual epochs (e.g. Holczer et al. 2016). This approach underestimates measurement uncertainties since it ignores the covariance between the transit shape parameters and the individual transit times. Rather than underestimating errors, we prefer to overestimate them and so adopt a different strategy. Instead, we split up our light curves in segments of ~ 10 epochs each, providing a manageable number of free parameters for each. The downside is that each segment is not able to utilize the information about the global transit shape learnt from other segments, and so the uncertainties will be larger (but more robust) than templating. Accordingly, in stage A3 our script automatically segments the data up where necessary.

We find a total of 346 out of our sample of 665 KOIs require segmenting, whereas the rest are treated in a single fit due to the tractable number of epochs available.

STAGE A3.5

Stage A3.5 was primarily performed using NASA's Pleiades cluster, and essentially involved loading, compiling, executing and then retrieving the over two thousands light curve regression jobs required. In total, we estimate that approximately $\sim 100,000$ CPU hours were used during this phase of the analysis and spanned several months of wall time.

STAGE A4

In the fourth stage, we segue into pass B by laying the ground-work needed using the results from our light curve fits. The first task is to decide for each KOI whether it is dynamically "hot" or "cold", by which we mean whether $\mathcal{H}_{\rm LIN}$ model (cold) or $\mathcal{H}_{\rm TTV}$ model (hot) is preferred. In cases where the fits were completed using a single segment, the evidences from MultiNest can be directly used to compute the Bayes factor and rigorously assign the preferred model. For segmented models, direct evidence calculation is not possible since the TTV model has multiple copies of the same parameters for the transit shape. Instead, we use weighted linear regression to find the maximum likelihood linear ephemeris through the posterior transit times and then inspect the residuals for evidence of TTVs. This is simply done using a p-value test searching for a excessive χ^2 (cut off used was $3\,\sigma$).

Formally, model assessments using a p-value are incorrect since they are never actually compared to another model. More precisely, the p-value test is prone to inferring significant evidence for the alternative hypothesis even in cases where it should not. For example, excess noise from other sources or a single poor measurement could lead to the p-value test favoring the TTV model erroneously. Let us consider the effect of this by imagining a linear ephemeris fit to a set of transits with the TTV model. The times of transit and basic transit parameters will all still come out formally correct, just with inflated uncertainties. Giving each epoch a free transit time is still able to recover the original linear ephemeris solution. Therefore, despite the p-value's tendency to overestimate significance, this merely acts to conservatively inflate our uncertainties and does not formally invalidate our inferences.

In total, we find 551 of the 665 KOIs are "cold", with the remaining 114 being dubbed hot. For comparison to later, we point out that one might expect pass B to increase the hot fraction due to the improved detrending and thus greater sensitivity to even small TTVs.

Stage 4 ends by duplicating all of the KOI folders into a new directory with a small text file recording the favored model and ephemeris parameters. The transit duration is also recorded in this file, where for $\mathcal{H}_{\rm LIN}$-favored KOIs is computed directly from the joint

posteriors and for $\mathcal{H}_{\rm TTV}$-favored KOIs from a weighted sum of each segment's marginalized credible interval for the duration.

2.5.3 Pass B

PASS B GOALS

Before describing each data processing step for pass B, we first re-assert the objectives, which represents the backdrop against which all decisions in pass B are framed.

Ultimately, the data product from pass B should be high-quality, cleaned light curves with accurate estimates of the transit times and basic parameters needed for stacking and re-scaling. Our tolerance for poor-quality light curve is necessarily more stringent here, since unlike pass A, there is second-chance for these light curves and they have to be of sufficient quality for stacking by the time pass B is complete.

STAGE B2 & 2.5

Downloading the data (stage 1A) and removing jumps (stage A1.5) does not need to be repeated since the raw data product is unchanged. Accordingly, we skip straight to stage B2. Mirroring stage A2, we detrend the SAP light curves using `CoFiAM` but now using the duration and transit times determined earlier in stage A4 (specifically we use the maximum *a posteriori* values).

In stage 2.5 we again inspect these light curves visually for poorly detrended examples and find 12 KOIs end up with fewer than two usable transits after this process. Removing these objects reduces our sample from 665 to 653 KOIs.

STAGE B3 & 3.5

As with stage A3, stage B3 collates the good transits and sets up folders for models $\mathcal{H}_{\rm LIN}$ and $\mathcal{H}_{\rm TTV}$ ready for fitting with MultiNest. Once again this results in just over two thousand distinct jobs to run, with 375 of the KOIs being segmented.

Unlike stage A, we here allow the light curve model to account for any known blending from nearby sources. These are collated from Everett et al. (2015); Kolbl et al. (2015); Adams et al. (2012, 2013); Dressing et al. (2014); Law et al. (2014), and we use *Kepler*-converted magnitudes to estimate the appropriate contamination factor for each band. These blending factors are treated as Gaussian priors, with a standard deviation set by the uncertainty on the converted magnitude. In total, 39 of the targets required a blending term to be included.

In stage 3.5, we again load, compile, execute and retrieve these runs on the NASA Pleiades cluster, requiring another round of 100,000 CPU hours and several months of wall time.

STAGE B4

Finally stage B4 repeats stage A4 with the new light curve fits, performing model selection using the same framework described earlier. After completion, we found that the fraction of cold KOIs indeed decreased as expected from 83% to 59%, giving 265 hot KOIs in our sample. In this work, these TTVs represent purely a nuisance but we acknowledge that this data set represents a rich and interesting catalog for others in the community. We therefore make all of the transit times, both for the hot and cold samples, publicly available at `github.com/alexteachey/TTV_posteriors`.

For each KOI, we export the maximum *a posteriori* transit fit and the corresponding vector of out-of-transit baseline fluxes (OOTs), which are found by linear minimization of the maximum *a posteriori* model and the data (this is actually done on the fly during every regression step, following the approach using by Kundurthy et al. 2011). These OOTs will be useful later for stacking the final light curves.

2.5.4 Post-Processing

In post-processing, we aim to export a single file for each KOI which contains a phase-folded light curve suitable for planet-stacking. First, the transit times are removed using the favored model and the maximum *a posteriori* parameter vector. Next, the planetary transit

is removed by excluding all data which falls within the $2\,\sigma$ upper limit of the derived full duration, T_{14}. Each epoch is then corrected for any residual DC power detected by the OOT vector regressed during Stage B4. Global blending factors, as well as quarter-to-quarter aperture flux contamination factors are corrected for following the approach described in Kipping & Tinetti (2010).

Next, we subtract unity from the normalized fluxes and multiply by a million to create a ppm residual light curve for each object. Finally, the time column is converted to t' using Equation (2.8) and the maximum *a posteriori* transit parameters derived from the preferred model regressed in Stage B4.

2.5.5 Filtering

Before stacking these planets together, we first remove KOIs and individual transits which fail to satisfy several criteria. First, we require a $2\,\sigma$ confidence that the planet is not a grazing event, which would mean that $b > (1 - p)$. Grazing events have degenerate planetary radii and could be far larger in size, potentially even a false-positive. Erring on the side of caution, we remove any such KOI which filters out 288 KOIs, leaving us with 365 targets.

We next test for excessive autocorrelation using the DW metric. For each transit, we generate 10,000 mock realizations where the data are drawn from perfect normal distributions at the exact same sampling observed in the data and the reported uncertainties. We use these synthetic transits to generate an expected distribution for the DW metric and flag any transits for which the real DW metric is more than two standard deviations away from our synethetic population. If a KOI has 50% or more of it's transits flagged in this way, the entire KOI is dropped from the sample. This filtering removes a further 12 KOIs.

Next, we compare the transit derived stellar density (which assumes a circular orbit) to an independent estimate, in order to exploit the photoeccentric effect (Kipping et al., 2012b; Dawson & Johnson, 2012) to infer a minimum eccentricity, $e_{\rm min}$, of each KOI (Kipping,

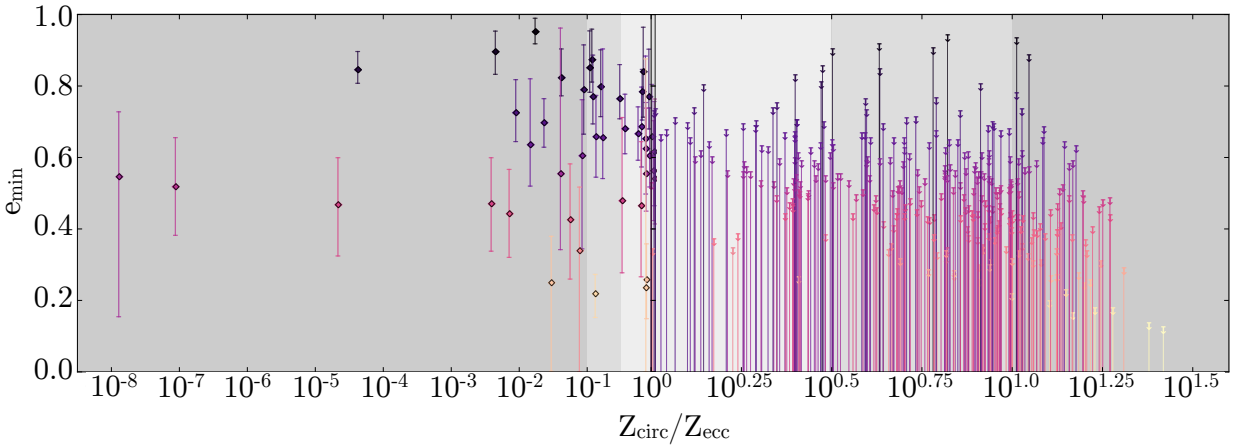

Figure 2.4: Minimum eccentricity of 353 KOIs derived using the photoeccentric effect as a function of the Bayes factor for a circular vs eccentric orbit. KOIs to right of unity are depicted as upper limits on eccentricity, whereas we plot 1 σ credible intervals for the others. The 284 KOIs favoring a circular orbit are considered further as viable exomoon candidates in this work.

2014). We draw a random sample from the transit derived posterior found in stage B4 and divide it by a random sample drawn from the KOI's corresponding stellar density posterior derived in Mathur et al. (2017). This ratio is then converted into a minimum eccentricity using Equation (39) of Kipping (2014), and the process is repeated until we have derived 40,000 posterior samples for $e_{\rm min}$ for each KOI. For each KOI, we also construct a prior for $e_{\rm min}$ based off the prior used in the transit fits and Mathur et al. (2017) distribution.

We next evaluate the Savage-Dickey ratio between the posterior and the prior to estimate the Bayes factor, $Z_{\rm circ}/Z_{\rm ecc}$. We find that 284 KOIs have $Z_{\rm circ}/Z_{\rm ecc} > 1$, implying a near-circular orbit, whereas the other 69 KOIs are rejected for further analysis, on the basis that eccentric planets likely result from scattering which would disrupt moon systems (Gong et al., 2013). We plot e_{min} as a function of the Bayes factor in Figure 2.4.

Finally, we elected to remove transits for which we are unable to measure the transit time to within a precision of $10^{-0.75}$ days, chosen to remove unconverged posteriors given the prior width, which is necessary to ensure we are able to reasonably phase-fold transits together. This did not change the number of KOIs from 284, but did reduce the number of transits in

our sample from 6144 to 6096.

2.5.6 Constructing a grand light curve

With each target now having a fully processed phase-folded light curve, we are finally ready to stack different targets together to create a grand light curve. This stacking can be across all 284 surviving targets, or a subset of them, as explored later. Although we describe here the case for the complete ensemble, the planet-stacking methodology is the same when dealing with subsets.

Across the 284 KOIs, we have 6096 unique transits comprising of 364059 photometric measurements. The re-scaled times are well-described by a half-normal distribution with a standard deviation of 113. We elect to remove any points which fall outside of the range $t' > 150$, leaving us with 309750 points.

The grand light curve photometry shows no evidence for correlated noise structure, as expected from averaging so many independent data sets together. This is verified in Figure 2.5, where we plot the root mean square (r.m.s.) of the time series as a function of bin size, which displays excellent agreement with the expected root N scaling.

Dividing the fluxes by their formal reported uncertainties, we find that the robust r.m.s. (given by 1.4286 multiplied by the median absolute deviation) equals 1.09, indicating only a small amount of extra noise above the reported uncertainties. We re-scaled the errors by this factor and then performed $3\,\sigma$ clipping, removing just under one percent of the points. The final time series is found to have a standard deviation of 5.1 ppm when binned to a scale of $\Delta t' = 0.5$.

2.6 Modeling

2.6.1 Choosing a modeling formalism

In addition to manipulating the *Kepler* data to construct the grand light curve, we also require the ability to model its shape, as a function of various exomoon parameters of interest.

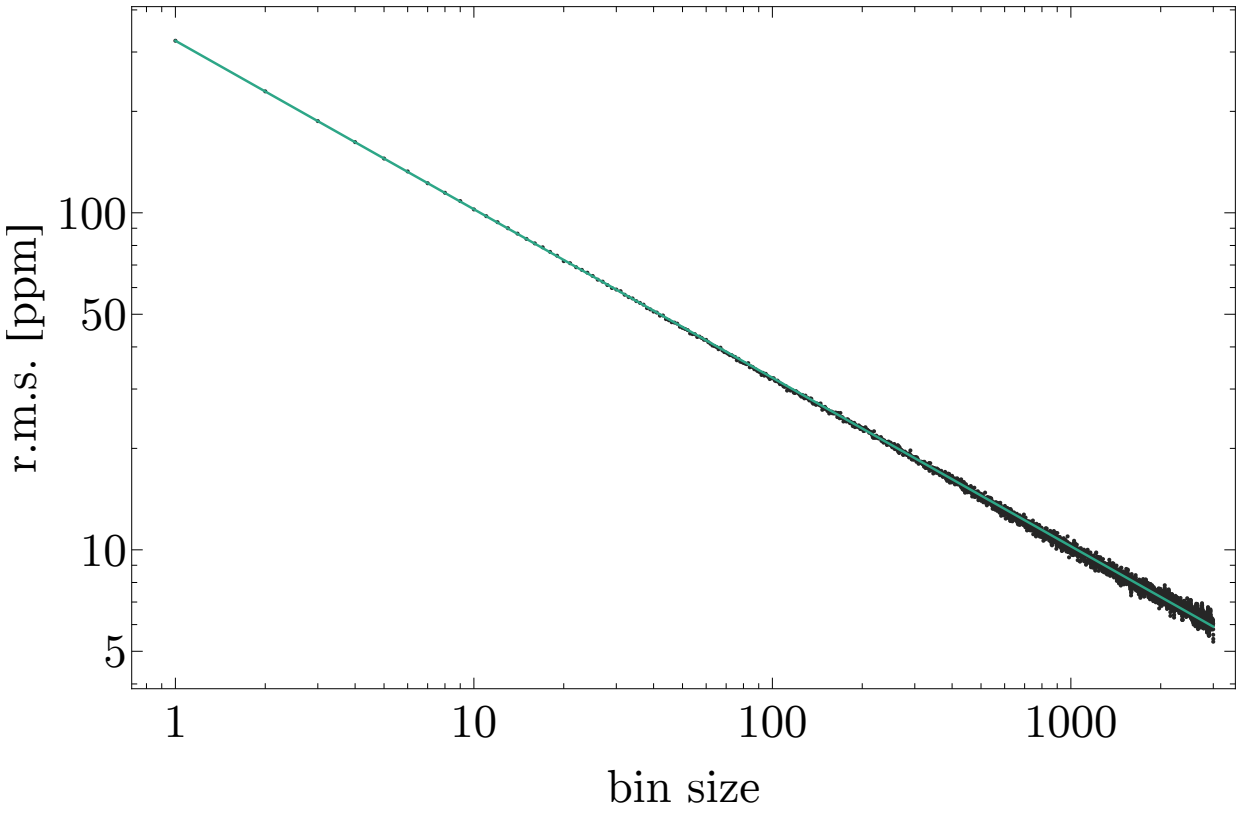

Figure 2.5: RMS of the grand light curve as a function of bin size, demonstrating the expected $N^{-1/2}$ scaling of white noise.

There are two possible avenues to modeling the grand light curve. The first is to model the individual systems then combine them to create an ensemble model, and the second is to use a model describing the ensemble from the outset. The latter approach describes the model proposed in Heller (2014) and later modified in Heller et al. (2016), who refer to this model as the "orbital sampling effect" (OSE). The great advantage of this approach is that one can employ analytic expressions described the ensemble signal without having to laboriously simulate each of the individual systems and then combine later. Thus, in principle, the OSE approach has the advantage of speed and being more straight-forward in application. Indeed, this was the model used in Hippke (2015).

The alternative approach would be to use a detailed "photodynamical" light curve simulator, such as `LUNA` (Kipping, 2011), to predict the light curve of each system with some trial set of moon parameters and then later combine them to produce a grand model. Photodynamics, a test first coined in Carter et al. (2011), refers to a light curve simulator which evolves a planetary system at each time step and computes the corresponding shadows cast onto the sky-projected stellar disk. Unlike the OSE model of Heller (2014), this model is not specific to phase-folded events but of course can be easily used to simulate such a case by simply folding the final predicted light curve.

In general, `LUNA` provides a physically detailed light curve simulation, but comes at the expense of greater computational cost than the simple closed-form expressions of OSE. For these reasons, if the accuracy of OSE is validated, it would be far simpler and thus preferable to employ the OSE formalism for our model fits of the grand light curve. However, after photodynamical testing of the OSE predictions and consideration of the specifics of our problem hand, we came to the conclusion that OSE would not be an accurate modeling tool for our data product. In particular, we argue that three key barriers prevent us from directly using the OSE models to describe our grand light curve:

- Inter-population variation,
- Heteroscedastic weighting,

- Laplace resonances.

We briefly describe these three reasons in what follows.

Inter-population variation

The OSE model is derived assuming one co-adds many transits of the same planet-moon system i.e. that the basic parameters of the system are not changing. However, in our case we co-add different systems together which have distinct planet and moon parameters. For example, in our Galilean moon fits, we assume that the moons have inclinations representative of Io, Europa, Ganymede & Callisto, but each planet's moon system will have unique moon inclinations, despite being drawn from a common underlying distribution. While we could co-add many OSE signals together modeled individually with the corresponding parameters, each OSE curve would be modeling only a small number of transits and thus would be formally invalid - since it is by definition an ensemble model. Without detailed investigation, it was unclear that one could simply co-add across a population in this way and recover the correct phase stacked signal and thus we preferred to use `LUNA` which provided an accurate model of the individual events.

Laplace Resonances

A subtle and minor point of concern was dealing with the Laplace resonance in the OSE framework when modeling Galilean analogs. In `LUNA` each individual moon transit is generated and thus we are able to assign relative phases between the satellites such that they reside in not only the correct mean motion resonance but also share the Laplace resonance in terms of their mutual phases: $\pi = \lambda_{\rm Io} + 3\lambda_{\rm Europa} + 2\lambda_{\rm Ganymede}$. In contrast, the OSE framework never models the individual events, rather just the ensemble, and we were unable to demonstrate that OSE was correctly accounting for such a phase lock.

Heteroscedastic weighting

Finally, OSE is an average of light curves, which by definition means each light curve is given precisely the same weight. Second, each light curve contribution is assumed to be uniformly and densely sampled. Our real data products do not satisfy such constraints, since first we co-add the different planets together using weights based off the root mean square of the photometric residuals. Second, light curves are not uniformly sampled, featuring data gaps and removed outliers, as well as being non-uniformly transformed in time via our temporal re-scaling. Since `LUNA` models individual events, data gaps, integration time effects (Kipping, 2010b) and re-scaling are easily accounted for before applying any co-addition, enabling us to ensure our model is representative of the data.

2.6.2 Photodynamic Look-Up Tables (LUTs)

As a result of the myriad of complicated effects influencing the final model yet the relative low-dimensionality of the model itself, we elected to build a grid of pre-computed models which accounted for all of the effects described above. For each KOI, we took the maximum *a posteriori* transit parameters from the planet-only fits (using the favored model) and generated a planet+moon light curve using `LUNA` with the same planet parameters but adding in one or more moons. The model curve is generated at precisely the same cadence as the data used for the planet-stack and accounts for the long-cadence integration time using 1 minute numerical re-sampling (Kipping, 2010b). After all of the KOI model light curves have been computed for a specific choice of underlying moon population, they are co-added with the same weighting used for the real data. In other words, we inflict precisely the same transformations to the model as we do to the data, to ensure a like-for-like comparison at the end.

Galilean Analogs

The moons are generated in two ways. The first case was for a Galilean analog. Here, we assume that four moons orbit each KOI with properties resembling those of Io, Europa, Ganymede & Callisto. To inject some stochastic variation between each moon, yet maintain the 1:2:4 resonance of the inner three, we randomly place each Io-analog to have a semi-major axis of $(a_{SP}/R_P) \sim \mathcal{U}[0.8 \times 6.1, 1.2 \times 6.1]$, where 6.1 is the actual value for Io around Jupiter and $\mathcal{U}[a, b]$ is a uniform distribution from $a \rightarrow b$. The next moon along is then assumed to lie in a 2:1 resonance, such that $(a_{SP}/R_P)_{\rm Europa} = 2^{2/3}(a_{SP}/R_P)_{\rm Io}$, and similarly for the Ganymede-analog with respect to the Europa-analog. The semi-major axis of the Callisto-analog, which does not reside in the resonance chain, is then randomly drawn as $(a_{SP}/R_P)_{\rm Callisto} \sim \mathcal{U}[3 + a_{SP}/R_P)_{\rm Ganymede}, 1.2 \times 27.2]$, where 27.2 is the actual value for Callisto around Jupiter. Any moon systems generated where two moons have semi-major axes within 3 planetary radii of each other are rejected.

To generate moon radii which were stochastic yet representative, we adopted the radius-separation power-law model of Kane et al. (2013). The authors note that the radii of moons tend to increase with respect to semi-major axis following a power-law model. We took the four Galilean moons in isolation and calibrated a least squares power-law model to it, giving $\log(R/R_\oplus) \sim \mathcal{N}(6.95 + 0.27\log(a_{SP}/R_P), 0.17)$, where the standard deviation quoted is that resulting from the residuals of the best-fitting line. To protect against peculiar draws, we required that the quadrature sum of the radii was within 20% of the actual sum for the Galilean system, that the minimum radius moon was at least 80% the radius of Io and the maximum radius moon was no more than 120% the radius of Ganymede. The resulting covariant distribution is illustrated in Figure 2.6.

To go from physical radii, R_S, to R_S/R_P needed for the modeling (or equivalently $R_S/R_\star = (R_S/R_P)(R_P/R_\star)$), we used planetary radii derived in Chen & Kipping (2017b) to make the conversion. We also included the slight TTV and TDV effects induced by the Galilean moons by including (M_S/M_P), which was computed by using `forecaster` (Chen & Kipping, 2017a)

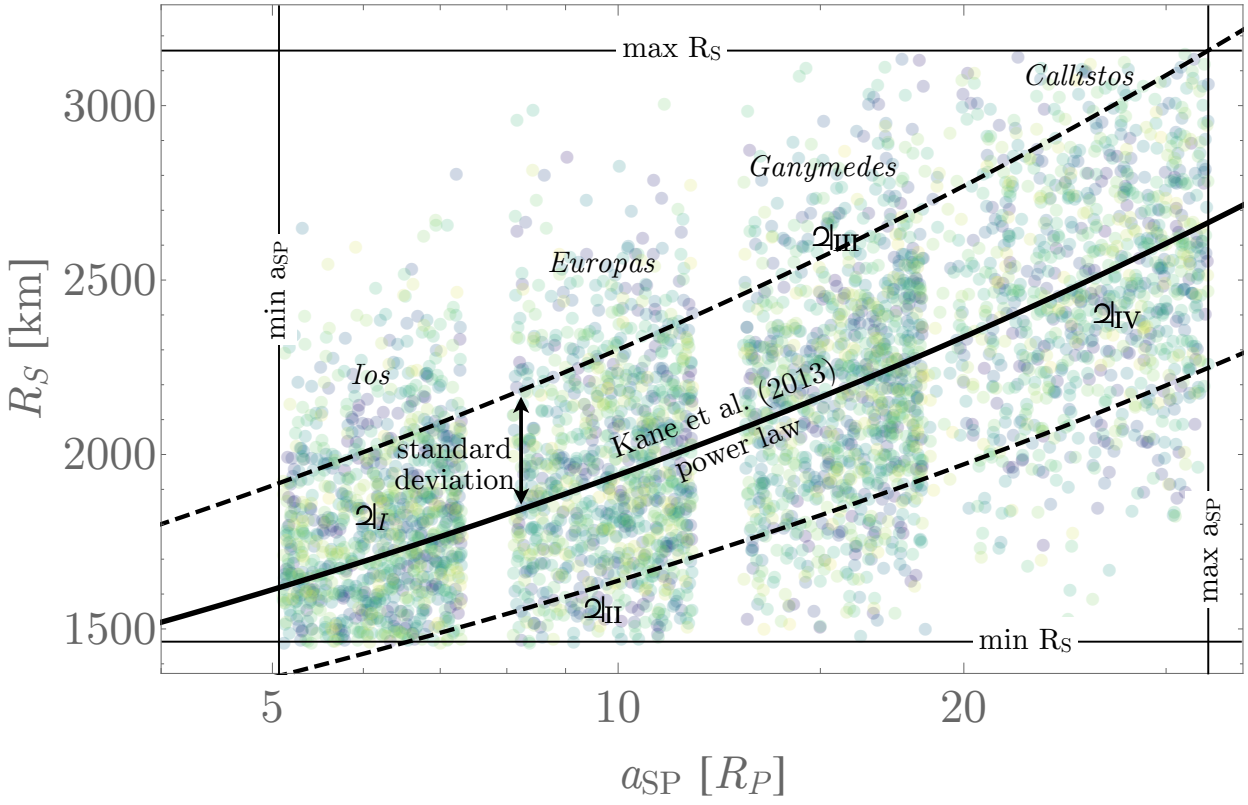

Figure 2.6: One thousand randomly generated Galilean-moon system analogs using the method described in the main text. Each color represents a unique system comprising of four moons.

applied to the moon radius to predict a mass, and then using the physical planetary masses predicted in Chen & Kipping (2017b).

The mean longitude of the inner two moons are randomly generated uniformly but the third is enforced to satisfy a Laplacian resonance. The inclination of the moons are randomly drawn from a Von Mises distribution with $\kappa = 42637$, which we found to maximize the likelihood of a Von Mises distribution conditioned on the real Galilean moons. Eccentricities were kept fixed at zero and the moons all follow perfect keplerian orbits i.e. we do not model gravitational interactions between the moons.

After stacking the resulting model light curves with correct cadence and weightings, we tried varying the fraction of systems which harbor moons. Treating each system as a Bernoulli experiment with a probability of having a moon system given by η, we found varying η was equivalent to simply scaling the $\eta = 1$ resulting light curve by the same factor. Having demonstrated this, we were able to exploit it to aid in later fits.

In cases where a subset of systems were modeled, the process described above was repeated creating unique model light curves for each specific subset. Generally, the shape of each resulting light curve were very similar, but ended up with different amplitudes as a result of the differing weights, data gaps and stellar radii for the host stars. As noted these factors represent data specific properties for each subset and were saved for later use with the single moon simulations.

Single Moons

While a single moon is four times quicker to generate than four Galilean moons, the **Galilean moons** follow an expected distribution in terms of their sizes and orbital semi-major axes. For a single hypothetical moon, we have no idea what these properties are *a priori* and thus our grid cannot simply span η, as before, but now must also span R_S and a_{SP} leading to a three-dimensional look-up table. Fortunately, the effect of η is a simple scaling and thus can be applied easily during the fits themselves, yet this still means we need to

generate a two-dimensional grid of models, rather than just a single look-up example in the case of the Galilean-analogs.

We setup a logarithmically-spaced grid from $R_S = 0.2\,R_\oplus$ to $R_S = 2.0\,R_\oplus$, with 16 unique grid points. For a_{SP}, our grid is again logarithmic, defined as $a_{SP} = 2^x$ where x is stepped through from 1 to 7 in 0.1 steps, leading to a total grid size of 976 elements. The moon is treated as being exactly coplanar and circular with random phase and as before we generate unique light curves for each KOI and then co-add with the appropriate weightings to create our final models. For these simulations we set $(M_S/M_P) = 0$, since some simulations permit very massive moons which would cause noticeable TTVs, which would then be subsequently removed anyway by our data processing pipeline described earlier.

When dealing with subsets, we apply the scaling factors found earlier with the Galilean-analog experiments, since the computation time to create the grid required many weeks. During the actual fits, we used a bi-linear interpolation of every unique binned photometric data point, conditioned upon the LUT. We also added an extra grid point of $R_S = 0.0\,R_\oplus$, corresponding to a flat line, to provide numerical stability if fits attempted to compute the likelihood of radii below our $R_S = 0.2\,R_\oplus$ limit.

2.7 Analysis

2.7.1 Galilean Global Fits

We first discuss our results from regressing our photodyanmical phase-folded planet-stacked planet+moon light curve models (see Section 2.6) to all 284 KOIs deemed to be of suitable quality for our analysis (see Section 2.3). As discussed, the fits are conducted for two different light curve models, a Galilean-analog and a single moon.

For the Galilean-analog, the only parameter directly affecting the light curve model is η, the fraction of KOIs which harbor a Galilean-analog. In addition, we added two other free parameters into our fits. The first was to account for excess photometric scatter, σ, and was simply added in quadrature to our derived uncertainties in the planet-stacked light curve.

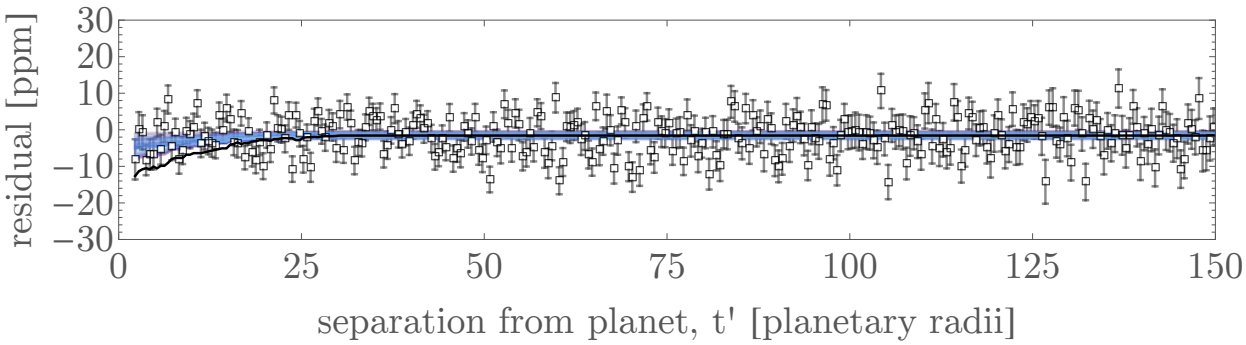

Figure 2.7: Phase-folded planet-stacked light curve of all 284 KOIs deemed to be of acceptable quality. Temporal axis has been re-scaled and binned, with uncertainties shown given by the standard deviations within each bin. Black solid line represents the expected signature if $\eta = 100\%$ of the KOIs had a Galilean-analog moon system. Blue lines show 100 posterior samples from our fits, giving $\eta = 0.16^{+0.13}_{-0.10}$.

The second was an offset term, γ, to allow for a re-normalization of the data set. While η and γ were assigned uniform priors, σ was assumed to follow a log-uniform prior from 0.1 to 10 ppm. The regression was performed using MultiNest with 1000 live points. The fits were repeated ten times each, from which the posteriors were combined. Since MultiNest estimates the marginal likelihood, we repeated our fits with η fixed to zero and removed as a free parameter, giving us a direct estimate of the Bayes factor, B_{SP}, for the moon model.

We find that the null model is slightly favored, with $\log B_{SP} = -0.84$, or an odds ratio of 2.3-to-1 preference for the null model. The resulting light curve and model fitting lines are shown in Figure 2.7, and the associated posterior distribution is plotted in Figure 2.8. Our results imply that $\eta < 0.38$ to 95% confidence for the 284 KOIs considered, with a 68.3% confidence interval of $\eta = 0.16^{+0.13}_{-0.10}$.

2.7.2 Single Moon Global Fits

As discussed in Section 2.6.2, the single moon case required on-the-fly interpolation of a look-up table for the likelihood calls. For this reason, we found it more practical to use a Markov Chain Monte Carlo (MCMC) algorithm instead of MultiNest. Our regression was performed using a simple MCMC that we wrote, which used the Metropolis rule for

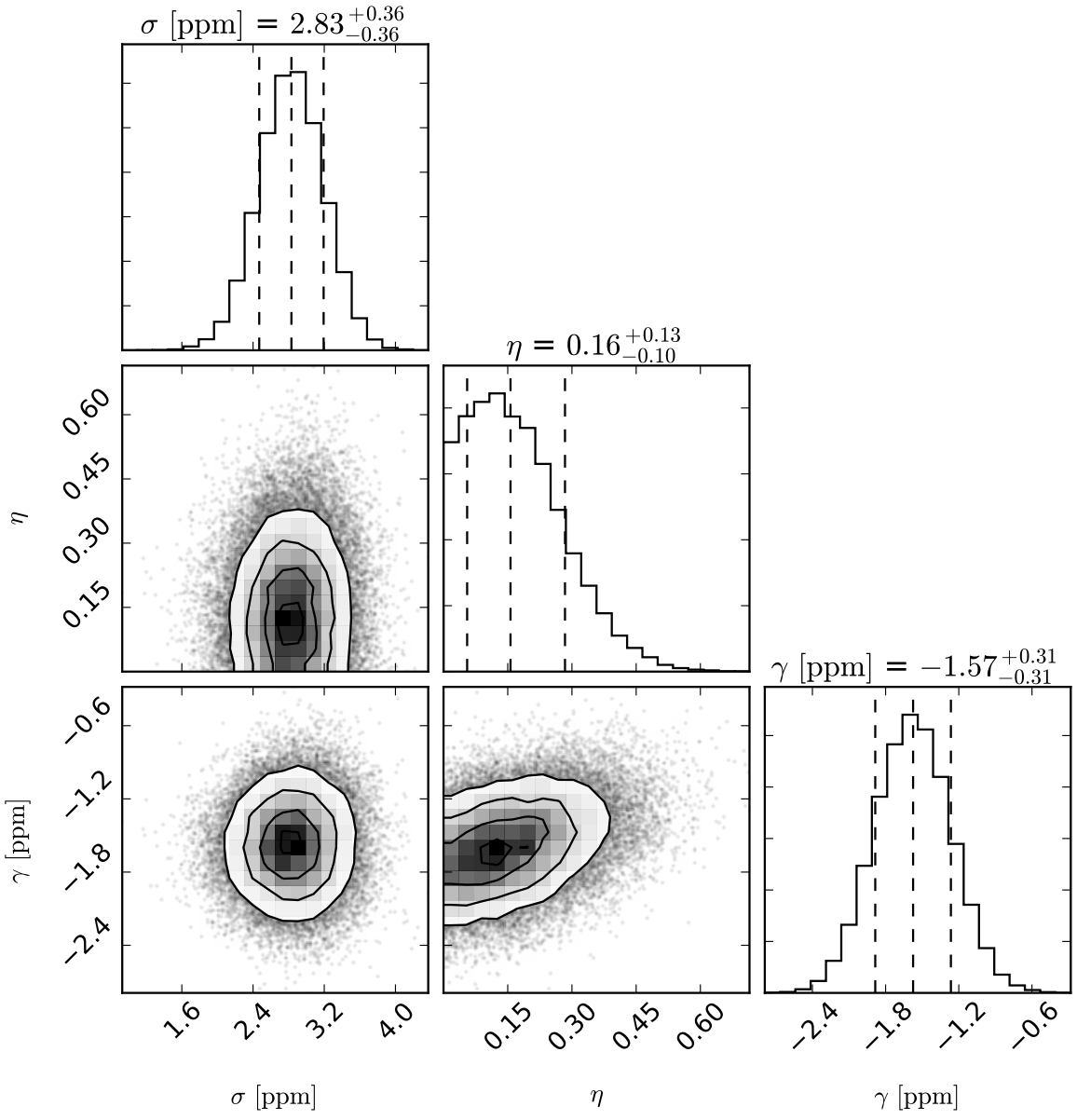

Figure 2.8: Corner plot of the three-parameter joint posterior distribution from our Galilean-analog moon fit. This fit was disfavored over a null fit with a Bayes factor of $B_{SP} = 0.43$.

sampling and normal proposal functions tuned by hand to give a ~50% acceptance rate. Ten independent chains were seeded from random locations within the prior volume, all of which converged within 100 steps, and were then allowed to propagate for 10^5 accepted steps, giving $\sim 10^6$ posterior samples in total.

In addition to the three free parameters used in the Galilean fit, we added in a parameter controlling the semi-major axis of the moon, (a_{SP}/R_P), and the moon size in Earth radii, $(R_S/R_\oplus)$. Both were assumed to follow log-uniform priors spanning the limits in our LUT. The η term was allowed to follow a log-uniform prior spanning 0.01 to 100, since technically it represents an effective moon in this case and thus should be interpreted as the average number of moons per system.

The fits converged to a solution of $R_S = 0.51^{+0.59}_{-0.23} R_\oplus$ and $(a_{SP}/R_P) = 6.3^{+7.6}_{-3.1}$ for $\eta = 0.43^{+0.33}_{-0.28}$. This fit does not directly return a marginal likelihood, since MCMC was used, nor is the Savage-Dickey ratio suitable given that three extra covariant free parameters have been added. However, the kernel-approach shown later reveals that the evidence favoring the single moon fit in this region is modest at $B_{SP} \simeq 2$.

2.7.3 Single Moon Kernel

While the single moon fit is useful for identifying the maximum *a posteriori* region of parameter space, it does not provide a clear view of the overall likelihood trends occurring within the prior volume. To address this, we repeated our fits for a single moon but fixed a_{SP}/R_P and R_S to a specific choice and just regressed η, σ and γ. Since no interpolation was necessary, it was straightforward to use MultiNest with η-rescaling on a single interpolated model each time. R_S was varied across a grid from 0.02 to 2 Earth radii in 100 log-evenly spaced steps. Similarly, a_{SP}/R_P was varied from 2 to 100 in 100 log-even steps.

At each point, we derived a three-dimensional joint posterior distribution and marginalize over σ and γ to directly measure the occurrence rate of exomoons at each location. The posterior derived is mathematically equivalent to

$$\mathrm{P}(\eta|R_S, a_{SP}/R_P) = \int\int \mathrm{P}(\eta, \sigma, \gamma|R_S, a_{SP}/R_P)\mathrm{P}(\sigma)\mathrm{P}(\gamma)\mathrm{d}\sigma\mathrm{d}\gamma \tag{2.9}$$

In addition to deriving a posterior at each grid point, we also estimate the evidence against the null model, allowing us to compute the Bayes factor. In Figure 2.9 we plot the Bayes factor and exomoon frequency for the ensemble as a function of effective moon radius R_S and semi-major axis a_{SP}. The Bayes factor (left) indicates whether the moon model is favored over the model without a moon. Red represents regions of parameter space where the moon model is disfavored, while green regions are areas where the moon model is favored, and intensity represents our degree of confidence in that model selection. We emphasize to the reader that paying attention to the contours in this plot is essential for an accurate interpretation; while much of the plot appears green, the moon model is in fact only weakly favored ($B_{SP} \simeq 2$) on a small island in parameter space. By contrast, large values of R_S and a_{SP} are strongly ruled out (B_{SP} around 0.01). A value of 1 in this plot means we can make no statement about one model being a better fit to the data than the other.

The right side of Figure 2.9 should then be read in the context of the left side. The similarity between the contours on both sides is readily apparent. For large values of R_S and a_{SP} we find an exceptionally low occurrence rate, while at the lower end of these variables the occurrence rate shoots up. It is tempting to read this as a moon signal, but in the context of the Bayes factor on left it is clear these occurrence rates are not at all constrained or well supported by the evidence. Only in regions of high confidence (B_{SP} much greater or much less than 1) should the exomoon frequency values be given much credence. It is worth pointing out, perhaps, that there is little qualitative difference between a very low exomoon occurrence rate and a very low value for B_{SP}. Both are consistent with virtually no signal in this region of parameter space.

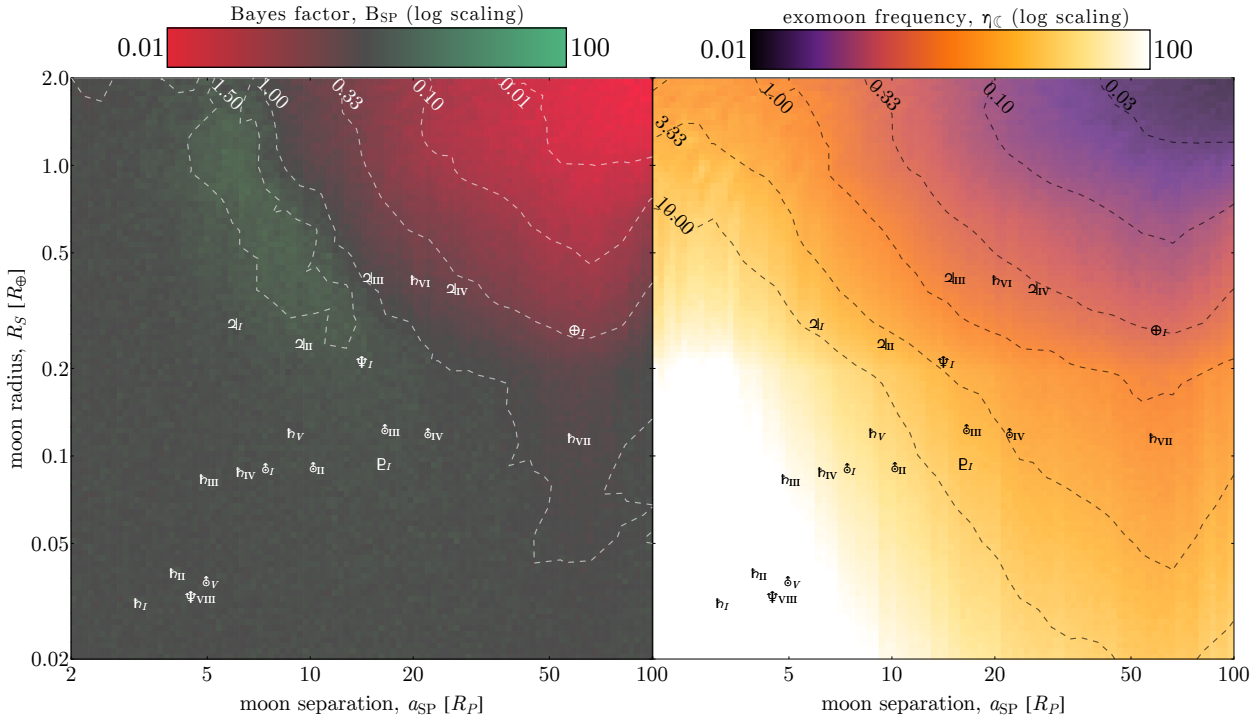

Figure 2.9: *Left:* Heatmap of the Bayes factor B_{SP} as a function of single effective moon radius R_S and semi-major axis a_{SP} for the ensemble. Red indicates regions of parameter space where the moon model is disfavored, while green represents regions where the moon model is favored. Greater color intensity corresponds to greater confidence in the selected model. *Right:* Exomoon frequency in the ensemble as a function of R_S and a_{SP}. A collection of solar system moons are plotted for context.

2.7.4 Evidence for a Population of Super-Ios?

The island on the left side of Figure 2.9 where $B_{SP} > 1.50$ is intriguing, if only marginally significant. We hesitate to make any strong statement about this region of parameter space, but it is worth pointing out that theoretical modeling (e.g. Namouni, 2010) suggests that while planets migrating inward will tend to lose their moons in the process due to a shrinking Hill sphere, they are more likely to retain moons orbiting closer to the host planet (that is, close-in moons tend to survive planetary migration to smaller a_P). Recall that we are probing planets within roughly 1 AU of their host star, suggesting that a large fraction of these planets may have migrated from beyond the snow line. The island of modest moon signal could therefore be evidence (albeit inconclusive) of a population of moons that have survived migration by virtue of their separation from their host planet. Note, however, that more recent theoretical work (Spalding et al., 2016) suggests by contrast that satellites closer in to their planet ($a_{SP} \lesssim 10R_P$) are also vulnerable to dynamical moon-loss during migration. It is unclear, then, whether one or both of these mechanisms could be at play here, and indeed, the strength of these mechanisms rely in part of the evolutionary history of each system which will of course be unique. In any case, our results suggesting a dearth of exomoons at small a_P appear to provide observational support for the findings of both Namouni (2010) and Spalding et al. (2016), and more broadly, could be evidence of giant planet migration.

2.7.5 Subset Fits

In addition to fitting the entire sample with effective moon and Galilean analog GLC models, we also performed GLC fits on a number of physically motivated subsets. The aim here was to identify whether a certain class of planets in the sample preferentially hosts moons over another. As such we divided the sample into several equally-populated pairs and fit the GLC model to these subsets. These pairs were small/large planets, cold/hot planets, cold/hot stars, and inner/outer planets. We also split the sample into single/multi-planet

Galilean Analog Subset Fits							
Group	$\eta_{\leftmoon}$ $[1\,\sigma]$	$\eta_{\leftmoon}^{\max}$ $[2\,\sigma]$	B_{SP}	Group	$\eta_{\leftmoon}$ $[1\,\sigma]$	$\eta_{\leftmoon}^{\max}$ $[2\,\sigma]$	B_{SP}
Smaller Planets	[0.06, 0.35]	0.48	0.353 ± 0.035	Larger Planets	[0.07, 0.44]	0.61	0.448 ± 0.035
Colder Planets	[0.15, 0.68]	0.86	1.003 ± 0.070	Hotter Planets	[0.07, 0.39]	0.53	0.426 ± 0.043
Colder Stars	[0.07, 0.33]	0.44	0.639 ± 0.063	Hotter Stars	[0.06, 0.47]	0.66	0.411 ± 0.034
Inner Planets	[0.21, 0.64]	0.80	2.564 ± 0.250	Outer Planets	[0.03, 0.28]	0.42	0.224 ± 0.018
Single-Planet Systems	[0.08, 0.50]	0.68	0.689 ± 0.066	Multi-Planet Systems	[0.07, 0.33]	0.44	0.420 ± 0.038
Habitable Zone	[0.23, 0.88]	0.97	1.679 ± 0.083	Non-habitable Zone	[0.08, 0.37]	0.50	0.653 ± 0.066

Table 2.1: Table of occurrence rates η for various subsets of the 284 planets examined in this work. Here η represents 1σ credible interval values from the posterior distributions while 95 pct is the 95$^{\text{th}}$ percentile upper limit. B_{SP} is the Bayesian evidence computed by MultiNest.

systems, and habitable zone/non-habitable zone planets (most of the latter residing inside the innermost edge of the habitable zone). These last two categories, of course, are not equally populated. Insolations were taken from the NASA Exoplanet Archive and anything less than the maximum (inner-edge) insolation given in Yang et al. (2014) equation 2 was considered to be in the habitable zone.

The results for Galilean moon fits can be seen in Figure 2.10 and Table 2.1. The thick black line in the figure represents the peak posterior value, while the light red lines represent 50 fair draws from the posterior. From these plots we can make a number of comparisons. Dividing the sample in two by size, we see a marginally higher occurrence rate for the larger planets. For planet temperature we see a higher occurrence rate on the colder end. We also see a higher exomoon occurrence rate for colder stars, which we can take to mean later-type or evolved stars. All of these observations are in line with what we might expect *a priori*.

We see very little difference in the occurrence rate for single- and multi-planet systems. Interestingly, though, the inner 50% of planets (those closest to their star) show a significantly higher exomoon occurrence rate than those farther away, where the maximum *a posteriori* value indicates a total non-detection. This is unexpected, since the Hill sphere shrinks with smaller semi-major axis. Finally, and perhaps most intriguingly, the maximum *a posteriori* values for habitable zone planets ran away to the maximum, while non-habitable zone planets have a much lower occurrence rate. This should be read with caution, however, considering the size of the error bars in the habitable zone planet case. While the comparisons above

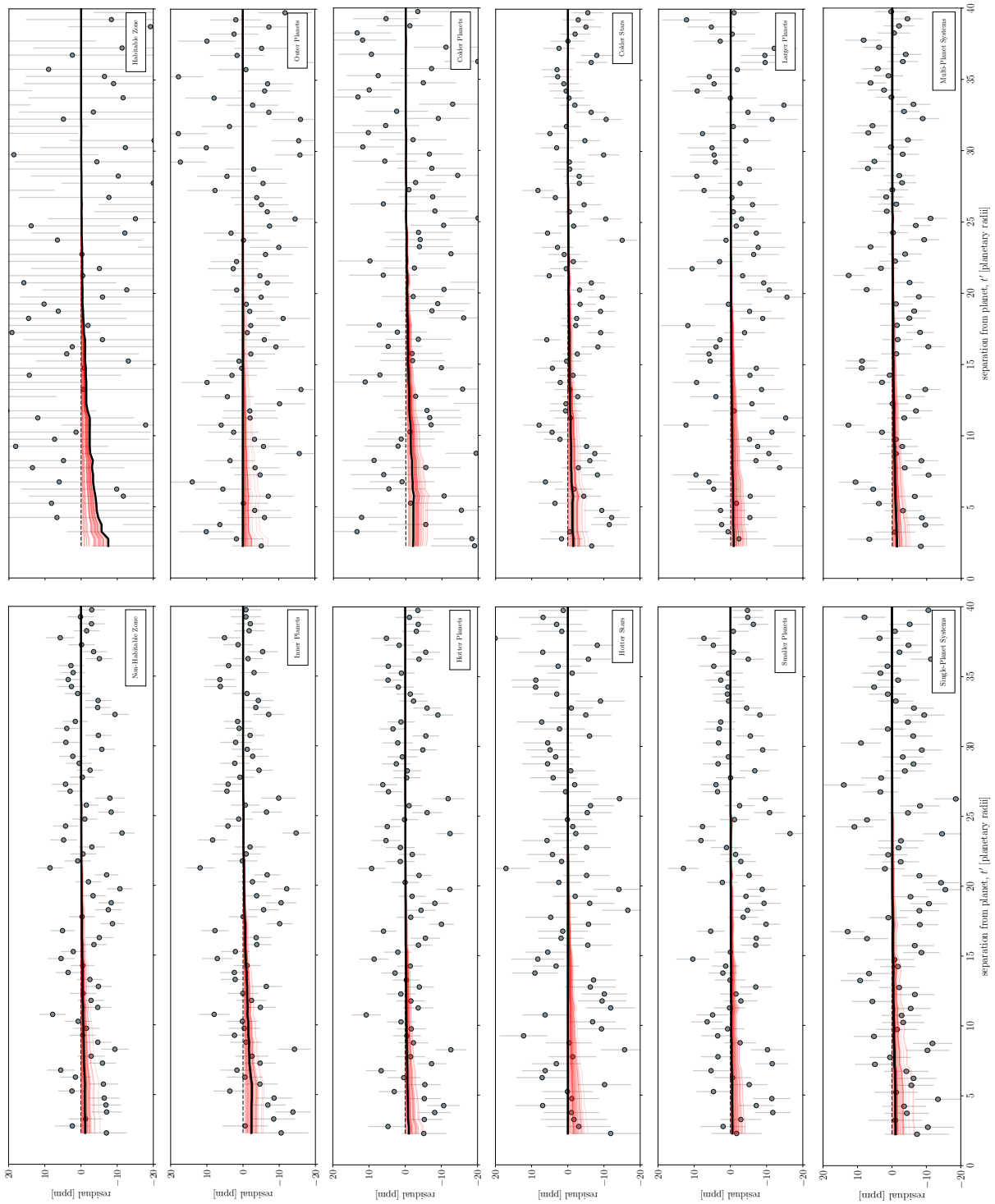

Figure 2.10: Galilean analog GLC plots for a variety of sample subsets.

Single Effective Moon Subset Fits							
Group	$R_S\,[1\,\sigma]$	$R_S^{\max}\,[2\,\sigma]$	B_{SP}	Group	$R_S\,[1\,\sigma]$	$R_S^{\max}\,[2\,\sigma]$	B_{SP}
Smaller Planets	[0.02, 0.36]	0.90	0.75	Larger Planets	[0.02, 0.41]	1.18	0.76
Colder Planets	[0.02, 0.41]	1.12	0.81	Hotter Planets	[0.02, 0.42]	1.03	0.79
Colder Stars	[0.02, 0.35]	0.83	0.79	Hotter Stars	[0.02, 0.36]	1.07	0.74
Inner Planets	[0.03, 0.81]	1.40	1.04	Outer Planets	[0.02, 0.29]	0.83	0.69
Single-Planet Systems	[0.02, 0.42]	1.05	0.79	Multi-Planet Systems	[0.02, 0.34]	0.96	0.72
Habitable Zone	[0.03, 1.10]	1.66	1.12	Non-habitable Zone	[0.02, 0.33]	0.92	0.74

Table 2.2: Table of effective moon sizes R_S for various subsets of the 284 planets examined in this work, in units of Earth radii. We present 1σ credible interval values from the posterior distributions while 95 pct is the 95^{th} percentile upper limit. Here B_{SP} is the Savage-Dickey ratio computed from the $\log(R_S)$ posteriors.

are made between equally populated subsets, there are far more non-habitable zone planets than there are planets in the habitable zone, making the results in the latter case far more uncertain.

Single, effective moon fits were also performed for these same OSE subsets. Results can be found in Table 2.2. Unlike the case for Galilean analog fits, we cannot meaningfully quote an occurrence rate in this case because the depth of the moon signal is controlled by the size of the effective moon. There is a degeneracy between moon size and occurrence rate, so we model the size of the moon as a proxy for occurrence rate. In essence, a smaller effective moon can mean either a) we have a high occurrence rate, with small moons, or b) we have a low occurrence rate of larger moons. The truth, of course, is probably somewhere in the middle.

To characterize whether our fits exclude a null detection to high confidence we compute the Savage-Dickey ratio between a uniform prior and the posterior distribution for $R_S^2 = 0$. We treat this number as our Bayes factor. For the colder stars we find $B_{SP} = 1.53$, for outer planets $B_{SP} = 0.7$, large planets $B_{SP} = 1.29$, multi-planet systems $B_{SP} = 0.77$, and habitable zone planets $B_{SP} = 3.11$. These values suggest that there is only marginal evidence in support of an effective moon signal in the cold star, large planet, and habitable zone planet subsets, while the null hypothesis is favored for the outer planet and multi-planet subsets.

2.8 Exomoon Candidate Kepler-1625b I

2.8.1 Individual Fits

Thus far, our analysis has taken a population-based approach to seeking exomoons, in contrast to the conventional method adopted in previous HEK papers where candidates were interrogated individually. Although a full suite of photodynamical Bayesian fits to each planetary candidate is beyond the scope of this work, representing a formidable computational challenge, we do here investigate systems individually with a simpler model.

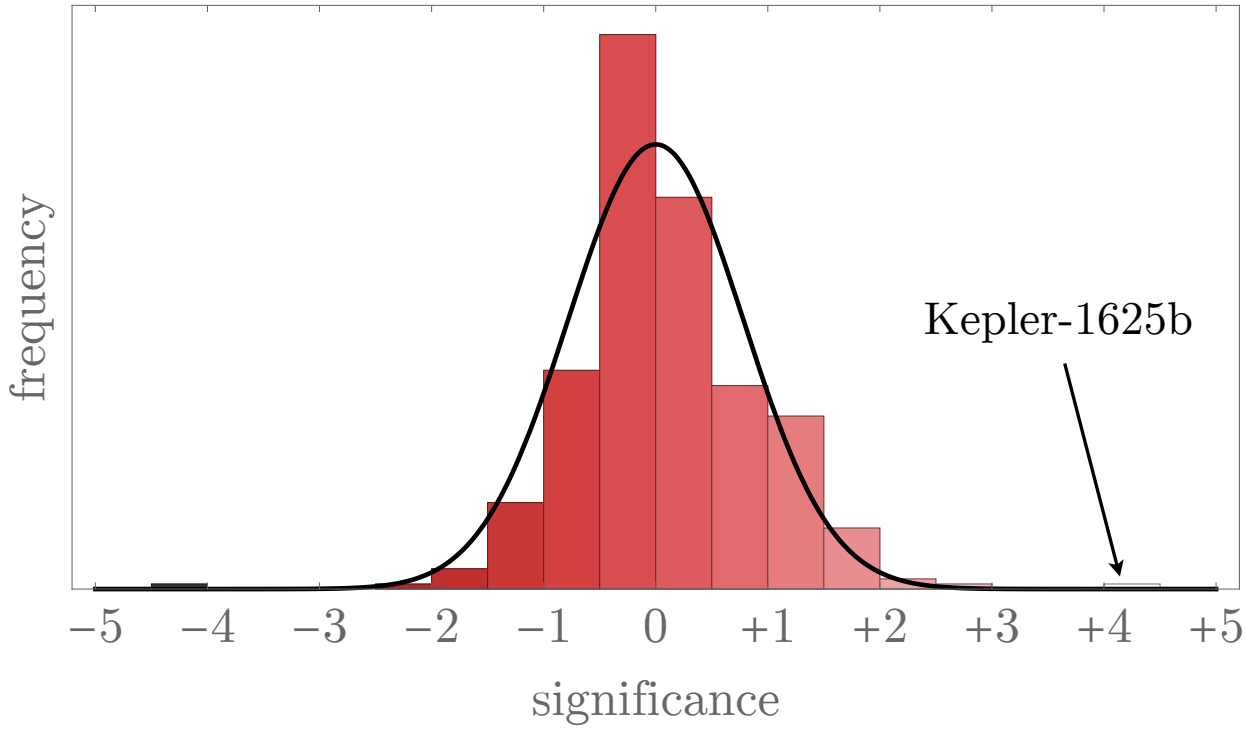

Figure 2.11: Histogram of the "significance" of an OSE detection for several hundred KOIs, the test which revealed the presence of a possible candidate around Kepler-1625b. The vertical axis scale is linear.

For each KOI, we took the final phase-folded light curve and used MultiNest to fit the Heller (2014) OSE model through the data. As discussed earlier, OSE does have drawbacks, yet it remains a useful and quick tool for checking for any significant flux decreases surrounding the phase-folded transits. For each KOI, we fit for an offset term, a photometric jitter,

the moon size, and a semi-major axis, (a_{SP}/R_P) with MULTINEST using 1000 live points. In the earlier ensemble fits, we used a log-uniform prior on the moon-size (see Section 2.7.1), which means negative moon radii cannot be sampled. In these fits, we wished to allow for negative radius moon, which correspond to inverted transit signals, to provide insights into possible biases affecting the results. Accordingly, we modified our moon parameter to be the satellite-to-star ratio-of-radii, s, squared, to account for the fact that transit detection bias is approximately proportional to s^2.

We evaluated the median s^2 value for each KOI from the posterior chain and divided by the lower quantile bounding a 68.3% confidence interval, a metric which we loosely refer to as "significance" in what follows (although it is best interpreted as a power). In Figure 2.11, we show a histogram of the resulting significances for 353 KOIs, where we have also included the likely eccentric KOIs to provide a wider sample to assess distribution properties. Inspection of the results reveals a sizable spread centered around zero, as might be expected. However, we note that KOI-5084.01 (Kepler-1625b) appears quite deviant from the bulk population with a $+4.4\,\sigma$ significance. Similarly, KOI-4202.01 (Kepler-1567b) is an outlier on the negative scale with a $-4.4\,\sigma$ significance.

The negative outlier clearly cannot be a genuine exomoon and thus we do not consider it further it in what follows. The positive outlier though could be an isolated candidate missed by the ensemble analyses. Excluding the two outliers, the histogram shown in Figure 2.11 is well-described by a normal distribution centered on zero with a standard deviation of 0.8. If we generate a list of 353 random variates from that distribution and take the maximum value, +4.4 is highly improbable; we used Monte Carlo simulations to estimate the probability, which was found to be 4×10^{-6}. On this basis, Kepler-1625b appears quite unexpected and thus worthy of more detailed investigation.

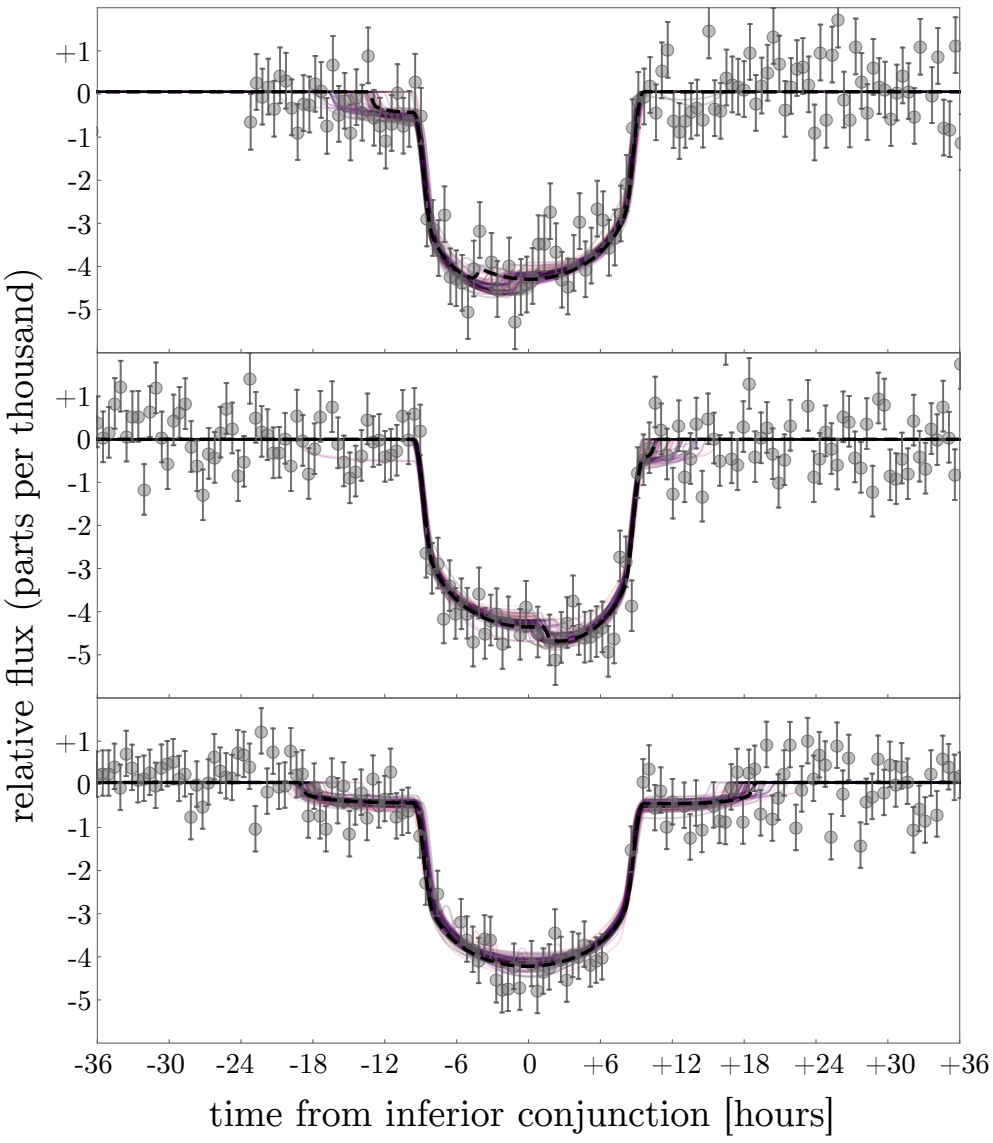

Figure 2.12: The three transits of Kepler-1625b observed with *Kepler*, overlaid with 100 draws from the model posteriors. The black line is the maximum *a posteriori* model.

2.8.2 Detailed Investigation of Kepler-1625b

To investigate further, we performed detailed photodynamical fits of Kepler-1625b using the `LUNA` model and MultiNest, in the same manner as that conducted in our previous series of HEK papers (e.g. see Kipping et al. 2013b). This enables a rigorous Bayesian model selection to ensure not only a physically plausible model can explain the photometry, but that the moon parameters are justified given the extra complexity they introduce.

Comparing the evidences between a planet-only and planet+moon model revealed that the moon model was favored with $\log B_{SP} = 10.2$, or a $4.1\,\sigma$ preference, consistent with the level found previously. The light curve fits are illustrated in Figure 2.12.

Given the limited numbed of transits available, just three, the conventional HEK approach of cross-validating by removing a transit was not tenable and thus we adapted this strategy somewhat in what follows. Instead, we performed k-fold cross-validations, where we omit one-half of a transit (centered about the time of inferior conjunction) and re-fit the remaining data blindly each time. Since there are three transits in the *Kepler* data, this yields six unique ways of conducting the cross-validation. In each case, we performed a new blind fit and found a positive detection consistent with the original signal in all cases, with Bayes factors indicating 3.9, 2.4, 4.6, 4.0, 3.1, and 2.7 σ preference for the moon model (comparable to the original $4.1\,\sigma$ detection when using all of the data).

As with all previous moon candidates we attempted to rule out all other possible explanations for the signal. For any candidate moon signal there will be one of two possible explanations: the signal is either an instrumental artifact or it has a true astrophysical origin (be it an exomoon or something else). To test the possibility of an instrumental aberration we performed an independent and manual detrending on the *Kepler* DR25 data (our OSE survey used DR24) using `CoFiAM` on the PA and PDC data. In both cases the planet-moon model remained the favored hypothesis over the planet-only model. In contrast, polynomial-based detrending was found to remove the signal, likely due to very long-timescale nature of the driving event occurring in epoch three.

We also examined data from individual *Kepler* pixels to determine whether there might have been anomalous behavior in the vicinity of the transit events (as occurred for false-positive Kepler-90g.01; Kipping et al. 2015b), but no unusual spikes or drop outs were detected in any relevant pixel. We also verified that there were no bad data flags at the time of the transits, which was a source of the false-positive result for the moon candidate around PH2-b (Kipping et al., 2015).

If the signal were astrophysical in origin, then there are several possible hypotheses, including a moon. A ring is not likely since it should produce a coherent signal in all events, which is not seen, unless the ring precession rate is very fast. Rotating spots on the surface of the star also affect the light curve, primarily producing a long-term undulation in the data. This long-term trend is removed via our detrending (repeated several times independently) and thus the only remaining possible starspot-induced signal would be crossing events. However, starspot crossings (when a transiting planet occults a dark spot; e.g. see Rabus et al. 2009) cannot be responsible for the observed moon-like dips. This is because spot crossings can only occur inside the main planetary transit and can never produce out-of-transit flux decreases, as seen for Kepler-1625b, purely from a geometrical argument. If the signal were confirmed then, this would leave the exomoon hypothesis as the leading explanation based on current information.

The quoted stellar properties of Kepler-1625 in the NASA Exoplanet Archive changed significantly from DR24 to DR25, owing to the addition of updated information in the latter data release (Mathur et al., 2017). This update pushed the star from a sub-Solar to a super-Solar radius ($R_\star = 0.838^{+0.366}_{-0.079} \rightarrow 1.793^{+0.263}_{-0.488}\, R_\odot$), enhanced the metallicity from sub- to super-Solar abundances, and lowered the density substantially ($\rho_\star = 2.059^{+0.4626}_{-1.306} \rightarrow 0.2636^{+0.3257}_{-0.0768}$ g/cm^3), indicating that this star is likely climbing the giant branch. Critically, our planet-only and planet+moon fits favor a low stellar density of $\rho_\star = 0.387^{+0.034}_{-0.083}$ g/cm^3 and $\rho_\star = 0.405^{+0.028}_{-0.054}$ g/cm^3 and if the true density were much higher, then Kepler-1625b would need to be either highly eccentric or blended (Kipping, 2014), both of which would be

severely detrimental to the exomoon hypothesis. Determining the true nature of this star is critical as it will also dictate the sizes of the planet and moon derived from the transit depth (which we describe shortly).

We also attempted to recover a rotation period for the star (following the methodology described in Torres et al. 2015) but the amplitude of variability appears too small to recover a consistent period across each quarter, with best-fitting periods ranging from 4.5 days to 21 days. Attempting to regress a coherent signal across all quarters gives an amplitude of 66 ppm, and when performed on each quarter independently, the median amplitude was 136 ppm. Given the lack of strong evidence for rotation, the weak amplitudes in comparison to the candidate moon transit depth (570 ppm), and the arguments made earlier as to why rotational modulations are unlikely to be a source of false-positive, we deem it unlikely that activity is responsible for the signal observed.

Our photodynamical fits combined with the DR25 stellar properties indicate that Kepler-1625b is likely a Jupiter-sized planet with approximately ten times Jupiter's mass, orbited by a moon roughly the size of Neptune. We calculate the radii of the planet and moon by measuring the depth of the flux dip ($\Delta F/F = (R_o/R_*)^2$, where R_o is the radius of the object in question) and we are able to derive a mass based on the photodynamical model fit. We note that both the planet and the moon show good agreement between mass and radius estimates and lie in physically reasonable parameter space based on the mass-radius forecaster model of Chen & Kipping (2017a). We find the semi-major axis of the moon $a_S = 19.1^{+2.1}_{-1.9} R_P$, which is well outside the Roche limit and comfortably within the Hill sphere for this planet. It is dynamically stable and should not have spun out / escaped over 5 Gyr (Barnes & O'Brien, 2002).

While the existence of a Neptune-sized moon has largely not been anticipated in the literature (however, see Cabrera & Schneider 2007), we cannot readily rule out its existence on these grounds. Indeed, the existence of Hot-Jupiters was also wholly unexpected prior to their discovery in the mid-1990s. It seems clear that a moon of this type could not have

formed in a circumplanetary accretion disk akin to that which is thought to have formed the regular moons of Jupiter and Saturn. It is conceivable however that the moon could have been captured by the planet, a kind of intermediate process between typical capture scenarios (e.g. Neptune and Triton; see Agnor & Hamilton 2006) and the cataclysmic impact event that is believed to have formed Earth's Moon (Ćuk & Stewart, 2012). In this scenario a grazing impact might be experienced as a kind of extreme atmospheric drag sufficient to capture the passing body. Observation of this system might therefore not only produce the first unambiguous detection of an exomoon, but could also go a long way in demonstrating once again that what we observe in our Solar System is not all that is possible.

2.8.3 Validating the Exomoon Candidate Kepler-1625b I

At this time, we remain cautious about the reality of this signal, given the relatively small number of transits available. This is particularly true because the third transit appears to be crucial to the exomoon interpretation and can be removed using polynomial-based detrending approaches. Detrending the photometric time series of long-period transits is more challenging than their shorter-period counterparts and it remains wholly plausible that the signal observed is nothing more than an artifact of our detrending process. We strongly emphasize these points and encourage the community to not treat this signal as genuine until it can be confirmed.

Fortunately, our photodynamic moon fit yields a testable prediction for the morphology of the next transit event occurring October 2017. With such a long event duration from anticipated exomoon ingress to egress, the event cannot be observed in its entirety by any single optical/NIR instrument on the ground south of latitude $\sim 78^\circ$ N (north of which the long Arctic night has already begun on the date of observation). A space-based observation is clearly essential to characterize the system. We have therefore secured HST observations to validate the signal during the next transit of the planet, and we strongly advocate treating this object as no more than a candidate at this time, similar to previous moon candidates

discussed in earlier HEK papers.

2.9 Conclusion

In this work we have examined 284 *Kepler* exoplanets (from an original sample of 4098 KOIs) in search of an exomoon signal in the ensemble. We performed a rigorous multi-stage analysis to select only the highest quality data, measure and correct for TTVs, and stack a total of 6096 transit events to characterize the exomoon population. As a byproduct of our work we present new TTV posterior distributions, along with a handful of stellar properties, and make them available online to the community.

Our results place new upper limits on the exomoon population for planets orbiting within about 1 AU of their host star, upper limits that are remarkably low. We have also analyzed subsets of the ensemble to test the effect of various data cuts. Our analysis suggests that exomoons may be quite rare around planets at small semi-major axes, a finding that supports theoretical work suggesting moons may be lost as planets migrate inward. On the other hand, if the dearth of exomoons can be read as a reliable indicator of migration, our results suggest a large fraction of the planets in the ensemble have migrated to their present location.

Finally, we have briefly highlighted our identification of an exomoon candidate in the Kepler-1625 system, for which we have secured a follow-up observation with HST. This candidate has passed a thorough preliminary inspection, but we emphasize again our position that the *Kepler* data are insufficient to make a conclusive statement about the existence of this moon. Only after the HST observation is made should any claim about this moon's existence be given much credence.

2.10 Acknowledgments

This paper includes data collected by the Kepler mission. Funding for the Kepler mission is provided by the NASA Science Mission directorate.

Resources supporting this work were provided by the NASA High-End Computing (HEC) Program through the NASA Advanced Supercomputing (NAS) Division at Ames Research Center.

This research has made use of the NASA Exoplanet Archive, which is operated by the California Institute of Technology, under contract with the National Aeronautics and Space Administration under the Exoplanet Exploration Program.

This research has made use of the `corner.py` code by Dan Foreman-Mackey at github.com/dfm/corner.py.

This work made use of the Michael Dodds Computing Facility, for which we are grateful to Michael Dodds, Carl Allegretti, David Van Buren, Anthony Grange, Cameron Lehman, Ivan Longland, Dell Lunceford, Gregor Rothfuss, Matt Salzberg, Richard Sundvall, Graham Symmonds, Kenneth Takigawa, Marion Adam, Dour High Arch, Mike Barrett, Greg Cole, Sheena Dean, Steven Delong, Robert Goodman, Mark Greene, Stephen Kitt, Robert Leyland, Matthias Meier, Roy Mitsuoka, David Nicholson, Nicole Papas, Steven Purcell, Austen Redman, Michael Sheldon, Ronald Sonenthal, Nicholas Steinbrecher, Corbin Sydney, John Vajgrt, Louise Valmoria, Hunter Williams, Troy Winarski and Nigel Wright.

We thank members of the Cool Worlds Lab for helpful conversations in preparing this manuscript. Finally, we thank the anonymous referees for their constructive comments. DMK acknowledges support from NASA grant NNX15AF09G (NASA ADAP Program). AT acknowledges support from the NSF GRFP grant DGE 16-44869.

Chapter 2: Bibliography

Adams, E. R., Ciardi, D. R., Dupree, A. K., Gautier, T. N., III, Kulesa, C. & McCarthy, D., 2012, AJ, 144, 42

Adams, E. R., Dupree, A. K., Kulesa, C. & McCarthy, D., 2013, AJ, 146, 9

Agnor, C. & Hamilton, D. P., 2006, Nature 441, 192

Agol, E., Steffen, J., Sari, R. & Clarkson, W., 2005, MNRAS, 359, 567

Akeson, R. L., Chen, X., Ciardi, D., et al., 2013, PASP, 125, 989

Barnes, J. W. & O'Brien, D. P. 2002, ApJ, 575, 1087

Bennett, D. P., Batista, V., Bond, I A., et al. 2014, ApJ, 785, 155

Cabrera, J., & Schneider, J. 2007, A&A, 464, 1133

Canup, R. M. & Ward, W. R., 2002, ApJ 124, 3404

Canup, R. M. & Ward, W. R., 2006, Nature 441, 834

Carter, J. A., Fabrycky, D. C.; Ragozzine, D. et al., 2011, Science, 331, 562

Chen, J. & Kipping, D. M., 2017a, ApJ, 834, 17

Chen, J. & Kipping, D. M., 2017b, ApJ, submitted (astro-ph:1706.01522)

Christiansen, J. L., Jenkins, J. M., Caldwell, D. A., et al. 2012, PASP, 124, 1279

Ćuk, M. & Stewart, S. T. 2012, Science, 338, 1047

Dawson, R. I. & Johnson, J. A., 2012, ApJ, 756, 122

Dickey, J. O., Bender, P .L., Faller, J. E., et al., 1994, Science, 265, 482

Domingos, R. C., Winter, O. C. & Yokoyama, T., 2006, MNRAS, 373, 1227

Dressing, C. D., Adams, E. R., Dupree, A. K., Kulesa, C. & McCarthy, D., 2014, AJ, 148, 78

Durbin, J. & Watson, G. S., 1950, Biometrika, 37, 409

Everett, M. E., Barclay, T., Ciardi, D. R., Horch, E. P., Howell, S. B., Crepp, J. R., Silva, D. R., 2015, AJ, 149, 55

Gong, Y.-X., Zhou, J.-L., Wie, J.-W. & Wu, X.-M., 2013, ApJL, 769, L14

Gould, A., Gaudi, S. B. & Han, C., 2004, arXiv e-prints:0405217

Feroz, F. & Hobson, M. P., 2008, MNRAS, 384, 449

Feroz, F., Hobson, M. P. & Bridges, M. 2009, MNRAS, 398, 1601

., Hendrix, A., Esposito, L., et al. 2005, AGU Fall Meeting Abstracts,

Hansen, C. J., Esposito, L., Stewart, A. I. F., et al. 2006, Science, 311, 1422

Heller, R., ApJ, 787, 14

Heller, R., Hippke, M. & Jackson, B., 2016, ApJ, 820, 88

Holman, M. J. & Murray, N. W., 2005, Science, 307, 1288

Huber, D., Silva A. V., Matthews, J. M. et al., 2014, ApJS, 211, 2

Hippke, M., 2015, ApJ, 806, 51

Hogg, D. W., Myers, A. D. & Bovy, J., 2010, ApJ, 725, 2166

Holczer, T., Mazeh, T., Nachmani, G., et al., 2016, ApJS, 225, 9

Ida, S., Canup, R. M., & Stewart, G. R. 1997, Nature, 389, 353

Kane, S. R., Hinkel, N. R. & Raymond, S. N., 2013, AJ, 146, 122

Kipping, D. M., 2009a, MNRAS, 392, 181

Kipping, D. M., 2009b, MNRAS, 396, 1797

Kipping, D. M., 2010a, MNRAS, 407, 301

Kipping, D. M., 2010b, MNRAS, 408, 1758

Kipping, D. M. & Tinetti, G., 2010, MNRAS, 407, 2589

Kipping, D. M., 2011, MNRAS, 416, 689

Kipping, D. M., Bakos, G. Á., Buchhave, L. A., Nesvorný, D. & Schmitt, A. R., 2012, ApJ, 750, 115

Kipping, D. M., Dunn, W. R., Jasinski, J. M. & Manthri, V. P., 2012, MNRAS, 421, 1166

Kipping, D. M., 2013, MNRAS, 435, 2152

Kipping, D. M., 2013, Proc. for the Frank N. Bash Symp. 2013: New Horizons in Astronomy, held October 6–8, 2013 in Austin, TX

Kipping, D. M., Hartman, J., Buchhave, L. A., Schmitt, A. R., Bakos, G. Á., Nesvorný, D., 2013, ApJ, 770, 101

Kipping, D. M., Forgan, D., Hartman, J., Nesvorný, D., Bakos, G. Á., Schmitt, A. R., Buchhave, L. A., 2013, ApJ, 777, 17

Kipping, D. M., 2014, MNRAS, 440, 2164

Kipping, D. M., Nesvorný, D., Buchhave, L. A., Hartman, J., Bakos, G. Á., Schmitt, A. R., 2014, ApJ, 784, 28

Kipping, D. M., Schmitt, A. R., Huang, X., Torres, G., Nesvorný, D., Buchhave, L. A., Hartman, J., Bakos, G. Á., 2015, ApJ, 813, 14

Kipping, D. M., Huang, X., Nesvorný, D., et al. 2015, ApJ, 799, L14

Kipping, D. M. & Sandford, E., 2016, MNRAS, 463, 1323

Kipping, D. M., Cameron, C., Hartman, J. D. et al., 2016, AJ, 153, 93

Kolbl, R., Marcy, G. W., Isaacson, H. & Howard, A. W., 2015, AJ, 149, 18

Kovács, G., Zucker, S., Mazeh, T., 2002, A&A, 391, 369

Kozai, Y., 1968, PASJ, 20, 24

Kundurthy, P., Agol, E., Becker, A. C., Barnes, R., Williams, B. & Mukadam, A., 2011, ApJ, 731, 123

Lainey, V., Arlot, J.-E., Karatekin, Ö., et al., 2009, Nature, 459, 957

Lainey, V., Karatekin, Ö., Desmars, J., et al., 2012, ApJ, 752, 14

Law, N. M. Morton, T., Baranec, C., Riddle, R., Ravichandran, G., Ziegler, C., Johnson, J. A., Tendulkar, S. P., Bui, K., Burse, M. P., Das, H. K., Dekany, R. G., Kulkarni, S., Punnadi, S., Ramaprakash, A. N., 2014, ApJ, 791, 35

Mayor, M. & Queloz, D., 1995, Nature, 378, 355

Mathur, S., Huber, D., Batalha, N. M., Ciardi, D. R., Bastien, F. A., Bieryla, A., Buchhave, L. A., Cochran, W. D., Endl, M., Esquerdo, G. A., Furlan, E., Howard, A., Howell, S. B., Isaacson, H., Latham, D. W., MacQueen, P. J. & Silva, D. R., 2017, ApJS, 229, 30

Morabito, L. A., Synnott, S. P., Kupferman, P. N., & Collins, S. A. 1979, Science, 204, 972

Namouni, F. 2010, ApJ, 719, L145

Rabus, M., Alonso, R., Belmonte, J. A., Deeg, H. J., Gilliland, R. L., Almenara, J. M., Brown, T. M., Charbonneau, D., Mandushev, G. 2009, A&A, 494, 391

Sartoretti, P. & Schneider, J., 1999, A&AS, 134, 553

Seager, S., Mallén-Ornelas, G., 2003, ApJ, 585, 1038

Sheets, H. A. & Deming, D., 2014, ApJ, 794, 133

Simon, A. E., Szabó, Gy. M., Kiss, L. L. & Szatmáry, K., 2012, MNRAS, 419, 164

Smith, J. C., Stumpe, M. C., Van Cleve, J. E., Jenkins, J. M., Barclay, T. S., Fanelli, M. N., Girouard, F. R., Kolodziejczak, J. J., McCauliff, S. D., Morris, R. L. & Twicken, J. D., 2012, PASP, 124, 1000

Spalding, C., Batygin, K., & Adams, F. C. 2016, ApJ, 817, 18

Squyres, S. W., Reynolds, R. T., & Cassen, P. M. 1983, Nature, 301, 225

Sparks, W. B., Hand, K. P., McGrath, M. A., et al. 2016, ApJ, 829, 121

Stumpe, M. C., Smith, J. C., Van Cleve, J. E., Twicken, J. D., Barclay, T. S., Fanelli, M. N., Girouard, F. R., Jenkins, J. M., Kolodziejczak, J. J., McCauliff, S. D. & Morris, R. L., 2012, PASP, 124, 985

Torres, G., Kipping, D. M., Fressin, F. et al., 2015, ApJ, 800, 99

Trafton, L., 1974, ApJ, 193, 477

Waldmann, I. P., Tinetti, G., Drossart, P., Swain, M. R., Deroo, P. & Griffith, C. A., 2012, ApJ, 744, 35

Yang, J., Boué, G., Fabrycky, D. C., & Abbot, D. S. 2014, ApJ, 787, L2

Chapter 3: Evidence for a Large Exomoon Orbiting Kepler-1625b

"It's Mr. Moon who plays in tune, Mr. Moon who knows,
and if it's Mr. Moon that gives the sign then that's the sign that goes."
– Jamiroquai

3.1 Introduction

The search for exomoons remains in its infancy. To date, there are no confirmed exomoons in the literature, although an array of techniques have been proposed to detect their existence, such as microlensing Han & Han (2002); Han (2008); Liebig (2010), direct imaging Cabrera & Schneider (2007); Agol et al. (2015), cyclotron radio emission Noyola et al. (2014), pulsar timing Lewis et al. (2008) and transits Sartoretti & Schneider (1999); Kipping (2009a,b). The transit method is particularly attractive however since many small planets down to lunar radius have already been detected Barclay et al. (2013), and transits afford repeated observing opportunities to further study candidate signals.

Previous searches for transiting moons have established that Galilean-sized moons are uncommon at semimajor axes of 0.1 to 1 astronomical unit (AU) Teachey et al. (2018). This result is consistent with theoretical work that has shown that the shrinking Hill sphere Namouni (2010) and potential capture into evection resonances Spalding et al. (2016) during a planet's inward migration could efficiently remove primordial moons. Nevertheless, amongst a sample of 284 transiting planets recently surveyed for moons, one planet did show some evidence for a large satellite, Kepler-1625b Teachey et al. (2018). The planet is

---

This chapter was first published in Science Advances, vol. 4, issue 10, id. eaav1784 (October 2018). co-authored with David Kipping. https://ui.adsabs.harvard.edu/abs/2018SciA....4.1784T/abstract
The original formatting, which begins with a short summary followed by detailed methodology in the supplementary materials, has been retained.

a Jupiter-sized validated world Morton et al. (2016) orbiting a solar-mass star Mathur et al. (2017) close to 1 AU in a likely circular path Teachey et al. (2018), making it a prime *a priori* candidate for moons. On this basis, and the hints seen in the three transits observed by *Kepler*, we requested and were awarded time on the Hubble Space Telescope (HST) to observe a fourth transit expected on 28 to 29 October 2017. In this work, we report on these new observations and their impact on the exomoon hypothesis for Kepler-1625b.

3.2 Materials and Methods

Our original analysis was the product of a multiyear survey and thus utilized an earlier version of the processed photometry released by the *Kepler* Science Operations Center (SOC). In that study Teachey et al. (2018), we used the simple aperture photometry (`SAP`) from SOC pipeline version 9.0 Jenkins et al. (2010), but the most recent and final data release uses version 9.3. In this work we reanalyzed the *Kepler* data using the revised photometry, which includes updated aperture contamination factors that also affect our analysis. During this process, we also investigated the effect of varying the model used to remove a long-term trend present in the *Kepler* data.

We detrended the revised *Kepler* photometry using five independent methods. The first method is the `CoFiAM` (Cosine Filtering with Autocorrelation Minimization) algorithm Kipping et al. (2013) which was the approach used in the original study, since it was specifically designed with exomoon detection in mind. In addition, we considered four other popular approaches: a polynomial fit, a local line fit, a median filter, and a Gaussian process (see the Supplementary Materials for a detailed description of each). The detrended photometry is stable across the different methods (see Fig. 3.1), with a maximum standard deviation (SD) between any two SAP time series of 250 parts per million (ppm), far below the median formal uncertainty of ~ 590 ppm. Although we verified that the Presearch Data Conditioning (`PDC`) version of the photometry Stumpe et al. (2012); Smith et al. (2012) produces similar results (as evident in Fig. 3.1), we ultimately only used the five `SAP` reductions in what follows. We

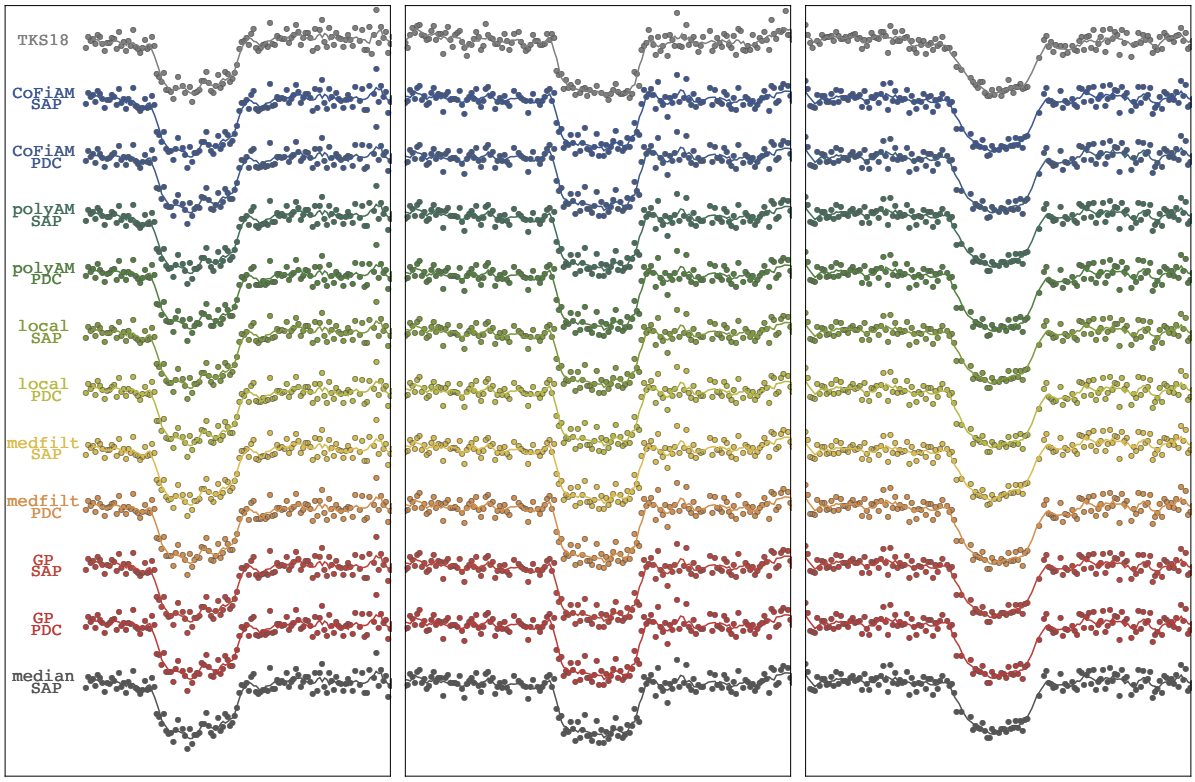

Figure 3.1: **Method marginalized detrending.** Comparison of five different detrending methods on two different data *Kepler* products. Top curve shows the *Kepler* reduction used in Teachey et al. (2018) and the bottom curve shows the method marginalized product used in this work.

produced a "method marginalized" final time series by taking the median of the ith datum across the five methods and propagating the variance between them into a revised uncertainty estimate (see the Supplementary Materials for details). In this way, we produced a robust correction of the *Kepler* data accounting for differences in model assumptions.

We fit photodynamical models Kipping (2011) to the revised *Kepler* data, using the updated contamination factors from SOC version 9.3, before introducing the new HST data. Bayesian model selection revealed only a modest preference for the moon model, with the Bayes factor (K), going from $2 \log K = 20.4$ in our original study down to just 1.0 now. Detailed investigation revealed that this is not due to our new detrending approach, as we applied our method marginalized detrending to the original version 9.0 data and recovered a

similar result to our original analysis (see the Supplementary Materials for details). Instead, it appeared that the reduced evidence was largely caused by the changes in the `SAP` photometry going from version 9.0 to 9.3, and to a lesser degree by the new contamination factors. This can be seen in Fig. 3.1, where the third transit in particular experienced a pronounced change between the two versions, and it was this epoch that displayed the greatest evidence for a moon-like signature in the original analysis.

With a much larger aperture than *Kepler*, HST is expected to provide several times more precise photometry. Accordingly, the question as to whether Kepler-1625b hosts a large moon should incorporate this new information and in what follows we describe how we processed the HST data and then combined them with the revised *Kepler* photometry.

HST monitored the transit of Kepler-1625b occurring on 28 to 29 October, 2017 with Wide Field Camera 3 (WFC3). A total of 26 orbits, amounting to some 40 hours, were devoted to observing the event. The observations consisted of one direct image and 232 exposures using the G141 grism, a slitless spectroscopy instrument that projects the star's spectrum across the detector. This provides spectral information on the target in the near-infrared from about 1.1 to 1.7 μm. Of these 232 exposures, only three were unusable, as they coincided with the spacecraft's passage through the South Atlantic Anomaly, at which time HST was forced to use its less-accurate gyroscopic guidance system. Each exposure lasted roughly 5 min, resulting in about 45 min on target per orbit. Images were extracted using standard tools made available by the Science Telescope Space Institute (STScI) and are described in the Supplementary Materials.

Native HST time stamps, recorded in the Modified Julian Date system, were converted to Barycentric Julian Date (BJDUTC) for consistency with the *Kepler* time stamps. The BJDUTC system accounts for light travel time based on the position of the target and the observer with respect to the solar system barycenter at the time of observation. As the position of HST is constantly changing we set the position of the observer to be the center of the Earth at the time of observation, for which a small discrepancy of ±23 ms is introduced.

This discrepancy can be safely ignored for our purposes.

While the telescope performed nominally throughout the observation, three well-documented sources of systematic error were present in our data that required removal. First, thermal fluctuations due to the spacecraft's orbit led to clear brightness changes across the entire CCD (sometimes referred to as "breathing"), which were corrected for by subtracting image median fluxes (see the Supplementary Materials for details). After computing an optimal aperture for the target, we observed a strong intra-orbit ramping effect (also known as the "hook") in the white light curve (see Fig. 3.2), which has been previously attributed to charge trapping in the CCD Agol et al. (2010); Berta et al. (2012). We initially tried a standard parametric approach for correcting these ramps using an exponential function, but found the result to be suboptimal. Instead, we devised a new nonparametric approach described in the Supplementary Materials that substantially outperformed the previous approach.

We achieved a final mean intra-orbit precision of 375.5 ppm (versus 440.1 ppm using exponential functions), which was about 3.8 times more precise than *Kepler* when correcting for exposure time. The transit of Kepler-1625b was clearly observed even before the hook correction. After removal of the hooks, an apparent second decrease in brightness appeared towards the end of the observations, which was evident even in the noisier exponential ramp corrected data (see Fig. 3.2). Repeating our analysis for the only other bright star fully on the CCD, KIC 4760469, revealed no peculiar behavior at this time indicating that the dip was not due to an instrumental common mode. Similarly, the centroids of both the target and the comparison star showed no anomalous change around this time (see Figure 3.13).

A detailed analysis of the centroid variations of both the target and the comparison star revealed that the 10 millipixel motion observed was highly unlikely to be able to produce the ~500 ppm dip associated with the moon-like signature. Further, we found that the signal was achromatic appearing in two distinct spectral channels, which was consistent with expectations for a real moon. Finally, a detailed analysis of the photometric residuals revealed that the fits including a moon-like transit were consistent with uncorrelated noise equal to

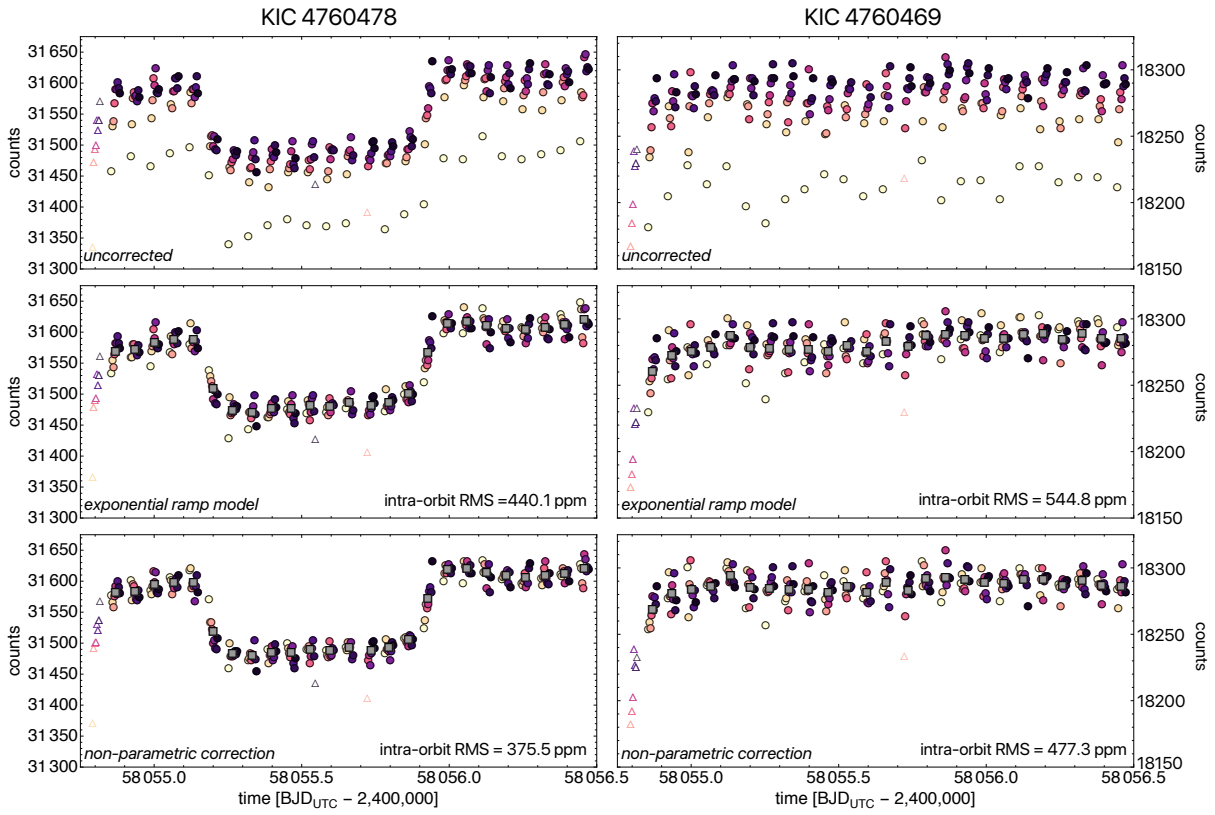

Figure 3.2: **Hook corrections**. (Top) The optimal aperture photometry of our target (left) and the best comparison star (right), where the hooks and visit-long trends are clearly present. Points are colored by their exposure number within each HST orbit (triangles represent outliers). (Middle) A hook-correction using the common exponential ramp model on both stars. (Bottom) The result from an alternative and novel hook-correction approach introduced in this work.

the value derived from our hook correction algorithm. These three tests, detailed in the Supplementary Materials, provide no reason to doubt that the moon-like dip is astrophysical in nature and thus we treat it as such in what follows.

Upon inspection of the HST images we identified a previously uncataloged point source within 2 arcseconds of our target. The star resides at position angle 8.5° east of north, with a derived *Kepler* magnitude of 22.7. We attribute its new identification to the fact that it is both exceptionally faint and so close to the target that it was always lost in the glare in other images. Using a *Gaia*-derived distance to the target we found that, were this point source to be at the same distance, it would be within 4500 AU of Kepler-1625. It is not known whether the two sources are physically associated, however. Its faintness means that it produces negligible contamination to our target spectrum. We estimated that the source has a variability of 0.33% and contributes less than 1 part in 3000 to our final WFC3 white light curve, which means that the net contribution to our target is 1 ppm and can be safely ignored.

In addition to the breathing and the hooks, a third well-known source of WFC3 systematic error we see is a visit-long trend (apparent in Fig. 3.2). These trends have not yet been correlated to any physical parameter related to the WFC3 observations Wakeford et al. (2016), and thus the conventional approach is a linear slope (for example, 25-27) although a quadratic model has been used in some instances (for example, 28,29) The time scale of the variations is comparable to the transit itself and thus cannot be removed in isolation; rather, any detrending model is expected to be covariant with the transit model. For this reason, it was necessary to perform the detrending regression simultaneous to the transit model fits. We considered three possible trend models; linear, quadratic and exponential. All models include an extra parameter describing a flux offset between the 14^{th} and 15^{th} orbits. This is motivated by the fact that the spacecraft performed a full guide star acquisition at the beginning of the 15^{th} orbit (a new "visit"), and ended up placing the spectrum ~0.1 pixels away from where it appeared during the first 14 orbits. Although the white light curve

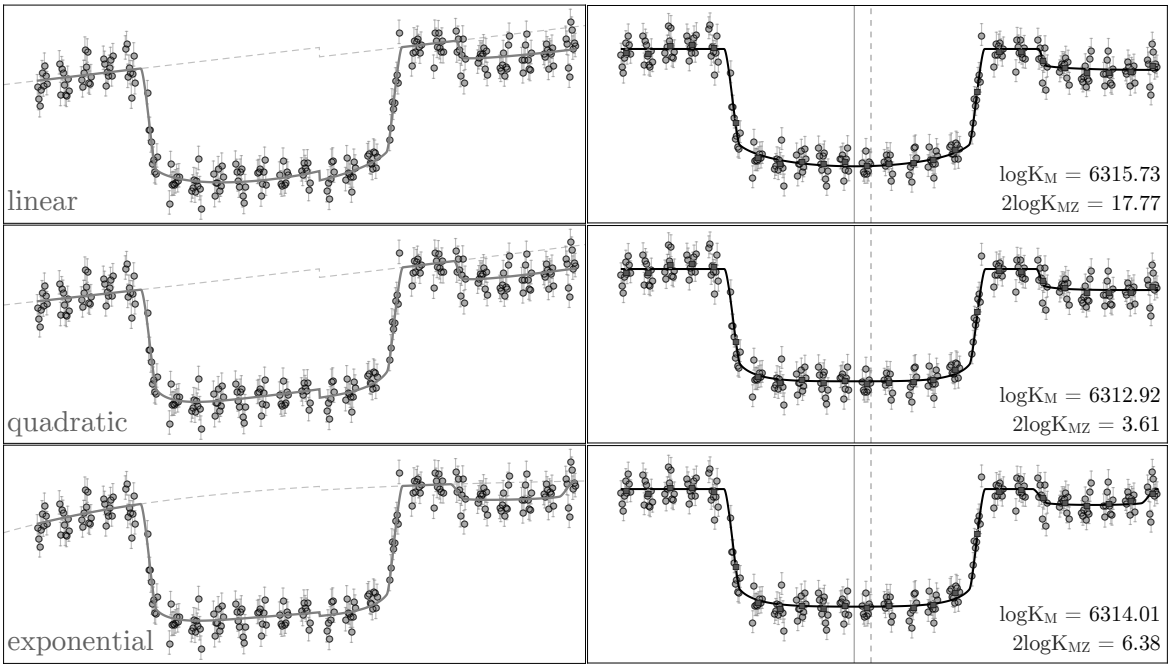

Figure 3.3: **HST detrending.** The HST observations with three proposed trends fit to the data (left) and with the trends removed (right). Bottom-right numbers in each row give the Bayes factor between a planet plus moon model (model M) and a planet plus moon model where the moon radius equals zero (model Z), which tracks the significance of the moon-like dip in isolation.

shows no obvious flux change at this time, the reddest channels display substantial shifts motivating this offset term.

Finally, we extracted light curves in nine wavelength bins across the spectrum in an attempt to perform transmission spectroscopy. As a planet transits its host star, the atmosphere may absorb different amounts of light depending on the constituent molecules and their abundances Seager & Sasselov (2000). This makes the planet's transit depth wavelength-dependent. An accurate measurement of these transit depths not only provides the potential to characterize the atmosphere's composition; it is also potentially useful in providing an independent measurement of the planet's mass De Wit & Seager (2013). While a low surface gravity planet will show very pronounced molecular features and a steep slope at short wavelengths due to Rayleigh scattering, a high surface gravity world will yield a substantially flatter transmission spectrum.

With the HST WFC3 data prepared, we are ready to combine them with the revised *Kepler* data to regress candidate models and compare them. We considered four different transit models, which, when combined with three different visit-long trend models, leads to a total of 12 models to evaluate. The four transit models here were designated as P, for the planet-only model; T, for a model that fits the observed transit timing variations (TTVs) in the system agnostically; Z, for the zero-radius moon model, which may produce all the gravitational effects of an exomoon without the flux reductions of a moon transit; and M, which is the full planet plus moon model. Models were generated using the `LUNA` photodynamical software package Kipping (2011) and regression was performed via the multimodal nested sampling algorithm MultiNest Feroz & Hobson (2008); Feroz et al. (2009). For each model, we derived not only the joint *a posteriori* parameter samples, but also a Bayesian evidence (also known as the marginal likelihood) enabling direct calculation of the Bayes factor between models.

3.3 Results

One clear result from our analysis is that the HST transit of Kepler-1625b occurred 77.8 min earlier than expected, indicating TTVs in the system. Bayes factors between models P and T support the presence of significant TTVs for any choice of detrending model (see Table 5.3.3), with the T fits returning a χ^2 decreased by 17 to 19 (for 1048 data points). Further, if we fit the *Kepler* data in isolation and make predictions for the HST transit time, the observed time is $> 3\,\sigma$ discrepant (see Figure 3.16 in the Supplementary Materials). For reference, each *Kepler* transit midtime has an uncertainty on the order of 10 min and the standard deviation on linear ephemeris predictions is 25.2 min derived from posterior samples. Identifying TTVs was among the first methods proposed to discover exomoons Sartoretti & Schneider (1999), but certainly perturbations from an unseen planet could also be responsible. We find that the $\simeq$25 min amplitude TTV can be explained by an external perturbing planet (see the Supplementary Materials), although with only four transits on

	linear	quadratic	exponential
$\log \mathcal{Z}_{\rm P}$	6302.79 ± 0.11	6306.68 ± 0.11	6308.41 ± 0.11
$\log \mathcal{Z}_{\rm T}$	6304.86 ± 0.11	6308.81 ± 0.12	6310.71 ± 0.11
$\log \mathcal{Z}_{\rm Z}$	6306.84 ± 0.11	6311.12 ± 0.12	6310.82 ± 0.12
$\log \mathcal{Z}_{\rm M}$	6315.73 ± 0.12	6312.92 ± 0.12	6314.01 ± 0.12
$2\log K(\mathcal{Z}'_{\rm M}/\mathcal{Z}'_{\rm P})$	1.00 ± 0.22		
$2\log(\mathcal{Z}_{\rm M}/\mathcal{Z}_{\rm P})$	25.88 ± 0.32	12.47 ± 0.33	11.19 ± 0.32
$2\log(\mathcal{Z}_{\rm M}/\mathcal{Z}_{\rm T})$	21.72 ± 0.33	8.21 ± 0.34	17.81 ± 0.33
$2\log(\mathcal{Z}_{\rm M}/\mathcal{Z}_{\rm Z})$	17.77 ± 0.33	3.61 ± 0.33	6.38 ± 0.34
$\Delta\chi'^2_{\rm PM} = 2\log(\hat{\mathcal{L}}'_{\rm M}/\hat{\mathcal{L}}'_{\rm P})$	18.66		
$\Delta\chi^2_{\rm PM} = 2\log(\hat{\mathcal{L}}_{\rm M}/\hat{\mathcal{L}}_{\rm P})$	54.93	41.04	41.57
$\Delta\chi^2_{\rm TM} = 2\log(\hat{\mathcal{L}}_{\rm M}/\hat{\mathcal{L}}_{\rm T})$	35.69	23.97	23.97
$\Delta\chi^2_{\rm ZM} = 2\log(\hat{\mathcal{L}}_{\rm M}/\hat{\mathcal{L}}_{\rm Z})$	33.68	19.59	19.22

Table 3.1: **Model performance.** Bayesian evidences ($\mathcal{Z}$) and maximum likelihoods ($\hat{\mathcal{L}}$) from our combined fits using *Kepler* and new HST data. *Kepler*plus HST fits. The subscripts are P for the planet model, T for the planetary TTV model, Z for the zero-radius moon model and M for the moon model. The three columns are for each trend model attempted. The primed values correspond to those derived the *Kepler* data in isolation.

hand it is not possible to constrain the mass or location of such a planet, and no other planet has been observed so far in the system.

We also found that model Z consistently outperforms model T, though the improvement to the fits is smaller at $\Delta\chi^2 \simeq$ 2-5 (see Table 5.3.3). This suggests that the evidence for the moon based on timing effects alone goes beyond the TTVs, providing modest evidence in favor of additional dynamical effects such as duration changes Kipping (2009a) and/or impact parameter variation Kipping (2009b), both expected consequences of a moon present in the system. This by itself would not constitute a strong enough case for a moon detection claim, but we consider it to be an important additional check that a real exomoon would be expected to pass.

The most compelling piece of evidence for an exomoon would be an exomoon transit, in addition to the observed TTV. If Kepler-1625b's early transit were indeed due to an exomoon, then we should expect the moon to transit late on the opposite side of the barycenter. The previously mentioned existence of an apparent flux decrease towards the end of our

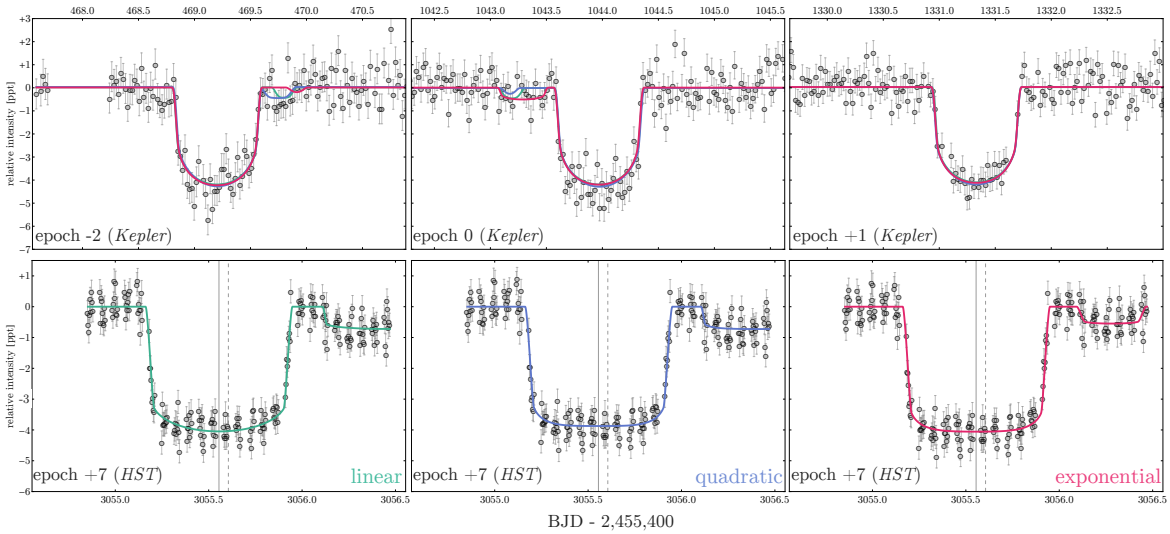

Figure 3.4: **Moon solutions.** The three transits in *Kepler* (top) and the October 2017 transit observed with HST (bottom) for the three trend model solutions. The three colored lines show the corresponding trend model solutions for model M, our favored transit model. The shape of the HST transit differs from that of the *Kepler* transits owing to limb darkening differences between the bandpasses.

observations is therefore where we would expect it to be under this hypothesis. Although we have established that this dip is most likely astrophysical, we have not yet discussed its significance or its compatibility with a self-consistent moon model.

We find that our self-consistent planet plus moon models (M) always outperform all other transit models in terms of maximum likelihood and Bayesian evidences (see Table 5.3.3). The moon signal is found to have a signal-to-noise ratio of at least 19. The presence of a TTV and an apparent decrease in flux at the correct phase position together suggest that the exomoon is the best explanation. However, as is apparent from Fig. 3.3, the amplitude and shape of the putative exomoon transit vary somewhat between the trend models, leading to both distinct model evidences and associated system parameters.

3.4 Discussion

Although the overall preference of the moon model is arguably best framed by comparison to model P, the significance of the moon-like transit alone is best framed by comparing M

and Z alone. Such a comparison reveals a strong dependency of the implied significance on the trend model used. In the worst case, we have the quadratic model with $2 \log K \simeq 4$, corresponding to "positive evidence" Kass & Raftery (1995) - although we note that the absolute evidence $\mathcal{Z}_M$ is the worst amongst the three. The linear model is far more optimistic yielding $2 \log K \simeq 18$, corresponding to "very strong evidence" Kass & Raftery (1995), whereas the exponential sits between these extremes. The question then arises, which of our trend models is the correct one?

Because the linear model is a nested version of the quadratic model, and both models are linear with respect to time, it is more straightforward to compare these two. The quadratic model essentially recovers the linear model, apparent from Fig. 3.3, with a curvature within $1.5\,\sigma$ of zero, and yields almost the same best χ^2 score to within 1.2. This lack of meaningful improvement causes the log evidence to drop by 2.8, since evidences penalize wasted prior volume. The exponential model appears more competitive with a log evidence of 1.72 lower, but a direct comparison of two different classes of models, such as these, is muddied by the fact that these analyses are sensitive to the choice of priors. The most useful comparison here is simply to state that the maximum likelihoods are within $\Delta\chi^2 = 0.68$ of one another and thus are likely equally justified from data-driven perspective.

Another approach we considered is to weigh the trend models using the posterior samples. Given a planet or moon's mass, there is a probabilistic range of expected radii based on empirical mass-radius relations Chen & Kipping (2017). Although we exclude extreme densities in our fits, parameters from model M can certainly lead to improbable solutions with regard to the photodynamically inferred Kipping (2010) masses and radii.

To investigate this, we inferred the planetary mass using two methods for each model and evaluated their self-consistency. The first method combines the photodynamically-inferred planet-to-star mass ratio Kipping (2010) with a prediction for the mass based on the well-constrained radius using `forecaster`; an empirical probabilistic mass-radius relation Chen & Kipping (2017). The second method approaches the problem from the other side, taking the

Table 3.2: **System parameters.** Median and ±34.1% quantile range of the *a posteriori* model parameters from model M, where each column defined a different visit-long trend model. The top panel gives the credible intervals for the actual parameters used in the fit, and the lower panel gives a selection of relevant derived parameters conditioned upon our revised stellar parameters. The quoted inclination of the satellite is the inclination modulo 90°.

Parameter	Linear	Quadratic	Exponential
Photodynamics only			
$R_{P,\mathrm{Kep}}/R_\star$	$0.06075^{+0.00062}_{-0.00065}$	$0.06061^{+0.00068}_{-0.00073}$	$0.06072^{+0.0062}_{-0.00063}$
$R_{P,\mathrm{HST}}/R_{P,\mathrm{Kep}}$	$0.998^{+0.013}_{-0.013}$	$1.009^{+0.019}_{-0.017}$	$1.006^{+0.014}_{-0.014}$
$\rho_{\star,\mathrm{LC}}$ [g cm^{-3}]	424^{+9}_{-16}	424^{+9}_{-15}	425^{+9}_{-14}
b	$0.104^{+0.084}_{-0.066}$	$0.099^{+0.088}_{-0.063}$	$0.096^{+0.078}_{-0.058}$
P_P [days]	$287.37278^{+0.00075}_{-0.00065}$	$287.3727^{+0.0022}_{-0.0015}$	$287.37269^{+0.00074}_{-0.00076}$
τ_0 [BJD$_{\mathrm{UTC}}$]	$2456043.9587^{+0.0027}_{-0.0027}$	$2456043.9572^{+0.0033}_{-0.0093}$	$56043.9585^{+0.0025}_{-0.0029}$
$q_{1,\mathrm{Kep}}$	$0.45^{+0.19}_{-0.14}$	$0.44^{+0.19}_{-0.15}$	$0.45^{+0.18}_{-0.14}$
$q_{2,\mathrm{Kep}}$	$0.31^{+0.19}_{-0.15}$	$0.32^{+0.20}_{-0.16}$	$0.31^{+0.19}_{-0.15}$
$q_{1,\mathrm{HST}}$	$0.087^{+0.057}_{-0.041}$	$0.096^{+0.064}_{-0.045}$	$0.087^{+0.056}_{-0.040}$
$q_{2,\mathrm{HST}}$	$0.25^{+0.25}_{-0.15}$	$0.21^{+0.23}_{-0.14}$	$0.22^{+0.22}_{-0.14}$
P_S [days]	22^{+17}_{-9}	24^{+18}_{-11}	22^{+15}_{-9}
a_{SP}/R_P	45^{+10}_{-5}	36^{+10}_{-13}	42^{+7}_{-4}
ϕ_S [°]	179^{+136}_{-70}	141^{+161}_{-65}	160^{+150}_{-60}
i_S [°]	42^{+15}_{-18}	49^{+21}_{-22}	43^{+15}_{-19}
Ω_S [°]	0^{+142}_{-83}	12^{+132}_{-113}	8^{+136}_{-81}
(M_S/M_P)	$0.0141^{+0.0048}_{-0.0039}$	$0.0196^{+0.0294}_{-0.0071}$	$0.0149^{+0.0052}_{-0.0038}$
(R_S/R_P)	$0.431^{+0.033}_{-0.036}$	$0.271^{+0.150}_{-0.099}$	$0.363^{+0.048}_{-0.079}$
Δa_0 [ppm]	330^{+120}_{-120}	180^{+170}_{-210}	220^{+130}_{-140}
+ Stellar properties			
$R_\star$ [$R_\odot$]	$1.73^{+0.24}_{-0.22}$	$1.73^{+0.24}_{-0.22}$	$1.73^{+0.24}_{-0.22}$
$M_\star$ [$M_\odot$]	$1.04^{+0.08}_{-0.06}$	$1.04^{+0.08}_{-0.06}$	$1.04^{+0.08}_{-0.06}$
$\rho_{\star,\mathrm{iso}}$ [kg m^{-3}]	$0.29^{+0.13}_{-0.09}$	$0.29^{+0.13}_{-0.09}$	$0.29^{+0.13}_{-0.09}$
$e^\dagger_{\mathrm{min}}$	$0.13^{+0.11}_{-0.09}$	$0.13^{+0.11}_{-0.09}$	$0.13^{+0.11}_{-0.09}$
R_P [$R_\oplus$]	$11.4^{+1.6}_{-1.5}$	$11.4^{+1.6}_{-1.4}$	$11.4^{+1.6}_{-1.4}$
$\log_{10}(M_P/M_\oplus)$	$2.86^{+0.48}_{-0.50}$	$2.40^{+0.70}_{-0.72}$	$2.75^{+0.53}_{-0.54}$
a_P [AU]	$0.98^{+0.14}_{-0.13}$	$0.98^{+0.14}_{-0.12}$	$0.98^{+0.14}_{-0.12}$
R_S [$R_\oplus$]	$4.90^{+0.79}_{-0.72}$	$3.09^{+1.71}_{-1.19}$	$4.05^{+0.86}_{-1.01}$
$\log_{10}(M_S/M_\oplus)$	$1.00^{+0.46}_{-0.48}$	$0.74^{+0.56}_{-0.52}$	$0.93^{+0.49}_{-0.50}$
S_{eff} [$S_\oplus$]	$2.65^{+0.19}_{-0.16}$	$2.64^{+0.18}_{-0.16}$	$2.64^{+0.18}_{-0.16}$
+ forecaster			
$\log_{10}(M_P/M_\oplus)$	$3.12^{+0.26}_{-0.27}$	$2.65^{+0.50}_{-0.52}$	$3.01^{+0.26}_{-0.30}$
$\log_{10}(M_S/M_\oplus)$	$1.27^{+0.29}_{-0.30}$	$1.11^{+0.55}_{-0.58}$	$1.20^{+0.32}_{-0.34}$
M_P [M_J]	$[1.2, 12.5]$	$[0.2, 9.0]$	$[0.6, 10.5]$
M_S [$M_\oplus$]	$[4.4, 68]$	$[1.0, 140]$	$[2.6, 76]$
K [m/s]	$[35, 380]$	$[6, 280]$	$[18, 320]$

moon's radius and predicting its mass with `forecaster` and then calculating the planetary mass via the photodynamically-inferred moon-to-planet mass ratio. Our analysis (discussed in more detail in the Supplementary Materials) reveals that all three models have physically plausible solutions and generally converge at $\sim 10^3\, M_\oplus$ for the planetary mass, with the exception of the quadratic model that had broader support extending down to Saturn-mass. We ultimately combined the two mass estimates to provide a final best-estimate for each model in Table 3.2.

As a consistency check, we used our derived transmission spectrum to constrain the allowed range of planetary masses for a cloudless atmosphere De Wit & Seager (2013). Using an MCMC (Markov chain Monte Carlo) with `Exo-Transmit` Kempton et al. (2017), we find that masses in the range of > 0.4 Jupiter masses (to 95% confidence) are consistent with the nearly flat spectrum observed, assuming a cloudless atmosphere (see the Supplementary Materials for details).

In conclusion, the linear and exponential models appear to be the most justified by the data and also lead to slightly improved physical self-consistency, although we certainly cannot exclude the quadratic model at this time. For this reason, we elected to present the associated system parameters resulting from all three models in Table 3.2. The maximum *a posteriori* solutions from each, using model M, are presented in Fig. 3.4 for reference.

We briefly comment on some of the inferred physical parameters for this system. First, we have assumed a circular moon orbit throughout due to the likely rapid effects of tidal circularization. However, we did allow the moon to explore three-dimensional orbits and find some evidence for noncoplanarity. Our solution somewhat favors a moon orbit tilted by about 45° to the planet's orbital plane, with both pro- and retrograde solutions being compatible. The only comparable known large moon with such an inclined orbit is Triton around Neptune, which is generally thought to be a captured Kuiper Belt object Agnor & Hamilton (2006). However, we caution that the constraints here are weak, reflected by the posterior's broad shape, and thus it would be unsurprising if the true answer is coplanar.

One jarring aspect of the system is the sheer scale of it. The exomoon has a radius of $\simeq 4\,R_\oplus$, making it very similar to Neptune or Uranus in size. The measured mass, including the `forecaster` constraints, comes in at $\log(M_S/M_\oplus) = (1.2 \pm 0.3)$, which is again compatible with Neptune or Uranus (although note that this solution is in part informed by an empirical mass-radius relation). This Neptune-like moon orbits a planet with a size fully compatible with that of Jupiter at $(11.4 \pm 1.5)\,R_\oplus$, but most likely a few times more massive. Finally, although the moon's period is highly degenerate and multimodal, we find the semimajor axis is relatively wide at $\simeq$40 planetary radii. With a Hill radius of (200 ± 50) planetary radii, this is well within the Hill sphere and expected region of stability (see the Supplementary Materials for further discussion).

The blackbody equilibrium temperature of the planet and moon, assuming zero albedo, is $\sim$350 K. Adopting a more realistic albedo can drop this down to $\sim$ 300 K. Of course, as a likely gaseous pair of objects there is not much prospect of habitability here, although it appears that the moon can indeed be in the temperature zone for optimistic definitions of the habitable zone.

What is particularly interesting about the star is that it appears to be a solar-mass star evolving off the main sequence. This inference is supported by a recent analysis of the *Gaia* DR2 parallax by Berger et al. (2018), as well as our own isochrone fits (see the Supplementary Materials). We find that the star is certainly older than the Sun, at $\simeq$9 gigayears in age, and that insolation at the location of the system was thus lower in the past. The luminosity was likely close to solar for most of the star's life, making the equilibrium temperature drop down to $\sim$250 K for Jovian albedos for most of its existence. The old age of the system also implies plenty of time for tidal evolution, which could explain why we find the moon at a fairly wide orbital separation.

The origins of such a system can only be speculated upon at this time. A mass ratio of 1.5% is certainly not unphysical from in-situ formation using gas-starved disk models, but it does represent the very upper end of what numerical simulations form Cilibrasi et al. (2018).

In such a scenario, a separate explanation for the tilt would be required. Impacts between gaseous planets leading to captured moons are not well-studied but could be worth further investigation. A binary exchange mechanism would be challenged by the requirement for a Neptune to be in an initial binary with an object of comparable mass, such as a super-Earth Agnor & Hamilton (2006). Formation of an initial binary planet, perhaps through tidal capture, seems improbable due to the tight orbits simulation work tends to produce from such events Ochiai et ala. (2014). If confirmed, Kepler-1625b-i will certainly provide an interesting puzzle for theorists to solve.

3.5 Conclusion

Together, a detailed investigation of a suite of models tested in this work suggests that the exomoon hypothesis is the best explanation for the available observations. The two main pieces of information driving this result are (i) a strong case for TTVs, in particular a 77.8 min early transit observed during our HST observations and (ii) a moon-like transit signature occurring after the planetary transit. We also note that we find a modestly improved evidence when including additional dynamical effects induced by moons aside from TTVs.

The exomoon hypothesis is further strengthened by our analysis that demonstrates that (i) the moon-like transit is not due to an instrumental common mode, residual pixel sensitivity variations, or chromatic systematics; (ii) the moon-like transit occurs at the correct phase position to also explain the observed TTV; and (iii) simultaneous detrending and photodynamical modeling retrieves a solution that is not only favored by the data, but is also physically self-consistent.

Together, these lines of evidence all support the hypothesis of an exomoon orbiting Kepler-1625b. The exomoon is also the simplest hypothesis to explain both the TTV and the post-transit flux decrease, since other solutions would require two separate and unconnected explanations for these two observations.

There remain some aspects of our present interpretation of the data that give us pause.

First, the moon's Neptunian size and inclined orbit are peculiar, though it is difficult to assess how likely this is *a priori* since no previously known exomoons exist. Second, the moon's transit occurs towards the end of the observations and more out-of-transit data could have more cleanly resolved this signal. Third, the moon's inferred properties are sensitive to the model used for correcting HST's visit-long trend and thus some uncertainty remains regarding the true system properties. However, the solution we deem most likely, a linear visit-long trend, also represents the most widely agreed upon solution for the visit-long trend in the literature.

Finally, it is somewhat ironic that the case for observing Kepler-1625b with HST was contingent on a previous data release of the *Kepler* photometry that indicated a moon Teachey et al. (2018), while the most recent data release only modestly favors that hypothesis when treated in isolation. Despite this, we would argue that planets like Kepler-1625b – Jupiter-sized planets on wide, circular orbits around solar-mass stars – were always ideal targets exomoon follow-up. There are certainly hints of the moon present even in the revised *Kepler* data, but it is the HST data – with a precision four times superior to *Kepler*– that are critical to driving the moon as the favored model. These points suggest that it would be worthwhile to pursue similar *Kepler* planets for exomoons with HST or other facilities, even if the *Kepler* data alone do not show large moon-like signatures. Furthermore, our work demonstrates how impactful the changes to *Kepler* photometry were, at least in this case, as it suggests other results over the course of the *Kepler* mission may be similarly affected, particularly for small signals.

All in all, it is difficult to assign a precise probability to the reality of Kepler-1625b-i. Formally, the preference for the moon model over the planet-only model is very high, with a Bayes factor exceeding 400,000. On the other hand, this is a complicated and involved analysis where a minor effect unaccounted for, or an anomalous artifact, could potentially change our interpretation. In short, it is the unknown unknowns that we cannot quantify. These reservations exist because this would be a first-of-its-kind detection – the first exomoon.

Historically, the first exoplanet claims faced great skepticism because there was simply no precedence for them. If many more exomoons are detected in the coming years with similar properties to Kepler-1625b-i, it would hardly be a controversial claim to add one more. Ultimately, Kepler-1625b-i cannot be considered confirmed until it has survived the long scrutiny of many years, observations and community skepticism, and perhaps the detection of similar such objects. Despite this, it is an exciting reminder of how little we really know about distant planetary systems and the great spirit of discovery that exoplanetary science embodies.

3.6 Supplementary Materials and Methods

3.6.1 *Kepler* Re-analysis

Background

The original analysis of the candidate exomoon signal is described extensively in Section 8 of Teachey et al. (2018) (Chapter 2). We refer readers to that work for the details of our original interpretation of the data. We briefly recap here the case for Kepler-1625b as a candidate exomoon host.

The analysis in Teachey et al. (2018) was the largest and most ambitious search for exomoons to date, requiring many years to complete. Consequently, the data used in that study were first downloaded on November 10th 2014. Since the *Kepler* Science Processing Pipeline (see Jenkins et al. (2010)), built by the *Kepler* Science Operations Center (SOC), evolved over the years of the primary mission, the data analyzed in Teachey et al. (2018) does not represent the most up-to-date data product at the time of writing. The three quarters in which Kepler-1625b is observed to transit are quarters 7, 13 and 16. In Teachey et al. (2018), the simple aperture photometry (SAP) time series used were produced by SOC pipeline v9.0.3 (corresponding to data release DR21) for quarters 7 and 13 and v9.0.7 (DR22) for quarter 16. At the time of writing, the most up-to-date (and the final) data release is

DR25, for which quarters 7, 13 and 16 were processed by SOC v9.3.24, v9.3.29 and v9.3.31.

Based on our prior experience with *Kepler* data products, we did not anticipate the *Kepler* photometry or case for the exomoon candidate would significantly change as a result of going from SOC v9.0 to v9.3. Nevertheless, in this more detailed and focused study, we decided to revisit the *Kepler* photometry and verify that this was true, as well as ensure that our results are robust against choice of detrending method.

Joint modeling of the original data had suggested the presence of a large moon in the system, with statistically significant drops in flux appearing on the wings of these transit events. Based on DR25 estimates of the star's radius ($1.793 \pm 0.263\,R_{\odot}$), the planet is approximately the size of Jupiter. Meanwhile the photodynamical fits to the data, which require a self-consistent moon model for all transit events, suggested the moon's radius was comparable to that of Neptune (although with sizable uncertainty). No bad data flags (such as reaction wheel zero crossing events) or anomalous pixel behavior (such as that seen a previous candidate Kepler-90g; Kipping et al. (2015b)) could explain the candidate signal at the time, nor could we identify any astrophysical explanation besides the moon hypothesis that accorded well with the data in hand.

Despite this, we argued that three transits were insufficient to claim strong evidence for an exomoon detection, and so an additional validation observation was sought and awarded on the Hubble Space Telescope (HST). Twenty-six orbits amounting to ~40 hours were awarded and the observations were executed on 28-29 October 2018.

In our analysis of the HST data, it was necessary to perform a joint fit with the *Kepler* photometry. To that end, we conducted a revised analysis of the *Kepler* data, now with the SOC v9.3 data product. Upon initial inspection of these results it was found that significant differences do exist between the revised *Kepler* data and the earlier release used for this target. In addition to the photometry, which has undergone noticeable modification, one changing value in particular – the crowding or blend factor (`CROWDSAP`), a measure of aperture contamination by nearby sources – appeared to play an important role in determining

whether the moon model is favored. This is because blending can introduce transit depth variations which could be explained by a moon in proximity to the planet. Because this single value was such an important part of the moon fit we sought to determine the cause of this change from SOC v9.0 to SOC v9.3.

KIC 4760471 - The Phantom Star

In investigating the source of the modified crowding values we examined the Data Validation (DV) report for Kepler-1625b closely. This examination revealed that there is star included in the optimal aperture sky maps that does not exist. The source, KIC 4760471, is clearly present in the DV maps, situated almost directly between our target and the neighboring KIC 4760469, and in fact, the star is purportedly brighter than our target by about half a magnitude (see Figure 3.5). However, we find no evidence of this source in images from 2MASS, UKIRT, Pan-STARRS, *Gaia*, or our HST images. Nor is the star included in the catalogue of nearby sources available on the *Kepler* Exoplanet Follow-Up Observing Program (ExoFOP) website. We must conclude that this star is a spurious inclusion in the KIC, and this raises the question of how many other such stars there may be in the catalogue that may be producing erroneous contamination estimates. Indeed, this is not the first "phantom star" to be identified in the KIC; Dalba et al. (2017) also reported the discovery of a spurious KIC inclusion that resulted a $\sim 50\%$ transit depth change for Kepler-445c.

After modeling what *Kepler* should be seeing based on the published Pixel Response Function Bryson et al. (2010), we introduced KIC 4760471 into the model to test whether the star was actually included in the models for SOC v9.3 and earlier releases. We find that it cannot have been a part of the SOC model, despite its inclusion in the DV report, as the brightness and proximity of this phantom star would contaminate the optimal aperture by $\sim 50\%$ or more, and this is not reflected in the CROWDSAP numbers published by the SOC. This ultimately led us to conclude that the SOC v9.3 were accurate even if the DV report appears in error.

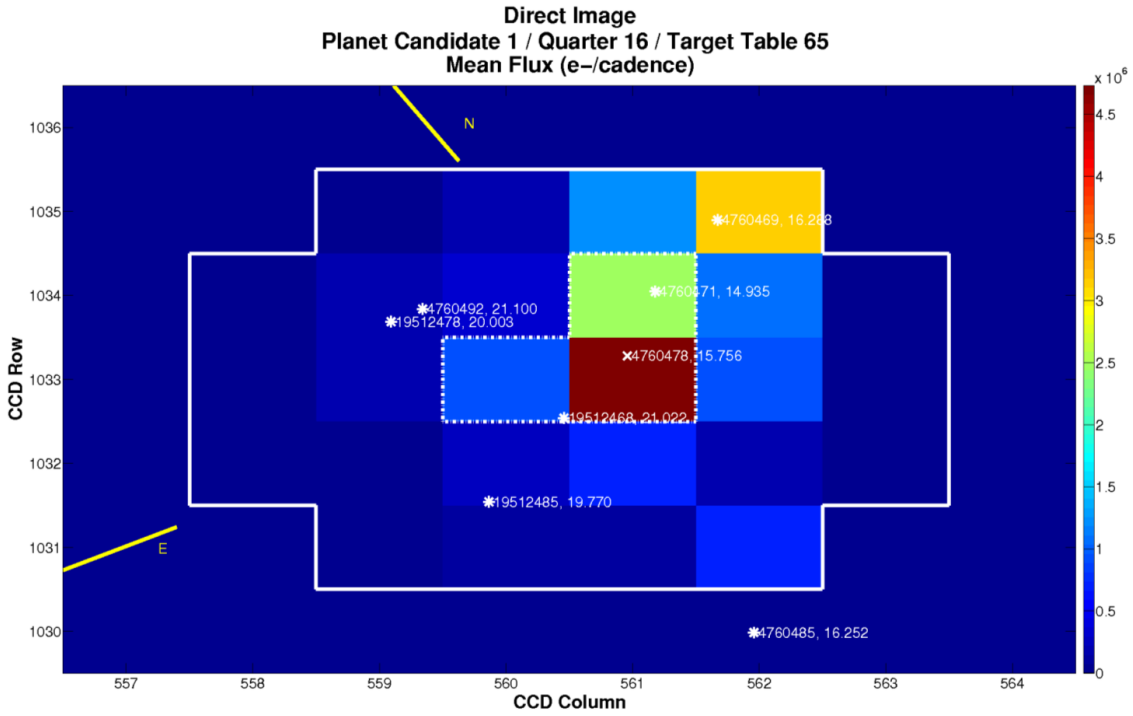

Figure 3.5: **The "Phantom" Star.** Model of the *Kepler* optimal aperture taken from the Data Validation Report (Q16 aperture), with the model star field overlaid. KIC 4760471 is clearly marked within the green pixel, but the star apparently does not exist.

Method-Marginalized Detrending the *Kepler* Data

A critical aspect of light curve analysis is the removal of systematic trends in the data. For the exomoon search this is especially important, as we must be careful to neither produce nor remove the subtle signatures of the moon which, unlike planetary transits, will not show morphological repetition from one epoch to the next. Various approaches to detrending have been utilized in the literature, but no single method is considered the gold standard.

The Hunt for Exomoons with Kepler (HEK) project developed the Cosine Filtering with Autocorrelation Minimization (`CoFiAM`) method Kipping et al. (2013), which was specifically designed for the moon search, as it focuses on preserving small / short duration features (possible exomoon transit signals) in the vicinity of the planetary transit. The algorithm represents the long-term trends as a sum of harmonic cosine functions with the longest period component being equal to the twice the baseline. The algorithm forbids components with periods (i.e. timescales) less than twice the transit duration, meaning that the Fourier decomposition of a transit is not disturbed by `CoFiAM`. This removes any long-term trends in the data, be they instrumental or astrophysical, while preserving short-duration events like those expected from an exomoon transit. We refer the reader to Kipping et al. (2013) for a more in-depth discussion of the method.

In addition, it is worthwhile to explore other detrending approaches to see what effect they may have on the final results, precisely because the signal we seek is so subtle. To that end we examined the results from a number of fairly standard detrending approaches: a polynomial fit, a local line fit, a median filter, and a Gaussian process. The polynomial method is identical to `CoFiAM` except we replace the basis set from cosines to polynomials, exploring up to twentieth order and selecting the locally-minimized autocorrelation result. The local line fit is a polynomial fit up to twentieth order but only training on data immediately surrounding (we used ± 80 hours) the transits and selecting the order which minimizes the Bayesian Information Criterion (BIC). The median filter uses a bandpass equal to five times the transit duration to remove long-term trends. The Gaussian Process regression adopts a

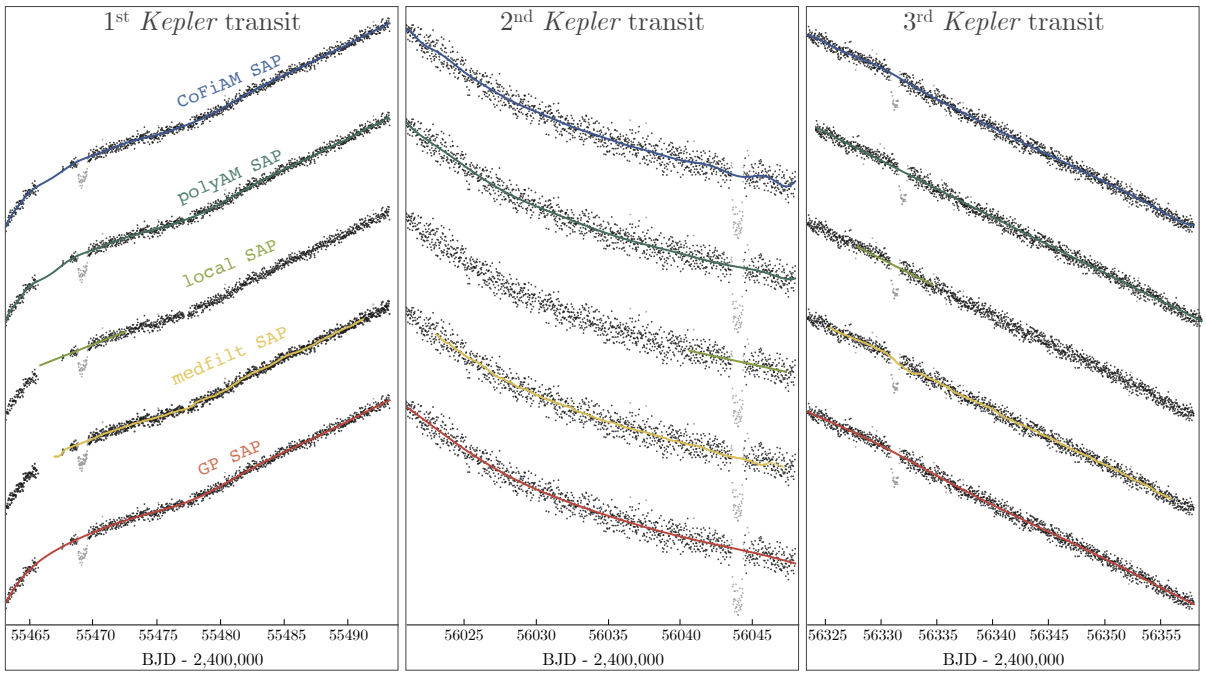

Figure 3.6: ***Kepler* detrending.** Comparison of five different methods used for detrending the `SAP` *Kepler* data. Baselines shown represent the full training set used, except for the `local` method which is trained on only data immediately surrounding the transits of interest.

square exponential kernel and trains on the entire quarter masking the transits.

The regressed trend models for the SAP photometry are shown on Figure 3.6. We detrended both the SAP and PDC data with all five methods, giving a total of ten light curves as shown in Figure 1 of the main text. A visual examination of these results suggests no clear favorite or obviously problematic detrending, but we may compare residuals from the difference of two methods to gain a sense of where and to what extent they differ. A matrix representation of these residuals can be seen in Figure 3.7, where the maximum standard deviation peaks at ~ 250 ppm, much less than the formal photometric errors on the *Kepler* data of ~ 590 ppm. This suggests that the result is robust against choice of detrending method.

We decided to marginalize over the different detrendings to create a robust light curve to work with in what follows. To do this, we took the five SAP light curves and computed the median flux at each time stamp. The PDC data was not used here because the data have

been corrected for contamination effects already and thus are expected to be slightly offset from the SAP results. The expectation then is that any anomalous features produced by one detrending method will be mitigated. The uncertainties on each data point are appropriately scaled up by quadrature addition of 1.4286 multiplied by the median deviation across the different methods at each time step (which had a median value of 34.6 ppm), thereby yielding uncertainties accounting for detrending differences. This effectively imposes a more stringent requirement for the moon to make its presence obvious in the data, since larger uncertainties provide more flexibility for the planet-only model. The final light curve is presented in Figure 1 in the main text.

Revised Photodynamical Fits of the *Kepler*-only Data

Although we have new HST data in hand, we decided to re-analyze the revised *Kepler* data in isolation before combining the data sets. The main purpose of this exercise was that later results can be placed in better context to assess which data set is primarily responsible for any interesting signals found. In Teachey et al. (2018), Bayesian photodynamical fits were conducted on Kepler-1625b to test for the presence of a possible moon. This led to a Bayes factor between the planet+moon (M) and planet-only (P) model which would be classified as "very strong" evidence on the Kass & Raftery (1995) scale ($2 \log K = 20.4$). The third transit in particular seemed to dominate the signal and the strong dependency of our inference on just a single epoch led us to conclude that the case for an exomoon remained ambiguous despite the high Bayes factor (and ultimately led us to pursue follow-up observations).

Using the revised SOC v9.3 data, we performed the same fits for models M and P, as well as a transit timing variation model, T, and a zero-radius moon model, Z. The evidences, derived using MultiNest Feroz et al. (2009), are presented in Table 3.4.

From this table, it is immediately apparent that the case for an exomoon is dramatically weakened using the revised *Kepler* data, going from $2 \log K \simeq 20$ to $2 \log K \simeq 1$. This naturally raises the question as to what exactly caused such a large decrease. In total, four

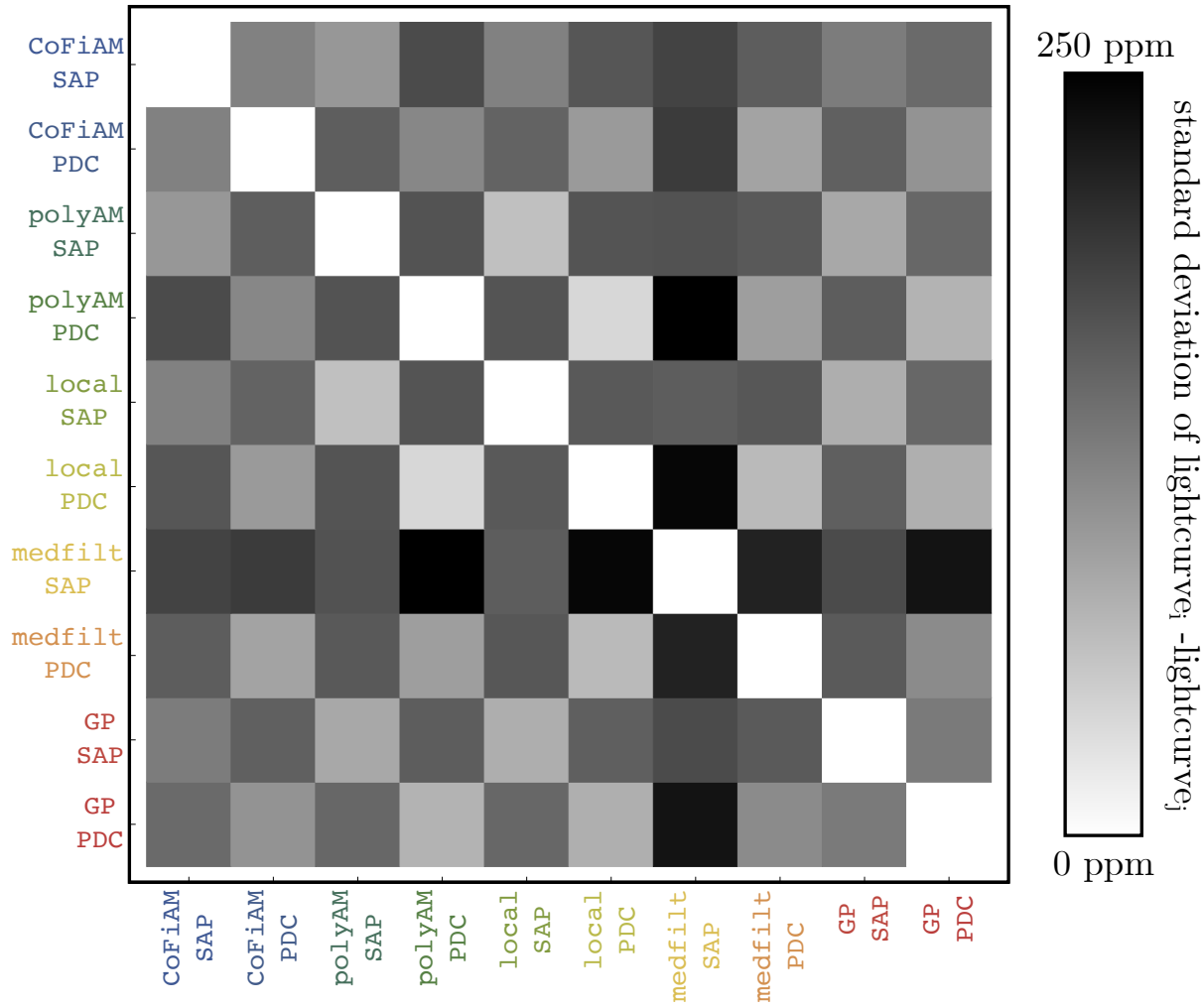

Figure 3.7: ***Kepler* detrending comparison.** Matrix plot of the standard deviations obtained when taking the differences between the resulting fluxes from each detrending approach. The plot is scaled such that white equals 0 ppm and black equals 250 ppm. For comparison the typical photon noise uncertainty in 600 ppm.

aspects of the analysis have changed:

1. Revised analysis used a slightly contracted baseline
2. Revised analysis used method marginalized detrending, rather than just relying on `CoFiAM`
3. Revised analysis used updated and thus different crowding factors
4. Revised analysis used the SOC v9.3 data products, rather than the SOC v9.0

In principle, one or more of these must be responsible for the decrease in the evidence for the exomoon hypothesis and here we attempt to distill by elimination which one(s) it is.

The baseline (which is the temporal window around each transit used in the MultiNest fits) in our revised analysis was contracted from the original analysis. Because Teachey et al. (2018) were blindly searching for moons in an ensemble out to 100 planetary radii from the host, transits were detrended with a baseline equal to an estimated 150 planetary radii. That calculation, detailed in Teachey et al. (2018), was not only based on now out-dated system parameters, but also is excessive for the purpose of validating the exomoon candidate Kepler-1625b-i which has a hypothesized semi-major axis in the range of 20 planetary radii. We estimated that 80 hours on either side of the transit would easily accommodate the features of interest in our new analysis (whereas 98.8 hours was used in Teachey et al. (2018)). Nevertheless, this difference could somehow explain the change and so we conducted a controlled experiment where we fit the exact original data from Teachey et al. (2018), but contracted the baseline to ±80 hours. Rather than reducing the evidence, this actually boosted it slightly, with $2 \log K$ increasing from 20.6 to 24.4 (here and in what follows we quote the Bayes factor between the M and P models only). This increase can be understood when considering the fact that a larger fraction of the data are affected by the moon-like signals. We also verified we recovered the same evidence as Teachey et al. (2018) with all inputs left the same in two independent fits, which both gave the same result to

within 0.6 in $2 \log K$. These tests demonstrate that the contracted baseline is certainly not responsible for the decreased evidence, as we might reasonably expect.

Our second hypothesis was that it was the different detrending algorithms used that caused the difference. We therefore compared the fit that used the original data but truncated baseline (from the previous paragraph), with a fit that was identical except that the photometry was detrended using the method marginalized approach, rather than `CoFiAM` alone (but still using the SOC v9.0 data products and blend factors in both cases). This again only led to a higher evidence for the moon, with $2 \log K$ now reaching 29.2. This therefore establishes that the decreased evidence for the moon model described at the beginning of this subsection, is not caused by either the baseline or detrending method used by Teachey et al. (2018).

This leaves two remaining possibilities: the blending factors and/or the actual photometric data products. We took the method marginalized light curve produced from SOC v9.0 and fit it in two ways; one using the crowding factors produced by v9.0 and one using those from v9.3 - but everything else kept the same. The former of these two fits corresponds to the $2 \log K = 29.2$ case of the previous paragraph. The latter case gives a greatly reduced $2 \log K = 14.8$. Thus, the contamination factors must be, in part, responsible - likely as a result of transit depth variations.

The other possibility was investigated by repeating the previous experiment but instead comparing to a fit where the blend factors are those of v9.0 but the data input to the method marginalized detrending algorithm comes from the SAP light curve of v9.3. This causes the Bayes factor to decrease from $2 \log K = 29.2$ to $2 \log K = 6.8$, a decrease even greater than that due to blending.

We are therefore able to deduce that our revised analysis of the *Kepler* data alone leads to decreased evidence for an exomoon as a result of two changes introduced between SOC v9.3 and v9.0: i) the new blend factors and ii) the new SAP photometry. The choice of baseline and/or detrending method are certainly not responsible. Comparing the light curve

from Teachey et al. (2018) and/or the revised method marginalized version on the same v9.0 data products reveals that the largest difference, versus the v9.3 product, is the third transit in quarter 16. This transit appears distorted and asymmetrical in the original analysis, explained by a large moon carving out flux around the planetary transit. The new data shows a much cleaner signal more closely resembling the other two transits.

3.6.2 Hubble Observations

Observation Design

In a planet-moon transit event, an in-situ observer would see the moon either leading or trailing its host planet, and in rarer cases the event could begin with both bodies passing in front of the star simultaneously. To adequately observe an exomoon transit it is therefore imperative that observations begin well before the anticipated planetary ingress and conclude well after planetary egress, as we expect to see flux reductions due to the presence of a moon in these parts of the light curve. The separation in time between the planet and moon ingresses / egresses is directly connected to their sky-projected separation, which will be a function of the moon's semi-major axis.

With a photodynamical moon model in hand from fits to the *Kepler* data it was possible to run the model forward in time to generate expected light curves for the October 2017 transit, though the morphology of the transit was poorly constrained when projected five epochs into the future. To determine the best start and end times for the HST observation we drew from the model posteriors and generated forward models based on these inputs, generating a range of possible outcomes. We set our start and end times such that exomoon ingress and egress features would be captured in the observing window to 95% confidence. We requested a start time as close as possible to Barycentric Julian Date ($\mathrm{BJD_{UTC}}$) 2458054.8 and ending no earlier than $\mathrm{BJD_{UTC}}$= 2458056.5.

For the observation we selected the G141 grism on Wide Field Camera 3 (WFC3), which is sensitive from ~ 1.1 to 1.7 microns. This choice was motivated in part by the expectation

that stellar variability, to the extent it is present, should be suppressed towards the infrared. In addition, because the observation would amount to some 40 hours on target, we knew that HST would pass through the South Atlantic Anomaly (SAA) for a significant fraction of this time, and WFC3 is one of the few instruments aboard HST that may be used during passage through the SAA. The grism creates a dispersion spectrum such that the light from each source is spread across the detector, providing spectral information on the target and its neighbors. Use of the grism also allows for longer exposures, as it takes much longer for a bright target to saturate. It is worth pointing out that time on target is of paramount interest in carrying out such an observation; we wish to minimize telescope overheads, which can include data readouts that interrupt the observation, and the RMS of the observation is directly tied to the time on target.

The drawback to using the grism is primarily the fact that a suitable roll angle for the telescope must be selected, one that minimizes overlap between the target spectrum and neighboring spectra. With each spectrum illuminating roughly 135 pixels in the direction of dispersion and several pixels on either side of the spectrum's central line, crowded fields must be modeled with care. Depending on the time of observation and the orientation of HST with respect to the Sun at that time, suitable roll angles may not be available. For the field containing Kepler-1625 we found only $\sim$ 20 of 360 degrees suitable for a grism observation.

We used the HST Exposure Time Calculator to plan the observations. Inputs were 1) the star's J-band magnitude (14.364±0.029) from 2MASS; 2) $E(B-V) = 0.19$, taking line of sight extinction A_V to target = 0.594 (NED value taken from Schlafly et al. (2011), and assuming $R_V = 3.1$; 3) a built-in Pickles model spectrum for a G0III star with $T_{eff} = 5610$ K (the closest match available to Kepler-1625); and 4) zodiacal light, Earth shine and air glow models for the date of observation, also built-in options in the ETC. We found that a 300 second exposure would fall well short of the time to saturation, which the ETC calculated to be 508.74 seconds. We opted not to utilize HST's spatial scanning mode, which moves the spacecraft perpendicular to the direction of spectral dispersion during each exposure,

thereby spreading the light onto more pixels. Spatial scanning was inappropriate for our observation, both because we were not close to reaching saturation (typically the primary motivation for spatial scans), and because the crowded field meant the spatial scan would have to be extremely short, and potentially unachievable by the spacecraft given the length of each exposure. We note also that the software provided by STScI for reduction of the HST data does not currently support spatial scanning mode.

The standard data reduction pipeline distributed through STScI, aXe Kümmel et al. (2009) requires at least one direct image of the target (i.e. without the grism) so that spectral calibrations can be made. Using source locations derived from the direct image using SourceExtractor Bertin et al. (2010), the aXe software is designed to calculate the position of each spectrum, the wavelength solution for each pixel, and contamination from nearby sources. A single direct image using the F130N filter was made at the start of the observing run with an exposure time of 103.129 seconds, followed by a total of 232 grism exposures. All data were taken in the 256×256 subarray to reduce overheads. Each grism exposure lasted 290.776 seconds, well short of the saturation time. `SAMP-SEQ` was set at SPARS25 and `NSAMP` was 14, meaning there were 14 non-destructive readouts of each exposure at roughly 25 second intervals. These multiple readouts are useful in part for rejecting cosmic rays within the front-end data processing pipeline `calwf3`.

Roughly 3100 seconds, or 51.67 minutes, were available for observing the target during each orbit. The remaining time in each orbit (about 44 minutes) were unusable as the target was occulted by the Earth. For all but two orbits, there were only 72 seconds of unused visibility time. For the first orbit 130 seconds went unused, and for the 15^{th} orbit, 270 seconds went unused. This was due to increased overheads, namely a full guide star acquisition, which takes longer than re-acquisition and reduces the number of exposures that can be made in an orbit. Each exposure was long enough that data dumps could be made in parallel; therefore, there were no gaps between the first and last exposure in a given orbit.

It was found that only three of our exposures would occur during passage through the SAA, which was fortuitous. For most of the observation SAA passage was restricted to times when the target was occulted by the Earth, when no data could be taken. For these SAA-affected exposures the Fine Guidance System (FGS) cannot be used; instead, HST must use gyro control to stay on target. The gyros are known to have a pointing drift on the order of 1.5 mas per second, meaning that for our ~290 second exposure we could expect the target to drift by approximately half an arcsecond.

Execution

Observations began at 6:52:15 (UTC) on 28th October 2017 and ended at 23:20:08 on 29th October 2017. No malfunctions of the spacecraft were detected. Three of the 232 exposures were indeed affected by passage through the SAA, in line with expectations. While the SAA had no discernible effect on the photometry itself, pointing drifted considerably during these exposures, as expected. The spectra from these images, landing on neighboring pixels and significantly dilluted by smearing, showed significantly different flux levels, even with appropriate pixel sensitivity corrections, and were therefore left out of our final analysis.

The final orientation of the spacecraft (provided in the header as the position angle of HST's V3 axis) was ~ 249 East of North and produced a clean spectrum of the target with minimal contamination from nearby stars.

Data Preparation

For initial processing of the raw data files from STScI we followed the WFC3 IR Data Reduction Cookbook, using updated configuration and reference files where available, and making minor updates to source code when deprecated packages were still in use and caused fatal errors. We refer interested readers to the Cookbook for details on the data reduction. We utilize the `.flt` files produced by the front-end `calwf3` pipeline, which account for a number of known systematics and perform cosmic-ray rejection. We point out that while

the `calwf3` performs a flat field correction, for grism observations the flat field division is unity everywhere, as the flat fielding for grism images is intended to be performed later as a part of the spectral extraction process.

The standard aXe software is unable to handle images taken in subarray mode, so each 256×256 image must be embedded in a larger 1014×1014 array before processing. It is important that the images be embedded in the correct position, as subsequent flat fielding and background removal are performed on a pixel-by-pixel basis. We modeled our own embedding code after a python script provided by STScI, which we were unable to run on our machines. We verified the embedding was done correctly by eye, matching obvious artifacts in the master sky image and the flat field cube to the same artifacts that are readily apparent in the images pre-correction. With this step completed we could follow the instructions of the cookbook.

Unfortunately we were unable to follow the prescribed reduction beyond the `axeprep` stage of the cookbook, as we experienced a persistent breakdown with the `axecore` task which we were never able to resolve. The `axecore` tool is responsible for automated spectral extraction. From here on we wrote our own code to analyze the data.

Following the STScI prescription, the background flux levels are calculated by finding the median flux of the image, masking out pixels with greater than 5 electron counts. The Master Sky Image Kümmel et al. (2011), calculated from in-orbit observations, is then scaled up by the median background flux level at each time-step and subtracted. This step removes the well-documented "breathing" effect that appears in HST time-series observations, which has been attributed to thermal fluctuations in the telescope as it orbits the Earth. An additional flat-fielding must then be performed, using the flat field data cube produced from pre-flight laboratory testing. The flat field cube encodes four polynomial coefficients at each pixel location to model the pixel sensitivity as a function of wavelength.

The `axecore` task is designed to calculate a wavelength solution for each pixel in the spectrum, and with this information a wavelength-specific flat-fielding can be performed.

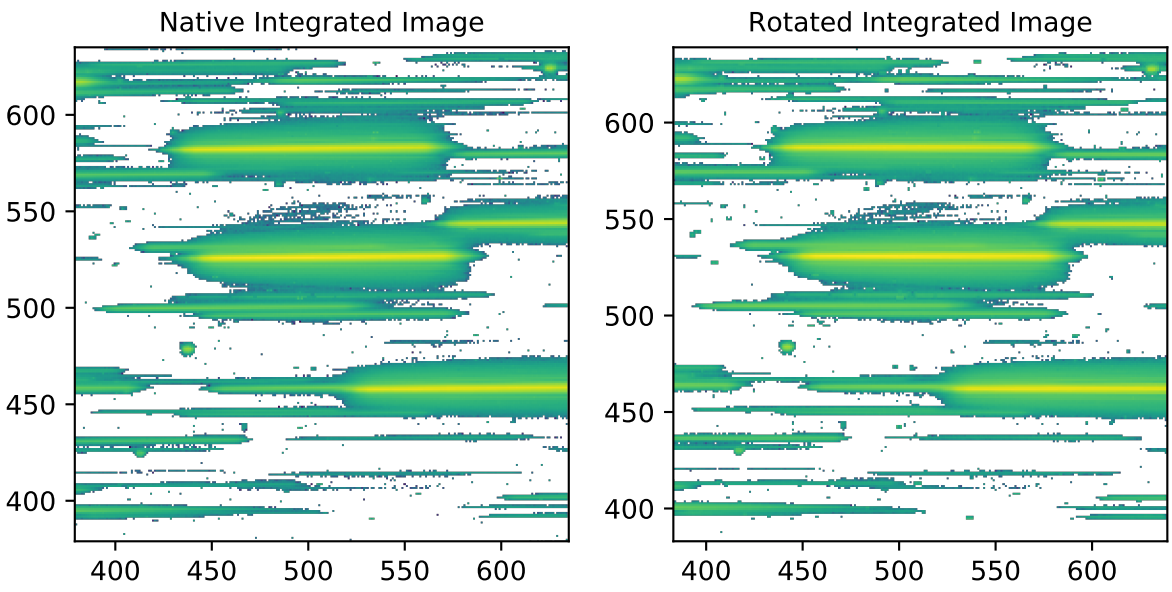

Figure 3.8: **HST image rotation.** *Left:*Integrated HST image with the spectra natively inclined 0.5 degrees with respect to the x-axis. *Right:* the rotated image for simplifying the spectral extraction. Pixel values are logarithmic to show the full extent of the spectra; white space indicates backgrounds integrating to values < 0, for which the logarithm is undefined.

However, due to the code breakdown we did not have wavelength solutions in hand until later in the analysis. While the pixels show significant sensitivity variation across the detector, we found there was minimal variation across the wavelength range for any given pixel (typically less than 1%), so we opted to compute the median sensitivity for each pixel and we used these values to perform the flat-fielding. We are thus marginalizing over intra-pixel sensitivity variations, but we point out that STScI documentation also suggests G141 grism observations may be flat-fielded using the F140W flat field which also does not encode intra-pixel sensitivity variations. We compared our median G141 flat field image to the F140W flat and determined that flattening the G141 cube was a more faithful rendering of the wavelength-sensitivity information for the grism.

Imputation of Bad Pixels

Following the flat field corrections there remained $O(10^3)$ pixels that were clearly outliers, as evident from a median stack of the images. Wherever possible, we elected to perform

imputation of all time stamps associated with these outlier pixels. The tasks of outlier identification and imputation are treated independently.

To identify the outliers we start by calculating median pixel values across our observations, which should reveal pixels that behave anomalously in more than 50% of the exposures. This superstack is done on all grism images, except for numerous images which were identified during preliminary analyses as behaving in a non-representative way: namely cadences 107, 116, 125 and 126; as well as orbits 1 (telescope settling), 7 (transit ingress) and 18 (transit egress). Each superstack pixel has an uncertainty equal to 1.4286 multiplied by the median absolute deviation (MAD) across all times.

Since the grism spreads the target spectrum along the pixel rows, we expect (and indeed observe) that the observed flux follows a smoother pattern along the rows than the columns. Exploiting this fact, we extract each row's flux vs column index from the superstacked image and look for outliers along each row independently.

We do this by constructing a 3-point moving median and then training a Gaussian process (GP) with a squared exponential kernel on the result. This GP is then evaluated on all column indices, computing residuals as we move along. A median-version of the reduced χ^2 is computed to re-scale. We define residuals as the difference from the GP normalized by the pixel's uncertainty. We also scale the uncertainties such that the median version of the residual's reduced χ^2 equals unity. Outlier pixels are then flagged as those which depart more than $10\,\sigma$ (chosen after some experimentation) away from the GP model. Note that our GP lacks predictive power on the edge columns and this process does not consider column pixels during the search.

In addition to this procedure, we flag pixels as an outlier if the pixel's derived uncertainty across all images exceeds $20\,\sigma$ of the median error of the row's pixels, where σ is again coming from another MAD.

If a pixel is flagged as an outlier but has an immediate neighbor of similar flux, we remove the outlier flag. This is done by first computing the maximum deviation of each superstacked

pixel with its row-wise precursor and successor and then seeing if the candidate outlier is less than $10\,\sigma$ away from the median deviations seen in that row (where again σ comes from the MAD). This was necessary to avoid killing zeroth-order spectrum features which look like islands of outliers in a single row (though we do not attempt to use zeroth-order features for analysis, as the target's zeroth-order is off the detector).

Imputation is not performed using medians as this essentially represents a zeroth order polynomial which is not sufficient to capture the gradients observed across the pixels. Instead, for each pixel in the image we produce a predicted flux based on a 1-dimensional spline interpolation with two pixels preceding and two pixels following it in the row. We found that a 1-D row interpolation is superior to a 2-D interpolation, as the latter does not adequately handle pixel replacement across the spectrum peaks, that is, perpendicular to the direction of dispersion. If the pixel has been flagged as an outlier it will be replaced with the predicted flux. An exception to this rule is if one of these four training pixels is itself an outlier. In such cases, we flag the pixel as an irreplaceable outlier.

After the first round of outlier identification and imputation, we ran the algorithm a second time. This process led to the identification of 1756 outlier pixels, of which 634 were irreplaceable whose fluxes were simply set to NaN after this point to mask them.

Finally, we note that the spectra produced by the grism are inclined 0.5 degrees with respect to the pixel grid, per WFC3 Grism documentation. We therefore use a standard SciPy package Jones et al. (2001) to rotate the image clockwise by 0.5 degrees, performing a 3rd-order spline interpolation, thereby aligning the spectra with the x-axis. This simplifies the extraction of the spectra considerably, as the optimal aperture may be neatly aligned with the image grid, and produces no discernible artifacts in the spectra. The pixel errors must also be rotated, which is potentially problematic if the errors across the image were random. However, since the errors scale predictably with flux levels this rotation is also well behaved, and the resulting distribution of errors across the image is unchanged from the native images.

Optimal Aperture

The target's point spread function is centered in the rotated images at approximately $x = 515$ and $y = 531$ (where x represents the column index and y the row index). To extract photometric time series, we elected to employ simple aperture photometry rather than modeling the complex point spread function (PSF) observed. This is well justified since the high angular resolution of HST, combined with our observational design, means that we do not see significant overlap of neighboring sources with the target.

In choosing an aperture, we could simply draw a broad box around the target by hand, but instead we elected to choose an optimal aperture which minimizes the scatter in the final target light curve. The optimal aperture was found in a two-step process. First, we setup a grid of 105,840 candidate apertures where each permutation has a unique aperture defined by four parameters, $\{x_{\rm min}, x_{\rm max}, y_{\rm min}, y_{\rm max}\}$, such that the aperture is bounded by $x_{\rm min} \leq x \leq x_{\rm max}$ and $y_{\rm min} \leq y \leq y_{\rm max}$. The grid of candidates spans the range $426 \leq x_{\rm min} \leq 480$, $481 \leq x_{\rm max} \leq 599$, $515 \leq y_{\rm min} \leq 530$, and $532 \leq y_{\rm max} \leq 544$, where we step between the extrema in 2-pixel intervals. We remind the reader that these pixel values do not correspond to the native images from HST, but to our rotated image, for which there is an offset. In each of these candidate apertures, we extract a white light curve for the target and correct for the hook effect using a simple exponential ramp (explained in detail in Section 3.6.2). While we eventually developed our own approach to removing the hook trends, this simple model does reasonably well correcting for charge trapping, and thus observations are expected to be stable within each orbit, although visit-long trends have not been corrected for at this point.

As visit-long trends persist, and there are of course flux decreases caused by Kepler-1625b's transit, as well, comparing the raw root-mean-square (RMS) of each candidate aperture's is not an appropriate cost function to score the different apertures. Instead, we reasoned that if we mask the times during the ingress and egress of the planetary transit (which take up one HST orbit each; orbits 7 and 18 respectively), then we should expect

the photometry to be stable within each orbit (but not necessarily between each orbit). We further mask the first orbit, which appears to represent a settling-orbit for the photometric behavior and is typically discarded in similar studies. Finally, we also mask time stamps 107, 116, 125 and 126 where we later came to suspect outlier behavior. The remaining 202 points (of the original 232) are then grouped into their respective orbits and the RMS of each is computed. We then define a cost function as the mean of these RMS values (23 in total).

This process identified an optimal aperture defined by $456 \leq x \leq 581$ and $526 \leq y \leq 542$. However, the grid search used a resolution of 2-pixels in its search and further more used a fast but sub-optimal hook correction method. As described in Section 3.6.2, we found a novel non-parametric hook correction method is able to out-perform the exponential model and better capture the sharp hook morphology. We therefore performed a second-stage in our search where we essentially walk the aperture in 1-pixel intervals away from the previously found solution. Each bound (i.e. $x_{\rm min}$, $x_{\rm max}$, $y_{\rm min}$, $y_{\rm max}$) is perturbed by ± 1 pixel to create 8 candidate grids, as well as the original solution to give a ninth. Across these 9 possibilities, we extract photometry as before but this time perform the more computationally intensive non-parametric hook correction described in Section 3.6.2. If a better aperture is found amongst the 9 options, we walk to that solution and repeat, else we stop.

In practice, we perturbed the optimal aperture from stage one adding a random integer between -2 and +2 to each bound and walked from that position, in order to test if the walker would return to the same solution. Indeed this is what happened and the final optimal aperture returned to $456 \leq x \leq 581$ and $526 \leq y \leq 542$, which has mean intra-orbit RMS of 375.5 ppm. In what follows, we set value that value, 375.5 ppm, as the standard photometric error for this optimal time series.

Modeling the Hooks

A well-known feature of time-series observations on HST are the exponential ramps or hooks, e.g. Berta et al. (2012); Deming et al. (2013); a phenomenon that's also been observed in *Spitzer* data, e.g. Deming et al. (2006); Knutson et al. (2007); Charbonneau et al. (2008); Agol et al. (2010). As the observation begins, the flux readings ramp up with each subsequent exposure towards a saturation asymptote. This is thought to be due to charge trapping in the detector Agol et al. (2010). Once the observations are interrupted, either due to occultation of the target by the Earth or through a non-parallel data dump, the ramps resume.

Common previous approaches for removing this systematic include templating the ramps from the out-of-transit orbits to detrend them in all other orbits Berta et al. (2012), and assuming a parametric model fit (typically using exponential functions) to each ramp for removal Agol et al. (2010). The templating approach is not ideal for our observations since we do not know which orbits are in- or out-of-transit *a priori*, due to the candidate exomoon. For this reason we initially pursued an exponential ramp model. The results from this approach were certainly reasonable and provided a clear transit recovery. Despite this, in our quest to extract as much information as possible from these observations, we devised an alternative strategy that ultimately provided a superior correction.

The inspiration behind our new approach can be seen in Figure 2 of the main text. Each orbit is typically comprised of 9 exposures, which are shown with distinct colors in the figure, and can be labelled with the index $\nu = 1, 2, ..., 9$. If one considers just the $\nu = 1$ points, the light curve appears remarkably clean, and the same argument holds true for any specific choice of ν. This is understandable if we consider the fact that each observation ν shares a common observational history; that is, the $\nu = 1$ exposures all occur immediately following target acquisition after occultation, the $\nu = 2$ exposures all have a common history following $\nu = 1$ observations, with all the charge trapping associated with those exposures, and so on. These common histories will then act as a baseline flux level for each ν that may be independently corrected. We therefore hypothesized that these nine light curves could

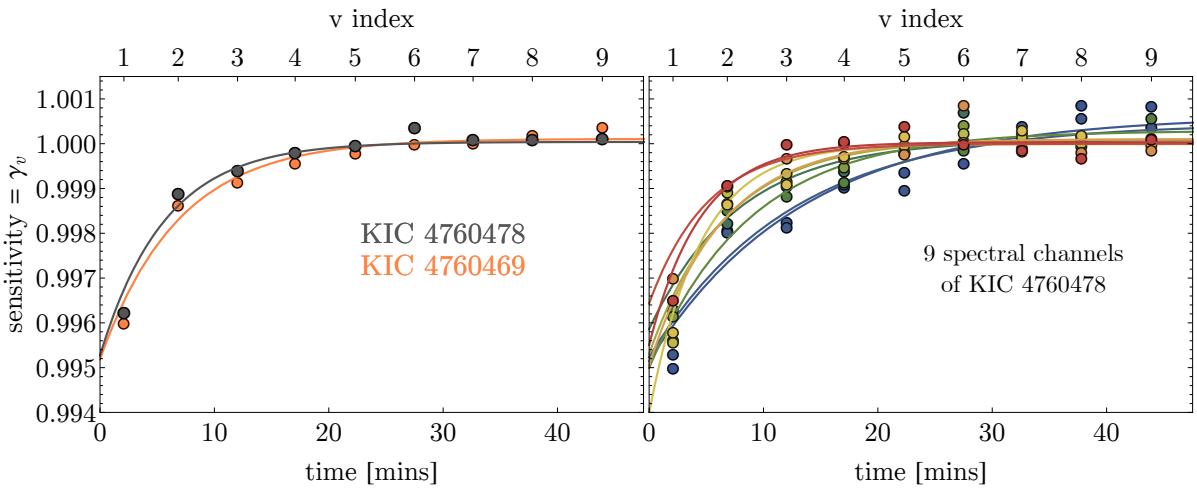

Figure 3.9: **HST hook model comparison.** *Left*: Comparison of the two models for the WFC3 hook; an exponential ramp fit (solid) and a novel discrete model introduced in this work (points). The left panel shows the results from the white light curve of the target and a comparison star. *Right:* Same as left except we show the 9 different spectral elements used in this work for the target. The v index is the exposure number within a given orbit.

be combined by simply scaling them independently to create one coherent light curve. We assign 9 scaling factors, γ_v, to each and treat these as unknown parameters to be solved for.

Following our earlier argument in Section 3.6.2, we expect the hook-corrected light curve to exhibit stable intra-orbit photometry (but not inter-orbit). We therefore use the same cost function as used earlier, namely the mean intra-orbit RMS. We iteratively optimize for the γ_v terms until the cost function improves negligibly. The final γ_v terms are shown in Figure 3.9, where we also overplot the optimized exponential ramp model for comparison. This plot reveals that the exponential ramp model is not able to fully capture the very sharp turn-on of the hook. The exponential hook correction light curve is also shown in Figure 2 of the main text, where the mean intra-orbit RMS is considerably higher at 440.1 ppm (versus 375.5 ppm). Nevertheless, an exomoon-like decrease in brightness is observed in both versions following the planetary transit.

We highlight that our non-parametric approach is somewhat guaranteed to out-perform the ramp model due to more degrees of freedom. However, the ability to capture sharper hooks, combined with the more agnostic nature of the method's assumptions ultimately

led to us to use this method for our final hook-correction. We highlight that 375.5 ppm per 300 seconds corresponds to 154.8 ppm per *Kepler* long-cadence, which is 3.8 times lower than the median *Kepler* uncertainty resulting from our method marginalized detrending (589.9 ppm). Thus, HST greatly out-performs *Kepler* on this target. For this reason, one might reasonably expect that the HST data will be the dominant transit epoch for constraining putative moons.

From Figure 2 in the main text, an apparent decrease in brightness is evident with both versions of the hook correction for Kepler-1625, occurring a few hours after the primary transit has finished. The precise shape of this moon-like dip appears dependent upon the trend assumed in the data, which is discussed in detail in the main text.

Wavelength Solution

To derive a wavelength solution for each pixel we extracted the spectra for the target and comparison star and found the best fits (minimizing χ^2) to a model spectral profile (Figure 3.10). The model is produced by multiplying the G141 response function by a blackbody curve, the latter produced using published values for the stars' effective temperatures. By examining the wings of the model curve (where sensitivity falls off rapidly) and comparing to the extracted spectra the fits are excellent. We note however that this simple model overpredicts the flux at shorter wavelengths and underpredicts at longer wavelengths, as can be seen in Figure 3.10. We examined whether this discrepancy could be due to our marginalizing over the wavelength information in the flat field, as there is a distinctive wave-like structure in the pixel sensitivities that propagates across the detector in the direction of dispersion. In some places on the detector shorter wavelengths are more sensitive than longer wavelengths, where in other places the reverse is true (see Figure 3.11). As we marginalize over the wavelength information in the flat field, this information is lost.

However, we find that this intra-pixel flat-field structure cannot explain the spectrum-model discrepancy in Figure 3.10, as flux errors across the spectrum average out to be less

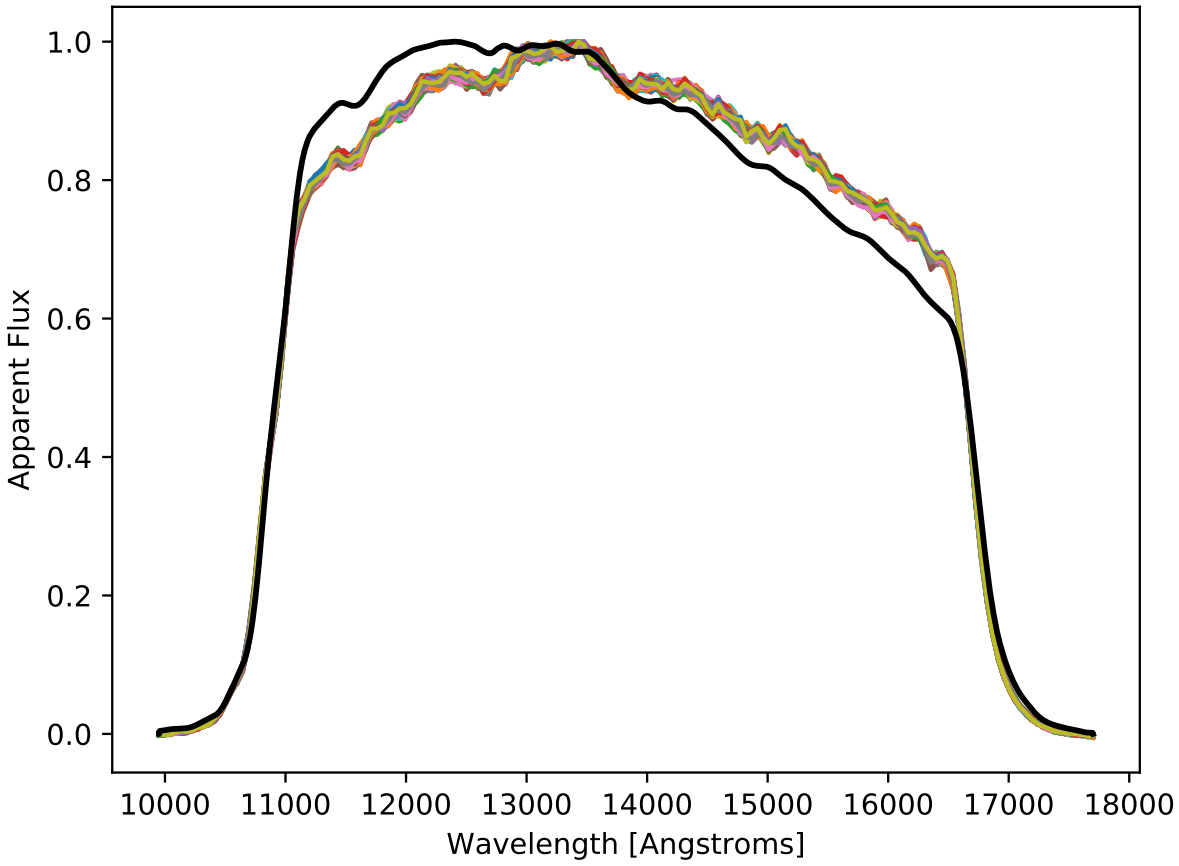

Figure 3.10: **Wavelength solution.** Calculated blackbody for Kepler-1625 multiplied by the G141 response function (black). The spectrum of Kepler1625 extracted from each HST image is overlaid in multiple colors (for each exposure). Both curves are normalized by dividing out their maximum values. Note the overprediction of the model at shorter wavelengths and underprediction at longer wavelengths.

than 0.2% for any flux bin when accounting for the wavelength dependence, whereas the discrepancies are clearly well in excess of that. We speculate that the discrepancy merely arises from the fact that stars are not perfect blackbodies. In any case, these discrepancies do not invalidate the wavelength solution, as the rapid fall-off in sensitivity at either side of the spectrum clearly matches the sensitivity curve very well.

Nearby Uncatalogued Source

Upon inspection of the HST images it became clear that a previously uncatalogued point source is present in close proximity to the target. The object shows up in every image obtained by HST and cannot be an artifact on the detector, as it moves like all the other point sources during the three SAA-affected exposures. By every indication it is a another spectrum for a nearby point source.

To estimate its position on the sky we simply take the first pixel along the spectrum for which Flux > 0. Note that this is *not* the sky position of the star. However, we may do the same for the target star and nearby KIC 4760469, for which the sky separation and position angle is known, and thereby orient ourselves to calculate the sky separation and position angle of the uncatalogued source.

To calculate the magnitude of the new source, we step through each column along the spectra and fit a three-Gaussian model to the data. One Gaussian is fit to the new (contaminating) source while two Gaussians are fit to the target – one narrow, and another wide. The combination of these two Gaussians do an excellent job fitting the peaky-but-broad profile of the target star, with the wider component accounting for what appears to be flux bleeding into neighboring pixels (see Figure 3.12). Having stepped through every column, the areas under these three Gaussians can be computed, and the flux of the uncatalogued source may be compared to that of the target. With this information the relative magnitude with respect to the target may be computed.

We find that the star is located $\sim 1.78''$ away from the target at Position Angle 8.5 degrees

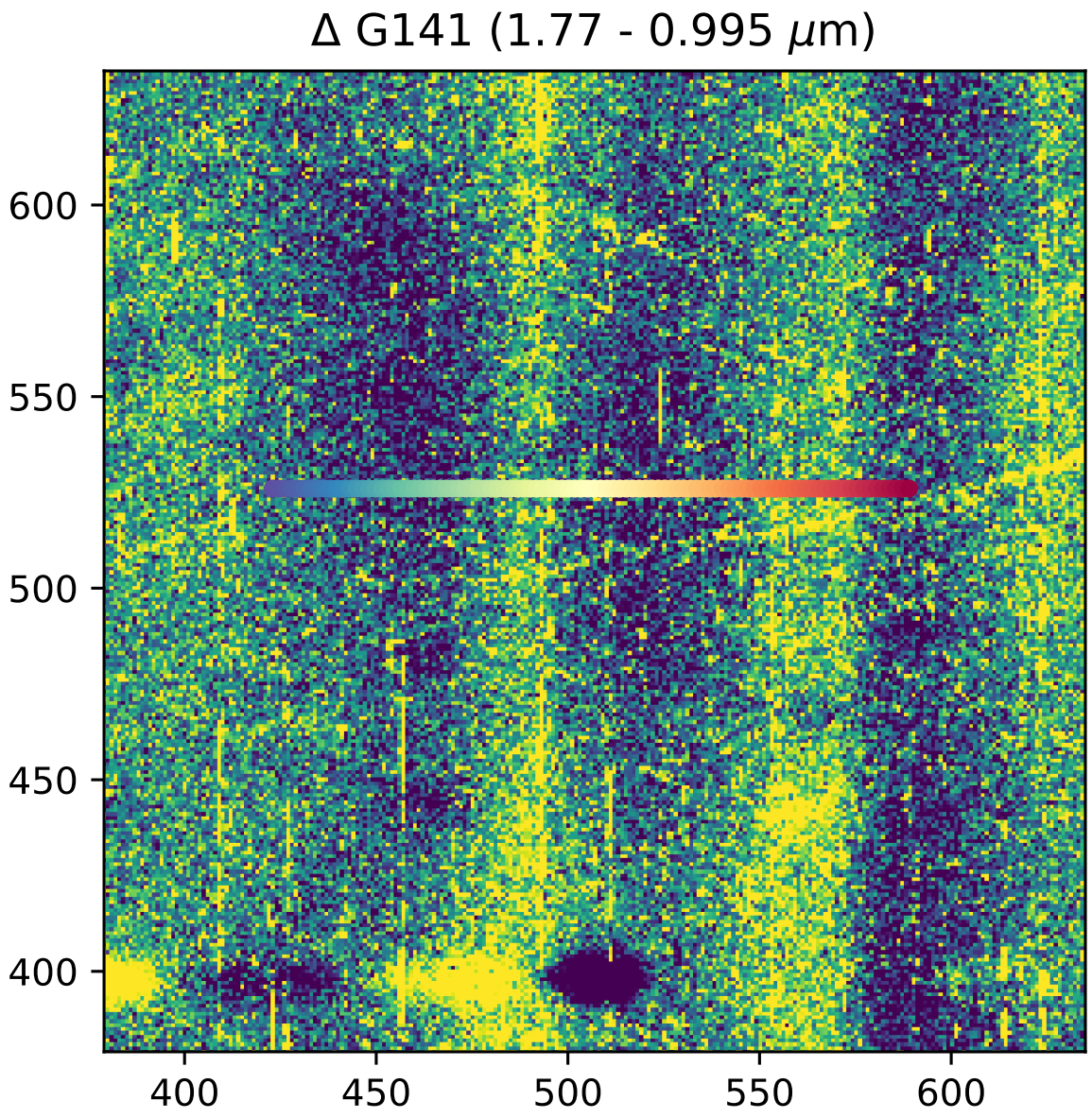

Figure 3.11: **Wavelength-dependent pixel sensitivity**. Change in pixel sensitivity across the G141 wavelength range. The color range is ± 1%. Towards the purple end short wavelengths are more sensitive than longer wavelengths, while at the yellow end longer wavelengths are more sensitive. The full width of the target spectrum's response function is overplotted.

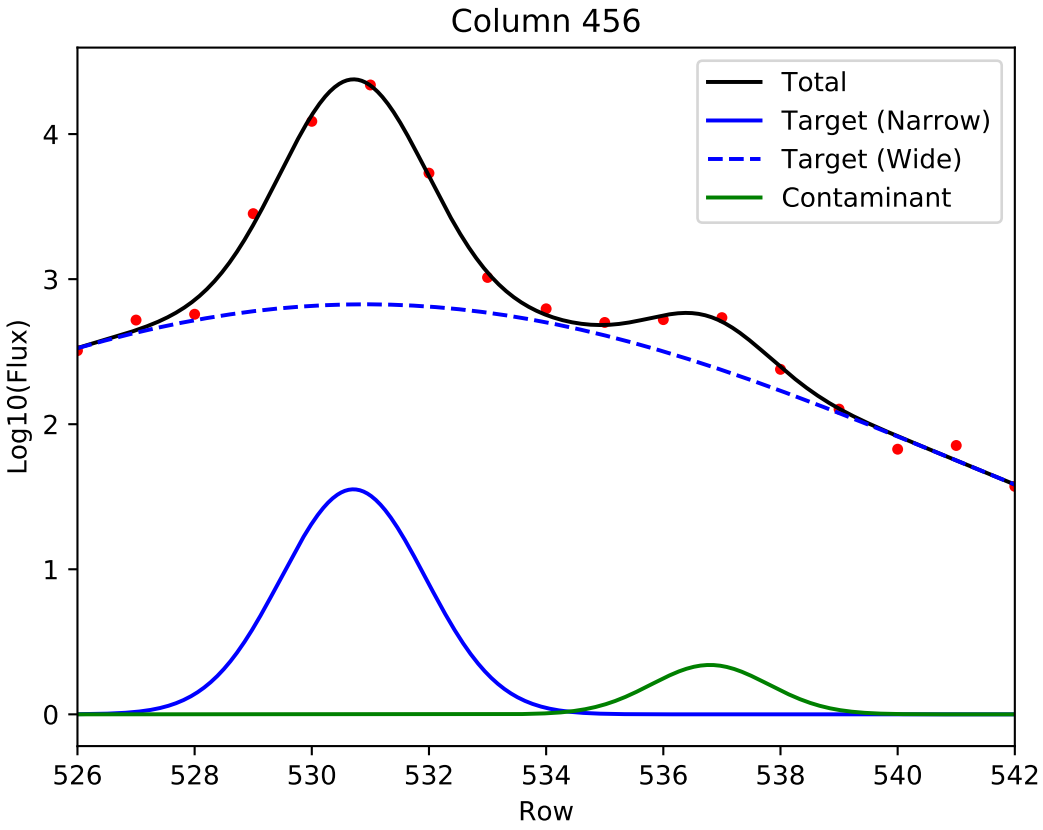

Figure 3.12: **Modeling the uncatalogued source contamination.** A single column within the HST optimal aperture, fitting three Gaussians to the source and the contaminating uncatalogued source. Pixel fluxes are shown by the red data points. The blending is calculated by taking the inverse of the starlight fraction within the optimal aperture originating from the target star.

East of North. We compute a *Kepler* bandpass magnitude of 22.7. To compute a blend factor we take the same approach as before, i.e. modeling the target and the contaminant with three Gaussians, only now we restrict the window to the optimal aperture see Figure 3.12. We compute a blend factor of 1.000328, indicating an extremely small contribution from the uncatalogued source on our light curve. Furthermore, an extracted light curve from the uncatalogued source shows a variability of 0.33%, so its contribution to the target light curve is 1 ppm. We therefore ignore it in subsequent analysis.

Using a *Gaia*-derived distance of (2460 ± 220) pc to the target, we may calculate the physical separation of this uncatalogued source from Kepler-1625, under the assumption that the two objects are at the same distance. We find that separation would be ~4400 AU, placing it well within the gravitational influence of Kepler-1625. It is however impossible with the data in hand to determine whether these objects are in fact physically associated.

Visit-long trends

In addition to the breathing and hook effects, one other well-known source of systematic error requires correction - the visit long trend Wakeford et al. (2016). Visit-long trends with WFC3 are typically modeled as a linear slope (e.g. see Huitson et al. (2013); Ranjan et al. (2014); Knutson et al. (2014)). However, our observations are unusual in that they span 40 hours, far more than the few hours used when observing transiting planet on short orbital periods. These trends have not yet been correlated to any physical parameter related to the WFC3 observations Wakeford et al. (2016), and indeed not all observations appear affected. Simple inspection of our white light curve, using either the exponential or non-parametric hook correction, show clear evidence for a visit-long trend (see Figure 2 in the main text).

Although a linear trend is the most common approach (e.g. see Huitson et al. (2013); Ranjan et al. (2014); Knutson et al. (2014)), we note that Stevenson et al. (2014a,b) report improved fits using a quadratic model and so we considered both models in this work. We further extended our investigation to include an observation-long exponential ramp model.

This last model is motivated by visual inspection of the light curve, which appears to ramp up and flatten, as well as the asymptotic behavior it introduces which is more physically motivated than an ever-increasing/decreasing trend. We speculate that it could perhaps be caused by the same charge trapping that causes orbit-long ramps, only operating on a much longer timescale.

If we look at specific spectral channels, rather than the white light curve, a flux offset occurs between the 14^{th} and 15^{th} orbits for the reddest wavelengths (see Figure 3.14). This moment in time corresponds to the HST visit change, during which the spacecraft performed a full guide star acquisition at the beginning of the 15^{th} orbit, which placed the target spectrum $\simeq$0.1 pixels away from where it appeared during the first 14 orbits. This discontinuity appears in the raw photometry mid-planetary transit and is potentially problematic due to the fact that it can mimic a moon signal. Any model placing the moon at the kink would be immediately suspect.

In the white light curve, a flux offset at this time is barely noticeable but since the red spectral channels contribute to the white, then we deemed it necessary to allow for an offset term in our three trend models.

Since only centroid position changes during the visit switch, whether the visit-long trend be astrophysical or due a long-term instrumental effect, there is no reason to expect a different functional form or function parameters to become introduced at the instant of the visit change. For this reason, we generally expect a smooth continuous function (such as the linear, quadratic or exponential models) but with an offset term to account for any remaining pixel sensitivity variations. We therefore did not consider models described by two completely independent polynomials, for example, on either side of the visit change.

Since the visit-long trend occurs on a long timescale, it is inextricably mixed with the transits of the planet and possible moon. In such a case, strong covariances are expected between the trend parameters and the transit parameters and thus joint fitting is required. Our joint fits are described later in Section 3.6.3, but for now we point out that all six

visit-long trend models were regressed in conjunction with the transit models considered (e.g. models M, P, T and Z), but independent of one another. In this way, we can rank the different approaches based on their Bayesian evidences, as well as the resulting associated physical parameters.

The results of these fits using the planet+moon (M) transit model may be seen in Figure 3.4, with Bayesian evidences tabulated in Table 3.3. All of the fits are able to explain the previously noted decrease in brightness towards the end of the observations as being due to a moon transit, although the duration and depth of the event vary somewhat between the three trend models. In the exponential trend model the moon fully egresses before the end of the observation, while in the linear and quadratic fits the moon is still in the process of transiting. However, in virtually every case the dip remains discernible after detrending has been performed. Note that all of these models were found to be preferable to any other transit model (e.g. T, Z or P) attempted using the Bayes factor (we direct the reader to Section 3.6.3 for a more detailed discussion of model comparison).

Is the moon-like dip instrumental or astrophysical?

As noted earlier, there appears to be a decrease in flux in our WFC3 photometry towards the end of our observations. Since we are primarily interested in the possible existence of exomoons in this work, that decrease is of particular importance as it will greatly affect photodynamical model fits, if real.

To assess whether this dip is instrumental or astrophysical in nature, we considered three tests: i) inspection of the comparison star ii) inspection of the centroids iii) a chromatic test. The chromatic test is described later in Section 3.6.3, but we here describe i) and ii) in more detail than possible in the main text.

The only other bright star with a full spectrum in the HST images is KIC 4760469. This star was not observed by *Kepler*, but is listed as a $0.84\,R_{\odot}$ 5555 K main sequence star in the KIC. Accordingly, it is expected to be photometrically stable and provide a good test for

our correction algorithm. We therefore applied the exact same routines to this star as was done for Kepler-1625.

As seen in Figure 3.2, the comparison star displays no obvious long-term trends that might be attributed to instrumental systematics, apart from a gentle upward slope at the beginning of the observation which is also seen in the light curve for Kepler-1625b and is cleanly corrected with our detrending. We used MULTINEST to compare the linear, quadratic and exponential trend models and found all three were similar in evidence, with the quadratic model being slightly favored.

In terms of assessing if the moon-like dip associated with Kepler-1625b is real or not, the relevant region is visible from BJD 2,458,456 of Figure 3.2. The photometry appears quite stable at this time and certainly no indications of an instrument-induced flux change.

The second check we performed was to look at the centroids. Column (x) and row (y) centroids were computed from the optimal apertures of the target and comparison star and are shown in Figure 4.3. We highlight that the y position has been decreased by 0.1 pixels after the visit change (for both sources) to fit them on the same scale, but the x position has not been altered. Both the target and the comparison star show nearly identical centroid behavior, as expected.

The moon-like dip occurs across several orbits and thus the only way centroid variations could explain the dip would be an inter-orbit centroid variation. We therefore take the median centroid position of each orbit as binned points, shown by the black data in Figure 4.3. Aside from the visit change shift, there are no substantial changes in the inter-orbit centroid position, and certainly not around the time of interest (highlighted by the vertical lines).

Comparing the centroids to the flux variations observed can also be used to gauge how feasible it is that the moon-like is a product of these clearly small centroid changes. From our later fits of the target, we find that the visit-change flux offset is (330 ± 120) ppm, (180 ± 190) ppm and (220 ± 140) ppm for the linear, quadratic and exponential models respectively. That corresponds to a decipixel (0.1 pixels) change in the y-position and about half that in

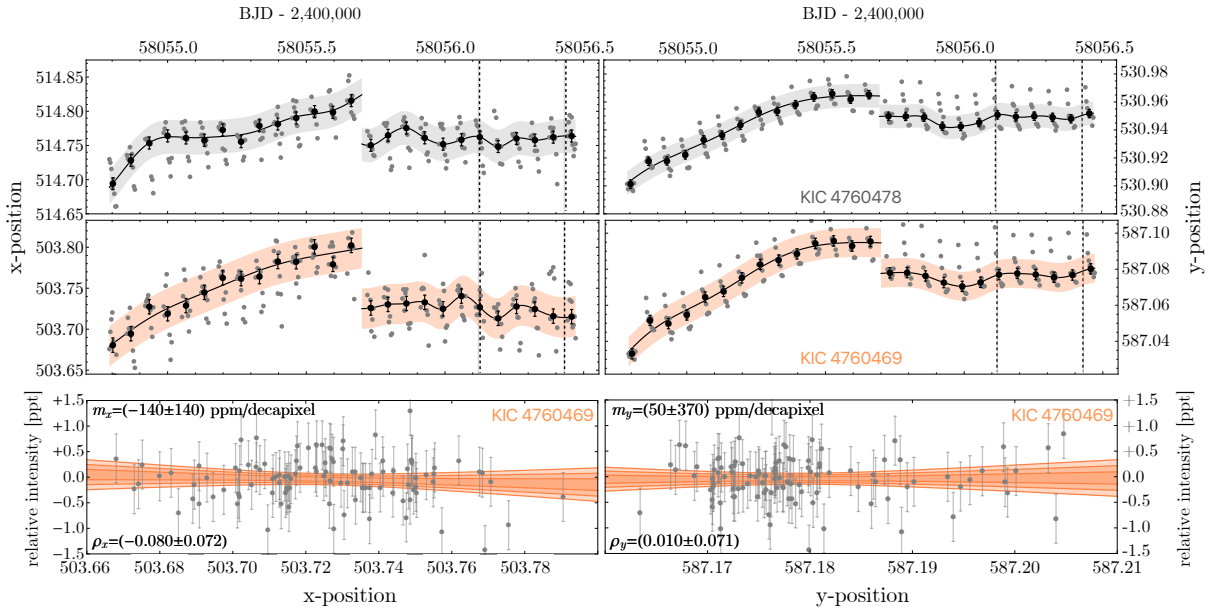

Figure 3.13: **HST centroids.** Top row shows the centroid position of our target in both row (left) and column (right) pixel index, with a GP model overlaid (shaded region). Middle row shows that the same but for our best comparison star (right). The column positions have been offset by 0.1 pixels after the visit change to more easily fit them on a single scale. The vertical grid lines mark the location of the moon-like dip seen in the photometry of the target, where we note that no peculiar behavior is evident. The lowest row shows a correlation plot of intensity versus centroid position for the comparison star (second visit), where no clear dependency is evident either.

the x direction. Together then, this indicates that a decipixel shift in centroid position may slightly affect the photometry at the ~ 200 ppm level.

Inspection of the centroid variations after the visit-change - where the moon-like dip occurs - reveals a shift of approximately one-tenth of a decipixel (see Figure 3.13). Since a one decipixel shift is associated with a 200 ppm photometric change, we argue that it is highly unlikely that ~500 ppm moon-like dip is a product of centroid shifts. However, we highlight that a quantative calculation of the centroid-induced photometric change in this region is not possible without knowledge of the functional form governing the residual inter-pixel sensitivity.

The above considered the centroid-flux correlation as observed using just the target, but the comparison star is also worth considering. If we assume the comparison star is stable,

a flux versus centroid plot reveals much more information than that from above. However, we caution that such a plot only maps the sensitivity at this part of the detector and this may not necessarily be the same as that on the target itself. Taking the second visit data only, which is the region of interest, we computed a cross-correlation of normalized intensity versus x and y centroid positions, as shown in Figure 4.3.

No clear relationship is apparent from inspection of this plot. If there is a functional relationship between flux and pixel position, it appears to operate at a level below the noise of this data. Certainly no linear relationship is detected, with the Pearson's correlation coefficient being consistent with zero for both x and y positions. Computing a linear slope in both cases implies that flux depends on x and y centroid position as (-140 ± 180) ppm/decipixel and $(+50 \pm 480)$ ppm/decipixel respectively. These numbers are consistent with the 200 ppm change observed in the target for a decipixel shift during the visit change, despite being located at a different part of the detector.

We can make some simplifying assumptions in order to have an approximate estimate of the centroid-induced flux changes expected around the moon-like event. These assumptions should not be treated as truth, but rather as plausible and necessary for quantitative progess. Although we don't know the true functional form, let's assume that intensity indeed maps linearly with centroid position for small changes (exploiting a Taylor expansion logic). We further assume that the detector's behavior on the comparison star is representative of the source. The x range in the source's second visit is 0.25 decipixels, which would be associated with a < 50 ppm change using the assumptions above. Similarly, the y range is 0.1 decipixels, implying a maximum variation of < 50 ppm. Accordingly, although this is certainly a simplified model, it suggests a 500 ppm flux change is quite unlikely to arise from the centroid variations.

Spectral analysis

With a low-resolution transmission spectrum we may also attempt to characterize the planet's atmosphere. A transmission spectrum, measured as the changing transit depth as a function of wavelength, can potentially reveal molecular absorption features in the atmosphere, and with sufficient sensitivity it may also be used to infer the planet's mass, as atmospheric abundances and Rayleigh scattering will be sensitive to the planet's surface gravity.

We split our optimal aperture in 10 even segments along the column direction and corrected for the hook independently in each using our non-parametric algorithm. We found that the tenth and reddest channel was quite unstable and thus neglect it in what follows. Light curves from the other nine channels are shown in Figure 3.14, and final transmission spectrum is shown in Figure 3.15.

We utilized our own MCMC code to explore parameter space and generate a best-fit transmission spectrum using the Exo-Transmit code Kempton et al. (2017); Freedman et al. (2008, 2014); Lupu et al. (2014). The generated high-resolution spectrum is binned at the appropriate wavelengths to test against the data, which also includes the *Kepler* transit depth, for a total of 10 data points across the wavelength range, 9 of which are derived from the WFC3 spectrum.

The variable inputs are radius of the planet R_P, surface gravity of the planet g_P, and (optionally) a cloud deck atmospheric pressure. The code treats the atmosphere as opaque at pressures higher than the cloud deck pressure value, corresponding to greater depths and effectively increasing the radius of the planet. All variables have uniform priors. Three other inputs (stellar radius, temperature, and metallicity) are fixed, as discussed below.

R_P was allowed to range from half to twice the radius of our best fit planet radius, while g_P is restricted between 1 and 1000 m s^{-1}. The mass may then be inferred from the combination of surface gravity and planetary radius. The cloud top pressures (when applied) could range from 1 to 10^7 Pa, allowing for virtually no cloud deck down to $\sim$ 100 atmospheres. To speed

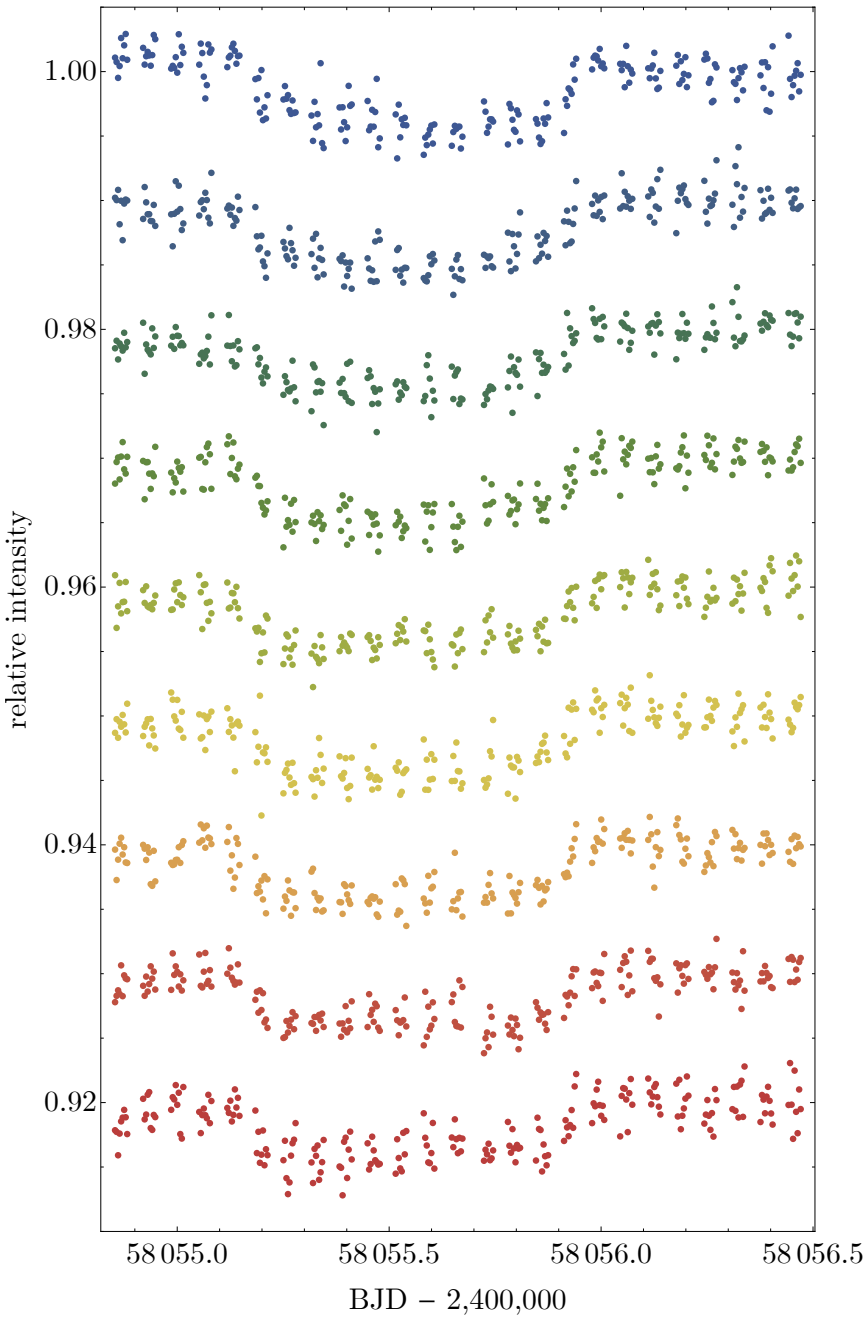

Figure 3.14: **Spectral analysis**. Nine spectral channels, color-coded in wavelength from the bluest channel (top) to the reddest (bottom), extracted from our WFC3 photometry of Kepler-1625b. Naturally the noise in each channel is considerably higher than the white light curve.

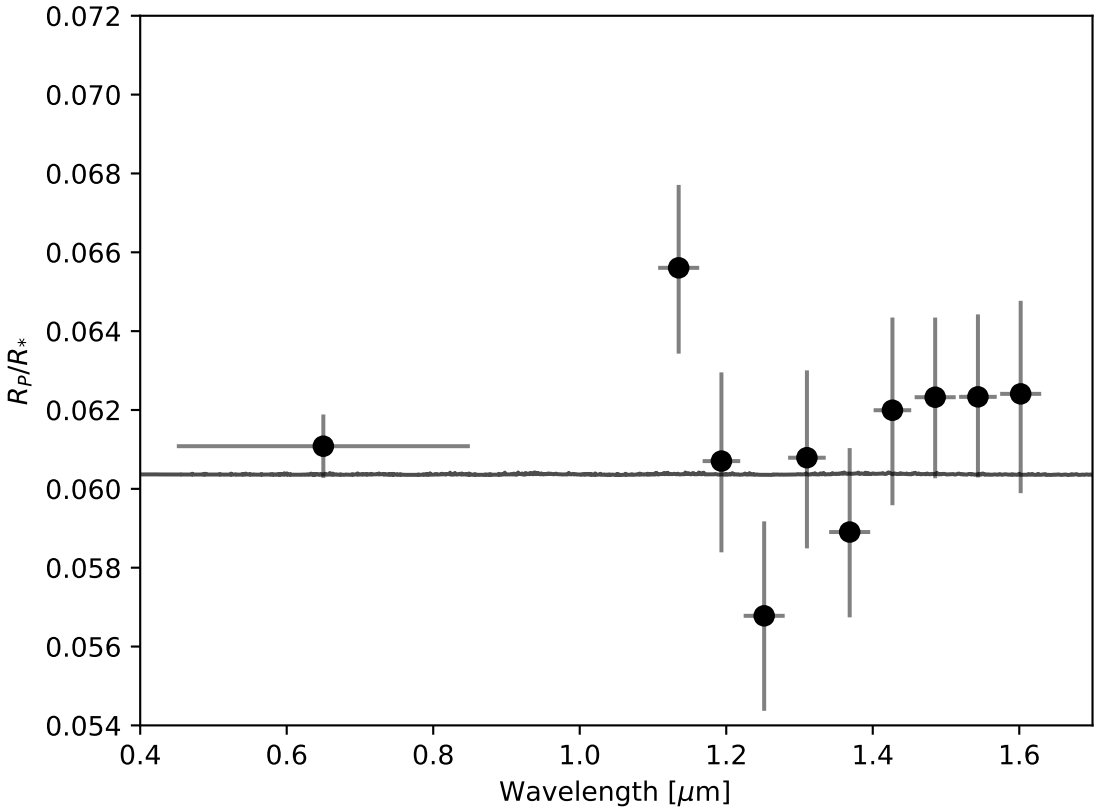

Figure 3.15: **Transmission spectrum.** The transmission spectrum measured as the ratio of the planet radius to the stellar radius, utilizing the *Kepler* bandpass at far left and the spectral channels extracted from the WFC3 grism photometry. For reference a model spectrum assuming $M_P = M_{\rm Jup}$ is also plotted.

λ	$\Delta\lambda$	$R_P/R_\star$	$\sigma_{R_P/R_\star}$
0.65	0.200	0.06102	0.0008
1.135	0.027	0.06438	0.0015
1.192	0.025	0.06198	0.0015
1.252	0.027	0.05770	0.0017
1.310	0.024	0.06157	0.0015
1.368	0.028	0.05960	0.0014
1.426	0.026	0.05990	0.0017
1.485	0.027	0.05924	0.0019
1.543	0.023	0.06019	0.0020
1.600	0.027	0.05910	0.0026

Table 3.3: **Transmission spectrum.** Marginalized ratio-of-radii derived from a Bayesian model averaged joint-posteriors of the linear, quadratic and exponential HST detrending models, using the averaged ("AVG") *Kepler* data detrending model.

up convergence the MCMC was initialized with reasonable first guesses for the radius (from the transit depth) and surface gravity of the planet (randomly chosen to be some value less than 30 m s^{-1}.

We fix the stellar radius at 1.793 $R_\odot$, as there is nothing in the transmission spectrum that can constrain the size of both the planet and the star. In addition, we fix the metallicity for the model at solar abundances and the planet temperature at 300K. Unlike other parameters in the Exo-Transmit code, for which any number can be specified, there are a narrow range of options to choose from for metallicity and temperature, owing to the fact that large files for abundances and temperature-pressure profiles must be used. While the Exo-Transmit authors state that interpolations between two files may be performed (for example, a 350K T-P profile could be obtained by interpolating between the 300K and 400K models), there is considerable uncertainty for the target with respect to both parameters; ExoFOP lists Fe/H for the target at 0.12 ± 0.15 (i.e. consistent with solar), and the equilibrium temperature at 350K. The albedo of the target is of course unknown, but likely pulls the temperature down closer to 300K. Meanwhile, generating interpolated files at each MCMC step would be considerably more expensive computationally.

After some experimentation we opted to model the spectrum without use of a cloud

deck. Motivating the elimination of clouds is the fact that it will act to suppress molecular features in the spectrum, thereby confusing the situation; if the spectrum is consistent with a flat line, we cannot know whether the molecular features are suppressed because of a cloud deck, or because the planet is very massive. Eliminating the cloud deck thereby allows us to characterize what the spectrum is doing were it to be influenced by planet mass alone.

As the uncertainties are quite large across all wavelength bins, due to the faintness of the target, the system parameters derived from this test are poorly constrained. We find the spectrum is consistent with a flat line / featureless atmosphere. If we assume no clouds present, atmospheric absorption would be potentially detectable only for very low surface gravity worlds, with sub-Saturn masses.

3.6.3 Joint Fits

Stellar Parameters

In our previous analysis Teachey et al. (2018), our source for fundamental stellar parameters came from *Kepler* DR25 Mathur et al. (2017). Since that time, data release two from *Gaia* has been released providing parallax information which should be expected to yield an improved inference Luri et al. (2018). Although these parallax constraints have been incorporated in a prior publication for *Kepler* planet hosts Berger et al. (2018), that work does not include stellar masses necessary for this work.

We therefore decided to use isochrone modeling to derive revised stellar parameters including the *Gaia* parallax. To do this, we use the `isochrones` package Morton et al. (2015) with Dartmouth tracks coupled to the `emcee` Bayesian inference algorithm Foreman-Mackey et al. (2013). The *Gaia* parallax of (0.406 ± 0.035) mas corresponds to a distance of (2460 ± 220) pc. Following the recommendations on the *Gaia* DR2 portal, the global systematic offset can be neglected since it is far less than the measurement error, and the measurement error is not expected to be an underestimate given the target's brightness. In addition to the parallax, we used the stellar effective temperature, surface gravity and metal-

licity inferred by Mathur et al. (2017) from spectroscopic constraints. Finally, we included the *Kepler* apparent magnitude (15.756) with an uncertainty set to 0.1 mag Huber et al. (2016).

Our revised stellar parameters yield an approximately Solar mass star ($1.04^{+0.08}_{-0.06}\,M_{\odot}$) with an enlarged radius ($1.73^{+0.24}_{-0.22}\,R_{\odot}$), implying that the star has evolved off the main sequence. This is consistent with the prior classification using the *Gaia* parallax Berger et al. (2018), yielding physical dimensions highly consistent with both previous estimates Mathur et al. (2017); Berger et al. (2018). The evolved state of the star means that the its age can be constrained to be $8.7^{+1.8}_{-1.8}$ Gyr, some 4 Gyr older than our Sun. This means that although the present day luminosity is $2.55^{+0.72}_{-0.58}\,L_{\odot}$, the star would have been 2.5 times less luminous for most its life, meaning Kepler-1625b would have received very close to Earth's present-day insolation during the main-sequence lifetime, given its semi-major axis.

Model Description

A transiting planet model represents a nested model of the more general planet+moon model, where the moon mass and radius equals zero. For this reason, a moon fit is guaranteed to provide a lower χ^2, or (more rigorously) a higher maximum likelihood. This basic fact forces exomoon hunters to adopt methods able to account for model complexity, such as Bayesian model selection, in order to make any progress, something long advocated by the Hunt for Exomoons with Kepler (HEK) project since its inception Kipping et al. (2012).

In this work, we fit light curve models to the data using a normal likelihood function with the MultiNest regression package Feroz et al. (2009). MultiNest is designed to estimate the Bayesian evidence of any model attempted, using multimodal nested sampling (see Skilling (2004) and Feroz & Hobson (2008)) to conduct inference. The Bayes factor between two models is then evaluated by taking the ratio of two evidences. A by-product of this process is the parameter posteriors, which are also useful checks when comparing models against one another.

Our light curve model is generated by `LUNA` Kipping (2011), a photodynamical Fortran code for simulating planet-moon light curves. In total, we consider four basic transit models, which we designate as P, T, M, and Z. We describe these models in turn.

The planet-only model P is described by seven parameters for a lone planet: the ratio-of-radii ($p = R_P/R_\star$), the stellar density ($\rho_\star$), the transit impact parameter (b), the time of transit minimum of the second observed *Kepler* transit (τ_0), the orbital period (P) and two quadratic-law limb darkening coefficients (q_1 and q_2; see Kipping (2013)). However, the presence of HST data meant chromatic differences were expected and so we included three extra terms, $p_{\rm HST}/p_{\rm Kep}$ to describe the ratio of the HST-to-*Kepler* ratio-of-radii and two new limb darkening terms for the WFC3-band, giving 10 parameters in total.

Model T is the same as model P except that each transit epoch (four in total) is given its own unique time of transit minimum, thereby allowing for timing variations. Duration variations or any other kind of dynamical change are not modeled. Model T requires four extra free terms but also two fewer (no orbital period and no τ_0), thereby giving a 12-parameter model.

Model M is the planet+moon model, which is similar to model P except that a moon is included. As such, seven additional parameters are added: the radius ratio, R_S/R_P; the mass ratio, M_S/M_P; the orbital period, P_S; the semi-major axis, a_{SP}/R_P; a term describing the orbital phase at time τ_0, ϕ_S; the orbital inclination angle, i_S; and the longitude of the ascending node, Ω_S. These seven new parameters give a total of 17 terms in model M (see Kipping et al. (2013b) for details on these definitions).

Finally, model Z is identical to model M except that R_S/R_P is fixed to zero and thus has one fewer free parameter. Model Z does not simply reproduce model T because moons can induce duration variations Kipping (2009a) as well as impact parameter changes Kipping (2009b), both of which are modeled by Z but not T.

We highlight that the number of free parameters described above only represent the transit-model parameters, and in practice the total number of free parameters is higher due

to the inclusion of visit-long trend terms. We direct the reader to our previous papers (e.g. Kipping et al. (2015)) for a description of the priors used.

It is important to note that model M also includes some constraints on physically acceptable parameter combinations. The planet and moon density can be derived as described in Kipping (2010) and we reject any samples for which these exceed $28\,\mathrm{g\,cm^{-3}}$ or drop below $0.08\,\mathrm{g\,cm^{-3}}$ in a bid to remove physically unsound combinations. This has two important consequences. First, by penalizing a part of the parameter volume, particularly at small signal sizes compatible with very marginal signals, model M will generally obtain a lower Bayesian evidence than it otherwise would. By demanding the model is physically sound, we are thus being more stringent in our calculation of Bayes' factors when assessing the case for a moon. Second, there is strong evidence for a timing offset in the HST data, which requires a non-zero M_S/M_P to explain for model M. Since infinitesimal radii would lead to infinite densities, R_S/R_P cannot approach zero and is therefore forced to always take on a positive value.

Without careful consideration, these positive values could be misinterpreted as evidence for a moon. Instead, the correct procedure should be to compare the evidences between the different models. In order to weigh up the evidence for a moon-like transit, we introduced the Z model which handles all of the moon-induced timing effects but without the radius effect. Comparing the evidences between models M and Z is the most direct way to infer the case for the moon-like transit signature. Comparing the evidences between models M and P is the most direct way to evaluate the overall preference for the moon model over the lone planet model.

We also take a brief aside to mention a subtlety with model T. Model T is somewhat unphysical. There is a single lone planet in the system which exhibits essentially arbitrary transit times. In practice, something must be causing these timing variations, be it a moon or another planet in the system. Any such body would require at least six new parameters to describe it and yet we here model the case with just (net) two additional parameters. For this

model	$\log \mathcal{Z}$	$\log \hat{\mathcal{L}}$
P	4924.84 ± 0.07	4950.26
T	4924.38 ± 0.08	4954.38
Z	4927.53 ± 0.08	4956.42
M	4925.34 ± 0.08	4959.59
M − P	0.50 ± 0.11	9.33
M − T	0.96 ± 0.11	5.22
M − Z	-2.19 ± 0.11	3.17

Table 3.4: **Kepler-only fits**. P for planet model. T for planetary TTV model. Z for a zero-radius moon model. M for moon model.

reason, model T is able to do better than it really should in terms of the Bayesian evidences, which generally penalize models for using extra degrees of freedom. For this reason, it is always worth keeping this point in mind when considering the T model and looking at the likelihoods as well.

The case for TTVs

Resulting Bayesian evidences ($\mathcal{Z}$) from the joint fits are presented in Table 1 of the main text, along with the maximum likelihoods ($\hat{\mathcal{L}}$) amongst the derived (and finite) posterior samples.

We begin by first considering the evidence for transit timing variations. As can be seen in Table 1 of the main text, model T is consistently favored over model P for all choices of the visit-long trend with $2 \log K \simeq 4$, indicating positive evidence for TTVs. Leaving the Bayesian evidences aside, the $\Delta\chi^2$ between model P and model T ranges from 17.1 to 19.2. Since model T uses just two more parameters than model P, the Akaike Information Criterion would range from 13.1 to 15.2, which indicates a very strong case for TTVs.

Figure 3.4 in the main text shows the maximum *a posteriori* predicted time of the transit, as inferred using a linear ephemeris P model fit to the *Kepler*-only data (described in Section 3.6.1) with a dashed vertical line in the right-hand-side panels. This time is clearly visually offset from the center of the planetary transit signal irrespective of visit-long trend

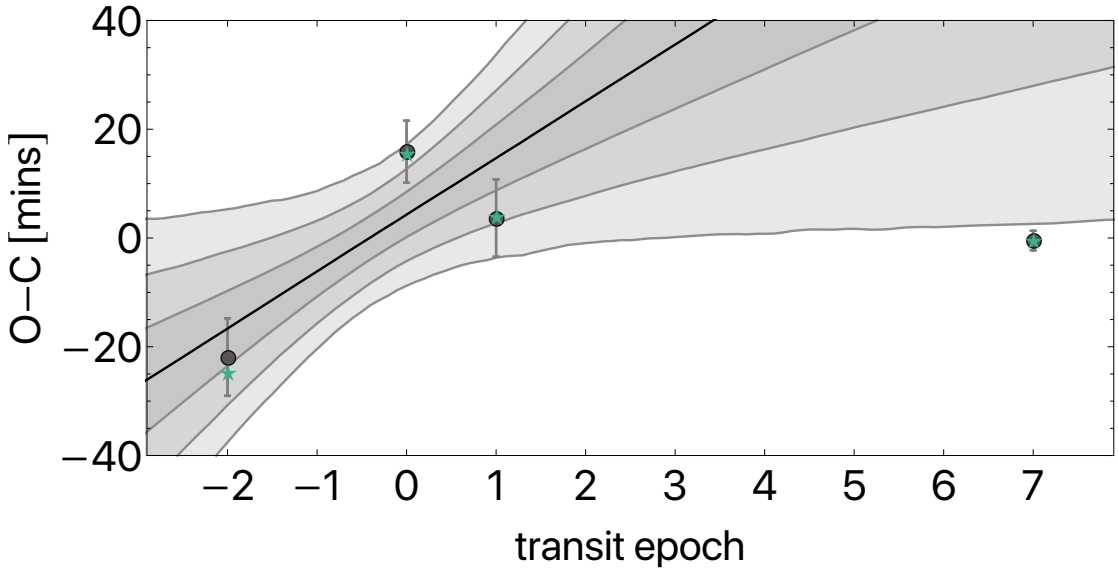

Figure 3.16: **Transit timing variations.** TTVs for Kepler-1625b, defined as observed times minus calculated times, where calculated times come from a linear ephemeris fit (model P) to all of the data marginalized over the various visit-long trend models attempted. The sloped line and three shaded regions represent the median, one-, two- and three-sigma credible intervals for the *a posteriori* linear ephemeris when conditioned upon the *Kepler* data alone, which reveals how deviant the HST epoch is. The green stars indicate the O-C values produced by the moon model, M.

Epoch	τ	O-C [mins]
-2	$55469.2037^{+0.0048}_{-0.0049}$	$-21.9^{+7.1}_{-7.1}$
0	$56043.9715^{+0.0040}_{-0.0040}$	$+15.9^{+5.6}_{-5.8}$
1	$56331.3337^{+0.0049}_{-0.0051}$	$+3.7^{+7.0}_{-7.2}$
7	$58055.5563^{+0.0013}_{-0.0014}$	$-0.4^{+1.7}_{-1.8}$

Table 3.5: **Transit timings.** Marginalized transit times derived from a Bayesian model averaged joint-posteriors of the linear, quadratic, and exponential HST detrending models, using the averaged *Kepler* data detrending model.

model. The vertical dashed lines on those plots denote the location of the time of transit minimum resulting from model T (and now including the HST data), and indicates that the transit came in ≃78 minutes earlier than expected.

Based on the *Kepler* data alone, we found that the P and T models had a Bayes factor of approximately unity (see Section 3.6.1 and Table 3.6.3). Therefore, the first three transits observed appear compatible with a linear ephemeris. It is only the new HST transit which appears offset. However, including the HST epoch causes the maximum likelihood linear ephemeris to change substantially. This is visualized in Figure 3.16, where one can see the first three *Kepler* epoch times with the corresponding maximum *a posteriori* linear ephemeris plotted in black. The overall O-C diagram uses the updated maximum *a posteriori* ephemeris from model P using the HST data (marginalized over all of the visit-long trend models). Similarly, each transit time plotted (and also presented in Table 3.6.3) is that from a Bayesian model averaged posterior across the various visit-long trend models using model T (for details on the model averaging see Section 3.6.3). It is clear from Figure 3.16 that predicted *Kepler*-based prediction is inconsistent with the HST epoch with a one-sided p-value of $> 3\,\sigma$. For reference, each *Kepler* transit midtime has an uncertainty on the order of 10 minutes and the standard deviation on linear ephemeris predictions is 25.2 minutes derived from posterior samples. On this basis, we consider there to be a strong case for TTVs after introducing the HST epoch.

Against the maximum *a posteriori* linear ephemeris, the TTVs have a semi-amplitude

of $\simeq 25$ minutes. With four transit times alone, is it perhaps not surprising that it is easy to find good solutions to the TTVs using an external perturber, although the solutions are extremely degenerate. We used `TTVfaster` Agol & Deck (2016) with an MCMC exploration of the parameter space and found a wide range of plausible solutions assuming an external perturber.

Other timing effects

An important point is that no matter how well a perturbing planet can explain the observed TTVs, there are additional effects that an exomoon is expected to impart on a light curve that a perturbing planet, in general, will not. These are all the other photodynamical effects that may be seen in the transit of Kepler-1625b, such as transit duration variations Kipping (2009a), transit impact parameter variations Kipping (2009b) and transits of the moon itself Kipping (2011). Therefore, the way to establish whether a putative TTV is due to a moon or a perturbing planet is to look for these additional characteristic signatures expected from the presence of a moon.

Taking the dynamical signatures first, these are generally expected to be much smaller than the TTVs Kipping (2009a), and accompanied by larger measurement errors (for example, transit durations have twice the uncertainty of transit times; Carter et al. (2008)). Given the fact that the TTVs are only significant at the $3\,\sigma$ level, we certainly do not expect, nor do we observe, noticeable duration variations. Nevertheless, these small effects are accounted for in a photodynamical model such as `LUNA` and thus we can see if they lead to any improvement in the fits. Model Z serves this purpose by fixing the moon radius to zero, but otherwise describing a full three-dimensional moon orbit with six more terms than model P.

It is instructive to compare the maximum likelihoods from models T and Z. Although our formalism for model T only uses two extra free parameters, as noted earlier in Section 3.6.3, this is somewhat artificial and in reality timing variations would require a planet described by six orbital parameters too. In this way, models T and Z can be directly compared as

having essentially the same number of free parameters. Model Z consistently out-performs T in this regard. However, we consider these improvements as being necessary for any successful moon model rather than being convincing evidence in isolation. This is because model Z has greater flexibility than model T to explain light curve changes and thus in many ways is guaranteed to lead to an improvement. Nevertheless, the evidences show model Z consistently out-performs T by around $2 \log \boldsymbol{K} \simeq 4$, meaning there is some evidence favoring the moon hypothesis over a planet perturbation model. We caution that we have assumed planets do not induce short-period TDVs and/or impact parameter changes here but in some rare cases such changes have been detected (e.g. Nesvorný et al. (2013); Szabó et al. (2012)).

The case for an exomoon

So far we have established that a) Kepler-1625b exhibits an early transit in the HST epoch indicating TTVs in the system, and b) a zero-radius moon model, which explains TDVs and other effects in addition to TTVs, leads to a modest improvement in the fits. A true exomoon would be expected to exhibit both points, and so while this is not convincing enough on its own, the exomoon hypothesis is certainly "on-track". The next step is to consider the full planet+moon models.

We first note that model M, the full planet+moon model, is the favored hypothesis in each and every visit-long trend model attempted (see Table 1 of the main text), which already formally establishes the moon hypothesis as the leading candidate solution. With a strong case for TTVs already established (see Section 3.6.3), and some modest evidence for other moon-induced dynamical effects (Section 3.6.3), the mass signature of the putative moon is measurable and significant. To build a compelling case for an exomoon, we also need to detect the radius signature, since the dominant dynamical effect (TTVs) could be plausibly caused by a perturbing planet instead. Not only should we detect this radius signature, but that radius signature must be consistent with the moon solution in terms of both phase and physical parameters.

As a result of the strong TTV, model M has no other way to explain the timing offset except for using the moon, which leads to a positive M_S/M_P, which in turn demands a positive R_S/R_P (due to our rejection of unphysical moon densities; see Section 3.6.3). Accordingly, inspection of the marginalized posterior distributions of R_S/R_P is not a useful strategy for evaluating the case for a detection of the moon's radius signature. The most direct method would be to directly compare the Bayes factor between the full moon model (i.e. model M) and an identical model for which the radius effect is turned off (i.e. model Z).

In all visit-long trend models attempted, we find that model M is favored over model Z. However, there is considerable range in the Bayes factors which result (see Table 3.3 in the main text). This can be linked to Figure 3.4 in the main text, where one can see that the high Bayes factor cases (e.g. linear) tend to correspond to cases where the moon transit is noticeably larger. In all cases, the moon transit occurs towards the end of the observations.

We would argue that a moon-like transit is immediately obvious even in the non-detrended data (see Figure 2 in the main text) and corresponds to the location where all of the moon models place the signal. This location is important because Kepler-1625b transits early in the HST epoch, for which we would therefore expect a corresponding moon to be transiting late. In other words, just combining the fact that we see an early transit along with the extra dip observed in the non-detrended data already presents a strong case for a self-consistent moon solution (and indeed our detailed modeling verifies this statement). Although the phasing is indeed aligned as expected then, this does not address whether the amplitude of the TTV (implying a certain moon mass) is compatible with the amplitude of the moon-like transit (implying a certain moon radius). We can therefore summarize the situation as being that the best explanation to the data is an exomoon, driven by a timing offset and moon-like dip with self-consistent phases, although the properties of the moon (in particular the size) appear to be dependent upon the trend model adopted and one (or all) of these may not be physically permissible. This latter point forms the subject of investigation in Section 3.6.3.

Before discussing whether the moon properties are physically sound or not, we briefly

highlight that the χ^2 improvement in going from model M to Z is quite high, ranging from 19.2 to 33.7 (note that we have $n = 1048$ data points). On this basis, we find the SNR of the moon-like transits to be at least 19.

Evaluating the physicality of putative system parameters

We now have a TTV signal, modest evidence for other moon-induced dynamical effects and a sizable χ^2 improvement when including the transit of a moon at the correct phase to explain the TTV. It is not yet clear, however, whether the candidate moon actually has physically sound parameters, beyond falling within the very generous density constraints imposed by our model (and described in Section 3.6.3).

Each trend model clearly leads to distinct moon parameters, as evident by the different moon depths in Figure 4 of the main text. Not only is the depth different, but the moon sometimes displays an egress feature and other times the egress occurs after the observations have ceased, which would require a greater planet-to-moon semi-major axis. Since the semi-major axis multiplied by the moon-to-planet ratio dictates the TTV amplitude, these longer semi-major axis solution necessarily require a smaller M_S/M_P. Furthermore, these longer semi-major axis solutions tend to be correlated to the greater moon transit depths as a result of the interplay with the trend model. This leads to a situation where R_S/R_P and M_S/M_P are somewhat anti-correlated between the various trend models, although admittedly there are just three models under inspection here.

The differing planet+moon transit parameters between each trend model imply different physical parameters then, with no guarantee of such parameters being necessarily physically plausible. To investigate this, we have to go from ratios to absolute dimensions. For the radius, this is easy. All solutions lead to good constraints on R_S/R_P and $R_P/R_\star$ and since the stellar radius is known we can derive absolute sizes with reasonable precision.

Deriving absolute mass solutions is far trickier, since although M_S/M_P is well-constrained from the TTVs, $M_P/M_\star$ is found to be only weakly constrained photodynamically in all

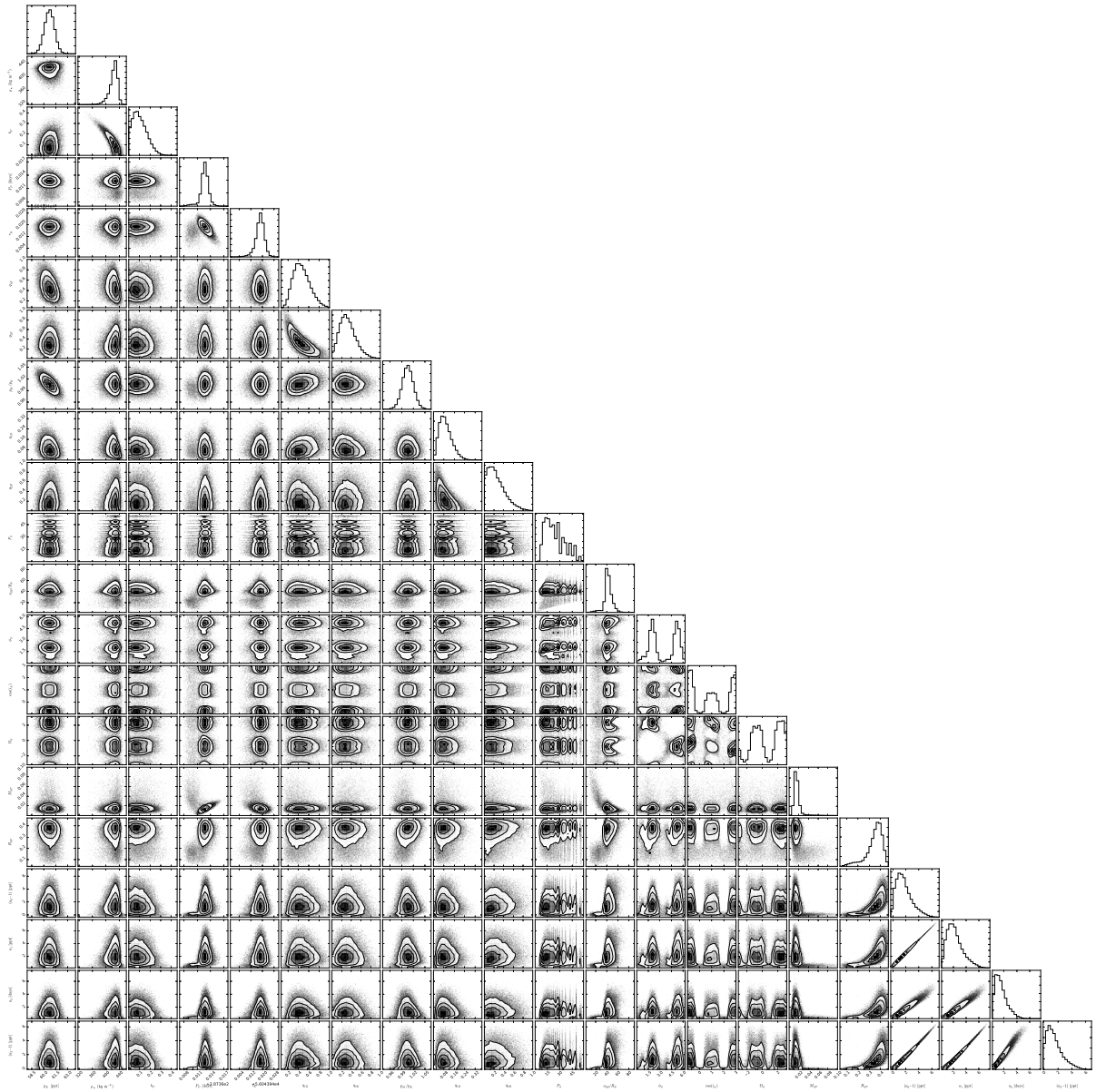

Figure 3.17: **Model posteriors.** Model posteriors of the parameters explored in the moon model. Shown here the results from the exponential trend model.

models. This can be seen in Figure 3.19, by inspection of the dotted lines. In an attempt to shore-up this planetary mass measurement, we fed the absolute radius samples into the probabilistic empirical mass-radius relation `forecaster` Chen & Kipping (2017), which are the dashed lines in that plot. Taking the product of the two PDFs (using a Gaussian kernel density estimator) leads to the solid black posteriors in main text Figure 3.19, which represents one method of estimating the planetary mass from our data.

A second way to measure the planetary mass is to come from the exomoon side. Using the moon radius, we feed the samples into `forecaster` to get a moon mass and then convert that into a planetary mass using the M_S/M_P distribution inferred by our fits. This second planetary mass solution is plotted in solid orange in Figure 3.19. A physically self-consistent solution would correspond to no tension between the two distributions. Visual inspection of Figure 3.19 reveals that the two estimates are consistent with one another for all three trend models, with the only difference being support for lower masses in the quadratic case.

To quantify the compatibility, we draw a random sample from one of the distributions (the moon-forecast based solution) and evaluated its likelihood using the product PDF from the other solution. We repeated this for all available samples ($\sim$ 40000) and then evaluated the mean likelihood, which is reported in the top-right corners of Figure 3.19. In agreement with simple inspection, the linear model appears to provide the greatest degree of physical self-consistency. All three provide broadly consistent mean likelihoods, none of which appear unphysical.

Combining the two mass estimates

Although the absolute dimensions of the system are constrained from photodynamics and our stellar properties, we are able to obtain more precise constraints by folding in the `forecaster` results from the previous paragraphs. For each trend model, we take the two planetary mass distributions (represented by the black solid and orange solid lines in Figure 3.19) and combine them to form a single planetary mass solution. This is achieved

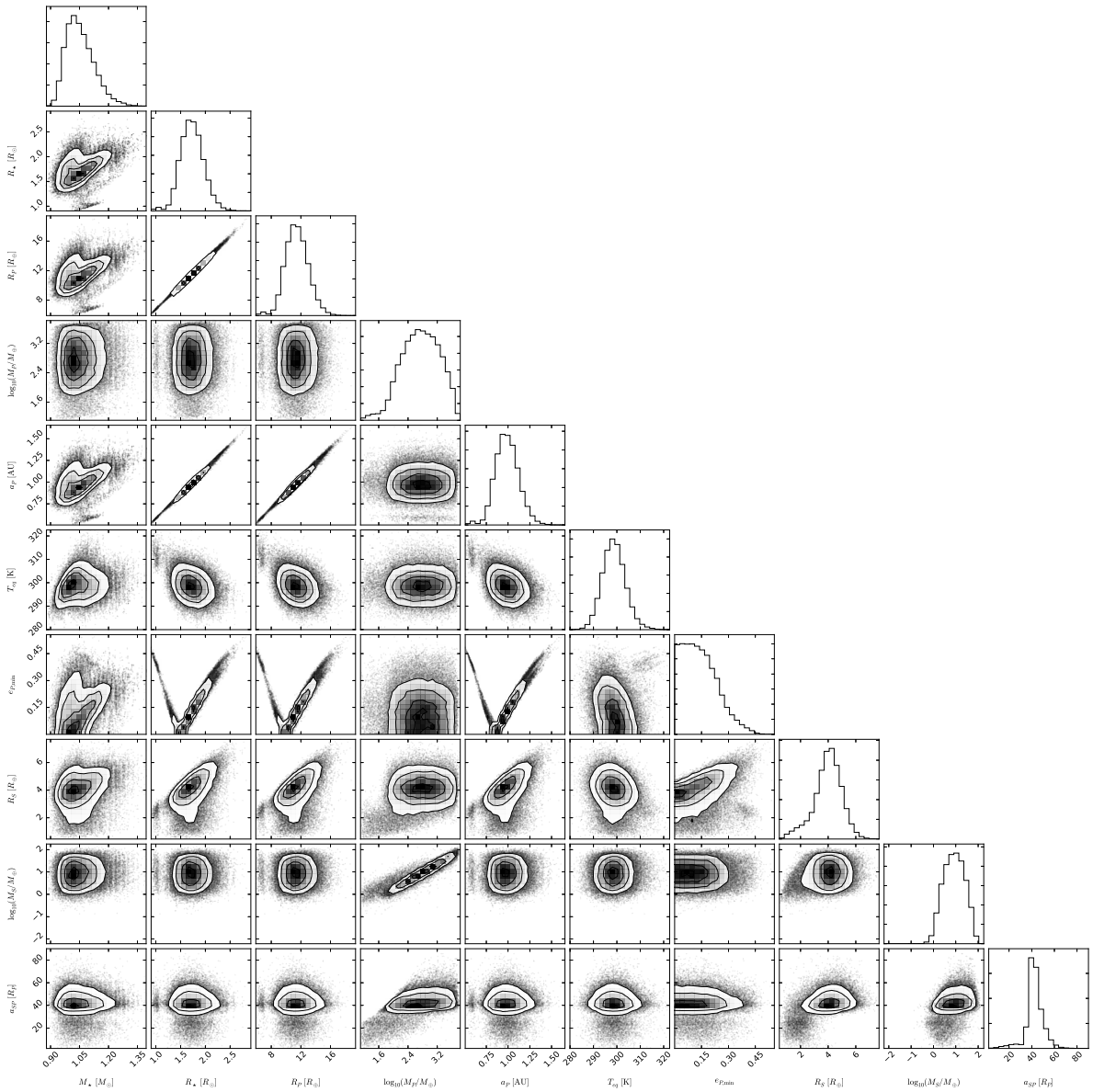

Figure 3.18: **Physical posteriors.** Physical system parameter posteriors derived from the exponential model results.

by taking the product of two PDFs and then defining the result as a likelihood function. We then sample from the function with a simple Markov Chain until 40,000 samples have been computed. The resulting planetary mass posterior is converted into a satellite mass using the M_S/M_P parameter from the model M fits, which can be found in Table 2 of the main text.

Bayesian Model Averaging

The moon models display subtle but important variations in the associated parameters for each trend model. In contrast, we find the depths and transit times from model T are highly stable. In order to compile a single posterior for the transit times (presented in Table 3.6.3), we decided to marginalize over the three models using Bayesian Model Averaging.

The basic idea is that each model (in our case the three visit-long trend models) are assigned a weight based on their Bayesian evidence, $\mathcal{Z}$ (which is also known as the marginal likelihood). The odds-ratio between model i and j may be computed as

$$O_{ij} = \exp(\log \mathcal{Z}'_i - \log \mathcal{Z}'_j)$$

And the final weights are now defined as

$$w_i = \frac{O_{i1}}{\sum O_{i1}},$$

such that $\sum w_i = 1$. With the weights assigned, we draw a random integer from a multinomial distribution using the weights vector to choose a model. We then choose a random posterior sample from that model and append it to a new array. After many iterations, we construct a Bayesian Model Averaged posterior for model T.

We repeated this process on the nine spectral channels computed earlier using model T with the three trend models. The resulting marginalized ratio-of-radii is treated as our final spectral retrieval for Kepler-1625b and is presented in Table 3.6.2.

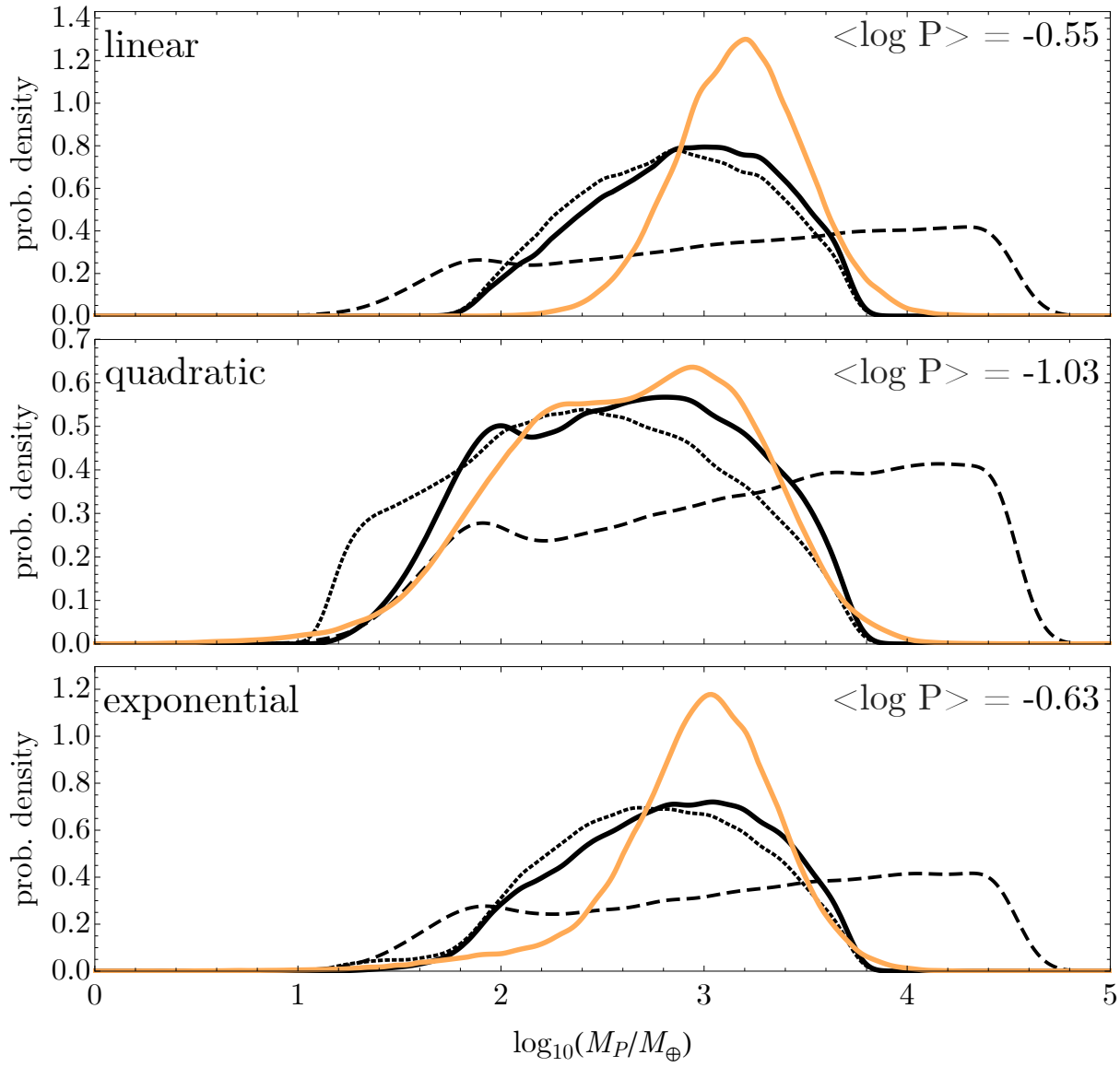

Figure 3.19: **Mass constraints.** Mass solutions for the planet from the various HST detrendings. The dotted lines represent the photodynamical posterior probability distribution while the dashed lines are posteriors generated by `forecaster` Chen & Kipping (2017). The solid black is the product of these probabilities. The orange lines represent the mass solution for the planet derived from that of the moon, which is well constrained based on the inferred radius. The mean likelihood, tracking the compatibility of the two solid curves, is shown in the upper-right corner.

Stability Analysis

Moons are generally considered stable if they reside beyond the Roche limit and within $\sim 0.4895\ R_{Hill}$ for prograde moons and $0.9309\ R_{Hill}$ for retrograde moons Domingos et al. (2006). Our solutions place the moon at $a_s = 0.265^{+0.123}_{-0.081}$, $0.275^{+0.126}_{-0.095}$, and $0.258^{+0.111}_{-0.081}\ R_{Hill}$ for the linear, quadratic, and exponential models, respectively.

However, satellites on inclined orbits may be less stable over the long term. Our three moon solutions all suggest significant inclinations with respect to the planet's orbital plane, so we carry out a stability analysis using the analytical formula provided by Donnison (2014). Drawing from our joint model posteriors we find the moon solution has a stable configuration for 73.1% of the draws with the linear model, 73.6% of the time for the quadratic model, and 78.3% of the time for the exponential model.

Analysis of Residuals

As an extra check, we here describe an analysis of the light curve residuals. Strong time-correlated noise structure could potentially explain the moon-like dip without invoking a satellite. In total, we considered six sets of relevant residuals, the three trend models applied to the moon model, M, and the same three applied to the zero-radius moon model, Z. These are shown in Figure 3.20.

The M model residuals appear consistent with Gaussian noise, displaying no obvious trends or time-correlated structure. In contrast, the Z models consistently fail to explain the moon-like dip leading to a noticeable excursion in the residuals at this time.

We binned the residuals into progressively larger bins and monitored the effect this has on the root mean square (RMS). Gaussian residuals would be expected to bin as $\sigma_n = \sigma_1 n^{-1/2}\sqrt{m/(m-1)}$, where n is the number of points binned, m is the number of bins and σ_1 is the RMS of the unbinned data.

Setting σ_1 to our standard photometric error, derived from our hook correction procedure, reveals that this curve yields a close match to the M model residuals (see lower panels of

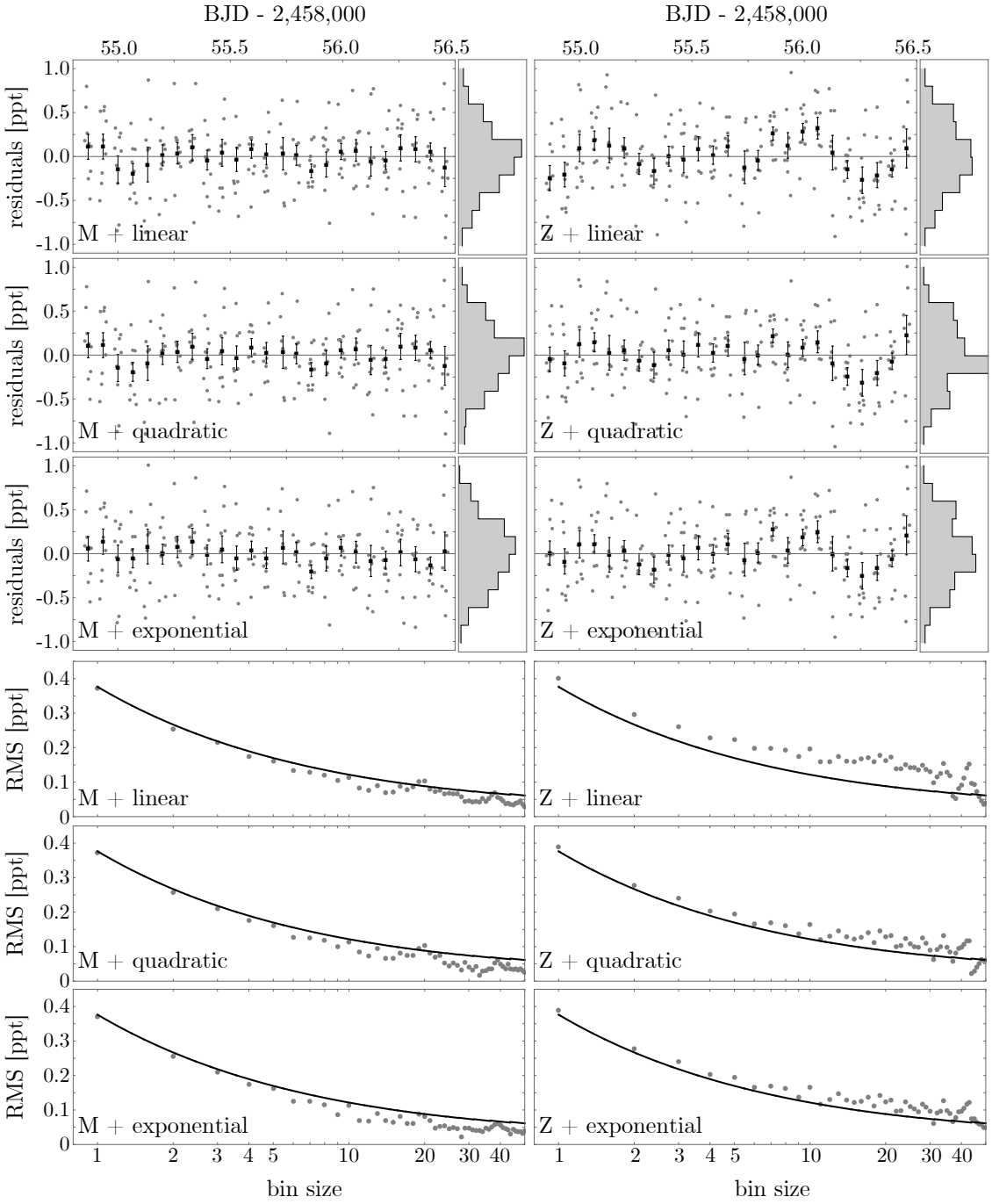

Figure 3.20: **Residual analysis.** Analysis of the white light curve from WFC3 residuals for six different models. Upper six panels show the photometric residuals, with orbit binned points, alongside a histogram of the unbinned scatter. Lower panels show the root mean square (RMS) as a function of bin size, where the solid line is that expected for pure Gaussian noise equal to the assumed photometric noise in our fits. In both sets, excess noise in the Z-models is visible, caused by the moon-like dip being ignored in those fits.

Figure 3.20). If anything, the noise appears to behave even better than Gaussian suggesting that we have slightly overestimated our standard uncertainty value.

In contrast, all three Z models show excess power at large bin sizes, indicative of time-correlated structure in the residuals. This structure must be caused by the moon-like dip since models M and Z are otherwise identical.

In summary, our analysis of the residuals adds weight to the case that the moon-like dip is a real feature in the data.

Chromatic test

In Section 3.6.2, we discussed how inspection of the comparison star's photometry and the centroids revealed no evidence to suspect the moon-like dip is an instrumental artifact. A third test, mentioned in that section, was to consider if the dip was chromatic in nature. An exomoon transit might be expected to be slightly chromatic due to the atmosphere, but large transit depth variations would be indicative of a blend or some previously unexplained instrumental effect. Accordingly, we here describe a brief investigation into this possibility.

Already we have discussed how spectral elements can be extracted by dividing the white light curve aperture up into slices (see Section 3.6.2). As a result of dividing up the aperture in this way, the noise naturally increases in each channel. For the sake of testing if the moon-like dip is chromatic or not, we only require two different colors, allowing us to use broader slices with higher precision. Accordingly, we decided to split up the optimal aperture into two colors for a chromatic test.

When divided into ten slices, we found that the reddest and bluest channels displayed less stable photometry, likely as a result of the sharp drop off in WFC3's sensitivity towards the wavelength extremes. For this test then, we elected to omit those most extreme channels and ultimately took the central two-thirds of the aperture, split into two, as our two colors. This led to a "blue" channel defined from columns 476 to 518 (inclusive) and a "red" channel from 519 to 561, which corresponds to 1.2-1.4μm and 1.4-1.6μm respectively.

We first independently correct for the hooks using the same method as described in Section 3.6.2. The white light curve led to a final inter-orbit RMS of 375.5 ppm, the blue and red channels here yielded 598.8 ppm and 612.8 ppm respectively. This is slightly better than the naive $\sqrt{3}$ scaling one would expect if the product of the source and WFC3 had a perfectly flat wavelength response, reflecting how the response actually peaks in the center and drops off towards the edges.

With the chromatic photometry in hand, our objective is now to test whether each channel is more consistent with either the model including or excluding an exomoon transit; formally models M and Z. To accomplish this, we took the maximum *a posteriori* solution from model M and plotted the light curve morphology with the trend parameters turned off, thus creating a "template" of the moon solution. We then multiplied this template by a trend model with unset trend parameters. These unknown trend parameters were then fit by regressing (least squares) the template multiplied by the trend model to the blue/red channel photometry. By doing this, we allow for the fact that the visit-long trend model is itself chromatic. Since there are three possible trend models considered in this work, we tried all three and selected the best one, since it was clear sometimes a particular trend model was a poor representation of the data. It is interesting to note that the quadratic trend won out every time.

One complication with the above is that there is not a single moon model template, but rather three - since we originally attempted three trend models and have been unable to definitively select a single model. We therefore repeated the above the templates resulting from M + linear, M + quadratic and M + exponential trends. This was done for both the blue and red channel data, and in each case we recorded the (six) resulting χ^2 values of the (three) trend-regressed templates. We then repeated the entire excercise using the Z model templates, which do not allow for an exomoon transit. The results, including a comparison of the χ^2 values, is shown in Figure 3.21.

In all six cases, the moon template provides a closer match to the red/blue channel data

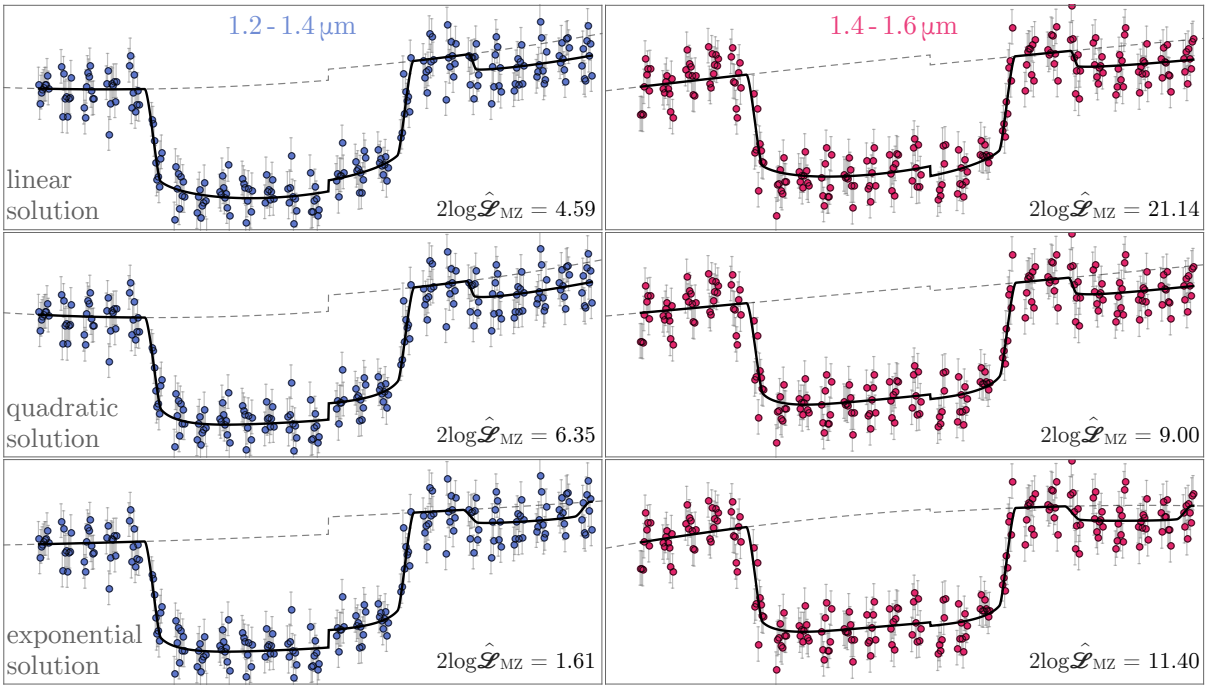

Figure 3.21: **Chromatic test.** Tests to see if the moon-like dip is present in two independent spectral regions of the WFC3 bandpass (each column). Each row shows our maximum *a posteriori* moon model (model M) plotted in black, multiplied by a simple trend model regressed to each channel. Since we have three different moon models depending on which trend model is used on the white light curve, we show all three (one per row). In every case, these templates give a closer match to the data than those resulting from model Z (no moon transit).

than the Z model. Although the moon model appears favored in every case, the red channel does appear to yield a more significant preference.

In general, one should expect the red channel to be less affected by stellar activity, such as rotating star spots, plages, granulation and micro-flares. If the blue channel, which is more sensitive to such effects, were to exhibit a stronger preference for a moon-like dip, this could have been a cause for concern. As it stands, however, the results appear to be consistent with an overall preference for model M over Z.

We also fitted the two channels with independent Z and M models using the same procedure as for the white light curve. In all six cases (three trends, two channels), the moon model is favored over model Z with evidences of 1.51 ± 0.31, 1.84 ± 0.31 and 3.78 ± 0.31 for the blue channel and 6.68 ± 0.32, 2.59 ± 0.32 and 4.58 ± 0.31 for the red, where the three numbers correspond to $2\log(\mathcal{Z}_M/\mathcal{Z}_Z)$ for the linear, quadratic and exponential trends, respectively. We do note that the solutions are less well-converged than before, as a result of the reduced precision, with credible intervals inflating by up to a few times.

In conclusion, the moon-like dip occurring post-egress is statistically favored in both the red and blue channels and appears consistent with the solution derived from the white light curve alone.

Predictions

As noted in the main text, we consider that future observations will be critical in assessing the true nature of Kepler-1625b and the reality of the exomoon signature. To this end, we took 100 random samples from our model M posteriors and used them to predict the morphology of the next transit event in May 2019 (epoch 9). This prediction is shown in Figure 3.22. The model suggests a high likelihood of observing a pre-transit feature due to the exomoon and thus we strongly encourage observations around this time.

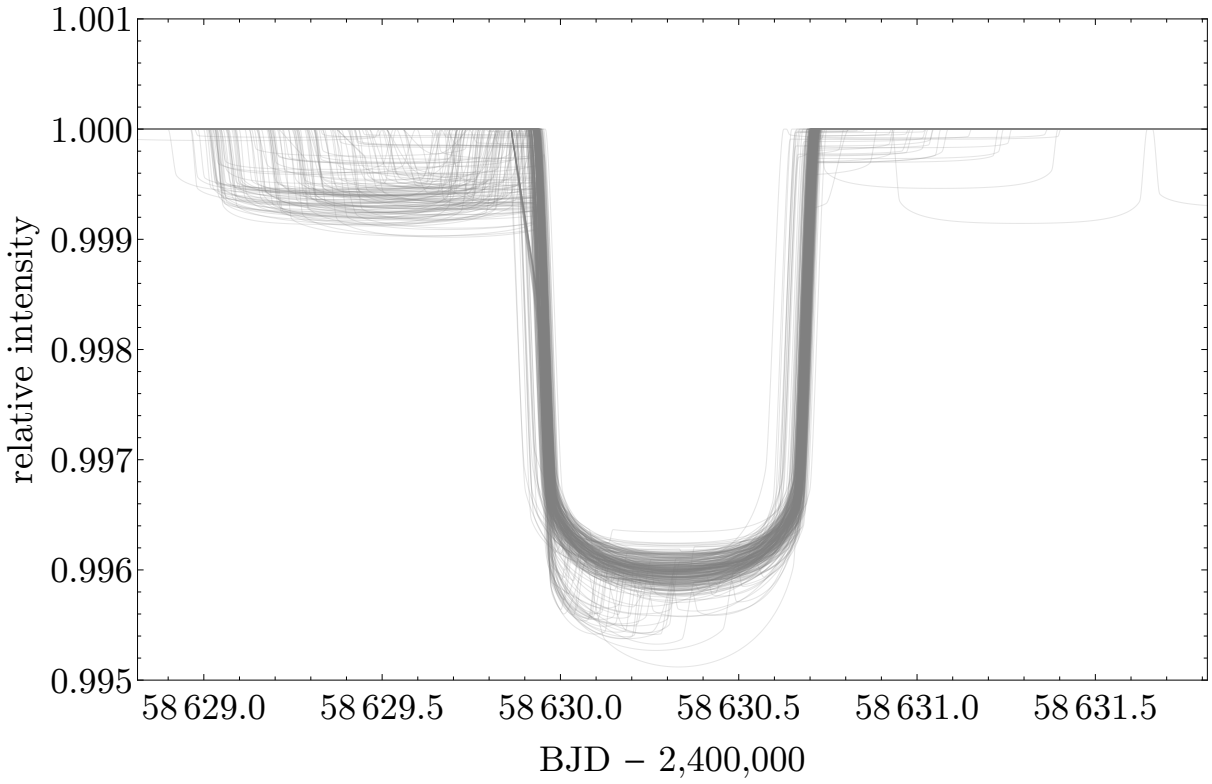

Figure 3.22: **The May 2019 transit.** Predictions for the May 2019 transit of Kepler-1625 assuming the planet-moon model is correct. 100 random draws from each of the three instrumental trend model posteriors are overlaid showing broad consistency between the three model predictions.

Acknowledgements

We wish to thank STScI staff scientists Bill Januszewski and Kevin Stevenson for their critical contributions during the planning and execution of the HST observation. We also thank Jon Jenkins at NASA and Paul Dalba at Boston University for useful discussions regarding source contamination in the *Kepler* data. Members of the Cool Worlds Lab at Columbia University (Ruth Angus, Jingjing Chen, Jorge Cortes, Tiffany Jansen, Moiya McTier, Emily Sandford, and Adam Wheeler) provided valuable feedback at every stage of this analysis. We are also grateful to members of the Hunt for Exomoons with Kepler project for their continued support throughout the early years of our program. Finally, we thank Travis Berger and collaborators for sharing their *Gaia*-derived posteriors for the target's radius.

Data and materials availability: The raw data from both the *Kepler* and HST observations are freely available for download at the Mikulski Archive for Space Telescopes (https://archive.stsci.edu). All relevant information required for replication of these results and to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors. This work made use of Numpy, Scipy, Pandas, Matplotlib, Astropy, `TTVfaster`, `Exo-Transmit`, `forecaster`, `LUNA` and MultiNest.

Chapter 3: Bibliography

Han, C., Han, W. 2002. On the Feasibility of Detecting Satellites of Extrasolar Planets via Microlensing. The Astrophysical Journal 580, 490.

Han, C. 2008. Microlensing detections of moons of exoplanets. The Astrophysical Journal 684, 684.

Liebig, C. 2010. Detectability of Extrasolar Moons as Gravitational Microlenses. Astronomy & Astrophysics 520, 68.

Cabrera, J., Schneider, J. 2007. Detecting companions to extrasolar planets using mutual events. Astronomy and Astrophysics 464, 1133.

Agol, E., Jansen, T., Lacy, B., Robinson, T. D., Meadows, V. 2015. The Center of light: Spectroastrometric detection of exomoons. The Astrophysical Journal 812, 5.

Noyola, J. P., Satyal, S., Musielak, Z. E. 2014. Detection of Exomoons through Observation of Radio Emissions. The Astrophysical Journal 791, 25.

Lewis, K. M., Sackett, P. D., Mardling, R. A. 2008. Possibility of Detecting Moons of Pulsar Planets through Time-of-Arrival Analysis. The Astrophysical Journal 685, L153.

Sartoretti, P., Schneider, J. 1999. On the detection of satellites of extrasolar planets with the method of transits. Astronomy and Astrophysics Supplement Series 134, 553.

Kipping, D. M. 2009. Transit timing effects due to an exomoon. Monthly Notices of the Royal Astronomical Society 392, 181.

Kipping, D. M. 2009. Transit timing effects due to an exomoon - II. Monthly Notices of the Royal Astronomical Society 396, 1797.

Barclay, T., and 57 colleagues 2013. A sub-Mercury-sized exoplanet. Nature 494, 452.

Teachey, A., Kipping, D. M., Schmitt, A. R. 2018. HEK. VI. On the dearth of Galilean analogs in Kepler, and the exomoon candidate Kepler-1625b I. The Astronomical Journal 155, 36.

Namouni, F. 2010. The Fate of Moons of Close-in Giant Exoplanets. The Astrophysical Journal 719, L145.

Spalding, C., Batygin, K., Adams, F. C. 2016. Resonant Removal of Exomoons during Planetary Migration. The Astrophysical Journal 817, 18.

Morton, T. D., and 7 colleagues 2016. False Positive Probabilities for all Kepler Objects of Interest: 1284 Newly Validated Planets and 428 Likely False Positives. The Astrophysical Journal 822, 86.

Mathur, S., and 16 colleagues 2017. Revised Stellar Properties of Kepler Targets for the Q1-17 (DR25) Transit Detection Run. The Astrophysical Journal Supplement Series 229, 30.

Jenkins, J. M., and 29 colleagues 2010. Overview of the Kepler Science Processing Pipeline. The Astrophysical Journal 713, L87.

Kipping, D. M., Hartman, J., Buchhave, L. A., Schmitt, A. R., Bakos, G. Á., Nesvorný, D. 2013. The Hunt for Exomoons with Kepler (HEK). II. Analysis of Seven Viable Satellite-hosting Planet Candidates. The Astrophysical Journal 770, 101.

Stumpe, M. C., and 10 colleagues 2012. Kepler Presearch Data Conditioning I—Architecture and Algorithms for Error Correction in Kepler Light Curves. Publications of the Astronomical Society of the Pacific 124, 985.

Smith, J. C., and 10 colleagues 2012. Kepler Presearch Data Conditioning II - A Bayesian Approach to Systematic Error Correction. Publications of the Astronomical Society of the Pacific 124, 1000.

Kipping, D. M. 2011. LUNA: an algorithm for generating dynamic planet-moon transits. Monthly Notices of the Royal Astronomical Society 416, 689.

Agol, E., and 6 colleagues 2010. The Climate of HD 189733b from Fourteen Transits and Eclipses Measured by Spitzer. The Astrophysical Journal 721, 1861.

Berta, Z. K., and 9 colleagues 2012. The Flat Transmission Spectrum of the Super-Earth GJ1214b from Wide Field Camera 3 on the Hubble Space Telescope. The Astrophysical

Journal 747, 35.

Wakeford, H. R., Sing, D. K., Evans, T., Deming, D., Mandell, A. 2016. Marginalizing Instrument Systematics in HST WFC3 Transit Light Curves. The Astrophysical Journal 819, 10.

Huitson, C. M., and 16 colleagues 2013. An HST optical-to-near-IR transmission spectrum of the hot Jupiter WASP-19b: detection of atmospheric water and likely absence of TiO. Monthly Notices of the Royal Astronomical Society 434, 3252.

Ranjan, S., and 6 colleagues 2014. Atmospheric Characterization of Five Hot Jupiters with the Wide Field Camera 3 on the Hubble Space Telescope. The Astrophysical Journal 785, 148.

Knutson, H. A., and 9 colleagues 2014. Hubble Space Telescope Near-IR Transmission Spectroscopy of the Super-Earth HD 97658b. The Astrophysical Journal 794, 155.

Stevenson, K. B., and 7 colleagues 2014. Transmission Spectroscopy of the Hot Jupiter WASP-12b from 0.7 to 5 μm. The Astronomical Journal 147, 161.

Stevenson, K. B., Bean, J. L., Fabrycky, D., Kreidberg, L. 2014. A Hubble Space Telescope Search for a Sub-Earth-sized Exoplanet in the GJ 436 System. The Astrophysical Journal 796, 32.

Seager, S., Sasselov, D. D. 2000. Theoretical Transmission Spectra during Extrasolar Giant Planet Transits. The Astrophysical Journal 537, 916.

de Wit, J., Seager, S. 2013. Constraining Exoplanet Mass from Transmission Spectroscopy. Science 342, 1473.

Feroz, F., Hobson, M. P. 2008. Multimodal nested sampling: an efficient and robust alternative to Markov Chain Monte Carlo methods for astronomical data analyses. Monthly Notices of the Royal Astronomical Society 384, 449.

Feroz, F., Hobson, M. P., Bridges, M. 2009. MULTINEST: an efficient and robust Bayesian inference tool for cosmology and particle physics. Monthly Notices of the Royal Astronomical Society 398, 1601.

Kass, R. E. & Raftery, A. E. 1995. Bayes Factors. *Journal of the American Statistical Association*, 90, 773.

Chen, J., Kipping, D. 2017. Probabilistic Forecasting of the Masses and Radii of Other Worlds. The Astrophysical Journal 834, 17.

Kipping, D. M. 2010. How to weigh a star using a moon. Monthly Notices of the Royal Astronomical Society 409, L119.

Kempton, E. M.-R., Lupu, R., Owusu-Asare, A., Slough, P., Cale, B. 2017. Exo-Transmit: An Open-Source Code for Calculating Transmission Spectra for Exoplanet Atmospheres of Varied Composition. Publications of the Astronomical Society of the Pacific 129, 44402.

Agnor, C. B., Hamilton, D. P. 2006. Neptune's capture of its moon Triton in a binary-planet gravitational encounter. Nature 441, 192.

Berger, T. A., Huber, D., Gaidos, E., van Saders, J. L. 2018. Revised Radii of Kepler Stars and Planets using Gaia Data Release 2. ArXiv e-prints arXiv:1805.00231.

Cilibrasi, M., Szulágyi, J., Mayer, L., Drażkowska, J., Miguel, Y., Inderbitzi, P. 2018. Satellites Form Fast & Late: a Population Synthesis for the Galilean Moons. ArXiv e-prints arXiv:1801.06094.

Ochiai, H., Nagasawa, M., Ida, S. 2014. Extrasolar Binary Planets. I. Formation by Tidal Capture during Planet-Planet Scattering. The Astrophysical Journal 790, 92.

Dalba, P. A., Muirhead, P. S., Croll, B., Kempton, E. M.-R. 2017. Kepler Transit Depths Contaminated By a Phantom Star. The Astronomical Journal 153, 59.

Bryson, S. T., and 10 colleagues 2010. The Kepler Pixel Response Function. The Astrophysical Journal 713, L97.

Schlafly, E. F., Finkbeiner, D. P. 2011. Measuring Reddening with Sloan Digital Sky Survey Stellar Spectra and Recalibrating SFD. The Astrophysical Journal 737, 103.

Kümmel, M., Walsh, J. R., Pirzkal, N., Kuntschner, H., Pasquali, A. 2009. The Slitless Spectroscopy Data Extraction Software aXe. Publications of the Astronomical Society of the Pacific 121, 59.

Bertin, E., Arnouts, S. 2010. SExtractor: Source Extractor. Astrophysics Source Code Library ascl:1010.064.

Kümmel, M., Kuntschner, H., Walsh, J. R., Bushouse, H. 2011. Master sky images for the WFC3 G102 and G141 grisms.Space Telescope WFC Instrument Science Report.

Jones, E., Oliphant, T., Peterson, P., et al. 2001, http://www.scipy.org

Deming, D., and 20 colleagues 2013. Infrared Transmission Spectroscopy of the Exoplanets HD 209458b and XO-1b Using the Wide Field Camera-3 on the Hubble Space Telescope. The Astrophysical Journal 774, 95.

Deming, D., Harrington, J., Seager, S., Richardson, L. J. 2006. Strong Infrared Emission from the Extrasolar Planet HD 189733b. The Astrophysical Journal 644, 560.

Knutson, H. A., and 8 colleagues 2007. A map of the day-night contrast of the extrasolar planet HD 189733b. Nature 447, 183.

Charbonneau, D., and 7 colleagues 2008. The Broadband Infrared Emission Spectrum of the Exoplanet HD 189733b. The Astrophysical Journal 686, 1341.

Freedman, R. S., Marley, M. S., Lodders, K. 2008. Line and Mean Opacities for Ultracool Dwarfs and Extrasolar Planets. The Astrophysical Journal Supplement Series 174, 504.

Freedman, R. S., Lustig-Yaeger, J., Fortney, J. J., Lupu, R. E., Marley, M. S., Lodders, K. 2014. Gaseous Mean Opacities for Giant Planet and Ultracool Dwarf Atmospheres over a Range of Metallicities and Temperatures. The Astrophysical Journal Supplement Series 214, 25.

Lupu, R. E., and 8 colleagues 2014. The Atmospheres of Earthlike Planets after Giant Impact Events. The Astrophysical Journal 784, 27.

Luri, X., Brown, A. G. A., Sarro, L. M., et al. 2018, "Gaia Data Release 2: using Gaia parallaxes", arXiv e-print:1804.09376

Morton, T. 2015, isochrones: Stellar model grid package, Astrophysics Source Code Library.

Foreman-Mackey, D., Hogg, D. W., Lang, D., Goodman, J. 2013, "emcee: The MCMC Hammer", PASP, 125, 306.

Huber, D., Bryson, S. T., Haas, M. R., et al. 2016, "The K2 Ecliptic Plane Input Catalog (EPIC) and Stellar Classifications of 138,600 Targets in Campaigns 1-8", ApJS, 224, 2.

Kipping, D. M., Bakos, G. Á., Buchhave, L., Nesvorný, D., Schmitt, A. 2012. The Hunt for Exomoons with Kepler (HEK). I. Description of a New Observational project. The Astrophysical Journal 750, 115.

Skilling, J. 2004. Nested Sampling. American Institute of Physics Conference Series 395.

Kipping, D. M. 2013. Efficient, uninformative sampling of limb darkening coefficients for two-parameter laws. Monthly Notices of the Royal Astronomical Society 435, 2152.

Kipping D. M., Forgan, D., Hartman, J., Nesvorný, D., Bakos, G. Á., Schmitt, A. R., Buchhave, L. A. 2013, ApJ, 777, 134

Kipping, D. M., and 7 colleagues 2015. The Hunt for Exomoons with Kepler (HEK): V. A Survey of 41 Planetary Candidates for Exomoons. The Astrophysical Journal 813, 14.

Agol, E., Deck, K. 2016. TTVFaster: First order eccentricity transit timing variations (TTVs). Astrophysics Source Code Library ascl:1604.012.

Carter, J. A., Yee, J. C., Eastman, J., Gaudi, B. S., Winn, J. N. 2008. Analytic Approximations for Transit Light-Curve Observables, Uncertainties, and Covariances. The Astrophysical Journal 689, 499.

Nesvorný, D., Kipping, D., Terrell, D., Hartman, J., Bakos, G. Á., Buchhave, L. A. 2013. KOI-142, The King of Transit Variations, is a Pair of Planets near the 2:1 Resonance. The Astrophysical Journal 777, 3.

Szabó, G. M., Pál, A., Derekas, A., Simon, A. E., Szalai, T., Kiss, L. L. 2012. Spin-orbit resonance, transit duration variation and possible secular perturbations in KOI-13. Monthly Notices of the Royal Astronomical Society 421, L122.

Kipping, D. M. 2014. Characterizing distant worlds with asterodensity profiling. Monthly Notices of the Royal Astronomical Society 440, 2164.

Domingos, R. C., Winter, O. C., & Yokoyama, T. 2006, MNRAS, 373, 1227.

Donnison, J. R. 2014, Earth Moon and Planets, 113, 73.

Chapter 4: Loose Ends for the Exomoon Candidate Host Kepler-1625b

"Biggs, Wedge, let's close it up. We're going in, we're going in full throttle."
– Luke Skywalker

4.1 Introduction

Last year, Teachey, & Kipping (2018) (Chapter 3, TK18 hereafter) presented evidence for a large exomoon orbiting the gas giant Kepler-1625b. That work was based on a joint analysis of three transits of the planet observed with *Kepler*, and a fourth transit observed with the Hubble Space Telescope (HST) in October 2017 (GO-15149, PI Teachey). The conclusion was based on the presence of significant transiting timing variations (TTVs) in the system, as well as a sustained dip in the star's brightness following planetary egress. These two lines of evidence were interpreted as self-consistent indications that a large moon is present in the system. A number of alternative explanations for these two signals were explored and the likelihoods of these alternatives were considered. Taken together, the exomoon hypothesis emerged as the best explanation for the data in hand.

Since the publication of TK18, discussions with and amongst colleagues have highlighted open questions and unresolved issues emerging from the analysis. In this work we take the opportunity to address some of these points and present an update on the prospects of confirming or rejecting the exomoon hypothesis for Kepler-1625b.

This paper is structured as follows. In Section 4.2, we explore other systematic models to

This chapter was first published in The Astronomical Journal, Volume 159, Issue 4, id.142 (April 2020). Co-authored with David Kipping, Christopher J. Burke, Ruth Angus, and Andrew W. Howard.
https://ui.adsabs.harvard.edu/abs/2020AJ....159..142T/abstract

account for the long-term trend seen in the TK18 light curve and the effects they have on the interpretation. In Section 4.3 we address differences between our work and that of another group (Kreidberg et al., 2019), whose independent reduction and analysis we became aware of during the course of writing this paper. In Section 4.4, we discuss the possibility that the moon-like dip is in fact caused by a second, previously undetected transiting planet in the system. In Section 4.5, we provide a more detailed assessment of the host star's activity and investigate the possibility that it could be responsible for the moon-like dip. In Section 4.6, we use forward propagation of the TK18 solution to determine the location and probability of seeing exomoon transits in future epochs. Conclusions are summarized in Section 4.7.

4.2 Other Systematic Models

4.2.1 Overview

TK18 employed three different models to account for the long-term trend seen in their data. These were broadly motivated to follow as closely as possible the most standard approaches in the literature for previous WFC3 analyses (see Wakeford et al. (2016) for an overview of WFC3 systematics). Most authors have previously elected to use a simple linear trend for this correction, of the form $a_0 + a_t(t - t_0)$ (e.g. see Huitson et al. 2013; Ranjan et al. 2014; Knutson et al. 2014). In some rarer cases, a quadratic model has been invoked (Stevenson et al., 2014a,b) and thus both of these models were attempted. A third exponential model was also attempted giving three trend models in total, all with time as the independent variable.

Gaussian processes (GPs) have also been utilized in, for example, Evans et al. (2018) to handle WFC3 systematics. In general, however, GPs are not obviously appropriate for the moon search unless there is reason to suspect the data are not drawn from a sequence of independent Gaussians. As we show later, we see no evidence of time-correlated noise structure. But the flexibility of GPs mean they will inevitably fit out a moon-like dip, and insomuch as less flexible detrending models explored here are also capable of attenuating

or removing the moon signal, invoking GPs here is neither well-motivated nor particularly illuminating.

TK18 argued that it was crucial to perform this detrending simultaneous to the transit fits, repeating for each model (planet/moon/TTV), to account for the fact that the trend model appeared highly covariant with the moon-like dip. In comparing these models, we re-emphasize here that full Bayesian evidences should be used. As a non-linear model (Kipping, 2011), the number of degrees of freedom cannot be estimated and thus reduced χ^2 comparisons are certainly invalid (Andrae et al., 2010). Another popular alternative to computing evidences is the Bayesian Information Criterion (BIC, Schwarz, 1978). This was used, for example, in Kreidberg et al. (2019). But again here there are serious concerns about its use for this problem. By invoking a Laplacian approximation on the posterior, one approximates the posterior to a Gaussian centered on the maximum likelihood estimator, which is inappropriate for highly multi-modal posteriors such as those resulting from exomoon fits (Kipping et al., 2012). Further, the BIC is not guaranteed to yield a Bayes factor which is close to one computed using priors an observer would consider appropriate, since it assumes the unit information prior on the model parameters (Weakliem, 1999). For these reasons, model comparisons are performed using the Bayesian evidence in what follows.

The three trend models considered by TK18 allowed for an offset between the two visits, which was most apparent in the extreme channels and to a lesser degree in the white light curve. Clearly the models explored by TK18 are a small subset of an essentially infinite number of possible models one could try. In general, the more flexible the model, the easier it is to fit out the moon-like dip when assuming no-moon present. It was for this reason that more flexible trend models were not explored by TK18, since any sufficiently flexible instrument model can fit-out interesting astrophysics.

Nevertheless, this was neither demonstrated nor investigated in detail in that work. For this reason, we re-visit the trend modeling here exploring a) the effect of going to higher-order polynomials b) the effect of allowing for discontinuous trend models c) the effect of

changing the dependent variable.

4.2.2 Higher-order polynomials

Although an infinite number of polynomials exist beyond a quadratic trend, we here perform a cubic model as a simple extension to illustrate the effects. We re-ran the moon (M) and zero-radius moon (Z) models on the TK18 using MultiNest (Feroz & Hobson, 2008; Feroz et al., 2009) and `LUNA` (Kipping, 2011), as was done in TK18, except we add an additional cubic term to the quadratic trend model.

The resulting maximum *a posteriori* light curve is shown in the second row of Figure 4.1. The shape closely matches that of the exponential model shown above it (and indeed the quadratic model form TK18), reflecting the fact that the cubic term is almost zero. In fact, we can quantify this statement by evaluating the $1\,\sigma$ credible interval of the marginal posterior of the cubic coefficient to be 630^{+430}_{-450} ppm, reflecting how the posterior is only $1.4\,\sigma$ from zero.

It is therefore not surprising that a) the cubic fit returns a similar Bayes factor in favor of the exomoon to the original models of TK18, and b) the cubic fit has an overall lower evidence than the other models ($\log \mathcal{Z}_M = 6311.34 \pm 0.16$) since it includes effectively wasteful parameter volume. In other words, the model has been penalized for additional complexity.

4.2.3 Discontinuous polynomials

TK18 only considered models where the function is continuous across the visit change, except for a flux offset. This means that the higher-order polynomial coefficients, a_t and a_{t2} in the case of a quadratic model, are the same on both sides of the visit change. The motivation for this was that a) the small trends either side of the visit change did not correlate with centroid position and thus did not appear to be instrumental in nature, and b) the star's intrinsic variability should not change dramatically on either side of the visit change.

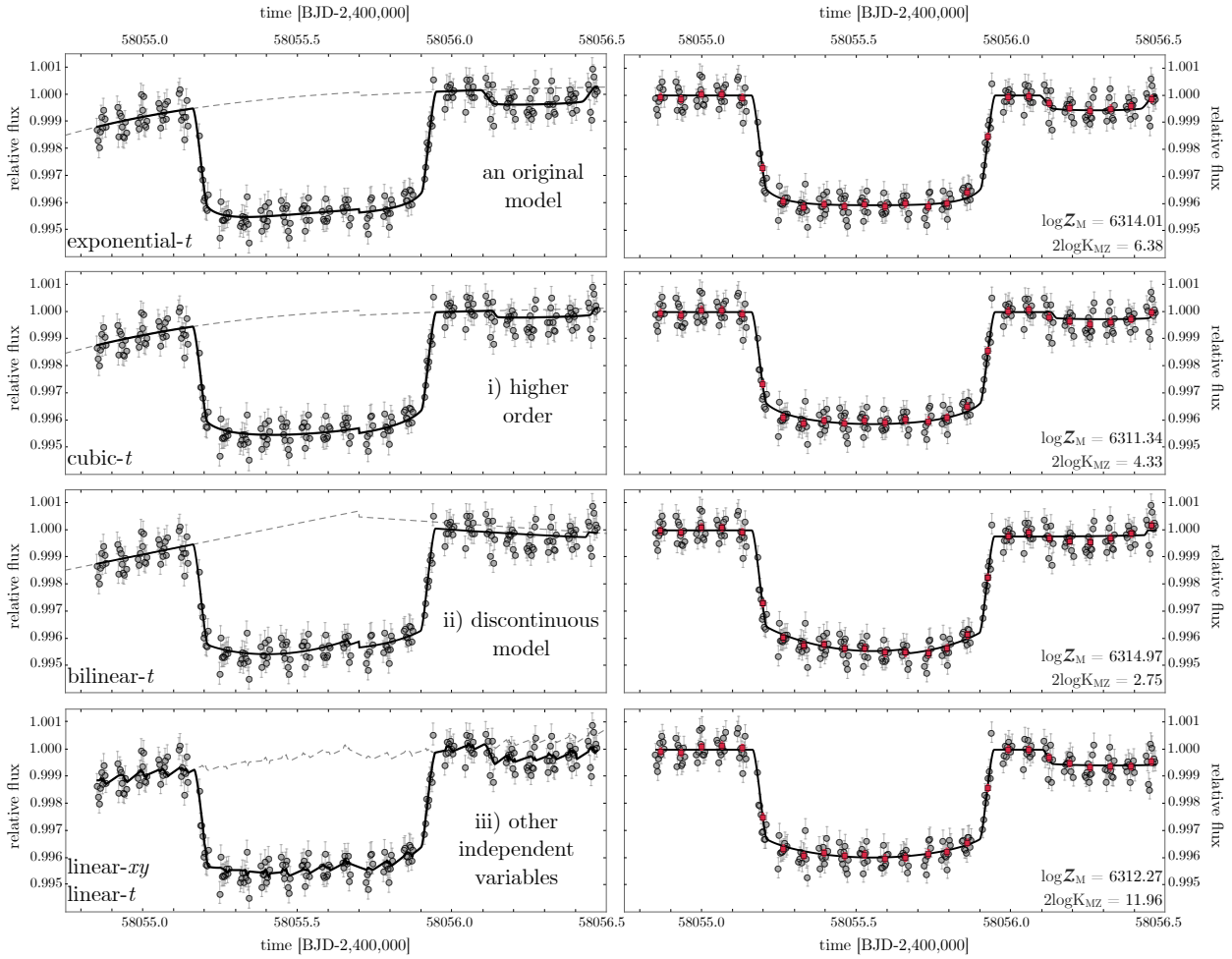

Figure 4.1: Comparison of three different long-term trend models applied to the TK18 data. The left column shows the uncorrected data with the trend model overlaid while the right column shows the post-correction data. More flexible systematic models than those considered by TK18 attenuate the formal exomoon evidence and also find alternative modes that are inconsistent with the TK18 candidate signal.

We discuss here the effect of relaxing this assumption. As in the case of higher-order polynomials, this essentially represents a more flexible model. The simplest discontinuous polynomial is two independent straight lines ("bilinear" in what follows). As in TK18, all detrending choices explored in this work are implemented after the hook correction has been applied. Because the hook correction minimizes intra-orbit RMS independent of the model under consideration, there is no covariance between them and therefore the evidences are not impacted. We refer readers to TK18 for a more thorough discussion of this choice.

We find that the fits favor a very pronounced reversal in the systematic gradient located at the visit-change, as can be seen from the third row of Figure 4.1. It is unclear how this behavior could manifest physically, since flux is apparently uncorrelated with centroids in each individual visit for the comparison star KIC 4760469 (see Figure S10 and Section 1.2.10 of Teachey, & Kipping 2018). That is, while there is no known explanation for the visit-long trends, we would *a priori* expect the target star and the comparison star to display similar systematic morphologies. The fact that they do not leads us to question whether a downturn in the target star trend can be attributed to instrumental systematics. At the same time, we see no reason to expect the star to exhibit a pronounced reversal coincidental with HST's visit change. The second visit slope has a negative gradient that absorbs the decrease in brightness caused by the moon-like dip, and for this reason the evidence is significantly attenuated for the exomoon model. Indeed, the moon solution is quite distinct from the original paper and can be immediately dismissed as suspicious because the moon ingress is nearly coincident with the flux offset associated with the visit change.

The bilinear model has two free parameters per visit, giving four in total - the same as the number of free parameters describing the TK18 quadratic and exponential models. Despite having the same number of free parameters, it does not necessarily have the same degrees of freedom. These two concepts are distinct if the underlying model is non-linear (Andrae et al., 2010), which is true here because of the non-linear step function occurring at the visit change. In fact, it can be seen that the bilinear model in fact has more freedom, because

it does not require a continuous gradient across the boundary, unlike the case of the TK18 quadratic model.

It is therefore perhaps not surprising that this more flexible model is able to fit-out the moon-like dip sufficiently well to find no evidence for the putative moon. This analysis does not particularly add to or detract from the exomoon hypothesis, since the behavior can be understood as a byproduct of employing more flexible trend models. Clearly the attenuation of the moon-like dip does not necessarily imply that the trend model is incorrect; strictly speaking the conclusion is simply that the moon+flexible model is not supported by the data. Put another way, the moon model and the planet-only model are essentially equiprobable with this trend model, and therefore we would not claim evidence for a moon. We find no physical or data-driven motivation for adopting the bilinear model, nor is there precedent for doing so in the literature. On the other hand, the observation is unprecedented in several ways (e.g. the faintness of the target, the 40 hour duration, and the objective itself), so we cannot rule out the possibility that we are observing unique or heretofore only marginally important systematic effects.

4.2.4 Changing the independent variable

We now consider a third and final modification to the systematic model, namely, modeling the systematics as a function of both time and centroid position, rather than simply a function of time. We start by taking our simplest model, the linear-t model, given by

$$S(t) = a_0 + a_{t1}(t - t_0), \tag{4.1}$$

and extending it to include a linear dependency on x and y centroid positions:

reduction	$2\log(\mathcal{Z}_{M,\mathrm{coplanar}} - \mathcal{Z}_Z)$
linear-t	1.05 ± 0.32
quadratic-t	-0.28 ± 0.33
exponential-t	-0.01 ± 0.33

Table 4.1: Repeat of the TK18 model fits but enforcing the condition that the moon must be coplanar i.e. $i_S = 90°$.

$$\begin{aligned} S(t,x,y) =& a_0 + a_{t1}(t - t_0) \\ &+ a_{x1}(x - x_0) + a_{y1}(y - y_0), \end{aligned} \tag{4.2}$$

where a_i are coefficients to fit for and the subscript 0 variables represent the median time/centroid positions. Using the same photodynamical MULTINEST fitting software from TK18, the resulting maximum *a posteriori* light curve is shown the fourth row of Figure 4.1.

The figure, as well as the evidences quoted in the panel, show that the same moon is again detected. We therefore conclude that adding x and y as linear independent variables to the systematic model does not significantly affect the conclusions of TK18.

4.2.5 Fixing orbits to coplanar

The inclined solution for the exomoon candidate K1625b-i is particularly curious. The fact the posteriors favor an inclined solution suggests that it should be very difficult to fit the same moon to the existing data (both *Kepler* and HST) if one imposes coplanarity. To investigate this, we repeat the three trend model fits of TK18 for the M models but fix $i = 90°$. Comparing the resulting evidences to the original Z models of that work ($2\log\mathcal{Z}_{MZ} = 17.77 \pm 0.33$, 3.61 ± 0.33, and 6.38 ± 0.34 for the linear, quadratic and exponential models, respectively) indeed show the case for an exomoon is removed (see Table 4.2.5). This highlights the importance of including inclination in such fits.

label	systematics model	$\log \mathcal{Z}$
linear-t^*	$a_0 + a_{t1}(t-t_0) + (a'_0 - a_0)\mathcal{H}[t-t_{II}]$	-0.60 ± 0.06
quadratic-t^*	$a_0 + a_{t1}(t-t_0) + a_{t2}(t-t_0)^2 + (a'_0 - a_0)\mathcal{H}[t-t_{II}]$	0.00 ± 0.06
exponential-t^*	$a_0 + a_{e1}\exp(\frac{t-t_0}{a_{e2}}) + (a'_0 - a_0)\mathcal{H}[t-t_{II}]$	-0.38 ± 0.06
linear-$xy^\dagger$	$a_0 + a_{x1}(x-x_0) + a_{y1}(y-y_0)$	-0.61 ± 0.06
linear-xy linear-t	$a_0 + a_{t1}(t-t_0) + a_{x1}(x-x_0) + a_{y1}(y-y_0)$	-4.29 ± 0.06
linear-xy quadratic-t	$a_0 + a_{t1}(t-t_0) + a_{t2}(t-t_0)^2 + a_{x1}(x-x_0) + a_{y1}(y-y_0)$	-3.96 ± 0.07
quadratic-xy	$a_0 + a_{x1}(x-x_0) + a_{x2}(x-x_0)^2 + a_{y1}(y-y_0) + a_{y1}(y-y_0)^2$	-1.27 ± 0.06
quadratic-xy linear-t	$a_0 + a_{t1}(t-t_0) + a_{x1}(x-x_0) + a_{x2}(x-x_0)^2 + a_{y1}(y-y_0) + a_{y1}(y-y_0)^2$	-4.97 ± 0.07
quadratic-xy quadratic-t	$a_0 + a_{t1}(t-t_0) + a_{t2}(t-t_0)^2 + a_{x1}(x-x_0) + a_{x2}(x-x_0)^2 + a_{y1}(y-y_0) + a_{y1}(y-y_0)^2$	-3.43 ± 0.07

Table 4.2: Bayesian evidences from applying various systematic models to the comparison star KIC 4760469. All evidences are quoted with 1334.18 subtracted - the absolute value obtained for the second model listed - "quadratic-t". Models with a * indiciate that this one of the original models used by TK18. Models with a † are those used by Kreidberg et al. (2019). The model considered by Kreidberg et al. (2019) is formally indistinguishable from the systematic models used in TK18, and are therefore not favored over those used in that work.

4.2.6 Using the comparison star as model benchmark

Since the comparison star is expected to be stable (TK18), it provides a useful control test for comparing the different possible systematic trend models. Expanding to quadratic order in x, y and t, we fitted nine different models to the comparison star (assuming an intrinsically flat baseline) using MultiNest. The various models and resulting evidences are listed in Table 4.2.

In TK18, only three of these models were considered but it turns out none of the other six models proposed here yield an evidence superior to the simple time models. We conclude that analysis of the comparison star indicates that systematic models using x and y centroid positions are not supported by the current data.

4.3 Comparison to Kreidberg et al. (2019)

During the final preparations of this paper, it came to our attention that Kreidberg et al. (2019) (henceforth KLB19) had conducted an independent reduction of the HST WFC3 observations of Kepler-1625 and concluded that there was no evidence for an exomoon based

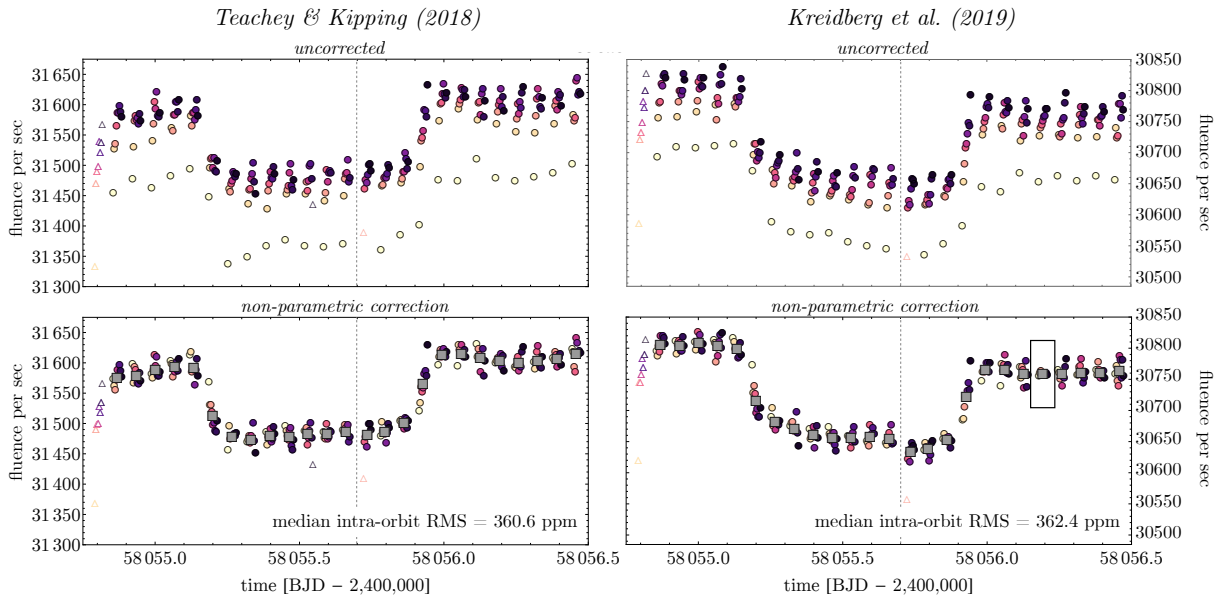

Figure 4.2: A comparison of light curves before and after the hook correction from TK18 and KLB19. Compare to Figure 2 in TK18. As in that work, the data are color coded by the observation number within each HST orbit (light yellow for the first observation, dark purple for the last). The grey squares in the bottom panel represent the binned flux for each orbit. Triangles indicate observations from the first orbit, which are left out of the hook correction normalization. the anomalously low scatter in the 22nd orbit of the KLB19 analysis is highlighted with a rectangle.

on the apparent lack of a moon-like dip following planetary egress. We will compare the KLB19 reduction and results to that of TK18 in what follows.

4.3.1 Raw photometry comparison

It is instructive to make a side-by-side comparison of the raw photometry presented in TK18 and the new reduction by KLB19 before any hook or trend corrections have been applied, which is shown in the top panel Figure 4.2.

Before any systematic effects have been corrected, the photometry from both groups is apparently quite similar, yet important differences are evident. Quick inspection of both light curves reveals a much more pronounced offset in flux for KLB19 occurring at the instant of the visit change between orbits 14 and 15 (marked by the vertical dashed lines in that

figure). As described in TK18, the full guide star acquisition performed at the beginning of orbit 15 was responsible for the introduction of this offset. Detailed modeling (described in Section 4.3.6) finds that the amplitude of the offset increase in every case from the TK18 reduction to that found in KLB19:

- from (20 ± 110) ppm in TK18 to (-900 ± 120) ppm in KLB19, for the linear-t Z model;
- from (-140 ± 120) ppm in TK18 to (-850 ± 130) ppm in KLB19, for the quadratic-t Z model; and
- from (-20 ± 110) ppm in TK18 to (-880 ± 110) ppm in KLB19, for the exponential-t Z model.

There is of course only one ground-truth in terms of the motion of the telescope, and astrophysical variation. Because we can reasonably assume the star itself is not exhibiting a sudden change in flux after the 14^{th} HST orbit, the discontinuity there must be systematic. Thus, a larger discontinuity could be viewed as being farther from the star's ground-truth, requiring a more substantial correction that could impact the results of KLB19.

4.3.2 Hook correction

We next applied the exact same hook correction algorithm described in TK18 to the KLB19 reduction. KLB19 also uses the non-parametric approach of TK18, thereby providing a fair comparison of the two, and this is shown in the bottom panel of Figure 4.2).

The mean intra-orbit photometric RMS from KLB19 is somewhat smaller at 360.7 ppm, versus 374.8 ppm for TK18[1]. At first glance, this appears to indicate that the KLB19 reduction is more precise. However, inspection of Figure 4.2 reveals that the 22nd HST orbit appears to display an anomalously low scatter of just 85 ppm. While TK18 also find that

[1]TK18 quote 375.5 ppm but that value is the mean intra-orbit RMS after 10 rounds of hook correction iterations, whereas the final light curve actually uses 100 rounds.

this orbit has the lowest scatter, the RMS is much more consistent with the other orbits, at 210 ppm.

For normally distributed data, the standard deviation of sample standard deviations equals $\sigma/\sqrt{2(n-1)}$. Since the mean RMS for TK18 is σ = 374.8 ppm, and the average number of points per orbit is 8.8, then one should expect RMS values with a standard deviation of 94.9 ppm. The actual standard deviation of RMS values is less than one percent larger at 95.5 ppm. This in turn means that the 210 ppm smallest RMS value is only 1.7 σ from the mean.

For the KLB19 orbits, the expected standard deviation in RMS values is 91.3 ppm and the observed value is 8.6% higher at 99.2 ppm. Critically, the 22nd orbit now appears to be a 3 σ outlier. Strictly speaking, the formula above is only valid for $n \gg 1$, so we are at the limit of applicability in the present case. Thus, a better estimate for how surprising this orbit is can be obtained by masking the orbit, taking the mean of the other RMS values, and then generating fake Gaussian data for all orbits and Monte Carlo evaluating the expected distribution of RMS values. This reveals that the 22nd orbit from KLB19 is anomalously low at the 4.0 σ level.

This seems highly implausible from a statistical perspective and would make the 22nd orbit intra-orbit RMS an outlier by most definitions. In the presence of outliers, a more robust summary statistic to compare the precision of each reduction is the *median* intra-orbit RMS, rather than the mean. On this basis, the original TK18 reduction is marginally more precise at 360.6 ppm versus 362.4 ppm for KLB19. That is, they are essentially indistinguishable from the perspective of their noise profile. The source of this improbably low scatter in the 22$^{\rm nd}$ orbit is unknown, but it appears to be present in the data before the application of the hook correction.

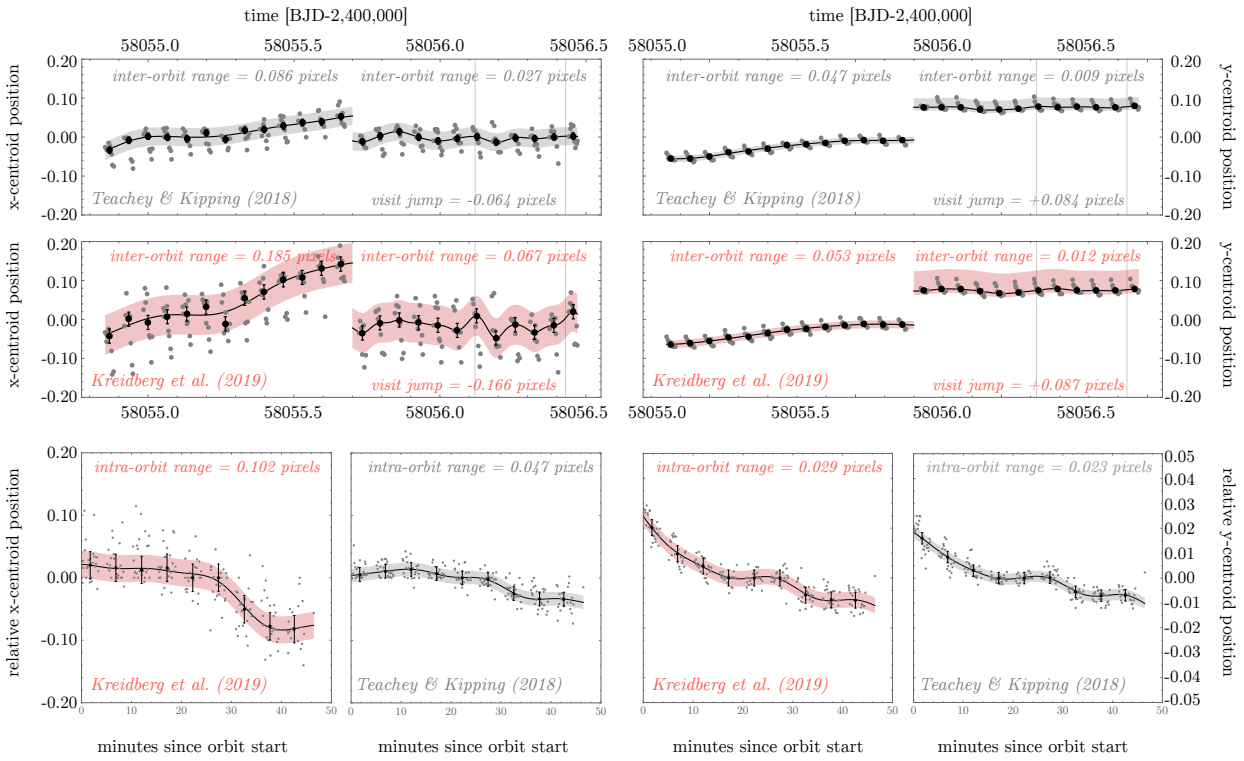

Figure 4.3: Comparison of the centroids reported by TK18 (gray) and those of KLB19 (red). In all cases, we find that the KLB19 centroids exhibit larger variations.

4.3.3 Centroids

Centroids deserve special attention since KLB19 use the target's position on the detector as the basis for their systematics model correction. TK18 presented their centroid variability in Figure 3.13, for both Kepler-1625 and the comparison star KIC 4760469. Figure 4.3 directly compares the centroids of TK18 to those of KLB19 for Kepler-1625, where morphological similarities are apparent.

As was found in TK18, there is substantial variation of the apparent centroid position within an orbit, which we attribute to the hook and / or breathing effects rather than a real variation. For this reason, long-term behavior (associated with pointing drift) is best tracked using the orbit median centroid positions, shown in black in Figure 4.3.

We find that the range in inter-orbit Y-centroid position is 10% higher for the first visit

and 30% higher for the second visit in the KLB19 reduction than that of TK18, and 2.2 to 2.5 times higher for the X-centroid position. Similarly, the change in centroid position after the visit change is 2.6 times higher in X for KLB19 than TK18.

We tried offsetting the median centroid of each orbit and then orbit-folding (see lowest row of Figure 4.3) to look at intra-orbit centroid variations, rather than inter-orbit. As before, we find higher intra-orbit centroid variability for the KLB19 reduction, increased by similar levels.

The origin of these centroid discrepancies is unclear. Systematic effects such as the hook and breathing effects likely play a role in the calculated position for each image, as suggested by the intra-orbit centroid variations which do not appear to be associated with pointing drift. Different handling of these systematics could therefore reasonably explain the discrepancy. due to the image rotation peformed in TK18, it is not possible to apply our centroid corrections to the KLB19 reduction, nor is it possible to use the KLB19 centroids for detrending our extracted light curve.

We also point out that the calculation of these centroids is handled differently in TK18 and KLB19. TK18 simply calculates the flux weighted centroid of the optimal aperture at every time step as

$$X_{centroid} = \frac{\sum_i^N x_i f_i}{\sum_i^N f_i} \tag{4.3}$$

where x_i a an f_i are the pixel coordinate and flux of pixel i, respectively. Calculation of the centroid in Y is identical. TK18 performed this operation for both the target star and the comparison star KIC4760469, which showed good agreement.

By contrast, KLB19 perform a more complicated analysis to compute the motion, with different methodologies for the x and y directions. For the y (or "spatial") direction, KLB19 sum the flux in each column of the image at each time step, perform a 4-pixel Gaussian convolution of the resulting array and then an interpolation to compute a best fitting offset from a template at each time step. It is not obvious how spatial information is recovered

reduction	a_{x1}	a_{y1}
TK18, Kepler-1625	$+2700^{+910}_{-900}$	$+4780^{+440}_{-450}$
TK18, KIC 4760469	$+580^{+970}_{-960}$	$+3050^{+610}_{-620}$
KLB19, Kepler-1625	$+830^{+400}_{-400}$	-10170^{+410}_{-410}

Table 4.3: Using the systematic model "linear-xy" only (which assumes no part of the trend is dependent on time), we compare here the parameters a_{x1} and a_{y1} (i.e. the centroid gradient terms) which result from three different reductions. Elements list median and 68.3% credible intervals in units of parts per million per day.

from this algorithm as described in KLB19, nor whether comparison to a template could introduce biases. KLB19 perform a similar operation for the x (or "spectral") direction, though now only summing up along the target spectrum instead of along each row in the image. The result, as shown in Figure 4.3 is morphologically similar to TK18, though with larger systematics.

4.3.4 Systematic trend comparison

KLB19 use systematic models which decorrelate against X and Y centroid position rather than just time. Since the previous subsection has conducted a like-for-like comparison of this model, it is instructive to inspect the systematic parameter posteriors that result. We list these values in Table 4.3.

As can seen be seen from the table, KLB19 find an overall stronger dependency between flux and centroid position than TK18, with almost all of the variability coming in via the Y-direction. It is also worth noting that the sign of a_{y1} reverses for the KLB19 reduction.

We also remind the reader that our earlier comparison of different trend models applied to the comparison star found that models including X-Y pixel position were disfavored (Section 4.2.6) over temporal models. In any case, it is difficult to determine the degree to which these discrepancies arise from the different centroiding approach, and how much is due to differences in the raw fluxes owing to the reduction itself.

4.3.5 Model evidences comparison

TK18 perform model comparison using the Bayes factor calculated using Bayesian evidences (marginal likelihoods). In contrast, KLB19 perform their model selection using reduced χ^2 and the Bayesian Information Criterion (BIC). As discussed earlier in Section 4.2.1, model comparison using the reduced χ^2 is invalid for non-linear models and it is thus not appropriate for transit light curve fits. The BIC is also inappropriate due to the multimodality of the posterior, which is poorly described by the Laplacian approximation used by BIC. Further, it is generally not guaranteed to produce an approximation of the Bayes factor (Weakliem, 1999), and indeed it has been argued to not even represent an approximation to any exact Bayesian solution - including the Bayes factor (Gelman & Rubin, 1995). Accordingly, we strongly urge the avoidance of these tools for exomoon model selection.

To perform a full comparison between the two reductions, it is instructive to repeat the full photodynamical MultiNest fits conducted by TK18 on the KLB19 reduction. This allows us to evaluate what the Bayes factor would be for the exomoon had we used this data set instead.

We fit the hook-corrected light curves of KLB19 using the same three models used by TK18 - linear in time, quadratic in time and exponential in time, all of which also include a flux offset parameter at the visit change. Further, we ran the photodynamical MultiNest fits for both the TK18 and KLB19 adopting a fourth systematics model - one motivated by the choice of KLB19 to decorrelate against centroid position. Specifically, this model is linear in time as well as in X and Y centroid position i.e. an example of changing the independent variable. The results of these fits are summarized in Table 4.4.

As can be seen from the table, the KLB19 reduction consistently yields lower Bayes factors for the moon solution versus that found by TK18. Although a moon-dip is favored in all cases (contrasting with the BIC and reduced χ^2 testing of KLB19), the strength of the evidence is diminished to such a degree that we would not consider it justifiable to claim evidence for an exomoon. Combined with the investigation described earlier in Section 4.2.4,

model	TK18	Kreidberg et al. (2019)
linear-t^*	17.77 ± 0.33	1.08 ± 0.32
quadratic-t^*	3.61 ± 0.33	1.38 ± 0.32
exponential-t^*	6.38 ± 0.34	1.88 ± 0.33
linear-xy linear-t	11.96 ± 0.34	0.56 ± 0.34

Table 4.4: Top: Bayesian model evidences using different formalisms for the systematic model, and comparing two different reductions of the HST WFC observations of K1625. Each element represents $2\log(\mathcal{Z}_M - \mathcal{Z}_Z)$ - the Bayes factor for the exomoon. * = original fits from TK18.

this strongly suggests that the differing conclusions between TK18 and KLB19 is not due to the choice of systematic model, but rather due to the reduction itself. This is the same conclusion reached by KLB19. However, we do not agree that the moon's existence has been ruled out, particularly in light of a second independent reduction and analysis carried out by Heller et al. (2019), which also finds evidence for the moon-like dip following the planet's transit.

We point out that Nelson et al. (2020) found through an extensive comparison of approaches to computing model evidences that uncertainties are likely to be underestimated. As such the uncertainties quoted in Tables 4.2 and 4.4 may be too low. For each run we used 4000 live points, which is twice the recommended number for accurate evidence uncertainties (Feroz & Hobson, 2008). In any case, artificially low uncertainties would not invalidate the salient features of our argument here, namely, that 1) we see no strong impetus to adopt a detrending model based on centroids (see Section 4.3.3), and 2) evidence for the moon is considerably weaker based on the KLB19 light curve.

4.3.6 Model residuals comparison

The null hypothesis is that no moon is present around K1625 and so the obvious place to conduct a residual analysis is on the no-moon models (model Z).

The original residual analysis conducted by TK18 (see Figure 3.20) shows that without a moon there appears to be high time-correlated noise when inspecting simple RMS vs bin-

size style diagrams. However, as shown in that same figure, the origin of the time-correlated noise excess is apparently localized in time to the specific point where TK18 claim evidence for a moon-like dip.

A fairer test of residual noise is then to continue using the null hypothesis but mask out the region where TK18 claim a photometric anomaly associated with a possible moon. To accomplish this, we compute the maximum *a posteriori* model residuals for the exponential-t model (since this is the model used for light curve comparison by KLB19 in their Figures 3 & 4), for both the TK18 and KLB19 reductions, and then mask out the region $t > 2458056.1$ BJD. This also conveniently removes orbit 22 of KLB19, which is argued to show anomalously low scatter in Section 4.3.2.

We then compute an RMS vs bin-size diagram, as shown in Figure 4.4. We find that both reductions display Gaussian-like behavior with no clear indications of excess noise. Without any binning, the RMS values are 369.0 ppm for TK18 and 370.3 ppm for KLB19, which are effectively identical.

4.3.7 Presence of TTVs

The sustained moon-like dip in the HST observation observed by TK18 is one important element of the case for the exomoon. However, another critical, self-consistent component of the case for the exomoon presented in TK18 is the presence of TTVs in the system. A large moon like the one described in TK18 is expected to exert a significant gravitational influence on the planet, detectable in the photometry in the form of timing variations.

As described in TK18, the transit of Kepler-1625b in the HST observation occurred a full 78 minutes earlier than anticipated based on a linear ephemeris calculated from the three transits of the planet observed by *Kepler*. This indicates the presence of TTVs to $\sim 3\sigma$ confidence.

We fit the transit timings for the KLB19 light curve and find $\tau = 58055.5539^{+0.0013}_{-0.0012}$, $58055.5538^{+0.0013}_{-0.0012}$, and $58055.5539^{+0.0013}_{-0.0012}$ for the linear-t, quadratic-t and exponential-t mod-

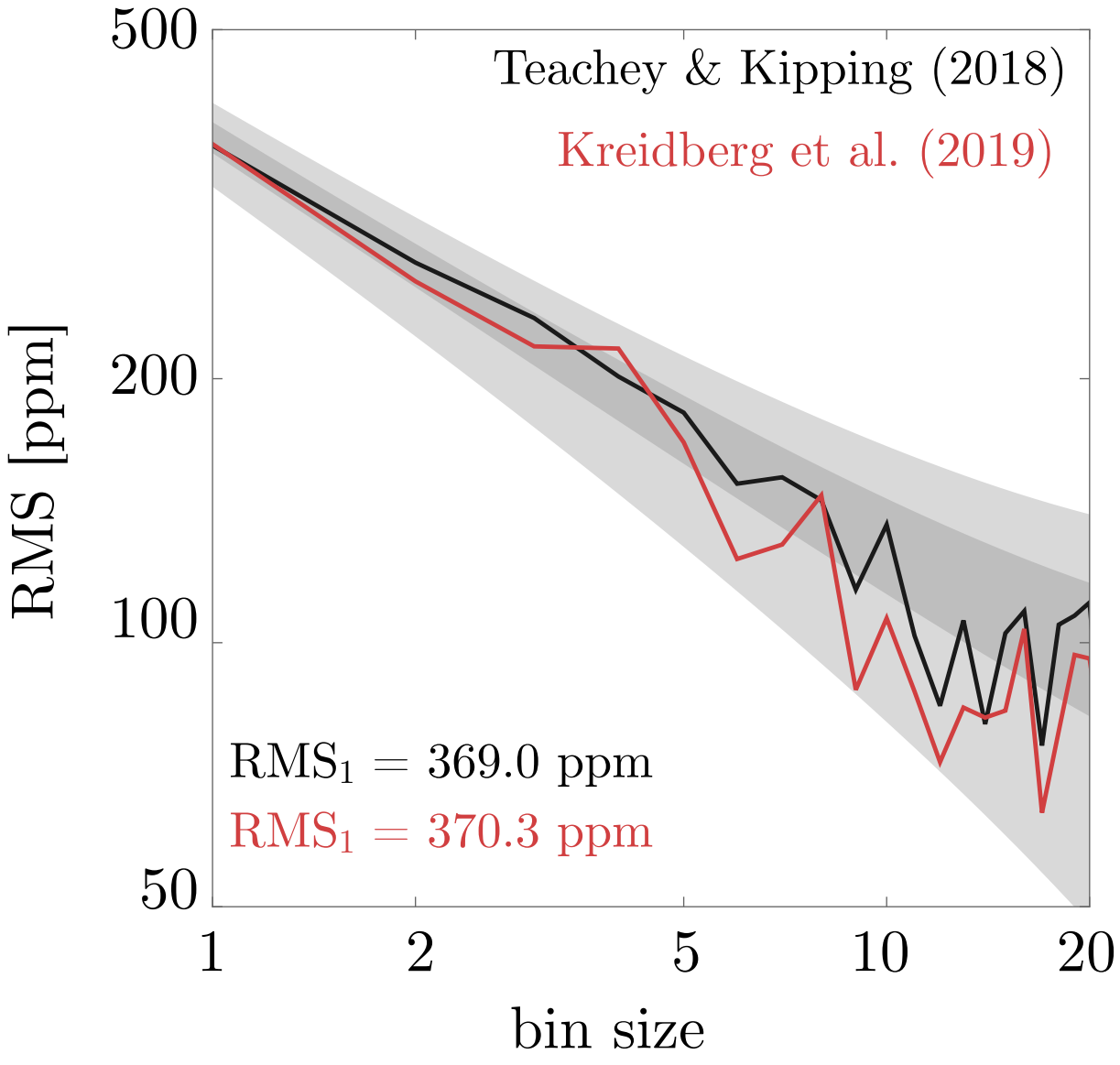

Figure 4.4: RMS vs bin-size diagrams for the exponential-t model regressed to the TK18 reduction (black) and the KLB19 reduction (red). In both cases the model assumes no moon but masks the region $t > 2458056.1$ BJD where the moon-like dip is seen by TK18. Both reductions appear consistent with Gaussian noise properties (the gray 1 and $2\,\sigma$ regions shown).

els, respectively. Comparing this to the value in TK18 of $\tau = 58055.5563^{+0013}_{-0014}$, we consider the presence of TTVs to be validated by the new analysis, and it is worth noting that the new reduction actually suggests the HST transit occurred even earlier than was found in TK18.

4.3.8 Summary

We have executed a detailed comparison of the KLB19 reduction and that of TK18. KLB19 argue that there is in fact no evidence for a moon in their light curve based on the absence of the moon-like dip. We also find that their reduction does not strongly support the presence of the moon-like dip (although it is still formally favored using a Bayesian model comparison), after applying the same hook correction and full Bayesian photodynamical model selection methods used by TK18 (see Section 4.3.6). We note, however, that the KLB19 reduction validates the presence of TTVs in the system, though TTVs alone do not constitute sufficient evidence for a moon.

The question naturally arises as to why the two studies yield different results and which one is ultimately correct. We have argued that there are two major differences between the TK18 analysis and that of KLB19, and so presumably one (or both) of these is responsible for the discrepancy. The first is the choice of systematic model and the second is the independent reduction itself.

Systematic model?

The first major difference between TK18 and KLB19, as explored in Section 4.2.6, is that KLB19 use a systematic model to correct for the long-term trend correlating flux with X and Y centroid position, while TK18 only decorrelate against time. TK18 found no correlation between flux and centroid for the comparison star KIC 4760469 and in this work we have shown that amongst a broad suite of possible models, some with and some without such correlations, models including X and Y correlations are consistently disfavored (see

Section 4.2.6).

Even so, this does not address whether this different choice in detrending is ultimately responsible for the overall differing conclusions. We conclude that the detrending choice is unlikely to be the underlying cause, since re-fitting the original TK18 data including centroid correlation terms still recovers the same exomoon signal to comparable confidence as before (see Section 4.2.4).

Reduction?

With the detrending choice shown to be an unlikely explanation for the discrepant conclusions, we turn our attention to the reduction itself. There are certainly differences between the two reductions, both with respect to the methodologies (described in their respective papers) and the results.

With regard to the methodology, the KLB19 pipeline clearly has a track record that the TK18 reduction does not. Even so, the present observation is unprecedented in several ways. The star is significantly fainter than previous HST targets, the duration of the observation is far longer than typical transmission spectroscopy observations, and the nature of the pursued signal is fundamentally different. Therefore, it is reasonable to ask whether the KLB19 pipeline is guaranteed to be better than the one we have developed.

We note also that the procedure for selecting an optimal aperture as described in KLB19 is potentially problematic for the moon search. Their approach is to explore various apertures until they find the one which minimizes scatter with respect to the transit model. This differs from our approach which does not assume a model. We can only guess that given the computational expense of running a full exploration of parameter space with an MCMC simultaneous to the selection of an aperture, a static planet-only model was assumed and the scatter was minimized with respect to it. This approach could inadvertently incentivize the selection of an aperture for which the moon signal is attenuated. Nevertheless, the final aperture selected by KLB19 is quite similar to that of TK18, the primary differences being

a 13% smaller aperture for KLB19, which probes slightly farther into the blue and a bit less into the red than the aperture of TK18.

We also identified anomalous behavior with orbit 22 of the HST observation as produced by the KLB19 pipeline, which shows suspiciously low photometric scatter. We are unable to determine the source of that anomaly, however.

KLB19 states that the moon-like signal presented in TK18 is "likely an artifact of the data reduction." However, no faults with the original reduction pipeline were found, nor was any step in the reduction pipeline identified as being the source of the moon-like dip. Therefore it is perhaps more accurate to conclude (as we do here) simply that the different pipelines have produced different results. Of note, a recent analysis by Heller et al. (2019), using their own independent reduction pipeline, also recovered a moon-like signal very similar to that presented in TK18. As such, the original interpretation of the data presented in TK18 has now been both validated and called into question in the literature. We thus argue that the existence of the moon remains an open question and additional observations are warranted.

To summarize Sections 4.3.1, 4.3.2 and 4.3.3, we find that the product of the KLB19 reduction:

- exhibits marginally higher median intra-orbit RMS (362.4 ppm versus 360.6 ppm) after correcting for the hooks,
- has a ≃900 ppm larger flux offset at the visit change,
- has ≃2 times larger variations in the y-centroid positions,
- has a x-centroid flux correlation coefficient ≃ 3.5 times greater, and with opposite sign to KIC 4760469,
- exhibits a marginally higher residual RMS (370.3 ppm vs 369.0 ppm) after fitting out a "no-moon model" and masking the claimed moon region in both reductions

Accordingly, we argue that the KLB19 reduction is not obviously superior in any measurable way.

4.4 Second Transiting Planet?

4.4.1 Overview

One possible false-positive scenario for the moon-like dip that was not discussed in TK18 was the possibility that the dip is real but caused by a second transiting planet, not a moon. This scenario was not investigated in the original paper because of the location of the dip with respect to the TTV offset - indicating a strong case for the exomoon hypothesis - as well as the inherently unlikely possibility that a planet could have evaded detection by *Kepler* but appear in this small segment of HST data. Nevertheless, this is certainly a valid concern, and the probability of this scenario was not quantified in the original paper, so we address it here.

We express the probability that the moon-like dip was caused by a second (hypothetical) transiting planet, K1625c, with orbital period P_c, as

$$\mathbb{P}_c = \Pr(\mathcal{T}, \bar{\mathcal{D}}_{\rm Kep}, \mathcal{D}_{\rm HST} | P_c), \tag{4.4}$$

where $\mathcal{T}$ is short-hand for the probability that $b < 1$ (i.e. that planet c has the correct geometry to transit), $\mathcal{D}_X$ denotes "detected by X". Via Bayes' theorem we can express the probability as

$$\begin{aligned} \mathbb{P}_c &= \Pr(\mathcal{T}|P_c)\Pr(\bar{\mathcal{D}}_{\rm Kep}|\mathcal{T}, P_c)\Pr(\bar{\mathcal{D}}_{\rm HST}|\mathcal{T}, \bar{\mathcal{D}}_{\rm Kep}, P_c), \\ &= \Pr(\mathcal{T}|P_c)\Pr(\bar{\mathcal{D}}_{\rm Kep}|\mathcal{T}, P_c)\Pr(\bar{\mathcal{D}}_{\rm HST}|\mathcal{T}, P_c), \end{aligned} \tag{4.5}$$

where on the second-line we remove the conditional $\bar{\mathcal{D}}_{\rm Kep}$ since there is no causal dependency.

To simplify the analysis, we will assume that any other planets in the system are coplanar with Kepler-1625b, whose low impact parameter essentially guarantees that these planets will be transiting too. Accordingly, we assume $\Pr(\mathcal{T}|P_c) \simeq 1 \; \forall \; P_c \in \{P_{\rm min}, P_{\rm max}\}$, where $P_{\rm min}$ and $P_{\rm max}$ are some yet-to-be-determined minimum/maximum limits on the period of planet c.

This optimistic assumption of coplanarity means that we will tend to overestimate the chance that the moon-like dip is caused by a second planet - which is the conservative option - and reduces the overall complexity of the problem.

4.4.2 Basic properties of a hypothetical K1625c

The depth of the moon-like dip varies between the three different long-term trend models adopted by TK18. In all three cases, the radius is approximately Neptune-like, yielding $4.90^{+0.79}_{-0.72}\,R_\oplus$ for the linear model, $3.09^{+1.71}_{-1.19}\,R_\oplus$ for the quadratic model, and $4.05^{+0.86}_{-1.01}\,R_\oplus$ for the exponential model. The last value is not only the median of the three but also represents the favored model by TK18. For this reason, we will assume here that the hypothetical second transiting planet has a radius of $4\,R_\oplus$ in what follows.

The moon-like dip is approximately flat-bottomed, indicating that if it were due to a transiting planet, the impact parameter is small, that is, the planet must be non-grazing. This means the inferred radius from the depth is a fair estimator of the true radius.

The duration of the moon-like dip varies between the models from 8.5 hours for the linear and quadratic models to 7.8 hours for the exponential (using the $\tilde{T}$ transit duration definition of Kipping 2010). This therefore establishes that the duration of the hypothetical K1625c must exceed 7.8 hours. This is still a relatively long transit duration and implies that the orbital period is not short.

For any given orbital period, the longest possible duration corresponds to a zero impact

parameter. Therefore, for any given duration, the shortest allowed orbital period corresponds to a zero-impact parameter. We can therefore take this duration value and convert it into a minimum period. Assuming a circular orbit, one may solve the Kipping (2010) $\tilde{T}$ duration equation for P in the limit of $b \rightarrow 0$, and also transform $a/R_\star$ into stellar is density, $\rho_\star$, using Kepler's Third Law. Since $\rho_\star$ is well-constrained from *Gaia* and isochrone modeling to be $0.29^{+0.13}_{-0.09}\,\mathrm{g\,cm^{-3}}$ (TK18), we can solve for the minimum period numerically to find $P > 16$ days, as shown in Figure 4.5.

4.4.3 Probability of a missed TCE

$\Pr(\bar{\mathcal{D}}_{\mathrm{Kep}}|\mathcal{T})$ denotes the probability that a Neptune-sized transiting planet was undetected by the *Kepler* pipeline - i.e. a missed threshold crossing event (TCE). There are no detected TCEs for Kepler-1625 aside from Kepler-1625b in DR25 (Thompson et al., 2017), but this fact alone does not provide a probability that one was missed by the *Kepler* pipeline.

The probability of missed TCEs is most directly computed by using the per-target detection contours for DR25 reported by Burke & Catanzarite (2017). The `KeplerPORTs` software, first discussed in Burke et al. (2015), computes detection completeness contours for a given *Kepler* target through transit injection and recovery tests and provides the most realistic estimate of completeness available. The stellar parameters used by Burke & Catanzarite (2017) are the DR25 Mathur et al. (2017) values, for which Kepler-1625 is reported as a $1.79^{+0.26}_{-0.49}\,R_\odot$ - which is approximately the same as the *Gaia*-based value found by TK18 of $1.73^{+0.24}_{-0.22}\,R_\odot$. This therefore demonstrates that the `KeplerPORTs` detection contours for a given planetary size do not require any significant update since the minor revision of TK18.

After running `KeplerPORTs` on our target (see Figure 4.6), we extracted a slice along the radius axis of $4\,R_\oplus$, corresponding to the dip seen in the HST data by TK18. This is shown in the left-most panel of Figure 4.7.

`KeplerPORTs` natively computes completeness only out to 500 days, and indeed by this point the probability of missing a $4\,R_\oplus$ exceeds 99% and is effectively unity - meaning there

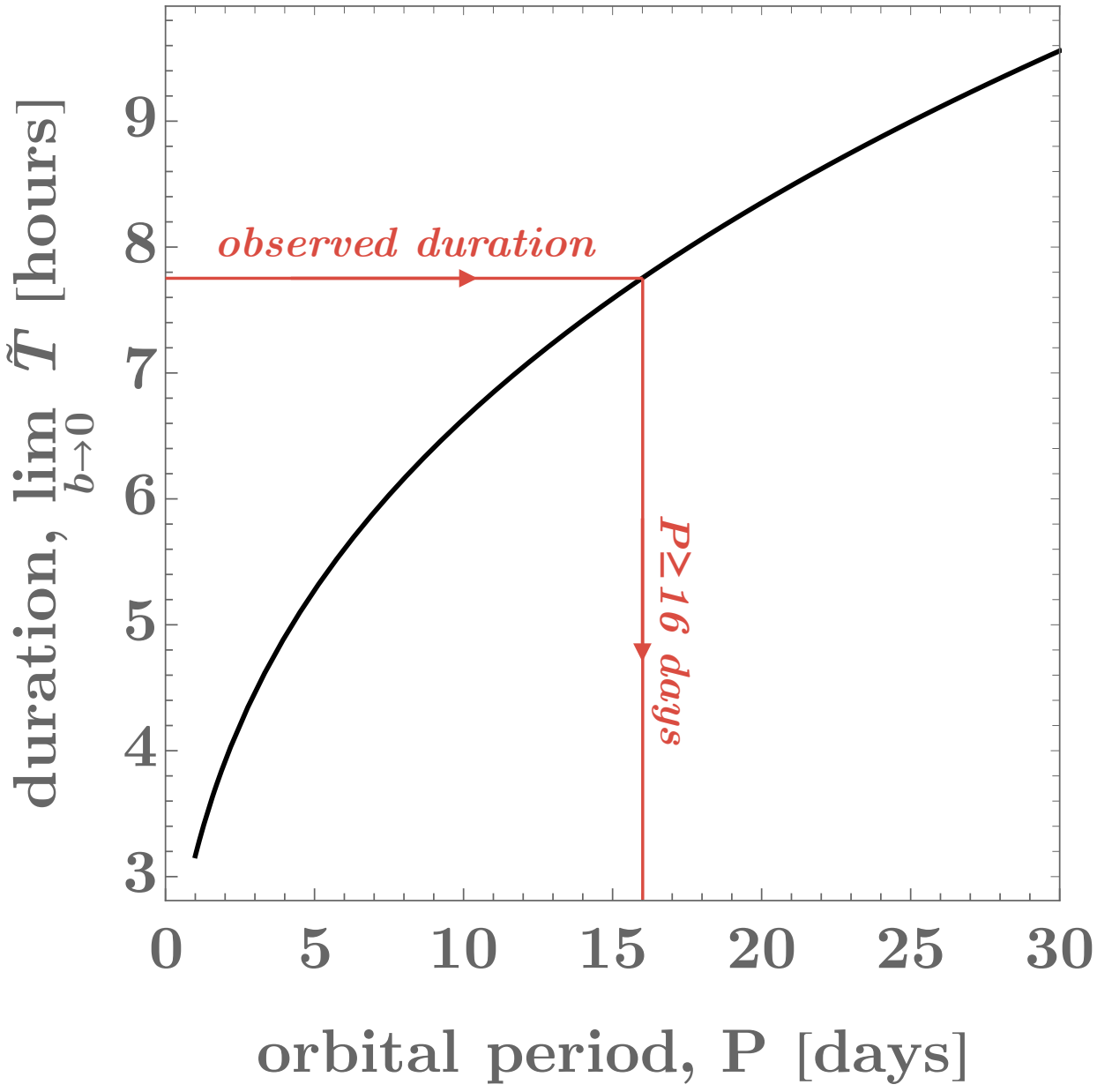

Figure 4.5: The moon-like dip reported by TK18 has a duration of at least 7.8 hours. Plotting the maximum transit duration (for a circular orbit) as a function of period for a planet around Kepler-1625, one can see that the period cannot be smaller than 16 days to explain the dip.

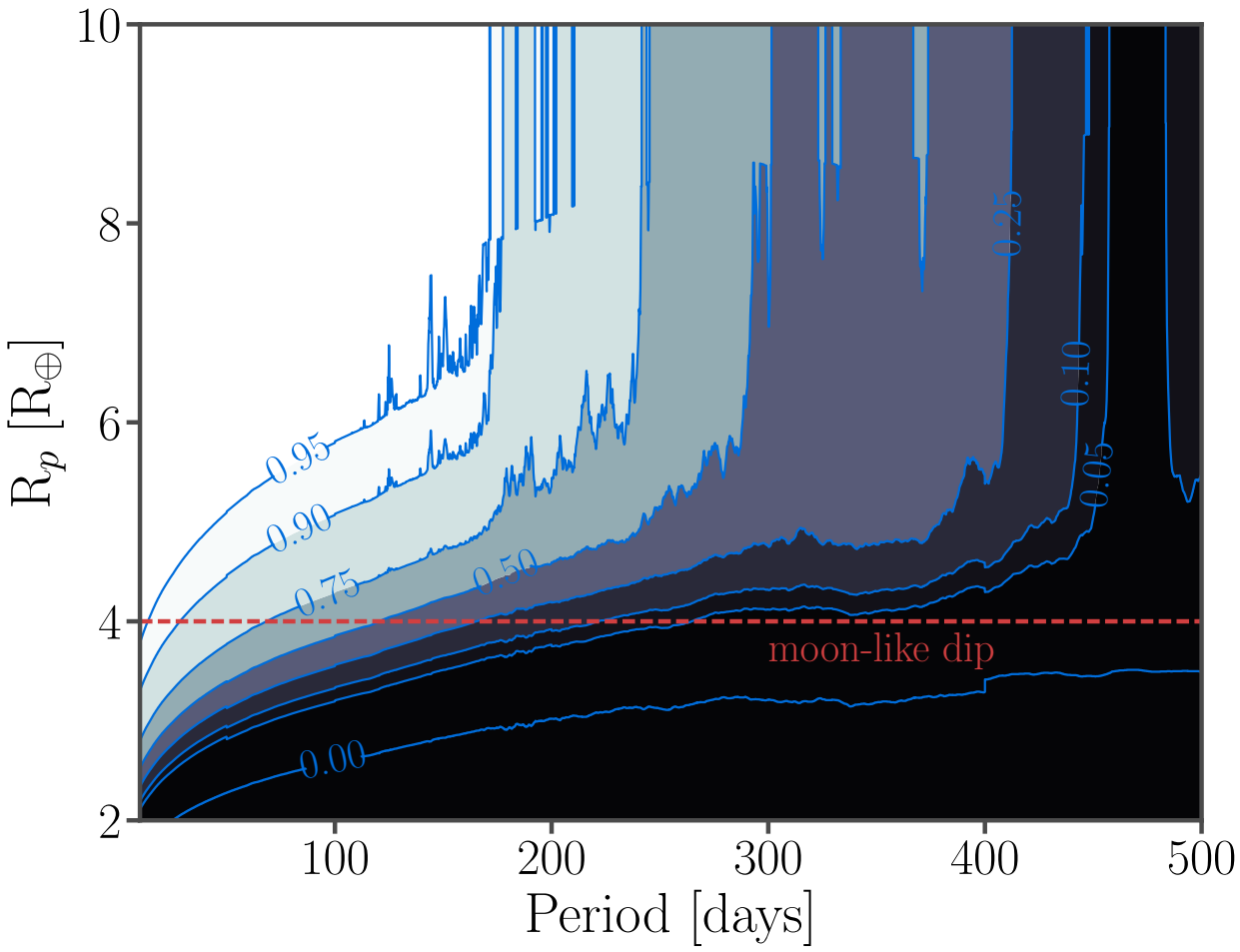

Figure 4.6: Detection completeness contours plot generated by `KeplerPORTs` (Burke et al., 2015; Burke & Catanzarite, 2017) for the target Kepler-1625.

is little point in extending past this period.

4.4.4 Probability of K1625c transiting in the HST window

If the moon-like dip were due to another planet, then within the HST window of $W = 38.8$ hours we would have observed a single transit of our hypothetical planet K1625c. The HST photometry is approximately four times superior to that of *Kepler* and thus HST is effectively complete to a Neptune-sized transit of the observed duration. $\Pr(\mathcal{D}_{\rm HST}|\mathcal{T})$ then simply reduces to the probability that the planet will have the correct phasing to transit within the 38.8 hour observing window.

Consider the possibility that K1625c has an orbital period of 100 years. The chance of seeing this world transit in a fixed window of observations is clearly going to be very low. Indeed, the chance of seeing a planet with period P transit at least once in a window is $\propto 1/P$. This is known as the window effect and is described in detail in Kipping (2018), who shows that

$$\begin{aligned} &\Pr(n = 1|P_c, W, \mathcal{T}) = \\ &\Pr(n \geq 1|P_c, W, \mathcal{T})(1 - \Pr(n \geq 2|P_c, W, \mathcal{T})), \end{aligned} \tag{4.6}$$

where n is the number of transits observed in the window of duration W and the components probabilities are

$$\Pr(n \geq 1|P_c, W, \mathcal{T}) = \begin{cases} 1 & \text{if } P_c \leq W, \\ \frac{W}{P_c} & \text{if } P_c > W, \end{cases} \tag{4.7}$$

and

$$\Pr(n \geq 2|P_c, W, \mathcal{T}) = \begin{cases} 1 & \text{if } P_c \leq \frac{W}{2}, \\ \frac{W-P_c}{P_c} & \text{if } \frac{W}{2} < P_c \leq W, \\ 0 & \text{if } P_c > W. \end{cases} \tag{4.8}$$

Kipping (2018) shows how a lower limit on the period can be derived from the relative phase of the transit within the window but in our case a far more constraining lower limit on the period comes from the duration argument earlier in Section 4.4.2. Imposing this as a hard limit simplifies Equation (4.6) to

$$\Pr(n = 1|P_c, W, \mathcal{T}) = \frac{W}{P}. \tag{4.9}$$

This is shown in the middle panel of Figure 4.7. And finally we may write that $\Pr(\mathcal{D}_{\rm HST}|\mathcal{T}) = \Pr(n = 1|P_c, W, \mathcal{T})$ since we treat HST as effectively complete to Neptune-sized transits.

4.4.5 Combining the constraints

The final step is to combine the probabilities from above using Equation (4.5), which is shown in the right-most panel of Figure 4.7. The probability peaks at $P_c = 133.3$ days with $\mathbb{P}_c = 0.74\%$, and decreases monotonically either side. Given the presence of TTVs in the system (TK18), the most plausible planet-scenario to explain both the dip and the TTVs would be an interior transiting planet close to a mean motion resonance (e.g. 2:1 would lead to $P_c \simeq 144$ days).

The probability computed above suggests the existence of another transiting planet causing the moon-like dip is quite low, which might lend additional credence to the exomoon hypothesis. At the same time, this probability should be weighed against the probability of observing such a large exomoon. This comparison unfortunately eludes us for the time

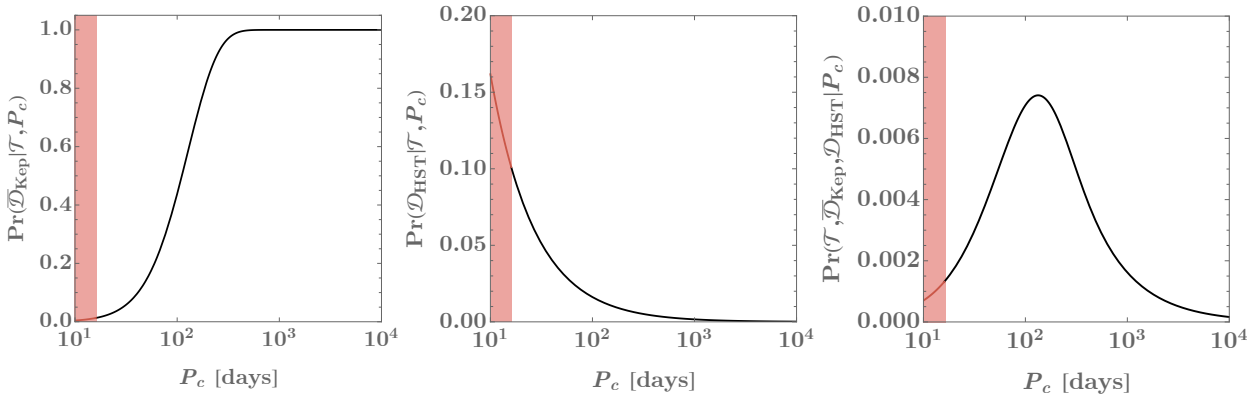

Figure 4.7: Probability that a Neptune-sized transiting planet evaded detection by *Kepler* (left), was seen to transit in the HST window of TK18 (middle), and the probability of both of these statements being true - as a function of the planet's orbital period.

being, as there are so far no other verified exomoons in the literature, and the occurrence rate of such an unanticipated object cannot be quantified at this time.

4.5 Stellar Activity

4.5.1 Rotation

There is no known rotation period for Kepler-1625 at the time of writing. The star is included within the autocorrelation function (ACF) catalog of McQuillan et al. (2014), but no clear rotation period was found in that work.

We attempted to search for the rotation period using a Lomb-Scargle (LS) periodogram, applying the algorithm to each *Kepler* quarter (PDC data) independently. Since each quarter is treated independently, and each quarter has a duration of $\simeq 90$ days, it is not possible to detect periods longer than approximately half this value. The results out to 50 days are therefore shown in Figure 4.8.

Consistent with the analysis of McQuillan et al. (2014), we are unable to identify any clear rotation period from the LS periodogram. We find that the maximum amplitude of a periodic signal < 50 days must be less than 200 ppm, significantly lower than the amplitude

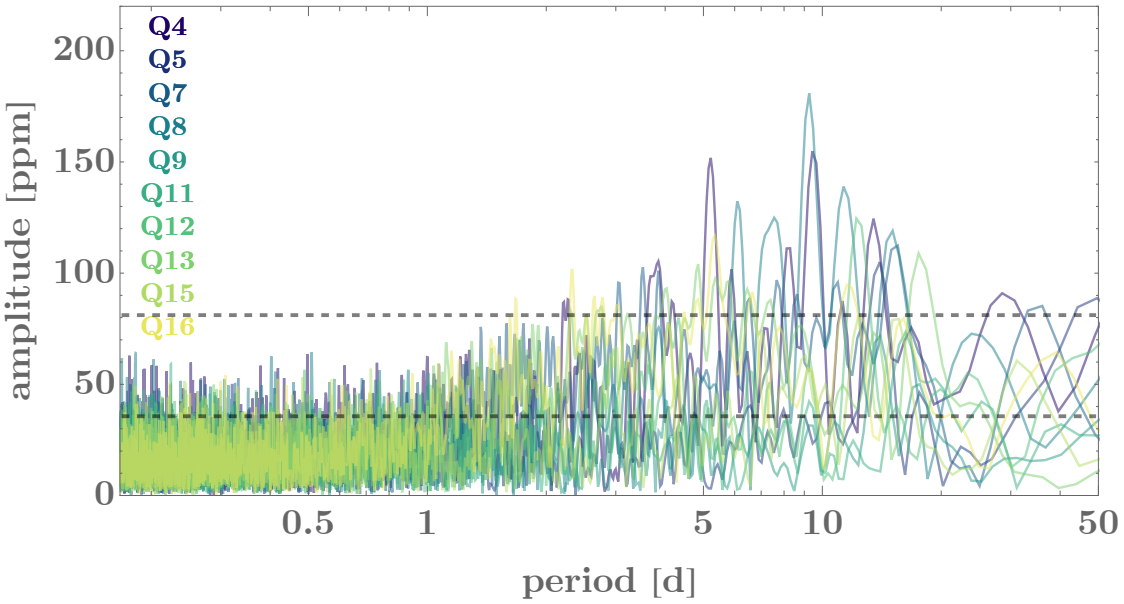

Figure 4.8: LS periodogram for the PDC data of Kepler-1625 for ten quarters. Activity appears lower than 200ppm for periods < 50 days. Dashed lines represent the p-values of 0.05 for the most active and most quiescent quarters based on bootstrapping. Activity above these limits may be considered real, but a clear rotation period across all quarters is not detected.

of the moon-like dip reported by TK18.

We also attempted to recover a rotation period using Gaussian process regression. We used the `celerite` software package (Foreman-Mackey et al., 2017) to model the light curve, the kernel function consisting of a mixture of two simple harmonic oscillators with periods separated by a factor of two [2]. Exploring the posterior PDF of the star's rotation period using `PyMC3` (Salvatier et al., 2016) we infer a rotation period of $12.9^{+0.7}_{-0.6}$ days. However, for this period the natural log of the Q factor, or damping ratio, was -3.1 ± 0.3. This means that the light curve, when modeled as a damped harmonic oscillator, is overdamped ($Q \sim 0.05$), indicating that the stellar brightness variations are incoherent, which suggests that the star spot lifetimes are shorter than the rotation period of the star on average. This, combined with the inability of both the LS periodogram and the ACF to recover a reliable period, implies that the signal is aperiodic and non-sinusoidal. Taken together with the low RMS of the light curve (~ 200 ppm), this indicates that Kepler-1625 is an inactive star and that there is little evidence of short-timescale (sub-hour) variability that could mimic the ingress of a moon.

Although not a direct measure of the rotation period, the $v \sin i$ can provide some useful information on rotation too. We obtained two Keck *HIRES* spectra without iodine in October and November 2018 to attempt to measure the velocity broadening. Using the `SpecMatch` pipeline described in Petigura et al. (2017), we obtain a marginal detection of $v \sin i = (1.9 \pm 1.0)$ km/s. Combining this with the isochrone posteriors from TK18 yields a minimum rotation period of 45^{+44}_{-15} days (a minimum because we don't know $\sin i$). It is therefore probable that the rotation period falls within the 50 day range that *Kepler* is sensitive to, but that the amplitude of rotational modulations is simply too small to reliably recover.

[2] see https://exoplanet.readthedocs.io/en/latest/tutorials/stellar-variability/

4.5.2 Activity-induced dips

Stellar activity can produce complex morphological signatures in photometric time series (Dumusque et al., 2014). Although the photometric periodic behavior of Kepler-1625 appears limited to < 200 ppm (see Section 4.5.1) – too small to induce an effect comparable to the moon-like transit reported by TK18 – shorter, non-periodic variations deserve our attention.

The moon-like dip is characterized by a transit depth of $\simeq$500 ppm in the integrated light (white) bandpass of WFC3 (TK18). As a near-infrared instrument, stellar activity is generally expected to be suppressed by WFC3 versus an optical bandpass like that of *Kepler*. To estimate the magnitude of this effect, we took the isochrone posterior chains for Kepler-1625 (TK18) and extracted the median effective temperature of the star, $T_{\rm eff} = 5563$ K. We then assume spots on the surface with a temperature approximately 2000 K cooler than the photosphere, typical of sunspots. Integrating a Planck function multiplied by the bandpass response function for *Kepler* and WFC3 reveals that spots would appear 1.3 times larger in amplitude, as viewed by *Kepler* than WFC3. Accordingly, if the moon-like dip is due to spots, then one should expect to see frequent dips of amplitudes of $\simeq 650$ ppm (500×1.3) in the eleven quarters of *Kepler* data for Kepler-1625.

To test this hypothesis, we extracted a random segment from a random quarter of the available *Kepler* PDC time series, with the segment duration being equal to 26 HST orbits (which was the time window observed by TK18). Each quarter was first detrended using a median filter of window size equal to 5 times the minimum duration of the moon-like dip, approximately 7.8 hours (TK18). We then performed a blind search for the best-fitting box-like transit within this segment, forcing the box to have a duration equal to that of the TK18 moon-like dip. The central time and depth were optimized for in a least squares sense. An example of this is shown in lower panel of Figure 4.9.

The best-fitting box was saved and then a new random segment was picked, repeating 10^5 times. A histogram of the best-fitting depths is shown in the upper panel of Figure 4.9. Because this is the best-fitting depth within a segment, these depths always deviate from

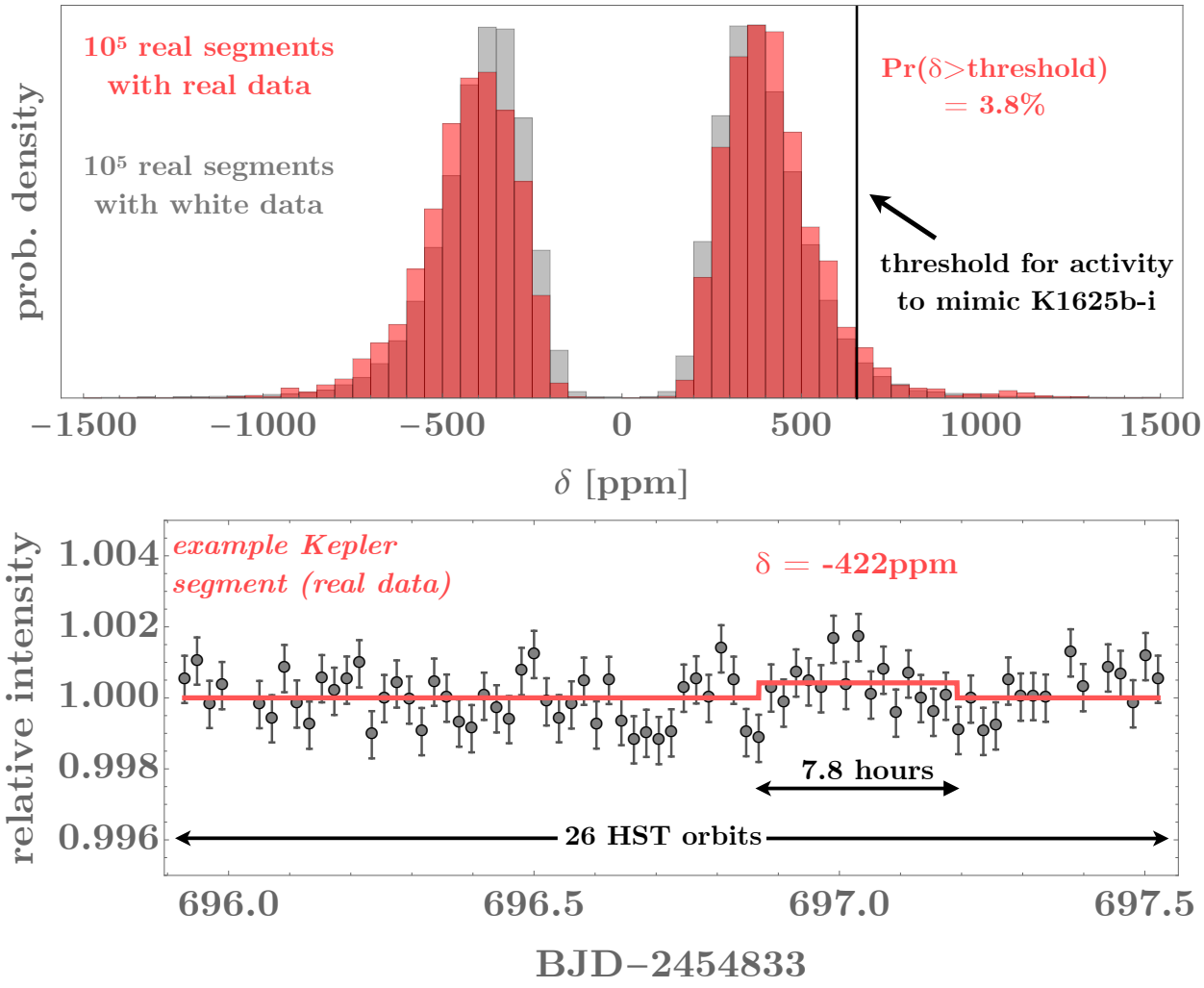

Figure 4.9: Bottom: Example of a random *Kepler* segment of Kepler-1625 with a duration equal to that of the TK18 HST observing window. Regressing the best fitting box with a duration of 7.8 hours (same as the moon-like dip) finds an inverted transit in this case of 422 ppm depth. Top: Repeating this exercise on 10^5 random segments, we obtain a nearly symmetric distribution of best-fitting box amplitudes (red histogram). For comparison, we repeated the simulations assuming pure Gaussian noise only (gray histogram), which is nearly identical.

zero since the regression routine is allowed to try many different possible central times. There is symmetry about zero, with just as many inverted transits as positive transits being recovered. We find that 3.8% of the experiments run on the *Kepler* data are able to produce a best-fitting transit of depth exceeding 650 ppm (and 3.5% produce depths < -650 ppm). Naively, one might interpret this as indicating that the moon-like dip reported by TK18 is only 2.1 σ significant (3.8%). However, these simulations were conducted for the 0.95 m *Kepler* telescope data, and not the 2.4 m HST data set in which the dip is actually observed. To interpret this 3.8% number, one must consider the plausible origin of these spurious (i.e. moon-mimicking) events.

If indeed the signals are spurious, there are two possible causes for these random quasi-dips. Either 1) time-correlated noise structure caused by intrinsic stellar activity is able to produce > 650 ppm dips, or 2) the noise is not significantly correlated (i.e. white noise) but the noise budget of the *Kepler* photometry is sufficiently large that the best-fitting boxes can infrequently exceed 650 ppm.

If the former were true, then the dip observed by HST could be explained as one of these 3.8% instances of an activity-driven false-positive. If the latter were true, then one could expect it to be highly improbable for the HST moon-like dip to be a product of Gaussian noise, as the measurement uncertainties are 3.8 times smaller than that of *Kepler*.

Clearly this is an important distinction. To distinguish between them, we can set up another experiment where we repeat our previously described Monte Carlo experiments except we replace the real *Kepler* photometric fluxes with artificial fluxes computed assuming pure Gaussian (white) noise. The artificial data are drawn from a normal distribution with a mean of unity and standard deviation equal to the standard deviation of a randomly picked real *Kepler* segment.

After drawing 10^5 segments and replacing the photometric fluxes with white noise, we make a histogram of the best-fitting box depths as before and find a very similar distribution, as shown in Figure 4.9. The 650 ppm threshold is exceeded in a similar number of trials,

2.4%. We interpret the similarity between these two distributions as evidence favoring the hypothesis that the spurious, moon-mimicking detections are simply a product of Gaussian-like noise controlled by photon counting statistics, rather than being due to intrinsic stellar activity. Since the HST data is much more precise, the probability of a white noise driven box is far smaller, and is in fact accurately accounted for in our evidence calculations since we assumed a normal likelihood function in TK18. We therefore conclude that there is no evidence from the *Kepler* analysis that activity is a plausible explanation for the moon-like reported by TK18 in the HST data.

4.6 Follow-Up

4.6.1 Photometric follow-up

The best way to confirm the presence of the exomoon candidate would be to see it transit again. To this end we have explored various avenues to observe future transits of Kepler-1625b. Unfortunately, this is a very challenging target for transit observations because of its faintness ($K_P = 15.756$) and the very long duration of the planet's transit (~ 19 hours). These challenges are exacerbated by the fact that the exact location of the exomoon cannot be known ahead of time for any given transit; as we project into the future, our predictions are naturally degraded as the uncertainties in our posterior samples propagate. A wide range of times before planetary ingress and after planetary egress must therefore be monitored to cover as many geometries as possible.

These limitations generally restrict any efforts to detect the exomoon transit to space-based telescopes. However, targeted observations of this sort clearly require considerable dedicated resources to the exclusion of other priorities. While *Spitzer* may be a suitable alternative to observing with HST, the former can only observe $\sim 35\%$ of the sky at any given time due to pointing restrictions, and cannot observe the May 2019 transit as the target falls within the zone of avoidance. Future observations carried out with a survey (non-targeted) spacecraft could potentially bear fruit, though we note that the Transiting

Exoplanet Survey Satellite (TESS) observation of the *Kepler* field will occur in July 2019, missing the May 2019 transit of Kepler-1625b.

On the other hand, transit timings could potentially be measured from the ground more easily, and continued monitoring of the TTVs and (to the extent possible) transit duration variations (TDVs) would be valuable. A single instrument may be able to monitor in its entirety either planetary ingress or egress, but likely not both, due to the time separation of these two events. Of course, this requires the target to be up at night long enough for the observation to be made, and the telescope must be located at a longitude where the event can be observed in its entirety without sunset or sunrise encroaching. Latitude is also a consideration; while northern latitudes place the target above the horizon for longer durations, they also experience a greater range of night lengths.

Radial velocity (RV) measurements of the system may also provide additional evidence for or against the moon hypothesis. On the one hand, RVs could potentially yield evidence for a second planet in the system, in which case the observed TTVs might be attributable to that planet and the case for the moon would be weakened. On the other hand, if an additional massive planet can be ruled out, or strong constraints can be placed on the mass and location of an undetected planet, the moon could emerge as a stronger candidate insomuch as an alternative mechanism for the timing variations is weakened or removed altogether.

Of course RV measurements should also provide a reliable measurement of the target planet's mass. If Kepler-1625b is revealed to be significantly less massive than anticipated, this could also weaken the moon case, as it would be more difficult to support such a large moon. Conversely, a mass measurement consistent with the inferred mass presented in TK18 could lend additional credence to the moon hypothesis. Figure 4.10 presents our best estimate for the planet's mass which may be tested by the acquisition of RVs.

An additional complication for photometric confirmation through transit observations arises from the exomoon candidate's inferred large inclination with respect to the planet's

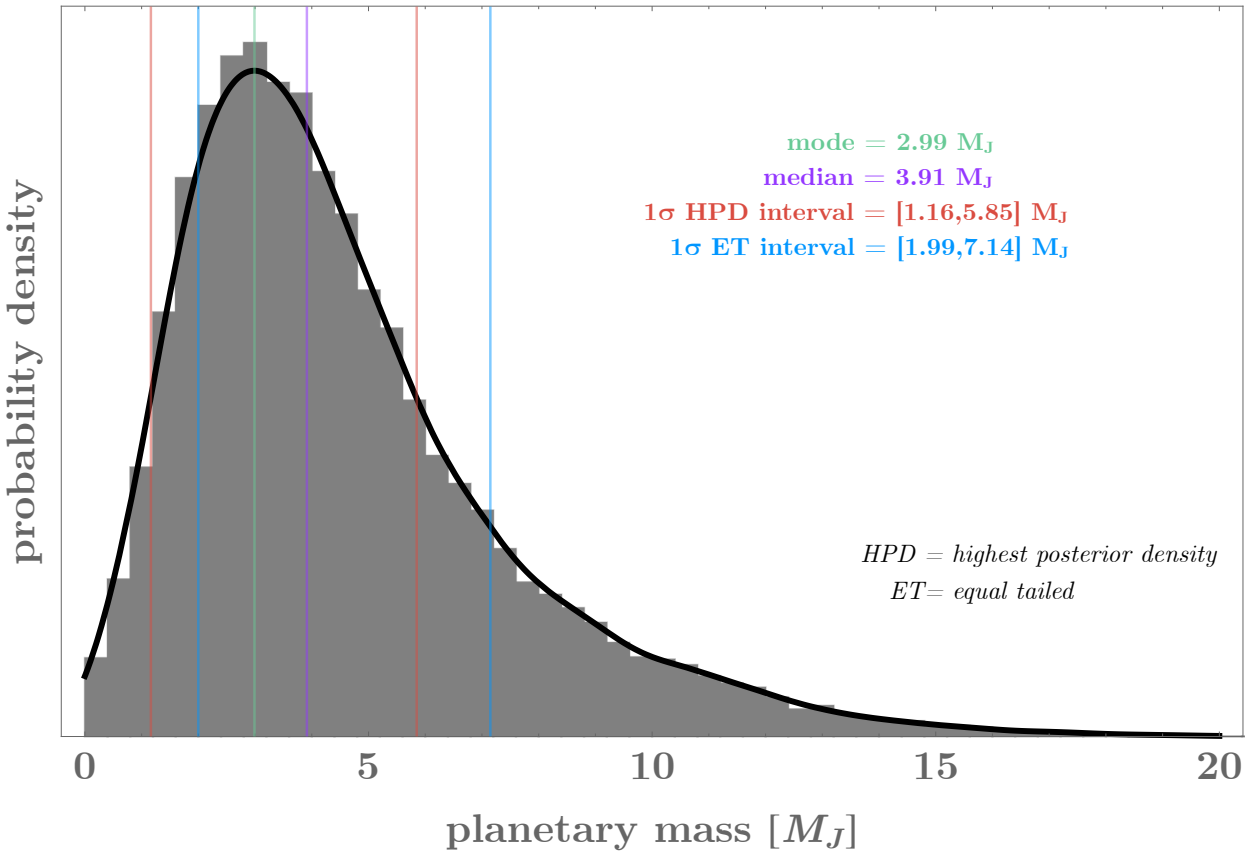

Figure 4.10: Combined mass posterior distribution for Kepler-1625b.

orbital plane (TK18 found $i_S = 42^{+15}_{-18}$, 49^{+21}_{-22}, and 43^{+15}_{-19} degrees for the linear, quadratic, and exponential detrendings, respectively). This has the effect of sending the moon high above or below the disk of the star for a significant fraction of its orbit, precluding the possibility of a transit when the moon is in these positions. Coupled with the uncertainty in the moon's true anomaly, this means that for any given transit observation there is no guarantee of seeing the moon transit at all. Thus, a null detection of the moon for any given epoch cannot be interpreted as definitive evidence that the moon does not exist. Only with many repeated observations, all lacking evidence for the moon, could the moon truly be ruled out to high confidence (Martin et al., 2019). Clearly this has a multiplicative impact on the telescope requirements, and naturally leads to the conclusion that follow-up transit observations are only worthwhile to the extent that they are not excessively expensive.

Using the posterior samples from Teachey, & Kipping (2018), it is possible to predict the morphology of the combined planet and moon system for future transit events, though as noted these predictions naturally deteriorate with every epoch. We elected to consider ten epochs, including the original observed epoch for comparison, and calculate 1000 projections of the transit light curve for Kepler-1625b. For this purpose, we used model $\mathcal{M}$ and repeated for each of the three trend models used by Teachey, & Kipping (2018). The light curves can be viewed in Figure 4.11.

For each epoch, we find the time of transit minimum for the planet component only and use these times to compute a median mid-transit time and an associated standard deviation, which is quoted in the panels of Figure 4.11. We also consider the moon component in isolation and count how often the moon presents any deviation in flux away from unity - the probability that the moon will transit at all in each epoch. These probabilities are again added to the panels of Figure 4.11. Finally, we use the moon component only to compute a probability distribution for the most likely location one should expect to observe the exomoon (assuming it transits at all). These curves are shown in gray in Figure 4.11.

As expected, we find that the uncertainty in the time of transit grows as we project further

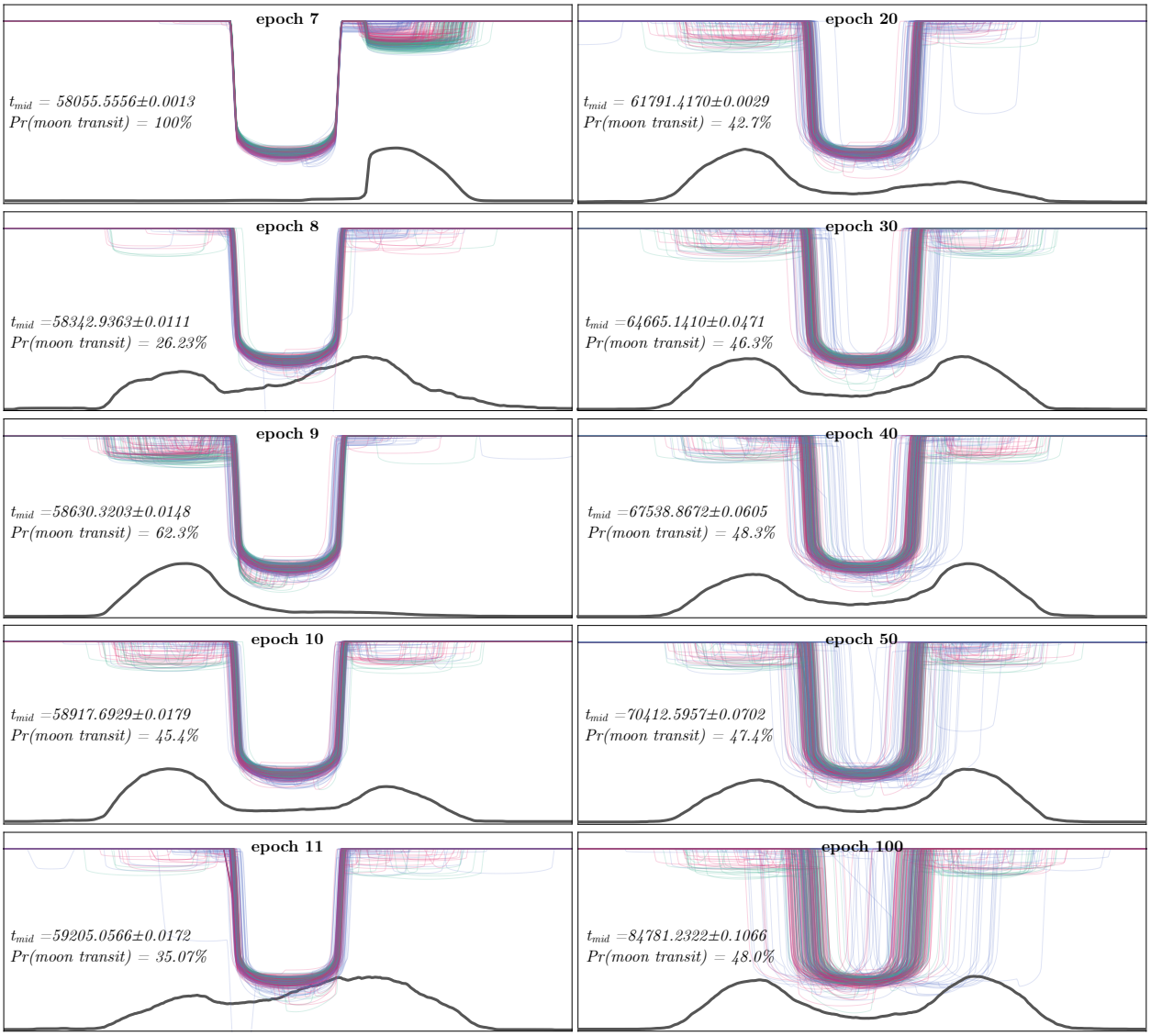

Figure 4.11: Projections of the light curve of Kepler-1625b and its candidate moon into the future, using the posterior samples of TK18. We show 100 light curves from each trend model (green = linear, blue = quadratic and pink = exponential). Uncertainties can be seen to visually grow over time. The gray distributions reflect the most likely location for a moon transit.

into the future. Regressing a power-law to the uncertainties, we find that uncertainty grows as $t^{2/13}$ to an excellent approximation. The probability of a moon transiting oscillates for the first few epochs but then tends towards slightly less than 50%, which is broadly consistent with the findings of Martin et al. (2019). Observers can therefore treat the chances of seeing the moon in a given future transit as approximately 1 in 2.

We also highlight that epoch 9 (May 2019) appears to be the most favorable for follow-up. It has the highest probability of seeing a moon transit out of any future epoch and a clean prediction for the location of said transit (before planetary ingress). A proposed observation of this transit with HST was not awarded, and as pointed out, there are no other viable space-based options for this event.

In light of these challenges, photometric confirmation of the exomoon candidate may remain elusive for some time, until repeated observations may be performed at relatively low cost. Of course, if the moon is real, eventual confirmation is probably inevitable, but in the near term it will likely remain merely a candidate.

4.7 Conclusions

In this work we have examined a number of alternative hypotheses put forth by the community to explain the two critical pieces of the exomoon case for Kepler-1625b, namely, the presence of significant TTVs, and a sustained flux reduction in the HST light curve following planetary egress. We have explored various additional detrending models, employing more degrees of freedom, and found that while some of these approaches are able to attenuate the purported moon signal, this is to be expected given their flexibility. That is, from the standpoint of their Bayesian evidences, more flexible detrending models paired with planet-only transit models are in some cases indistinguishable from simpler detrending models combined with system models that include a moon. While we cannot rule out the presence of unprecedented systematic effects, we also see no evidence for them, and therefore the adoption of more flexible detrending models that attenuate the moon signal are not particularly well

motivated.

We have investigated the differences between the light curve presented in TK18 and a new reduction from Kreidberg et al. (2019), and find that while the source of the discrepancy is not readily identifiable, our light curve displays effectively identical noise properties, and therefore, the KLB19 light curve is not demonstrably superior. We also highlight once again the work of Heller et al. (2019) which, through their own independent reduction and analysis, also found evidence of a moon-like dip following planetary egress.

In terms of a possible additional transiting planet in the system, we have calculated the probability that such a planet could have gone undetected in the *Kepler* data and transit in the short time window of the HST observation and find the maximum probability of this scenario to be $< 0.75\%$.

To determine whether the dip in brightness measured with HST could be due to stellar activity, we have attempted to measure a rotation period for the star in the *Kepler* data using a variety of standard methods and are unable to recover it, indicating that the star exhibits negligible periodic variability. We have also searched for photometric dips that might be associated with (non-periodic) star spot crossings and find that such dips, while possible to find in the *Kepler* data, are consistent with Gaussian noise.

Finally, we discussed the outlook for confirming the presence of the exomoon using space-based transit monitoring, radial velocity observations, and ground-based measurement of transit timings. We find that the system poses a number of substantial challenges to observational confirmation in the near-term, and conclude that while modest ground-based observations may be worthwhile for 1) constraining the mass of the planet, 2) quantifying the probability of an unseen perturber in the system, and 3) measuring TTVs, additional targeted observations from space likely fail a reasonable cost-benefit analysis. Confirming or refuting the moon to high confidence may therefore require many years and the advent of additional space-based time-domain survey data that can be acquired at minimal cost.

4.8 Acknowledgements

We would like to thank Laura Kreidberg for useful discussions and providing data products in advance of her paper, which allowed us to investigate in greater detail the source and degree of the discrepancies between her findings and ours. We also thank Erik Petigura for performing the $v \sin i$ measurement. We thank the anonymous reviewer for thorough comments which strengthened this paper. Finally, we wish to thank past and present HST and *Kepler* scientists and engineers, mission support personnel, and the crews of STS-31, 61, 82, 103, 109, and 125, who through their dedication have been jointly responsible for making this work possible.

Analysis was carried out in part on the NASA Supercomputer PLEIADES (Grant #HEC-SMD-17-1386). AT is supported through the NSF Graduate Research Fellowship (DGE-1644869). DK is supported by the Alfed P. Sloan Foundation Fellowship. This work is based in part on observations made with the NASA/ESA Hubble Space Telescope, obtained at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-26555. These observations are associated with program #GO-15149. Support for program #GO-15149 was provided by NASA through a grant from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-26555." This paper includes data collected by the *Kepler* mission. Funding for the *Kepler* mission is provided by the NASA Science Mission directorate.

Chapter 4: Bibliography

Andrae, R., Schulze-Hartung, T., & Melchior, P. 2010, arXiv e-prints, arXiv:1012.3754

Burke, C. J., Christiansen, J. L., Mullally, F., et al., 2015, ApJ, 809, 8

Burke, C. J. & Catanzarite, J. 2017, "Planet Detection Metrics: Per-Target Detection Contours for Data Release 25", KSCI-19111-001

Dumusque, X., Boisse, I., Santos, N. C., 2014, ApJ, 796, 132

Feroz, F. & Hobson, M. P., 2008, MNRAS, 384, 449

Evans, T. M., Sing, D. K., Goyal, J. M., et al. 2018, AJ, 156, 283

Feroz, F., Hobson, M. P. & Bridges, M., 2009, MNRAS, 398, 1601

Foreman-Mackey, D., Hogg, D. W., Lang, D., Goodman, J., 2013, PASP, 125, 306

Foreman-Mackey, D., Agol, E., Ambikasaran, S., et al. 2017, AJ, 154, 220

Gelman, R. & Rubin, D., 1995, Sociological Methodology, 25, 165

Heller, R., Rodenbeck, K., & Bruno, G. 2019, A&A, 624, A95.

Huitson, C. M., Sing, D. K., Pont, F., et al., 2013, MNRAS, 434, 3252

Kipping, D. M., 2010, MNRAS, 407, 301

Kipping, D. M., 2011, MNRAS, 416, 689

Kipping, D. M., Bakos, G. Á., Buchhave, L., Nesvorný, D., Schmitt, A., 2012, ApJ, 750, 115

Kipping, D. M., 2018, RNAAS, 2, 223

Knutson, H. A., Dragomir, D., Kreidberg, L., et al., 2014, ApJ, 794, 155

Kreidberg, L., Luger, R., & Bedell, M. 2019, ApJ, 877, L15

Martin, D., Fabrycky, D., Montet, B., 2019, ApJ Letters, submitted

Mathur, S., Huber, D., Batalha, N. M., et al., 2017, ApJS, 229, 30

McQuillan, A., Mazeh, T., Aigrain, S., 2014, ApJS, 211, 24

Nelson, B. E., Ford, E. B., Buchner, J., et al. 2020, AJ, 159, 73

Petigura, E. A., Howard, A. W., Marcy, G., et al., 2017, AJ, 154, 107

Ranjan, S., Charbonneau, D., Désert, J.-M., Madhusudhan, N., Deming, D., Wilkins, A., Mandell, A. M., 2014, ApJ, 785, 148

Salvatier J., Wiecki T.V., Fonnesbeck C. (2016) Probabilistic programming in Python using PyMC3. PeerJ Computer Science 2:e55

Schwarz, G., 1978, The Annals of Statistics, 6, 461

Stevenson, K. B., Bean, J. L., Seifahrt, A., Désert, J.-M., Madhusudhan, N., Bergmann, M., Kreidberg, L., Homeier, D., 2014a, AJ, 147, 161.

Stevenson, K. B., Bean, J. L., Fabrycky, D., Kreidberg, L., 2014b, ApJ, 796, 32.

Teachey, A. & Kipping, D. M., 2018, Science Advances, 4, 1784 (TK18)

Thompson, S. E., Coughlin, J. L., Hoffman, K., et al., 2017, ApJS, 235, 38

Weaklieam, D. L., 1999, Sociological Methods & Research, 27, 359

Wakeford, H. R., Sing, D. K., Evans, T., et al. 2016, ApJ, 819, 10

Chapter 5: Identifying Candidate Exomoon Signals with Convolutional Neural Networks

"Logic, logic, logic. Logic is the beginning of wisdom, Valeris, not the end."
– Spock

5.1 Introduction

The identification of candidate exomoon signals in time-domain photometry remains a significant challenge. With such an enormous data volume to comb, and rigorous Bayesian model selection so computationally expensive (e.g. Kipping et al., 2015), how does one efficiently identify potentially interesting signals? Increasingly, the astronomical community has turned to the use of machine learning for approaching such problems: indeed, mentions of the term in refereed publication abstracts has grown exponentially in the last decade, from about 200 articles in 2010 to more than 2600 in 2019 – a growth rate far exceeding that of the yearly literature increase. In some applications, machine learning may be able to identify patterns that human beings cannot, particularly in recognizing patterns in high dimensional spaces. In other contexts, machine learning may be only as good as what the human eye can discern, but can nevertheless be applied to vast amounts of data that would require an unreasonable amount of human hours to examine.

Recent work (e.g Shallue & Vanderburg, 2018; Pearson et al., 2018; Osborn et al., 2020) with the *Kepler*, K2, and TESS datasets have shown that convolutional neural networks (CNNs) can be effective in distinguishing genuine exoplanet transits from false positives. These efforts employed parallel networks trained on both a "local" and a "global" view of the light curve, and in some cases also incorporated additional observables as inputs, achieving up to 96% accuracy and confirming a number of new planet candidates.

In these works, the networks were trained on a sample of previously confirmed systems. This is certainly a reasonable approach, since it incorporates the great diversity of systems so far discovered and avoids the requirement of producing simulated systems. At the same time, there are some downsides to this methodology, namely, that it 1) runs the risk of inadvertently teaching the networks about (and therefore encoding) the inherent biases of previous detection / confirmation methods, 2) limits the training sample to those systems that have already been detected, and 3) prevents analysis based on ground-truth values.

In this work, we apply an ensemble of CNNs to the problem of identifying candidate exomoon signals in single transit events. We ask the CNNs to make a simple binary classification: For a given light curve segment, is there a moon-like transit present, or not? Because we do not yet have an abundance of exomoon detections in the literature, we must train the the CNNs using artificial light curves, being careful to make sure these light curves are as authentic as possible and represent the full range of possible system architectures and geometries.

In Section 2 we describe our methodology for construction of the training set and the CNN architecture itself. In Section 3 we examine the results of the CNN classifications and highlight several targets that were deemed worthy of additional vetting. In Section 4 we discuss the performance of the CNN and the ongoing challenges for this work, and we conclude in Section 5.

5.2 Network Construction, Training and Testing

The goal of this work was to produce a fast and flexible methodology for identifying systems worthy of closer examination in the search for exomoons. To that end, we elected to utilize CNNs, which have been quite successful in searching for exoplanet transits and distinguishing them from false positives (e.g. Shallue & Vanderburg, 2018; Pearson et al., 2018; Zucker & Giryes, 2018; Dattilo et al., 2019; Yu et al., 2019; Osborn et al., 2020).

For maximum flexibility, we wanted the CNN to take as its input a light curve segment

containing a single planetary transit. This approach necessarily sacrifices encoded transit timing and duration variation information, which is typically an important component of identifying candidate exomoons (and indeed, plays a key role in our subsequent vetting). However, there are several motivations for opting for a single-transit approach.

First, the CNN simply learns to identify moon transits based on transit morphology alone and therefore simplifies the problem considerably, since the classifier need not know anything about other transits in the light curve, nor do timings need to be known precisely ahead of time. Second, training on single events provides maximum flexibility for the classifier, equally capable of identifying moons in a planet with 3 transits as one with 30. Third, it is trivial to apply a single transit classifier to any number of planet transits in a given system, and if the CNN ensemble determines that multiple transits of the same planet contain moon-like signals, this can be an especially powerful indicator that the system deserves more scrutiny. Fourth, the advent of survey missions such as K2 and TESS, with more limited temporal baselines, means that many planets discovered in the future will have only a single transit recorded, or perhaps only a few. Finally, recent work (Hippke, 2015; Teachey et al., 2018) appears to corroborate theoretical studies (Namouni, 2010; Spalding et al., 2016) suggesting that moons are more likely to be found at a greater distances from their host stars. These planets at larger semimajor axes will therefore transit much more rarely and are likely to have one or only a few transits in existing datasets.

5.2.1 Training and Validation Sets

The first step for producing training and validation datasets is to generate noise-free light curves. Because we currently lack a body of exomoon detections in the literature, these light curves must be generated artificially, but should be as physically plausible as possible and represent the full range of possible system architectures and geometries. We are ultimately pursuing a binary classification – "moon" or "no moon" – so we produce an equal number of light curves with and without moons.

The light curves were made using the `LUNA` code (Kipping, 2011), which produces a photodynamical model of the planet-moon system through nested-two body integration and modeling of 3-body syzygies. Each model consisted of a planet or planet-moon system with the planet's transit midtime centered halfway through the time series. Each light curve length is fixed at 10 days, or 5 days on each side of the planet's transit midtime, since the CNN requires inputs of uniform size. We considered setting the window size to that of the Hill sphere (the maximum distance at which a moon could be found) but opted against this approach because 1) reliable masses are only available for a fraction of transiting planets, so reliable Hill sphere estimates are frequently unavailable, and 2) the light curves would have to be binned, therefore making the time sampling heterogeneous across the training sample. Meanwhile, the 10-day time window provides ample coverage to observe the entire Hill sphere for even very long period planets.

To produce a simulated light curve sample that matches closely the observed population of systems, we turned to the *Kepler* sample itself for inputs. For each model light curve, a known planet-hosting star was selected at random from the list of Kepler Objects of Interest (KOIs) on the NASA Exoplanet Archive. The relevant estimates for that star (mass, radius, density, surface gravity, effective temperature, metallicity, as well as the orbital period of the KOI) were extracted, and the appropriate limb darkening coefficients for that star were selected from a table of *Kepler* values calculated by Sing (2010). A simulated planet or planet+moon system was then generated, using these stellar properties as a frame of reference where appropriate. For example, the transit depth was dictated by the ratio of the planet size (randomly generated) to that of the star (drawn from the real distribution of *Kepler* stars). However, we do not draw the planet *sizes* from these same KOIs, as that would severely limit the number of possible systems that could be simulated. The consequence of this is that we may in some cases be generating systems that are physically allowed but rare in nature, for example, low-mass stars with very massive planets. Still, it is worthwhile to generate training and validation sets that encompass the full range of possibilities, including

systems we may be unlikely to find (or have not yet found) in nature.

There are a total of 14 inputs for the ratio of radii $p = R_P/R_*$, the stellar density ρ_*, the planet's impact parameter b, its orbital period P_P, two limb darkening coefficients q_1 and q_2 (see Kipping, 2013), the planet's density ρ_P, the semimajor axis of the moon a_S, a reference phase for the moon ϕ_S, the moon's inclination i_S, longitude of the moon's ascending node Ω_S, the ratio of radii for the moon and planet R_S/R_P, and the ratio of masses M_S/M_P.

As previously noted, stellar attributes were drawn from the real KOIs. Planet sizes were selected randomly from a uniform distribution between 0.5 and 16 $R_\oplus$. With this in hand, the planet's mass is produced using the empirical mass-radius code `forecaster` (Chen & Kipping, 2017), after which the planet's density can be computed. The planet's period was chosen randomly from a log-uniform distribution between 10 and 1500 days; shorter period planets would not be considered due to self-contamination in the light curve. The impact parameter was also drawn from a uniform distribution.

The semimajor axis of the moon was selected randomly from a uniform distribution but was required to be somewhere between the Roche limit and $0.4895\,R_{Hill}$ for prograde moons and $0.9309\,R_{Hill}$ for retrograde moons (Domingos et al., 2006), based on the computed Hill sphere for the planet's mass and semimajor axis. The orbital parameters for the injected moon were all drawn from a uniform distribution, as were the impact parameter and orbital period of the planet.

Initially moons were generated by selecting a mass ratio (from 0.01% up to 1% the mass of the planet) and then producing a realistic moon radius `forecaster`. However, it became clear that this resulted in a large fraction of very massive moons which would have in practice never avoided detection in the first place (that is, we would not need the CNN to identify them). From that point on we generated moons by specifying a size first, with a radius between that of Europa and the Earth. Of course, smaller moons are likely lurking in the data, but if we have no hope of identifying them through Bayesian model selection, there is little point in training to detect them. For the sake of producing the light curves and

calculating a moon transit SNR, we set the mass ratio to zero. This ensures that the moon does not disturb the planet's transit time. The moon's orbit is circular in all cases ($e_S = 0$).

To generate realistic light curves, we randomly selected *Kepler* light curves from stars for which no planets have been reported, into which noise-free light curve models would be injected. From these light curves a 10 day segment was chosen at random, our only requirement being that it lacked large data gaps, and into this light curve a simulated planet or planet+moon light curve could be inserted. Of course, the fact that a planet has not been reported for the donor light curve does not necessarily mean there are no planets in the system, and particularly small planets that have gone undetected could certainly mimic a moon signal if it appears near an injected planet. Identification of yet-undiscovered planets in these light curves is clearly beyond the scope of this work, and so we proceeded with the expectation that a small minority of our light curves designated as "planet only" could in fact have additional signals embedded in them that could potentially confuse the CNN.

We also point out that the spectral type of the *Kepler* donor light curve selected for injection was not required to match the spectral type of the star used to produce the light curve model, and therefore the ratio of radii p of the injected model could be unphysical for the true size of the "donor" *Kepler* light curve. For example, if the donor *Kepler* light curve belonged to a particularly massive star, the transit depth of the injected model might suggest the transiter is a low-mass star rather than a planet. To the extent that photometric variability is correlated with spectral type, this could produce a light curve that is once again unlikely to be found in nature. The limb darkening coefficients, likewise, may not match the spectral type of the donor light curve. In any case, since we detrend the light curves prior to training, the astrophysical signatures of the donor star should in theory be effectively removed and therefore any discrepancy between the selected star and the modeled system should have also been removed. Furthermore, we do not train the CNN by providing any system parameters, only the light curve itself, so the CNN has no knowledge about the star apart from the noise profile that it sees, the transit depths, and sometimes residual

short-duration variability, which we attempt to screen as described below.

The light curves were detrended using the Cosine Filtering Autocorrelation Minimization (`CoFiAM`) algorithm (Kipping et al., 2013), which optimizes a superposition of sinusoids for the purpose of removing astrophysical variability, while preserving short-duration signals (that is, those that may be produced by the transit of the planet or a moon). The planet's transit is also masked during detrending to ensure that the algorithm does not attempt to fit out the transit signal. We briefly experimented with employing a method-marginalized detrending approach as employed in Teachey, & Kipping (2018), but ultimately decided against this as it was more prone to individual failures of one or more methods, and therefore could not be relied upon for a systematic approach without individual inspection of the detrended light curves.

Despite the robustness of `CoFiAM`, stars with variability on short timescales (i.e. comparable to the transit duration) can slip through with detrended light curves showing residual variability that are too pronounced for our purposes, as they are capable of mimicking moon transits. Because we produced $\mathcal{O}(10^5)$ simulated light curves, it was not feasible to examine each light curve by eye. Instead, we elected to screen badly-detrended light curves by comparing the median absolute deviation (MAD) to the expected standard deviation of a clean light curve (a flat out-of-transit baseline), assuming the photometric noise is Gaussian. For a Gaussian distribution MAD is equal to 1.4826σ, but we point out that for a finite set of measurements there is an expected distribution of measured standard deviations, given by $\sigma_{sample} = \sigma/\sqrt{2(n-1)}$, where n is the number of data points. After some experimentation we elected to reject detrended light curves for which MAD $> 0.6745\,(\sigma + 10\sigma_{sample})$. This does an excellent job rejecting most poorly-detrended light curves while also rejecting a small number of light curves which a by-eye treatment might have allowed for inclusion.

After screening, a total of 64912 generated light curve segments were used in the training, amounting to 80% of the total number of light curves utilized in training and validation. 10% of the light curves were held back for validation "under the hood" (to be used during training

as a check against over-fitting), and a further 10% were held back for additional validation.

Each light curve was normalized for input as

$$F_{norm} = \frac{F - min(F)}{\tilde{F} - min(F)} - 1 \tag{5.1}$$

where F is the array of fluxes and $\tilde{F}$ is the median value of the input segment. Because moons on high inclinations are not always guaranteed to transit, we ensured that any injected model that contained a moon but that did not in fact show a moon transit were reclassified as "planet only." We also ensured that the planet SNRs are in good agreement with the distribution of SNRs found in the real *Kepler* data. The true distribution of moon sizes in nature, and thus their associated moon SNRs, are currently unknown, and therefore cannot be used to ensure the training and validation sets are in good agreement with the real sample.

5.2.2 Preparation of the real *Kepler* data

It is important that the training set match the data to be classified as closely as possible. To prepare the real *Kepler* data, we first pulled planet transit timings from Holczer et al. (2016) to extract the light curve with the identical 5-day window on either side of that transit. These light curves would then be detrended using the same `CoFiAM` algorithm as utilized on the simulated data, rejecting once again badly detrended light curves, and normalized as before.

A large fraction of KOIs are multiplanet systems. Systems containing more than one transiting planet present a complication for the moon classifier, as transits of a neighboring planet can and will mimic a genuine moon signal. Restricting our moon search to planets without neighbor contamination would severely limit the number of systems / transits that could be examined, and would rule out the examination of some planets entirely. As such, we elected to remove contaminating transits of neighboring planets. To achieve this we computed a moving median trend line for the light curve, masking any neighboring planets' transits, and generating Gaussian-noise around the median trend line based on the noise

profile of the light curve (see Figure 5.1). While this approach has the potential to introduce a localized removal of any correlated noise, particularly variability with duration shorter than the neighbor's transit duration, we found this to be the most robust approach to the problem. Gaussian processes, by contrast, were significantly more computationally expensive, did not always produce a reasonable result, and were locally flat in the region of the masked transit, thus obviating the need for this more sophisticated approach.

We produced 10-day segment light curves for 5400 KOIs, regardless of the planet's period and its NASA Exoplanet Archive disposition (confirmed, candidate, or false positive). This resulted in 1,190,008 transit segments to be vetted with the CNN ensemble. In our final analysis we vetted 3223 KOIs, comprised of 1716 that are classified as 'confirmed' and 1507 classified as 'candidate' by the NASA Exoplanet Archive, and which had periods longer than 10 days. This amounted to 172,274 transit events.

5.2.3 False-positive test light curves

As discussed later in Section 5.3.3, we required an additional set of light curves to further discriminate against false positive exomoon identifications. Because each light curve presents unique astrophysical variation, which despite detrending and screening may have been preserved to some degree, there is the potential that these systematics may be capable of fooling the CNN ensemble into an erroneous moon classification. To guard against this scenario, we produced a false positive test sample for each individual KOI, consisting of at least 100 planet-only transits for each target to also be vetted by the CNN ensemble.

Following the neighbor-masking procedure described above, we masked every transit present in the light curve, including the target planet and any of its neighbors. We then injected planet transits at random locations throughout the light curve, adopting the fiducial parameters of the target planet from the NASA Exoplanet Archive. These light curves were then detrended and segmented in the same way as the real *Kepler* light curves, and fed through the same CNN ensemble classification pipeline to produce a false positive rate for

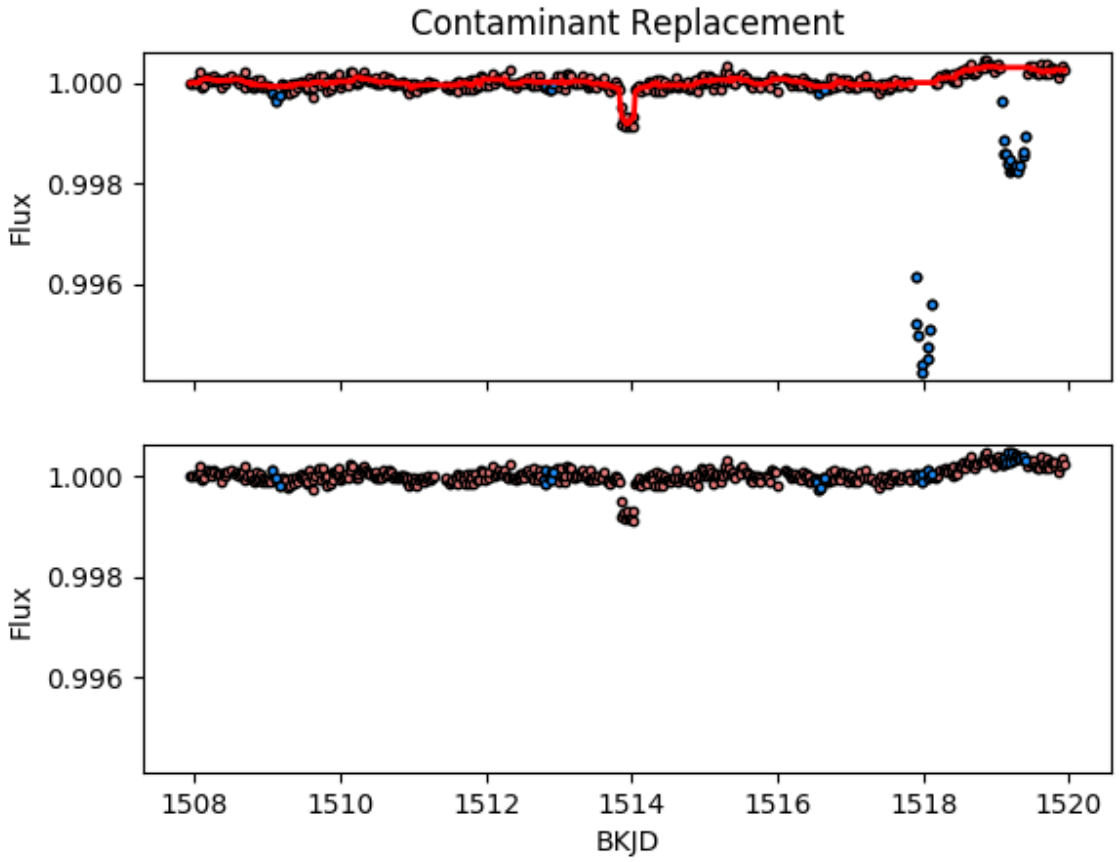

Figure 5.1: An example light curve demonstrating the removal of contaminating transits by neighboring planets in the system. These transits will mimic moon signals if they are not removed. The red line is the median filter, while transit locations are marked with blue points.

each KOI. That is, since every injected transit is known to only contain a planet transit (no moon), moon classifications will all be erroneous, and we therefore have a metric for how often the astrophysical variation of any given star in the sample produces a false positive.

5.2.4 CNN Design

Our CNN was constructed using the Python deep learning library (Chollet et al., 2018), utilizing the backend (Abadi et al., 2016). Our initial architecture and hyperparameters of the CNN were adopted from those described in Shallue & Vanderburg (2018). However, early testing suggested that this exact architecture was not able to produce the desired classification accuracy.

We experimented with a variety of hyperparameters. As each training run can take up to half an hour, it is computationally intractable to explore every possible hyperparameter combination without relying on a high performance computing cluster. We therefore elected to explore parameter space with an MCMC-type walker, initializing with hyperparameters known to achieve decent results through experimentation.

Hyperparameter variables explored by the walker were 1) the number of filters, 2) kernel size, 3) pool size, 4) pool type (average or max pooling), 5) stride length, 6) dropout rate, 7) the number of convolutional layers, and 8) the number of dense layers.

A sample architecture can be seen in Figure 5.2. The kernel initializer was orthogonal, and all activations were `relu` (rectified linear unit), except for the final activation which used `sigmoid`. Since the problem is binary classification, the loss function was set to "categorical crossentropy". The optimizer was "adam", and we optimized for classification accuracy. These settings were adopted from Shallue & Vanderburg (2018).

We found that in fact a variety of different hyperparameter combinations produced very similar validation accuracies, though there is no apparent pattern that points towards an optimum hyperparameter combination. Importantly, trained CNNs with effectively identical classification accuracies did not necessarily make the same predictions for a given transit.

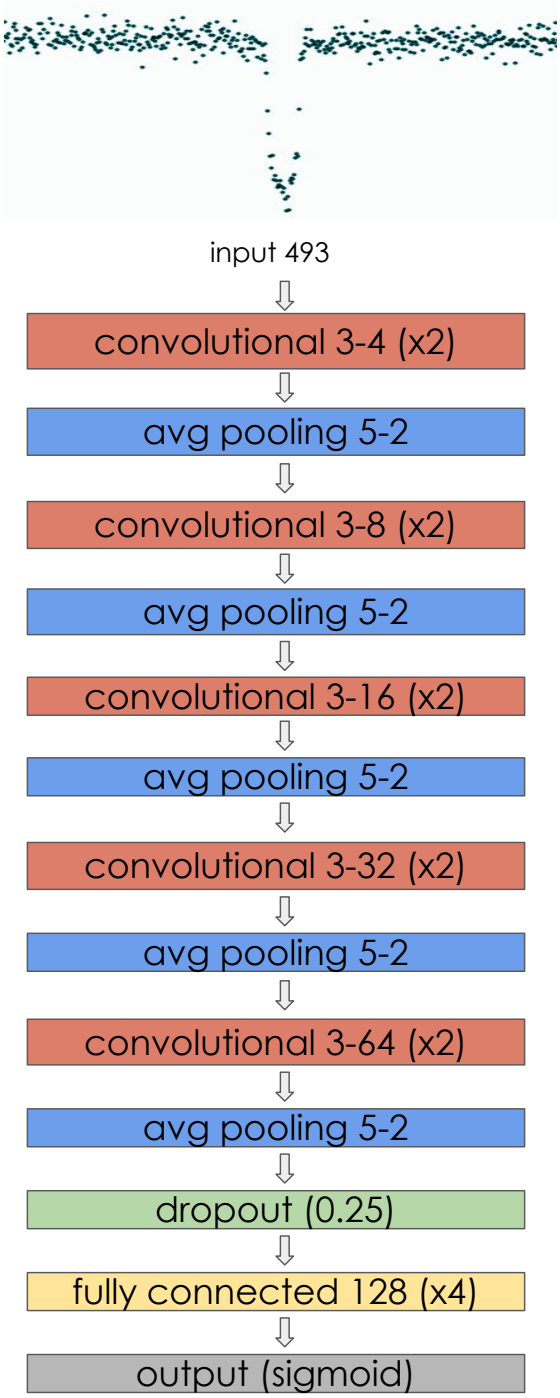

Figure 5.2: Example architecture one convolutional neural network in the ensemble. Following Shallue & Vanderburg (2018), the convolutional layers are marked as *kernel size* - *# of filters*, while the pooling layers are marked as *pool size* - *stride length*. Each convolutional layers is actually two layers in sequence, before each pooling (which may be average of max pooling). All hyperparameters indicated here were variable.

This suggested that it might be possible to leverage the power of an *ensemble* of classifiers to boost precision, analogous to the production of a random forest through an ensemble of decision trees. Altogether we used the best 50 CNN classifiers, all with individual classification accuracies > 80%. The distribution of hyperparameters for every CNN achieving > 80% accuracy can be seen in Figure 5.3. It is apparent that there is no obvious pattern to hyperparameter combinations that produced good results.

For our purposes, we are chiefly interested in identifying the most promising moon candidates. Computing population statistics are beyond the scope of this work; if that were our goal, we would be considerably more concerned with understanding the underlying performance of the CNN classifiers. Given our aim, we are comparatively less concerned by discarding those systems that may have moons present, but which the networks either fail to identify them or do not have sufficient confidence in a "moon". This is to say, we wish to minimize the false positive rate and maximize *precision*, defined as the ratio of true moon classifications over the sum of all moon classifications (true and false positives). This comes at the expense of *recall*, defined as the ratio of true moon classifications to the total number of moons in the sample (true positives and false negatives).

Every viable transit – that is, those with a 5-day baseline present on either side of the planet's transit – was analyzed for every planet in our sample. For each transit, a classification and confidence is produced for every voting model. The confidences are then used to augment the binary classification (1 for a moon, 0 for no moon) through a weighted average. That is:

$$\overline{x} = \frac{\sum_i^n w_i x_i}{\sum_i^n w_i}, \tag{5.2}$$

where $\overline{x}$ is the final classification of the transit in quest(> 0.5 for a moon, < 0.5 for no moon). The classification is further modified by using the "agreement metric", which allows us to tune precision and recall as desired. The agreement metric is defined such that total agreement of the voting models has a value of 1, and maximum disagreement (a 50/50 split)

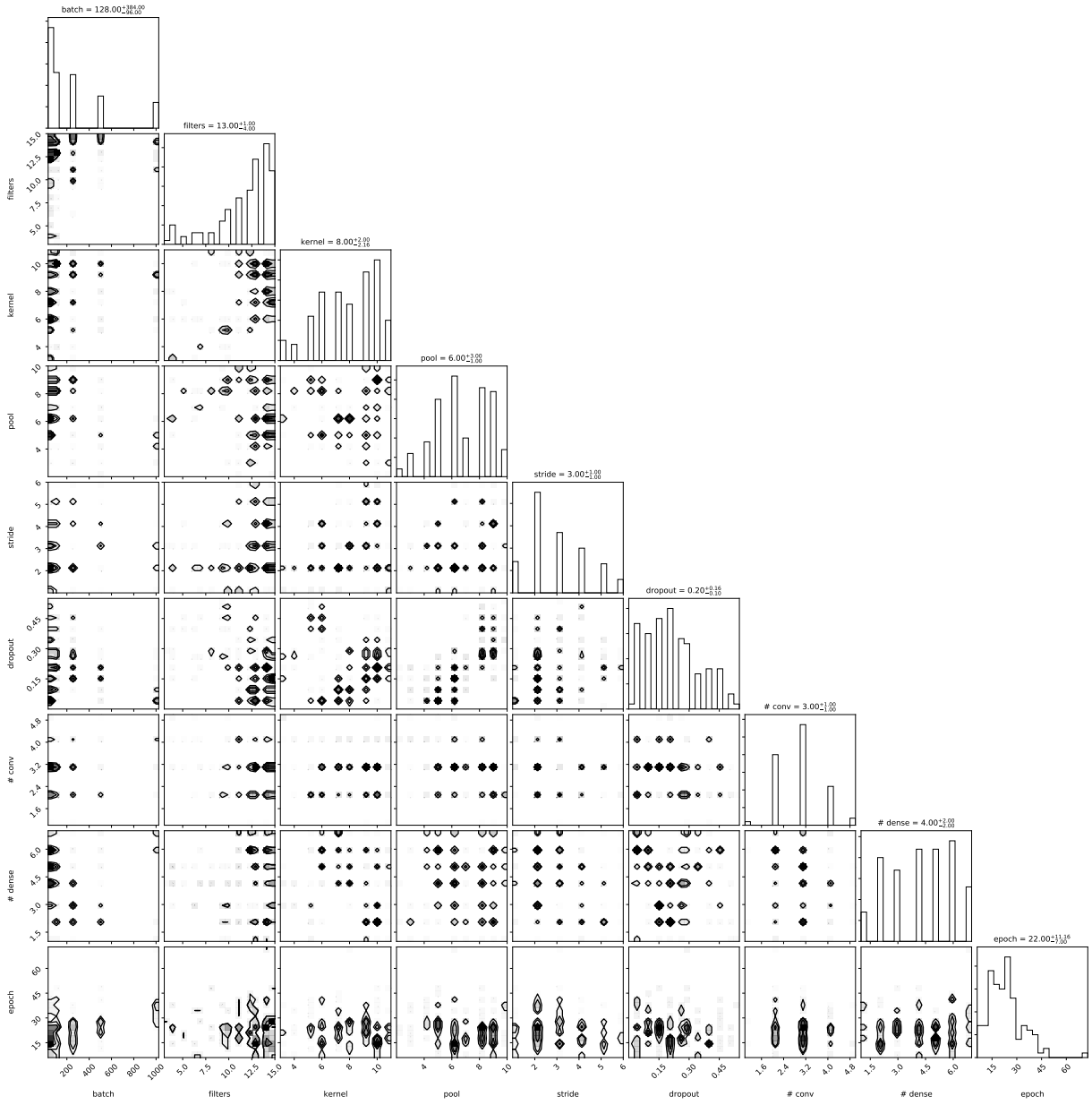

Figure 5.3: Distribution of hyperparameters for every CNN with validation accuracy > 80%. There appears to be a weak covariance between the dropout rate, and the kernel size and pool size.

has a value of zero. As shown in Figure 5.4, we can achieve higher confidence in our moon predictions when we require the models to be in good agreement (setting the agreement metric close to 1). Thus, a final moon prediction is made only in cases where the agreement metric surpasses our set threshold, otherwise it is re-classified as a no-moon prediction.

Unlike Shallue & Vanderburg (2018) and follow-on papers, we opt for using a single light curve input, rather than two parallel light curve inputs. In that work, the authors sought to distinguish genuine planets from false positive scenarios, and were therefore interested in the information that extensive out-of-transit data might reveal (for example, a phase curve and/or a secondary transit) as well as the morphology of a zoomed-in transit. Both of these input formats required binning of the light curve in order to make the input sizes uniform across a range of planet periods. In our case, however, binning is an inappropriate choice for the moon search, as it has the potential to wash out moon transits, and at the same time, there is little to be gained by employing a "global" view, since the moon will necessarily be located in close physical and temporal proximity to the host planet.

5.3 Results

5.3.1 CNN training results

As shown in Figure 5.4, we are able to achieve precision of up to 98% when the agreement metric for the CNN ensemble = 1. Once again, we define the agreement metric such that 1 corresponds to all CNN networks in agreement, and 0 means maximum disagreement (the number of "moon" classifications is equal to the number of "no moon" classifications). We note that even in the cases of large disagreement in the classifications, precision remains in the range of 90%, and recall is substantially higher.

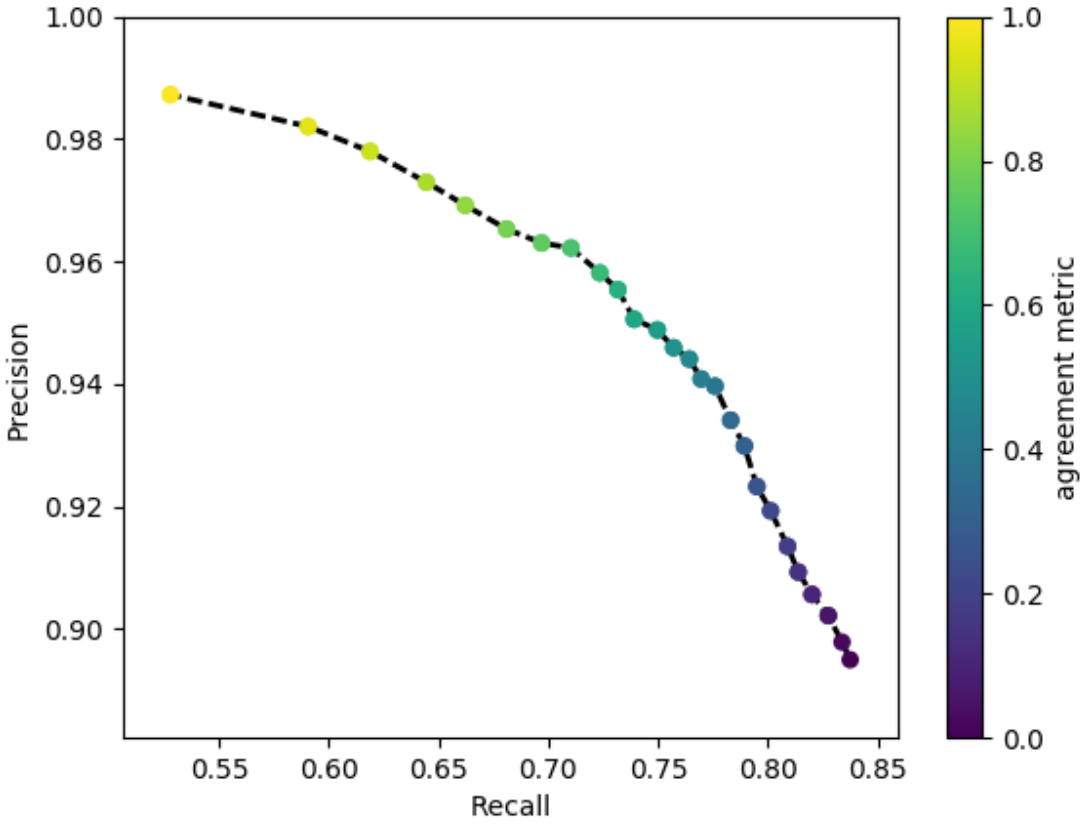

Figure 5.4: Precision-recall plot of the CNN ensemble on (simulated) validation data, indicating that maximum precision is acquired when the ensemble is also in total agreement.

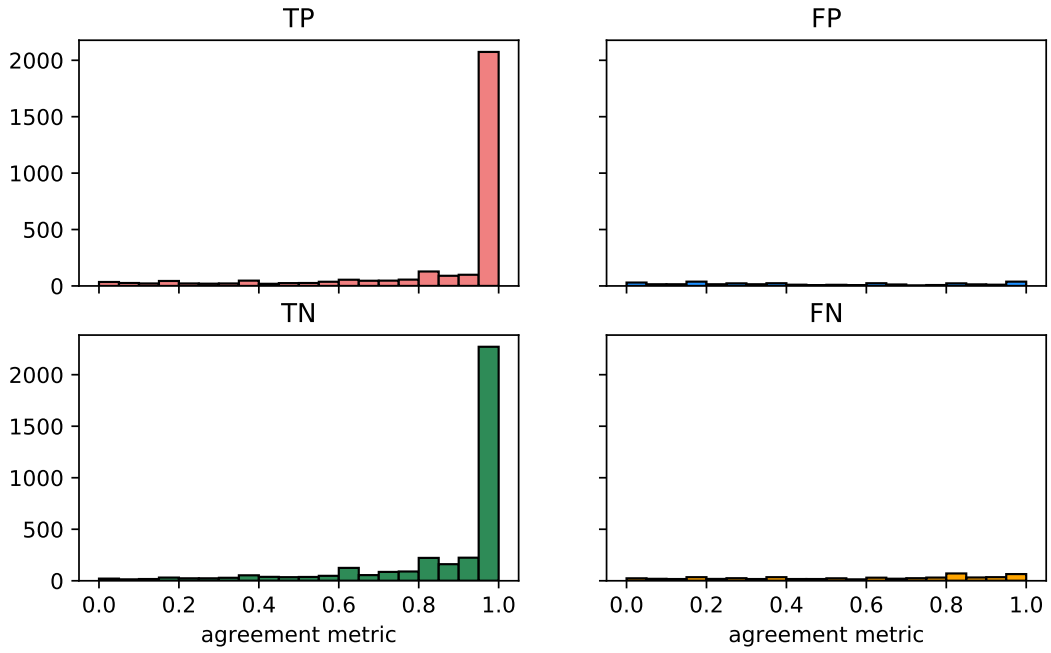

Figure 5.5: Distribution of true/false positives/negatives as a function of ensemble agreement for the validation light curves (simulated data). The majority of classifications come with maximum ensemble agreement.

5.3.2 Application to real light curves

Classification statistics

A total of 1716 "confirmed" planets and 1507 "candidates" (as classified by the NASA Exoplanet Archive) were vetted – 3223 KOIs in all. This constituted 172,274 transit events in the final sample. The fraction of these transits classified as having a moon by the CNN ensemble was 21.6% (note that this is a fraction of transits, not planets). However, only 1.9% of these transits met our highest standard of confidence, with agreement metric greater than 0.95. This metric therefore provides considerable discrimination power in identifying the best candidates to pursue for follow-up, though of course, given the reliability of the CNN even with somewhat lower agreement, we might reasonably suspect that up to $\sim 20\%$ of transits could in fact contain moon signals. The mean false positive rate across all vetted systems based on the planet-only light curve injection test was 0.81%.

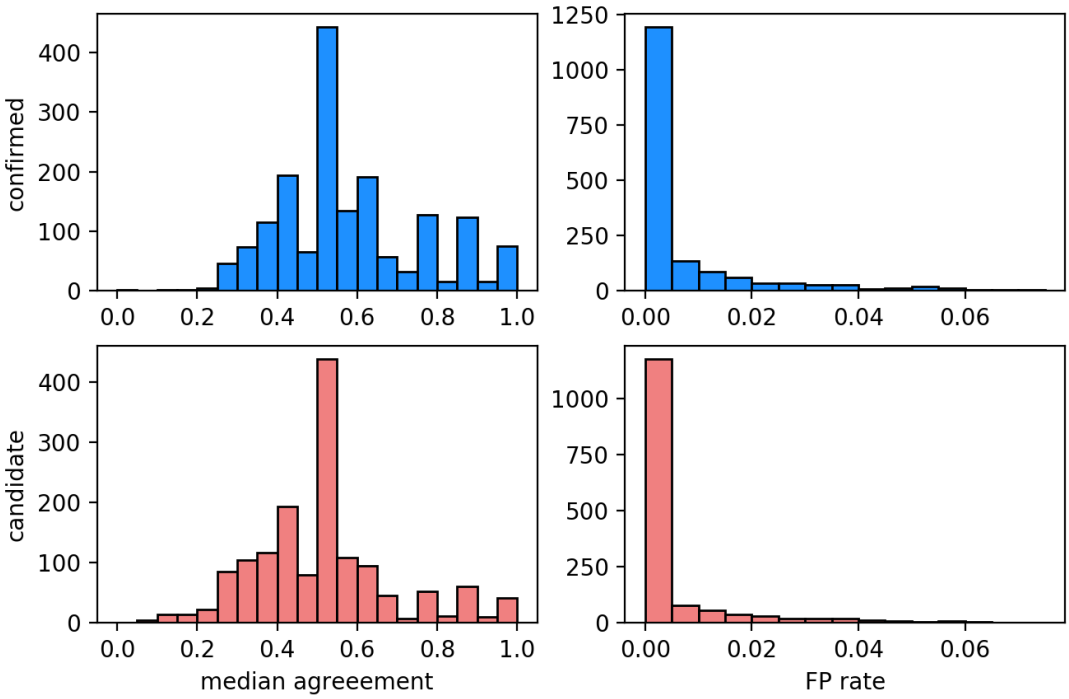

Figure 5.6: Median agreement and false positive rates for confirmed and candidate planet KOIs.

A total of 394 confirmed planets (23.0%) were classified as having more than 50% of their transits containing moons. Meanwhile, 332 candidate planets (22.0%) were classified as having more than 50% of transits containing a moon signal. However, only 22 confirmed planets and 22 additional planet candidates were classified as having more than 50% of their transits containing a moon when applying our 0.95 agreement metric threshold.

Comparison in agreement metrics

We note that the real sample of KOIs shows markedly lower median agreement metrics on average (peaking around 0.5, see Figure 5.6) than does the training sample, which in the majority of cases showed an agreement metric near 1 (Figure 5.5). Median agreement for the false positive test light curves also peak around 0.5, but with a tail towards even lower values. We find that the median agreement for the real "confirmed" KOIs is higher than

that of the false positive test light curves in 78% of cases, and for the "candidate" KOIs the agreements are better than the false positive agreements 59% of the time.

The cause or causes of these differences are not readily apparent, though it could suggest underlying differences in the samples. One possible difference could be related to stellar activity. As the training sample is taken from stars for which no planets have been detected, we must ask whether the absence of planets in these light curves is due to a real absence of transiting planets or to our inability to detect these transiting planets (likely because the planet is small and has been lost in the noise, or because the star exhibits short-duration variability that hides these transits).

Further discrimination of moon classifications

Despite the relatively high confidence we may have in the CNN's performance on simulated data, the ensemble's application to real data requires further scrutiny. Because the underlying exomoon population in the *Kepler* sample remains essentially unknown, it is difficult to know if the number of moon predictions the CNN ensemble produces is too high or too low. We note, however, that (Teachey et al., 2018) provided some basis for comparison, having found a low occurrence rate $\eta = 0.16^{+0.13}_{-0.10}$ of Galilean analogs in the *Kepler* data.

When applied to the real data, the CNN ensemble returns a large number of moon predictions to high confidence. At first glance this is encouraging, since it suggests there is an abundance of moons in the sample waiting to be confirmed. but if our purpose is to identify the most promising candidates, to be vetted with computationally expensive model selection, this abundance of potential signals to follow-up is unwieldy. The question naturally arises as to whether and to what extent these classifications may be due to other features in these light curves that are mimicking exomoon signals.

5.3.3 Further vetting

From our full list of KOIs we identified a number of promising systems deemed worthy of a closer look. In total 49 systems were identified in our final CNN classification run which had a significant fraction of moon predictions, made to high confidence and with low false positive rates. Nine targets were selected subjectively from this list for further vetting, based on visual inspection of their light curve morphologies as well as physical and/or observational properties. Below we discuss the the systems we have examined in greater detail through photodynamical model selection.

Photodynamical model fitting

We perform model fits for nine of the most promising systems identified by our CNN ensemble pipeline, listed in Table 1. Four models, labeled 'P', 'T', 'Z', and 'M', were generated with `LUNA` and fit with MultiNest. As described in (Teachey, & Kipping, 2018, Chapter 3[), the planet model 'P' is simply a planet with no moon in the system, and assumes strict linear ephemeris (there are also no other perturbing planets in the system). The 'T' model adds an additional transit midtime parameter to be fit for each observed transit, thus testing for the presence of TTVs while being agnostic about their cause (it could be a moon, or a perturbing planet, for example). The 'Z' model places a zero-radius moon into the system, thereby allowing the model to capture all the dynamical effects that a moon present in the system would produce, without requiring the presence of a moon transit signal in the photometry. Finally, the 'M' model is a fully physical moon model, fitting both the dynamical effects (timing and duration variations) as well as the moon transits.

To start we note the encouraging result that in every case for the planets examined below, the 'M' model evidences are significantly better than those of the 'P' model. It is worth remembering that the model evidence is the key metric for model selection, and accounts for (penalizes) the additional parameter space generated by more complex models.

The 'T' and 'Z' models become important additional checks, providing us a window into

which signatures in the photometry are driving the case for the moon. With the CNN, we should naturally expect to see favorable evidence for the 'M' model over the 'Z' model, because the CNN is trained to look for moon transits and not for dynamical signatures. We note that in some cases, non-convergence of 'Z' models forced us to revert to an auxiliary approach, employing the Savage-Dickey ratio, to compute M-Z. We indicate these instances as M-Z*.

It is perhaps worth keeping in mind, however, that in cases where M-Z is not positive (i.e., the 'M' model is not favored over the 'Z' model), we cannot, strictly speaking, conclude that the CNN identification is a false positive. It may very well be that a moon transit is present, and that the CNN has detected it, but there is insufficient evidence to adopt the more complex moon model. Consider a particularly noisy light curve, for example: in such a case, a moon could easily be lost in the noise and possibly visible, but because of the noise profile we cannot justify the additional complexity of the moon model.

For clarity and simplicity, we apply a subjective, four-tiered classification system based on the model evidences and our own interpretation of the model fits. They are

- Class A: candidate, should follow-up
- Class B: plausible, has potential
- Class C: ambiguous, some tension
- Class D: unlikely, low-priority

In what follows we will first assign a classification based on the rubric above, and then report the details of our model fits. The casual reader may wish to skip these details, but we report them here for the sake of posterity and transparency. We examine not only the global model evidences, but also the predictions for each transit, evaluating whether the 'M' model places a moon transit in that epoch. We also explore whether we see an improvement in the Bayesian Information Criterion (BIC) indicating the presence of significant TTVs and TDVs, comparing a best fitting sinusoidal model against a flat line.

KOI	Period	Moons / Transits	M - P	M - T	M - Z	P_{TTV}	A_{TTV}
622.01	155.04	4/8	5.04 ± 0.13	169.25 ± 0.15	−2.78 ± 0.12	6.1	6.2
806.02	60.32	8/12	4470.06 ± 0.16	632.70 ±0.19	3.87*	16.4	28.3
865.01	119.02	5/9	14.83 ± 0.14	478.78 ± 0.18	10.3*	9.6	2.2
902.01	83.92	5/9	1102.96 ± 0.14	49.17 ± 0.19	-2.3*	15.2	53.8
1005.01	35.62	18/27	30.69 ± 0.14	-170.09 ±0.17	31.65 ± 0.15	7.8	3.5
1162.01	158.69	3/4	26.1 ± 0.13	24.1 ± 0.16	-1.4*	9.5	15.5
1553.01	52.76	11/18	6.9 ± 0.15	-112.75 ±0.19	1.6*	4.7	1.7
2046.01	23.90	6/12	13.76 ± 0.1	-58.80 ±0.16	3.7*	5.4	15.8
2677.01	237.79	4/4	11.03 ± 0.12	23.08 ± 0.14	15.5*	4.1	9.5

Table 5.1: Summary of KOIs identified by the CNN ensemble that underwent full Bayesian model selection. The "Moons / Transits" column indicates the ratio of high confidence moon predictions to the total number of transits analyzed. Columns 4 through six are Bayes factors based on comparison to the planet-only, TTV model, and zero-radius moon models, respectively. In cases where multiple TTV models were run, the results for each model evidence difference is shown in parentheses. The last two columns provide the inferred TTV period and its amplitude. M-Z values with an asterisk (*) indicate M-Z models for which the Bayes factor is derived from computing the Savage-Dickey ratio. P_{TTV} is the period of the best fitting TTV model, in number of epochs. $ATTV$ is the best fitting amplitude of the TTV, in minutes.

For the sake of clarity, we will refer to a given epoch for which a moon has been predicted by the CNN ensemble and where it is also modeled as *true positive* (TP), an epoch where a moon is not predicted and also not seen in the model as a true negative (TN), etc. It should be kept in mind, however, that we cannot know the ground truth in these cases, so strictly speaking TP / TN / FP / FN classifications are not applicable in this case.

KOI-622.01

CLASS D. KOI-622.01 was identified as an intriguing candidate for follow-up modeling due to its relatively long period (155 days), the number of transits identified by the CNN ensemble as having moon transits (75%, with 50% above the 95% agreement threshold), high median agreement, low false positive rate, and visual inspection of the light curve itself. Specifically, there appeared to be some asymmetry in the phase-folded transit, as well as evidence of flux reduction in the wings of the transit.

We measure a TTV period of 6.3 cycles, with an amplitude of 8.6 ± 0.9 minutes. We find ΔBIC(TTV) = 8.1 and ΔBIC(TDV) = 8.5, indicating evidence for TTVs and TDVs. We find

the TTV is favored with Bayes Factor K of 2.5, and the TDV with a $K = 2.7$. Utilizing the framework introduced in Kipping & Teachey 2020 for calculating minimum moon semimajor axis $f = a_S/R_{Hill}$, we find $f = 0.045^{+0.026}_{-0.014}$.

MultiNest found two modes in the 'M' model, with mode 2 having the higher evidence, evidently associated with prograde and retrograde solutions. Both 'M' and 'P' models find $R_P/R_* = 0.072 \pm 0.002$, placing the inferred radius of KOI-622.01 at ~2.4 Jupiter radii (adopting a fiducial value of $R* = 3.32^{+0.53}_{-0.65} R_\odot$). This is consistent with published values ($26.44^{+4.23}_{-5.18}$), but is quite puffy for a planet with $T_{eq} = 511$ K. In fact, there are only 10 planets listed in the NASA Exoplanet Archive as of this writing with larger radii. Taking this as the planet's radius, the inferred moon in the 'M' model would then be $\sim 1.9\ R_\oplus$. We note that the stellar radius was revised upward considerably for *Kepler* DR25, from a previous sub-Solar value of $R_* = 0.77^{+0.23}_{-0.07}$. Therefore, it is conceivable that this planet, though not yet listed as false positive, could in fact be a low-mass star, though spectroscopic and / or RV analysis would be needed to confirm this hypothesis. We see no evidence for a secondary eclipse, however, which would be expected for an eclipsing binary.

The CNN ensemble classified six of eight analyzed transits as having moon signals, though only four of those six were classified with sufficient agreement to be considered robust. Modeling of the moon shows moon transits in six of eight epochs for the best mode, and transits in all eight epochs for mode 1. If we take mode 2 as the standard, the CNN then has four true positive identifications, two false positive and two false negative predictions. The transits modeled in mode 1 are of sufficient depth to be detectable in the future (~ 1900 ppm), while the transits in the favored mode 2 are significantly shallower (~ 200 ppm).

In any case, while the 'M' model is favored over the 'P' model, it does not outperform the 'Z' model(M-Z = -2.78, M-Z* = -0.69) , which is the key test for the presence of a moon dip. Therefore, we cannot consider this a particularly attractive target for follow-up at this time. The discrepant evolving stellar radius value and the the large size of the planet inferred also give us pause.

KOI-806.02 (Kepler-30c)

CLASS C. KOI-806.02 (Kepler-30c) was identified as an intriguing candidate for follow-up modeling due to the number of transits identified as having moon transits (75%, with 67% above the 95% agreement threshold), high median agreement (100%), low false positive rate, and high SNR for the planet.

KOI-806.02 is one of three known planets in the system, and exists in a near 2:1 resonance with an inner planet, KOI-806.01 Panichi et al. (2018). The presence of significant TTVs is therefore to be expected, and indeed, we measure a TTV period of 16.4 cycles, with some evidence of non-sinusoidal morphology. The amplitude for the TTV signal is calculated to be 28.3 minutes The SNR of the planet's transit suggest the system is attractive for follow-up with higher precision telescopes, but examination of the light curves suggests some short-duration variability in the light curves that could therefore mimic a moon signal. A rotation period of 16.004 ± 0.017 days was measured in McQuillan et al. 2013, with a variability amplitude of 14.75 millimag.

The 'M' model solution is highly multimodal (16 modes in all), all with very similar evidences. The 'M' model is favored over the planet model with $K = 4470.06$, indicating a very strong preference for the planet-only model. However, the moon model is favored over the zero-radius moon model with a far more modest $K = 3.87$ (Savage-Dickey). This indicates that the M-P value is driven primarily by the presence of dynamical effects seen in the light curve, which are due at least in large part to the other perturbing planets in the system. The effects of these other planets have not been jointly modeled in this work.

The CNN ensemble classified 16 out of 21 transits as having evidence for a moon, though for many of the 'M' model modes there are no moon transits at all, suggesting an inclined moon solution, manifesting as a purely dynamical signature. Several modes show some moon transits however (up to 10 out of 21).

We also note diverges between the 'M' and 'P' results for a variety of parameters. We see discrepant results in limb darkening coefficients q_1 and q_2, reference transit time τ_0, period

of the planet P, impact parameter b (b_M near zero, b_P at ~ 0.7), density of the star ρ_* (~ 900 kg m^{-3} for the planet, ~ 2400 for the moon model), and R_P/R_*. We note that fiducial values of ρ_* are in closer alignment with the moon model result. The published impact parameter ($0.224^{+0.067}_{-0.146}$), meanwhile, agrees with neither the planet nor the moon model result, though it is somewhat closer to the moon solution. These discrepancies could be due in part to the fact that none of our models attempt to account for the effects of other perturbing planets in the system.

We conclude that though there is some evidence of moon transits in this light curve, the system is not particularly attractive for follow-up work, due to 1) multimodality of the 'M' model solutions, including the disagreement as to whether the moon transits at all, and 2) the challenges of disentangling transit timings due to a moon and to near-resonant perturbations. This system highlights the outstanding need to jointly model planet-planet perturbations alongside moon-induced dynamical effects in future work.

KOI-865.01

CLASS B. This planet was selected for additional vetting due to its long period (119 days), the large fraction of transits diagnosed by the CNN ensemble as containing a possible moon transit (8 of 9, 5 meeting the agreement metric threshold), perfect median agreement, a low false positive rate (1%), decent SNR, and a possible OSE signal seen in the wings of the planetary transit.

Our best fitting TTV model finds period of 9.6 cycles, but with a small amplitude (2.2±0.7 minutes), and in fact, the TTV model is formally disfavored through the BIC (Δ BIC = -1.5). We also compute a minimum semimajor axis for the moon based on the TDV floor, which is at $f = 0.47 R_{Hill}$, nearly at the limit of stability for a prograde moon.

Like KOI-806.02, the 'M' model for this planet is highly multimodal, in this case, 11 modes with similar evidences. The 'M' model is favored over the 'P' model with a $K = 14.83$, And importantly, the 'M' model also outperforms the 'Z' model with $K = 10.3$ – the key

test for the presence of a moon dip in the light curve. The period of the satellite is weakly constrained in the 'M' model.

If we take highest evidence mode 1 as the true model of the system, the CNN ensemble registers three true positive detections, one false negative, and five false positives. Across the various modes, the CNN identified as many as six transits "correctly" and as many as eight "incorrectly".

The system is attractive from a follow-up standpoint, given its high SNR, and that the moon model is formally favored (though with mixed results on how often the moon transits, and lack of evidence for transit timing variations, which remain an important component of testing the moon hypothesis). Future transit observations with higher precision could be fruitful, but the system does not stand out to such a degree that targeted observations should be pursued in the near term.

KOI-902.01

CLASS C. KOI-902.01 was selected for its relatively long period (83.9 days), high fraction of moon transit predictions (67%, with 55.6% meeting the agreement threshold), perfect median agreement, 0% false positive rate, and presence of TTVs in the system despite there not being any other known planets in the system.

We measure a TTV period of 15.2 cycles, which shows some evidence of non-sinusoidal morphology. The amplitude of the TTVs is measured to be 53.8 minutes, and the TTV is favored with $2 \log K = 3289.7$. McQuillan et al. 2013 measured a rotation period of 15.392 ± 0.02 days, with a variability amplitude of 8.62 millimag.

MultiNest finds 5 modes, with best evidence for mode 3. The 'M' model is strongly favored over the 'P' model with M-P = 1102.96. However, the evidence for the moon is strongly driven by the presence of TTVs in the system, as the 'M' model is formally disfavored against the 'Z' model with M-Z* = -2.3. It is interesting to note the presence of TTVs in the system despite there being no other known planets in the system. This could suggest

the presence of a moon, but of course the presence of TTVs by themselves are not sufficient evidence of a moon. The detection of some other object in this system is anticipated with future observations.

The best fitting limb darkening coefficients change between the planet model and planet+moon model. Our planet only model infers a much higher impact parameter ($b \sim 0.9$), whereas the moon model places it at a more modest $b \sim 0.5$. We also see wildly divergent values for ρ_*: the planet model finds $\rho_* = 390.61^{+30.49}_{-25.77}$ kg m^{-3}, whereas the moon model derives a value of $2993.54^{+101.81}_{-95.69}$. Meanwhile, ExoFOP reports $\rho_* = 1459 \pm 226$ kg m^{-3}, so neither model accords particularly well with the published value, suggesting a potentially eccentric planet. The inferred semimajor axis for the moon is reasonably large, at $a_S = 87.03^{+9.11}_{-15.22} R_P$.

The CNN predicted six of nine transits as containing the presence of a moon. Across the various modes, this corresponds to as many as seven, and as few as two, accurate predictions. For the highest evidence mode, we register four true positives, one true negatives, two false positives and two false negatives. However, it should be remembered that the moon model is disfavored against the zero-radius moon model in this case.

KOI-1005.01 (Kepler-747 b)

CLASS B. Kepler-747b was selected due to its large number of moon transit predictions (21 of 27, with 18 meeting the agreement threshold), perfect median agreement, low false positive rate (4%), and possible orbital sampling effect signal as seen in the phase-folded light curve. The planet's orbital period is among the shortest of our sample, at 35.6 days, but like most of the other systems we've examined, is the only known planet in the system.

The 'M' model is favored over the 'P' model with a M-P = 30.69. Inferred system parameters are broadly similar across the two models, though impact parameter and stellar density are somewhat better constrained in the moon model, with $\rho_* = 546.26^{+29.16}_{-73.14}$, which is in good agreement with the the published value on ExoFOP (593.2 ± 5.7). Remarkably, the 'M' model is favored over the 'Z' model with even greater significance, at 31.65. This

suggests that the the absence of significant dynamical effects are somewhat in tension with the moon solution. We measure a TTV period of 7.8 cycles, but with only marginal evidence for the TTVs (ΔBIC = 2.5), and a small TTV amplitude of only 3.9 ± 0.9 minutes. These values by themselves are suggestive of a moon. On the other hand, the presence of TDVs are disfavored with ΔBIC = -1.7, and we calculate a minimum semimajor axis for the moon from the TDV floor to be 0.55 R_{Hill}, requiring a retrograde moon for stability. The 'M' model is also formally disfavored against the 'T' model. To bring these results into alignment in favor of a genuine moon, we require an large primary mass ($\sim 30\, M_J$, suggesting it lies in or is approaching the brown dwarf regime.

Two other scenarios could be responsible for a strong M-Z value without the presence of associated timing variations. The first is the presence of multiple moons, which could have the effect of washing out TTVs by complicating the barycentric wobble of the planet. A low density moon could also potentially be responsible for a lack of substantial dynamical effects. With an inferred radius for the planet of 0.47 R_J, it would be on lower (denser) end of sizes for this mass, but this is not so far off as to be inconceivable. Future RV measurements of this system would be worthwhile to test whether the high mass primary prediction holds water.

The CNN ensemble predicted 21 of 27 as showing moon transits. 18 of the 21 moon predictions met the agreement threshold. If we take the 'M' model as the ground truth, the CNN achieves 21 true positives, 1 true negative, 6 false negatives, and 0 false positives. In fact, the moon model places a transit feature of some kind in 33 of of 37 epochs (as a reminder, not every transit is vetted by the CNN). This high fraction of moon transits makes it potentially attractive for follow-up.

In summary, the case for the presence of a moon based on the model evidences alone is strong, and the CNN predictions may lend additional credence to the moon hypothesis. However, the relatively short orbital period of the planet, and the requirement of a very massive primary to account for the absence of TDVs, arouses skepticism.

KOI-1162.01 (Kepler-782b)

CLASS C. This planet was selected for additional follow-up due to its long period (158.7 days), large fraction of transits showing a moon feature (3 of 4, all meeting the agreement threshold), perfect median agreement, low false positive rate (2%), and observed features in the light curve that looked like potential moon signals.

Kepler-782b is the only known planet in the system, and is among the smaller planets analyzed here, with $R_P = 3.08 \pm 0.44\, R_\oplus$. We measure a TTV period of 9.5 cycles, with an amplitude of 15.5 minutes. The TTV is favored over linear ephemeris with $2 \log K = 53.3$, suggesting that there is another object in the system, either a moon or another yet-undetected planet.

MultiNest finds only a single mode in the 'M' model, with a particularly modest moon dip of only ~ 50 ppm, though this represents a sizeable planet-to-moon radius ratio R_S/R_P of ~ 0.2. We see broad agreement between the 'M' and 'P' models, though we note that R_P/R_* and ρ_* are better constrained in the 'M' model ($rho_* = 921.26^{+33.32}_{-88.16}$ kg m^{-3}.). The 'M' model is favored over the 'P' model with M-P = 26.10, but we find that it is marginally disfavored with respect to the 'Z' model at -1.4 (Savage-Dickey), indicating insufficient evidence for the presence of a moon dip in the light curve. The evidence for the moon is thus driven by the presence of dynamical effects which are not captured in the linear ephemeris 'P' model.

If we were to take the 'M' model as ground truth, the CNN predicted 3 of four transits correctly: 2 true positives, 1 true negative, and 1 false positive. That false positive occurs in the first epoch ($\tau = 2455165.6$). In any case, since M-Z is underwater, we must conclude that this planet is not particularly attractive for follow-up observations.

KOI-1553.01 (Kepler-890b)

CLASS D. This planet was selected for additional vetting due to its large fraction of transits diagnosed as having a moon feature (14 of 18, with 11 meeting the agreement metric), perfect median agreement, and low false positive rate (5%). Visual inspection suggested the

presence of in-transit asymmetries, as well, which could be due to the presence of a moon.

Kepler-890b is the only known planet in the system, with an orbital period of 52.76 days. We measure a TTV period of 4.7 cycles, but with a very modest amplitude (1.7 minutes), which by themselves are suggestive of a moon. However, we find only marginal evidence in favor of the TTV over linear ephemeris ($2 \log K = 0.9$).

MultiNest found 5 modes in the 'M' model, with 4 out of 5 showing effectively identical evidences. The 'M' model is favored over the 'P' model with M-P = 6.90. However, the presence of a moon dip is not favored; though we calculated M-Z* = 1.6, this is for a *negative* radius moon, which is of course unphysical. In general, we see cleaner posterior solutions with the planet+moon model, and the moon mass prediction is among the more "plausible" of our sample, at $M_S = 0.004^{0.003}_{-0.002}\, M_P$. But the inferred period of the moon is poorly constrained. If we were to take the moon models as the ground truth, the CNN made as many as seven and as few as three accurate predictions across the various modes.

We must conclude that the absence of a discernible transit feature in the light curve renders this target unattractive for follow-up moon searches.

KOI-2046.01 (Kepler-1069b)

CLASS C. This planet was selected for additional vetting due to the large number of planet transits diagnosed as having a moon signal present (8 of 12, with 6 meeting the agreement threshold), perfect false positive rate, its fairly clean transit with possible asymmetrical morphology, and possible moon dips seen by eye. An additional four transits were vetted, but were excluded for failing our MAD variability filter. Of those, 3 were classified as containing a moon, 1 was not.

The planet has the shortest period of our sample, at 23.9 days. It is also the smallest, at $1.6^{+0.31}_{-0.14} R_\oplus$. We find broad agreement between the 'M' and 'P' model parameter solutions, though we see a bimodal solution in R_P/R_* in the 'M' model that is not present in the 'P' model. We measure a TTV period of 5.4 cycles, with amplitude of 15.8 minutes, both

suggestive of a potential moon. The TTV is only marginally favored, however, with $2 \log K = 1.4$.

MultiNest finds 5 modes, of which the 1st mode has the highest evidence. The 'M' model is favored over the 'P' model with M-P = 13.76. Critically, we find the 'M' model is also favored over the 'Z' model, with M-Z* = 3.7. Thus, we see evidence of both dynamical interactions and moon-like transit features.

We see two peaks in ρ_* in both 'M' and 'P' models, and both quite low, at around 30 and 50 kg m^{-3}. These values accord with the (also poorly constrained) value found on ExoFOP, where $\rho_* = 281.5 \pm 263.5$, indicative of a star climbing the giant branch. We also find very large ratio of radii R_S/R_P (two peaks around ~ 0.4 and 0.78), which is probably unphysical given the modest TTV amplitudes. The inferred mass ratio, which is of course tied directly to the dynamical effects of the modeled moon, is quite low by comparison ($M_S = 0.014^{+0.053}_{-0.010} M_P$). Unlike KOI-1005.01 (Kepler-747b), where we can potentially invoke a massive but compact primary to explain a large satellite with weak dynamical evidence, in this case the solid nature of both the planet and moon makes this mass ratio / radius ratio discrepancy likely unphysical.

The 1st (highest evidence) mode places a transit feature in all 16 epochs analyzed by the CNN. As the CNN predicted 11 transits in all, this would correspond to 11 true positives and 5 false negatives. Among all the transits fit in MultiNest, we see 23 out of 28 epochs showing moon transit features. We remind the reader that CNN ensemble only analyzed epochs for which sufficient out-of-transit baseline was available on either side of the planet transit, while the moon models have no such constraint and are therefore fit to every existing transit.

We conclude that while the moon model is formally favored over both the 'P' and 'Z' models, and the best fit TTV model is consistent with the presence of a moon, the tension between the satellite mass ratio and radius ratio lead us to remain skeptical of the moon hypothesis for this system. The relatively short period of the planet also cuts against our

expectations of where we ought to expect a moon *a priori*, though we cannot exclude it on those grounds. The low SNR of the planet would make follow-up confirmation of a moon challenging.

KOI-2677.01

CLASS B. KOI-2677.01 was selected for further analysis due to its long period (237.8 days), and the prediction that all four of the analyzed transits contained a moon signal to high confidence. There were also no false positive detections, and visual inspection of the light curve suggested some possible asymmetry in its V-shaped light transit.

The host star is estimated to be around a solar mass, but with an inflated radius of $1.853^{+0.343}_{-0.801}\,R_{\odot}$. We note that this estimate was revised upward in DR25 from $1.123^{+0.582}_{-0.133}\,R_{\odot}$. Mazeh et al. 2015 measured $P_{rot} = 90.06 \pm 2.88$ days.

Our best fit TTV has a period of 4.3 cycles and amplitude of 9.4 minutes, favored over linear ephemeris with ΔBIC of 9.4. These features, combined with the lack of any other known planet in the system, are suggestive of a moon's presence. However, we do not find evidence for a TDV, computing ΔBIC = -1.5. We calculate a minimum moon semimajor axis of $0.2\,R_{Hill}$ based on the Kipping & Teachey 2020 methodology.

The 'M' model outperforms the 'P' model with M-P = 11.03. Meanwhile, the 'M' model is a greater improvement on the 'Z' model, with M-Z* = 15.5. This suggests that the evidence is driven primarily by the morphology of the light curve, and that the absence of dynamical effects on the primary are somewhat weakening the case for the moon. That is, the 'P' model is also preferred to the 'Z' model.

Our 'M' model places tighter constraints on the limb darkening coefficients, which are unconstrained in the 'P' model. The 'M' model also reduces the impact parameter considerably ($b = 0.20^{+0.17}_{-0.13}$), whereas the 'P' model is consistent with a grazing transit ($0.944^{+0.029}_{-0.014}$). Values for stellar density are also considerably different across the two models, where the 'P' model finds a median value closer to the published value ($163.60^{+38.80}_{-54.79}$ m^{-3}), and the 'M'

model is more than an order of magnitude higher, at $3306.70^{+275.92}_{-468.49}$ kg m^{-}3. This value is poorly constrained on ExoFOP (239.2 ± 885.1 kg m^{-3}).

Critically, we find a moon-planet radius ratio $R_S = 0.94^{+0.04}_{-0.06} R_P$ (a binary solution), though with a poorly constrained mass solution ($M_S = 0.66^{+0.25}_{-0.45} M_P$), likely owing to the tension between the morphological evidence of a moon and the lack of a dynamical counterpart. We note however that these masses are not totally inconsistent with a radius ratio approaching 1 given the broad range of masses possible around the radius of Jupiter. The 'M' model also places the satellite close to the planet ($R_S = 15.47^{+2.81}_{-1.93} R_P$), which may be in tension with the minimum semimajor axis computed above. For this planet $\sim 0.5 R_J$ planet around 0.8 AU, the Hill sphere is likely to be quite large, though if the planet is on an eccentric orbit this could be reduced to a degree that the tension is removed.

The CNN ensemble classified all four analyzed transits as having evidence of a moon transit, with universal agreement. This was consistent with the 'M' model, which placed a moon in every epoch, with a depth of $\sim$ 800 ppm, roughly half the total transit depth. A curious result of this analysis is the fact that we have no false positives when we inject planet-only transits into the light curve. As a reminder, we adopt fiducial values for the planet properties when injecting these transits into the light curve. Thus, the false positive test knows about the size and impact parameter of the planet, yet the CNN is evidently not fooled by this. And still, it identified all four planet transits as having evidence of a moon.

Despite the model preference for the 'M' model over both 'P' and 'Z' models, the absence of substantial dynamical signatures of a satellite in this system give us pause. Moreover, the tension between the photodynamical stellar density derived in the 'M' model and the density derived from spectroscopic analysis of this target is difficult to reconcile if the moon hypothesis holds water. The latter is of course independent of the planet/moon solution, while the former is jointly modeled. Driving the photodynamical density value is likely the fact that the impact parameter is greatly reduced in the moon solution – in this scenario the V-shaped transit is due to the transit of two objects close together rather than a single

grazing transit. For this smaller impact parameter to be physical, however, the star must be significantly smaller (denser) than the fiducial value to preserve both the transit duration and the orbital period. Radial velocity measurements of this target would thus be fruitful to the extent that they could potentially reveal a massive planet, one which is capable of hosting a satellite of comparable radius, as well as constraining the eccentricity of the planet. It is conceivable that a significant eccentricity could bring the photodynamical planet+moon solution back in line with the spectroscopic solution.

We conclude that while the moon model is formally favored over alternative models, the tight binary planet solution which lacks evidence for TDVs defies our expectations sufficiently as to induce a skeptical treatment. That said, radial velocity measurements could be worthwhile, particularly insomuch as they could constrain the eccentricity of the planet and potentially ameliorate the tensions in our modeling.

5.4 Discussion

5.4.1 Applicability of the simulated sample

The question invariably arises, to what extent can we trust the classifications made by the CNN? As detailed above, we have tried to address this concern in several ways: first, by maximizing classification accuracy during the training process; second, by utilizing many, independent CNN architectures to vote on a classification, thereby boosting accuracy through classification agreement; third, by performing an individualized false positive test on each light curve, thus testing the extent to which unique photometric variability of a given target may impact the moon classification for that star; and finally, by performing a full suite of model fits on select targets to evaluate how well the CNN has performed in identifying promising targets.

We have seen that the CNN ensemble is capable of reaching up to 98% accuracy in the training and validation stage. But do the simulated light curves accurately reflect the systems we see in the real data?

There are three ways in which the (simulated) training and validation samples can differ from the real sample. First, the stellar activity could be different. Second, the systems found in these light curves could be different. Finally, the processing of these light curves could be different. Let us handle these in turn:

Stellar sample

The "donor" light curve stars and the planet-hosting stars we vetted in this work differ in one obvious way: the latter are known to host planets, the former are not. Of course, this does not mean the donor stars do not host planets; rather, planets have not been discovered, and if the star does host a planetary system, these planets are not transiting. Because the chances of seeing a planet transit around a random star are geometrically and temporally quite low, and the estimated occurrence rate of planets is quite high, it is likely that many, indeed most, of the donor stars host non-transiting planets.

A detailed comparison of the stellar properties for light curves with and without planet discoveries is beyond the scope of this work. However, because we did not make any cuts on the KIC stars used as donor light curves (they were selected at random), the most obvious potential difference in the sample is likely the presence of some larger stars (early-type and giants) which are in general more challenging targets for exoplanet transit detection.

On the other hand, the stars selected for monitoring by *Kepler* were mostly hand-picked to be attractive targets for transit detection, and only a small fraction of the targets were O / B type stars (< 200) or giants (~ 5000) (Batalha et al., 2010). It is also reasonable to assume, to first order, that because transit probabilities are a matter of geometry (system inclinations are randomly distributed), the planet-hosting stars represent a random subset of the target KICs. Thus we expect the donor light curve sample to largely reflect the planet-hosting sample. Once again, we expect that our pipeline effectively removed peculiar astrophysical signals through detrending and screened those targets with inadequate removal of systematics. And of course, our injected planet transits were large enough to be detectable,

so in this respect they are not comparable to planets around giants, for which the primary difficulty is their extremely shallow transit depths.

Ground truth architectures

Until we begin to accumulate a number of exomoon detections in the literature, we cannot know for sure what the underlying population looks like. At present, we can only base our expectations on observations of the Solar System, and theoretical work, which, by and large, has been premised on the goal of explaining the satellites we find in our own Solar System. We should keep in mind, though, that systems unlike our Solar System may very well be out there and waiting to be found, so we should not restrict ourselves to systems we find *plausible*. Any physically possible system should be in the sample, including large satellites. However, as previously mentioned, exceptionally large satellites were deemed to be unlikely to have evaded detection in the first place, and we therefore limited simulated moons radii to being no larger than Earth-sized (already on the large end of *expected* moons, but considerably smaller than the exomoon candidate Kepler-1625b-i.). We further restricted our training sample to moons with single transit SNRs ≥ 5. It is probably reasonable to assume that smaller moons, with lower SNRs, are hiding in the data. But if a moon is so small as to be lost in the noise, we have little hope of recovering it during model selection, nor can we expect to convince a skeptical community that a detection of such a low SNR object is real.

Our results for KOI-2677.01, with its binary solution, suggest that the CNN ensemble trained on smaller moons is potentially still capable of identifying signals that are larger than those found in the training sample. However, we cannot be entirely sure that what the CNN is seeing, and what the moon models are ostensibly finding, are the same signals.

Processing comparison

The final possible difference between training and validation samples could be the processing pipeline. But as we described above, we have made every effort to process them

identically, using the same detrending algorithm and filtering approach for identifying short-duration variability and/or inadequate detrendings. The key difference between the two sets of light curves, then, could have to do with transit timings. In the simulated sample, we are able to place the planet transit precisely in the middle of the input window. For the real systems, we rely on established transit timings. To the extent that these could be inaccurate, placing some transits someplace other than the center of the segment, our samples could differ, but we expect these differences will be negligible.

5.4.2 Selecting candidates for follow-up

A difficulty with the initial predictions produced by the CNN is that a great many systems were predicted to contain a sizeable fraction of moon transits. This is challenging, because we are asking the CNN to be our primary discriminator, identifying the best possible candidates. Underlying this challenge is that we do not have a firm grasp at present on the occurrence rate of exomoons, nor can we say much about which systems are *a priori* more likely to have moons. At most, we can argue that 1) longer period planets are more attractive, by virtue of their larger Hill spheres, and 2) more massive planets are more attractive, by virtue of their ability to host larger moons, and larger moons will be easier to detect in transit. On the other hand, longer period planets will have fewer transits to work with, and larger planets are more readily identified than smaller planets when orbiting fainter stars, thus making the identification of small moons challenging for these targets.

The CNN ensemble is capable of making rapid moon predictions, processing the entire sample of light curves in a matter of minutes. Bayesian model selection through `MultiNest`, on the other hand, is computationally expensive, sometimes requiring tens of thousands of CPU hours. With $\sim$ 23% of the sample classified as showing moon transits at least 50% of the time, it is impossible to go after all these with current methods. As described above, we selected our final targets based on network agreement, false positive rate, and our own subjective analysis, including the planet's period, transit morphology, and light curve

quality. This is not entirely satisfactory, because the final sample is inherently biased by human input.

What is ultimately needed, and left for future work, is to develop new methods for faster vetting of exomoon candidates, such that we need not filter subjectively. In practice we cannot simply require the CNN ensemble to be "more selective"; because we are ignorant of the underlying population of exomoons in the data, we cannot necessarily argue that the CNN has let too many candidates through, and if we are to trust the training and validation, it has not. Instead, the CNN has produced an abundance of real exomoon candidates, some that we will be unable to confirm through modeling simply because the data quality is inadequate to support the more complex model. The results from the nine candidates examined in this work are clearly somewhat murky, but this could be due in part to our own biases for target selection. And once again, the most accurate interpretation of the results is not that moons are not present; rather, it is the more subtle conclusion that the moon hypothesis is not supported by the data.

5.4.3 Preponderance of long period candidates?

It is interesting to note a preponderance of long-period planets in the sample of the 49 most promising targets (many of which were not selected for additional vetting at this time). Is this indicative of planets at longer periods preferentially hosting moons? It is too early to conclude this based on the present study, but the possibility is no doubt intriguing. There is however another possibility, and that is, longer period planets are more suitable to identification of moon signals. This would arise for a few reasons. First, the transit durations will be longer. Not only does this mean an increased SNR, it also means that, since the light curves have not been re-scaled, a greater fraction of the light curve has a moon transit feature present. Second, short period planets will tend to be on average somewhat noisier light curves (though this is a weak effect), owing to the fact that short period planets will yield a greater number of transit events.

A key challenge for vetting long period planets is simply that we lack a large number of transits with which to work. In terms of promising targets for follow-up, how do we compare two planets, one with, say, 12 out of 20 transits showing a moon, versus one with 3 out of 5? They obviously have the same percentage of transits showing a moon feature. The shorter period planet has four times the number of moon predictions, and could therefore appear more promising; 3 could be a fluke, 12 seems less likely to be so.

We can quantify this by appealing to the binomial theorem to compute a false positive probability for these two scenarios, as a function of the probability P_{moon} that a single transit moon prediction is made by pure chance. Since this is a binary classification problem, the probability that a no-moon prediction is made will simply be $1 - P_{moon}$. Since we have trained on a balanced set, and optimized for accuracy – not a minimized false positive rate – it may be reasonable to suggest $P_{moon} = 0.5$, a flip of the coin. On the other hand, the mean moon prediction rate for the false positive test set, which consisted entirely of transits without moons, is a mere 0.8%.

We display the full range of possibilities in Figure 5.7 for the two scenarios explored above. Here the false positive probability (FPP) is the chance a system with no moons at all has gotten 3 out of 5 (or 12 out of 20) transits erroneously classified as containing a moon signal. Unsurprisingly, the FPP peaks when the random moon prediction rate matches the fraction of transits classified as having a moon. But importantly, our false positive rate drops significantly when we have more transits vetted, and approaches zero when our random moon prediction rate is low. In light of this, we can say that the longer period planets may be more attractive *a priori* from the standpoint of physical considerations, but are less attractive targets in view of their follow-up opportunities and greater chance of a false positive scenario.

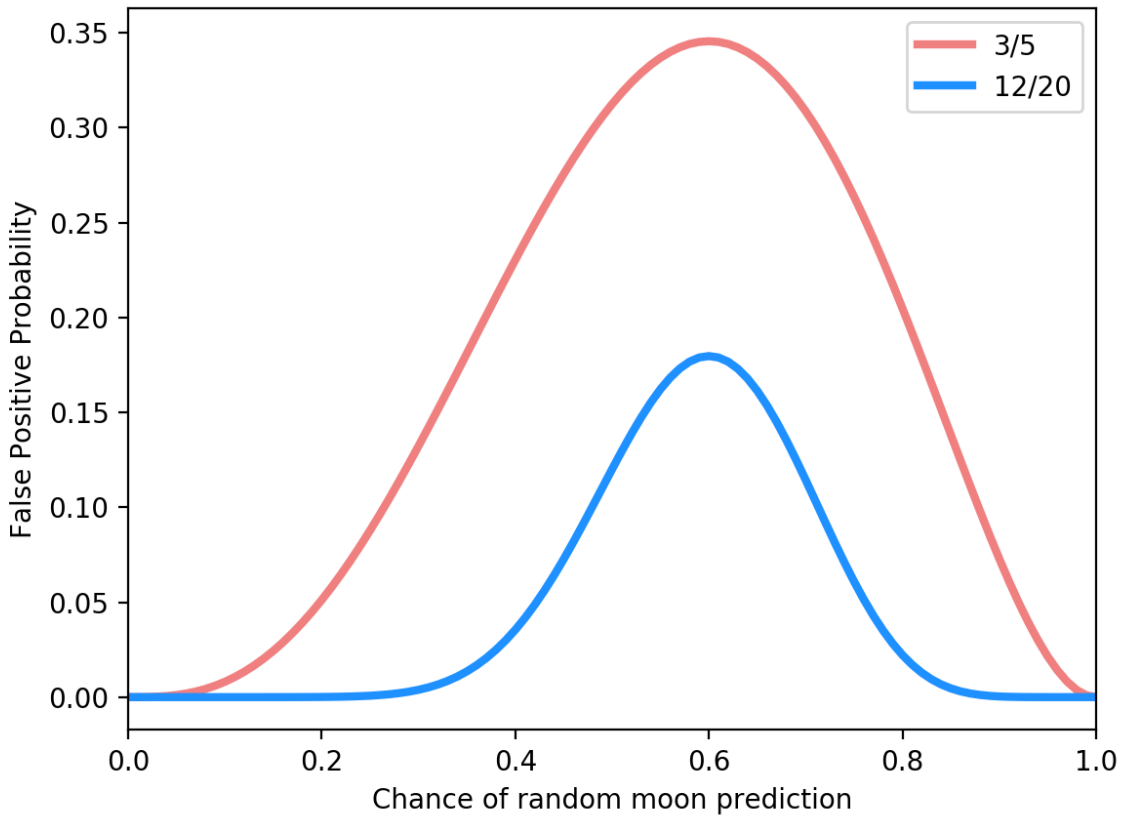

Figure 5.7: Chances of a false positive moon classification for two systems with a 60% prediction rate, shown as a function of the chance moon prediction rate.

5.4.4 Fractional transit classifications

A recurring issue in the identification of potential moon transits is that only in rare cases are 100% of the analyzed transits for a given planet classified as having a moon transit present. Frequently, for the systems we have considered here to be "promising", somewhere between 50% and 75% of the planet transits are classified as having a moon transit. Is this problematic? Yes and no.

We would presumably have the highest confidence in a system that displays a moon transit in every epoch. Not only would the chances of a false positive be reduced by virtue of such consistency across all epochs, it would also accord better with our (perhaps naive) expectation that a moon should be more-or-less co-planar with the orbital plane of the planet. By contrast, a planet that does not register a moon transit during every epoch is worrisome. First, it suggests an inclined orbit, which makes it more challenging from a follow-up perspective. If the moon doesn't transit every epoch, it becomes difficult to argue for expensive telescope resources to be devoted, since a transit observation cannot be guaranteed. On the other hand, we do not have good reason to reject the moon hypothesis based on an inclined moon solution, and indeed, we see large inclinations in the Solar System with respect to the ecliptic, both because the moon itself is inclined (Triton), or the entire system is tipped over with respect to the orbital plane (Uranus).

Beyond this, we face a difficulty with the modeling. Because we cannot put a reasonable prior on the inclination of a moon (we do not know what they should be), and an inclined moon does not constitute a more complex model, the model will not be penalized for placing a moon in such an orbit. Thus, a moon model can be flexible enough to ignore several, possibly even most, epochs that lack a moon signal, and it still accords with the data. The moon model may then be supported primarily by 1) the presence of dynamical effects, which do not have to be produced by a moon, and 2) by the occasional appearance of a moon-like dip. This is partly why we must condition our interpretation of the moon model evidence on its performance against both the 'P' and 'Z' models.

5.4.5 Future work

Because an inclined moon is not a more complex model, and is therefore not penalized during model selection, what is needed instead is improved moon priors. At present we have good reason to be agnostic about features like the size and semimajor axis distribution, occurrence rate, the relative frequencies of prograde and retrograde orbits, the presence of resonances, and inclinations. Conditioning these solely on the Solar System, or on theoretical results, could bias us against unanticipated architectures. This is to say, to some extent we must begin to find exomoons so as to refine our priors for future exomoon searches – a chicken and egg problem. Another useful line of research relates to the axial tilt of exoplanets. Predictions of an exoplanet's obliquity, or even better, direct observations of it, could be quite useful in tempering our expectations for the moon inclinations.

In addition to the models tested here, we recommend future work also include a model which enforces moon transits in every epoch – not necessarily co-planarity, but a geometry that ensures planet and moon transits. Like the 'Z' model utilized here and in our previous work, this is a useful test to perform to understand the extent to which the 'M' model "wants" to be inclined. Another test which we leave for future work is the inclusion of a model that jointly fits planet-planet interactions. It is clear in the case of Kepler-30c, for example, that planet-planet interactions are present, and this can obscure whatever dynamical effects a moon may have. But to the extent that a moon is still physically allowed in such a system, a joint modeling is required to disentangle these two effects. Of course, prior knowledge of the multiplanet system will be desirable.

Despite the somewhat ambiguous results from the present study, some of the systems we have identified with the CNN as being promising targets likely deserve additional scrutiny in future work. CNNs, and other machine learning approaches, will remain attractive options in the search for exomoons in large data sets, and will certainly be worthwhile to apply to the K2 and TESS data sets. As these data pose their own unique challenges, we leave application of the tools developed here to future work. In any case, the present results also demonstrate

the continued need for rigorous model testing and close inspection of the results.

5.5 Conclusion

In this work we have carried out a systematic search for exomoons in the *Kepler* data, utilizing an ensemble of convolutional neural networks trained on a large sample of simulated data. We are able to achieve up to 98% precision on the validation sample when the voting ensemble is in total agreement. We have applied this CNN ensemble to a total of 3223 KOIs, and vetted 9 targets deemed promising through a full Bayesian model selection process. Of these, a few (KOI-865.01, KOI-1005.01, and KOI-2677.01) emerged as intriguing targets that potentially deserve additional work, though we stop short of calling these exomoon candidates at this time. We elevate none of the targets investigated here to the status of exomoon "candidate", though how one defines this word is a matter of opinion and some debate.

This work highlights some of the ongoing challenges with identifying exomoons, and in particular, the difficulty of applying a machine learning framework to a problem with so many lingering unknowns. While the experimenter may be confident in the ability of machines to correctly classify simulated light curves through training and validation, application to real data sets poses some additional challenges. The unique features of each light curve, and underlying ignorance of everything happening in the system, can lead to predictions that require the individualized scrutiny which the machine learning approach is intended to avoid. Old-fashioned vetting, including checks on the physicality and plausibility of moon solutions, remains important. In the case of KOI-2677.01, for example, both the CNN ensemble and the model selection process find favorable evidence for something interesting going on with the system; still, when the system architecture defies expectation, or if a single model parameter solution for the system is at odds with fiducial values (in the case of KOI-2677.01, this parameter was ρ_*), there will be lingering skepticism around the moon hypothesis.

Nevertheless, CNNs are and should remain a useful framework for approaching the prob-

lem of identifying candidate exomoon signals in large data sets, and we intend to extend this work to other large surveys in the near future. Additional machine learning approaches may also be leveraged, but we will still require diligent and sober analysis of these targets.

Acknowledgements

AT has been supported through the NSF Graduate Research Fellowship (DGE-1644869). DK is supported by the Alfred P. Sloan Foundation. This paper includes data collected by the Kepler mission. Funding for the Kepler mission is provided by the NASA Science Mission directorate. This research has made use of the NASA Exoplanet Archive, which is operated by the California Institute of Technology, under contract with the National Aeronautics and Space Administration under the Exoplanet Exploration Program.

Chapter 5: Bibliography

Abadi, M., Agarwal, A., Barham, P., et al. 2016, arXiv e-prints, arXiv:1603.04467

Batalha, N. M., Borucki, W. J., Koch, D. G., et al. 2010, ApJ, 713, L109

Chen, J., & Kipping, D. 2017, ApJ, 834, 17.

Chollet, F., & others 2018, Keras: The Python Deep Learning library, ascl:1806.022

Dattilo, A., Vanderburg, A., Shallue, C. J., et al. 2019, arXiv e-prints , arXiv:1903.10507.

Domingos, R. C., Winter, O. C., & Yokoyama, T. 2006, MNRAS, 373, 1227.

Hippke, M. 2015, ApJ, 806, 51.

Holczer, T., Mazeh, T., Nachmani, G., et al. 2016, ApJS, 225, 9

Huber, D., Silva Aguirre, V., Matthews, J. M., et al. 2014, ApJS, 211, 2

Kipping, D. M. 2011, MNRAS, 416, 689.

Kipping, D. M., Hartman, J., Buchhave, L. A., et al. 2013, ApJ, 770, 101.

Kipping, D. M. 2013, MNRAS, 435, 2152

Kipping, D. M., Schmitt, A. R., Huang, X., et al. 2015, ApJ, 813, 14

Kipping, D., & Teachey, A. 2020, arXiv e-prints, arXiv:2004.04230

Kreidberg, L. 2015, PASP, 127, 1161

Mazeh, T., Nachmani, G., Holczer, T., et al. 2013, ApJS, 208, 16

Mazeh, T., Perets, H. B., McQuillan, A., et al. 2015, ApJ, 801, 3

McQuillan, A., Mazeh, T., & Aigrain, S. 2013, ApJ, 775, L11

Morton, T. D. 2015, isochrones: Stellar model grid package, ascl:1503.010

Namouni, F. 2010, ApJ, 719, L145.

Osborn, H. P., Ansdell, M., Ioannou, Y., et al. 2020, A&A, 633, A53

Panichi, F., Goździewski, K., Migaszewski, C., et al. 2018, MNRAS, 478, 2480

Pearson, K. A., Palafox, L., & Griffith, C. A. 2018, MNRAS, 474, 478

Shallue, C. J., & Vanderburg, A. 2018, AJ, 155, 94.

Sing, D. K. 2010, A&A, 510, A21.

Spalding, C., Batygin, K., & Adams, F. C. 2016, ApJ, 817, 18.

Teachey, A., Kipping, D. M., & Schmitt, A. R. 2018, AJ, 155, 36.

Teachey, A., & Kipping, D. M. 2018, Science Advances, 4, eaav1784.

Yu, L., Vanderburg, A., Huang, C., et al. 2019, arXiv e-prints , arXiv:1904.02726.

Zucker, S., & Giryes, R. 2018, AJ, 155, 147

Chapter 6: Conclusion

"I sometimes catch myself looking up at the Moon, remembering the changes of fortune in our long voyage, thinking of the thousands of people who worked to bring the three of us home. I look up at the Moon and wonder, when will we be going back, and who will that be?"
– Jim Lovell, *Apollo 13*

In this dissertation we began by examining the motivations, challenges, and methods for approaching the problem of finding and characterizing exomoons. In the following chapters, we explored in depth some of the key questions, and developed some new tools for the search. We examined the exomoon population in *Kepler*, performed two detailed studies of one particularly intriguing exomoon candidate system (Kepler-1625b), and built new tools for streamlining the exomoon search in large survey datasets.

Before we bring this dissertation to a close, it is worthwhile to consider some of the ways forward with this work, highlighting in particular some key questions that need answering, and to suggest some avenues for future research.

6.1 Open Questions

A lingering challenge in interpreting our results, both on an individual basis and when surveying the population of exoplanets, is our ignorance about the underlying architecture of exomoon systems. We have raised this point several times throughout this work, but it bears repeating one more time. Of course, we would like for the observations to guide this understanding; we are inclined to infer the population's properties from the observations themselves. But the difficulties in finding exomoons, and in particular, all the ways we know that moons can elude detection, means that our inferences about the systems we observe are intimately tied up with our assumptions.

Let us be concrete. In Chapter 2, we produced an estimate for the occurrence rate of Galilean-analog moons in the *Kepler* data. What that estimate lacked was a significant discussion about the role of moon inclination. Suppose exomoons are, on average, inclined with respect to the planet's orbital plane, and have reasonably large semimajor axes, such that moons will generally transit only 50% of the time. In the analysis of Chapter 2, this would lead us to infer an occurrence rate of only 50%, even if the underlying distribution is close to a 100% occurrence rate. Our inferences about the exomoon population are hobbled by this degeneracy. Going forward, then, we will need to be explicit about assumptions of co-planarity when discussing these rates, and examine the effect of inclination and semimajor axis in drawing conclusions about the population. What is ultimately needed, however, is a better understanding of these architectures, thereby obviating the need for building assumptions into our model.

A related challenge is the question of what to make of individual system models that place a moon on a large and inclined orbit, such that the moon only transits in a fraction of the observed transits. Do we trust such a model solution? This was of course one reason some in the community treated the Kepler-1625b moon candidate was some skepticism: the moon solution did not require it to transit every time, which may seem a bit too convenient. "Sure, we don't see evidence of the moon transit in every epoch, but that's what it *has* to do... it's inclined!" It is certainly not the most satisfying answer, but at the same time, do we have any good reason for rejecting such a conclusion? It would be a mistake to assume the moons must have co-planar orbits. What we need, then, is a better theoretical understanding of what we can expect from these moons, hopefully one tied into observables; say we're looking at a giant planet interior to 1 AU orbiting a low-metallicity K-dwarf. What should we expect from a system in *this* case? To my knowledge such theoretical studies have not been carried out along these lines yet.

Another open question regarding moon inclinations is their relationship to planet spin axes, and the intrinsic distribution of planetary axial tilts for transiting planets. Some in-

triguing work has been done along these lines in for example Millholland & Laughlin 2019, who examined the relationship between planet obliquity and orbital resonances. This naturally also ties into planetary migration. Could moons thus provide a window into (currently unobservable) planet obliquities? Or, might new techniques for determining planetary obliquities inform what we can expect from any moons present? We will want to continue a dialog between these rather different but also dynamically related investigations.

As I argued in the introduction, moons stand to provide any number of insights into the planetary systems in which they are embedded. But clearly, we want to have predictions for what should be happening in these systems as a function of a variety of system parameters, so that those predictions can be tested by observation. We have just mentioned a few questions: what does the evolution of moons look like as a function of spectral type? metallicity? We have discussed moons as a tracer of planet migrational history; are there other such observational indicators of migration that we may test against, or use to make predictions about the chances of observing a moon? And what exactly happens to an icy moon that migrates with its planet to an orbit inside of the snow line? A recurring theme here is this: theory informs observation, and observation informs theory.

6.2 What's needed

I have just highlighted a few of the open questions related to exomoons. So what do we need? Answers to the aforementioned questions would be nice. But let us get specific about the resources that would aid in the exomoon search.

6.2.1 Better priors

It is crucial that we test the exomoon hypothesis in a rigorous Bayesian framework. To that end, we desire informative priors. Specifically, we would like better priors for 1) how many detectable moons we should expect to find in a given system, 2) the distribution of moon orbit inclinations, 3) the semimajor axis distribution, and 4) the size distribution. For

the first of these, we have so far always modeled only a single moon in the system, and as we have said, adding additional moons presents a significant computational hurdle. For the other three mentioned here, the priors have been uninformative, because we cannot at present place any observational or theoretical constraints on them, apart from requiring the Moon to reside somewhere roughly between the Roche limit and the Hill radius, and being a physically plausible world (i.e. having a reasonable density). But informative priors are important, and would go a long way in (for example) quantifying the likelihood of seeing a very large moon in a wide and inclined orbit, as we saw with Kepler-1625b. With these priors also, it would be possible to make better inferences into the population of exomoons along the lines of our work in Chapter 2.

6.2.2 Modeling improvements.

At present our exomoon modeling is limited somewhat by the computational expense of our Bayesian model selection. A single moon model generated with `LUNA` has 14 parameters. Meanwhile, `MultiNest` can handle no more than ~ 20 parameters before it becomes computationally intractable, and as we pointed out, running a single moon model can cost tens of thousands of CPU hours. This requires access to a high performance computing cluster, which is of course not available to every researcher. But we know that the single moon assumption is not necessarily a good one; we have adopted it as a matter of expediency.

To what extent might a system with more than one moon fool the one-moon model into concluding that there is nothing there at all? Are there indications of multiple moons that have shown up, right under our noses, when modeling a single moon? More work is needed along these lines. In any case, we should keep the modeling of multiple moons as a goal, potentially achievable in the future, because this is a major assumption in the modeling that is clearly not well supported by our observations of the Solar System. Critically, a single moon system will induce transit timing signals that multiple moons likely will not, and we therefore might dismiss the exomoon hypothesis on the grounds that the inferred

moon solution is too low density for its inferred radius; erasing the presence of TTVs would of course lead us to a low mass solution that might be unphysical given the observed moon dips.

While we're on the subject of transit timings, we will also want to begin including joint planet and moon perturbation analysis. We know that generally speaking exomoons will induce TTVs with amplitudes of an hour or less, and recent (soon-to-be published) work by David Kipping suggests that we will find moon displaying short period TTVs, with fully 50% of moons having TTV periods between 2 and 4 epochs in a so-called "moon corridor". At the same time, planet-planet perturbations can induces TTVs with considerably larger amplitudes, and will generally have longer periods. This distinguishing characteristic of moons makes it a very attractive tool for identifying new candidate moon signals, as well as excluding TTV signals that cannot be attributed to an exomoon alone due to its unphysically large amplitude.

Even so, we must remember that the presence of planet-planet perturbations do not preclude the presence of moons. Therefore, what we may be seeing in practice is the superposition of two TTV signals, one from a moon and one from a perturbing planet. Going forward, we will want to model both of these effects simultaneously, whether or not another planet is known to be in the system. Indeed, another sticking point in the case for Kepler-1625b turned out to be whether the observed TTVs could be due to another planet instead of a moon. This is not trivial, because we are adding additional parameters to the model to be fit, approaching the limit of what `MultiNest` can support. But leaving this analysis behind could erroneously conclude a moon is not present if, say, we observe a convincing moon transit with a TTV that simply cannot be explained by a physical moon solution. A physical solution could simply be overshadowed by a planet-induced signal.

One more useful (but again, potentially unwieldy) addition to the model would be the inclusion of star spots. In the introduction we discussed the plausibility of a starspots mimicking exomoon signals; as a reminder, they are unlikely to masquerade as out-of-transit

moon signals, but could be more confounding for putative in-transit moon signals, which can have morphologies reminiscent of starspot crossings. While we therefore do not consider starspots especially worrisome from the standpoint of exomoon modeling, and we have a variety of other ways to interrogate the stellar activity, the starspot question does arise enough in the minds of other researchers that it would be worthwhile to handle within the same model testing framework. Until now, the significant challenges of tackling this task have evidently been more trouble than they're worth; how many spots do we include? Do we model their growth and decay across the observation baseline? This is already looking computationally intractable, but there could be a way forward.

Thus, there are a variety of additional models we could introduce beyond the four that have become standard in our exomoon work (the P, T, Z, and M models employed in Chapters 3-5). These additional models are 1) adding more moons, 2) modeling planet-planet TTVs, and 3) including starspots. Any one of these additions already presents significant new challenges from a computational perspective – hence their omission in the present work. For a more satisfying investigation of these systems we will want to pursue these avenues, but progress will need to be made in reducing the computational expense before such approaches can be implemented. This is easier said than done, but represents another key goal in advancing the exomoon search.

6.2.3 Observational efforts

The observational strategy for the exomoon search continues to evolve. When *Kepler* launched there was a great deal of optimism about the possibility of finding exomoons in this mountain of data. Now, more than a decade later, without any unambiguous detections, it may seem that those hopes were in vain. But the reality is, the computational expense of examining an individual planet in detail, and the relative dearth of observers working on the problem, means that even today only a small number of exoplanets have been seriously scrutinized. In short, the *Kepler* data remains a valuable, and in some ways still unmatched,

resource. We will want to continue exploring this rich dataset in the years to come. Two comparable datasets, K2 and TESS, are of course also available now. They will pose their own challenges, but they too should be squeezed for every ounce of exomoon science possible, in no small part because obtaining adequate time-domain photometry elsewhere will continue to be quite difficult.

In Chapters 3 and 4 we carried out the first ever investigation of an exomoon candidate through a targeted observation with Hubble. We have thus begun to demonstrate the great potential of targeted observations, and to develop ways of overcoming their unique challenges. Obtaining adequate time on instruments like Hubble, and eventually JWST, requires considerable additional effort in the form of proposal writing, and clearly these proposals are not always awarded (in the last five years our success rate for securing Hubble time has been 16.7%). Meanwhile, the amount of time requested routinely places these single-target proposals in the "Medium General Observer" category (> 35 orbits), thereby placing substantial pressure on the science case to justify such a large request. When it launches, there is little doubt that JWST will be heavily oversubscribed as well, and it remains to be seen whether there will be any drop in the demand for Hubble at that time. More likely, observers will simply start proposing for both telescopes.

From experience, the difficulty in having Hubble proposals awarded appears to stem from two challenging questions: 1) will Hubble actually see what we're proposing to look for? and 2) is there adequate science return for a successful observation? When it comes to the first question, we must demonstrate that a) we have good reason to believe there is a moon to be found in this system, and b) Hubble is capable of detecting it. The case is immediately hobbled if we have not already seen reasonably convincing evidence of a moon in another dataset; a blind search for a moon in a system that is merely attractive from a physical or observational point of view is unlikely to get very far with a risk-averse time allocation committee and heavy demands on the telescope's use. But this can be frustrating, because we are asking for Hubble time in no small part because of its unique capability to see things

that simply cannot be seen with other observatories. Then of course comes the perpetually confounding issues: how much out-of-transit observation do we need? Covering the entire Hill sphere can be exceptionally expensive time-wise. And how can you guarantee that the moon isn't inclined such that we won't see it transit anyway? We have no good way of guaranteeing this at present.

The second question regarding the science case is not exactly easy to answer, either – at least not within the confines of a few pages. The science case for finding and characterizing exomoons *in general* has been laid out at length (*ad nauseam?*) in this dissertation, but what is the science return on the discovery of any single object? To start with, we learn a great deal simply by trying to make the observation in the first place; we are mostly in uncharted territory here, so like any pioneering mission, successes and failures alike inform future efforts. Indeed, we learned a great deal from our experience observing Kepler-1625b with Hubble, an experience which has shaped our thinking and subsequent efforts considerably. This is however not exactly an appealing argument for a time allocation committee.

What we can argue, though, is that if the system displays any peculiar properties, this should significantly enhance our understanding of what else is possible out there. And of course, when an exomoon is finally detected unambiguously, it will immediately become among the more intriguing targets in exoplanet science – just as Kepler-1625b has elicited a great deal of interest in terms of follow-up work. But beyond this, of course, we cannot make much in the way of population inferences. We need many more such objects. But we have to discover one before we can discover many.

In the meantime, we must prepare to expand our efforts beyond those we have already imagined. We detailed at length the relative contributions that can be made by a variety of observational techniques in Chapter 1, but let us now get specific with an action plan.

We should aim to have analysis pipelines that continue to run on data coming in from a variety of instruments. *Kepler* and K2 are finished, of course, but TESS is expected to continue delivering data to us for years to come. With shifting observational strategies in

the extended mission, some of the currently very short-baseline light curves may become considerably longer in the years to come, and thus more attractive for the moon search. At the same time, it is conceivable that the enormous Evryscope dataset could be ripe for a moon search, at least around bright stars in uncrowded regions – if / when the data become available. In both cases, since these are active surveys, we will not simply want to run one search for exomoons and be done with it; rather, the search should be fast, automated, and carried out regularly.

Similarly, we should be prepared for rapid response to data coming down from the Nancy Grace Roman Space Telescope, to search for microlensing signals. By now the benefits of understanding the statistical moon population should be clear, and microlensing stands to help us enormously in this regard. And as direct imaging from the ground and in space continues to improve, we should be prepared to carry out the moon search as we described in the first chapter. In particular, we will want to utilize and further develop the planet reflex motion and center of light approaches previously discussed. In both cases, ideally, experts in these other fields would become accustomed to carrying out the moon search as a standard component of their exoplanet characterization, and exomoon researchers should become better acquainted with the tools of these other fields.

And we should continue to brainstorm ways in which targeted observations on the ground can play a role in follow-up transit observations, so that we can stop relying so heavily on space-based solutions. If coordinated observations around the world is what it will take, let us take steps towards making this less daunting. If we are asking for time on premiere telescopes that have never performed transit observations, so be it! It may be a hard sell, but we will have to make the case. Of course, more community engagement and support would go a long way. Questions of importance to the astronomical community are, by and large, questions of importance to the astronomical community.

Finally, we will need to imagine and develop new tools and tricks for identifying moons in the variety of data sources we already have on hand. This is clearly not a concrete

prescription, because I can hardly predict the thoughts and ideas I will have a year from now, let alone those of others. But there is reason to be optimistic. We need only survey the incredible ingenuity of our community in the last few decades of exoplanet science. We have made the most astonishing discoveries, through improvements in data quality, certainly, but most fundamentally, through the new ideas arriving on a daily basis from our amazing community of astronomers. I am awed by the prospect of what we will learn in the years to come.

6.3 Final thoughts

Well, we've come to the end, at long last. And yet this feels like just the beginning. There are so many questions left to answer, so many avenues to explore, and plenty of space for newcomers. My enthusiasm for these elusive worlds has steadily deepened since I undertook this work five years ago, and I can't wait to see what wonders we will discover in the years to come. And I am encouraged by the growing number of researchers in the field who are taking ever greater interest in investigating the key questions laid out in this work. Exomoons will be revealed in time, through painstaking effort and our persistent will to know. Triumphant discoveries await.