Super-Earth and Sub-Neptune Exoplanets: a First Look from the MEarth Project

Super-Earth and Sub-Neptune Exoplanets: a First Look from the MEarth Project

A dissertation presented

by

Zachory K. Berta

to

The Department of Astronomy

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Astronomy & Astrophysics

Harvard University

Cambridge, Massachusetts

April 2013

Dissertation Advisor: Professor David Charbonneau Zachory K. Berta

Super-Earth and Sub-Neptune Exoplanets: a First Look from the MEarth Project

Abstract

Exoplanets that transit nearby M dwarfs allow us to measure the sizes, masses, and atmospheric properties of distant worlds. Between 2008 and 2013, we searched for such planets with the MEarth Project, a photometric survey of the closest and smallest main-sequence stars. This thesis uses the first planet discovered with MEarth, the warm $2.7R_\oplus$ exoplanet GJ1214b, to explore the possibilities that planets transiting M dwarfs provide.

First, we perform a broad reconnaissance of the GJ1214b planetary system to refine the system's physical properties. We fit many transits to improve the planetary parameters, use starspots to measure GJ1214's rotation period (> 50 days), and search for additional transiting planets, placing strong limits on habitable-zone Neptune-sized exoplanets in the system.

We present Hubble Space Telescope observations of GJ1214b's atmosphere. We find the transmission spectrum to be flat between 1.1 and 1.7μm, ruling out at 8σ the presence of a clear hydrogen-rich envelope that had been proposed to explain GJ1214b's large radius. Additional observations will determine whether the absence of deep absorption features in GJ1214b's transmission spectrum is due to the masking influence of high altitude clouds or to the presence of a compact, hydrogen-poor atmosphere.

We describe a new algorithm to find transiting planets in light curves plagued

by stellar variability and systematic noise sources. This Method to Include Starspots and Systematics in the Marginalized Probability of a Lone Eclipse (MISS MarPLE) reliably assesses the significance of individual transit events, a necessary requirement for detecting habitable zone planets from the ground with MEarth.

We compare MEarth's achieved sensitivity to planet occurrence statistics from the NASA *Kepler* Mission, and find that MEarth's single discovery of GJ1214b is consistent with expectations. We find that warm Neptunes are rare around mid-to-late M dwarfs (< 0.15 planets/star). Capitalizing on knowledge from *Kepler*, we propose a new strategy to boost MEarth's sensitivity to smaller and cooler exoplanets, and increase the expected yield of the survey by $2.5\times$.

Contents

Acknowledgments

I thank my advisor, David Charbonneau, for teaching me how to be an astronomer. I thank him for sharing his passion with me, for taking me seriously, for allowing me to learn, for inspiring me to wonder, for gently guiding me through the toughest times, and for helping me to establish confidence in myself as a scientist. I thank him also for demonstrating that one can unravel the mysteries of the Universe by day and go home to cook dinner for his daughters in the evening. I am so grateful that I have had the privilege to work with him for these five years.

I thank Jonathan Irwin for so many reasons: for building a truly magnificent observatory, for delightful and productive scientific conversations, for thoughtful comments on all aspects of my work, for guiding my voice as a scientist, for magnificent loaves of bread, and for his support and friendship.

I thank the other members of the MEarth team – Philip Nutzman, Chris Burke, Elisabeth Newton, and Jason Dittmann – for making this thesis a possibility and for sharing in all the joys and woes of planets transiting small stars with me. I thank the other members of Dave's group, past and present, for entertaining conversations, both scientific or otherwise, over spicy string beans.

I thank David Latham for taking me to the 60" on my first observing run and for teaching me how to make every photon count. I thank Pavlos Protopapas, Dimitar Sasselov, and Alyssa Goodman, and Eric Mamajek for their helpful advice as members of my thesis committee. I thank Jacob Bean, Drake Deming, Andrew West, Peter McCullough, Alice Shapley, Neta Bahcall, David Spergel for the roles they have played

in shaping my career as an astronomer.

I thank Peg Herlihy, Jean Collins, Donna Adams, and Robb Scholten for managing the chaos with such friendly smiles. I thank Alison Doane for her always enthusiastic responses to my requests for tours of the Harvard plate stacks.

I thank all the astronomy graduate students for years of laughter, inspiration, and support. I especially thank Robert Harris, Greg Snyder, Sarah Rugheimer, Wen-fai Fong, Maggie McClean, Paul Torrey, Bob Penna, Nick Stone, Sarah Ballard, Laura Blecha, Diego Muñoz, Lauranne Lanz, Elisabeth Newton, Jason Dittmann, Anjali Tripathi, Courtney Dressing, Bekki Dawson, Ragnhild Lunnan, Katherine Rosenfeld, and Xinyi Guo. You helped me through the hardest parts of grad school.

I thank my parents, Madeline Kaczmarczyk and Jerry Berta, for making me who I am. I thank them for teaching me, with their words and with their actions, to do what I love. I thank them for instilling in me the very true knowledge that Art is Everywhere! I thank my aunt Darlene Kaczmarczyk for a lifetime of wonderful conversations. I thank my sister Amy Bengtson, and Brian and Claire, for support, laughter, and inspiration.

I thank Jessie Weidemier Thompson, for sharing her trajectory with me through space and time, for the Sun shining on an autumn afternoon, for growing together, for putting down roots, and for love.

For my family, for sharing the beauty of the Universe with me.

Chapter 1

Introduction

As an astronomer, I have had the great privilege to stand tall on mountain tops, peering outward from the edge of our homey little planet. Telescope domes at my back, I have watched the Sun glow red as it approached $\sec z = \infty$ and finally set below the horizon. From Mount Hopkins, I have seen the shadows of flowering ocotillo plants (*Fouquieria splendens*) grow long against the desert. From Las Campanas, I have stared breathlessly out over the Pacific Ocean, a vast marine pasture full of microscopic phytoplankton, as the Sun disappeared behind it.

The light of those sunsets carried an important message, albeit one my eyes were too coarse of spectrometers to see. Photosynthetic organisms – those ocotillos, the microbes filling that ocean – profoundly influence Earth's atmosphere. They pump it full of oxygen, as they have done on a globally meaningful scale for the past two billion years (Holland 2006). Those oxygen molecules, and the ozone molecules they produce, imprint absorption features on the spectrum of light transmitted through the atmosphere. As sunlight travels through Earth's atmosphere, over our heads at sunset, and back out into

space, it learns that our planet is rich in molecular oxygen. It learns that our planet is a place where photosynthesis happens, and it carries that message with it as it departs the Solar System, broadcasting into the cosmos that our planet teems with life.

If global photosynthetic life thrives on an extrasolar planet, we might be able to detect it. A large telescope, equipped with a sensitive spectrograph, could observe the spectrum of the planet's atmosphere, and we could attempt to decode the information it carries. With careful measurements and detailed modeling, we might one day be able to point to the spectrum of a planet and attribute its features to strange new plants growing on a world parsecs away. It will not be an easy endeavor, but it is an important one.

For centuries, we have imagined that stars other than our own might host planets. For decades, we have known that at least a few stars have planets orbiting them (Latham et al. 1989; Wolszczan & Frail 1992; Mayor & Queloz 1995). Results from the NASA *Kepler* satellite indicate that main-sequence stars on average host, at a minimum, about one planet per star (Fressin et al. 2013; Dressing & Charbonneau 2013). For the first time in human history, we have incontrovertible evidence that planets are common around other stars in the Milky Way.

With this knowledge secure, we are at a transitional moment for exoplanetary science. The exploratory questions can evolve from "How many are out there?" to "What are these worlds like? What are they made of? What's on them?" With the James Webb Space Telescope (JWST) and the next generation of giant segmented mirror telescopes coming online within the decade, we can begin to seriously consider our prospects for studying the atmospheres of habitable extrasolar worlds. In addition to these telescopes

not existing, several barriers stand in the way of our making these observations robustly and interpreting them accurately. The work that I present in this thesis aims to confront these barriers, to test the required techniques and probe their limitations. It is grounded in what is possible now, but strives to make substantial progress toward enabling the scientific investigation of life outside our Solar System in the not-too-distant future.

1.1 Transiting Planets as Observational Labs

Observations of an exoplanet's atmosphere encode information about the physical, chemical, and (potentially) biological processes that have shaped the planet. However, interpreting atmospheric measurements also requires context, and is easiest for those systems about which we know most. Empirical constraints on planets' bulk physical properties set the boundary conditions for theoretical modeling efforts, providing an opportunity to test hypotheses about planetary physics and chemistry. Of all exoplanetary systems, those that have a transiting geometry are the ones that provide the richest context. More knowledge can be gleaned from a transiting exoplanet than from a planet that does not transit.

- **Radii** can be estimated for transiting planets, by observing how much starlight the planet blocks during transit. To first order, a transit light curve for a planet of a given period P has three observable features: a depth, an ingress/egress duration, and a total duration (Seager & Mallén-Ornelas 2003; Winn 2010). These observables can be transformed into three physical parameters: the planet-to-star radius ratio ($k = R_p/R_\star$), the orbital inclination (i), and the mean stellar density

($\rho_\star$), as long as the eccentricity is known[1]. Combining these directly measured quantities with an external constraint on either the stellar mass $M_\star$ or the stellar radius $R_\star$, we can estimate the planetary radius R_p. A planet's radius cannot be measured if it does not transit; any possible estimate of the size of a non-transiting planet would have to be model-dependent.

- **Masses** can be estimated for transiting planets. Doppler measurements of a star's radial velocity can detect the star's wobble due to the presence of a massive planet. By itself, a radial velocity orbit measures only the mass function $M_p \sin i/(M_p + M_\star)^{2/3}$. For a transiting system, we know the inclination to be edge-on ($\sin i = 1$), and thus can estimate the true planet mass M_p (again assuming an external constraint on $M_\star$). The detection by Charbonneau et al. (2000) and Henry et al. (2000) of transits from the hot Jupiter HD209458b constituted the first direct estimates not only of R_p but also of M_p for a planet orbiting a Sun-like star. Masses of transiting planets have been estimated even without radial velocity orbits, from analyses of transit timing variations due the gravitational interactions among planets (a so-called photodynamical approach; Lissauer et al. 2011; Carter et al. 2012). For non-transiting systems, arguments of dynamical stability can be

[1]The light curve constrains the distance between the star and the planet at the time of inferior conjunction, scaled to the stellar radius ($d/R_\star$). If the planet's orbital eccentricity and argument of periastron are known, then $d/R_\star$ provides the scaled semimajor axis $a/R_\star$. Combining $a/R_\star$ with Kepler's Third Law gives

$$\rho_\star + k^3\rho_p = \frac{3\pi}{GP^2}\left(\frac{a}{R_\star}\right)^3$$

where $\rho_\star$ and ρ_p are the mean stellar and planetary densities. Provided that mass of the planet is small compared to the star ($k^3\rho_p << \rho_\star$), this is a direct constraint on the mean stellar density $\rho_\star$

used to limit the inclination parameter space (e.g. Rivera et al. 2010), but these do not constitute direct mass measurements *per se.*

- **Atmospheres** of transiting planets can be studied. One method probes a planet's transmission spectrum by gathering precise measurements of how the transit depth $[D]$ varies across different wavelengths $[\Delta D(\lambda)]$. During transit, a small fraction of stellar light will travel a slant path through the planet's upper atmosphere before reaching us. Wavelength-dependent scattering or absorption by the atmosphere will attenuate the light along that path, causing the planet to appear larger or smaller at different wavelengths (Seager & Sasselov 2000; Brown 2001; Hubbard et al. 2001). A rough estimate for the typical amplitude of features expected in $\Delta D(\lambda)$ is given by the ratio of the projected area of the atmosphere (an annulus surrounding the planet) to the projected area of the star. If the planet has an atmospheric scale height H, this ratio is $n \times 2HR_p/R_\star^2$, where n represents how many scale heights are probed in the particular wavelength range, and depends on the opacities. This factor can be $n = 1 - 10$ for strong absorption lines. Transmission spectroscopy probes the line-of-sight composition near the planet's day-night terminator, and has revealed the signatures of Rayleigh-scattering and atomic/molecular absorption in the atmospheres of hot Jupiters (see Huitson et al. 2012; Deming et al. 2013, for recent compilations). On synchronously rotating tidally locked planets, the day-night terminator marks the physical interface between permanent illuminated and permanently dark hemispheres and is the site of complicated atmospheric chemistry (Fortney et al. 2010; Burrows et al. 2010) and dynamics (Snellen et al. 2010; Miller-Ricci Kempton & Rauscher 2012).

  A complementary method probes the thermal emission from the planet. By

observing a transiting planet when it passes behind its star and is temporarily blocked, we can detect the planet's secondary eclipse and separate the flux emitted by the planet from that emitted by the much brighter star (Deming et al. 2005; Charbonneau et al. 2005). In the Rayleigh-Jeans limit, the depth of a planet's secondary eclipse is given roughly by $(R_p/R_\star)^2 \times T_p/T_\star$, where T_p and $T_\star$ are the brightness temperatures of the planet and the star. Multiwavelength secondary eclipse observations can be used to constrain the composition and pressure-temperature profile of the planet's day side (Madhusudhan & Seager 2009). Continuous infrared observations over a planet's orbit can detect the thermal phase curve from the planet (Knutson et al. 2007a), providing information on the global heat transport processes active in the planet's atmosphere (see Seager & Deming 2010, and references therein).

Taken together, these techniques mean that transiting planets offer a unique opportunity. The ability that transiting planets provide – to observationally probe the size and the mass and the atmosphere of the same planet – is a critically important one. Other tools can be used to constrain subsets of these fundamental planet characteristics for non-transiting planets, but the fact remains that joint measurements of all three are possible only for transiting systems.

1.2 Small Stars, Big Opportunities

The signal strengths of the measurements described above are small. For hot Jupiter exoplanets transiting Sun-like stars, transit depths are roughly 1%, secondary eclipses

are roughly 0.1% deep in the IR, and transmission spectra have features with amplitudes of about 0.01%. An overarching goal of exoplanetary science is to study smaller, cooler planets that are more like the Earth, planets that could potentially host life as we know it. However, smaller planets present smaller signals. For Sun-like exoplanet hosts, a shift from studying hot Jupiters to studying Earth-sized planets is accompanied by at least two orders of magnitudes decrease in the strength of each of the three above signals.

NASA's *Kepler* space telescope has met the first challenge, that of detecting Earth-sized planets transiting Sun-like stars. *Kepler* was designed to reach a photometric precision of 20 ppm over transit timescales for V=12 solar-type stars, and met those specifications once in space (Koch et al. 2010; Jenkins et al. 2010). *Kepler*'s exquisite photometry has enabled the discovery of thousands of probable small planets (Batalha et al. 2013), including the first validated discovery of an Earth-sized planet transiting a Sun-like star (Fressin et al. 2012).

Unfortunately, studying the atmospheres of the Earth-sized exoplanets *Kepler* is finding around Sun-like stars will be very difficult. The challenge is particularly pronounced for those planets that orbit in the habitable zone, the range of orbital separations in which a planet might be able to host liquid water at is surface (Kasting et al. 1993). The signals from habitable Earth-like planets' atmospheres have typical amplitudes smaller than 1 ppm, in both thermal emission and transmission, for Sun-like host stars. The precision that would be required to detect and characterize these signals is well beyond the capabilities of planned facilities.

Fortunately, the properties of the star can play a helpful role. Signal strengths do not depend simply on the absolute properties of a planet; rather, they scale with the

planet-to-star radius, mass, and temperature ratios. If a planet transits a star that is smaller, less massive, and cooler than the Sun, it becomes both easier to detect and more amenable to follow-up characterization studies. M dwarfs, as the smallest main-sequence stars, thus offer a unique opportunity. I demonstrate the M dwarf advantage here by presenting a quantitative comparison between planets transiting Sun-like stars and planets transiting M dwarfs. This comparison follows closely on that outlined by Nutzman & Charbonneau (2008). I consider the example of a planet orbiting in the habitable zone[2] of three different stars: a G2 dwarf star like the Sun ($1M_{\odot}$), an M0 dwarf ($0.58M_{\odot}$), and an M5 dwarf ($0.15M_{\odot}$). Throughout, I use stellar parameter estimates for these spectral types drawn from an online compilation maintained by Eric Mamajek[3].

- Habitable zone planets are very close-in around M dwarfs, making habitable zone planets more likely to transit. The *a priori* probability that a planet (on a circular orbit) will be geometrically aligned to transit is given by the ratio of its stellar radius to its semimajor axis ($R_{\star}/a$). Earth orbits at 215 solar radii, so this geometric probability is 0.5% for stars like the Sun. For a planet to receive the same bolometric flux that the Earth does, its semimajor axis must scale with

[2]Here, and elsewhere throughout this thesis, I take a fiducial habitable zone to correspond to the semimajor axis with which a planet in a circular orbit would receive the same bolometric flux as the Earth does from the Sun. This is a rough starting point. Updates by Kopparapu et al. (2013) to the Kasting et al. (1993) climate models indicate that Earth is very near the inner edge of the Solar System's habitable zone, and that this working habitable zone definition might be slightly too close-in for stars with redder spectral energy distributions than the Sun.

[3]This compilation was listed as "Version 2013.02.17" and was download from the following URL: `http://www.pas.rochester.edu/~emamajek/EEM_dwarf_UBVIJHK_colors_Teff.dat`

the stellar luminosity as $a \propto L_{\star}^{1/2}$. Using the definition of the stellar effective temperature $T_{\star,\rm eff}$, this relation can be transformed to $R_{\star}/a \propto T_{\star,\rm eff}^{-2}$. Scaling from the Sun ($T_{\star,\rm eff}$=5770K), the habitable zone transit probability increases to 1.1% for an M0 (3820K), and to 1.7% for an M5 (3000K). The 3.5× increase in transit probability to M0 stars means fewer stars need to be surveyed to find transiting habitable zone planets.

- The orbital periods of M dwarf's habitable zones are shorter. For a planet to receive Earth-equivalent bolometric flux, its period must scale as $P_{\rm HZ} \propto L_{\star}^{-1/4} M^{-1/2}$. With luminosities of $0.05 L_{\odot}$ and $0.003 L_{\odot}$, M0 and M5 dwarf stars yield habitable zone orbital periods of 52 and 12 days, respectively. For ground-based surveys, a planet in a 12 day orbital period is *much* easier to detect than a planet in a 365 day period.

- M dwarfs' smaller radii yield deeper transit depths, for planets of a given size. If a $2 R_{\oplus}$ planet transits a G2 dwarf, its transits are 0.03% deep. If the same planet transits an M0 ($0.58 R_{\odot}$), the transit depth is 0.1%. If it transits an M5 ($0.20 R_{\odot}$), the transit depths increase still further to 0.8%. The latter scenario is readily detectable from the ground, as I demonstrate throughout this thesis. The shallowness of the transit depths in the first two scenarios argues for the need for precise space-based photometry and was a major motivating force behind the *Kepler* mission (see Borucki & Summers 1984).

- M dwarfs' lower masses yield larger Doppler signals for habitable zone planets. If a habitable zone planet has a mass of $5 M_{\oplus}$, its orbit would impart a Doppler radial velocity signal with a semiampltiude of $K_{\star} = 0.4$ m/s on a G2 star. The precision

required to detect such a signal was recently demonstrated by the discovery of the planet α Centauri Bb (Dumusque et al. 2012), using the High-Accuracy Radial velocity Planetary Searcher (HARPS) spectrograph (Mayor et al. 2003). Even though the host star is very bright (V=1.3), the detection required a large investment of observing time and pushed limits of both the instrument's stability and the star's astrophysical jitter. The radial velocity signal of a habitable zone planet scales as $K_\star \propto L_\star^{-1/4} M^{-1/2}$. As such, a $5M_\oplus$ planet in the habitable zone of an M0 dwarf produces a stellar wobble with an semiamplitude of 1 m/s. For an M5 dwarf, the semiamplitude rises to 5 m/s. Historically, radial velocity capabilities have not been as mature for M dwarfs as for solar-type stars, but a 5 m/s orbit is possible for an M5 dwarf with existing instrumentation.

- Atmospheric studies are easier for planets that transit M dwarfs. Ambitious programs with JWST could measure transmission and emission spectra for habitable zone super-Earths, provided that they transit nearby M dwarfs (Deming et al. 2009; Kaltenegger & Traub 2009; Rauer et al. 2011; von Paris et al. 2011). For transmission spectroscopy, the signal-to-noise ratios achievable in near-IR absorption bands on an Earth-like planet increase between a G2 and M0 stellar host by factors of $1.7 - 2\times$, given the total time the planet could be observed in transit within a given year. In the same metric, the signal-to-noise ratios are higher by factors of $4 - 6\times$ for planets transiting M5 dwarfs, when compared to Sun-like stars (Kaltenegger & Traub 2009).

  Measurements of the atmospheres of Earth-like planets transiting M dwarfs may also be possible from the ground (Ehrenreich et al. 2006). A complicating factor is that an Earth-like exoplanet's atmosphere would closely resemble our own

planet's telluric absorption spectrum. To disentangle the two, observations could be gathered at high spectral resolution ($\lambda/\Delta\lambda = 100,000$), to resolve molecular features into individual lines. Earth's orbital motion and the exoplanet's systemic velocity would shift the wavelength scales of the two atmospheres with respect to each other, and allow them to be distinguished. Such studies would require massive photon collecting power feeding a high-dispersion echelle spectrograph, a task that would be well-suited to the planned G-CLEF spectrograph on the Giant Magellan Telescope (Szentgyorgyi et al. 2012).

Preceding the launch of JWST or the construction of the next generation of extremely large telescopes, atmospheric studies are already possible for non-habitable planets that transit nearby M dwarfs. The magnification of atmospheric signal strengths described above also applies to planets too hot and too big to be habitable, in some cases allowing them to be probed with the Hubble Space Telescope or existing, large ground-based telescopes. In this thesis, I present observations of the atmosphere of a $2.7R_{\oplus}$ planet transiting a nearby $0.2R_{\odot}$ M dwarf, observations that would not be possible if the planet transited a larger star.

- Finally, M dwarfs are common. The census of all known stars within 10 pc of the Sun (Henry et al. 2006)[4] contains 20 G dwarfs and 248 M dwarfs. The stellar mass function throughout the Galaxy peaks at $0.25M_{\odot}$ (Bochanski et al. 2010), and giant elliptical galaxies may host even higher fractions of M dwarfs (van Dokkum & Conroy 2010). To understand the most common planet environments in the Universe, we have to study planets around M dwarfs. More immediately, the

[4]Also, `http://www.chara.gsu.edu/RECONS/census.posted.htm`

abundance of M dwarfs in the solar neighborhood bodes well for finding nearby planets transiting such stars (in contrast to white dwarfs, which otherwise share some of the observational benefits; see Agol 2011; Loeb & Maoz 2013).

The above comparisons yield an overarching conclusion: *if we want to study the sizes, masses, and atmospheres of small, cool planets, we should look for such planets transiting small, cool stars.* In addition, these comparisons emphasize an important but occasionally overlooked point: the observability of a planet orbiting an early M dwarf (M0) is very different from one orbiting a mid-to-late M dwarf (M5). The M spectral type covers a huge range of physical properties on the main-sequence, with the spectral subtypes M0V to M8V spanning factors of 120 in luminosity, 7 in mass, and 4.5 in radius. For comparison, a G0V is only 2.6× more luminous, 1.2× more massive, and 1.3× larger than a G9V. In this thesis, I focus specifically on the smaller and cooler members of the M spectral type – the mid-to-late M dwarfs. They are the M dwarfs for which the small star advantage is most pronounced.

Within this discussion of planets in the habitable zones around M dwarfs, it bears noting that such planets would be very different from the Earth. They would experience very different dynamical, radiative, and plasma environments than the Earth does around the Sun, and aspects of these environments may make the planets surfaces unsuitable for life. Because these planets are close to their stars, strong tidal forces could lock a low-eccentricity planet into synchronous rotation (Kasting et al. 1993) or substantially alter the orbit of an eccentric planet over its habitable lifetime (Barnes et al. 2008). With a permanent day side, a tidally locked planet could be in danger of atmospheric collapse, although gases can be prevented from freezing out on the planet's night side as

long as advection can efficiently redistribute heat around the planet (Joshi 2003; Heng & Kopparla 2012). The long activity lifetimes of M dwarfs (West et al. 2008) would batter their planets with intense UV flux and particle flares, with unknown consequences for the survival and evolution of life on the surface (Segura et al. 2010). Tarter et al. (2007) reviewed these and other potential barriers to life for M dwarf habitable zone planets, concluding that the uncertainties in the processes involved are too large to make *a priori* predictions about M dwarf planets' suitability for life. Empirical constraints are needed to understand the influence of each of these processes. Because planets transiting M dwarfs are accessible to observational characterization, we have good prospects for gathering such constraints. Ultimately, observations of potentially habitable planets transiting M dwarfs could allow us to address the question – what conditions are required for life to arise and survive on a planet? Astrobiological experiments to test this fundamental question have been underway for billions of years on billions of planets in our Galaxy. If we can find habitable zone planets transiting nearby mid-to-late M dwarfs, we may eventually be able to collect the results of that ongoing experiment.

1.3 One Useful Planet

My thesis centers on the MEarth Project, a survey to find transiting planets around nearby mid-to-late M dwarfs. To provide an introduction and context for MEarth, I focus on one particular planet. I use this planet as a lens to illuminate the challenges and opportunities that MEarth faces.

GJ1214b is a warm super-Earth/sub-Neptune exoplanet that we discovered with MEarth in 2009. Planets with the same radius and equilibrium temperature as GJ1214b

are abundant in the *Kepler* candidate sample (Batalha et al. 2013), and planets with masses comparable to GJ1214b are inferred to be common from radial velocity surveys (Howard et al. 2010). Yet, GJ1214b is unique. It transits a very nearby (14.5 pc), very small ($0.2R_{\odot}$) M dwarf. The favorable geometry of this system makes it amenable to follow-up characterization measurements. Figure 1.1 demonstrates the remarkable opportunity that GJ1214b offers among the known transiting exoplanets. Thanks to the ease with which it can be observed, in the few years following its discovery, GJ1214b has become one of the most thoroughly studied exoplanets with a radius smaller than that of Neptune.

1.3.1 The Discovery of GJ1214b

The MEarth Project is a survey that was designed to find a planet like GJ1214b. The stated goal of MEarth is to find a few small planets transiting nearby mid-to-late M dwarfs, planets whose radii and masses could be measured and whose atmospheres could be studied (Nutzman & Charbonneau 2008). In the service of that goal, MEarth specifically targets the closest mid-to-late M dwarfs in the sky, so that any planets we find will be bright enough for follow-up characterization observations.

The MEarth survey has been operating since 2008 from the Fred Lawrence Whipple Observatory (FLWO) on Mt. Hopkins, AZ. MEarth employs an array of eight robotic 40cm telescopes to photometrically monitor nearby $0.1 - 0.35M_{\odot}$ stars. The survey was designed to be sensitive to transits of planets as small as $2R_{\oplus}$ and with periods out to the habitable zones of these stars ($10 - 20$ days). We are building a duplicate array of telescopes in the southern hemisphere at the Cerro Tololo Inter-American Observatory

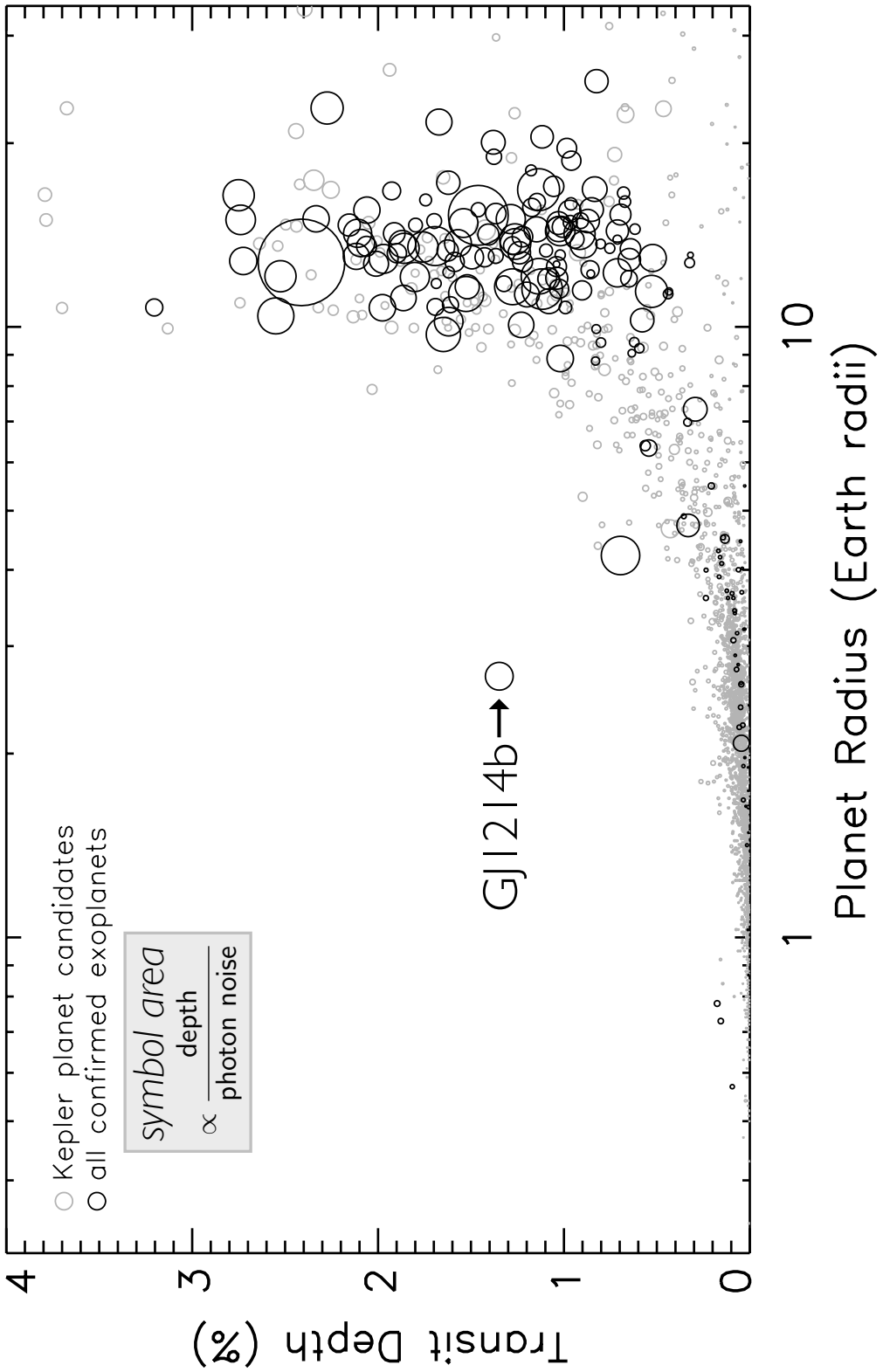

Figure 1.1: For known transiting planets from `exoplanets.org` (*black*) and *Kepler* planet candidates (*gray*; most of which are bonafide planets), a visualization of the ease with which each planet can be studied, as a function of its radius. Symbol areas are proportional to the signal-to-noise ratio with which a photon-limited telescope could detect the transit depth (estimated from 2MASS J magnitudes for confirmed exoplanets and *Kepler* magnitudes for *Kepler* candidates). Transmission and emission spectroscopy signals of exoplanet atmospheres scale with transit depth.

(CTIO) in Chile; this MEarth-South observatory should be on sky soon. Nutzman & Charbonneau (2008) defined the original strategy for MEarth, and later chapters of this thesis discuss the details of that strategy at great length. I do not reiterate those details here, except to highlight the specific niche that MEarth carves within the current landscape of planet searches:

- MEarth has more light-collecting power than small-aperture (10cm), wide-field transit surveys like HATNet (Bakos et al. 2004), SuperWASP (Pollacco et al. 2006), TrES (Alonso et al. 2004), XO (McCullough et al. 2006), or KELT (Siverd et al. 2012). This is necessary to achieve sufficient precision on our intrinsically faint M dwarf targets. The HATSouth survey (Bakos et al. 2013) and the currently under-construction Next Generation Transit Survey (NGTS; Wheatley et al. 2013) use 20cm apertures, bridging the gap on the ground between small-aperture surveys and MEarth. With a finer pixel scale (0.76"/px) than these other surveys ($10 - 20$"/px), MEarth suffers from fewer false positives due to blended eclipsing binaries than these other surveys.

- MEarth has less light-collecting power than several deep-field transit searches that also target M dwarfs. With MEarth, we want to observe the closest stars we can, within 33 pc, to maximize photons available for follow-up studies. Deep surveys like PTF/M-dwarfs (on the Palomar 48"; Law et al. 2012), the WFCAM Transit Survey (on the UKIRT 3.8m; Nefs et al. 2012) and *Kepler* (1m in space; Borucki et al. 2010) observe many M dwarfs, but those M dwarfs skew strongly toward earlier spectral types and very few of them fall within that volume limit. MEarth observes targets one-by-one in a pointed fashion, with the specific goal of targeting only the

brightest, smallest potential planet hosts (an all-sky, pencil-beam strategy).

- MEarth's targets overlap slightly with radial velocity surveys that observe low-mass stars, although those surveys have mostly focused on more massive stars than MEarth. M dwarfs have long been included in Doppler searches, for example from Mt. Wilson/Lick(Marcy & Benitz 1989), from ELODIE (Delfosse et al. 1999), from the Hobby-Eberly Telescope (Endl et al. 2003), from Keck (Butler et al. 2004; Johnson et al. 2007; Apps et al. 2010), and from HARPS (Bonfils et al. 2005b; Delfosse et al. 2013). Because these spectrographs operate in the optical, the stars in these surveys tend toward earlier spectral types, typically <M3. The HARPS M-dwarf survey, however, has started to probe later spectral types in earnest (Bonfils et al. 2013). Near-IR Doppler spectrographs now being built, such as CARMENES[5] (Quirrenbach et al. 2012) and HZPF[6] (Mahadevan et al. 2012), will soon survey many MEarth-like late M dwarfs with the precision necessary to detect super-Earth exoplanets. Planets found by Doppler surveys have the advantage that their host stars are usually bright, but the disadvantage that most of them will not transit.

- Microlensing and astrometry are also sensitive to planets around M dwarfs[7], but

[5]CARMENES = Calar Alto high-Resolution search for M dwarfs with Exo-earths with Near-infrared and Visible Echelle Spectrographs

[6]HZPF = Habitable Zone Planet Finder

[7]Microlensing planet hosts are drawn from lines of sight through the Galaxy, and thus biased by the mass function of stars toward less massive stars (roughly $0.5M_{\odot}$; Gould et al. 2010). Microlensing planets probe larger orbital separations than transiting planets can, but they are very distant and their signals non-repeating, making follow-up characterization studies difficult. Astrometry would be sensitive to nearby planets in long

are unlikely to find transiting planets whose mass, radii, and atmospheres can be studied.

At the time of this writing, MEarth telescopes have collected 2,175,901 exposures on 1885 stars spread over 993 independent nights. MEarth's total open-shutter time between spring 2008 and spring 2013 has been the equivalent of 3.05 continuous (24 hours/day) telescope-years. I portray in Figure 1.2 how these observations are distributed over the sky and over our target volume, using distance estimates from Nutzman & Charbonneau (2008). Observations are scarcer for stars that are highest during Arizona's summer monsoon (20^h <R.A. < 23^h) and denser for stars that are up in the spring, when the weather is best (10^h <R.A. < 14^h). Overall, the MEarth observations very well populate the known 0.1–0.35$M_\odot$ M dwarfs that are suspected to be within 33 pc.

If planetary transits were the only phenomenon that could cause an M dwarf to change its apparent brightness, discovering planets with MEarth would be a straightforward matter of signal detection in white, Gaussian, photon noise. However, M dwarfs brighten and dim for other reasons. Instrumental and telluric effects introduce structured noise that can mimic transits. Intrinsic stellar variability, from starspots and stellar flares, is common at 1% amplitudes in MEarth photometry. This variability can teach us about M dwarf physics (for example, Irwin et al. 2011a; Schmidt et al. 2012) but can also complicate the detection of transits. In Chapter 4 of this thesis, I present a new algorithm to robustly detect transiting planets amid these challenges: the Method

orbital periods, but astrometric planet searches for planets around M dwarfs have been fraught with difficulties. Several reported detections have been conclusively ruled to be false positives through radial velocity measurements (e.g. Anglada-Escudé et al. 2010; Bean et al. 2010b; Choi et al. 2013)

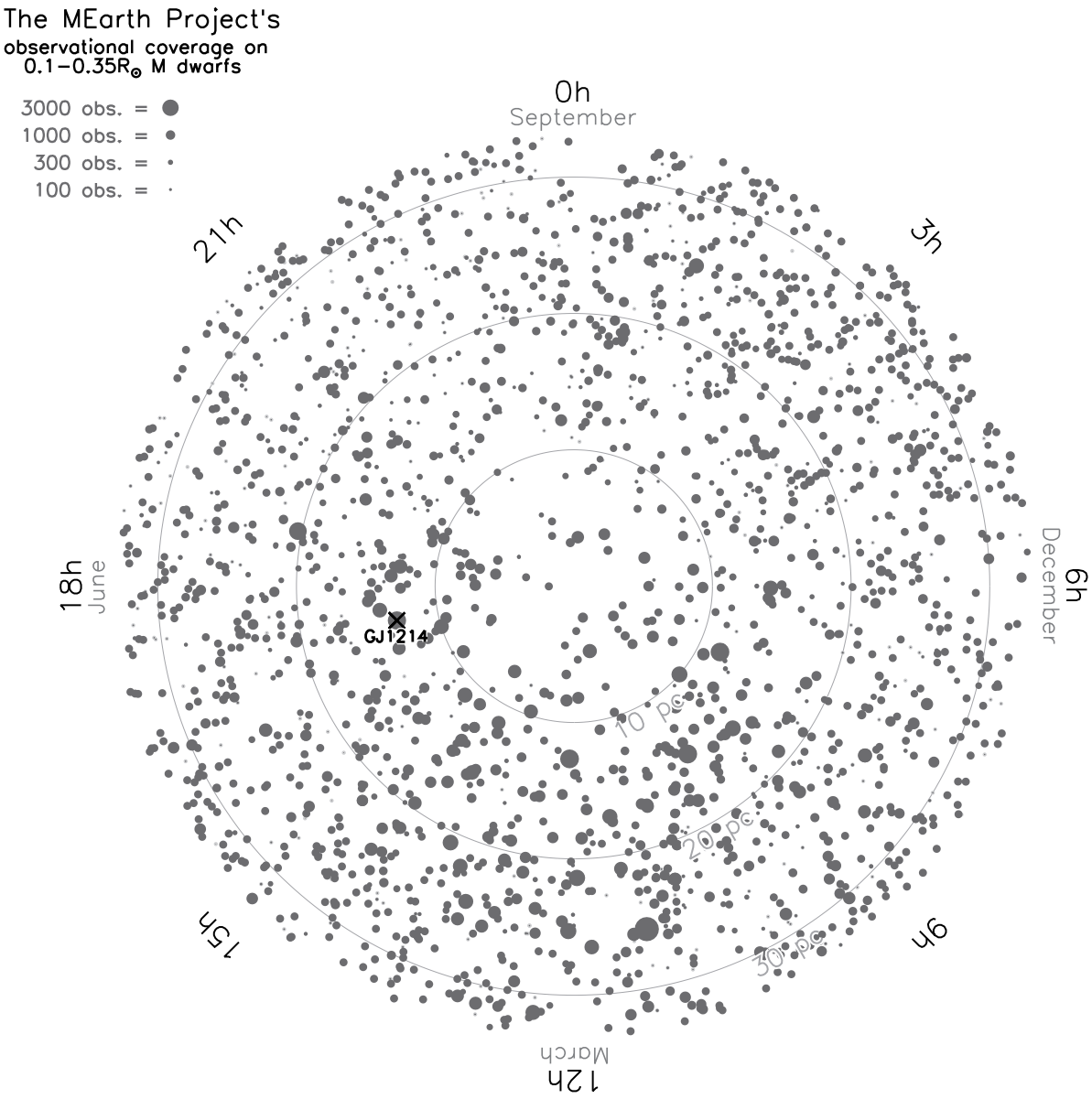

Figure 1.2: The total number of observations MEarth gathered for the M dwarfs in our sample, represented by the area of each symbol. MEarth M dwarfs are shown as a function their distance from the Sun (which sits at the center of the plot) and their Right Ascension (R.A.; clockwise from top). In this figure, the number of observations refers to the number of independent telescope pointings; in some cases, multiple exposures are taken per pointing (see Chapters 4 and 5.)

to Include Starspots and Systematics in the Marginalized Probability of a Lone Eclipse (MISS MarPLE). I developed this method over the course of my graduate career, in response to the observed characteristics of MEarth light curves.

Early, in the spring of 2009, I discovered a periodic signal in the MEarth light curve of the star GJ1214, using a very preliminary predecessor to this eventual MISS MarPLE pipeline. This signal comprised six anomalously dim measurements separated by integer multiples of a 1.58 day period. In Figure 1.3, I show these original discovery data. For visualization purposes the light curves have been re-processed by the current MISS MarPLE pipeline. The three transits shown in Figure 1.3 provided the first indication of the existence of the exoplanet GJ1214b.

1.3.2 The Confirmation of GJ1214b

Once I detected this signal, we scheduled MEarth to gather high cadence light curves at predicted times of transit[8]. With this strategy, we quickly observed a confirmation transit with a single MEarth telescope, followed by additional transits on all eight MEarth telescopes and with KeplerCam on the FLWO 48". The transits were 1.4% deep and flat-bottomed, indicating a transiting body much smaller than the size of the star. GJ1214 moves 1"/year across the sky (Lépine & Shara 2005), so by looking in archival images, we could rule out unassociated blended eclipsing binaries at GJ1214's position. By gathering reconnaissance spectra from the TRES spectrograph on the FLWO Tillinghast 60", we could disfavor physically associated blends. Already confident that

[8]I leave a lesson in this footnote for future graduate students. Julian Date (JD) and Modified Julian Date (MJD) are related by the definition MJD = JD − 2400000.5 (see McCarthy 1998). The half day matters.

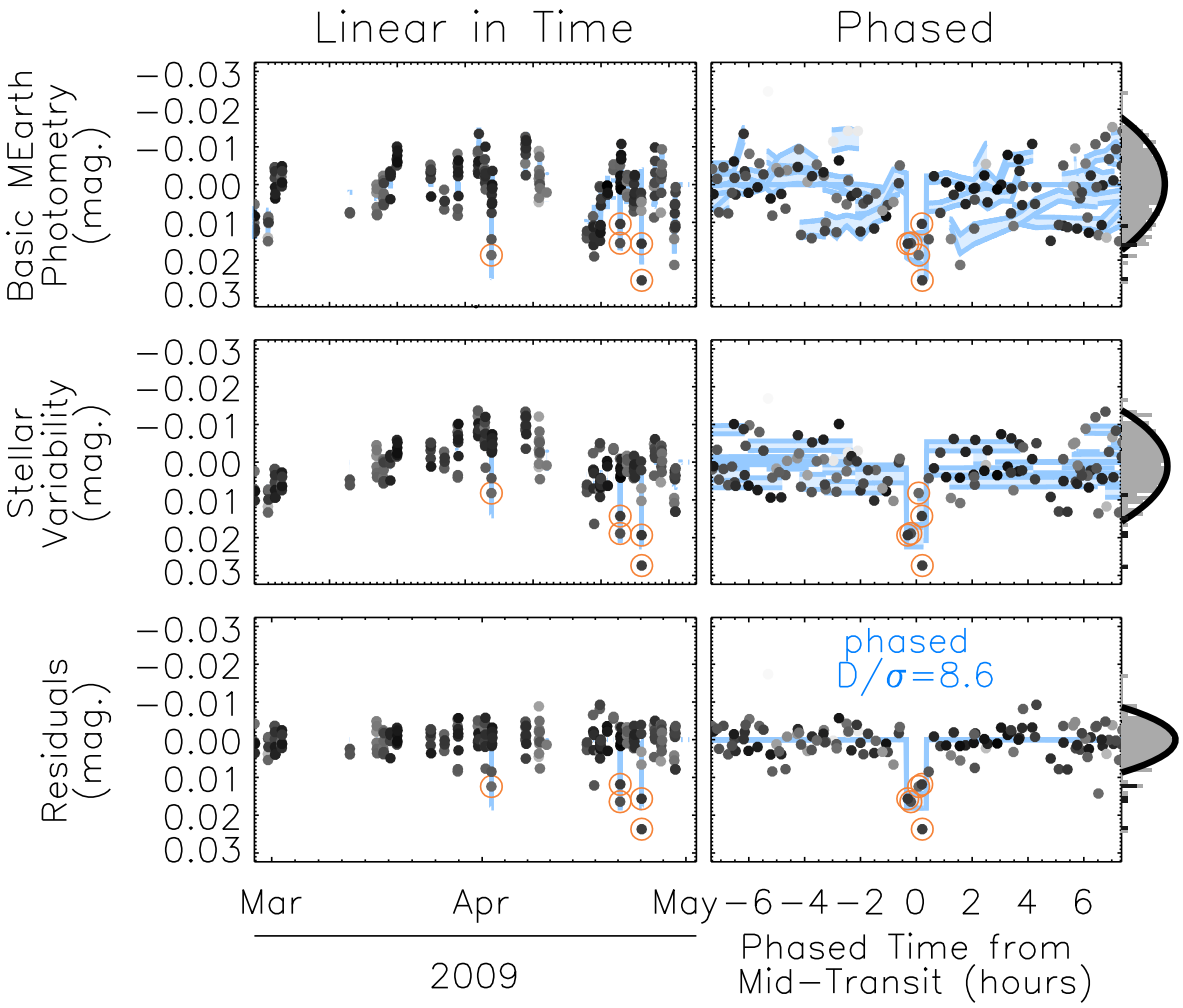

Figure 1.3: The original MEarth discovery light curve of GJ1214b, shown as a function of time (left) and phased to the planet's 1.6 day period (right). This MISS MarPLE framework searches for planets by generating a simultaneous probabilistic model (blue swaths) for instrumental/telluric systematics (dominating the top panels), stellar variability (middle panels), and planetary transits (bottom panels). Histograms of each light curve are shown at right, with a log scale (in which a Gaussian distribution appears as a parabola).

GJ12124b was a planet, the CfA team and I initiated a collaboration with astronomers at the Bohr Institute, the Geneva Observatory and the University of Grenoble to gather precise Doppler measurements from the HARPS spectrograph on the La Silla 3.6m to measure its mass.

Using these observations and estimating a mass[9] of $0.16M_{\odot}$ for GJ1214, we confirmed GJ1214b to be a planetary mass object (Charbonneau et al. 2009). We measured its radius to be $2.68 \pm 0.13R_{\oplus}$ and its mass to be $6.55 \pm 0.98M_{\oplus}$. In the year 2013, GJ1214b still is among a small sample of planets intermediate in size between Earth and Neptune for which both the mass and radius are known. Figure 1.4 shows the mass and radius of GJ1214b in the context of the other planets for which these properties are known, as well as brown dwarfs and stars that have mass and radius estimates from single- or double-lined eclipsing binaries. At the time of its discovery, GJ1214b's < 560K estimated equilibrium temperature placed it among coolest known transiting exoplanets. Since the launch of *Kepler* much cooler transiting planets have been discovered (e.g Borucki et al. 2012).

What can be learned from the confirmation experience of GJ1214b? As I demonstrate in Chapter 5 of this thesis, MEarth's next planet is likely to be smaller than GJ1214b and in a longer orbital period, in the range of $5 - 20$ days. It is instructive to imagine how the confirmation might play out for such a planet.

GJ1214b has a short (1.6 day) orbital period. As such, the original discovery signal contained multiple transits that constrained the planet's period to a well-defined family of possibilities, and confirmation transits of the planet were quick to recover. For longer

[9]See §1.3.3 for discussion of how the mass was derived.

periods, we hope to discover planets with *single* transits with MEarth (see Chapters 4 and 5 for a discussion of MEarth's "realtime trigger"). In such a case, confirmation transits will be more difficult to gather with MEarth itself for two reasons: the transits are more difficult to observe because they occur less frequently and we will have much weaker constraints on possible orbital periods, if any[10]. To confirm long-period planets, two different options are promising:

1. **Intensive Photometric Monitoring from Multiple Sites:** One way to confirm a single-transit candidate, and to measure its period, is to observe a subsequent transit. We could do so through brute force photometric monitoring, requiring as complete coverage of orbital phase as possible. Photometric monitoring has a major advantage – since MEarth identifies transits with simple equipment (40cm telescopes with CCD imagers), the requirements for follow-up facilities are similarly modest.

   The most efficient route (see Bakos et al. 2013) would be to monitor the star from telescopes at different sites (to ameliorate weather losses) and at different longitudes (to minimize daytime gaps). An ideal tool would be the growing Las Cumbres Observatory Global Telescope (LCOGT) network (Pickles et al. 2012). LCOGT will eventually consist of a worldwide array of 0.4m and 1m telescopes, capable of such longitudinally spread, continuous photometric monitoring. For now, the existing network is patchy in the northern hemisphere where overlap with MEath

[10]If the egress of the transit is resolved at decent signal-to-noise, the egress duration would roughly constrain the period range, assuming the stellar mass and radius are known. However, for realistic scenarios with MEarth, the uncertainties in these predictions will likely be large.

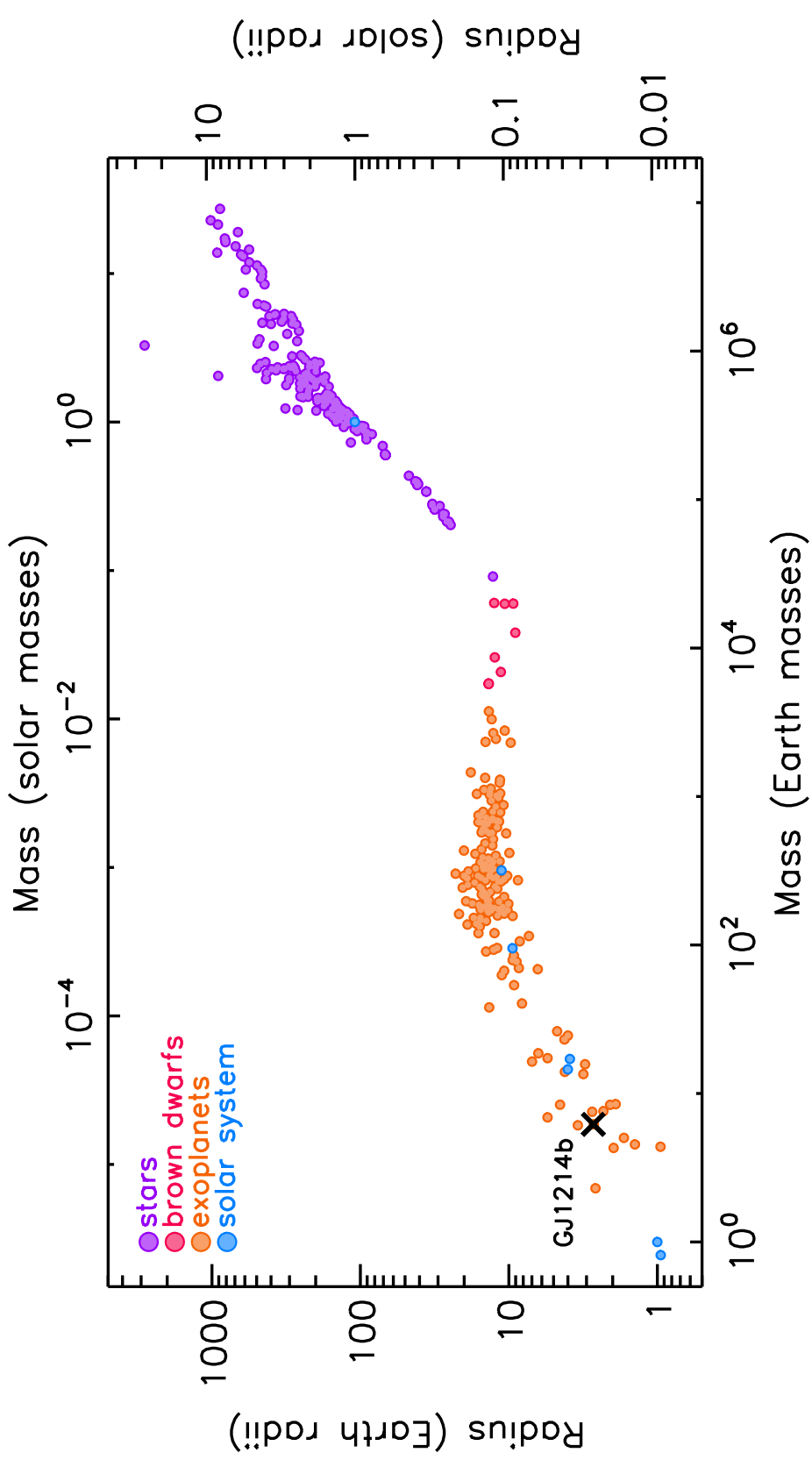

Figure 1.4: Masses and radii of stars, brown dwarfs, and planets as determined from eclipsing/transiting systems. Data were taken from `exoplanets.org` and from recent curated compilations by Torres et al. (2010), Irwin et al. (2011b), and Triaud et al. (2013b). The objects shown are meant to be representative and are not a complete sample. Most stars in this figure are in double-lined, double-eclipsing systems, for which masses and radii can be measured with minimal astrophysical assumptions. The brown dwarfs and planets transit stars and are typically single-lined systems, so their masses and radii are dependent on estimates for the primary star's mass. Optical/near-IR interferometric measurements of stellar radii (e.g Demory et al. 2009; Boyajian et al. 2012a,b) are not included.

is greatest, but a longitudinal ring of 8×1m telescopes is projected to be on-sky in the southern hemisphere (3 in Chile, 3 in South Africa, and 2 in Australia) by late 2013. This timescale is well matched to when we expect the MEarth-South clone we are building at CTIO to be complete; LCOGT could provide a valuable resource for the follow-up of candidates discovered by MEarth-South.

2. **Precise Doppler Monitoring with HARPS-N:** One disadvantage of the photometric monitoring is that proving the negative (that there is no transiting planet) would be difficult, requiring continuous orbital phase coverage over the entire (unknown) orbital period. An interesting alternative is to gather precise radial velocities to measure the orbital motion. A $5M_\oplus$ planet in the habitable zone of an M5 dwarf imparts a radial velocity wobble with 5 m/s semiamplitude. The HARPS spectrograph achieved a precision of $3 - 4$ m/s on GJ1214 in 40 minute exposures (Charbonneau et al. 2009). The predicted photon-limited uncertainties of these observations are smaller (1–2 m/s; Anglada-Escudé et al. 2013), suggesting shorter exposure times might be possible without loss of accuracy. The recently built HARPS-N spectrograph on the TNG 3.6m at La Palma has demonstrated similar performance to its southern predecessor (Cosentino et al. 2012; Desidera et al. 2013). With observations from HARPS-N, we could detect a radial velocity signal to confirm the planet's presence. We could then proceed to measure the planet's mass and period, use the orbital solution to narrow down possible transit windows, and target these windows intensively with photometry. Radial velocity noise from stellar variability may present a challenge for this strategy (Reiners et al. 2010) but could be partially mitigated with contemporaneous photometry from MEarth (Aigrain et al. 2012).

The investments of telescope time require for these follow-up efforts are not small. However, as I demonstrate in Chapter 5, the return on the investment could be huge – potentially providing confirmation of a habitable zone super-Earth whose atmosphere could be studied.

1.3.3 The Bulk Characterization of GJ1214b

How well we understand the properties of the planet GJ1214b relies crucially on how well we understand the properties of the star GJ1214. How well *do* we understand the star? Whereas the properties of main-sequence Sun-like stars are understood to exquisite precision, the problem of inferring a mass, radius, effective temperature and metallicity for an M dwarf remains a notoriously difficult one. This has posed a challenge for GJ1214, and will do so for other planet-hosting M dwarfs. I outline the tools that have been used for GJ1214b and reflect on opportunities for improvement in the years to come.

Before tackling an M dwarf like GJ1214, one question would help to set the context. How do we infer the properties of Sun-like planet hosts, and how accurately can we do so? Typically, a high-resolution, high signal-to-noise stellar spectrum is gathered. Such a spectrum contains information about conditions at the star's photosphere, including three physical parameters: the stellar effective temperature $T_{\star,\mathrm{eff}}$, the surface gravity $\log g$, and the iron abundances [Fe/H] as a tracer for overall metallicity. Torres et al. (2012) found that the combination of these spectroscopically derived parameters with a transit light curve[11] can be used to infer effective temperatures to accuracies of 1.5%,

[11]Determinations of surface gravity ($g \propto M_\star/R_\star^2$) from line profiles in stellar spectrum

surface gravities to 0.06 dex, and metallicities to 0.09 dex or better for solar-type stars. These spectroscopically determined parameters can then be mapped to stellar masses and radii, either through interior structure models (e.g. Yi et al. 2001) or through empirical relations. In their review of double-lined eclipsing binaries with precise mass and radius measurements, Torres et al. (2010) derived polynomial expressions for $M_\star$ and $R_\star$ given input estimates of $T_{\star,\rm eff}$, $\log g$, and [Fe/H]. They find these relations can predict stellar masses to 6% accuracy and stellar radii to 3% accuracy, for main-sequence and evolved stars more massive than $0.6 M_\odot$. Achieving these precisions does not require *a priori* knowledge of the distance to the star, and better precision is possible for Sun-like exoplanet hosts that are close enough for parallaxes and, in some cases, direct interferometric radius measurements. For example, von Braun et al. (2011) measured the mass, radius, and effective temperature of 55 Cancri to 1.6%, 1.1%, and 0.46% using interferometry.

Compared to Sun-like stars, M dwarfs offer additional challenges. In the optical, their spectra are too complicated for the kind of spectral synthesis techniques that lead to the precise $T_{\star,\rm eff}$, $\log g$, and [Fe/H] measurements possible in Sun-like stars, although some progress using empirical calibrations has been made in the infrared (Rojas-Ayala et al. 2012; Muirhead et al. 2012a). Even if these parameters could be well determined, the interior structure models required to map these to masses and radii have known

---

are often by themselves weakly constrained. In the case of a transiting planet, the light curve and radial velocity orbit provide an independent measurement of the stellar density ($\rho_\star \propto M_\star/R_\star^3$). Torres et al. (2012) advocate using this external constraint on $\rho_\star$ to restrict the range of possible $\log g$ values allowed when fitting stellar spectra, to narrow the degeneracies between $T_{\star,\rm eff}$ and [Fe/H] and the weak spectroscopic $\log g$ measurement. Stellar models are needed to translate between $\rho_\star$ and $\log g$; this does not pose a problem for solar-type stars, where the models perform well.

problems: they systematically underestimate the radii of M dwarfs by $5-10\%$ in eclipsing binaries where both masses and radii can be measured (López-Morales 2007; Torres 2013). Chabrier et al. (2007) have attributed this "radius inflation problem" to the fact that these eclipsing binaries were in short orbits, tidally locked, and rotating more rapidly than single stars would be. The strong magnetic fields generated by this rapid rotation would inhibit the efficiency with which energy could escape from the star. The resulting suppression of effective temperature would lead to an inflation of star's radius in order to compensate for an overall constant luminosity (set by reaction rates deep in the stars core). As such, the standard interior structure models used for single stars (Baraffe et al. 1998; Dotter et al. 2008) would underestimate the radius at a given mass. López-Morales (2007) demonstrated that the magnitude of the radius inflation over the models correlated with X-ray flux (a tracer for magnetic activity), as expected in the tidally induced rapid rotation hypothesis. Additional progress on this problem was long hampered by the small number of detached eclipsing binaries that were known at the bottom of the main-sequence.

Helping to improve this situation, we discovered three new M dwarf eclipsing binaries with MEarth. GJ3236 (Irwin et al. 2009a), is a $0.38 \pm 0.02 M_{\odot}$ and $0.28 \pm 0.02 M_{\odot}$ detached eclipsing binary in a 0.77 day orbit. As expected, the central values of the estimated radii for the components were inflated relative to models. However, I undertook a systematic investigation of the role that different starspot configurations on the components could play in our interpretation of the system and argued that the unknown location of those starspots caused large (5%) systematic uncertainties in the radii, limiting the significance of any inflation we might have seen. NLTT 41135B (Irwin et al. 2010) is a brown dwarf transiting an M5 dwarf, and in a hierarchical triple with

a visually resolved companion. I discovered this system as part of my transit search, originally mistaking the blended 2% depth for a planetary signal. LSPM J1112+7626 (Irwin et al. 2011b) is an $0.395 \pm 0.002 M_{\odot}$ and $0.275 \pm 0.001 M_{\odot}$ eclipsing binary in an extraordinarily long 41 day orbital period. To confirm the orbital period and to characterize the system, I gathered precise photometry of LSPM J1112+7626's secondary eclipses with the 40cm Clay telescope, located on the roof of the Harvard University Science Center in Cambridge, Massachusetts. Under the above rotation-activity hypothesis, we would expect a binary with such a long orbital period to more closely reflect the properties of single stars. Interestingly, this system still shows significant evidence for radius inflation, despite its long period, indicating that the radius inflation problem is not restricted to short period binaries. This finding is consistent with recent results for single M dwarfs provided by optical interferometry measurements (Boyajian et al. 2012b).

In the context of the still unsolved radius inflation problem and the general skepticism of M dwarf models, I outline the methods we used to infer the stellar parameters for GJ1214. We used a literature parallax estimate (van Altena et al. 1995) and 2MASS photometry (Skrutskie et al. 2006) to determine the absolute K-band luminosity of the star. We used an empirical luminosity-mass relation (Delfosse et al. 2000) to calculate a stellar mass. We used the constraint on the stellar density $\rho_{\star}$ from the transit light curve to estimate the stellar radius, and a relation between color and bolometric correction (Leggett et al. 2000) to estimate the star's effective temperature $T_{\star,\mathrm{eff}}$. Other authors have tried different approaches; I summarize some literature estimates of GJ1214's fundamental properties in Table 1.1. Notably, Anglada-Escudé

et al. (2013) remeasured the parallax[12] and found the star to be 10% more distant (about 1.5σ) than previously believed, and about 10% more massive. These authors also slightly revised the radial velocity orbit; coincidentally, the net effect of the changes made to the planetary properties are small. They do not significantly alter the interpretation of any of the analyses I present for GJ1214b in this thesis.

What are the limiting uncertainties in the above process? The scatter in the Delfosse et al. (2000) K-band mass-luminosity relation is 10% in $M_\star$ for stars less massive than $0.25\mathrm{M}_\odot$. As the mass and luminosity measurements that go into this relation typically have errors smaller than this, the 10% scatter likely represents true astrophysical variation and an underlying uncertainty to the relation. When determined from the light curve's constraint on the stellar density, errors on $R_\star$ are relatively insensitive to mass uncertainties, and an accuracy of roughly 5% is achievable. Because the distance to the star is known, the bolometric luminosity can be inferred directly from integrating over broadband photometry, minimizing its susceptibility to the otherwise large systematic uncertainties in M dwarf temperature scales (see discussion in Casagrande et al. 2008).

Estimates of an M dwarf's mass, radius, and luminosity will be much less precise if the distance to the star is unknown. Until recently this posed a major concern for MEarth, as literature parallaxes were only available for a small fraction of MEarth's targets. However, we are measuring parallaxes to MEarth targets using the MEarth

[12]The parallax measurement used in GJ1214b's discovery paper is almost a half century old. Our adopted absolute parallax measurement of $\pi = 77.2 \pm 5.4$ mas appeared first in the catalog of Gliese & Jahreiß (1979). Harrington & Dahn (1980) list the original source of the relative parallax measurement as coming from 44 photographic plates taken at the 61" telescope at Flagstaff between 1965 and 1971 (GJ 1214 = USNO parallax star #265 = G 139-21). Anglada-Escudé et al. (2013) recently updated this parallax to $\pi = 69.1 \pm 0.9$, using 6 observations from CAPSCam on the duPont 2.5m at Las Campanas.

Table 1.1. Inferred Fundamental Properties of the GJ1214 System

Analysis	$M_\star$ ($M_\odot$)	$R_\star$ ($R_\odot$)	$T_{\star,eff}$ (K)	$L_\star$ ($L_\odot$)
Charbonneau et al. (2009) $d = 1/\pi$ + K + Delfosse et al. (2000) $\rightarrow M_\star \bullet M_\star$ + light curve $\rho_{\star,e=0} \rightarrow R_\star \bullet R_\star$ + I-K + Leggett et al. (2000) bolometric corrections $\rightarrow T_{\star,\mathrm{eff}}$	0.157 ±0.019	0.2110 ±0.0097	3026 ±130	0.00323 ±0.00045
Kundurthy et al. (2011) UBVRIJHK[3.6][4.5μm] phot. + Hauschildt et al. (1999) model atmospheres $\rightarrow T_{\star,\mathrm{eff}} + \log g \bullet \pi$ + integral of model atm. $\rightarrow L_\star \bullet L_\star$ + Baraffe et al. (1998) $\rightarrow M_\star \bullet M_\star$ + light curve $\rho_{\star,e=0} \rightarrow R_\star$	0.153 ±0.01	0.210 ±0.005	2949 ±30	0.0028 ±0.0004
Carter et al. (2011), "Method A" Same as Charbonneau et al. (2009) but with the addition of new light curves.	0.157 ±0.012	0.210 ±0.007		
Carter et al. (2011), "Method B" $d = 1/\pi$ + JHK photometry + prior on age + Baraffe et al. (1998) $\rightarrow M_\star + R_\star$ (using *no* constraints from light curve).	0.156 ±0.006	0.179 ±0.006	3170 ±23	0.0029
Anglada-Escudé et al. (2013) New parallax measurement. $\bullet$ $d = 1/\pi$ + JHK + Delfosse et al. (2000) $\rightarrow M_\star \bullet M_\star + \rho_\star$ with RV-constrained $e \rightarrow R_\star \bullet$ JHK[W1][W2] + BT-Settl-2010 model atmospheres $\rightarrow T_{\star,eff}$	0.175 ±0.0087	0.210 ±0.011	3250 ±20	0.00398 ±0.00019
Anglada-Escudé – alternative π + BVR + Baraffe et al. (1998) $\rightarrow M_\star + T_{\star,eff}$	0.11		2880	
Anglada-Escudé – alternative π + JHK + Baraffe et al. (1998) $\rightarrow M_\star + T_{\star,eff}$	0.172		3225	

⋆Uncertainties are quoted from the original sources, where present. In some cases, the authors specifically acknowledged that they were statistical uncertainties only and did not, for example, account for systematic problems with stellar models.

telescopes themselves, in a project led by Harvard graduate student Jason Dittmann (Dittmann et al. 2012). Such measurements will relieve a major bottleneck in characterizing these stars. The uncertainties are still roughly a factor of two larger than is possible for Sun-like stars, but we may be able to shrink them in the years to come by improving our understanding of the astrophysics at the bottom of the main-sequence.

Additionally, I note that starspots play important roles in the overall characterization of planets transiting M dwarfs. Starspots can help us understand a planetary system better, by allowing us to measure the rotation period of the star and potentially constrain the age of the system (Irwin et al. 2011a). Starspots can also impinge our ability to infer the properties of a planet or its atmosphere, by biasing measurements from transit light curves. In Chapter 2 of this thesis, I explore the influence of spots in the GJ1214 system. I use them to infer a long rotation period and likely old age for the star, and I consider the impact that GJ1214's spots likely have on our overall understanding of the planet.

1.3.4 The Atmospheric Composition of GJ1214b

In Chapter 3 of this thesis, I present Hubble Space Telescope WFC3 observations of the transmission spectrum of GJ1214b's atmosphere. I designed these observations to probe a particular question: *Does GJ1214b have a hydrogen-rich outer envelope?* In this section, I briefly outline the context and the motivation for trying to address this question observationally:

- **GJ1214b sits at the poorly defined boundary between super-Earth and sub-Neptune exoplanets.** In papers describing radial velocity surveys (see Howard et al. 2010; Mayor et al. 2011), planets with minimum masses of

$m \sin i < 10 M_\oplus$ are called "super-Earths". In papers describing the *Kepler* transit survey (see Batalha et al. 2013; Fressin et al. 2013), planets with radii of $1.25 - 2 R_\oplus$ are called "super-Earths," and planets with radii of $2 - 4 R_\oplus$ are called "small Neptunes". With a mass of $6.5 M_\oplus$ and a radius of $2.7 R_\oplus$, is GJ1214b a super-Earth or is it a Neptune? These coarse distinctions obviously reflect the traits that each method probes, but the question of what to call GJ1214b is not just a question of nomenclature. It reflects a deeper curiosity about the composition of the planet.

No formal definition exists to distinguish a super-Earth from a sub-Neptune, but a useful working definition might be as follows: a sub-Neptune would be a planet that accreted and maintained a substantial H/He envelope from the primordial nebula, and a super-Earth would be a planet that lacked such an envelope, either because it never accreted one or because it lost such an envelope to atmospheric escape. Theoretical models can plausibly explain the mass and radius of GJ1214b (and its low density of 2 g/cm^3, compared to 5.5 g/cm^3 for Earth) either with or without the presence of a substantial H/He envelope (Rogers & Seager 2010b). If only measurements of its mass and radius were considered, GJ1214b would forever sit in limbo between super-Earth and sub-Neptune. For planets in this mass and radius regime, compositional degeneracies will always allow for a wide range of possible bulk compositions (Adams et al. 2008; Rogers & Seager 2010a). However, an observational determination of whether or not the outer envelope of GJ1214b is H/He-rich would break these degeneracies.

- **A method existed to measure an atmosphere's H/He content.** The year before we found GJ1214b with MEarth, Miller-Ricci et al. (2009) proposed an

observational test to measure the hydrogen content of a transiting exoplanet's outer atmosphere. The idea is as follows. The strength of features in a planet's transmission spectrum is directly proportional to the scale height of the atmosphere (see §1.1). The scale height is inversely proportional to the atmosphere's mean molecular weight μ. In scenarios such as GJ1214b's, the mean molecular weight could plausibly vary from $\mu = 2$ for an atmosphere dominated by molecular hydrogen to $\mu = 18 - 44$ for atmospheres dominated by heavier molecules like H_2O or CO_2. H/He-poor atmospheres would have features that were an order of magnitude smaller than in H/He-rich atmospheres, so the strength of transmission spectrum could serve as a proxy for the atmosphere's H/He content. Transmission spectroscopy probes only high altitudes (above roughly 100 mbar), but the composition of this outer atmospheric layer would provide a boundary condition to constrain the composition of whatever gaseous envelope lay underneath.

- **GJ1214b can serve as a useful representative for other planets at the boundary between super-Earth and sub-Neptune.** Thanks to the small radius of the host star, the transmission spectroscopy signal of a H/He-rich atmosphere around GJ1214b would have an amplitude of 0.1%. Thanks to the proximity of the system to the Sun, this precision is achievable with instruments on the Hubble Space Telescope. Figure 1.1 demonstrates that all other planets in GJ1214b's size range exhibit much shallower transit depths than it does, making these measurements more difficult. Many of these similarly-sized planets may have similar physical properties, but they transit stars that are too big and too distant for atmospheric studies to be feasible. In order to understand the properties of these other planets sitting at the boundary between super-Earths and sub-Neptunes, it

is best to study the one system where the measurements are easiest to make[13].

The results that I present in Chapter 3 indicate that the transmission spectrum of GJ1214b is flat. It does not show the large amplitude features that would be expected for a simple H/He-rich atmosphere. This result leaves us with two options regarding GJ1214b's atmosphere - either it is hydrogen-poor or its features are masked by a high-altitude cloud layer. Due to this ambiguity, we do not yet know whether GJ1214b is a super-Earth or a sub-Neptune. However, the ambiguity of the flat transmission spectrum can be resolved with additional observations (see §1.3.6). By probing GJ1214b more deeply, we may soon learn what processes shape planets in this interesting regime.

1.3.5 The Population Statistics for Planets like GJ1214b

In 2008, when we began the MEarth survey, we knew very little about the population of planets that might orbit mid-to-late M dwarfs. We knew neither what kinds of planets might be present nor how many of them there might be. Our understanding of the statistics of small planets orbiting more massive stars (FGK and early M dwarfs) has improved dramatically in the past five years, but substantial unknowns still remain at the very bottom of the main-sequence.

Kepler has provided an unprecendetedly detailed view of small planets orbiting Sun-like stars, and has revealed clear trends in the underlying distribution of planets.

[13]A few other systems stand out as promising targets in Figure 1.1. The two high signal-to-noise targets to the right of GJ1214b are the warm Neptunes Gl436b and GJ3470b, transiting early M dwarfs. The the high signal-to-noise target near $2R_\oplus$ is 55 Cancri e, transiting a bright G dwarf; this system's shallow transit depth is made up for by the star being *very* bright (V=6.0, J=4.8; Winn et al. 2011).

For planets larger than $2R_\oplus$ in orbits out to 0.25 AU, Howard et al. (2010) found that planets that are smaller and that orbit farther from their stars are more common. After carefully accounting for potential false positives, Fressin et al. (2013) demonstrated that 50% of stars have at least one planet with an orbital period < 85 days. These results broadly agree with findings from radial velocity surveys, which have shown that super-Earth and Neptune mass planets are significantly more common than gas giants (Howard et al. 2010; Mayor et al. 2011).

For early M dwarfs, the picture is qualitatively similar. Dressing & Charbonneau (2013) studied planet occurrence around M dwarfs in the *Kepler* sample, and Bonfils et al. (2013) presented results of a HARPS radial velocity survey of nearby M dwarfs. These two works find that planets around M dwarfs exhibit the same trends as around Sun-like stars, that planet occurrence increases toward smaller planets and longer periods. Although the *Kepler* studies of Sun-like stars were restricted to non-habitable planets too hot for life by the time baseline of the available data (< 1.5 years at the time of their analyses), both the *Kepler* and the HARPS studies of cooler M dwarfs were able to probe out to planets in their stars' habitable zones. Dressing & Charbonneau (2013) calculate that early M dwarfs host $0.15^{+0.13}_{-0.06}$ planets in the habitable zone, for planetary radii of 0.5-1.4$R_\oplus$. Bonfils et al. (2013) quote that $0.41^{+0.54}_{-0.13}$ of their M dwarfs host $1 < m \sin i < 10M_\oplus$ planets with periods in the habitable zone. These are the first direct estimates of the occurrence rate of habitable zone planets for any type of star.

GJ1214b fits at an interesting place among the *Kepler* results. In both the calculated planet occurrence distribution for Sun-like stars by Fressin et al. (2013) and that for early M dwarfs by Dressing & Charbonneau (2013), a planetary radius similar to GJ1214b's marks an important transition point. Both works find that planets smaller than $2R_\oplus$

are extremely common and that planets larger than $3R_\oplus$ are very rare, with planets like GJ1214b balanced on a steep slope in between. I explore the implications that such a steep rise in planet occurrence near GJ1214b's radius can have on MEarth's prospects for finding more planets in Chapter 5.

One big remaining uncertainty is exactly how planet occurrence results scale from early M dwarfs to the mid-to-late M dwarfs that are most interesting to MEarth. Neither *Kepler* nor existing radial velocity surveys strongly constrain the demographics of planets around the smallest M dwarfs. Any differences that may exist for mid-to-late M dwarfs' planet occurrence distribution would have important implications not only for our understanding of planet formation and evolution across different environments, but also for the prospects of finding more small planets amenable to spectroscopic characterization. I address this issue in further detail in Chapter 5.

1.3.6 The Open Questions on GJ1214b

We have much left to learn about the GJ1214b system. Here I outline what I view as some of the most interesting open questions about this system. I focus on questions that I believe are observationally tractable within the next 0–10 years. Answering these will help us understand GJ1214b in more detail, and they will provide insight into how to find and how to interpret other M dwarf planetary systems, and planets in general.

Are the skies cloudy or clear on GJ1214b?

In Chapter 3 of this thesis, I find that the transmission spectrum for GJ1214b is flat between 1.1 and 1.7μm. The question of clouds lingers at the end of that chapter, and

has not yet been resolved. Is the transmission spectrum flat because the atmosphere is dense and compact, or is it flat because the atmosphere hosts high-altitude clouds? The atmospheric scale height, the atmosphere's H/He content, the overall bulk composition of the planet, and whether GJ1214b is more aptly described as a super-Earth or a sub-Neptune all hinge upon the question of clouds.

Theoretically, what clouds might be present in GJ1214b's atmosphere that could affect the transmission spectrum? Condensate clouds, forming as water clouds do on Earth, could appear in GJ1214b's atmosphere for higher-temperature compounds like KCl and ZnS. Such clouds would likely form deep in the atmosphere; if they can be lofted higher in the atmosphere, they could explain the flat transmission spectrum (Miller-Ricci Kempton et al. 2012; Morley et al. 2013).

If organic tholin-like hazes (Sagan & Khare 1979) formed at high altitudes, they could mask transmission features in GJ1214b's atmosphere (Howe & Burrows 2012; Morley et al. 2013). Such hazes would form naturally as the byproducts of methane photolysis. Miller-Ricci Kempton et al. (2012) demonstrated such photolysis would occur in GJ1214b's atmosphere if GJ1214 exhibited as powerful UV emission as the active M dwarf AD Leo, but recent measurements of GJ1214's UV spectrum with the Hubble Space Telescope STIS and COS spectrograph revealed weak chromospheric emission overall and no Lyα emission (France et al. 2013). With such weak UV flux from the star, photochemical hazes now appear to be an unlikely explanation.

Since the publication of Chapter 3, new measurements have been added to the transmission spectrum. Fraine et al. (2013) published precise measurements of the 3.6 and 4.5μm transit depth that tightened the evidence for a flat spectrum. This result

was the first of several analyses planned for a very large Spitzer program led by Drake Deming that gathered 20 nearly continuous days of infrared photometry of GJ1214 (to look for habitable planets, as well as study GJ1214b's atmosphere). An observation by de Mooij et al. (2012) reported an increased transit depth in g-band, possibly indicative of Rayleigh scattering in a H/He-rich atmosphere for GJ1214b, but the large uncertainties on this measurement limit the significance of the result to at most 2σ.

Looking forward, how can we observationally distinguish whether GJ1214b's flat spectrum is caused by clouds or caused by the planet having a high-density atmosphere? Benneke & Seager (2012) have demonstrated, through extensive theoretical modeling, these two scenarios can be distinguished by improving the precision of the transmission spectrum measurements at wavelengths where they already exist. They show that both cloudy, H/He-rich atmospheres and cloud-free, H/He-poor atmospheres will show features in the transmission spectrum at the 0.01% level, and the shape of these features would allow them to be distinguished. The 0.01% amplitude of these features is smaller than the uncertainties of existing observations (and thus the spectrum currently appears flat).

I am participating as a Co-Investigator in two programs to gather new measurements with the precision required to detect the small atmospheric features predicted by Benneke & Seager (2012). In the first of these, we are duplicating and elaborating on my HST/WFC3 observations, gathering 15 additional transits in the same $1.1 - 1.7\mu$m wavelength range (P.I. = Jacob Bean). We are gathering these data in a much more efficient "spatial scanning" mode than was available for my original observations, so the overall improvement will be a factor of 25$\times$ increase in the number of photons recorded in transit. The improved precision at these wavelengths will enable us to detect and

measure the shape of water absorption features in an atmosphere that is either cloudy or H/He-poor.

We are also observing eight transits from the ground with the Gemini telescopes, using the Gemini Multiobject Spectrographs to measure the transmission spectrum at much bluer wavelengths (400-680 nm; P.I. = Laura Kreidberg). The spectrum in this wavelength range is sensitive to the Rayleigh-scattering slope, which provides an independent diagnostic of the scale height of the atmosphere (see Huitson et al. 2012; Benneke & Seager 2012). Data are already being collected in both of these programs; the results will hopefully clarify our still hazy view of GJ1214b's atmosphere.

What is GJ1214b's planet-wide energy budget?

If GJ1214b has a Bond albedo of 0 and efficiently redistributes heat from its day side to its night side, the planetary equilibrium temperature would be 575K (Anglada-Escudé et al. 2013). This zero-albedo equilibrium temperature gives a rough sense for the conditions on the planet, but we could eventually glean a more detailed view of its energy budget from observations of the system. What is the albedo? How much stellar insolation is actually absorbed by the planet? And where does that energy go? Is it quickly reradiated from the planets' day side, or does it advect to the night side?

Measuring the albedo by directly detecting reflected light from the planet will be difficult. Given a geometric albedo α, the secondary eclipse depth from reflected light would be $\alpha \times (R_p/R_\star)^2 \times (R_\star/a)^2$. If GJ1214b had an albedo of $\alpha = 1$, the corresponding reflected light eclipse depth would be 60 ppm. If the albedo is closer to 0.1, as seems to be the case for hot Jupiters (see Cowan & Agol 2011), the depth would drop to 6 ppm.

Signals this small are detectable in *Kepler* photometry (Kipping & Spiegel 2011); could they be detected from the ground for GJ1214b? In Chapter 2 of this thesis, I presented light curves of GJ1214 from ESO's Very Large Telescope (VLT) with an RMS scatter of 350 ppm and a cadence of 72 seconds. These light curves exhibited time-correlated noise, but if the systematics could be controlled on timescales longer than the 50 minute eclipse duration, the precision could bin down with $\sqrt{N}$ to the level of 60 ppm per eclipse[14]. If GJ1214b had an albedo of 1, nine eclipse observations would be required to achieve a 3σ detection of the eclipse. If the albedo is 0.1, nine hundred eclipses would be required. Meaningful direct constraints on α will be difficult, but not necessarily impossible, to gather in the near future.

Once JWST launches, the prospects will be bright for detecting thermal emission from the planet during secondary eclipse. The secondary eclipse depth from thermal emission is roughly $(R_p/R_\star)^2 \times B(\lambda, T_p)/B(\lambda, T_\star)$, where the second factor represents the ratio of blackbody emission from the planet and the star. At long wavelengths ($> 10\mu$m), the eclipse depths are 2×10^{-3} – such depths could be measured to high precision with JWST (Deming et al. 2009). Such observations could be used to determine the strength of molecular absorption features in the planet's thermal emission spectrum, and would constrain how much stellar flux the planet actually absorbs and reradiates.

Before JWST, however, the prospects are somewhat bleaker at the wavelengths accessible to current facilities. At Warm Spitzer's remaining 3.6 and 4.5μm channels,

[14]This is still above the predicted atmospheric scintillation limit for an 8.5m telescope on 50 minute timescales (Young 1967; Dravins et al. 1998). How well other systematic noise sources may behave over this timescale, and across multiple eclipses observed with the same instrument, remains to be seen.

the eclipse depths are 10^{-5}. In spite of 20 days of observations of the GJ1214 system with Spitzer (Fraine et al. 2013), eclipses this small are still unlikely to be detected (although the analysis to search these data for the secondary eclipse signal has not yet been published).

With JWST we will also be able to measure the thermal phase curve, as the planet's day and night sides rotate in and out of view throughout the orbit. Thermal phase curves have provided maps of the brightness distribution on hot Jupiters and probed how efficiently winds can transport heat on global scales (see, for example, Knutson et al. 2007a, 2009a,b, 2012). Menou (2012) proposed that estimates of the ratio of radiative to advective timescales in a phase curve for GJ1214b could provide an independent constraint on the bulk composition of the planet's atmosphere.

Like GJ1214b, planets in the habitable zones of M dwarfs may be tidally locked (Kasting et al. 1993). For a tidally locked planet, the efficiency of advective heat transport is of crucial importance for the planet's actual ability to maintain an atmosphere capable of supporting life (Joshi et al. 1997; Joshi 2003). Studies of GJ1214b with JWST will provide empirical constraints to inform global climate models of planets in this unique environmental regime (see Lewis et al. 2010), and improve our understanding of the challenges facing life on a planet in the habitable zone of an M dwarf.

What is GJ1214b's orbital eccentricity?

In Charbonneau et al. (2009), we placed an upper limit of $e < 0.27$ on GJ1214b's eccentricity. What is the actual eccentricity? Is it really close to zero, or is it closer to 0.1? Tighter constraints on the eccentricity would help us understand the history of the

planet and the properties of the star.

GJ1214b's orbit is close to its star ($a = 15R_\star$). Tidal effects have certainly played a role in the planet's evolution, one manifestation being that tides may have circularized initially eccentric orbits. Carter et al. (2011) state that plausible guesses for the efficiency of this process (parameterized by the tidal quality factor Q_p) yield characteristic circularization timescales that span orders of magnitude, from 10 Myr to 10 Gyr. If we can detect a measurable eccentricity, it would inform studies of the dynamical history of the system, and potentially either rule out some possible values of Q_p or point toward other perturbing bodies in the system, such as exterior planets. As tides play a strong role in the likely habitability of planets orbiting M dwarfs (Tarter et al. 2007; Barnes et al. 2008, 2009), every insight that we can make into still poorly understood tidal evolution processes will be helpful.

Whatever the eccentricity, a tighter constraint on the value would help inform the M dwarf radius inflation problem (see §1.3.3), an outstanding question in stellar astrophysics. Transiting planets are single-lined eclipsing binaries[15]. Transit light curves alone directly constrain physical quantity

$$\rho_\star \times \left(\frac{1 + e \sin \omega}{\sqrt{1 - e^2}} \right)^3$$

where $\rho_\star$ is the stellar density, e is the eccentricity, and ω is the argument of periastron (Winn 2010; Kipping 2010; Carter et al. 2011). If we can measure the eccentricity for GJ1214b's orbit, we can directly determine the stellar density of GJ1214.

[15]HD209458b is a notable exception. Snellen et al. (2010) detected the planets orbital motion at high resolution in CO absorption lines, effectively making the system double-lined.

For the sake of comparing to models, we would prefer to have direct measurements of $M_\star$ and $R_\star$ independently. However, since only four known stars less massive than $0.2M_\odot$ that have measured radii (see Boyajian et al. 2012b), every contribution in GJ1214's mass range is valuable. As we discover and characterize more planets transiting late M dwarfs, we will create more empirical tests of the physical properties of single stars at the bottom of the main-sequence, provided we measure their light curves and eccentricities to sufficient precision.

More radial velocity observations would be needed to tighten the constraints on eccentricity. The necessary data to measure the eccentricity have probably already been gathered. The HARPS team continued monitoring GJ1214b after measuring the initial radial velocity orbit for the discovery paper, but no analysis of these data has yet appeared in the literature.

Is GJ1214b's orbit aligned with its star's spin?

The orbital axes of Solar System planets are aligned to within seven degrees of the Sun's rotation axis. The Sun's stellar obliquity – the angle between its spin axis and the axis of the ecliptic – is small. Hot Jupiters are not always so well-aligned, with many showing high stellar obliquities (Winn et al. 2010). Hot Jupiters did not form in their present-day orbits; they must have migrated inwards from farther out. The presence of the misaligned hot Jupiters has been attributed to some or all of these planets having migrated through some mechanism that jostled the planets' inclinations as they moved inward (Fabrycky & Winn 2009; Albrecht et al. 2012).

With GJ1214b's low density and implied high volatile content, it probably formed

farther out in the protoplanetary disk than its current orbit. A measurement of whether its present orbit is aligned with its stellar spin may be able to teach us about the process by which it migrated inward.

The traditional way to measure stellar obliquities is through the Rossiter-McLaughlin (RM) effect (Rossiter 1924; McLaughlin 1924), the anomalous redshift or blueshift when a planet blocks the approaching or receding half of the stellar disk during transit. Unfortunately, RM measurements are unfeasible for the GJ1214b system. GJ1214's slow rotation period (see Chapter 2) and small radius give it very small $v \sin i$, in turn suppressing the predicted amplitude of the RM signal to below feasibility with current spectrographs, especially for such a faint star.

Fortunately, another technique exists. Sanchis-Ojeda et al. (2011) and Nutzman et al. (2011b) have used starspot occultations in transit light curves to cosntrain the stellar obliquity. The idea is that a planet that is well-aligned with its star's spin would transit the same spot at multiple rotational phases of the star, whereas a misaligned planet could occult the same spot only at a single location along the transit chord. Starspot occultations have been seen regularly in precise transit light curves of GJ1214b (Carter et al. 2011; Kundurthy et al. 2011; Bean et al. 2011, and Chapter 2 of this thesis). With precise transits populating enough of GJ1214's rotational phase, a starspot-based obliquity measurement may be possible for this system, enabling additional constraints on formation and migration scenarios.

Knowing whether planets like GJ1214b are well-aligned matters for another reason: our goal of devising an optimal strategy for finding more planets transiting nearby M dwarfs. With the indications from *Kepler* that most stars host planets, the challenge

of finding transiting planets becomes a challenge simply of finding which stars' planets are aligned to transit. If stellar obliquities are usually low for M dwarfs with small planets, then an efficient strategy for finding transiting planets would be to identify stars that appear to be edge-on (for example, through the combination of their $v \sin i$, their photometric rotation periods as already measured by MEarth, and their estimated stellar radii; see Beatty & Seager 2010). If multiple planets are mutually aligned with each other and with their stellar spin (as recently demonstrated in the Kepler-30 system; Sanchis-Ojeda et al. 2012), then this strategy would be even more profitable.

Is there a GJ1214c?

In the *Kepler* sample, 20% of planet host stars have more than one transiting planet (Batalha et al. 2013). Of the planets *Kepler* has found transiting M dwarfs, many are in multiples (for example, Muirhead et al. 2012b; Swift et al. 2013). Radial velocity surveys have found a similarly large fraction of multiple planet systems (e.g Bouchy et al. 2009; Wright et al. 2009). Does GJ1214b reside in a multiple system?

One way to find these planets would be to detect their transits. Fang & Margot (2012) have determined that the *Kepler* planet candidates exhibit a high degree of orbital coplanarity. If other planets are on coplanar orbits with GJ1214b, they may be likely to transit as well (Gillon et al. 2011). In Chapter 2 of this thesis, I search for additional transiting planets around GJ1214 using two years of MEarth photometry. I place limits on the presence of $4R_\oplus$ planets out to the system's habitable zone. The nearly continuous 20 day Spitzer monitoring program of GJ1214 is much more sensitive to the presence of planets than MEarth was, and could have found $1R_\oplus$ planets in the

habitable zone. The results of the search for transits in these data has not yet been published (see Fraine et al. 2013).

Non-transiting planets could be detected with radial velocity monitoring. The continued observations of GJ1214 by the HARPS team will be sensitive to planets at a wide range of orbital periods. GJ1214 will probably also be the target of additional observations with HARPS-N, as well as the CARMENES and HZPF spectrographs when they come online. Close Doppler scrutiny of this star will map out the population of planets that may or may not be present in the system. Measuring this population will be crucial to developing a holistic understanding of the conditions under which GJ1214b formed and evolved.

How has atmospheric escape shaped GJ1214b throughout its history?

Today, GJ1214 appears to be a quiet, old M dwarf. Optical indicators point toward weak magnetic activity on the modern-day GJ1214, and no chromospheric Lyα emission was detected in a 7000 second exposure with HST/STIS (France et al. 2013). In its youth, however, GJ1214 would have been more rapidly rotating, would have exhibited stronger magnetic fields, and would have been much more chromospherically active (see Mamajek & Hillenbrand 2008). The young GJ1214 would have roasted the young planet GJ1214b with intense UV radiation (Hawley et al. 2003; Walkowicz et al. 2008) and particle flares (Aarnio et al. 2012). The localized deposition of this energy in the upper atmosphere would have substantially enhanced rates of atmospheric escape (Murray-Clay et al. 2009). Activity lifetimes for late M dwarfs like GJ1214 can last for gigayears (West et al. 2008); such a prolonged barrage must have played an important role in the overall

evolution of GJ1214's outer envelope (Rogers & Seager 2010b; Rogers et al. 2011).

Atmospheric escape has been observed in progress for hot Jupiters transiting mildly to moderately active Sun-like stars, through the detection of Lyα absorption by extended hydrogen exospheres (Vidal-Madjar et al. 2003; Lecavelier Des Etangs et al. 2010; Ehrenreich et al. 2012). The opportunity to the observe the influence of activity on GJ1214b's atmosphere has passed, but we might be able to probe atmospheric escape on other super-Earths and sub-Neptunes transiting younger, more active M dwarfs. Shkolnik et al. (2009, 2012) compiled a sample of nearby M dwarfs that are bright in UV and X-rays, with the goal of identifying young systems favorable for direct imaging campaigns (e.g. Bowler et al. 2012). MEarth, or another dedicated search, could potentially find planets transiting these active stars; many of them already appear in the MEarth target list. Obtaining radial velocity orbits for such planets would be difficult (van Eyken et al. 2012), but might be possible in the infrared where the effects of stellar activity may be lower (Reiners et al. 2010; Barnes et al. 2011).

Transit surveys like MEarth should not ignore active stars – planets transiting them offer unique opportunities for testing new planetary physics. The issue is directly applicable to the question of habitability around M dwarfs, because whether an Earth or super-Earth planet can retain a significant atmosphere against the harsh environment of a young M dwarf is a hotly debated topic (see Tarter et al. 2007; Scalo et al. 2007). Empirical measurements at younger ages, of any planet, would inform the debate.

1.4 Where are the other transiting planets?

I conclude this introduction to my thesis with the biggest unanswered question: how do we find more planets transiting nearby mid-to-late M dwarfs? GJ1214b is a useful planet, but it is just one planet. If we really want to understand the processes that shape super-Earth and sub-Neptune planets, we will need to have a population of such objects for which direct measurements of the planet's mass, radius, and atmosphere are possible. What magnitude of effort will be required to build up this population?

A productive way of addressing this question is to consider what we have accomplished so far with MEarth. By characterizing the completeness with which MEarth has surveyed its stars for planets, I provide a reference point for what still remains to be done. I do so in the following way. I define a quantity $T_{50\%}$ – the half-completeness temperature – and calculate its value for all of the stars in the MEarth target sample. For planets of a given radius, I take a star's $T_{50\%}$ to be the planetary equilibrium temperature at which MEarth has surveyed 50% of the orbital phase with sufficient sensitivity to detect the planet. If planets hotter than $T_{50\%}$ transited the star, they would likely have been discovered, whereas planets cooler than $T_{50\%}$ could easily have been missed. In this section, I report $T_{50\%}$ based on a 7.5σ detection threshold in a phase-folded transit search (see Chapters 4 and 5). Data taken up to summer 2012 are included.

Figure 1.5 visualizes this half-completeness temperature $T_{50\%}$ for planets larger than $4.0R_\oplus$ for all stars in the MEarth sample. For Neptune-sized planets, a large fraction of MEarth's observed targets out to 25 pc have values of $T_{50\%}$ approaching 300K. If our targets hosted large numbers of transiting Neptune-sized planets, we should have found

them already. Importantly, if another survey were to look at these stars and was sensitive only to Neptune-sized planets, that survey would be battling diminishing returns.

One might interpret this statement with concern, worried that MEarth would run out of nearby targets before finding planets. This may be true for Neptune-sized planets, but it is certainly not true for smaller planets. Figure 1.6 shows MEarth's completeness to $2.7R_\oplus$ planets (the same size as GJ1214b), using the same $T_{50\%}$ diagnostic. At this smaller planetary radius, the number of stars decreases for which we can achieve sensitivity to planets in cool orbits. For many stars, we are 50% complete only out to equilibrium temperatures of 500-600K.

Finally, Figure 1.7 gives MEarth's completeness to $2.0R_\oplus$ planets. MEarth has observed only a few stars with sufficient intensity to address the presence or absence of $2R_\oplus$ transiting planets. We cannot rule out the presence of $2R_\oplus$ planets even for some of the most nearby stars we have surveyed. Below planetary radii of $2R_\oplus$, the local M dwarf population is still essentially uncharted territory.

In the triptych of Figures 1.5 – 1.7, MEarth's sensitivity falls off steeply. Importantly, this falloff in sensitivity is complemented from *Kepler* by a striking increase in the rate of planet occurrence over the same planet radius range. The uncharted territory below $2R_\oplus$ is teeming with planets, if the results from *Kepler* can be safely extrapolated to MEarth's mid-to-late M dwarfs. In Chapter 5, I outline a new strategy for MEarth that improves our sensitivity to smaller planets, in order to take advantage of this observation. This new strategy will help us wander a little deeper into this new planet-rich space, and could be accompanied by big gains for our expected planet yield.

The task of exploring this uncharted territory is going to be arduous. With this

revised strategy and with the construction of MEarth-South, we may be able to glean a few more planets out of the sky. But ultimately, MEarth will not complete the survey for transiting planets around M dwarfs in the local neighborhood by itself. Surveys beyond MEarth, with more telescopes or larger apertures, will be required to stare deeply enough at these small stars to detect the shallow transits of truly Earth-sized planets. These future surveys, descendants of MEarth, will be able to discover habitable planets whose atmospheres can be studied. The planets are out there; all we have to do is look closely enough to find them.

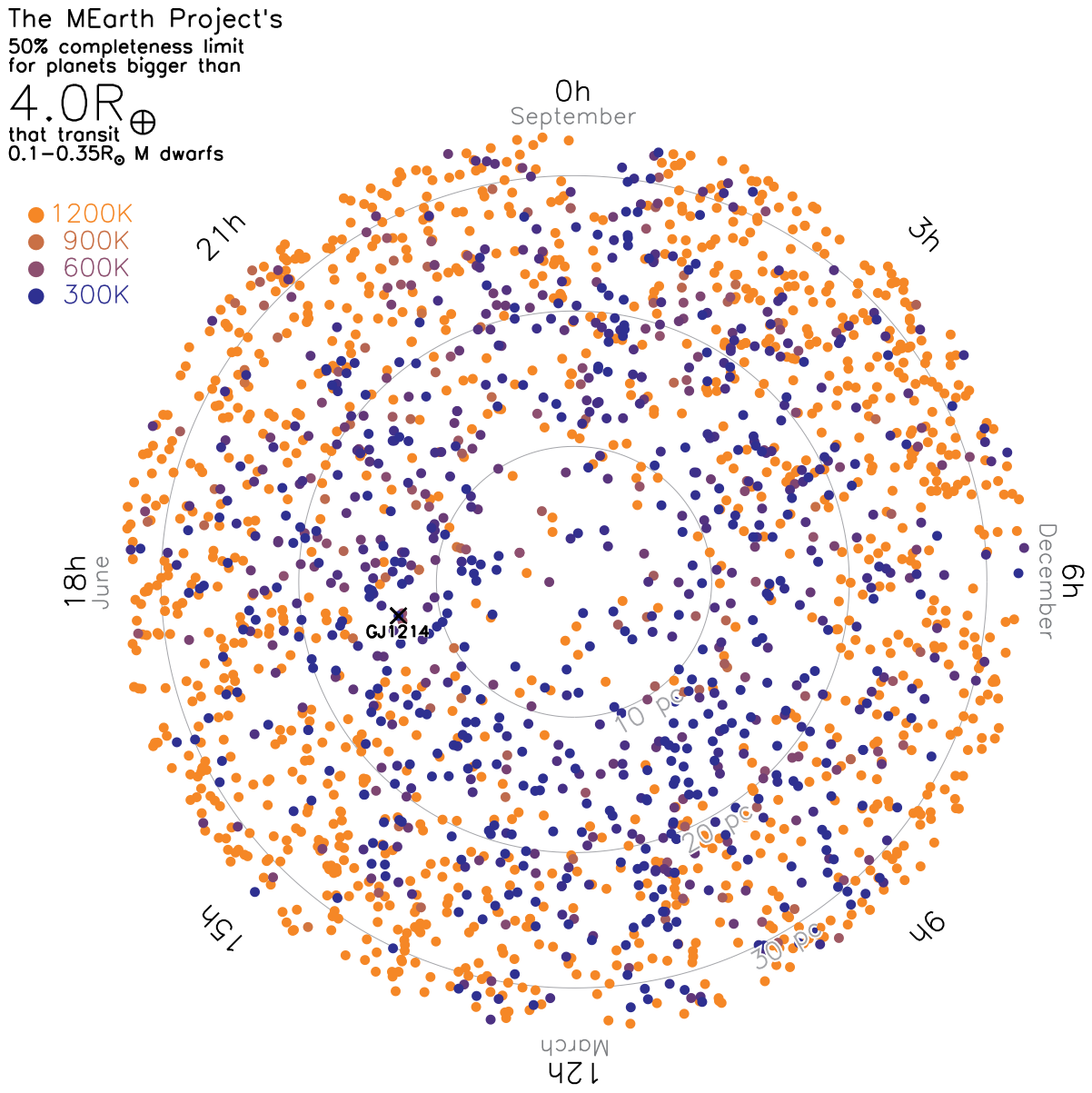

Figure 1.5: MEarth's completeness to $\geq 4R_{\oplus}$ transiting planets, the size of Neptune or larger. Each MEarth M dwarf is shown as a function of its distance from the Sun and its Right Ascension (R.A.). The color of each symbol represents $T_{50\%}$, the planetary equilibrium temperature (assuming zero albedo and efficient heat redistribution) to which MEarth has surveyed that star at 50% completeness. Planets hotter than $T_{50\%}$ should have been detected; planets cooler than $T_{50\%}$ are beyond the sensitivity we have achieved so far.

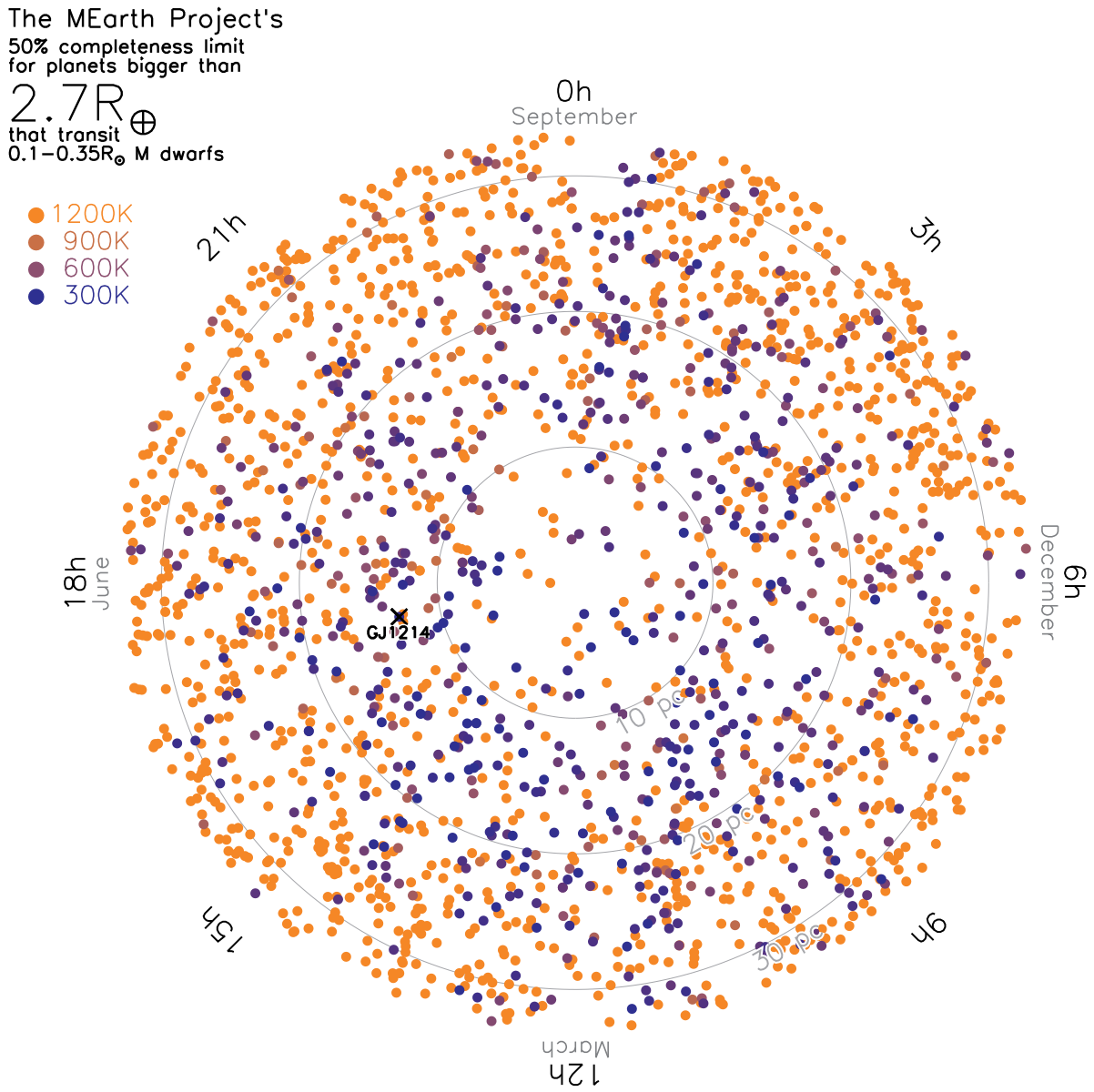

Figure 1.6: Same as Figure 1.5, but showing MEarth's completeness to $\geq 2.7R_{\oplus}$ transiting planets like GJ1214b.

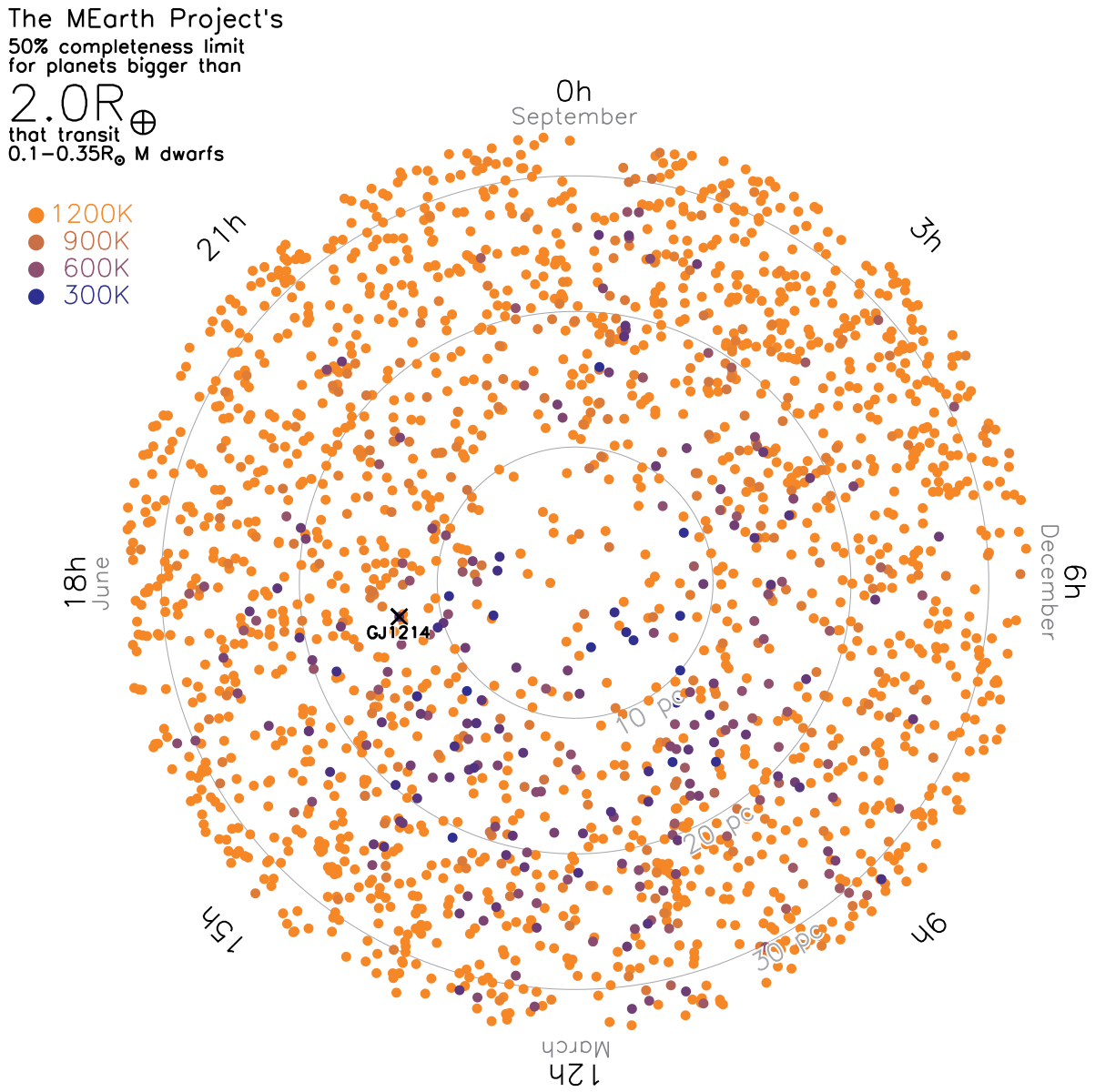

Figure 1.7: Same as Figures 1.5 and 1.6, but showing MEarth's completeness to $\geq 2R_{\oplus}$ transiting planets.

Chapter 2

The GJ1214 Super-Earth System: Stellar Variability, New Transits, and a Search for Additional Planets

This thesis chapter originally appeared in the literature as

Z. K. Berta, D. Charbonneau, J. Bean, J. Irwin, C. J. Burke, J.–M. Désert, *The Astrophysical Journal*, 736, 12, 2011

Abstract

The super-Earth GJ1214b transits a nearby M dwarf that exhibits 1% intrinsic variability in the near-infrared. Here, we analyze new observations to refine the physical properties of both the star and planet. We present three years of out-of-transit photometric monitoring of the stellar host GJ1214 from the MEarth Observatory and find the rotation

period to be long, mostly likely an integer multiple of 53 days, suggesting low levels of magnetic activity and an old age for the system. We show such variability will not pose significant problems to ongoing studies of the planet's atmosphere with transmission spectroscopy. We analyze 2 high-precision transit light curves from ESO's Very Large Telescope along with 7 others from the MEarth and FLWO 1.2 meter telescopes, finding physical parameters for the planet that are consistent with previous work. The VLT light curves show tentative evidence for spot occultations during transit. Using two years of MEarth light curves, we place limits on additional transiting planets around GJ1214 with periods out to the habitable zone of the system. We also improve upon the previous photographic V-band estimate for the star, finding $V = 14.71 \pm 0.03$.

2.1 Introduction

The transiting exoplanet GJ1214b offers an unparalleled opportunity to explore the physical properties of super-Earth planets. With a mass ($M_p = 6.6$ $M_\oplus$) and radius ($R_p = 2.7$ $R_\oplus$) between those of Earth and Neptune, and a likely equilibrium temperature ($T_{eq} = 500K$) cooler than for most transiting planets, GJ1214b represents an intriguing new kind of world with no Solar System analog (Charbonneau et al. 2009). Given intrinsic degeneracies in the mass-radius diagram in this regime (Seager et al. 2007; Adams et al. 2008; Rogers & Seager 2010a), the bulk composition of the planet cannot be uniquely determined from current measurements of the mass and radius alone. For example, Rogers & Seager (2010b) can explain the observed mass and radius to within 1σ with any of three generic physical models: (i) a mini-Neptune that accreted and maintained a low-mass H/He layer from the primordial nebula, (ii) a superfluid water-world with

a sublimating H_2O envelope, or (iii) a rocky planet with an H-dominated atmosphere formed by recent outgassing. Detailed calculations of GJ1214b's thermal evolution by Nettelmann et al. (2011) favor a metal-enriched H/He/H_2O envelope, finding that a water-only atmosphere would require an implausibly large water-to-rock ratio in the planet's interior.

Fortunately, because GJ1214b transits a very nearby (13 pc), bright ($K = 8.8$), low-mass M dwarf (0.16 $M_\odot$), it is amenable to follow-up observations that could distinguish among these hypotheses. In particular, the large ($D = 1.4\%$) transit depth enables transit studies of the planet's atmosphere. Miller-Ricci et al. (2009) show that measuring the amplitude of the planet's transmission spectrum (i.e., the wavelength-dependence of the transit depth $\Delta D(\lambda)$ caused by absorption at the limb of the planet) constrains the mean molecular weight of its atmosphere and, in turn, the hydrogen content of its outer envelope. Cases (i) or (iii) of Rogers & Seager (2010b) would produce $\Delta D(\lambda) \approx 0.1\%$ variations in the transit depth across wavelengths accessible from the ground as well as *Hubble* and *Spitzer Space Telescopes*, while case (ii)'s dense atmosphere would result in variations below the sensitivity of current instruments (Miller-Ricci & Fortney 2010).

Providing a potential complication, however, the host star GJ1214 shows roughly sinusoidal photometric modulations that are presumably due to an asymmetric distribution of spots on a rotating star. Such spots can bias planetary parameters as measured from transit light curves whether or not they are occulted by the planet (e.g. Pont et al. 2007; Désert et al. 2011b), partially decoupling the observed transit depth D from the actual planet-to-star radius ratio $R_p/R_\star$. Of particular importance for transmission spectroscopy studies, the change in transit depth induced by spots can vary with both time and wavelength, potentially mimicking the signal of a planetary

atmosphere.

Stellar spots have been observed in several transiting exoplanet systems around active stars (see Strassmeier 2009). *Hubble Space Telescope* (*HST*) photometry (Rabus et al. 2009) and later ground-based follow-up (Dittmann et al. 2009) of TrES-1b has shown evidence for spot occultations in transit light curves. The high photometric precision and continuous coverage provided by the CoRoT satellite enabled detailed modeling of spotty transit and out-of-transit light curves for the hot Jupiters CoRoT-2b (Wolter et al. 2009; Czesla et al. 2009; Huber et al. 2010) and CoRoT-4b (Aigrain et al. 2009; Lanza et al. 2009). For the former, joint fits to the transit and out-of-transit flux showed that initial estimates of the planet's $R_p/R_\star$ were 3% (9σ) too low (Czesla et al. 2009). The interpretation of the transiting super-Earth CoRoT-7b is obfuscated by the fact that both the transit depth and the reflex motion are well below the amplitude of activity-induced modulations (Léger et al. 2009; Queloz et al. 2009). Reanalyses of the CoRoT-7 radial velocities find changing values for the mass of CoRoT-7b (Hatzes et al. 2010; Ferraz-Mello et al. 2011; Lanza et al. 2010) and call into question the significance of the mass measurements for both CoRoT-7b and the claimed outer planet CoRoT-7c (Pont et al. 2011).

Like GJ1214b, the well-studied hot Jupiter HD189733b (Bouchy et al. 2005) is an ideal system for characterization studies, but requires corrections for stellar activity. The host HD189733 is an active K2 dwarf (Moutou et al. 2007) with 2% peak-to-peak variability in the optical (Croll et al. 2007; Miller-Ricci et al. 2008). Henry & Winn (2008) undertook a long-term photometric monitoring campaign from which they measured the 12 day stellar rotation period of HD189733. Extrapolation from their out-of-eclipse photometric spot characterization was useful for interpreting transmission

spectroscopy results of individual transits from *Hubble* (Pont et al. 2007; Swain et al. 2008) and measurements of the thermal phase curve from *Spitzer* (Knutson et al. 2007a, 2009a). Understanding the time-variable surface of the star was even more crucial for broadband transmission spectroscopy studies that rely on comparing transit depths at different epochs (e.g., Désert et al. 2009, 2011b; Sing et al. 2009); interpretation of these data rely heavily on the photometric monitoring of Henry & Winn (2008).

To aid ongoing and future studies of GJ1214b, we present new data (§2.2) to characterize the star GJ1214's variability and estimate its rotation period (§2.3). We compare the measured variability to a simultaneous analysis of 2 high-precision transit light curves from ESO's Very Large Telescope with 7 other new or previously published transits (§2.4). Additionally, we place upper limits on the radii of other possible transiting planets in the system (§2.5) and present a refined estimate of the star's V flux, which bears directly upon its metallicity as estimated using M_K and $V - K$ relations. Finally, we discuss the implications of the measured variability for the properties of the star and for transmission spectroscopy studies of GJ1214b's atmosphere (§2.6).

We also note the following correction. In Charbonneau et al. (2009), we quoted a systemic radial velocity for GJ1214 that had a typo in the sign; the actual velocity is $\gamma = +21.1 \pm 1.0$ km s^{-1} (i.e. a redshift).

2.2 Observations and Data Reduction

2.2.1 MEarth Photometry

We monitored the brightness of the GJ1214 system at a variety of cadences with the MEarth Observatory at Mt. Hopkins, AZ throughout the 2008, 2009, and 2010 spring observing seasons. As described in Nutzman & Charbonneau (2008), the MEarth Observatory was designed to detect transiting exoplanets around nearby M dwarfs, and consists of eight identical 40-cm telescopes on German Equatorial mounts in a single enclosure at the Fred Lawrence Whipple Observatory (FLWO). Each telescope is equipped with a thinned, back-illuminated 2048x2048 CCD with a pixel scale of 0.757"/pixel for a 26' field of view. For the bulk of the data presented in this work telescopes were equipped with a fixed, custom, 715 nm long-pass filter; the response is similar to a combination of the Sloan $i + z$ bandpasses and will be hereafter referred to as the "MEarth" bandpass. The MEarth Observatory is almost fully automated and operates on every clear night, observing target stars selected from a list of 2000 nearby late M dwarfs (Nutzman & Charbonneau 2008). In typical operating mode, each telescope observes its own list of 20-30 stars per night when they are above airmass 2 with the cadence and exposure times necessary to detect a transiting planet as small as 2 $R_\oplus$ in each star's habitable zone.

Light curves are extracted automatically from MEarth images by a modified version of the Monitor pipeline (Irwin et al. 2007), using nightly flat field (dawn and dusk), dark, and bias exposures for calibration. A differential photometry correction for each frame was calculated from a robust, weighted fit to 78 automatically selected field

comparison stars within 2.3 instrumental magnitudes of GJ1214. The mean MEarth - K color of these stars is 0.98; none were as red as GJ1214 (MEarth - $K = 2.19$). Predicted uncertainties for each measurement are calculated from a standard CCD noise model.

GJ1214 was observed with three main cadences. "Low" cadence (20-30 minutes between exposures) was that associated with the normal survey mode, and was employed in the 2008 and 2009 seasons. "Medium" cadence (5-10 minutes) was implemented after the discovery of GJ1214b and was intended to boost sensitivity both to other transiting planets and to characterizing the out-of-transit variability of the star. "High" cadence (40 seconds) was employed at predicted times of transit to determine the system parameters. High cadence transits were observed simultaneously with 7 or 8 MEarth telescopes for greater precision, as the systematic noise sources and scintillation patterns among pairs of MEarth telescopes appear to be largely uncorrelated, so the S/N improvement scales with the square root of the number of telescopes. In Table 2.1, we present one new MEarth transit light curve, along with the four MEarth and two KeplerCam light curves that were analyzed but not made electronically available in Charbonneau et al. (2009).

While we include data from 2008 for the rotation analysis, we caution that these observations took place during MEarth's early commissioning, before the observing strategy and software were finalized. Changes to the telescope throughout the season may have corrupted the season-long stability. Importantly, a field acquisition loop designed to mitigate flat-fielding errors by bringing each star back to the same pixel was not implemented until the late spring of 2008. During the 2008 season, a Bessell-prescription I filter (Bessell 1990) was used instead of the custom MEarth bandpass.

Table 2.1. Transit Light Curves

Mid-exposure Time[a]	Relative Flux[b]	Error[c]	Airmass	Instrument	Cycle[d]
2454964.8926699	0.99622	0.00317	1.1197	MEARTH	0
2454964.8933879	0.99777	0.00321	1.1198	MEARTH	0
2454964.8941049	1.00578	0.00325	1.1199	MEARTH	0
...					
2454980.7148766	0.99733	0.00185	1.6190	FLWO	9
2454980.7153966	1.00044	0.00185	1.6140	FLWO	9
2454980.7158946	0.99959	0.00185	1.6080	FLWO	9
...					
2455315.7660750	0.99786	0.00014	1.2125	VLT	221
2455315.7668492	0.99762	0.00014	1.2105	VLT	221
2455315.7702432	0.99763	0.00012	1.2020	VLT	221

Note. — This table is presented in its entirety in the electronic edition; a portion is shown here for guidance regarding its form and content.

[a]Times are given as BJD_{TDB}, Barycentric Julian Dates in the Barycentric Dynamical Time system (Eastman et al. 2010).

[b]Differential photometry corrections have been applied, but additional systematic corrections (see text) have not. Each light curve has been divided by the median out-of-transit flux.

[c]Theoretical 1σ errors have been calculated from a standard CCD noise model.

[d]Time measured from the reference epoch in units of the orbital period.

2.2.2 V-band Photometry from KeplerCam.

An independent out-of-transit light curve was obtained through Harris V and I filters[1] with KeplerCam on the 1.2 meter reflector at FLWO atop Mt. Hopkins, AZ. The observations discussed here were gathered in service mode from 26 March 2010 until 17 June 2010, after which date the mirror was taken off and put back on the telescope, introducing an uncorrectable systematic offset to the field light curves so later data had to be discarded.

Given the large night-to-night positional shifts of the field, we made an effort to quantify and ameliorate flat-fielding errors by sampling multiple regions of the detector, with each observation consisting of a set of three exposures offset by 3 arcminute dithers. Individual exposures had theoretical noise limits ranging from 0.3% to 2%, but the scatter among dither points suggested that calibration errors from flat-fielding introduced a 1% noise floor to the light curve. Dark sky flats were generated and corrected over time for changes at high spatial frequencies (i.e. dust donuts) by nightly dome flat exposures.

We measured calibrated V and I magnitudes (Table 2.2) to improve on previously published photographic estimates (Lépine & Shara 2005). Standard fields (Landolt 1992) were observed on the nights of 26, 27, and 28 March. Conditions were clear, although seeing as poor as 10" FWHM was witnessed. We estimate the calibration uncertainties for the nights from the scatter in multiple standard exposures.

[1] http://www.sao.arizona.edu/FLWO/48/CCD.filters.html.

2.2.3 Light Curves from VLT-FORS2

Spectra of GJ1214 and 6 comparison stars were gathered during three transits of GJ1214b using ESO Director's Discretionary Time on the VLT (Prog. ID #284.C-5042 and 285.C-5019). As described by Bean et al. (2010a), the primary purpose for obtaining these data was to measure the transmission spectrum of GJ1214b's atmosphere by generating multiwavelength transit light curves and determining the wavelength-dependence of the transit depth. In this work, we generate and analyze high precision "white" light curves by summing together all the photons collected in each spectrum.

Observations were performed in queue mode with the multiobject, low dispersion spectrograph FORS2 (Appenzeller et al. 1998) on VLT/UT1. The spectrograph was configured with the 600z+23 grism with a central wavelength of 900 nm and the red-sensitive (MIT) CCD in the standard 2x2 read mode. Exposure times were 20-40 seconds, and the readout time was 37 seconds. A custom slit mask was used; each slit was a rectangle 12" in the dispersion direction and 15-30" in the spatial direction, small enough to isolate GJ1214 and the comparison stars but large enough that changing slit losses due to variable seeing were negligible. Wavelength calibration exposures with a He, Ne, Ar emission lamp were taken through a 1" slit the day after each set of observations. Given the position of the comparison stars on the chip, the wavelength range 780 to 1000 nm was used in this analysis. The CCD response governs the red edge of this range, and the spectral response is similar to that of the MEarth bandpass.

After bias subtraction and flat fielding, we used the comparison stars to correct for the time varying zeropoint of the system. For each exposure, we extracted 1D spectra from the images using the optimal extraction of Horne (1986), and divided the total

flux (summed over wavelengths) of GJ1214 by the total flux in all the comparison stars. Theoretical error bars calculated from photon statistics alone were assigned to each point. Each exposure yielded $1 - 3 \times 10^8$ photons from GJ1214 and twice as many from comparison stars.

The corrected GJ1214 light curves exhibit systematic trends which we correct for by fitting a second-order polynomial function of time. In §2.4 we propagate the uncertainty from the systematics corrections through to the transit parameters. We searched for correlations between the relative flux and airmass, seeing, and positional shifts in the dispersion and cross-dispersion directions. The relationships were more complicated than low-order polynomials, so we did not attempt to remove them using common decorrelation techniques (e.g. Burke et al. 2010).

We also tested whether the observed drifts in flux could be caused by a changing color-dependence of the atmospheric extinction along the line of sight. To do so, we applied differential photometry corrections to individual spectral channels *before* combining them, to allow each wavelength its own extinction. This procedure did not remove the systematic trends.

We suggest the following as a more probable explanation for the systematics. Moehler et al. (2010) found that the linear atmospheric dispersion corrector (LADC) on the telescope has surface features that affect is sensitivity across the field of view. Because the LADC is positioned before the field rotator in the optical path and rotates relative to the sky, individual stars can drift across these features and encounter throughput variations that are not seen by the other comparison stars. No rotationally-dependent flat-fields were applied to these data, although Moehler et al. (2010) provide a route to a

possible correction.

The first two "white" light curves, normalized to their median out-of-transit flux level, are published in Table 2.1 and shown in Fig. 2.4.

A third transit observation was attempted on 2010 Jul 22. The brightest comparison star could not be used because it saturated mid-transit. The exposure times were cut in half immediately after egress, and a notable offset is visible in the transit light curve, perhaps due to an uncorrected non-linearity in the detector. Although the light curves of the first two transits were robust to the choice of comparison stars, the third changed significantly depending on which set of comparison stars was used. We show this transit in Fig. 2.4, but exclude it from all following analyses.

2.3 Rotation Period of GJ1214

With a growing understanding of the systematic effects present in MEarth data, we revisit the issue of GJ1214's intrinsic variability. In the discovery paper for GJ1214b (Charbonneau et al. 2009), we stated that the dominant periodicity seen in the out-of-transit light curve of the star GJ1214 had an 83 day period, implying that this was the rotation period of the star. Here, we revisit the question of GJ1214's rotation period with another season of observations.

Semi-stable spot complexes on the surface of a star imprint photometric modulations that can be approximated as a sinusoid with a fundamental period that matches the stellar rotation period. We search each year's light curve with a weighted, least-squares periodogram that has been modified to simultaneously fit for stellar variability along

with scaled templates of known systematic effects. These systematics are discussed in the next several subsections. To account for the likely evolution of spots with time, we investigate the 2008, 2009, 2010 data sets separately and do not require a coherent sine curve to persist over multiple years' data.

2.3.1 Avoiding Persistence in MEarth Light Curves

The MEarth detectors are subject to image persistence; pixels that are illuminated in one exposure can show enhanced dark current in subsequent exposures, which decays exponentially with a half hour time scale. Because the dark current in a given pixel depends on how recently that pixel was illuminated, differential photometry light curves can show baseline shifts between observations taken at different cadences, as well as 'ramps' at the start of a high-cadence sequence of exposures.

Correcting for these changing baseline shifts would require a simultaneous modeling of the complete photon detection history of every pixel and is impractical. During MEarth's normal survey mode, we purposely center subsequent targets on different pixels to avoid persistent charge stacking up. As the effect is most noticeable for data with cadence shorter than 5 minutes, we circumvent the problem by throwing out from the rotation period analysis all but the first point of any segment with such cadence.

2.3.2 Adding a Systematic 'Jitter'

Even having removed the highest cadence data, the sampling in the MEarth GJ1214 light curve can vary from N=1 to N=35 points per night, with a typical theoretical noise

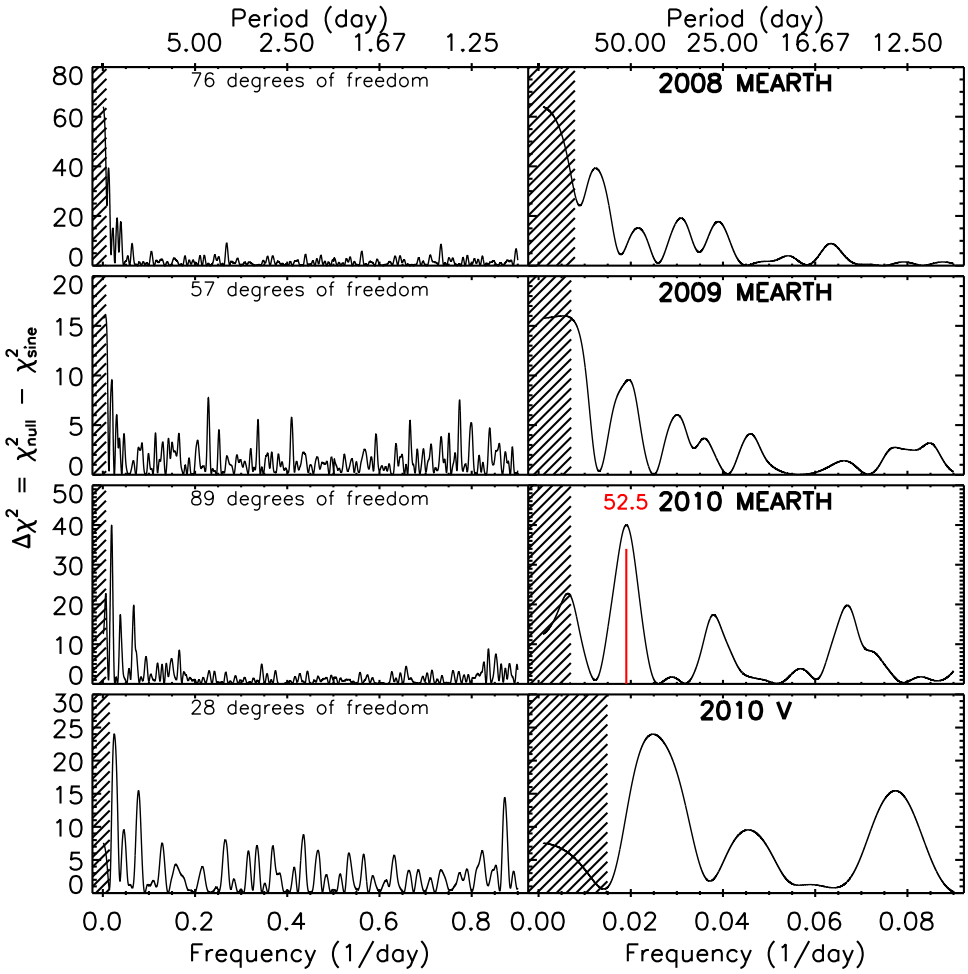

Figure 2.1: Periodograms showing the $\Delta\chi^2$ improvement achieved by fitting a sinusoid of a given period + common mode + meridian flip over a null model consisting only of common mode + meridian flip for MEarth (*top 3 panels*) and FLWO V-band photometric monitoring (*bottom*). Possible periods as short as 1 day (*left*) and a zoom in to longer rotational periods (*right*) are shown. Periods for which less than one full cycle is observed per season are shaded.

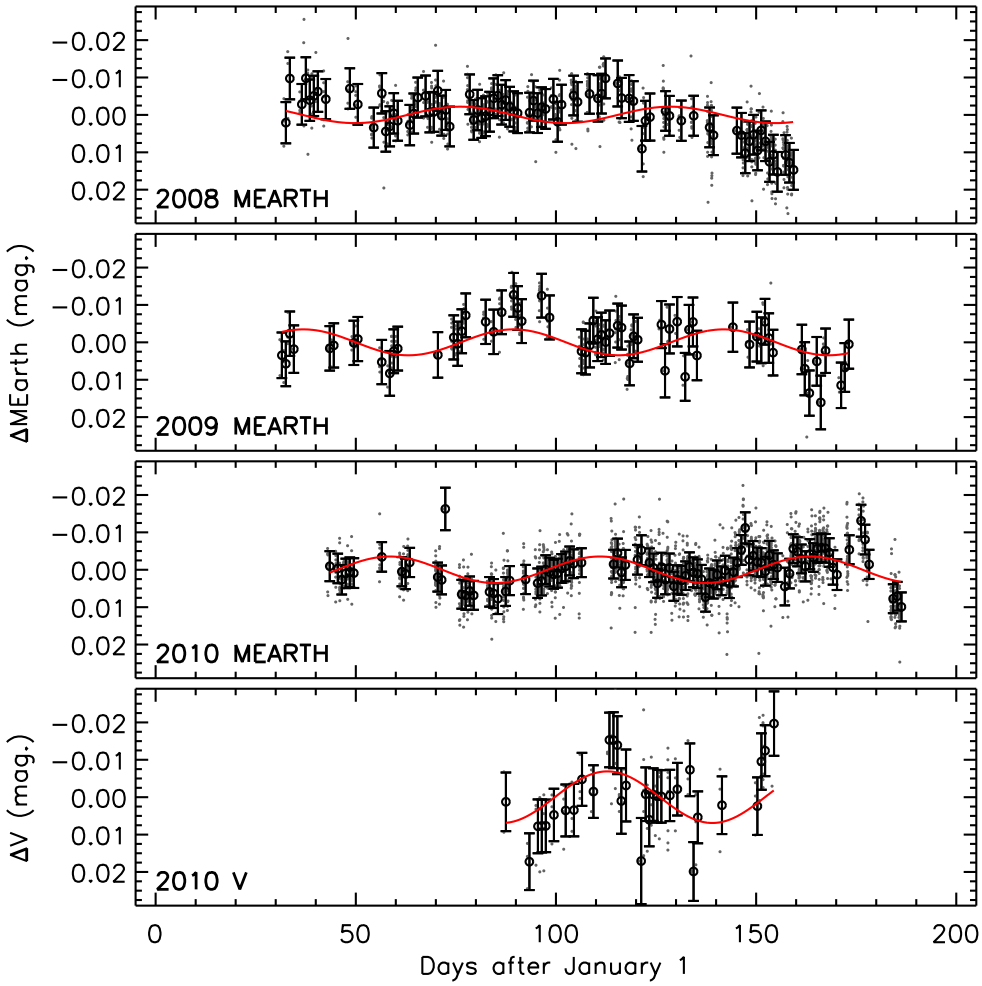

Figure 2.2: Out-of-transit light curves of GJ1214 from the MEarth Observatory (*top three panels*) and in V from the FLWO 1.2m (*bottom panel*). Individual exposures (*gray points*) and nightly binned values with errors that include the systematic 'jitter' (*black circles*, see text for details) are shown for each. A sine curve at the proposed 53 day period (derived from 2010 MEarth data) is shown (*solid lines*) at the best-fit phase and semiamplitude for each light curve. The MEarth points have been corrected for the best-fit common mode and meridian flip decorrelation.

limit per point of 3 millimagnitudes. In a strict least-squares sense, if the noise in our data were accurately described by an uncorrelated Gaussian process (e.g. photon noise), then every exposure should be allowed to contribute on its own to the period search, meaning the uncertainty associated with each night would go down as $1/\sqrt{N}$. Although MEarth telescopes have achieved such white noise down to millimagnitude levels within individual nights (Charbonneau et al. 2009), photometric variations between nights are most likely dominated by subtle changes in the telescope that are not corrected by our calibration efforts. The $1/\sqrt{N}$ weighting scheme would unfairly bias a period search to fit only a few well-sampled nights.

To account for this in each light curve, we remove all in-transit exposures and bin the data to a nightly time scale. For each night with N data points, we calculate an inverse variance weighted mean flux and time, using the theoretical errors calculated for each exposure in the weighting. To each of these nightly bins we assign an error given by

$$\sigma_{nightly} = \sqrt{\sigma_{bin}^2 + \sigma_{jitter}^2} \tag{2.1}$$

where σ_{bin} is the intrinsic standard error on the mean of the nightly bin ($RMS/\sqrt{N-1}$) and σ_{jitter} is a constant noise floor term to capture the night-to-night calibration uncertainty.

The signal lost in the binning process should be minimal. Preliminary searches of unbinned data and visual inspection of high cadence nights revealed no significant periodic signal at periods shorter than 1 day. Under the (untested) assumption that the stellar spin is roughly aligned with the orbital angular momentum, the upper limit on the projected rotation velocity of $v \sin i < 2$ km s^{-1} would correspond to a rotation period $P_{rot} > 5$ days.

We use observations on successive nights to estimate σ_{jitter} for the MEarth and the V-band light curves for each observing season. This assumes, on the basis of the apparent lack of short term variability, that the flux difference between pairs of nights is dominated by systematics. We find $\sigma_{jitter} = 0.0052$, 0.0058, 0.0038, 0.0067 magnitudes for the 2008, 2009, 2010 MEarth and 2010 V light curves. As these values are comparable to the predicted noise for most exposures, the quadrature addition of σ_{jitter} means we weight most nights roughly equally.

2.3.3 Correcting for Meridian Flips in MEarth Light Curves

MEarth's German Equatorial mounts rotate the detectors 180° relative to the sky when switching from negative to positive hour angles. Thus, the target and comparison stars sample two different regions of the detector. Given imperfect flat field corrections, an offset is apparent in many MEarth light curves between exposures taken on either side of the meridian.

To account for this effect, we allow different sides of the meridian to have different zeropoints. We construct a meridian flip template $m(t_i)$, which for an unbinned light curve would consist of binary values corresponding to the side of the meridian at each time stamp t_i. By extension, for each night of the binned data t_i, we define

$$m(t_i) = n_+/(n_+ + n_-) \tag{2.2}$$

where n_+ and n_- are the number of data points with positive or negative hour angles in a given nightly bin. We allow a scaled version of this template to be fit simultaneously with the period search.

2.3.4 Correcting for Water Vapor in MEarth Light Curves

Because the wide MEarth bandpass overlaps significant water absorption features in the telluric spectrum, the color-dependence of the throughput of our observing system is sensitive to the precipitable water vapor (PWV) in the column overhead. The fraction of stellar photons lost to water vapor absorption from a typical MEarth target M dwarf is much larger than the fraction lost from the (typically solar-type) comparison stars. When PWV along the line of sight to a star varies, a crucial assumption of simple differential photometry - that stars are experiencing the same losses - is violated.

Although the variations in any particular light curve might come either from the PWV induced noise or from intrinsic stellar variability, we can harness the ensemble of M dwarfs observed by MEarth at any particular time to characterize and correct for this effect. To do so, we construct a "common mode" template by robustly (median) binning all the differential photometry light curves of all M dwarfs observed on all eight MEarth telescopes into half hour bins. This averages out uncorrelated stellar variability and serves as an estimate of the atmospheric variation that is common to all red stars observed at a given time. We only use data during times when we have > 50 and > 30 targets contributing to a bin. The strongest periodicities in the common mode templates are 25.1 days for spring 2009 and 14.5 days for spring 2010 (Irwin et al. 2011a). If left uncorrected, such periodicities could appear as spurious intrinsic stellar variability.

To correct the GJ1214 light curve for this effect, we interpolate the "common mode" to the unbinned time stamps. We then perform the nightly binning on it to construct a common mode template $c(t_i)$, which we use for simultaneous decorrelation (see next section). This binning is justified because the typical common mode variation

within a night is typically at the level of 1-2 millimagnitudes, much smaller than the night-to-night or week-to-week changes we hope to correct.

2.3.5 Periodograms

We generate a periodogram by calculating the χ^2 of a weighted linear fit of the light curve $\Delta F(t_i)$ (in magnitudes) to a model

$$\Delta F_{sine}(t_i) = A \sin\left(\frac{2\pi(t_i - t_0)}{P_{rot}}\right) + B\ m(t_i) + C\ c(t_i) + D \tag{2.3}$$

where A is the semiamplitude of the sinusoidal variability, t_0 is an epoch, P_{rot} is the stellar rotation period, B and C are scale factors for the systematics, and D is a constant offset. We compare this χ^2 to that of the null hypothesis that ΔF is explained by the systematics alone

$$\Delta F_{null}(t_i) = B\ m(t_i) + C\ c(t_i) + D. \tag{2.4}$$

Mathematically, this procedure would be identical to traditional least-squares periodograms (Lomb 1976; Scargle 1982) if we fixed $B = C = 0$. In (Irwin et al. 2011a), we use a similar method to estimate photometric rotation periods for a sample of 41 MEarth M dwarf targets and test its sensitivity with simulations.

In Fig. 2.1 we plot the χ^2 improvement ($\Delta\chi^2 = \chi^2_{\rm null} - \chi^2_{sine}$) between these two hypotheses for each of the three MEarth seasons and the short 2010 V-band campaign. Periods for which less than one cycle would be visible have been masked.

The most prominent peak among all the periodograms in Fig. 2.1 is that at 53 days from the 2010 MEarth data, which corresponds to a semiamplitude of $A = 3.5 \pm 0.7$ millimagnitudes, where the uncertainty has been estimated from the covariance matrix of

the linear fit. We estimate the false alarm probability (FAP) for this period by running the complete period search on 10^4 time series that consist of the best-fit scaled versions of $m(t_i)$ and $c(t_i)$ and randomly generated Gaussian noise set by $\sigma_{nightly}$, recording the $\Delta\chi^2$ of the best peak from each iteration. We find that a FAP of 10^{-4} corresponds to $\Delta\chi^2 = 28$, much less than the achieved $\Delta\chi^2 = 41$.

There is a nearby, but statistically insignificant, peak at 51 days in the 2009 MEarth data. Both 2008 and 2009 MEarth light curves are dominated by long-period trends that are unresolvable in each year. Of the resolved peaks in the 2008, the strongest is at 81 days (FAP $< 10^{-4}$). In spite of the formal significance of this last peak, we caution that mid-season changes to the then still uncommissioned observatory might also account for the variations seen. One conclusion is robust; our 3 years of MEarth light curves show no evidence for any rotational modulation with a period shorter than 25 days.

Given that 2010 had the most uniform sampling and cadence, we tentatively suggest $P_{rot} = 53$ days as our current best estimate of GJ1214's likely rotation period. Fig. 2.2 shows the 3 MEarth and 1 V binned light curves with a sinusoid whose period has been fixed to our estimated $P_{rot} = 53$ days but whose amplitude and phase have been fitted to the data. The fit is acceptable for 2009 MEarth, but clearly fails for the 2008 MEarth data.

We stress the caveat that the true rotation period could instead be a longer multiple of our quoted 53 day period (e.g. $P_{rot} \approx 100$ days) if the star exhibits multiple, well-spaced active regions. This kind of harmonic confusion appeared and was addressed in studies of Proxima Cen (Benedict et al. 1998; Kiraga & Stepien 2007). Preliminary data for GJ1214 collected in 2011 while this paper was under review do not seem to

show evidence for a 53 day period, preferring instead a much longer one. Due to this factor-of-n uncertainty in the true rotation period, we do not quote a formal error bar on our 53 day period estimate.

2.3.6 Chromatic Spot Variation

By itself, the V-band light curve prefers a period of 41 days (Fig. 2.1; FAP = $5\times10^{-}4$). When forced to fit a 53 day period (Fig. 2.2), these data show a phase offset of only 2 days relative to the simultaneous MEarth data. The semiamplitude of this fit is 7 ± 3 millimagnitudes, twice that seen in the MEarth bandpass.

There are 28 nights when observations were obtained in both MEarth and V band. In Fig. 2.3 we plot the nightly bins against each other; the apparent correlation suggests that the two instruments are observing the same stellar variability across two bands and not telescope systematics. We fit a line to the relation, accounting for errors in both ΔMEarth and ΔV (Press 2002) and find a slope of 2.4$\pm$0.8. While the significance is marginal, we take this as further evidence that the amplitude of the variability in V is greater than that in the MEarth band.

If starspots have a temperature ($T_\bullet$) that is only modestly lower than the stellar effective temperature ($T_{\rm eff}$), the color-dependence will arise from the spectrum of the spot rotating in and out of view. The factor of 2 we see would be consistent with $T_\bullet/T_{\rm eff} \approx 90 - 95\%$ as is commonly assumed in M dwarf eclipsing binaries (e.g. Morales et al. 2009; Irwin et al. 2009a). Totally dark spots ($T_\bullet = 0$K) would produce less of a chromatic variation, but would still be sensitive to the spectral signature of the stellar limb-darkening (e.g. Poe & Eaton 1985).

Table 2.2. Photometry of GJ1214

Filter	Magnitude	Source
V	14.71 ± 0.03	this work
I	11.52 ± 0.03	this work
J	9.750 ± 0.024	2MASS[a]
H	9.094 ± 0.024	2MASS[a]
K	8.782 ± 0.024	2MASS[a]

[a]Skrutskie et al. (2006)

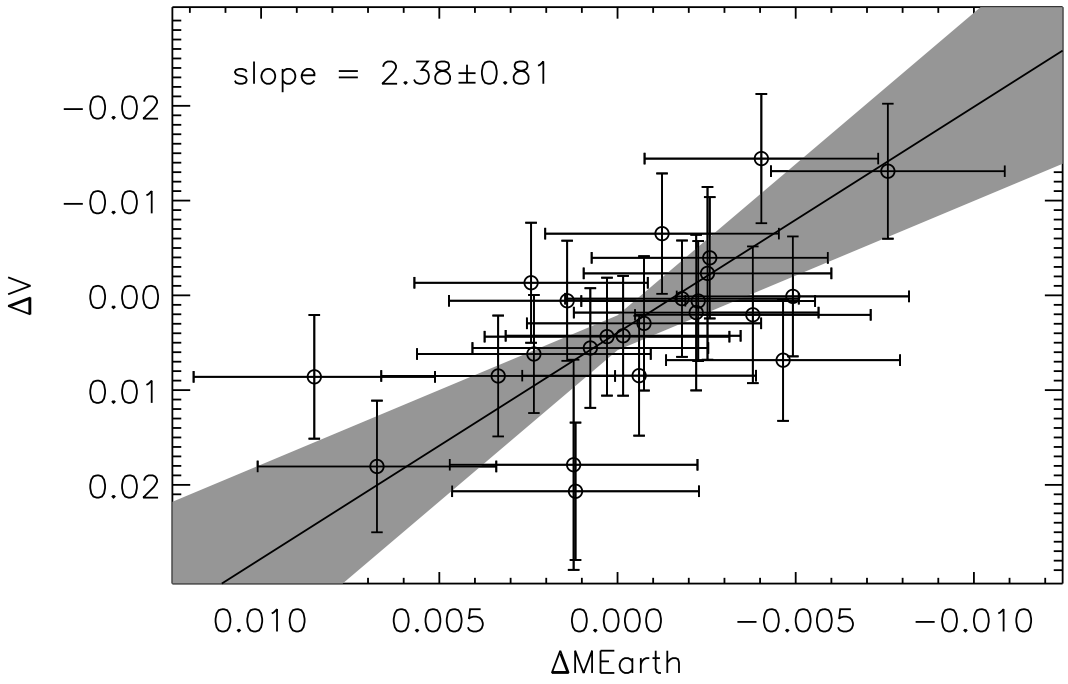

Figure 2.3: For a sub-sample of nights measured in both MEarth and V bandpasses, the nightly bins plotted against each other with the assigned σ_{bin} error bars. The best-fit slope (*black line*) and 1σ interval (*shaded region*) are shown.

2.4 Fitting the Transit Light Curves

We perform a simultaneous fit to the 4 transit light curves from MEarth and 2 from KeplerCam published by Charbonneau et al. (2009), 1 more transit collected by MEarth in spring 2010, and 2 high-precision transits from the VLT. This totals 9 light curves of 7 independent transits, as shown in Figures 2.4 and 2.5. We employ a model corresponding to a circular planet transiting a smooth, limb-darkened star (Mandel & Agol 2002) that has the following parameters: the planet-to-star radius ratio $R_p/R_\star$, the stellar radius $R_\star$, the total transit duration t_{14}, two quadratic stellar limb darkening parameters u_1 and u_2 for each of the 3 telescope systems used, and 7 mid-transit times T_c. The reparameterization of the scaled semimajor axis $a/R_\star$ and inclination i in terms of $R_\star$ and t_{14} substantially reduces the degeneracies in the problem, leading to an efficient exploration of the parameter space (Burke et al. 2007; Carter et al. 2008).

We fix the orbital period to $P = 1.5804043$ days (Sada et al. 2010). Given the upper limit on eccentricity from radial velocities and the short circularization time for GJ1214b (10^6 years for $Q'_p = 100$ and $Q'_* = 10^6$, following Raymond et al. 2008), we assume an eccentricity $e = 0$ throughout. Where necessary to derive physical parameters from the geometric parameters in the light curve fit, we adopt the the stellar mass $M_\star = 0.157$ $M_\odot$ (Charbonneau et al. 2009); we describe how we propagate the 0.019 $M_\odot$ uncertainty on this value to the other errors in the next section.

To account for the systematic trends present in the VLT light curves, we introduce a correction to the baseline stellar flux parameterized as a parabola in time ($a + bt + ct^2$) to each transit. Most of the MEarth and FLWO light curves showed no significant systematic trends, so only a single out-of-transit baseline flux level was fit to each night.

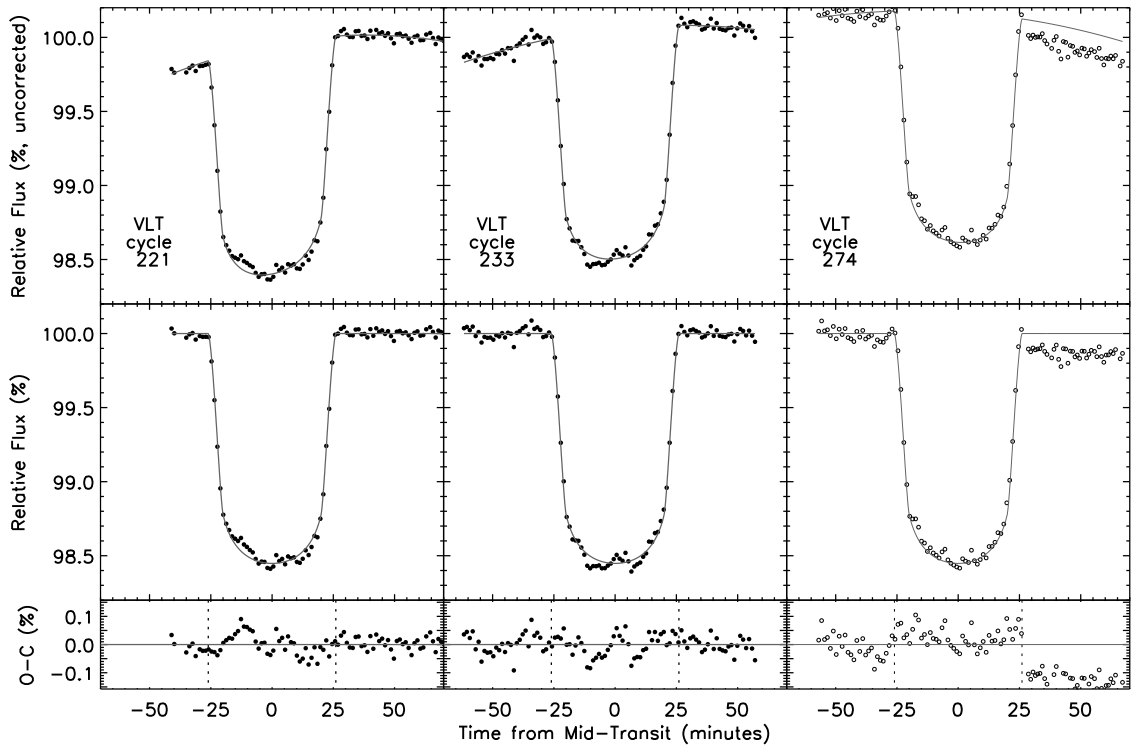

Figure 2.4: Transit light curves from summed FORS spectra from the VLT, before (*top*) and after (*middle*) subtracting a second-order polynomial function of time. The third transit (*open circles*) exhibited larger uncorrectable systematics than the first two (see text), and was excluded from further analysis. Residuals from the quoted model (*gray lines*) are also shown (*bottom*), with dotted lines indicating the in-transit duration.

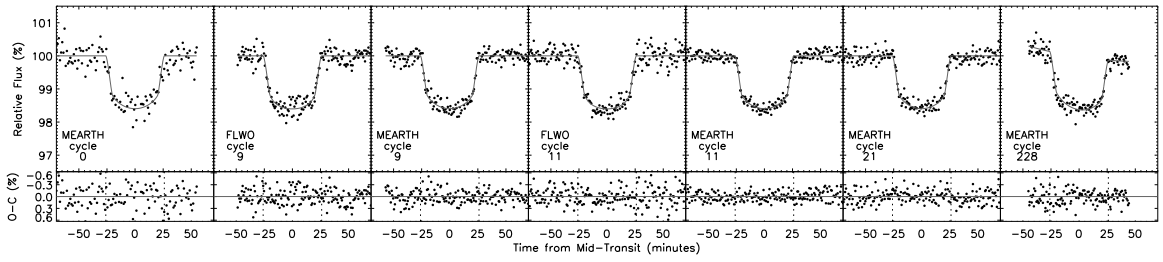

Figure 2.5: Transit light curves from MEarth and KeplerCam. Residuals from the quoted model (*gray lines*) are also shown (*bottom*), with dotted lines indicating the in-transit duration. The cycle 228 MEarth transit shows a linear trend with airmass that is included in the fit.

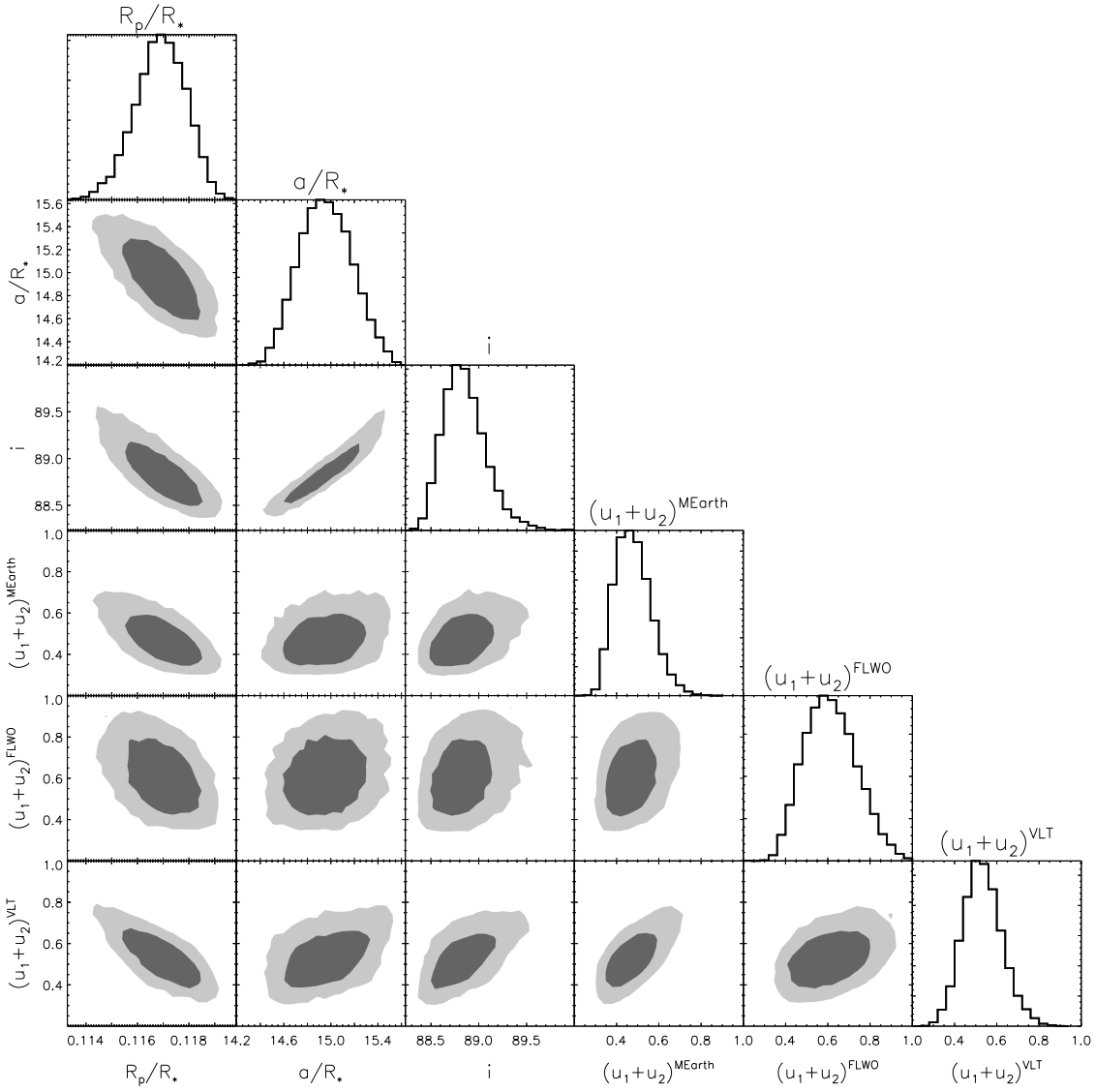

Figure 2.6: Matrix showing the correlations between selected transit parameters from a simultaneous residual permutation analysis of 5 MEarth, 2 FLWO, and 2 VLT light curves. Histograms are shown for each parameter, as well as contours that contain 68% (*dark gray*) and 95% (*light gray*) of the samples for each pair of parameters. For each telescope system, we show a linear combination of the limb-darkening coefficients $u_1 + u_2$ that is strongly correlated with the other parameters.

The MEarth light curve on the night of 2010 May 10 showed a strong correlation with airmass, so we also included a linear trend with airmass for this one night.

2.4.1 χ^2 minimization

We determine the best-fit values of the 30 model parameters by using an implementation of the Levenberg-Marquadt (LM) routine called MPFIT (Markwardt 2009) to minimize the value of

$$\chi^2 = \sum_{i=1}^{N} \frac{(F_i - F_{\rm model})^2}{\sigma_{\rm i}^2} \quad (2.5)$$

where F_i are the $N = 1495$ flux measurements (Table 2.1), σ_i are their uncertainties, $F_{\rm model}$ is the model described above.

After this initial fit, the uncertainty estimates for each light curve were increased until the reduced χ^2 of the out-of-transit residuals was unity and the fit was repeated.

While the LM fit provides a linearized estimate of the covariance matrix and errors of the parameters, this estimate is too precise because the method a) does not fully sample along non-linear correlations between highly-degenerate parameters and b) does not account for correlations between data points. In what follows, we calculate more conservative and realistic errors through a bootstrap method that addresses both these issues.

2.4.2 Error Estimates by Residual Permutation

Although the noise in the MEarth and FLWO light curves is well described by a white Gaussian process, the much lower photon noise in the VLT light curves reveals underlying

low level serial correlations among data points, or 'red' noise. The autocorrelation function of the VLT residuals is above 0.25 out to 4 and 2 data point lags respectively for the two transits, and binning the residuals by N points reduces the scatter more slowly than $1/\sqrt{N}$. To quantify the excess uncertainty in the fitted parameters due to this red noise, we perform a 'residual permutation' bootstrap simulation (Moutou et al. 2004; Gillon et al. 2007a; Désert et al. 2009) that fits resampled data while preserving the correlations between data points. This analysis is carried out simultaneously for the VLT, FLWO and MEarth light curves.

After subtracting the best-fit model from the ensemble of light curves, we perform 2×10^4 iterations of the following procedure. Preserving the time stamps for all the exposures, we cyclicly permute the residuals for each light curve by a random integer (shifting along the series and wrapping back from the last exposure to the first), inject the best-fit model back into the set of shifted residuals, perform the LM fit on the simulated light curve ensemble, and record the results. To initialize the parameters for the LM fit, we select guesses drawn from a multivariate Gaussian distribution whose covariance matrix is scaled up by 2^2 (i.e. 2σ) from the LM's estimate. We excised those bootstrap samples that found a best fit with unphysical limb-darkening parameters ($u_1 < 0, u_1 + u_2 > 1, u_1 + 2u_2 < 0$, Burke et al. 2007).

In Table 2.4, we quote the best-fit value for each parameter and uncertainty bars that exclude the lower and upper 15.9% of the bootstrap samples (i.e. the central 68.3% confidence interval), where parameters that were not directly fit have been calculated analytically from those that were (Seager & Mallén-Ornelas 2003). Having fit quadratic limb-darkening parameters for each instrument, we present similar confidence intervals of the coefficients u_1 and u_2 in Table 2.5.

Fig. 2.6 summarizes the correlations among the parameters $R_p/R_\star$, $a/R_\star$, i, and the linear combination of the limb-darkening parameters $u_1 + u_2$ for the three telescope systems. As the difference between central and limb specific intensities, $u_1 + u_2$ correlates strongly with the transit depth and $R_p/R_\star$. We also show in Fig 2.6 the bootstrap histograms for these parameters.

Although we use $R_\star$ as a fit parameter, the quantity that is actually constrained by the light curves is $\rho_\star$. To calculate the true uncertainty on $R_\star$ (and R_p), we calculate $\rho_\star$ for all of our bootstrap samples using the fixed $M_\star = 0.157\mathrm{M}_\odot$, assign values of $M_\star$ drawn from the appropriate Gaussian distribution, and recalculate $R_\star$ from $\rho_\star$ and $M_\star$.

We find results consistent with Charbonneau et al. (2009). Despite the high precision of the VLT light curves, the uncertainties for most parameters are comparable to the earlier work, and that on $R_p/R_\star$ is slightly larger. This is due in part to the correlated noise analysis we perform that Charbonneau et al. (2009) did not. More significantly, Charbonneau et al. (2009) fixed the quadratic limb-darkening parameters to theoretical values while we fit for them directly. The extra degrees of freedom allowed by our relaxation of astrophysical assumptions are known to increase the uncertainty on $R_p/R_\star$ (Burke et al. 2007; Southworth 2008). For comparison, Charbonneau et al. (2009) used coefficients appropriate for a 3000K, $\log g = 5$, PHOENIX atmosphere, specifically Cousins I (tabulated in Bessell 1990) coefficients ($u_1 = 0.303, u_2 = 0.561$; Claret 1998) as an approximation for the MEarth bandpass and Sloan z coefficients ($u_1 = 0.114, u_2 = 0.693$; Claret 2004) for the KeplerCam FLWO data. While our individual fitted values differ from these, the integral over the stellar disk ($1 - u_1/3 - u_2/6$) is very well reproduced.

2.4.3 Transit Timing Results

Mid-transit times (equivalent to times of inferior conjunction given the assumed circular orbit) are printed in Table 2.6. Our uncertainty estimate on the VLT transit times is 50% larger than it would be if we had ignored correlations in the data, but still only 6 seconds. This uncertainty is within a factor of two of the uncertainty on the highest-precision transit times yet measured from either the ground or space: the 3-second measurements of HD189733b with *Spitzer* (Agol et al. 2010). For stars of comparable brightness to GJ1214, large aperture ground based telescopes offer a powerful tool for precision transit times.

A revised linear ephemeris is shown in Table 2.4 derived with a weighted least squares method from our transit times and those of Sada et al. (2010). Residuals are shown in Fig. 2.7. The reduced χ^2 of the linear fit $\chi^2_\nu = 0.62$ gives no indication of transit timing variations over the time scales probed, and a Bayesian model comparison test (see Burke et al. 2010) does not show significant evidence for a model with linear and hypothetical sinusoidal components over a purely linear ephemeris. We note that a hypothetical 1.0 $M_\oplus$ planet in a 2:1 mean-motion resonance with GJ1214b would introduce ~ 100 second transit timing variations that would be easily detected in these data (following Bean 2009).

2.4.4 Occulted Spots

Occulted spots will appear in the transit light curve as a bump lasting roughly as long it takes the planet to move a distance $2(R_p + R_\bullet)$ across the spot, assuming a circular spot with radius $R_\bullet$. For small spots ($R_\bullet << R_p$), this is roughly the transit ingress/egress

Table 2.3. Summary of Transit Light Curves

Cycle	UT Date	Instrument	RMS (ppm)	Cadence (sec.)
0	2009 May 13	MEarth	2950	60
9	2009 May 29	FLWO	1960	45
9	2009 May 29	MEarth	1580	(binned) 45
11	2009 Jun 01	FLWO	2060	45
11	2009 Jun 01	MEarth	1240	(binned) 45
21	2009 Jun 17	MEarth	1620	(binned) 45
221	2010 Apr 29	VLT	380	72
228	2010 May 10	MEarth	1770	(binned) 45
233	2010 May 18	VLT	350	72

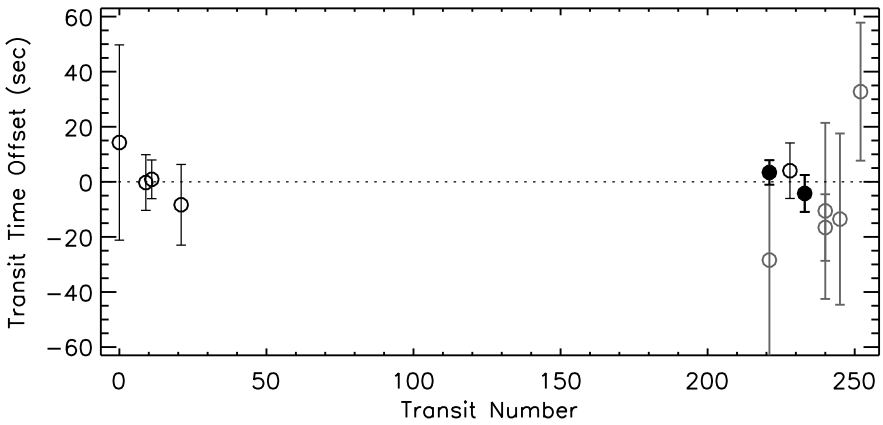

Figure 2.7: Deviations in the times of transit from the new best linear ephemeris, including transits from the VLT (*filled black circles*), MEarth (*open black circles*), and work by Sada et al. (2010) (*open grey circles*).

Table 2.4. GJ1214b Parameters and Uncertainties

Parameter	Previous Work	This Work[a]
R_p/R_*	0.1162 ± 0.0007	0.1171 ± 0.0010
a/R_*	14.66 ± 0.41	14.93 ± 0.24
i (deg.)	$88.62^{+0.35}_{-0.28}$	$88.80^{+0.25}_{-0.20}$
b	$0.354^{+0.061}_{-0.082}$	$0.313^{+0.046}_{-0.061}$
$\rho_*(g/cm^3)$	23.9 ± 2.1	25.2 ± 1.2
R_* ($R_\odot$)[b]	0.2110 ± 0.0097	$0.2064^{+0.0086}_{-0.0096}$
R_p ($R_\oplus$)[b]	2.68 ± 0.13	2.64 ± 0.13
$t_{12} = t_{34}$ (minutes)[c]	...	$6.00^{+0.24}_{-0.25}$
t_{23} (minutes)[c]	...	$40.10^{+0.46}_{-0.43}$
t_{14} (minutes)[c]	...	$52.11^{+0.25}_{-0.22}$
P (days)	1.5804043 ± 0.0000005	$1.58040490 \pm 0.00000033$
T_0 (BJD_{TDB})	$2454966.52506 \pm 0.00006$[d]	$2454966.525042 \pm 0.000065$

[a]Best-fit value and confidence intervals that exclude the upper and lower 15.9% of the residual permutation bootstrap samples. Where upper and lower limits differ by $> 10\%$, asymmetric errors are shown.

[b]Uncertainty calculated as described in the text assuming $M_\star = 0.157 \pm 0.019 M_\odot$

[c]Times between contact are $t_{12} = t_{34}$ = ingress/egress time, t_{23} = duration of full eclipse, t_{14} = total duration of transit. t_{14} was fit directly from the light curve, and the others were calcuated from the fitted parameters (see Seager & Mallén-Ornelas 2003; Carter et al. 2008).

[d]Converted from BJD_{UTC} to BJD_{TDB} and added 1 cycle for ease of comparison.

References. — Charbonneau et al. (2009), Sada et al. (2010)

Table 2.5. Inferred GJ1214 Limb-darkening Coefficients

Instrument	λ range (nm)	u_1[a]	u_2[a]
MEarth	715-1000	0.53 ± 0.13	-0.08 ± 0.21
VLT-FORS2	780-1000	0.34 ± 0.31	0.28 ± 0.46
KeplerCam z	850-1000	0.26 ± 0.11	0.26 ± 0.19

[a]Best-fit value and confidence intervals that exclude the upper and lower 15.9% of the residual permutation bootstrap samples.

Table 2.6. GJ1214 Transit Times

Cycle[a]	T_c (days)[b]	σ (days)
0	2454966.525207	0.000351
9	2454980.748682	0.000104
11	2454983.909507	0.000090
21	2454999.713448	0.000155
221	2455315.794564	0.000066
228	2455326.857404	0.000110
233	2455334.759334	0.000066

[a]Time from the first MEarth transit, measured in units of the orbital period.

[b]Mid-transit times T_c are given as Barycentric Julian Dates in the Terrestrial Time system ($\mathrm{BJD_{TDB}}$). To the best precision afforded by these transits, these times can be converted to a UTC based system via the approximation $\mathrm{BJD_{UTC}} \approx \mathrm{BJD_{TDB}} - (32.184 + N)/86400$ where N=34 is the number leap seconds between UTC and TAI at the time of these transits (Eastman et al. 2010).

time, or 6 minutes in the case of GJ1214b. The amplitude of a spot crossing event is determined by the fractional deviation in surface brightness occulted by the planet from the star's mean and could in principle be comparable to the transit depth itself for totally dark spots larger than the planet.

We see evidence for spot occultations in the first two VLT transits shown in Fig. 2.4, in which the residuals from a smooth model show a 0.1% brightening at the start of the first transit and near the middle of the second transit. These features last 5-10 minutes, as expected, and persist with comparable amplitudes in light curves generated separately from the blue and red halves of the FORS spectra. Their presence is robust to choice of comparison stars.

For the level of precision afforded by current data, we treat these possible spot occultations as excess red noise with correlations on the scale of the 6 minute spot crossing time. As such, the residual permutation method (§2.4.2) accounts for the uncertainty introduced by these features, regardless of their physical interpretation. Spots that are not occulted will not be accounted for in these errors estimates, but we discuss their influence in §2.6.2.

2.4.5 Unocculted Spots

Unocculted spots will also have an effect on the planetary parameters. By diminishing the overall flux from the star while leaving the surface brightness along the transit chord unchanged, increasing the coverage of unocculted spots on the star will make transits deeper. This is opposite the effect of occulted spots, which tend to fill in transits and make them shallower. Because we must apply a nightly normalization to each light curve

to avoid systematics and compare across telescopes, this depth change carries through to the implied $R_p/R_\star$.

To test whether unocculted spots are biasing individual measurements of $R_p/R_\star$, we repeat the fit and uncertainty estimation described in §2.4.1 and §2.4.2 but allow the 8 transit epochs to have different values of $R_p/R_\star$. As the visible spot coverage changes when the star rotates, we might detect changes in the inferred values of $R_p/R_\star$. For this experiment we are interested in relative changes among the $R_p/R_\star$ values, so we fix the other geometric parameters $R_\star$ and t_{14} and the limb-darkening parameters to their best fit values to effectively collapse along those dominant degeneracies.

The inferred $R_p/R_\star$ for each transit is shown in Fig. 2.8, with uncertainties estimated from the residual permutation method. Because the other parameters were fixed in this fit, the ensemble of points are free to move up and down slightly on this plot; the uncertainties shown are more relevant to comparisons between epochs. In §2.6.2, we will discuss our calculation of the predicted variations induced by unocculted spots consistent with the observed variability.

2.5 Limits on Additional Transiting Planets

After clipping out known transits of GJ1214b, we investigate the light curve of GJ1214 for evidence of other transiting planets in the system. The light curve contains 3218 points spanning spring 2009 and 2010.

To remove structured variability from the light curve before searching for transits, we employ an iterative filtering process that combines a 2-day smoothed median filter

(Aigrain & Irwin 2004) with a linear decorrelation against both external parameters (primarily the common mode, meridian flip, seeing, and pixel position) and light curves of other field stars (Kovács et al. 2005; Tamuz et al. 2005; Ofir et al. 2010). Such filtering decreases the light curve RMS from 9.6 to 4.7 millimagnitudes per point.

Using a variant of the box-fitting least squares (BLS) algorithm (Kovács et al. 2002), we search the filtered light curve for periodic rectangular pulses over a grid of period and transit phase. At each test period, we fix the transit duration to that of a mid-latitude transit of a circularly orbiting planet for GJ1214's estimated stellar mass (0.157 $M_\odot$) and radius (0.207 $R_\odot$). With the typical MEarth precision and cadence, our sensitivity depends only weakly on transit duration; violations of these assumptions will not substantially penalize our detection efficiency.

At every grid point, we determine the best fit transit depth D and the improvement $\Delta\chi^2$ over the $D = 0$ null hypothesis, using a weighted least-squares method that includes red noise estimated from the light curve itself using the σ_r formalism of Pont et al. (2006). Following Burke et al. (2006) we characterize candidates in terms of $\Delta\chi^2$ and $f = \max(\Delta\chi^2_{\text{each transit}}/\Delta\chi^2_{\text{total}})$, which is the largest fraction that any one transit event contributes to the signal.

The best candidate found in the clipped GJ1214 light curve exhibited a $\Delta\chi^2$ of 31.3 and is shown in Fig. 2.9. To estimate the significance of this value, we employ the bootstrap method of Jenkins et al. (2002). Strictly speaking, the presence of remaining correlated noise in our filtered light curve means this method gives an overestimate of the false alarm probability, but the complicated correlation structure of the light curve make a more accurate significance estimate difficult to calculate. In practice, we have

found the Jenkins method to provide an appropriate limit, even for light curves with substantial red noise. We generate 10^3 (so we can estimate the χ^2 associated with a 1% false alarm probability from 10 samples) transit-less light curves with Gaussian white noise and time sampling identical to that of the real light curve and performing the BLS search on these fake light curves. We find 20% of them show values of $\chi^2 > 31.3$, suggesting our best candidate shown in Fig. 2.9 is not significant.

These data place limits on the presence of other transiting planets in the GJ1214 system. Like Burke et al. (2006), Croll et al. (2007) and Ballard et al. (2010), we simulate our sensitivity by injecting 8000 randomly phased, limb-darkened transits of 1.0, 1.5, 2.0, and 4.0 $R_\oplus$ planets with random periods. We then attempt to recover them with our transit search using objective detection criteria. To account for possible suppression from the filtering, we inject the transits into the raw light curves and reapply the filter at each iteration. We adopt two criteria for detection: $\Delta\chi^2 > 50$ (corresponding to a formal probability of false alarm of 10^{-5} by the bootstrap analysis) and $f < 1.0$ (to ensure at least two events contribute).

To demonstrate this visually, we show in Fig. 2.9 examples of injected transits for each relevant radius, randomly selected from among the simulated planets with injected periods of 10 ± 1 days whose BLS results satisfied the detection criteria. Planets with periods near 10 days are of particular interest; given GJ1214's low luminosity, they would be in the star's habitable zone. Fig. 2.10 plots the fraction of injected planets that crossed the detection threshold as a function of period for each input planetary radius. The shape of the 4.0 $R_\oplus$ curve in Fig. 2.10 is driven largely by the $f < 1.0$ criterion requiring multiple events are observed, whereas the 1.5 $R_\oplus$ curve is dominated by the need for sufficient in-transit S/N to get to $\Delta\chi^2 > 50$. With these criteria, our

sensitivity falls below 50% beyond periods of 15, 8, and 2 days for 4.0 $R_\oplus$, 2.0 $R_\oplus$, and 1.5 $R_\oplus$ transiting planets. We had no sensitivity to 1.0 $R_\oplus$ planets.

Fig. 2.10 also shows the recovery fraction for 4.0 $R_\oplus$ planets without requirement that more than one event be observed. Neptune-sized transits of GJ1214 are so dramatic that they could be confidently identified by a single event (see Fig. 2.9). Our simulations show that we are 90% sensitive to transiting Neptunes around GJ1214 out to 10 days, and 80% sensitive out 20 days. For smaller planets requiring multiple events, however, a significant volume of parameter space remains unconstrained.

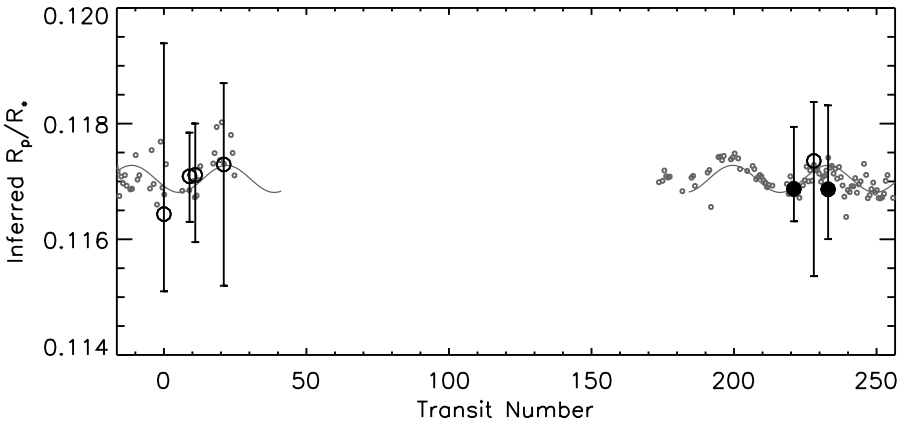

Figure 2.8: Estimates of the apparent planet-to-star radius ratio at each epoch (*black circles* for MEarth/FLWO (*open symbols*) and VLT transits (*filled symbols*)). Via the residual permutation estimate, the error bars include the uncertainty due to possible presence of occulted spots. The predicted variation in the apparent $R_p/R_\star$ due to the presence of unocculted spots (see §2.6.2) is shown (*grey*), calculated directly from the nightly-binned MEarth photometry (*points*) or the best-fit sine curves to those data (*lines*).

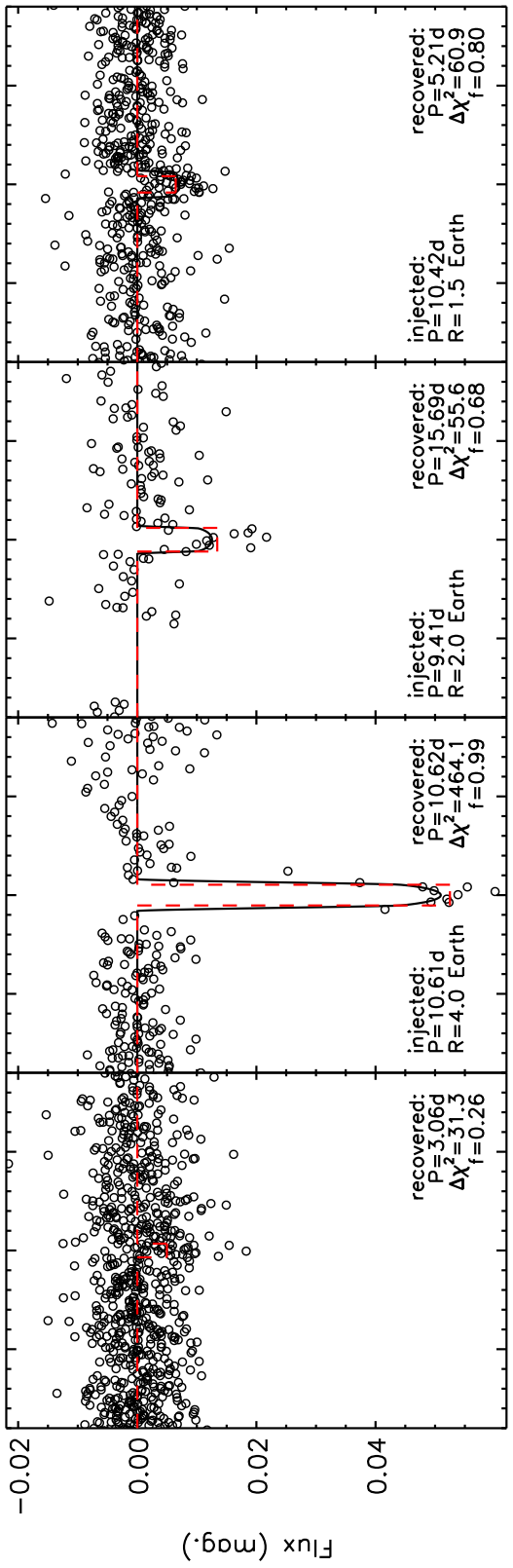

Figure 2.9: *Left:* The best candidate returned by the BLS search of the cleaned GJ1214 light curve. Bootstrap searches of simulated noise show the achieved $\Delta\chi^2$=31.3 has a large 20% probability of false alarm. *Right three:* Examples of simulated transiting planets with periods near 10 days where the transits injected (*black lines*) into prefiltered light curves have crossed the detection threshold in the BLS search. All four panels show light curves phased to the period of the best fit BLS result (*red dashed lines*); the BLS does not necessarily recover the correct period.

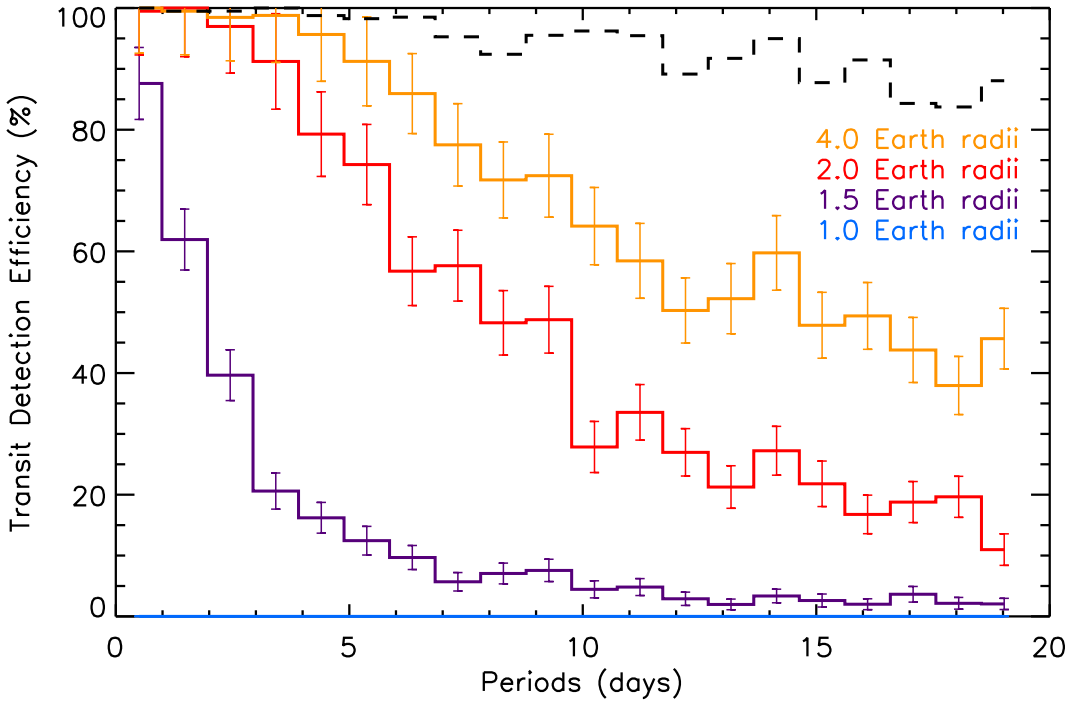

Figure 2.10: The recovery fraction as a function of period for the labeled planetary radii (*solid, colored lines*) assuming the detection criteria $\Delta\chi^2 > 50$ and $f < 1.0$. For the 4.0 $R_\oplus$ case, we also show the recovery fraction after lifting the $f < 1.0$ constraint (*black dashed line*); this is an estimate of the sensitivity to deep transits where only one event is necessary for a robust detection.

2.6 Discussion

2.6.1 GJ1214 as a Spotted Star

The slow rotation period we find implies a projected rotational velocity of 0.2 km s^{-1} that is well below the $v \sin i \simeq 1$ km s^{-1} detection limit of high-resolution rotation studies (Browning et al. 2010; Delfosse et al. 1998; Reiners 2007; West & Basri 2009) but not inconsistent with long photometric periods detected for other field M dwarfs. Benedict et al. (1998) estimated a rotation period of 83 days for Proxima Cen, and recent photometric work with ASAS (Kiraga & Stepien 2007), HATnet (Hartman et al. 2010) and MEarth (Irwin et al. 2011a) has confirmed the presence of many field M dwarfs with $P_{rot} > 10$ days.

Our P_{rot} further implies that GJ1214 should exhibit signs of only weak magnetic activity (e.g. Reiners & Basri 2007; Reiners 2007). Indeed, across three seasons of photometric monitoring, we see no evidence for flares in the MEarth bandpass, although their amplitude would be expected to be small in the near-IR. Activity induced chromospheric emission is not detected in either Hα or the Na I D doublet in the HARPS spectra used to measure the radial velocities presented in Charbonneau et al. (2009). The relation between magnetic activity and kinematic age (West et al. 2008) suggests that GJ1214 is > 3 Gyr old.

We calculate GJ1214's (U, V, W) space velocities (Johnson & Soderblom 1987) in a left-handed system where $U > 0$ in the direction of the Galactic anti-center to be (-47,-4,-40) km s^{-1}. These motions are consistent with membership in the Galactic old disk (Leggett 1992), lending further credence to an old age for GJ1214.

2.6.2 Implications for Transmission Spectroscopy

When inferring the transmission spectrum of a planet, one hopes to attribute changes in the transit depth across different wavelengths to atmospheric absorption by the limb of the planet. If the transmission spectrum is sensitive to 5 scale heights (H) of the planetary atmosphere, the amplitude of the transit depth variations are $\Delta D_{\rm planet}(\lambda) = 10HR_p/R_\star^2$ or 0.001 if GJ1214 has an hydrogen-rich atmosphere (Miller-Ricci & Fortney 2010). However, the presence of unocculted spots on the stellar surface can introduce transit depth variations $\Delta D_{\rm spots}(\lambda, t)$ that are a function of both wavelength and time. To aid ongoing and future work to study GJ1214b's atmosphere, we use a simple model to estimate the amplitude of the spot-induced contamination $\Delta D_{\rm spots}(\lambda, t)$.

We assume a fraction $s(t)$ of the star's Earth-facing hemisphere is covered with spots; $s(t)$ will change as the star rotates and the spots evolve. The observed out-of-transit spectrum $F_{\rm o.o.t.}(\lambda, t)$ is a weighted average of the spectrum of the unspotted photosphere $F_\circ(\lambda)$ and that of the presumably cooler spotted surface $F_\bullet(\lambda)$:

$$F_{\rm o.o.t.}(\lambda, t) = [1 - s(t)]\, F_\circ(\lambda) + s(t) F_\bullet(\lambda). \tag{2.6}$$

To simplify calculations, we neglect limb-darkening and treat each of the two components as having uniform surface brightness; tests with a limb-darkened spot model (Dorren 1987) indicate only pathological cases could change the following results by more than 10%.

When a planet with no atmosphere blocks light across a spot-free transit chord, it changes the relative weight of the two sources, causing the observed in-transit spectrum

$F_{\rm i.t.}(\lambda,t)$ to shift away from that of the unspotted photosphere toward that of the spots:

$$\begin{aligned}\frac{F_{\rm i.t.}(\lambda,t)}{F_{\rm o.o.t.}(\lambda,t)} &= 1-D(\lambda,t) \\ &= 1-\left(\frac{R_p}{R_\star}\right)^2-\Delta D_{\rm spots}(\lambda,t) \\ &= \frac{\left[1-s(t)-\left(\frac{R_p}{R_\star}\right)^2\right]F_\circ(\lambda)+s(t)F_\bullet(\lambda)}{\left[1-s(t)\right]F_\circ(\lambda)+s(t)F_\bullet(\lambda)}.\end{aligned} \tag{2.7}$$

Making the assumption that the total fraction of flux lost to the presence of spots is small ($s(t)[1-F_\bullet(\lambda)/F_\circ(\lambda)] << 1$), we solve to find

$$\Delta D_{spots}(\lambda,t) \approx s(t)\left[1-\frac{F_\bullet(\lambda)}{F_\circ(\lambda)}\right]\times\left(\frac{R_p}{R_\star}\right)^2. \tag{2.8}$$

If the quantity $s(t)[1-F_\bullet(\lambda)/F_\circ(\lambda)]$ were not small, we probably would have observed larger amplitude and more frequent spot occultation events in the transit light curves in §2.4. We note the significant possibility that $s(t)$ never reaches 0; that is, there may persist a population of symmetrically distributed spots that never rotates out of view. Such an unchanging population could cause us to overestimate the true value of $R_p/R_\star$ by up to several percent.

To match our observations of GJ1214's variability ($\Delta F_{\rm o.o.t.}(\lambda,t)/\overline{F_{\rm o.o.t.}}$) to this model, we write $s(t)=\overline{s}+\Delta s(t)$ where $\overline{s}$ is the mean Earth-facing spot covering fraction and $\Delta s(t)$ can be positive or negative. With the same assumption as above, we find

$$\frac{\Delta F_{\rm o.o.t.}(\lambda,t)}{\overline{F_{\rm o.o.t.}}} \approx -\Delta s(t)\left(1-\frac{F_\bullet(\lambda)}{F_\circ(\lambda)}\right). \tag{2.9}$$

In §2.3 we measured $\Delta F_{\rm o.o.t.}(\lambda,t)/\overline{F_{\rm o.o.t.}}$ to have a peak-to-peak amplitude of 1% in the MEarth bandpass ($715<\lambda<1000$ nm). Eq. 2.8 and 2.9 are equivalent to assuming a value of $\alpha=-1$ in Désert et al. (2011b)'s αf_λ formalism.

Désert et al. (2011b) show that measurements that rely on comparing photometric transit depths across multiple transits could potentially mistake time-variability of $\Delta D_{\rm spots}(t)$ for a feature in the transmission spectrum. From Eq. 2.8 and 2.9, we estimate the peak-to-peak time-variability of $\Delta D_{\rm spots}(t)$ to have an amplitude of 0.0001 in MEarth wavelengths over the rotation period of the star. This spot-induced variability is comparable to, but smaller than, the estimated uncertainty from the transits analyzed in this work and corresponds to an apparent change in planetary radius of $\Delta R_p = 70$ km or 1/2 the scale height of an H_2-dominated atmosphere on GJ1214b (Miller-Ricci & Fortney 2010). In Fig. 2.8 we show the expected variation in the apparent planetary radius from unocculted spots and our individual $R_p/R_\star$ measurements.

Using blackbody spectra for $F_\bullet(\lambda)$ and $F_\circ(\lambda)$, we can extrapolate with Eq. 2.8 from the MEarth observations ($\lambda \approx 0.85$ μm) to wavelengths accessible to *Warm Spitzer*. If we assume the spots are 300K cooler than the $T_{\rm eff} = 3000$K stellar photosphere, we find $\Delta D(t)$ variability amplitudes of 0.00004 and 0.00003 in *Spitzer*'s 3.6 and 4.5 μm bandpasses. This allows robust comparison of transit depths between these wavelengths (Désert et al., in prep.). In the conservative limit that $T_\bullet = 0$K, the variability amplitude would be achromatic.

Even when comparison across different epochs is unnecessary, as is the case for spectroscopic observations of individual transits (e.g. Charbonneau et al. 2002; Pont et al. 2008), unocculted spots can still introduce spurious wavelength features into the transmission spectrum. To place an upper limit on the amplitude of chromatic changes in $\Delta D_{\rm spots}(\lambda)$ within a given transit, we imagine an extreme scenario where $F_\bullet(\lambda)$ is identical to $F_\circ(\lambda)$ but with a very deep absorption line that does not appear in the unspotted spectrum, so the first factor in Eq. 2.8 can vary with λ between 0 and s.

Making the fairly conservative assumption that the population of spots that rotates in and out of view is comparable to the symmetric population ($s < 4\,\max[\Delta s]$) we find that $\Delta D_{\rm spots}(\lambda) < 0.0003$ for GJ1214 at wavelengths near 1 μm. In practice, most features will show significantly lower amplitudes, although precise calculations of them and extrapolation to other wavelengths will require knowledge of the spot temperature and reliable model atmospheres in the 2500-3000K temperature regime.

Given the sensitivity of current generation instruments, the known population of spots on GJ1214 do not pose a significant problem for ongoing studies of GJ1214b's atmosphere. Future transmission spectroscopy studies of GJ1214b, perhaps with the *James Webb Space Telescope*, comparing multiepoch, multiwavelength transit depths and aiming to reach a precision of $\sigma_{D(\lambda)} = 0.0001$ (see Deming et al. 2009) will have to monitor and correct for the stellar variability.

2.6.3 Limiting Uncertainties of GJ1214b

As shown in Fig. 2.8, stellar spots currently play a very small role in limiting our understanding of the bulk mass and radius of GJ1214b. Here we address what other factors provide the limiting uncertainties in the planet's physical parameters.

The 0.98 $M_\oplus$ (15%) uncertainty on the planetary mass M_p is the quadrature sum of 0.85 $M_\oplus$ propagated from the measured radial velocity semiamplitude and 0.5 $M_\oplus$ from the stellar mass uncertainty. Further spectroscopic monitoring may reduce the former, but to fully reap the benefits of the radial velocities, the 12% error on $M_\star$ should be improved. The current mass estimate is derived from an 2MASS photometry (2% uncertainty), an empirical M_K-mass relation ($\sim$ 10% scatter; Delfosse et al. 2000)

and the published system parallax (77.2 ± 5.4 mas; van Altena et al. 1995). Both improving M_K-mass relation for low-mass dwarfs and confirming GJ1214's parallax will be necessary to reduce GJ1214b's mass uncertainty.

The 0.12 $R_\oplus$ (5%) uncertainty on the planetary radius R_p is already dominated by the 5% uncertainty in the stellar radius $R_\star$, rather than the light curve parameters. This is currently constrained by the combination of the stellar density $\rho_\star$ that is measured directly from transit light curves (Seager & Mallén-Ornelas 2003) and the estimated mass. Even though $R_\star \propto (M_\star/\rho_\star)^{1/3}$, and so is relatively insensitive to uncertainty in $M_\star$, if we fix $M_\star = 0.157 \pm 0.019 M_\odot$ we find the errors on $R_\star$ and R_p shrink by a factor of two. Improving the estimate of GJ1214's mass is also the best way to improve our measurement of the radius of the planet.

We reiterate here that our measured out-of-transit photometric modulation probes only the spatially asymmetric component of the stellar spot distribution rotating around the star. If the star hosts a subtantial, unchanging, spatially symmetric population of unocculted spots, it will bias estimates of the true planetary radius (Czesla et al. 2009) too high. If the symmetric and asymmetric components of the spots are comparable, such a bias will be at the percent level of in R_p, smaller than the current uncertainties.

2.6.4 Metallicity of GJ1214

Several authors have recently developed empirical photometric calibrations to estimate M dwarf metallicities from absolute M_K magnitude and $V-K$ color (Bonfils et al. 2005a; Johnson & Apps 2009; Schlaufman & Laughlin 2010). As exoplanet surveys like MEarth lavish more attention on M dwarfs as exoplanet hosts, such studies hope to address

whether the giant planet vs. stellar metallicity correlation seen for AFGK stellar hosts (Fischer & Valenti 2005; Johnson et al. 2010) extends to smaller planets and smaller stars. Sousa et al. (2008) suggest the correlation does not persist down to Neptune-mass planets, but more data are needed.

Our improved estimate of $V = 14.71 \pm 0.03$ differs significantly from the $V = 15.1$ central value published in Charbonneau et al. (2009) but was the value used in the most recent analysis of this photometric metallicity calibration by Schlaufman & Laughlin (2010), who found [Fe/H] = +0.28 for GJ1214. This analysis agrees quite well with work by Rojas-Ayala et al. (2010) that estimates empirically calibrated metallicities from alkali metal lines in moderate resolution K-band spectra and finds [Fe/H] $= +0.39 \pm 0.15$ for GJ1214. This lends incremental evidence towards the persistence of the mass-metallicity correlation down to super-Earths around M dwarf hosts.

2.7 Conclusions

We have measured long-term photometric variability on GJ1214 to have a 1% peak-to-peak amplitude in the MEarth bandpass (715-1000 nm) and a long rotation period, most likely an integer multiple of 53 days. Fitting very high precision light curves from the VLT, we find likely instances of GJ1214b crossing small spots during transit. Treating these occultation events as correlated noise, we find parameters for the planetary system that are consistent with previous work.

We estimate the amplitude of time-variable changes in the apparent radius of the planet due to the observed stellar variability as $\Delta D_{\rm spots}(t) = 0.0001$ and place an upper

limit of $\Delta D_{\rm spots}(\lambda) < 0.0003$ on possible spot-induced spectral features in the planet's transmission spectrum. Stellar spots do not limit current studies (e.g. Bean et al. 2010a; Désert et al. 2011a), but could be important for future studies of GJ1214b with *JWST*.

Using two years of MEarth data, we have placed limits on the presence of other transiting planets around GJ1214. With 90% confidence, we rule out the presence of Neptune-radius transiting planets in orbits shorter than 10 days but cannot place strong constraints on planets smaller than 2.0 $R_\oplus$ at such long periods. In a system where a 1.0$M_\oplus$ planet in a 2:1 mean motion resonance would create 100 second perturbations to GJ1214b's transit times, we find no evidence for transit timing variations larger than 15 seconds. Further searches of the GJ1214 system for potentially habitable planets smaller and cooler than GJ1214b continue to be warranted.

Acknowledgements

We are extremely grateful to the KeplerCam observers Margaret Mclean, Gil Esquerdo, Mark Everett, Pete Challis, and Joel Hartman who collected V-band monitoring observations, as well as Gaspar Bakos, Dave Latham, and Matt Holman for donated time and Lars Buchhave and Xavier Bonfils for correcting the sign in the published γ velocity. We thank the referee for thoughtful and helpful comments that improved the paper significantly. J.B. acknowledges funding from NASA through the Sagan Fellowship Program. The MEarth team gratefully acknowledges funding from the David and Lucile Packard Fellowship for Science and Engineering (awarded to D.C.). This material is based upon work supported by the National Science Foundation under grant number AST-0807690. The MEarth team is greatly indebted to the staff at the Fred Lawrence

Whipple Observatory for their efforts in construction and maintenance of the facility, and would like to explicitly thank Wayne Peters, Ted Groner, Karen Erdman-Myres, Grace Alegria, Rodger Harris, Bob Hutchins, Dave Martina, Dennis Jankovsky and Tom Welsh for their support.

Chapter 3

The Flat Transmission Spectrum of the Super-Earth GJ1214b from Wide Field Camera 3 on the Hubble Space Telescope

This thesis chapter originally appeared in the literature as

Z. K. Berta, D. Charbonneau, J.–M. Désert, E. Miller-Ricci Kempton, P. R. McCullough, C. J. Burke, J. J. Fortney, J. Irwin, P. Nutzman, D Homeier, *The Astrophysical Journal*, 747, 35, 2012

Abstract

Capitalizing on the observational advantage offered by its tiny M dwarf host, we present HST/WFC3 grism measurements of the transmission spectrum of the super-Earth exoplanet GJ1214b. These are the first published WFC3 observations of a transiting exoplanet atmosphere. After correcting for a ramp-like instrumental systematic, we achieve nearly photon-limited precision in these observations, finding the transmission spectrum of GJ1214b to be flat between 1.1 and 1.7 μm. Inconsistent with a cloud-free solar composition atmosphere at 8.2σ, the measured achromatic transit depth most likely implies a large mean molecular weight for GJ1214b's outer envelope. A dense atmosphere rules out bulk compositions for GJ1214b that explain its large radius by the presence of a very low density gas layer surrounding the planet. High-altitude clouds can alternatively explain the flat transmission spectrum, but they would need to be optically thick up to 10 mbar or consist of particles with a range of sizes approaching 1 μm in diameter.

3.1 Introduction

With a radius of 2.7 $R_\oplus$ and a mass of 6.5 $M_\oplus$, the transiting planet GJ1214b (Charbonneau et al. 2009) is a member of the growing population of exoplanets whose masses and radii are known to be between those of Earth and Neptune (see Léger et al. 2009; Batalha et al. 2011; Lissauer et al. 2011; Winn et al. 2011). Among these exoplanets, most of which exhibit such shallow transits that they require ultra-precise space-based photometry simply to detect the existence of their transits, GJ1214b is

unique. The diminutive 0.21 $R_\odot$ radius of its M dwarf stellar host means GJ1214b exhibits a large 1.4% transit depth, and the system's proximity (13 pc) means the star is bright enough in the near infrared ($H = 9.1$) that follow-up observations to study the planet's atmosphere are currently feasible. In this work, we exploit this observational advantage and present new measurements of the planet's atmosphere, which bear upon models for its interior composition and structure.

According to theoretical studies (Seager et al. 2007; Rogers & Seager 2010b; Nettelmann et al. 2011), GJ1214b's 1.9 g cm^{-3} bulk density is high enough to require a larger ice or rock core fraction than the solar system ice giants but far too low to be explained with an entirely Earth-like composition. Rogers & Seager (2010b) have proposed three general scenarios consistent with GJ1214b's large radius, where the planet could (i) have accreted and maintained a nebular H_2/He envelope atop an ice and rock core, (ii) consist of a rocky planet with an H_2-rich envelope that formed by recent outgassing, or (iii) contain a large fraction of water in its interior surrounded by a dense H_2-depleted, H_2O-rich atmosphere. Detailed thermal evolution calculations by Nettelmann et al. (2011) disfavor this last model on the basis that it would require unreasonably large bulk water-to-rock ratios, arguing for at least a partial H_2/He envelope, albeit one that might be heavily enriched in H_2O relative to the primordial nebula.

By measuring GJ1214b's transmission spectrum, we can empirically constrain the mean molecular weight of the planet's atmosphere, thus distinguishing among these possibilities. When the planet passes in front of its host M dwarf, a small fraction of the star's light passes through the upper layers of the planet's atmosphere before reaching us; the planet's transmission spectrum is then manifested in variations of the transit depth

as a function of wavelength. The amplitude of the transit depth variations $\Delta D(\lambda)$ in the transmission spectrum scale as $n_H \times 2HR_p/R_\star^2$, where n_H is set by the opacities involved and can be 1-10 for strong absorption features, H is the atmospheric scale height, R_p is the planetary radius, and $R_\star$ is the stellar radius (e.g. Seager & Sasselov 2000; Brown 2001; Hubbard et al. 2001). Because the scale height H is inversely proportional to the mean molecular weight μ of the atmosphere, the amplitude of features seen in the planet's transmission spectrum places strong constraints on the possible values of μ and, in particular, the hydrogen/helium content of the atmosphere (Miller-Ricci et al. 2009).

Indeed, detailed radiative transfer simulations of GJ1214b's atmosphere (Miller-Ricci & Fortney 2010) show that a solar composition, H_2-dominated atmosphere ($\mu = 2.4$) would show depth variations of roughly 0.1% between 0.6 and 10 μm, while the features in an H_2O-dominated atmosphere ($\mu = 18$) would be an order of magnitude smaller. While the latter of these is likely too small to detect directly with current instruments, the former is at a level that has regularly been measured with the Hubble Space Telescope (HST) in the transmission spectra of hot Jupiters (e.g. Charbonneau et al. 2002; Pont et al. 2008; Sing et al. 2011).

Spectroscopic observations by Bean et al. (2010a) with the Very Large Telescope found the transmission spectrum of GJ1214b to be featureless between 0.78-1.0 μm, down to an amplitude that would rule out cloud-free H_2-rich atmospheric models. Broadband Spizer Space Telescope photometric transit measurements at 3.6 and 4.5 μm by Désert et al. (2011a) showed a flat spectrum consistent with Bean et al. (2010a), as did high-resolution spectroscopy with NIRSPEC between 2.0 and 2.4 μm by Crossfield et al. (2011). Intriguingly, the transit depth in K-band (2.2 μm) was measured from CFHT by Croll et al. (2011) to be 0.1% deeper than at other wavelengths, which would

imply a H_2-rich atmosphere, in apparent contradiction to the other studies.

These seemingly incongruous observations could potentially be brought into agreement if GJ1214b's atmosphere were H_2-rich but significantly depleted in CH_4 (Crossfield et al. 2011; Miller-Ricci Kempton et al. 2012). In such a scenario, the molecular features that remain (predominantly H_2O) would fit the CFHT measurement, but be unseen by the NIRSPEC and Spitzer observations. Explaining the flat VLT spectrum in this context would then require a broadband haze to smooth the spectrum at shorter wavelengths (see Miller-Ricci Kempton et al. 2012).. New observations by Bean et al. (2011) covering 0.6-0.85 μm and 2.0-2.3 μm were again consistent with a flat spectrum, but they still could not directly speak to this possibility of a methane-depleted, H_2-rich atmosphere with optically scattering hazes.

Here, we present a new transmission spectrum of GJ1214b spanning 1.1 to 1.7 μm, using the infrared slitless spectroscopy mode on the newly installed Wide Field Camera 3 (WFC3) aboard the Hubble Space Telescope (HST). Our WFC3 observations directly probe the predicted strong 1.15 and 1.4 μm water absorption features in GJ1214b's atmosphere (Miller-Ricci & Fortney 2010) and provide a stringent constraint on the H_2 content of GJ1214b's atmosphere that is robust to non-equilibrium methane abundances and hence a definite test of the CH_4-depleted hypothesis. The features probed by WFC3 are the same features that define the J and H band windows in the telluric spectrum, and cannot be observed from the ground.

Because this is the first published analysis of WFC3 observations of a transiting exoplanet, we include a detailed discussion of the performance of WFC3 in this observational regime and the systematic effects that are inherent to the instrument.

Recent work on WFC3's predecessor NICMOS (Burke et al. 2010; Gibson et al. 2011) has highlighted the importance of characterizing instrumental systematics when interpreting exoplanet results from HST observations.

This paper is organized as follows: we describe our observations in §3.2, our method for extracting spectrophotometric light curves from them in §3.3, and our analysis of these light curves in §3.4. We present the resulting transmission spectrum and discuss its implications for GJ1214b's composition in §3.5, and conclude in §3.6.

3.2 Observations

We observed three transits of GJ1214b on UT 2010 October 8, 2011 March 28, and 2011 July 23 with the G141 grism on WFC3's infrared channel (HST Proposal #GO-12251, P.I. = Z. Berta), obtaining simultaneous multiwavelength spectrophotometry of each transit between 1.1 and 1.7 μm. WFC3's IR channel consists of a 1024×1024 pixel Teledyne HgCdTe detector with a 1.7 μm cutoff that can be paired with any of 15 filters or 2 low-resolution grisms (Dressel et al. 2010). Each exposure is compiled from multiple non-destructive readouts and can consist of either the full array or a concentric, smaller subarray.

Each visit consisted of four 96 minute long HST orbits, each containing 45 minute gaps due to Earth occultations. Instrumental overheads between the occultations are dominated by serial downloads of the WFC3 image buffer, during which all science images are transferred to the telescope's solid state recorder. This buffer can hold only two 16-readout, full-frame IR exposures before requiring a download, which takes 6

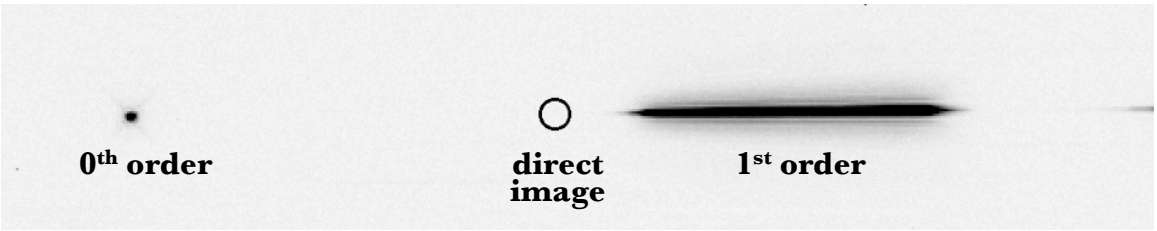

Figure 3.1: A 512x100 pixel cutout of a typical WFC3 G141 grism exposure of the star GJ1214. The 0^{th} and 1^{st} order spectra are labeled, and the start of the 2^{nd} order spectrum is visible on the right. The location of the star in the direct images (not shown here) is marked with a circle.

minutes. Exposures cannot be started nor stopped during a buffer download, so parallel buffer downloads are impossible for short exposures.

Subject to these constraints and the possible readout sequences, we maximized the number of photons detected per orbit while avoiding saturation by gathering exposures using the 512×512 subarray with the RAPID NSAMP=7 readout sequence, for an effective integration time of 5.971 seconds per exposure. With this setting, four 12-exposure batches, separated by buffer downloads, were gathered per orbit resulting in an integration efficiency of 10%. Although the brightest pixel in the 1^{st} order spectrum reaches 78% of saturation during this exposure time, the WFC3's multiple non-destructive readouts enable the flux within each pixel to be estimated before the onset of significant near-saturation nonlinearities.

A sub-region of a typical G141 grism image of GJ1214 is shown in Figure 3.1. The 512×512 subarray allows both the 0^{th} and 1^{st} order spectra to be recorded, and the 1^{st} order spectrum to fall entirely within a single amplifier quadrant of the detector. The 1^{st} order spectrum spans 150 pixels with a dispersion in the x-direction of 4.65 nm/pixel and a spatial full-width half maximum in the y-direction of 1.7 pixels (0.2"). The 0^{th} order spectrum is slightly dispersed by the grism's prism but is nearly a point source. Other stars are present in the subarray's 68" $\times$ 61" field of view, but are too faint to provide useful diagnostics of systematic trends that may exist in the data. For wavelength calibration, we gathered direct images in the F130N narrow-band filter; the direct images' position relative to the grism images is also shown in Figure 3.1.

To avoid systematics from the detector flat-fields that have a quoted precision no better than 0.5% (Pirzkal et al. 2011), the telescope was not dithered during any of the observations. We note that a technique called "spatial scanning" has been proposed to decrease the overheads for bright targets with WFC3, where the telescope nods *during* an exposure to smear the light along the cross-dispersion direction, thus increasing the time to saturation (McCullough & MacKenty 2011). We did not use this mode of observation as it was not yet tested at the time our program was initiated.

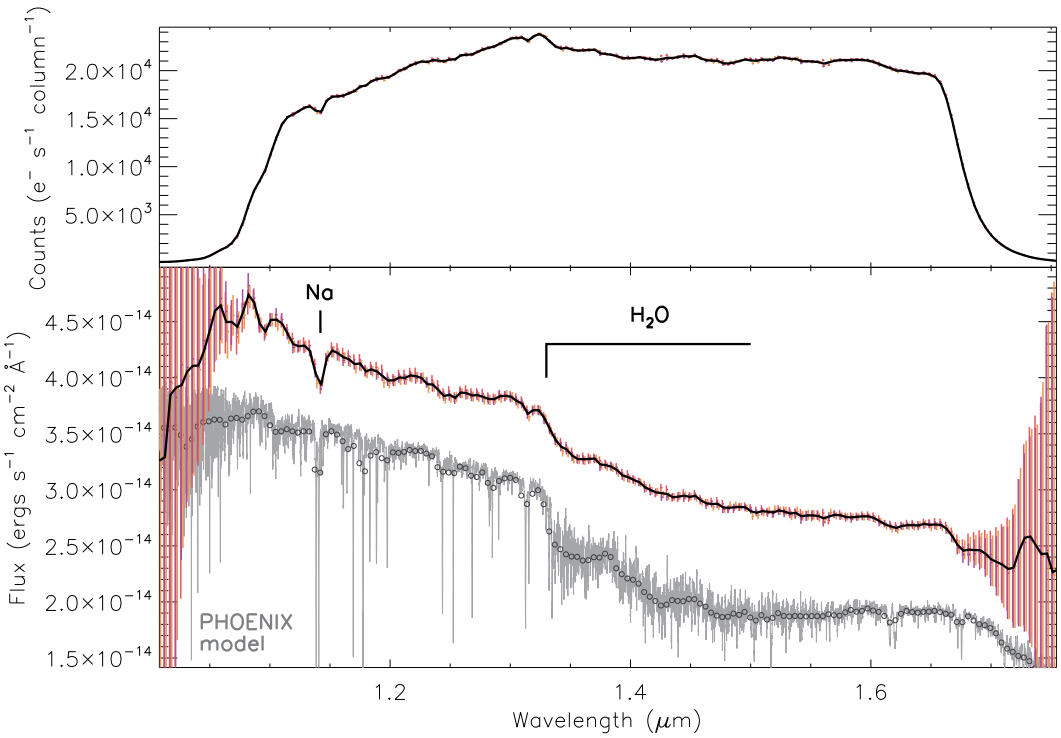

Figure 3.2: The mean out-of-transit extracted spectrum of GJ1214 (*black line*) from all three HST visits, shown before (top) and after (bottom) flux calibration. Individual extracted spectra from each visit are shown with their 1σ uncertainties (*color error bars*). For comparison, the integrated flux from the PHOENIX model atmosphere used to calculate the stellar limb darkening (see §3.4.2) is shown (*gray lines*) offset for clarity and binned to the WFC3 pixel scale (*gray circles*).

3.3 Data Reduction

The Python/PyRAF software package aXe was developed to extract spectra from slitless grism observations with WFC3 and other Hubble instruments (Kümmel et al. 2009), but it is optimized for extracting large numbers of spectra from full frame dithered grism images. To produce relative spectrophotometric measurements of our single bright source, we opted to create our own extraction pipeline that prioritizes precision in the time domain. We outline the extraction procedure below.

Through the extraction, we use calibrated 2-dimensional images, the "`flt`" outputs from WFC3's `calwf3` pipeline. For each exposure, `calwf3` performs the following steps: flag detector pixels with the appropriate data quality (DQ) warnings, estimate and remove bias drifts using the reference pixels, subtract dark current, determine count rates and identify cosmic rays by fitting a slope to the non-destructive reads, correct for photometric non-linearity (properly accounting for the signal accumulation before the initial "zeroth" read), and apply gain calibration. The resulting images are measured in $e^{-}s^{-1}$ and contain per pixel uncertainty estimates based on a detector model (Kim Quijano et al. 2009). We note that `calwf3` does *not* apply flat-field corrections when calibrating grism images; proper wavelength-dependent flat-fielding for slitless spectroscopy requires wavelength-calibrating individual sources and `calwf3` does not perform this task.

3.3.1 Interpolating over Cosmic Rays

`calwf3` identifies cosmic rays that appear partway through an exposure by looking for

deviations from a linear accumulation of charge among the non-destructive readouts, but it can not identify cosmic rays that appear between the zeroth and first readout. We supplement `calwf3`'s cosmic ray identifications by also flagging any pixel in an individual exposure that is $> 6\sigma$ above the median of that pixel's value in all other exposures as a cosmic ray. Through all three visits (576 exposures), a total of 88 cosmic rays were identified within the extraction box for the 1^{st} order spectra.

For each exposure, we spatially interpolate over cosmic rays. Near the 1^{st} order spectrum, the pixel-to-pixel gradient of the point spread function (PSF) is typically much shallower along the dispersion direction than perpendicular to it, so we use only horizontally adjacent pixels when interpolating to avoid errors in modeling the sharp cross-dispersion falloff.

3.3.2 Identifying Continuously Bad Pixels

We also mask any pixels that are identified as "bad detector pixels" (DQ=4), "unstable response" (DQ=32), "bad or uncertain flat value" (DQ=512). We found that only these DQ flags affected the photometry in a pixel by more than 1σ. Other flags may have influenced the pixel photometry, but did so below the level of the photon noise. In the second visit, we also identified one column of the detector ($x = 625$ in physical pixels[1]) whose light curve exhibited a dramatically different systematic variation than did light curves from any of the other columns. This column was coincident with an unusually low-sensitivity feature in the flat-field, and we hypothesize that the flat-field is more

[1]For ease of comparison with future WFC3 analyses, throughout this paper we quote all pixel positions in physical units as interpreted by SAOImage DS9, where the bottom left pixel of a full-frame array would be $(x, y) = (1,1)$.

uncertain in this column than in neighboring columns. We masked all pixels in that column as bad.

We opt not to interpolate over these continuously bad pixels. Because they remained flagged throughout the duration of each visit, we simply give these pixels zero weight when extracting 1D spectra from the images. This allows us to keep track of the actual number of photons recorded in each exposure so we can better assess our predicted photometric uncertainties.

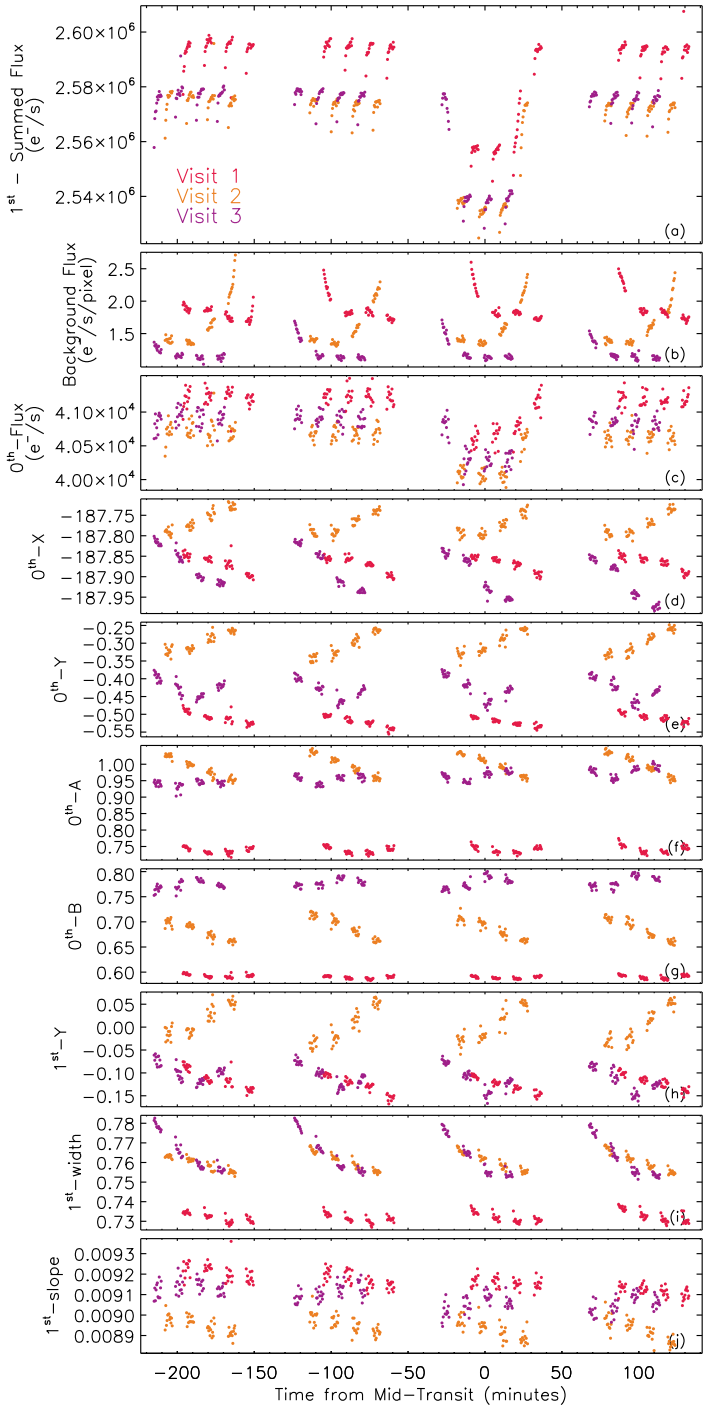

Figure 3.3: Extracted properties of the 0^{th} and 1^{st} order spectra as a function of time, including (a) the summed 1^{st} order photometric light curves; (b) the estimated sky background level; (c-g) the total flux, x and y position (measured relative to the reference pixel), and Gaussian widths in the x and y directions of the 0^{th} order image; and (h-j) the y offset, cross-dispersion width, and slope of the 1^{st} order spectrum. Geometric parameters are measured in physical pixels. All three visits are shown and are denoted by the color of the symbols. The 45 minute gaps in each time series are due to Earth occultations, the 6 minute gaps are due to the WFC3 buffer downloads.

3.3.3 Background Estimation

In addition to the target, WFC3 also detects light from the diffuse sky background, which comes predominantly from zodiacal light and Earth-shine, and must be subtracted. We draw conservative masks around all sources that are visible in each visit's median image, including GJ1214 and its electronics cross-talk artifact (see Viana & Baggett 2010). We exclude these pixels, as well as all pixels that have any DQ warning flagged. Then, to estimate the sky background in each exposure, we scale a master WFC3 grism sky image (Kümmel et al. 2011) to match the remaining 70-80% of the pixels in each exposure and subtract it. We find typical background levels of $1-3$ e$^-$ s^{-1} pixel^{-1}, that vary smoothly within orbits and throughout visits as shown in Figure 3.3 (panel b). As a test, we also estimated the background level from a simple mean of the unmasked pixels; the results were unchanged.

3.3.4 Inter-pixel Capacitance

The normal calibration pipeline does not correct for the inter-pixel capacitance (IPC) effect, which effectively couples the flux recorded in adjacent pixels at about the 1% level (McCullough 2008). We correct this effect with a linear deconvolution algorithm (McCullough 2008; Hilbert & McCullough 2011), although we find it makes little difference to the final results.

3.3.5 Extracting the Zeroth Order Image

The 0^{th} order image can act as a diagnostic for tracking changes in the telescope pointing and in the shape of the instrumental PSF. We select a 10×10 pixel box around the 0^{th} order image and fit a 2D Gaussian to it with the x position, y position, size in the x direction, size in the y direction, and total flux allowed to vary (5 parameters).

Time series of the 0^{th} order x and y positions, sizes in both directions, and total flux are shown for all three visits in Figure 3.3 (panels c-g). Thanks to the dispersion by the grism's prism, the Gaussian is typically 20% wider in the x direction than in the y direction. Even though the throughput of the 0^{th} order image is a factor of 60 lower than the 1^{st} order spectrum, the transit of GJ1214b is readily apparent in the 0^{th}-order flux time series.

3.3.6 Extracting the First Order Spectrum

To extract the first order spectra, we first determine the position of GJ1214 in the direct image, which serves as a reference position for defining the trace and wavelength calibration of the 1^{st} order spectrum. We adopt the mean position GJ1214 in all of the direct images as the reference position, which we measure using the same method as in extracting the 0^{th} order image in §3.3.5. The measured (x, y) reference positions for the first, second, and third visits are (498.0, 527.5), (498.6, 531.1), and (498.9, 527.1) in physical pixels.

Once the reference pixel for a visit is known, we use the coefficients stored in the

WFC3/G141 aXe configuration file[2] (Kuntschner et al. 2009), to determine the geometry of the 1^{st} order trace and cut out a 30 pixel tall extraction box centered on the trace. Within this extraction box, we use the wavelength calibration coefficients to determine the average wavelength of light that will be illuminating each pixel. We treat all pixels in the same column as having the same effective wavelength; given the spectrum's 0.5° tilt from to the x axis, errors introduced by this simplification are negligible.

Kuntschner et al. (2008) used flat-fields taken through all narrow-band filters available on WFC3/IR to construct a flat-field "cube" where each pixel contains 4 polynomial coefficients that describe its sensitivity as a function of wavelength[3]. We use this flat-field cube to construct a color-dependent flat based on our estimate of the effective wavelength illuminating each pixel, and divide each exposure by it. WFC3 wavelength calibration and flat-fielding is described in detail in the aXe manual (Kümmel et al. 2010).

To calculate 1D spectra from the flat-fielded images, we sum all the unmasked pixels within the extraction box over the y-axis. To estimate the uncertainty in each spectral channel, we first construct a per-pixel uncertainty model that includes photon noise from the source and sky as well as 22 e^- of read noise, and sum these uncertainties, in quadrature, over the y-axis. We do not use the `calwf3`-estimated uncertainties; they include a term propagated from the uncertainty in the nonlinearity correction that, while appropriate for absolute photometry, would not be appropriate for relative photometry. In each exposure, there are typically 1.2×10^5 e^- per single-pixel spectral channel and

[2]The aXe configuration file `WFC3.IR.G141.V2.0.conf` is available through `http://www.stsci.edu/hst/wfc3/`

[3]`WFC3.IR.G141.flat.2.fits`, through the same URL

a total of 1.5×10^7 e^- in the entire spectrum. Figure 3.3 (panel a) shows the extracted spectra summed over all wavelengths as a function of time, the "white" light curve.

For diagnostics' sake, we also measure the geometrical properties of the 1^{st} order spectra in each exposure. We fit one-dimensional Gaussians the cross-dispersion profile in each column of the spectrum and take the median Gaussian width among all the columns as a measurement of the PSF's width. We fit a line to the location of the Gaussian peaks in all the columns, taking the intercept and the slope of that line as an estimate of the y-offset and tilt of the spectrum on the detector. Time series of these parameters are shown in Figure 3.3 (panels h-j). These diagnostics are published in Table 1, along with the white light curve.

3.3.7 Flux Calibration

For the sake of display purposes only (see Figure 3.2), we flux calibrate each visit's median, extracted, 1D spectrum. Here we have interpolated over all bad pixels within each visit (contrary to the discussion in the §3.3.2), and plotted the weighted mean over all three visits. The calibration uncertainty for the G141 sensitivity curve (Kuntschner et al. 2011) is quoted to be 1%.

3.3.8 Times of Observations

For each exposure, we extract the `EXPSTART` keyword from the science header, which is the Modified Julian Date at the start of the exposure. We correct this to the mid-exposure time using the `EXPTIME` keyword, and convert it to the Barycentric Julian

Date in the Barycentric Dynamical Time standard using the code provided by Eastman et al. (2010).

Table 3.1. White Light Curves from WFC3/G141

Time[a] (BJD_{TDB})	Relative Flux[b]	Uncertainty	Sky (e^-/s)	0^{th}-X[c] (pix)	0^{th}-Y[c] (pix)	0^{th}-A[d] (pix)	0^{th}-B[d] (pix)	1^{st}-Y[c] (pix)	1^{st}-B[d] (pix)	1^{st}-Slope[e] (pix/pix)	Visit
2455478.439980	0.99381	0.00031	1.9546	-187.830	-0.474	0.790	0.613	-0.075	0.7484	0.00921	1
2455478.440270	0.99713	0.00031	1.9938	-187.839	-0.481	0.792	0.616	-0.082	0.7490	0.00925	1
2455478.440559	0.99787	0.00031	1.9718	-187.846	-0.484	0.789	0.615	-0.076	0.7486	0.00914	1
2455478.440848	0.99958	0.00032	1.8808	-187.827	-0.490	0.792	0.614	-0.087	0.7476	0.00917	1
2455478.441138	0.99989	0.00032	1.9379	-187.844	-0.490	0.785	0.612	-0.085	0.7492	0.00921	1
					...						

[a]Mid-exposure time.

[b]Normalized to the median flux level of the out-of-transit observations in each visit.

[c]Position measured relative to the Gaussian center of each visit's direct image.

[d]Gaussian width of the 0^{th} or 1^{st} order spectra in the horizontal (A) or vertical (B) direction.

[e]Slope of the 1^{st} order spectrum.

Note. — This table is published in its entirety in the electronic edition of the Astrophysical Journal. A portion is shown here for guidance concerning its form and content.

3.4 Analysis

In this section we describe our method for estimating parameter uncertainties (§3.4.1) and our strategy for modeling GJ1214's stellar limb darkening (§3.4.2). Then, after identifying the dominant systematics in WFC3 light curves (§3.4.3) and describing a method to correct them (§3.4.4), we present our fits to the light curves, both summed over wavelength (§3.4.5) and spectroscopically resolved (§3.4.6). We also present a fruitless search for transiting satellite companions to GJ1214b (§3.4.7).

3.4.1 Estimating Parameter Distributions

Throughout our analysis, we fit different WFC3 light curves with models that have different sets of parameters, and draw conclusions from the inferred probability distributions of those parameters; this section describes our method for characterizing the posterior probability distribution for a set of parameters within a given model.

We use a Markov Chain Monte Carlo (MCMC) method with the Metropolis-Hastings algorithm to explore the posterior probability density function (PDF) of the model parameters. This Bayesian technique allows us to sample from (and thus infer the shape of) the probability distribution of a model's parameters given both our data and our prior knowledge about the parameters (for reviews, see Ford 2005; Gregory 2005; Hogg et al. 2010). Briefly, the algorithm starts a chain with an initial set of parameters ($\mathbf{M}_{j=0}$) and generates a trial set of parameters ($\mathbf{M'}_{j+1}$) by perturbing the previous set. The ratio of posterior probability between the two parameter sets, given the data $\mathbf{D}$, is then

calculated as

$$\frac{P(\mathbf{M'}_{j+1}|\mathbf{D})}{P(\mathbf{M}_j|\mathbf{D})} = \frac{P(\mathbf{D}|\mathbf{M'}_{j+1})}{P(\mathbf{D}|\mathbf{M}_j)} \times \frac{P(\mathbf{M'}_{j+1})}{P(\mathbf{M}_j)} \tag{3.1}$$

where the first term (the "likelihood") accounts for the information that our data provide about the parameters and the second term (the "prior") specifies our externally conceived knowledge about the parameters. If a random number drawn from a uniform distribution between 0 and 1 is less than this probability ratio, then $\mathbf{M}_{j+1}$ is set to $\mathbf{M'}_{j+1}$; if not, then $\mathbf{M}_{j+1}$ reverts to $\mathbf{M}_j$. The process is iterated until j is large, and the resulting chain of parameter sets is a fair sample from the posterior PDF and can be used to estimate confidence intervals for each parameter.

To calculate the likelihood term in Eq. 3.1, we assume that each of the N flux values d_i is drawn from a uncorrelated Gaussian distribution centered on the model value m_i with a standard deviation of $s\sigma_i$, where σ_i is the theoretical uncertainty for the flux measurement based on the detector model and photon statistics and s is a photometric uncertainty rescaling parameter. Calculation of the ratio in Eq. 3.1 is best done in logarithmic space for numerical stability, so we write the likelihood as

$$\ln P(\mathbf{D}|\mathbf{M}) = -N \ln s - \frac{1}{2s^2}\chi^2 + \text{constant} \tag{3.2}$$

where

$$\chi^2 = \sum_{i=1}^{N} \left(\frac{d_i - m_i}{\sigma_i} \right)^2 \tag{3.3}$$

and we have only explicitly displayed terms that depend on the model parameters. Including s as a model parameter is akin to rescaling the uncertainties by externally modifying σ_i to achieve a reduced χ^2 of unity, but enables the MCMC to fit for and marginalize over this rescaling automatically. Unless otherwise stated for specific parameters, we use non-informative (uniform) priors for the second term in Eq. 3.1. We

use a Jeffreys prior on s (uniform in $\ln s$) which is the least informative, although the results are practically indistinguishable from prior uniform in s.

When generating each new trial parameter set $\mathbf{M'}_{j+1}$, we follow Dunkley et al. (2005) and perturb every parameter at once, drawing the parameter jumps from a multivariate Gaussian with a covariance matrix that approximates that of the parameter distribution. Doing so allows the MCMC to move easily along the dominant linear correlations in parameter PDF, and greatly increases the efficiency of the algorithm. While this procedure may seem circular (if we knew the covariance matrix of the parameter distribution, why would we need to perform the MCMC?), the covariance matrix we use to generate trial parameters could be a very rough approximation to the true shape of the parameter PDF but still dramatically decrease the computation time necessary for the MCMC.

To obtain an initial guess for parameters ($\mathbf{M_{j=0}}$), we use the `MPFIT` implementation (Markwardt 2009) of the Levenberg-Marquardt (LM) method to maximize $\ln P(\mathbf{M}|\mathbf{D})$. This would be identical to minimizing χ^2 in the case of flat priors, but it can also include constraints from more informative priors. The LM fit also provides an estimate of the covariance matrix of the parameters, which is a linearization of the probability space near the best-fit. We use this covariance matrix estimate for generating trial parameters in the MCMC, and with it, achieve parameter acceptance rates of 10-40% throughout the following sections. As expected, when fitting models with flat priors and linear or nearly-linear parameters (where the PDF should well-described by a multivariate Gaussian), the LM covariance matrix is identical to that ultimately obtained from the MCMC (see Sivia & Skilling 2006, for further discussion).

MCMC chains are run until they contain 1.25×10^5 points. The first 1/5 of the points are ignored as "burn-in", leaving 1×10^5 for parameter estimation. Correlation lengths for the parameters in the MCMC chains are indicated throughout the text; they are typically of order 10 points. A chain with such a correlation length effectively contains $1 \times 10^5/10 = 1 \times 10^4$ independent realizations of the posterior PDF. We quote confidence intervals that exclude the upper and lower 16% of the marginalized distribution for each parameter (i.e. the parameter's central 68% confidence interval), using all 1×10^5 points in each chain.

3.4.2 Modeling Stellar Limb Darkening

Accurate modeling of the WFC3 integrated and spectroscopic transit light curves requires careful consideration of the stellar limb-darkening (LD) behavior. GJ1214b's M4.5V stellar host is so cool that it exhibits weak absorption features due to molecular H_2O. Because inferences of the planet's apparent radii from transit light curves depend strongly on the star's limb-darkening, which is clearly influenced by H_2O as an opacity source, inaccurate treatment of limb-darkening could potentially introduce spurious H_2O features into the transmission spectrum.

If they were sufficiently precise, transit light curves alone could simultaneously constrain both the star's multiwavelength limb-darkening behavior and the planet's multiwavelength radii (e.g. Knutson et al. 2007b). For less precise light curves, it is common practice to fix the limb-darkening to a theoretically calculated law, even if this may underestimate the uncertainty in the planetary parameters (see Burke et al. 2007; Southworth 2008). Given the quality of our data, we adopt an intermediate solution

where we allow the limb-darkening parameters to vary in our fits, but with a Gaussian prior centered on the theoretical values (e.g. Bean et al. 2010a).

We model the star GJ1214's limb-darkening behavior with a spherically symmetric `PHOENIX` atmosphere (Hauschildt et al. 1999), assuming stellar parameters of $T_{\rm eff} = 3026K$, $\log g = 5$, and [M/H] = 0 (Charbonneau et al. 2009). As shown in Figure 3.2, the integrated flux from the PHOENIX model is in good qualitative agreement with the low-resolution, calibrated WFC3 stellar spectrum of GJ1214. From this model, we calculate photon-weighted average intensity profiles for the integrated spectrum and for each of the individual wavelength bins, using the WFC3 grism sensitivity curve and the `PHOENIX` model to estimate the photon counts. In the spherical geometry of the PHOENIX atmospheres the characterization of the actual limb (defined as $\mu = 0$, see below) is not straightforward, as the model extends beyond the photosphere into the optically thin outer atmosphere. The result is an approximately exponentially declining intensity profile from the outermost layers, that Claret & Hauschildt (2003) found not to be easily reproduced by standard limb darkening laws for plane-parallel atmospheres. These authors suggest the use of "quasi-spherical" models by ignoring the outer region. In an extension of this concept, we set the outer surface of the star to be where the intensity drops to e^{-1} of the central intensity, and measure $\mu = \cos\theta$ (where θ is the emission angle relative to the line of sight) relative to that outer radius.

We derive coefficients for a square-root limb-darkening law for each of these average intensity profiles using least-squares fitting. In this law, the intensity relative to the center of the star is given by

$$\frac{I(\mu)}{I(1)} = 1 - c(1-\mu) - d(1-\sqrt{\mu}), \tag{3.4}$$

where c and d are the two coefficients of the fit. We chose a square-root law over the popular quadratic law because it gave noticeably better approximations to the `PHOENIX` intensity profiles, while still having few enough free parameters that they can be partially inferred from the data. Indeed, van Hamme (1993) found the square-root law to be generally preferable to other 2-parameter limb-darkening laws for late-type stars in the near-IR. The square-root law matches the theoretical intensity profile nearly as well as the full nonlinear 4-parameter law introduced by Claret (2000) for the models we use here.

3.4.3 Light Curve Systematics

The summed light curve shown in Figure 3.3 (panel a) exhibits non-astrophysical systematic trends. The most obvious of these are the sharply rising but quickly saturating "ramp"-like features within each batch of 12 exposures between buffer downloads. To the eye, the ramps are very repeatable; the flux at the end of all batches asymptotes to nearly the same level. The amplitude of the ramp is 0.4% from start to finish for most batches, except for the first batch of each orbit, where the ramp is somewhat less pronounced.

These ramps are reminiscent of those seen in high-cadence Spitzer light curves at 8 and 16 μm (e.g. Deming et al. 2006; Knutson et al. 2007a; Charbonneau et al. 2008) which Agol et al. (2010) recently proposed may be due to "charge trapping" within the detector pixels. In their toy model, charge traps within each pixel become filled throughout an exposure and later release the trapped charge on a finite timescale, thereby increasing the pixel's dark current in subsequent exposures. The model leads to

exponential ramps when observing bright sources as the excess dark current increases sharply at first but slows its increase as the population of charge traps begins to approach steady state. We note this model also leads to after-images following strong exposures, i.e., persistence.

WFC3 has been known since its initial ground-testing to exhibit strong persistence behavior (McCullough & Deustua 2008; Long et al. 2010). Smith et al. (2008a) have proposed that persistence in 1.7 μm cutoff HgCdTe detectors like WFC3 is likely related to charge trapping. Measurements (McCullough & Deustua 2008) indicate that WFC3's persistence may be of the right order of magnitude (on $<$ 1 minute timescales) to supply the roughly 50 e^- s^{-1} $pixel^{-1}$ in the brightest pixels that would be necessary to explain the observed several millimagnitude ramp, although persistence levels and decay timescales can depend in complicated ways on the strength of previous exposures (see Smith et al. 2008b).

We were aware of this persistence issue before our observations and made an effort to control its effect on our light curves. When we planned the timing of the exposures, we attempted to make the illumination history of each pixel as consistent as possible from batch to batch and orbit to orbit. In practice, this means we gathered more direct images than necessary for wavelength calibration to delay some of the grism exposures.

Whether or not the ramps are caused by the charge trapping mechanism, they are definitely dependent on the illumination that a pixel receives. To demonstrate this, we construct light curves for each individual pixel over the duration of every out-of-transit 12-exposure batch that follows a buffer download and normalize each of these pixel light curves to the first exposure in the batch. Figure 3.4 shows the normalized pixel light

curves, grouped by their mean recorded fluence. Because it takes a finite time to read the subarray (0.8 seconds) and reset the full array (2.9 seconds), we note that each exposure actually collects 60% more electrons than indicated by these nominal, recorded fluences (see Long et al. 2011). The appearance of the ramp clearly becomes more pronounced for pixels that are more strongly exposed.

Buried beneath the ramp features, the summed light curve exhibits subtler trends that appear mostly as orbit-long or visit-long slopes with a peak-to-peak variation of about 0.05%. These are perhaps caused by slow drifts in pointing and focus (telescope "breathing") interacting with sensitivity variations across the detector that are not perfectly corrected by the flat field.

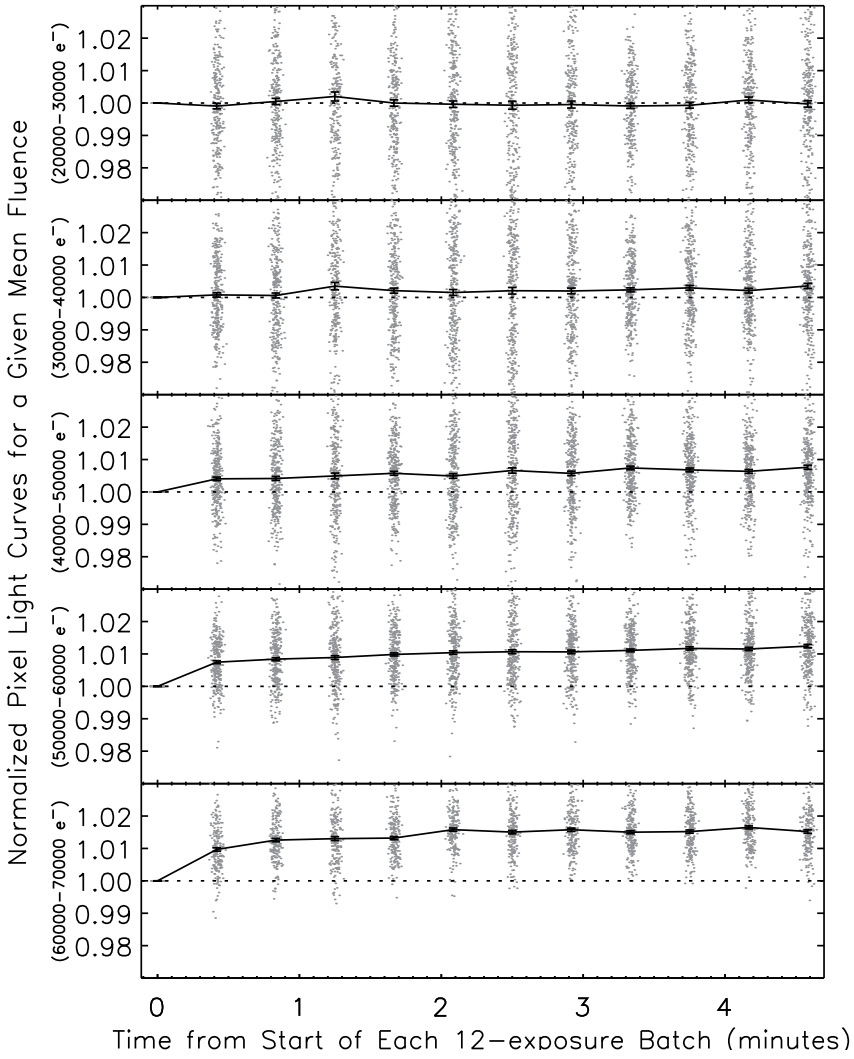

Figure 3.4: Single-pixel light curves within each 12-exposure batch following a buffer download (*gray points*), shown for different mean pixel illuminations. Pixel light curves have been normalized to the first exposure within each batch, and plotted with small random horizontal offsets for clarity. Only data from the first HST visit, which exhibited the smallest pointing drifts (see Figure 3.3), are shown. Error bars show the mean and its standard error for each time point and each illumination. An exponential ramp is apparent in the light curves of pixels that have a mean recorded fluence greater than 30,000 and 40,000 e^- (50% of the detector full well). Note, the nominal fluences quoted here do not include charge accumulated during detector flushing and initial readout (see text).

3.4.4 Correcting for Systematics

Fortunately, these systematics are extremely repeatable between orbits within a visit; we harness this fact when correcting for them. We divide the in-transit orbit of any photometric timeseries, either the white light curve or one of the spectroscopically resolved light curves, by a systematics correction template constructed from the two good out-of-transit orbits. This template is simply the weighted average of the fluxes in the out-of-transit orbits, evaluated at each exposure within an orbit. It encodes both variations in the effective sensitivity of the detector within an orbit and the mean out-of-transit flux level.

When performing the division, we propagate the template uncertainty into the photometric uncertainty for each exposure, which typically increases it by a factor of $\sqrt{1+1/2} = 1.22$. This factor, although it may seem like an undesired degradation of the photometric precision, would inevitably propagate into measurements of the transit depth whether we performed this correction or not, since $R_p/R_\star$ is always measured relative to the out-of-transit flux, which must at some point be inferred from the data.

Throughout this work, we refer to this process of dividing by the out-of-transit orbits as the `divide-oot` method. Because each point in the single in-transit orbit is equally spaced in time between the two out-of-transit exposures being used to correct it, the `divide-oot` method also naturally removes the 0.05% visit-long slope seen in the raw photometry. As we show in §3.4.5, when applied to the white light curves, the `divide-oot` treatment produces uncorrelated Gaussian residuals that have a scatter consistent with the predicted photon uncertainties. The white light curve published in Table 1 does not have this correction applied.

Unlike decorrelation techniques that have often been used to correct systematics in HST light curves, the `divide-oot` method does not require knowing the relationship between measured photometry and the physical state of the camera. It does, however, strictly require the systematics to repeat over multiple orbits. The `divide-oot` method would not work if the changes in the position, shape, and rotational angle of the 1^{st}-order spectrum were not repeated in the other orbits in a visit or if the cadence of the illumination were not nearly identical across orbits. In such cases, the Gaussian process method proposed by Gibson et al. (2012) may be a useful alternative, and one that would appropriately account for the uncertainty involved in the systematics correction.

3.4.5 White Light Curve Fits

Although the main scientific result of this paper is derived from the spectroscopic light curves presented in §3.4.6, we also analyze the light curve summed over all wavelengths between 1.1 and 1.7 μm. We use these white light curves to confirm the general system properties found in previous studies and quantitatively investigate the instrumental systematics.

We fit an analytic, limb-darkened transit light curve model (Mandel & Agol 2002) to the `divide-oot`-corrected white light curves. Only the in-transit orbits were fit; after the `divide-oot` correction, the two out-of-transit orbits contain no further information. Also, because the in-transit orbit's flux has already been normalized, we fix the out-of-transit flux level to unity in all the fits. Throughout, we fix the planet's period to $P = 1.58040481$ days and mid-transit time to $T_c = 2454966.525123$ $\mathrm{BJD_{TDB}}$ (Bean et al. 2011), the orbital eccentricity to $e = 0$, and the stellar mass to $0.157\mathrm{M}_{\odot}$ (Charbonneau

et al. 2009).

Combined White Light Curve

First, we combine the three visits into a single light curve, as shown in Figure 3.5, and fit for the following parameters: the planet-to-star radius ratio ($R_p/R_\star$), the total transit duration between first and fourth contact (t_{14}), the stellar radius ($R_\star$), and the two coefficients c and d of the square-root limb-darkening law[4]. Previous studies have found no significant transit timing variations for the GJ1214b system (Charbonneau et al. 2009; Sada et al. 2010; Bean et al. 2010a; Carter et al. 2011; Désert et al. 2011a; Kundurthy et al. 2011; Berta et al. 2011; Croll et al. 2011), so we fix the time of mid-transit for each visit to be that predicted by the linear ephemeris.

As in Burke et al. (2007), we use the parameters t_{14} and $R_\star$ to ensure quick convergence of the MCMC because correlations among these parameters are more linear than for the commonly fit impact parameter (b) and scaled semi-major axis ($a/R_\star$). Because nonlinear transformations between parameter pairs will deform the hypervolume of parameter space, we include a Jacobian term in the priors in Eq. 3.1 to ensure uniform priors for the physical parameters R_p, $R_\star$, and i (see Burke et al. 2007; Carter et al. 2008, for detailed discussions). For the combined light curves, the influence of this term is practically negligible, but we include it for completeness. In the MCMC chains described in this section, all parameters have correlation lengths of 6-13 points.

Initially, we perform the fit with limb-darkening coefficients c and d without

[4]The square-root law is a special case of the 4-parameter law and straightforward to include in the Mandel & Agol (2002) model.

any priors from the PHOENIX atmosphere model, enforcing only that $0 < c + d < 1$, which ensures that the star is brighter at its center ($\mu = 1$) than at its limb ($\mu = 0$). Interestingly, the quantity $(c/3 + d/5)$, which sets the integral of $I(\mu)$ over the stellar surface, defines the line along which c and d are most strongly correlated in the MCMC samples (see also Irwin et al. 2011b). For quadratic limb-darkening, the commonly quoted $2u_1 + u_2$ combination (Holman et al. 2006) has the same physical meaning. The integral of $I(\mu)$ can be thought of as the increase in the central transit depth over that for a constant-intensity stellar disk, so it makes sense that it is well-constrained for nearly equatorial transiting systems like GJ1214b. Planets with higher impact parameters do not sample the full range of $0 < \mu < 1$ during transit, leading to correspondingly weaker limb-darkening constraints that can be derived from their light curves (see Knutson et al. 2011). We quote confidence intervals for the linear combination $(c/3 + d/5)$ and one orthogonal to it in Table 3.2, along with rest of the parameters.

Heartened by finding that when they are allowed to vary freely, our inferred white-light limb-darkening coefficients agree to 1σ to those derived using the PHOENIX stellar model, we perform a second fit that includes the PHOENIX models as informative priors. For this prior, we say $P(\mathbf{M})$ in Eq. 3.1 is proportional to a Gaussian with $(c/3 + d/5) = 0.0892 \pm 0.018$ and $(c/5 - d/3) = -0.431 \pm 0.032$, which is centered on the PHOENIX model. To set the 1σ widths of these priors, we start by varying the effective temperature of the star in the PHOENIX model by its 130K uncertainty in either direction, and then double the width of the prior beyond this, to account for potential systematic uncertainties in the atmosphere model. The results from the fit with these LD priors are shown in Table 3.2.

The photometric noise rescaling parameter s is within 10% of unity, implying that the

376 ppm achieved scatter in the combined white light curve can be quite well-explained from the known sources of uncertainty in the measurements, predominantly photon noise from the star. As shown in Figure 3.6, for the `divide-oot`-corrected light curves, the autocorrelation function (ACF) of the residuals shows no evidence for time-correlated noise. Likewise, the scatter in binned `divide-oot` residuals decreases as the square-root of the number of points in a bin, as expected for uncorrelated Gaussian noise. If there are uncorrected systematic effects remaining in the data, they are below the level of the photon noise over the time-scales of interest here.

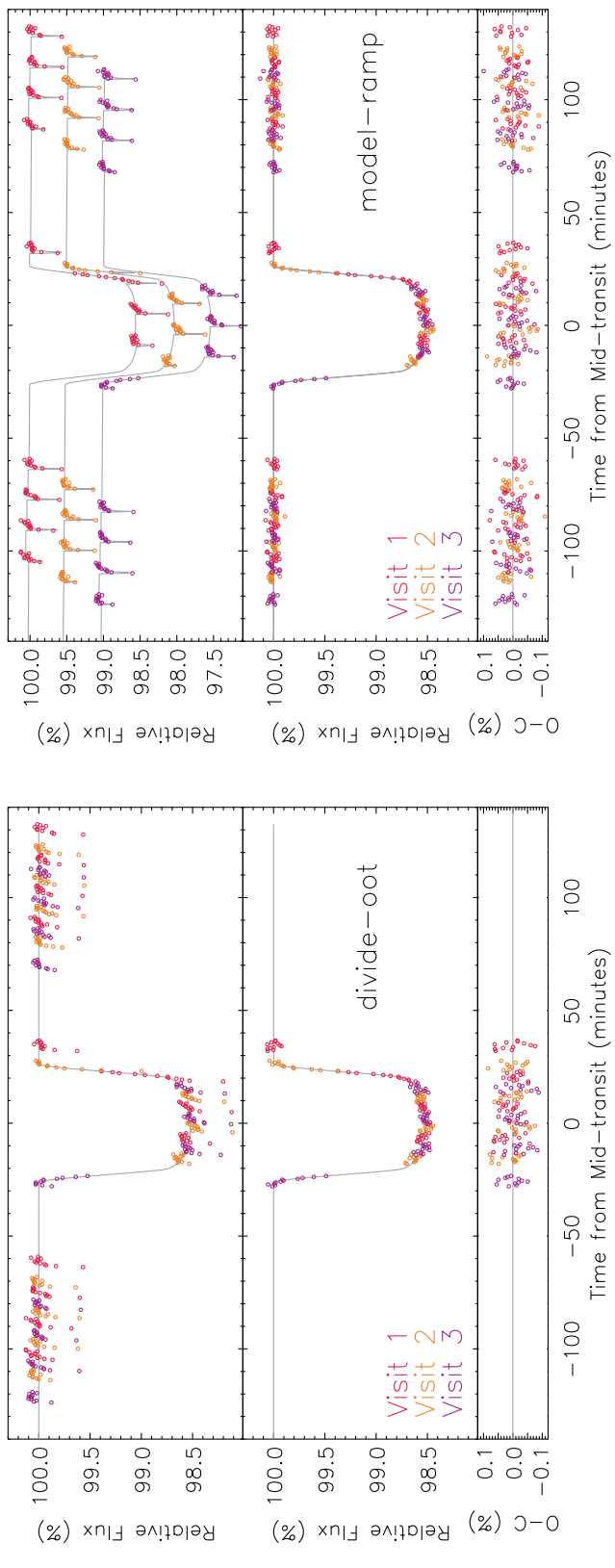

Figure 3.5: The white light curve of GJ1214b's transits before (*top panels*) and after (*middle panels*) removing the instrumental systematics using the `divide-oot` (*left*) and `model-ramp` (*right*, with offsets for clarity) methods described in §3.4.4 and §3.4.5. A transit model that was fit to the `divide-oot`-corrected light curve, constrained to the values of $a/R_\star$ and b used by Bean et al. (2010a), is shown (*gray lines*), along with residuals from this model (*bottom panels*). In the left panels, the out-of-transit orbits are not shown after the correction has been applied, because they contain no further information.

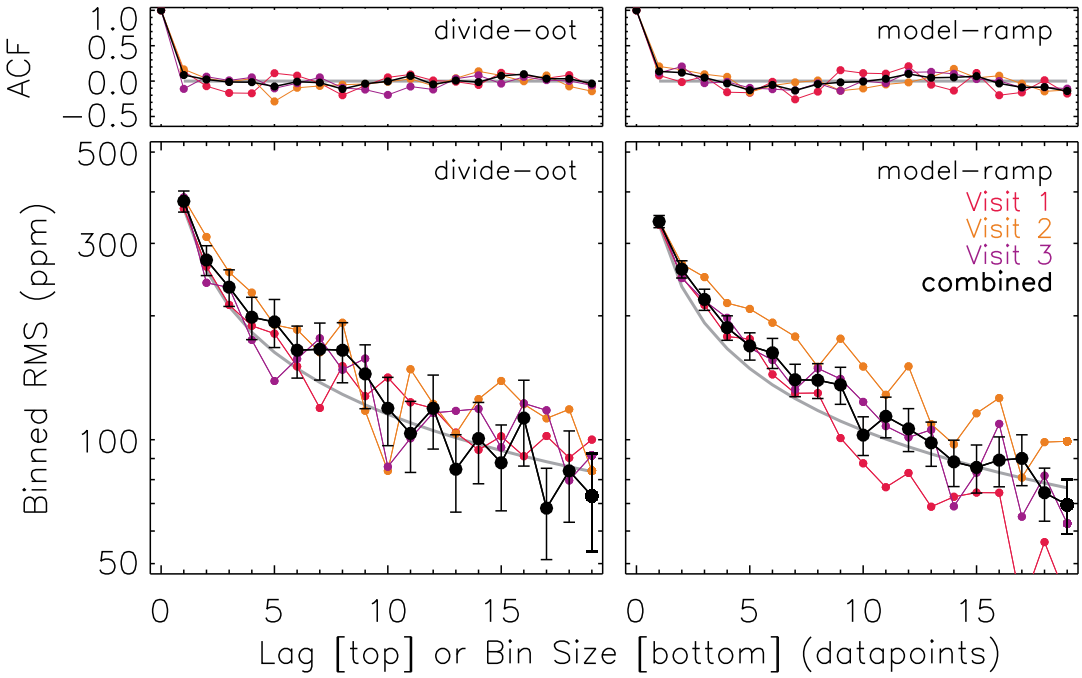

Figure 3.6: For the transit model in Figure 3.5 and both types of systematics treatments, the autocorrelation function of the residuals (ACF; *top*) and the scatter in binned residuals as a function of bin size (*bottom*). The residuals from the combined light curve are shown (*black points*), as well as the individual visits (*colorful points*). The expectations from uncorrelated Gaussian noise (0 in the top, $\propto 1/\sqrt{N}$ in the bottom) are overplotted (*dashed lines*).

Individual White Light Curves

To test for possible differences among our WFC3 visits, we fit each of the three `divide-oot`-corrected white light curves individually. In addition to $R_p/R_\star$, t_{14}, and $R_\star$, we also allow ΔT_c (the deviation of each visit's mid-transit time from the linear ephemeris) to vary freely. We allow c and d to vary, but enforce the same `PHOENIX`-derived priors described in §3.4.5.

Table 3.3 shows the results, which are consistent with each other and with other observations (Charbonneau et al. 2009; Bean et al. 2010a; Carter et al. 2011; Kundurthy et al. 2011; Berta et al. 2011). The three measured ΔT_c's show no evidence for transit timing variations. The uncertainties for the parameters t_{14}, $R_\star$, and ΔT_c are noticeably largest in the first visit; this is most likely because the first visit does not directly measure the timing of either 1^{st} or 4^{th} contact, on which these parameters strongly depend. Additionally, whereas the correlation lengths in the MCMC chains for these parameters in the two visits that do measure $1^{st}/4^{th}$ contact and for $R_p/R_\star$, c, and d in all three visits are small (10-30 points), the correlation lengths for t_{14}, $R_\star$, and ΔT_c in the first visit are very large (300-400 points), indicating these weakly constrained parameters are poorly approximated by the MPFIT-derived covariance matrix. On account of the large correlation lengths for these parameters, we ran the MCMC for the first visit with a factor of 10 more points. In each of the three visits, the uncertainty rescaling parameter s is slightly above but consistent with unity, indicating the photometric scatter is quite well explained by known sources of noise.

Table 3.2. Transit Parameters Inferred from the Combined White Light Curve

Parameter	No LD Prior	LD Prior[a]
t_{14} (days)	0.03624 ± 0.00013	0.03620 ± 0.00012
$R_\star$ ($R_\odot$)	$0.2014^{+0.0038}_{-0.0025}$	$0.201^{+0.004}_{-0.003}$
$a/R_\star$	$15.30^{+0.19}_{-0.29}$	$15.31^{+0.21}_{-0.29}$
i (°)	89.3 ± 0.4	$89.3^{+0.4}_{-0.3}$
b	$0.18^{+0.09}_{-0.11}$	$0.19^{+0.08}_{-0.11}$
$R_p/R_\star$[b]	$0.1158^{+0.0007}_{-0.0006}$	0.1160 ± 0.0005
$c/3 + d/5$	0.096 ± 0.008	0.095 ± 0.007
$c/5 - d/3$	$-0.52^{+0.22}_{-0.14}$	-0.433 ± 0.032
predicted RMS	337 ppm	337 ppm
achieved RMS	373 ppm	376 ppm
s	1.12 ± 0.07	1.12 ± 0.07

[a]The Gaussian limb-darkening priors of $(c/3 + d/5) = 0.0892 \pm 0.018$ and $c/5 - d/3 = -0.4306 \pm 0.032$ were derived from `PHOENIX` stellar atmospheres, as described in the text.

[b]Confidence intervals on $R_p/R_\star$ do not include the $\Delta R_p/R_\star = 0.0006$ systematic uncertainty due to stellar variability (see text).

Stellar Variability

GJ1214 is known to be variable on 50-100 day timescales with an amplitude of 1% in the MEarth bandpass (715-1000 nm; see Charbonneau et al. 2009; Berta et al. 2011). To gauge the impact of stellar variability in the wavelengths studied here, we plot in Figure 3.7 the relative out-of-transit flux as measured by our WFC3 data. For each HST visit, we have three independent measurements of this quantity: the F130N narrow-band direct image, the 0^{th}-order spectrum, and the 1^{st}-order spectrum. Consistent variability over these measurements that sample different regions of the detector within each visit would be difficult to reproduce by instrumental effects, such as flat-fielding errors. In Figure 3.7, GJ1214 appears brighter in the first visit than in the last two visits, with an overall amplitude of variation of about 1%.

This 1% variability, if caused by unocculted spots on the stellar surface, should lead to variations in the inferred planet-to-star radius ratio on the order of $\Delta R_p/R_\star = 0.0006$ (Berta et al. 2011). This is larger than the formal error on $R_p/R_\star$ from the combined white light curve (Tab. 3.2), and must be considered as an important systematic noise floor in the measurement of the absolute, white-light transit depth. We do not detect this variability in the individually measured transit depths (Tab. 3.3) because it is smaller than the uncertainty on each. Most importantly, while the spot-induced variability influences the absolute depth at each epoch, its effect on the relative transit depth among wavelengths will be much smaller and not substantially bias our transmission spectrum estimate.

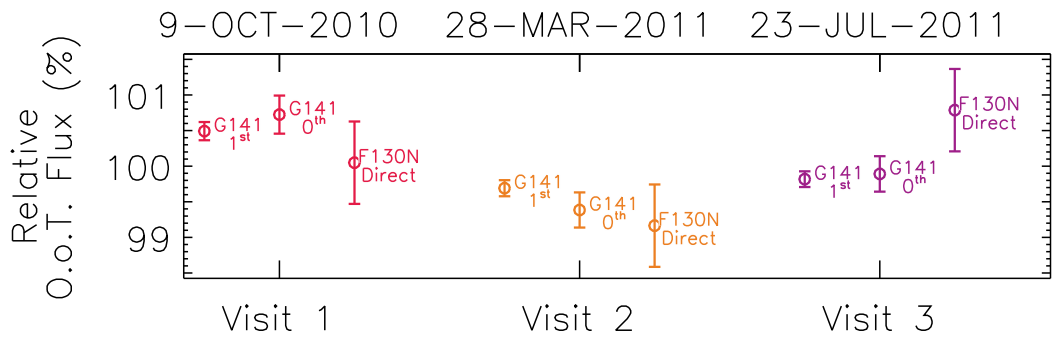

Figure 3.7: The relative out-of-transit (O.o.T.) flux for each HST visit, measured independently from three different groups of images: the summed 1^{st}-order spectrum, the 0^{th}-order image, and the narrow-band direct image, each normalized to its mean. Error bars denote the standard deviation of the out-of-transit measurements within each visit; they do not include the 0.5% uncertainty in the detector flat-field. The narrow-band measurements sample fewer photons, thus their larger uncertainties.

Table 3.3. Transit Parameters Inferred from Individual White Light Curves

Parameter	Visit 1	Visit 2	Visit 3
ΔT_c (days)[a]	$-0.0001^{+0.0012}_{-0.0020}$	0.00028 ± 0.00031	0.0002 ± 0.0004
t_{14} (days)	$0.037^{+0.004}_{-0.003}$	0.0357 ± 0.0007	0.0369 ± 0.0010
$R_\star$ ($R_\odot$)	$0.211^{+0.021}_{-0.014}$	$0.200^{+0.008}_{-0.006}$	$0.214^{+0.015}_{-0.011}$
$a/R_\star$	14^{+1}_{-1}	$15.4^{+0.5}_{-0.6}$	14.4 ± 0.9
i (°)	88.9 ± 0.7	89.2 ± 0.6	$88.5^{+0.8}_{-0.7}$
b	0.27 ± 0.17	0.21 ± 0.15	$0.38^{+0.13}_{-0.20}$
$R_p/R_\star$	$0.1164^{+0.0009}_{-0.0008}$	$0.1159^{+0.0011}_{-0.0009}$	$0.1175^{+0.0011}_{-0.0012}$
predicted RMS	337 ppm	337 ppm	337 ppm
acheived RMS	343 ppm	360 ppm	366 ppm
s	$1.07^{+0.13}_{-0.11}$	$1.12^{+0.13}_{-0.11}$	$1.14^{+0.14}_{-0.11}$

[a]Offset between the observed mid-transit time and that calculated from the linear ephemeris with $P = 1.58040481$ and T_c=2454966.525123 BJD_{TDB}.

Modeling Instrumental Systematics

Before calculating GJ1214b's transmission spectrum, we detour slightly to use the white light curves's high photometric precision to investigate the characteristics of WFC3's instrumental systematics. Rather than correcting for the instrumental systematics with the simple non-parametric `divide-oot` method, in this section we describe them with an analytic model whose parameters illuminate the physical processes at play. We refer to this treatment as the `model-ramp` method.

In this model, we treat the systematics as consisting of an exponential ramp, an orbit-long slope, and a visit-long slope. We relate the observed flux ($F_{\rm obs}$) to the systematics-free flux ($F_{\rm cor}$) by

$$\frac{F_{\rm obs}}{F_{\rm cor}} = (C + Vt_{\rm vis} + Bt_{\rm orb})\left(1 - Re^{-(t_{\rm bat}-D_{\rm b})/\tau}\right) \tag{3.5}$$

where $t_{\rm vis}$ is time within a visit (= 0 at the middle of each visit), $t_{\rm orb}$ is time within an orbit (= 0 at the middle of each orbit), $t_{\rm bat}$ is time within a batch (= 0 at the start of each batch), τ is a ramp timescale, and the term

$$D_{\rm b} = \begin{cases} D & \text{for the } 1^{st} \text{ batch of an orbit} \\ 0 & \text{for the other batches} \end{cases} \tag{3.6}$$

allows the exponential ramp to be delayed slightly for the first batch of an orbit.

The exponential form arises out of the toy model proposed by Agol et al. (2010), where a certain volume of the detector pixels has the ability to temporarily trap charge carriers and later release them as excess dark current. In quick series of sufficiently strong exposures, the population of charge traps approaches steady state, corresponding to the flattening of the exponential. Judging by the appearance of the ramp in the

2^{nd}-4^{th} batches of each orbit, the release timescale seems to be short enough that the trap population completely resets to the same baseline level after each 6 minute buffer download (during which the detector was being continually flushed each 2.9 seconds). Compared to these batches, the 1^{st} batch of each orbit appears to exhibit a ramp that is either weaker, or as we have parameterized it with the D_b term, delayed. We do not explain this, but we hypothesize that it relates to rapid changes in the physical state of the detector coming out of Earth occultation affecting the pixels' equilibrium charge trap populations.

The visit-long and orbit-long slopes are purely descriptive terms (as in Brown et al. 2001; Carter et al. 2009; Nutzman et al. 2011a), but relate to physical processes in the telescope and camera. The orbit-long slope probably arises from the combination of pointing/focus drifts (see Figure 3.3) with our imperfect flat-fielding of the detector. This effect of this orbital phase term could be equally well-achieved, for instance, by including a linear function of the 0^{th} order x and y positions (see Burke et al. 2010; Pont et al. 2007; Swain et al. 2008). The visit-long slope is not mirrored in any of the measured geometrical properties of the star on the detector, and is more difficult to associate with a known physical cause.

In order to determine the parameters C, V, B, R, D, and τ, we fit Eq. 3.6 multiplied by a transit model to the last three orbits of each visit's uncorrected white light curve. The transit parameters are allowed to vary exactly as in §3.4.5, including the use of the informative prior on the limb-darkening coefficients. The white light curves with the best `model-ramp` fit are shown in Figure 3.5, and the properties of the residuals from this model are shown in Figure 3.6. The transit parameters from this independent systematics correction method are consistent with those in Tab 3.3. We do not quite achieve the

280 ppm predicted scatter in the `model-ramp` light curves, and the residuals show slight evidence for correlated noise (Figure 3.6). More complicated instrumental correction models could almost certainly improve this, but we only present this simple model for heuristic purposes. In all sections except this one, we use the `divide-oot`-corrected data exclusively for drawing scientific conclusions about GJ1214b.

Figure 3.8 shows the inferred PDF's of the instrumental systematics parameters for all three HST visits, graphically demonstrating the striking repeatability of the systematics. As expected from the nearly identical cadence of illumination within each of the three visits, the ramp has the same $R = 0.4\%$ amplitude, $\tau = 30$ second timescale, and $D = 20$ second delay time across all observations. The values of τ and D are similar to the time for a single exposure, 25 seconds (including overhead). While the visit-long slope V is of an amplitude (fading by 0.06% over an entire visit) that could conceivably be consistent with stellar variability, the fact that it is identical across all three visits argues strongly in favor of it being an instrumental systematic. B is the only parameter that shows any evidence for variability between orbits; we would expect this to be the case if this term arises out of flat-fielding errors, since the 1^{st} order spectrum falls on different pixels in the three visits.

Figure 3.8: The *a posteriori* distribution of the instrumental systematics parameters from the analytic model, in each of the three visits. The MCMC results for single parameters (*diagonal;* histograms) and pairs of parameters (*off-diagonal;* contours encompassing 68% and 95% of the distribution) are shown, marginalized over all other parameters (including c and d with priors, t_{14}, and $R_\star$). V is measured in units of relative flux/$(3 \times 96$ minutes), B in relative flux/(96 minutes), R in relative flux, and both τ and D in seconds. All visits are plotted on the same scale; for quantitative comparison, the median values and 1σ uncertainties of each parameter are quoted along the diagonal. The systematics parameters are remarkably repeatable from visit to visit; also, they are largely uncorrelated with $R_p/R_\star$ (*left column*).

3.4.6 Spectroscopic Light Curve Fits

We construct multiwavelength spectroscopic light curves by binning the extracted first order spectra into channels that are 5 pixels ($\Delta\lambda = 23$ nm) wide. We estimate the flux, flux uncertainty, and effective wavelength of each bin from the inverse-variance (estimated from the noise model) weighted average of each quantity over the binned pixels. For each of these binned spectroscopic light curves, we employ the `divide-oot` method to correct for the instrument systematics.

To measure the transmission spectrum of GJ1214b, we fit each of these 24 spectroscopic light curves from each of the three visits with a model in which $R_p/R_\star$, c, d, and s are allowed to vary. We hold the remaining parameters fixed so that $a/R_\star = 14.9749$ and $b = 0.27729$, which are the values used by Bean et al. (2010a), Désert et al. (2011a), and Croll et al. (2011). For limb-darkening priors, we use the same sized Gaussians on the same linear combinations of c and d as in §3.4.5, but center them on the `PHOENIX`-determined best values for each spectroscopic bin (see §3.4.2). The correlation length of all parameters is < 10 in the MCMC chains.

For most spectroscopic bins, the inferred value of s is within 1σ of unity, indicating that the flux residuals show scatter commensurate with that predicted from photon noise (1400 to 1900 ppm across wavelengths). No evidence for correlated noise is seen in any of the bins, as judged by the same criterion as for the white light curves (see Figure 3.6).

Figure 3.9 shows the transmission spectra inferred from each of the three visits, as well as the `divide-oot`-corrected, spectrophotometric light curves from which they were derived. For the final transmission spectrum (shown as black points in Figure 3.9), we combine the three values of $R_p/R_\star$, and $\sigma_{R_p/R_\star}$ in each wavelength bin by averaging them

over the visits with a weighting proportional to $1/\sigma^2_{R_p/R_\star}$. Table 3.4 gives this average transmission spectrum, as well as the central values of the limb-darkening prior used in each bin. The wavelength grids in the three visits are offset slightly (by less than a pixel) from one another; in Table 3.4 we quote the average wavelength for each bin.

In §3.4.5 we found that GJ1214's 1% variability at WFC3 wavelengths causes $\Delta D = 0.014\%$ or $\Delta R_p/R_\star = 0.0006$ variations in the absolute transit depth. The starspots causing this variability would have a similar effect on measurements of the transmission spectrum, but unless GJ1214's starspot spectrum is maliciously behaved, the offsets should be broad-band and the influence on the wavelength-to-wavelength variations within the WFC3 transmission spectrum should be much smaller. Each visit's transmission spectrum is a differential measurement made with respect to the integrated stellar spectrum at each epoch; by averaging together three estimates to produce our final transmission spectrum, we average over the time-variable influence of the starspots. Importantly, if GJ1214 is host to a large population of starspots that are symmetrically distributed around the star and do not appear contribute to the observed flux variability over the stellar rotation period, their effect on the transmission spectrum will not average out (see Désert et al. 2011b; Carter et al. 2011; Berta et al. 2011).

If we fix the limb-darkening coefficients to the `PHOENIX` values instead of using the prior, the uncertainties on the $R_p/R_\star$ measurements decrease by 20%. If we use only a single pair of LD coefficients (those for the white light curve) instead of those matched to the individual wavelength bins, the transmission spectrum changes by about 1σ on the individual bins, in the direction of showing stronger water features and being less consistent with an achromatic transit depth. These tests confirm that the presence of the broad H_2O feature in the stellar spectrum (see Figure 3.2) makes it especially *crucial*

that we employ the detailed, multiwavelength LD treatment.

As a test to probe the influence of the `divide-oot` systematics correction, we repeat this section's analysis using the analytic `model-ramp` method to remove the instrumental systematics; every point in the transmission spectrum changes by much less than 1σ. We also experimented with combining the three visits' spectroscopic light curves and fitting for them jointly, instead of averaging together the transmission spectra inferred separately from each visit. We found the results to be practically identical to those quoted here.

Because the transmission spectrum is conditional on the orbital parameters we held fixed ($a/R_\star$, b), we underestimate the uncertainty in the *absolute* values of $R_p/R_\star$; the quoted $\sigma_{R_p/R_\star}$ are intended for *relative* comparisons only. Judging by the $R_p/R_\star$ uncertainty in the unconstrained white light curve fit (Table 3.2), varying $a/R_\star$ could cause the ensemble of $R_p/R_\star$ measurements in Table 3.4 to move up or down in tandem with a systematic uncertainty that is comparable to the statistical uncertainty on each. This is in addition to the $\Delta R_p/R_\star = 0.0006$ offsets expected from stellar variability (Berta et al. 2011).

3.4.7 Searching for Transiting Moons

Finally, we search for evidence of transiting satellite compansions to GJ1214b in our summed WFC3 light curves. The light curve morphology of transiting exomoons can be complicated, but they could generally appear in our data as shallow transit-shaped dimmings *or* brightenings offset from the planet's transit light curve (see Kipping 2011, for a detailed discussion). While the presence of a moon could also be detected in

temporal variations of the planetary transit duration (Kipping 2009), we only poorly constrain GJ1214b's transit duration in individual visits due to incomplete coverage.

Based on the Hill stability criterion, we would not expect moons to survive farther than 8 planetary radii away from GJ1214b so their transits should not be offset from GJ1214b's by more than 25 minutes, less than the duration of an HST Earth occultation. We search only the data in the in-transit visit, using the `divide-oot` method to correct for the systematics. Owing to the long buffer download gaps in our light curves (see §3.2), the most likely indication of a transiting moon in the WFC3 light curve would be an offset in flux from one 12-exposure batch to another. Given the 376 ppm per-exposure scatter in the `divide-oot` corrected light curve, we would have expected to be able to identify transits of 0.4 $R_\oplus$ (Ganymede-sized) moons at 3σ confidence. We see no strong evidence for such an offset. Also, we note that starspot occultations could easily mimic the light curve of a transiting exomoon in the time coverage we achieve with WFC3, and such occultations are known to occur in the GJ1214b system (see Berta et al. 2011; Carter et al. 2011; Kundurthy et al. 2011).

Due to the many possible configurations of transiting exomoons and the large gaps in our WFC3 light curve, our non-detection of moons does not by itself place strict limits on the presence of exo-moons around GJ1214b.

3.5 Discussion

The average transmission spectrum of GJ1214b from our three HST visits is shown in Figure 3.9. To the precision afforded by the data, this transmission spectrum is flat; a

simple weighted mean of the spectrum is a good fit, with $\chi^2 = 20.4$ for 23 degrees of freedom.

3.5.1 Implications for Atmospheric Compositions

We compare the WFC3 transmission spectrum to a suite of cloud-free theoretical atmosphere models for GJ1214b. The models were calculated in Miller-Ricci & Fortney (2010), and we refer the reader to that paper for their details. To compare them to our transmission spectrum, we bin these high-resolution ($R = 1000$) models to the effective wavelengths of the 5-pixel WFC3 spectroscopic channels ($R = 50 - 70$) by integrating over each bin. Generally, to account for the possible suppression of transmission spectrum features caused by the overlap of shared planetary and stellar absorption lines, this binning should be weighted by the photons detected from the system at very high resolution, but this added complexity is not justified for our dataset. The normalization of the model spectra is uncertain (i.e. the planet's true R_p), so we allow a multiplicative factor in $R_p/R_\star$ to be applied to each (giving 24-1=23 degrees of freedom for all models). Varying the bin size between 2 and 50 pixels wide does not significantly change any of the results we quote in this section.

A solar composition atmosphere in thermochemical equilibrium is a terrible fit to the WFC3 spectrum; it has a χ^2=126.2 (see Figure 3.9) and is formally ruled out at 8.2σ confidence. Likewise, the same atmosphere but enhanced $50\times$ in elements heavier than helium, a qualitative approximation to the metal enhancement in the Solar System ice giants (enhanced $30 - 50\times$ in C/H; Gautier et al. 1995; Encrenaz 2005; Guillot & Gautier 2009), is ruled out at 7.5σ ($\chi^2 = 113.2$). Both models assume equilibrium

molecular abundances and the absence of high-altitude clouds; if GJ1214b has an H_2-rich atmosphere, at least one of these assumptions would have to be broken.

Suggesting, along these lines, that photochemistry might deplete GJ1214b's atmosphere of methane, Désert et al. (2011a), Croll et al. (2011), and Crossfield et al. (2011) have noted their observations to be consistent with a solar composition model in which CH_4 has been artificially removed. With the WFC3 spectrum alone, we can rule out such an H_2-rich, CH_4-free atmosphere at 6.1σ (Figure 3.10). This is consistent with Miller-Ricci Kempton et al. (2012)'s theoretical finding that such thorough methane depletion cannot be achieved through photochemical processes, even when making extreme assumptions for the photoionizing UV flux from the star.

Previous spectroscopic measurements in the red optical (Bean et al. 2010a, 2011) could only be reconciled with a H_2-rich atmosphere if such an atmosphere were to host a substantial cloud layer at an altitude above 200 mbar (see Miller-Ricci Kempton et al. 2012). How far the flattening influence of such a cloud layer would extend beyond 1 μm to WFC3 wavelengths would depend on both the concentration and size distribution of the scattering particles. As such, we explore possible cloud scenarios consistent with the WFC3 spectrum in an *ad hoc* fashion, using a solar composition atmosphere and arbitrarily cutting off transmission below various pressures to emulate optically thick cloud decks at different altitudes in the atmosphere. Figure 3.11 summarizes the results. A cloud deck at 100 mbar, which would be sufficient to flatten the red optical spectrum, is ruled out at 5.7σ ($\chi^2 = 82.8$). Due to higher opacities between 1.1 and 1.7 μm, WFC3 probes higher altitudes in the atmosphere than the red optical, requiring clouds closer to 10 mbar to match the data ($\chi^2 = 23.4$). Note, with the term "clouds" we refer to all types of particles that cause broad-band extinction, whether they scatter or absorb, and

whether they were formed through near-equilibrium condensation (such as Earth's water clouds) or through upper atmosphere photochemistry (such as Titan's haze).

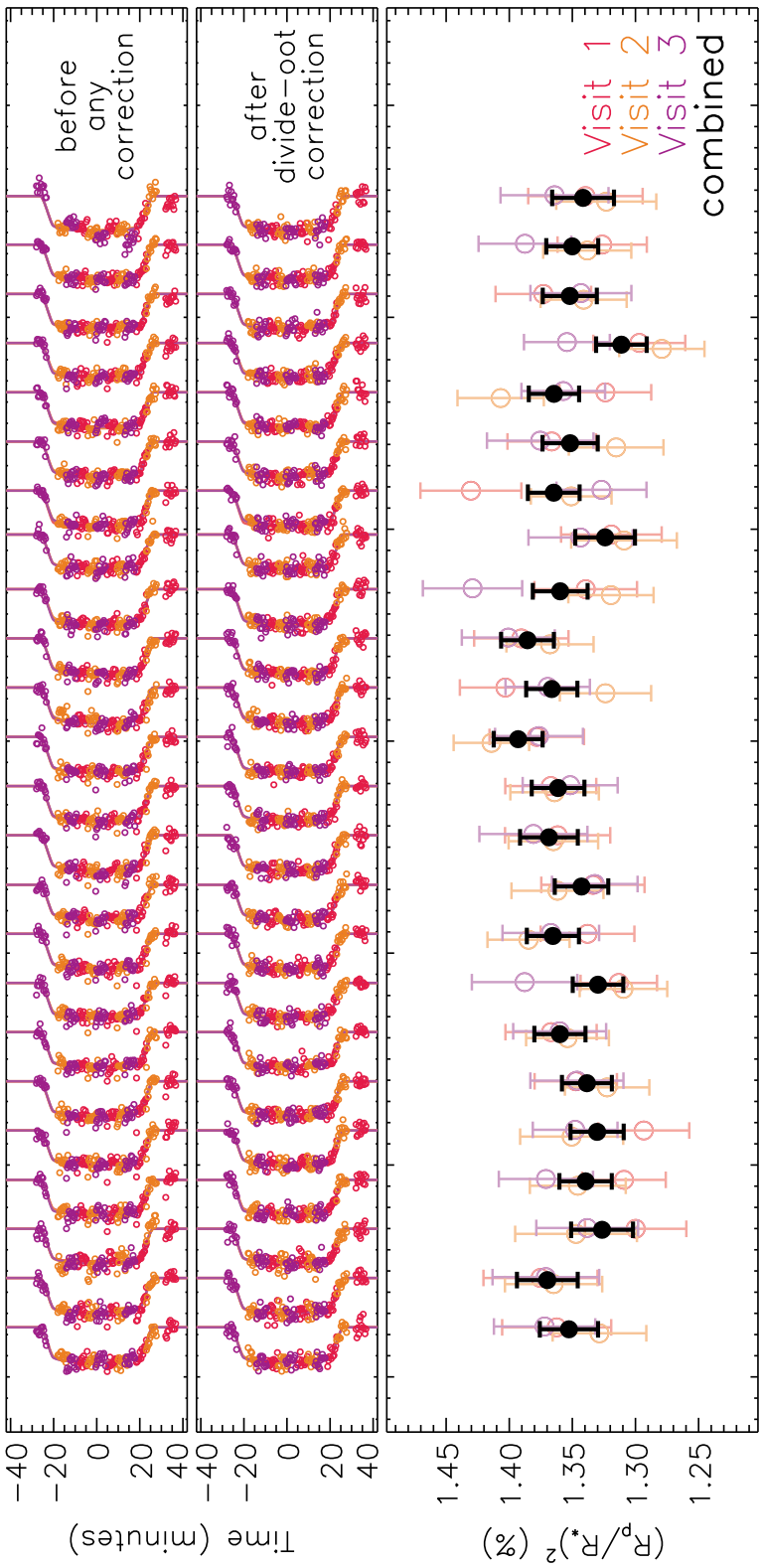

Figure 3.9: *Top panels:* Spectroscopic transit light curves for GJ1214b, before and after the `divide-oot` correction, rotated and offset for clarity. *Bottom panel:* The combined transmission spectrum of GJ1214b (*black circles with error bars*), along with the spectra measured for each visit (*colorful circles*). Each light curve in the top panel is aligned to its respective wavelength bin in the panel below. Colors denote HST visit throughout.

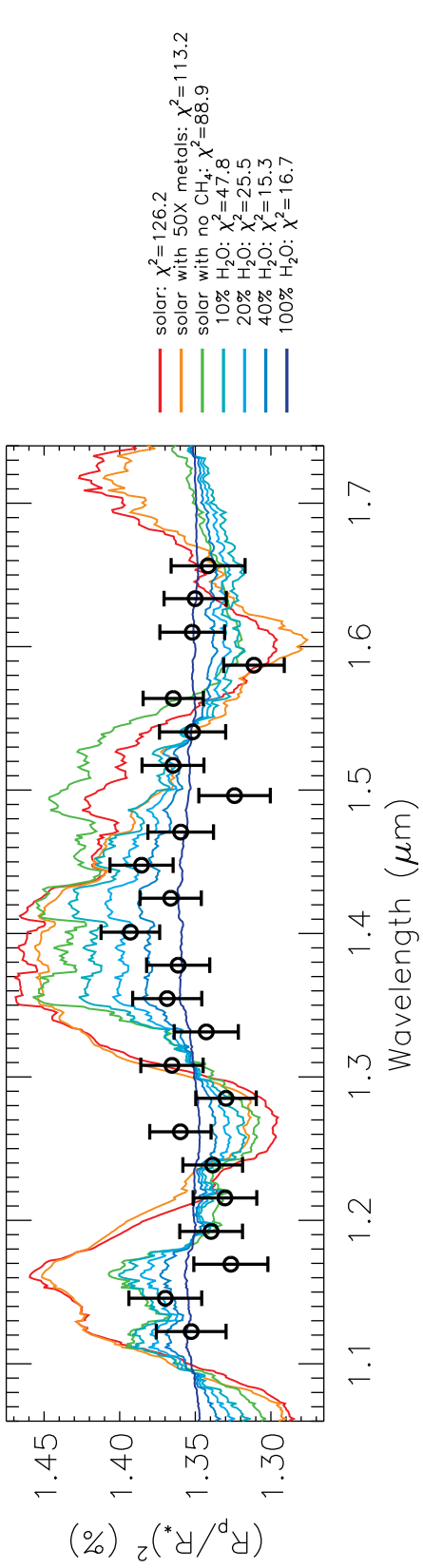

Figure 3.10: The WFC3 transmission spectrum of GJ1214b (*black circles with error bars*) compared to theoretical models (*colorful lines*) with a variety of compositions. The high resolution models are shown here smoothed for clarity, but were binned over each measured spectroscopic bin for the χ^2 comparisons. The amplitude of features in the model transmission spectra increases as the mean molecular weight decreases between a 100% water atmosphere ($\mu = 18$) and a solar composition atmosphere ($\mu = 2.36$).

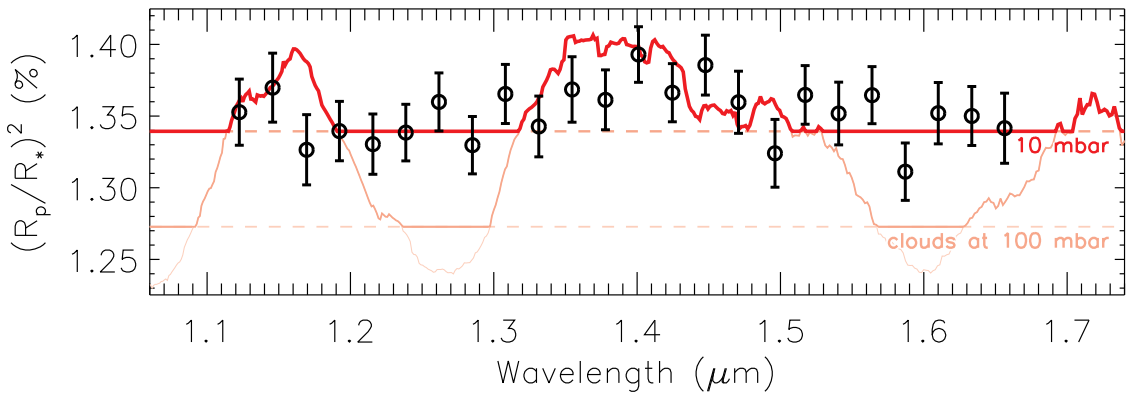

Figure 3.11: The WFC3 transmission spectrum of GJ1214b (*black circles with error bars*) compared to a model solar composition atmosphere that has thick clouds located at altitudes of 100 mbar (*pink lines*) and 10 mbar *red lines*). We treat the hypothetical clouds in an *ad hoc* fashion, simply cutting off transmission through that atmosphere below the denoted pressures.

Fortney (2005) and Miller-Ricci Kempton et al. (2012) identified KCl and ZnS as condensates that would be likely to form in GJ1214b's atmosphere, but found they would condense deeper in the atmosphere (200-500 mbar) than required by the WFC3 spectrum and would probably not be optically thick. While winds may be able to loft such clouds to higher altitudes, it is not clear that the abundance of these species alone would be sufficient to blanket the entire limb of the planet with optically thick clouds. The condensation and complicated evolution of clouds has been studied within the context of cool stars and hot Jupiters (e.g Lodders & Fegley 2006; Helling et al. 2008), but further study into the theoretical landscape for condensate clouds on planets in GJ1214b's gravity and temperature regime is certainly warranted. The scattering may also be due to a high altitude haze formed as by-products of high-altitude photochemistry; Miller-Ricci Kempton et al. (2012) found the conditions on GJ1214b to allow for the formation of complex hydrocarbon clouds through methane photolysis.

However such clouds might form, they would either need to be optically thick up to a well-defined altitude or consist of a substantial distribution of particles acting in the Mie regime, i.e. with sizes approaching 1 μm. Neither the VLT spectra nor our observations give any definitive indications of the smooth falloff in transit depth toward longer wavelengths that would be expected from Rayleigh scattering by molecules or small particles. This is unlike the case of the hot Jupiter HD189733b, where the uniform decrease in transit depth from 0.3 to 1 μm (Pont et al. 2008; Lecavelier Des Etangs et al. 2008; Sing et al. 2011) and perhaps to as far as 3.6 μm (see Sing et al. 2009; Désert et al. 2009) has been convincingly attributed to a small particle haze.

As an alternative, the transmission spectrum of GJ1214b could be flat simply because the atmosphere has a large mean molecular weight. We test this possibility

with H_2 atmospheres that contain increasing fractions of H_2O. This is a toy model, but including molecules other than H_2 or H_2O in the atmosphere would serve principally to increase μ without substantially altering the opacity between 1.1 and 1.7 μm, so the limits we place on μ are robust. We find that an atmosphere with a 10% water by number (50% by mass) is disfavored by the WFC3 spectrum at 3.1σ ($\chi^2 = 47.8$), as shown in Figure 3.9. All fractions of water above 20% (70% by mass) are good fits to the data ($\chi^2 < 25.5$). The 10% water atmosphere would have a minimum mean molecular weight of $\mu = 3.6$, which we take as a lower limit on the atmosphere's mean molecular weight.

For the sake of placing the WFC3 transmission spectrum in the context of other observations of GJ1214b, we also display it alongside the published transmission spectra from the VLT (Bean et al. 2010a, 2011), CFHT (Croll et al. 2011), Magellan (Bean et al. 2011), and Spitzer (Désert et al. 2011a) in Figure 3.12. Stellar variability could cause individual sets of observations to move up and down on this plot by as much as $\Delta D = 0.014\%$ for measurements in the near-IR (Berta et al. 2011); we indicate this range of potential offsets by an arrow at the right of the plot. We display the measurements in Figure 3.12 with no relative offsets applied and note that their general agreement is consistent with the predicted small influence of stellar variability. Depending on the temperature contrast of the spots, however, the variability could be larger by a factor of $2 - 3\times$ in the optical, and we caution the reader to consider this systematic uncertainty when comparing depths between individual studies. For instance, the slight apparent rise in $R_p/R_\star$ toward 0.6 μm, that would potentially be consistent with Rayleigh scattering in a low-μ atmosphere, could also be easily explained through the poorly constrained behavior of the star in the optical. Indeed, (Bean et al. 2011) found a significant offset

between datasets that overlap in wavelength (near 0.8 μm) but were taken in different years, suggesting variability plays a non-negligible role at these wavelengths.

Finally, we note that any model with $\mu > 4$, such as one with a $> 50\%$ mass fraction of water, would be consistent with the measurements from Bean et al. (2010a), Désert et al. (2011a), Crossfield et al. (2011), (Bean et al. 2011), and WFC3. The only observation it could not explain would be the deep K_s-measurement from Croll et al. (2011). Of the theoretical models we tested, we could find none that matched all the available measurements. We are uncertain of how to interpret this apparent incompatibility but hopeful that future observational and theoretical studies of the GJ1214b system may clarify the issue. In the meantime, we adopt an atmosphere with at least 50% water by mass as the most plausible model to explain the WFC3 observations.

3.5.2 Implications for GJ1214b's Internal Structure

If GJ1214b is not shrouded in achromatically optically thick high-altitude clouds, the WFC3 transmission spectrum disfavors any proposed bulk composition for the planet that relies on a substantial, unenriched, hydrogen envelope to explain the planet's large radius. Both the ice-rock core with nebular H/He envelope and pure rock core with outgassed H_2 envelope scenarios explored by Rogers & Seager (2010b) would fall into this category, requiring additional ingredients to match the observations. In contrast, their model that achieves GJ1214b's large radius mostly from a large water-rich core, would agree with our observations.

Perhaps most compellingly, a high μ scenario would be consistent with composition proposed by Nettelmann et al. (2011), who found that GJ1214b's radius could be

explained by a bulk composition consisting of an ice-rock core surrounded by a H/He/H_2O envelope that has a water mass fraction of 50-85%. Such a composition would be intermediate between the H/He- and H_2O-envelope limiting cases proposed by Rogers & Seager (2010b). The H/He/H_2O envelope might arise if GJ1214b had originally accreted a substantial mass of hydrogen and helium from the primordial nebula but then was depleted of its lightest molecules through atmospheric escape.

3.5.3 Prospects for GJ1214b

Future observations with the James Webb Space Telescope (Deming et al. 2009; Kaltenegger & Traub 2009), one of the immense next generation ground-based telescopes (GMT, TMT, ELT; see Ehrenreich et al. 2006), or possibly even a dedicated campaign with current facilities, could detect the 0.01% transmission spectrum features of a 100% water atmosphere on GJ1214b, and potentially distinguish between clear H_2-poor and cloudy H_2-rich atmospheres. Along another front, simulations by Menou (2012) show that observations of GJ1214b's thermal phase curve, such as those for HD189733b by Knutson et al. (2007a), would probe the ratio of radiative to advective timescales in GJ1214b's outer envelope and provide an independent constraint on the atmospheric composition. Detecting the thermal emission from this 500K exoplanet is currently very difficult, and will likely have to wait until the launch of JWST.

In the meantime, we advocate further study of the GJ1214 system in general. Confirming and refining the parallax for the system (van Altena et al. 1995) will improve our knowledge of the stellar mass, and in turn, the planet's mass and radius. Likewise, further radial velocity observations will empirically constrain the hypothesis by Carter

Table 3.4. GJ1214b's Transmission Spectrum from WFC3/G141

Wavelength (μm)	$R_p/R_\star$	c	d
1.123	0.11641 ± 0.00102	-0.372	1.068
1.146	0.11707 ± 0.00099	-0.397	1.088
1.169	0.11526 ± 0.00098	-0.404	1.089
1.192	0.11589 ± 0.00093	-0.410	1.090
1.215	0.11537 ± 0.00091	-0.406	1.075
1.239	0.11574 ± 0.00090	-0.403	1.063
1.262	0.11662 ± 0.00088	-0.407	1.064
1.285	0.11565 ± 0.00088	-0.403	1.045
1.308	0.11674 ± 0.00085	-0.411	1.042
1.331	0.11595 ± 0.00087	-0.390	1.068
1.355	0.11705 ± 0.00089	-0.368	1.101
1.378	0.11664 ± 0.00088	-0.371	1.110
1.401	0.11778 ± 0.00088	-0.338	1.075
1.425	0.11693 ± 0.00091	-0.295	1.029
1.448	0.11772 ± 0.00090	-0.319	1.056
1.471	0.11663 ± 0.00092	-0.322	1.061
1.496	0.11509 ± 0.00100	-0.345	1.084
1.517	0.11635 ± 0.00104	-0.305	1.013
1.541	0.11626 ± 0.00091	-0.330	1.042
1.564	0.11681 ± 0.00091	-0.351	1.059
1.587	0.11443 ± 0.00091	-0.372	1.072
1.610	0.11631 ± 0.00091	-0.399	1.082
1.633	0.11620 ± 0.00092	-0.399	1.075
1.656	0.11581 ± 0.00096	-0.415	1.081

et al. (2011) that a significantly non-zero orbital eccentricity could be biasing GJ1214b's inferred density.

3.6 Conclusions

In this work, we made new measurements of the GJ1214b's transmission spectrum using HST/WFC3. Reaching a precision of $\sigma_{R_p/R_\star} = 0.0009$ in 24 simultaneously measured wavelength bins, we found the transmission spectrum to be completely flat between 1.1 and 1.7 μm. We saw no evidence for the strong H_2O absorption features expected from a range of H_2-rich model atmospheres.

Given the lack of a known source for clouds or hazes that could create a truly achromatic transit depth across all wavelengths, we interpret this flat WFC3 transmission spectrum to be best explained by an atmosphere with a high mean molecular weight. Based on our observations, this atmosphere would likely consist of more than 50% water by mass or a mean molecular weight of $\mu > 4$. Such an atmosphere would be consistent with observations of GJ1214b's transmission spectrum by Bean et al. (2010a), Désert et al. (2011a), Crossfield et al. (2011), and Bean et al. (2011) although it would be difficult to reconcile with those by Croll et al. (2011).

Such a constraint on GJ1214b's upper atmosphere serves as a boundary condition for models of bulk composition and structure of the rest of the planet. It suggests GJ1214b contains a substantial fraction of water throughout the interior of the planet in order to obviate the need for a completely H/He- or H_2-dominated envelope to explain the planet's large radius. A high bulk volatile content would point to GJ1214b

forming beyond the snow line and migrating inward, although any such statements about GJ1214b's past are subject to large uncertainties in the atmospheric mass loss history (see Rogers et al. 2011).

Finally, this paper is the first published study using WFC3 for observing a transiting exoplanet. Aside from several instrumental systematics that were straightforward to correct and did not require a detailed instrumental model, the camera delivered nearly photon-limited performance both in individual spectrophotometric light curves and in summed white light curves. We are confident that WFC3 will serve as a valuable tool for exoplanet atmospheric characterization in the years to come.

3.7 Acknowledgements

We thank Jacob Bean, Bryce Croll, Ian Crossfield, Ernst de Mooij, Leslie Rogers, Drake Deming, Avi Mandell, Martin Kümmel, Ron Gilliand, Tom Wheeler, and the CfA Summer Statistics Club for valuable discussions regarding this work. We are extremely grateful to our Program Coordinator Patricia Royle, Contact Scientist Howard Bushhouse, the WFC3 instrument team, the entire staff at STScI and NASA, and the crew of STS-125 for their crucial roles in enabling these observations. We thank the anonymous referee for a thorough reading and thoughtful comments that improved the manuscript. E.K. acknowledges funding from NASA through the Sagan Fellowship Program. P.R.M. thanks CfA director Charles Alcock for enabling a sabbatical at SAO. We gratefully acknowledge funding from the David and Lucile Packard Fellowship for Science and Engineering (awarded to D.C.), the National Science Foundation (grant AST-0807690, awarded to D.C.), and NASA (grant HST-GO-12251). This work is based

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-12251. This research has made use of NASA's Astrophysics Data System.

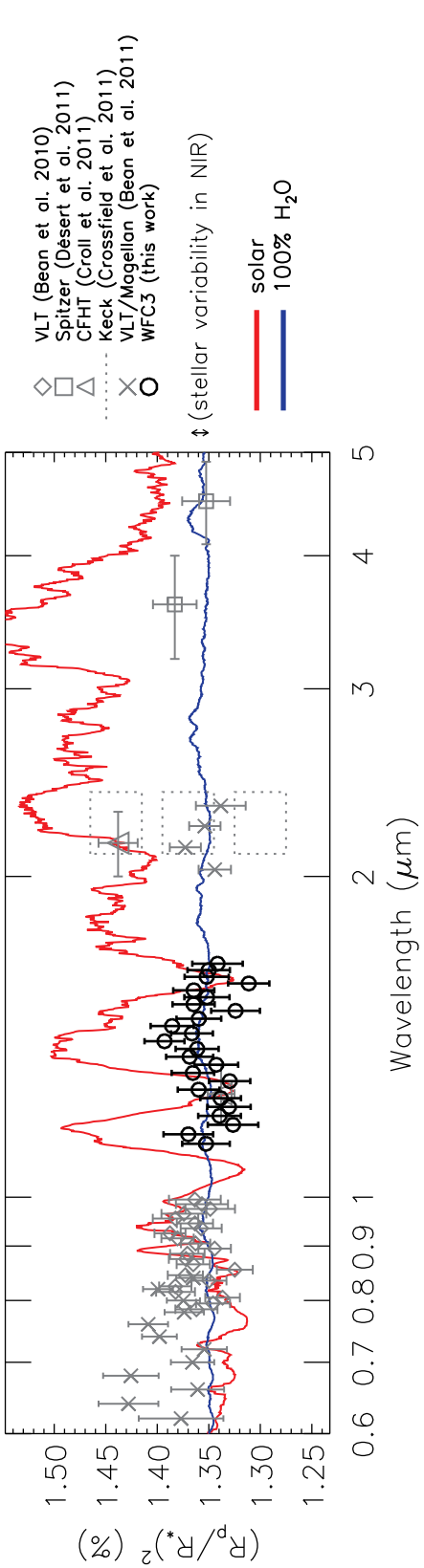

Figure 3.12: GJ1214b's transmission spectrum from WFC3 in the context of observations from the VLT (0.6-1 μm; Bean et al. 2010a, 2011), CFHT (1.25 + 2.15 μm; Croll et al. 2011), Magellan (2.0-2.3 μm Bean et al. 2011), and Spitzer (3.6 + 4.5 μm; Désert et al. 2011a). While they do not measure an absolute transit depth, observations from NIRSPEC on Keck (2.1-2.3 μm Crossfield et al. 2011) disfavor models they tested that had amplitudes larger than 0.05% in their wavelength range; we represent these constraints with the dashed rectangular boxes. Two extremes of the models explored in this paper are shown, normalized to the MEarth-measured transit depth (see Miller-Ricci & Fortney 2010). It is important to note that stellar variability could cause individual data sets to shift up or down on this plot as much as $\Delta D = 0.014\%$ in the near-IR or $2-3\times$ more in the optical, depending on the stellar spot spectrum.

Chapter 4

Transit Detection in the MEarth Survey of Nearby M Dwarfs: Bridging the Clean-first, Search-later Divide

This thesis chapter originally appeared in the literature as

Z. K. Berta, J. Irwin, D. Charbonneau, C. J. Burke, E. E. Falco *The Astronomical Journal*, 144, 145, 2012

Abstract

In the effort to characterize the masses, radii, and atmospheres of potentially habitable exoplanets, there is an urgent need to find examples of such planets transiting nearby M dwarfs. The MEarth Project is an ongoing effort to do so, as a ground-based photometric survey designed to detect exoplanets as small as $2R_\oplus$ transiting mid-to-late M dwarfs within 33 pc of the Sun. Unfortunately, identifying transits of such planets in photometric monitoring is complicated both by the intrinsic stellar variability that is common among these stars and by the nocturnal cadence, atmospheric variations, and instrumental systematics that often plague Earth-bound observatories. Here we summarize the properties of MEarth data gathered so far, emphasizing the challenges they present for transit detection. We address these challenges with a new framework to detect shallow exoplanet transits in wiggly and irregularly-spaced light curves. In contrast to previous methods that clean trends from light curves before searching for transits, this framework assesses the significance of individual transits simultaneously while modeling variability, systematics, and the photometric quality of individual nights. Our Method for Including Starspots and Systematics in the Marginalized Probability of a Lone Eclipse (MISS MarPLE) uses a computationally efficient semi-Bayesian approach to explore the vast probability space spanned by the many parameters of this model, naturally incorporating the uncertainties in these parameters into its evaluation of candidate events. We show how to combine individual transits processed by MISS MarPLE into periodic transiting planet candidates and compare our results to the popular Box-fitting Least Squares (BLS) method with simulations. By applying MISS MarPLE to observations from the MEarth Project, we demonstrate the utility of this

framework for robustly assessing the false alarm probability of transit signals in real data.

4.1 Introduction

Observationally, nearby M dwarf stars offer both opportunities and challenges as exoplanet hosts. M dwarfs' low masses and small sizes accentuate the radial velocity wobble and eclipse depths of any planets that transit them. Their low luminosities result in habitable zones at much smaller orbital distances than for more luminous stars, so planets in M dwarf habitable zones are more likely to transit and will transit more frequently. These advantages aid the initial discovery (Nutzman & Charbonneau 2008; Blake et al. 2008) and the later detailed characterization (e.g. Deming et al. 2009) of planets that could be small enough and cool enough to potentially host life. Mid-to-late M dwarfs offer a particularly compelling balance in that they have smaller statures than earlier-type stars but are still sufficiently bright to enable high precision followup studies, unlike later-type objects.

Exploiting this opportunity, the ground-based MEarth Project is using robotic, 40 cm telescopes to photometrically monitor nearby (< 33 pc), mid-to-late M dwarfs. MEarth has been operating since 2008 with eight telescopes on Mt. Hopkins, AZ, and will soon include 8 additional telescopes in the Southern hemisphere. By design, MEarth intends to be sensitive to planets as small as $2R_\oplus$ and with periods as long as 20 days, reaching the habitable zones of these stars (see Nutzman & Charbonneau 2008). Like MEarth, several additional ground-based surveys are attempting to capitalize on the M dwarf advantage, including PTF/M-dwarfs (Law et al. 2012), TRAPPIST (Jehin et al.

2011; Bonfils et al. 2012), APACHE (Giacobbe et al. 2012), and the WFCAM Transit Survey (Nefs et al. 2012).

MEarth's first discovered transiting planet, the 1.6 day, 2.7 $R_\oplus$, 6.6 $M_\oplus$ exoplanet GJ1214b (Charbonneau et al. 2009), is far too hot for habitability. But as the first planet in this size range accessible to atmospheric characterization, GJ1214b has proven a useful laboratory for theoretical work (e.g Miller-Ricci & Fortney 2010; Rogers & Seager 2010b; Nettelmann et al. 2011; Menou 2012; Miller-Ricci Kempton et al. 2012) and for observational studies, both from the ground (e.g. Bean et al. 2010a, 2011; Carter et al. 2011; Berta et al. 2011; Kundurthy et al. 2011; Croll et al. 2011; Crossfield et al. 2011; de Mooij et al. 2012) and from space (Désert et al. 2011a; Berta et al. 2012a). Yet, its period is still very short. What are the prospects for finding planets with longer periods, potentially habitable planets?

In light of the relative ease with which the space-based Kepler Mission can find transiting planets with periods longer than 100 days (Batalha et al. 2013), it is important to emphasize that 10-20 day habitable zone periods are long enough to pose significant detection challenges from the ground. Planets with these periods, even if geometrically aligned to transit, may offer only a single transit per season that can be observed from a single site, between weather losses and daytime gaps (e.g., Pepper & Gaudi 2005; von Braun et al. 2009). The scarcity of transits poses a two-fold problem: multiple transits are often necessary to build up sufficient signal-to-noise for detection (e.g. Bakos et al. 2010), and multiple transits are, at some point, almost always necessary for determining a planet's period.

MEarth attempts to address the first challenge with a novel, automated "real-time

trigger" mode of operation. This aids our ability to establish sufficient signal-to-noise to detect planet candidates from one or very few transits. While observing a target at low-cadence, MEarth can rapidly identify in-progress, marginally significant, single transit events from incoming observations. If an in-progress event crosses a low (3σ) threshold, MEarth can automatically trigger high-cadence followup to confirm the candidate event at higher confidence (see Nutzman & Charbonneau 2008; Irwin et al. 2009b). If a transit is real, the triggered observations could magnify its significance from mediocre to ironclad, without having to wait to observe subsequent transits. If no transit is present, the triggered observations generally wash out the importance of the original downward outliers.

The ability to confirm single events at high significance is crucial to MEarth's goal of finding long period planets. Our recent discovery of LSPM J1112+7626, a bright 0.4 + 0.3 $M_{\odot}$ double-lined eclipsing binary in a 41 day orbital period (Irwin et al. 2011b) highlights this point. We identified LSPM J1112+7626 from three exposures taken during a single primary eclipse. Due to the deep ($> 10\%$) eclipses, we were confident the system was real. In parallel with continued photometric monitoring, we began radial velocity observations, and the combination of these two efforts ultimately established the binary's 41 day period. We envision the discovery of a long period planet to follow the same trajectory: a shallow (1%) transit could be identified confidently using high-cadence observations from the real-time trigger, and follow-up scrutiny could be invested to measure the planet's period. This two-part strategy is the only way a ground-based survey like MEarth will have sufficient sensitivity to find planets with periods longer than 10 days.

For this strategy to work, we need a robust method for accurately assessing the

significance of individual transit events, both initially to trigger high-cadence observations of marginal events and later to assess whether an event is significant enough to warrant period-finding follow-up. This problem would be straightforward if MEarth's transit light curves exhibited no noise other than perfectly-behaved, uncorrelated, Gaussian, photon noise. This is not the case. MEarth light curves show astrophysical noise from the M dwarfs themselves, in the form of rotational modulation due to starspots or sporadic stellar flares. They show instrumental noise, such as that caused by pointing drifts, focus changes, and flat-fielding errors. They show extinction effects from Earth's dynamic atmosphere, some of which, as we discuss here, pose particularly pernicious problems for photometry of red stars. Often, these noise sources can mimic both the amplitude and the morphology of single planetary transits.

To invest MEarth's follow-up efforts wisely, we need a conservative method for assessing the significance of a transiting planetary signal in the face of these complicated noise sources. This method needs to both (a) suppress, remove, or correct for stellar variability and systematics to increase sensitivity to shallower transits and (b) accurately propagate the uncertainties associated with this cleaning process into the significance assigned to the candidate signal. This method needs to be able to do so, even if only one or few transits are observed. It does not need to accurately determine the period of a signal; we postpone that endeavor for the eventual follow-up of statistically promising candidates.

The exoplanet literature is teeming with well-established methods for cleaning variability and systematics from transit survey light curves and for searching those cleaned light curves for periodic transit signals. However, to our knowledge, of those methods appropriate for ground-based observations none is sufficiently well suited to this

challenge of estimating the significance of individual transits events. In this paper, we propose a new method, one that searches for transits simultaneously with a light curve cleaning process, so that the significance of candidates is marginalized over the cleaning's uncertainties.

If the rate of planet occurrence around mid-to-late M dwarfs rises sharply toward smaller planet sizes and long periods, as it does for FGK stars (Howard et al. 2012), then the development of even small improvements to our ability to detect shallow, rare transits could have a big payoff for MEarth. Additionally, the development of this method also provides a framework with which to estimate the ensemble sensitivity of the survey as a whole, thus enabling a statistical study from MEarth on the population of planets orbiting nearby mid-to-late M dwarfs. We intend to describe the results of such a study in a forthcoming work.

We begin by introducing the MEarth survey with a description of the observations we have gathered so far (Section 4.2). After reviewing the light curve cleaning and transit detection techniques that have been described in the literature to date (Section 4.3), we outline our new framework for MEarth, describing both how to estimate the significance of a single transit event and how to incorporate well-characterized single events into periodic planet candidates (Section 4.4). We test this method with simulations of injected transits and demonstrate that the candidates generated by its application to the existing MEarth dataset have the statistical properties we would expect (Section 4.5). We conclude by suggesting other potential applications and improvements that could be made with this method (Sections 4.6 and 4.7).

The reader should note that throughout this paper we use the terms "eclipse" and

"transit" completely interchangeably, referring to a planet passing in front of its star as seen from Earth.

4.2 Observations

To frame the observational problem we hope to address, we summarize the properties of the photometric data gathered by the MEarth Project since beginning its full operation in 2008. For completeness, we reiterate some of the points described in the MEarth design strategy (Nutzman & Charbonneau 2008), emphasizing the qualitative features of the MEarth data that present particular challenges to our goal of detecting transits of habitable super-Earths.

4.2.1 The Observatory

Each of MEarth's eight telescopes is an f/9 40-cm Ritchey-Chrétien mounted on a German Equatorial mount. The telescopes are located in a single enclosure with a roll-off roof at the Fred Lawrence Whipple Observatory (FLWO) at Mount Hopkins, Arizona. They are robotically controlled and observe every clear night, except for instrument failures. Due to the summer monsoon in Arizona, we never observe during the month of August when FLWO is closed, and we rarely gather much useful data in July or September.

Each telescope is equipped with a 2048×2048 CCD with a pixel scale of 0.76"/pixel, for a 26' field of view. Our target list contains 2,000 nearby M dwarfs (selected from Lépine & Shara 2005) that are spread all across the Northern sky ($\delta > 0°$), so they

must be observed one-by-one, in a pointed fashion. The field of view is large enough to contain ample comparison stars for each MEarth target, with typically at least ten times as many photons available from comparisons as from the target.

We use a custom 715 nm longpass filter, relying on the quantum efficiency of our back-illuminated e2v CCD42-40 detector to define the long-wavelength response of the system. Extending out to 1000 nm, the shape of this response resembles a combination of the Sloan $i + z$ filters (Fukugita et al. 1996). The broad wavelength range of this filter was designed to maximize our photon flux from M dwarfs, but it introduces an important systematic effect into our photometry, as outlined in Section 4.2.3.

4.2.2 Weather Monitoring

MEarth continuously monitors the conditions on Mt. Hopkins with a suite of weather sensors. At ground level, we measure temperature, humidity, and wind speed, as well as rain and hail accumulation. We detect cloud cover with a wide-angle infrared sensor (a TPS-534 thermopile) that measures the sky brightness temperature at wavelengths $> 5.5\mu$m (see Clay et al. 1998). The primary purpose of this monitoring is to prevent damage to the telescopes by keeping the observatory closed during inclement weather, but the timeseries from this monitoring are also useful in later analysis for identifying weather-related systematics in our data.

4.2.3 Calibrations

To go from raw images to reduced light curves, we follow the procedure and use modified code from Irwin et al. (2007). Here, we review those points in the process where calibration error could potentially lead to light curve systematics.

Non-linearity

The MEarth CCD's behave slightly non-linearly at all count levels, increasing up from a $1-2\%$ non-linearity at half of the detector full well up to $3-4\%$ near the onset of saturation. Because we often need to use comparison stars with different magnitudes than our target star, we must account for this non-linearity. When setting exposure times, we avoid surpassing 50% of the detector's full well, and estimate a correction for the non-linearity using sets of daytime dome flats taken with different exposure times. With these measures in place, we see no evidence that non-linearity limits our photometric performance.

Dark Current and Persistence

We scale dark exposures taken at the end of each night to remove some of the CCD dark current. However, until 2011 we operated our Peltier-cooled detectors at -20° to -15° C, and at these warm temperatures they showed significant persistence. That is, images of bright stars would persist as excess localized dark current in subsequent images, slowly decaying with a half hour timescale below an initial 1% fraction of the original fluence. This was a significant source of systematics: stars in incoming exposures could land on the same pixels as persistent ghost stars from previous exposures, thus gaining a hidden

amount of flux that depended on how recently and strongly those pixels were illuminated. As this effect depends on the entire recent 2-dimensional illumination history of the detector, correcting for it would be extremely complicated. We partially mitigated the persistence by ensuring different target stars were observed on different regions of the detector, but could not eliminate problems due to overlap with comparison stars or due to changes in cadence. Updating our camera housings in 2011, we now operate at -30° C where the amplitude of the persistence is lower. We also adopt a detector preflash before each exposure; this increases the overall dark current but suppresses localized persistent images. Between the lower temperature and this preflash step, persistence no longer has a substantial effect on MEarth photometry.

Flat-field Sensitivity Map

We gather flat-fields at evening and morning twilight, typically 8 per telescope per twilight, with empirically set exposure times estimated using the equations of Tyson & Gal (1993). To average out large-scale gradients in the illumination, we always take adjacent pairs of flats on opposite sides of the meridian, which has the effect of rotating the whole optical system relative to the sky (thanks to our German Equatorial mounts). Because our optical system shows high levels of centrally concentrated scattered light ($10-15\%$ of sky before 2011 and $< 5\%$ after; see Table 4.1) that corrupts the large-scale structure in twilight flat exposures, we estimate the sensitivity in the detector plane in two steps. First, we estimate a small-scale sensitivity map that accounts for dust donuts and pixel-to-pixel variations in the detector sensitivity by filtering out large-scale structure from the combined twilight flats. Second, we derive a large-scale map from dithered photometry of dense star fields to account for the non-uniform illumination

across the field of view. Additionally, our camera's leaf shutter takes a finite time to open and close, resulting in a varying exposure time across the field of view (on a 1 second exposure, the amplitude of this effect is 5%); we apply a shutter correction estimated from sets of twilight flats. Altogether, our flat-fielding procedure achieves a precision of 1% across the entire detector.

However, because we hope to perform photometry down to the level of 0.1%, this 1% knowledge of the sensitivity across the field is still imperfect and will inevitably be a source of systematics in our light curves. One unavoidable problem is that our German Equatorial mounts require the detector to flip 180° when crossing the meridian. In light curves, this causes offsets as large as 1% between opposites sides of the meridian, as stars sample different regions of the large-scale sensitivity of the camera. Notably, the step-function morphology of this systematic can mimic a transit ingress or egress. In addition to this "meridian flip" problem, we achieve a blind RMS pointing accuracy of 60-120". To improve on this, at each pointing we take a short binned image and use its astrometric solution to nudge the telescope to the correct pointing before science exposures, with a random error typically of 1-2". This minimizes the impact of these pointing errors, but does not completely remove the problem of stars sampling different pixels on an imperfectly flat-fielded detector.

Differential Photometry

We perform aperture photometry on all sources in the field of view. For each exposure, we derive a differential photometric correction from point sources in the field using an iterative, weighted, clipped fit that excludes variable stars from the comparison sample

(see Irwin et al. 2007, for details). We calculate a theoretical uncertainty estimate $\sigma_{\rm the}(t)$, in magnitudes[1], for each point:

$$\sigma_{\rm the}(t) = \frac{2.5}{\ln 10} \times \frac{\sqrt{N_\gamma + \sigma^2_{\rm sky} + \sigma^2_{\rm scint} + \sigma^2_{\rm comp}}}{N_\gamma} \tag{4.1}$$

where N_γ is the number of photons from the source, $\sigma^2_{\rm sky}$ is an empirically determined sky noise estimate for the photometric aperture that includes read and dark noise, $\sigma^2_{\rm scint}$ is the anticipated scintillation noise (Young 1967), and $\sigma^2_{\rm comp}$ accounts for the uncertainty in the comparison star solution. In some MEarth fields with very few comparisons, the $\sigma^2_{\rm comp}$ term can be a significant contribution to the overall uncertainty.

Precipitable Water Vapor

A crucial assumption of this differential photometry procedure is that atmospheric or instrumental flux losses are exactly mirrored between target and comparison stars. Our wide 715-1000 nm bandpass overlaps strong telluric absorption features due to water vapor, so as the level of precipitable water vapor (PWV) changes in the column over our telescopes, their effective wavelength response will also change. Red stars will experience a larger share of this time-variable PWV-induced extinction than stars that are blue in this wavelength range. As a typical MEarth field consists of one very red target star (median target $r - J = 3.8$)[2] amongst much bluer comparison stars (median comparison $r - J = 1.3$), most MEarth M dwarfs exhibit systematic trends caused by

[1]Technically, we convert from relative flux uncertainties into magnitude space as in Naylor et al. (2002), to which Eq. 4.1 is an accurate Taylor approximation.

[2]We take r magnitudes from the Carlsberg Meridian survey (Evans et al. 2002), and J magnitudes from 2MASS (Skrutskie et al. 2006).

this second-order extinction effect. This PWV problem has been noted before as a limitation for cool objects observed in the NIR (Bailer-Jones & Lamm 2003; Blake et al. 2008). Recently, Blake & Shaw (2011) showed that GPS water vapor monitoring could be used to correct for the influence of PWV variations, improving both relative and absolute photometric accuracy of SDSS red star photometry.

While we do not have a GPS water vapor monitor, we can track the impact of PWV variations on MEarth photometry using the ensemble of observations we gather each night, observations of red stars in fields of blue comparisons. Figure 4.1 shows all of the M dwarf light curves gathered by MEarth over one week, after applying basic differential photometry. These light curves (of different M dwarfs observed on different telescopes) move up and down in unison, reflecting water vapor changes in the atmosphere they all share. These trends correlate strongly with ground-level humidity and ambient sky temperature, which are rough tracers of PWV in the overlying column. As PWV variations within a night can mimic transit signals (e.g. the first panel of Figure 4.1), we must account for this effect when searching for planets. Fortunately, because these trends are shared among all our targets, we can estimate a "common mode" timeseries from the data themselves and use it to correct for these trends (see Section 4.4).

4.2.4 Science Observations

The observations of our target M dwarfs are scheduled automatically using an ad hoc dynamic scheduling algorithm. This algorithm weights the observability of targets with the usefulness of the data to the survey as a whole, prioritizing gap-free cadences while minimizing slewing overheads. Each star is tied to a particular telescope, for ease of

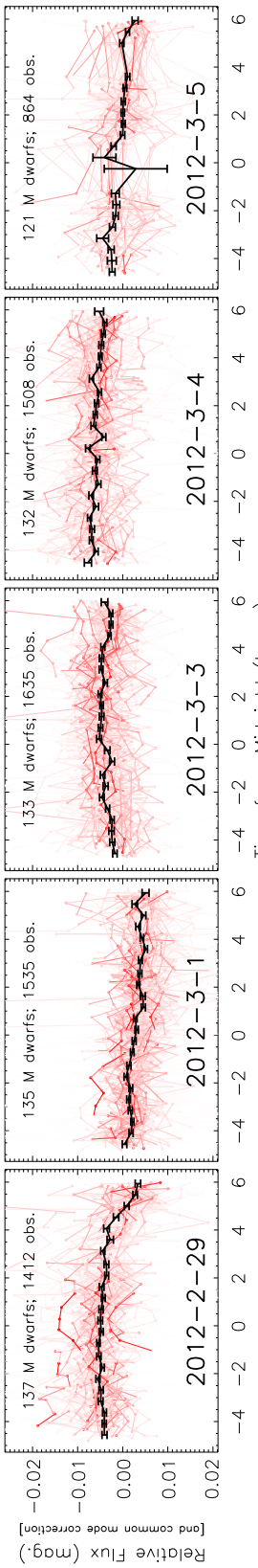

Figure 4.1: One week of light curves of all M dwarf targets observed on all MEarth telescopes (*red lines*), along with a 30-minute median-binned estimate of their shared behavior (*black line*, with uncertainty estimates). This "common mode" shows significant variations both within and between nights. We attribute this phenomenon to variations in the precipitable water vapor above our telescopes changing the effective shape of our wide bandpass, effectively causing more extinction for red M dwarfs than for their bluer comparison stars. The common mode correlates strongly with measured humidity and sky temperature (see Figure 4.2).

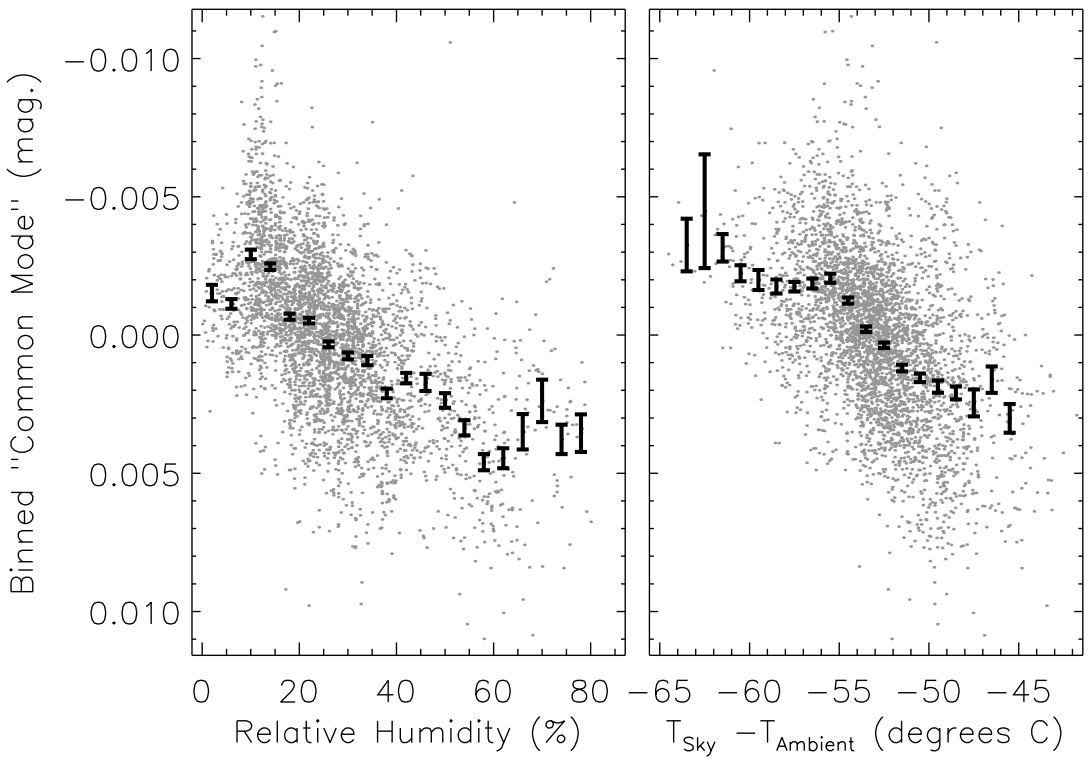

Figure 4.2: The relationship between the shared "common mode" behavior in MEarth photometry and ground-level relative humidity (*left*) and the difference between sky and ambient temperatures (*right*), both very rough (and not necessarily linear) tracers of the total precipitable water vapor in the overlying column. For the entire 2011-2012 season, we show each quantity averaged over independent half hour intervals (see Figure 4.1) as gray points. Error bars indicate the mean and its standard error for the common mode in subdivisions of humidity or sky temperature. In our bandpass, M dwarfs appear fainter when levels of precipitable water vapor are higher.

calibration and light curve production.

To inform this scheduling, we estimate masses, radii, and effective temperatures for all stars in the MEarth sample (Nutzman & Charbonneau 2008). Based on these estimates, we set the observational cadence to be sufficient to obtain two in-transit points from a mid-latitude transit of habitable zone planet. Because M dwarfs are dense stars, their transit durations are short (typically about 1 hour), requiring us to observe each star once every 20 minutes.

Based on our estimated stellar radii, we set our exposure time for each star so that we will record as many photons as are necessary to for the transit of a $2R_\oplus$ planet to have a 3σ transit depth. In cases where the required exposure time exceeds 2 minutes or would cause the peak counts in the star to exceed half of the detector full-well capacity, we split the observation into multiple sub-exposures. Stars requiring more than 7 minutes per pointing are never observed. If the time to reach $2R_\oplus$ is less than 60 seconds (i.e. bright, late M dwarfs), we artificially increase the exposure time. For the analyses presented in this paper, we combine all observations taken in a single pointing using scaled inverse-variance weighted means.

The scheduler input list can be updated in real-time, allowing us to "trigger" high-cadence observations of the egress of interesting transit events that are detected in progress. By immediately gathering more observations in candidate transits, we can greatly magnify the significance of an initial 3σ detection or refute it entirely without having to wait for future transits. As currently implemented, the real-time trigger assesses the significance of ongoing transits after subtracting a fixed systematics model and harmonic variability model and inflating the theoretical error on the in-transit mean

with an uncertainty estimate on the baseline out-of-transit level that is exponentially weighted toward recent observations. This practical estimator may eventually be replaced by the method explored in this paper. Skimming each star with a minimal cadence and triggering on marginal candidates maximizes our overall efficiency and increases our sensitivity to long-period planets.

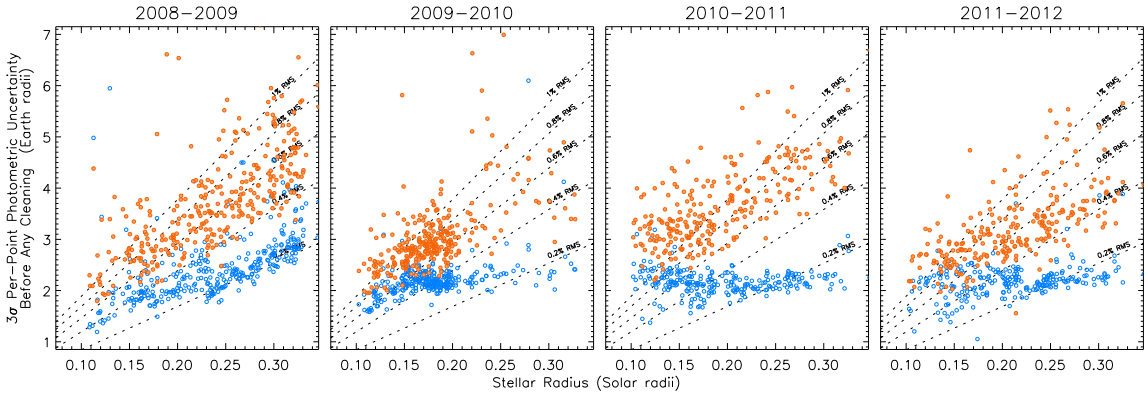

Figure 4.3: The per-point RMS photometric uncertainty as predicted from a CCD noise model (*blue open circles*) and the RMS actually achieved in the raw differential photometry (*orange filled circles*) as a function of estimated stellar radius for all MEarth targets. We cast the photometric uncertainty for each M dwarf target in units of the planet radius corresponding to a 3σ transit depth, given the inferred stellar radius; contours of constant RMS are shown for reference (*dashed lines*, equally spaced from 0.2 to 1%). Unlike wide-field surveys, we set exposure times individually for each target, minimizing the importance of apparent magnitude in these plots. The achieved RMS is shown before any treatment of systematics or stellar variability; see Figure 4.11 for comparison. One season of MEarth photometry is shown in each panel; Table 4.1 explains the causes of many of the year-to-year variations.

4.2.5 Morphological Description of the Light Curves

A typical MEarth light curve for a target M dwarf contains roughly 1000 observations spanning one observational season. Most of the time, the 20 minute cadence is continuous

within each night for the time a star has a zenith distance $< 60°$, but could be faster than this for up to several hours if a trigger occurred on the star. The cadence might also contain gaps within a night due to passing clouds or if a trigger occurred on another star observed by the same telescope. On longer timescales, in addition to gaps for daylight, light curves contain days- to months-long gaps from weather losses, instrumental failures, and scheduling conflicts (proximity to the Moon, other targets with higher priorities).

One useful summary of the challenge MEarth light curves present is our achieved RMS scatter. For all M dwarfs observed in the past four years, we show in Figure 4.3 both the RMS predicted with our CCD noise model (Eq. 4.1) and the RMS actually achieved after basic differential photometry has been performed. To emphasize the implications for planet detection, we cast the RMS in terms of the size of planet that could be identified at 3σ confidence in a single observation, given our stellar radius estimates.[3] This choice of parameter space (over the more common RMS vs. apparent magnitude space) reflects two of MEarth's unique aspects. First, we know more about our target stars than most wide-field surveys, enabling this translation from RMS (and a corresponding detectable transit depth) into detectable planetary radius. Second, because we set exposure times individually for each star, our sensitivity to transits does not depend on stellar apparent magnitude. The panels in Figure 4.3 show variations from year to year (see also Table 4.1), but all seasons of MEarth observations show significant gaps between the predicted and achieved noise. This indicates that stellar variability and systematics dominate over photon noise, highlighting the need for robust

[3]Specifically, the vertical axis in Figure 4.3 is given by $\sqrt{3\sigma} \times R_\star$, where σ is the relative flux uncertainty in a single observation (either predicted or achieved) and $R_\star$ is the estimated stellar radius.

method to correct for these complicated noise sources in our search for transits.

The source of excess scatter is sometimes known and sometimes unknown. By eye, some of the excess noise is clearly astrophysical, e.g. sinusoidal modulations from starspots rotating in and out of view or flares abruptly appearing then slowly decaying. Some is clearly instrumental, in that it can be associated with externally measured variables like position on the detector, weather parameters, or the behavior of other stars. Photometric outliers can often be associated with wind shake, where images exhibit broad and misshapen point spread functions. And lastly, some of the noise appears simply as unstructured excess scatter; this is caused either by astrophysical variability on timescales shorter than 20 minutes or by unidentified systematics.

4.3 Background

The problem of finding and assessing the significance of transiting exoplanet candidates in stellar photometry is an old one, and one that has already met many successful solutions. At their core, the majority of these solutions are variants of the matched filter Transit Detection Algorithm originally proposed by Jenkins et al. (1996), in which detection statistics are generated by matching light curves to families of templates consisting of periodic trains of transit-shaped pulses. The simplest and most intuitive of these methods is the Box-fitting Least Squares (BLS; Kovács et al. 2002), which models transits as simple boxcars in otherwise flat light curves. BLS identifies interesting candidates by folding individual photometric observations to trial periods, searching a grid of transit epochs and durations at each period, and picking the parameters that maximize the transit depth significance in a least-squares or χ^2 sense. As discussed by

Table 4.1:: Evolution of MEarth Hardware/Software

Season	Notes
2008-2009	Telescopes were operated purposely out-of-focus to minimize readout overheads, and exposure times were generally set to the maximum for a single (defocused) image, about 250,000 photons. The real-time trigger did not operate on-sky.
2009-2010	After repeated focus mechanism failures resulted in many light curves experiencing large focus drifts, telescopes have been operated in or near focus since early in this season with the use of sub-exposures to avoid overexposure. Smaller stars were prioritized in the scheduling queue, as noticeable in Figure 4.3. The real-time trigger began operating in November but was not always active due to development efforts.
2010-2011	In an attempt to remove systematics due to PWV, we operated during this year with a narrower filter ($715 - 895$nm, roughly I_C in shape) designed to avoid strong telluric water features. Unfortunately, the interference cutoff of this filter was found to be sensitive to humidity and temperature, resulting in larger common mode variations and higher systematic noise in the light curves (see Figures 4.3 and 4.6). Scattered light was also more pronounced with these filters, spurring our multi-part flat-fielding procedure. The real-time trigger improved in its response time and its treatment of variability and the common mode.
2011-2012	We returned to using the original MEarth 715 nm long-pass filter, but maintained the software improvements developed from the previous year. Dark flocking material affixed to the telescope baffles suppressed some of the scattered light. The real-time trigger operated normally for most of the season.

Aigrain et al. (2004), many other matched filter methods (Doyle et al. 2000; Defaÿ et al. 2001; Aigrain & Favata 2002; Street et al. 2003; Carpano et al. 2003) are essentially generalizations of BLS.

By assuming a flat out-of-transit light curve, BLS by itself can have a tendency to fold up any (non-planetary) time-correlated structures into seemingly significant candidates, when applied to real, wiggly light curves. As such, BLS is often paired with some sort of pre-search cleaning step to remove trends that could be caused either by instrumental effects or intrinsic stellar variability.

To deal with systematics, algorithms such as the Trend Filtering Algorithm (TFA; Kovács et al. 2005) and the principal component analysis-like Systematics Removal method (SysRem; Tamuz et al. 2005) were developed to remove trends that are present in multiple stars in a field and thus presumably not astrophysical. These algorithms use linear combinations of comparison star light curves to minimize the scatter in target stars. While these methods can remove trends without explicit knowledge of their causes, the trends do sometimes cluster into families that can be identified with physical processes (e.g. Kim et al. 2009). Unfortunately, strategies like TFA that work by constructing templates out of large numbers of field stars are of limited use for MEarth, with its small field of view and the substantial spectral type difference between our targets and comparisons. Methods that include known physical effects through linear models of externally measured variables (Bakos et al. 2010; Ofir et al. 2010) are more helpful for MEarth-like data.

Of course, systematics can also generally be minimized by improving various elements of the photometric reduction, observational strategy, or instrumentation. When

it can be done at reasonable cost, this is always preferable to applying filtering methods after the fact, because filtering inevitably suppresses the desired signal in addition to the noise.

To clean stellar variability from light curves, many methods were developed in preparation for space transit surveys like CoRoT and Kepler, where precision photometry makes it a dominant concern (e.g. Defaÿ et al. 2001; Jenkins 2002; Carpano et al. 2003; Aigrain & Irwin 2004; Régulo et al. 2007; Bonomo & Lanza 2008). These methods operate in the time, wavelet, or Fourier domains; many of them assume uniform photometric uncertainties and uniform cadence, as can realistically only be achieved from space. Running median filters (e.g. Aigrain & Irwin 2004) or piece-wise polynomial/spline fits (e.g. Croll et al. 2007) have also proven effective for removing smooth variability from high S/N light curves. Ground-based surveys for planets in open clusters motivated new methods to remove large amplitude variability from light curves with diurnal gaps (Street et al. 2003; Bramich et al. 2005; Burke et al. 2006; Aigrain et al. 2007; Miller et al. 2008), often by fitting series of sinusoids or allowing slowly varying baselines. We refer the reader to reviews and comparisons of these methods by Tingley (2003), Aigrain & Irwin (2004), and Moutou et al. (2005).

NASA's spaced-based Kepler Mission published the first Earth-sized planets (Fressin et al. 2012) and over 2,300 transiting planet candidates (Batalha et al. 2013) at the time of this writing. This success is thanks both to the design and stability of the spacecraft and to the sophistication with which the Kepler team accounts for its noise sources. To identify candidate transiting planets and assess their significance, Kepler employs a wavelet-based matched filter that is both optimal and efficient (Jenkins 2002; Tenenbaum et al. 2012). The Kepler Pre-search Data Conditioning pipeline can also

disentangle instrumental systematics from stellar variability using a linear model like the ones above paired with a Maximum A Posteriori approach employing empirical priors on the decorrelation coefficients to prevent over-fitting (PDC-MAP; Smith et al. 2012; Stumpe et al. 2012). Unfortunately, due to the need for uniformly spaced data to run the wavelet filter, the applicability of the Kepler transit-search method is limited in ground-based observations.

The coupling of many of the above cleaning methods with the BLS search has proven extremely successful for wide-field surveys. Using these methods, surveys such as TrES (Alonso et al. 2004), HATNet (Bakos et al. 2004), XO (McCullough et al. 2005), WASP (Pollacco et al. 2006) and KELT (Siverd et al. 2012; Beatty et al. 2012) have made the ground-based detection of hot gas giants transiting Sun-like stars routine. Amidst these successes, why should we bother to develop new methods?

Most of the above cleaning methods that are suitable for use from the ground work by subtracting some optimized model for systematics and variability from a target light curve. Subtracting this model inevitably introduces some extra uncertainty to the light curve: a cleaned light curve cannot possibly be as reliable as a light curve that did not need to be cleaned in the first place. However, these methods generally do not include a route for propagating the uncertainty from this cleaning into the significance of candidate transits. When many transits will be folded into into a planet candidate, this is okay. It is sufficient to know the average effect the cleaning has on the light curve, for example, that global filtering with TFA suppresses transit depths by 20% on average in HATNet (Bakos et al. 2013). In contrast, when only a single transit is available, knowing the average effect is not enough. We need to know: to what extent can we say that any one given dip is a bona fide eclipse and not the result of over- or under-correction by the

cleaning process?

We need a method that escapes the clean-first, search-later dichotomy of many of the previous methods. If the cleaning and the search are a two step process, the search knows nothing about how the cleaning has suppressed or exaggerated the apparent significance of transit-like features, making establishing rigorous detection thresholds very difficult. We need a reliable way to include our uncertainty in the corrections we make for systematics and variability in our search for planets; one way to do this is to combine the steps together, allowing the search to know about all the complicating details that go into the cleaning.

Here, we present a new method to detect single transits and robustly assess their significance. With 10–20 day M dwarf habitable zone planets offering at most a handful of observable transits per season, the ability to identify promising candidates with one or very few transits is absolutely necessary to our success. Here, we present a method for folding single transits into phased planet candidates, but we do not focus extensively on the problem finding the true periods of candidate systems. Although the challenge MEarth faces is not as bad as for the most sparsely sampled light curves (see Dupuy & Liu 2009; Tingley 2011; Dzigan & Zucker 2012), it will generally be extremely difficult to find accurate periods for 10–20 day planets from MEarth survey data alone. Rather, our goal is to be able to assess which candidates have high enough significance that they warrant the allocation of follow-up resources to, eventually, establish their periods and confirm their planetary nature.

4.4 Investigating a Single Eclipse: MISS MarPLE

We start by assessing the significance of an individual transit event within the context of a single night of observations of a star. We do so in the context of a parameterized, generative model for each target star light curve. This model contains parameters describing a simple box-shaped eclipse model, as well as parameters describing systematic effects plaguing the light curve and the star's intrinsic stellar variability.

Of the many parameters in this model, the depth D of a putative planetary eclipse is particularly important. We are interested in answering the following question: given a hypothetical lone planetary transit, with an epoch p_E and duration p_T, what is the probability distribution of the planetary eclipse depth D that the data imply? The integral of the normalized probability distribution $P(D|p_E, p_T)$ over the range $D > 0$ would provide a measure of the detection significance of the single eclipse. While $P(D)$ could generally take on any shape, we will approximate its shape to be Gaussian, so that we can completely characterize the distribution with two numbers, the maximum probability depth $\overline{D}$ and a width σ. In usual astronomical parlance, if $\overline{D}/\sigma > n$ then we have detected the eclipse "at $n\sigma$."

We want $P(D)$ to be conditional only on the parameters p_E and p_T; it should be marginalized over all other parameters to account for the additional uncertainty that each of these add to the width of the distribution. That is, we want $P(D)$ to be the Marginalized Probability of a Lone Eclipse (MarPLE), whose Gaussian width we will refer to as σ_{MarPLE}. In particular, because the transit depth could conceivably be quite correlated with the stellar variability or systematics parameters, marginalizing over these parameters will be crucial for a robust measure the eclipse depth uncertainty and thus

the significance of the detection. To achieve this goal, we outline a Method to Include Starspots and Systematics in the Marginalized Probability of a Lone Eclipse (MISS MarPLE) below.

4.4.1 The Model

At the core of MISS MarPLE is a model that attempts to describe every aspect of a single night of MEarth photometry of a single M dwarf. We use $d(t)$ to refer to the "data" sampled at time t: the relative flux measurements of the target star after basic differential photometric corrections have been applied. The two main aspects of the model are an idealized, noiseless light curve $m(t)$ and the uncertainty associated with a data point at any given time $\sigma(t)$. This model is generative, in the sense that fake light curves created with this model aim to be statistically equivalent to real MEarth light curves. Even if the model is an incomplete description of $d(t)$, it will still be useful for estimating the significance of a given candidate by allowing us to fit for and marginalize over the model parameters.

Throughout the following sections, light curves such as $m(t)$ and $d(t)$ will be expressed in magnitudes, so that effects that are multiplicative in flux can be described as linear models.

We write the model for the idealized, noiseless light curve as

$$m(t) = S(t) + V(t) + P(t) \tag{4.2}$$

where $S(t)$ models trends caused by instrumental systematics, $V(t)$ models the variability of the star in the absence of planetary transits, and $P(t)$ models the signal from a

hypothetical transiting planet.

Systematics Model

The $S(t)$ term in Eq. 4.2 enables us to include systematic trends that show clear correlations with externally measured variables. We construct $S(t)$ as a linear combination of $N_{\rm sys}$ relevant external templates:

$$S(t) = \sum_{j=1}^{N_{\rm sys}} s_j E_j(t). \tag{4.3}$$

Here $E_j(t)$ represent timeseries of the external variables, sampled at the times as the photometric observations, and the s_j are systematics coefficients. For MEarth, at a bare minimum, we include $N_{\rm sys} = 6$ terms in this sum: the "common mode," the "meridian flip", and the x and y pixel positions on either side of the meridian.

$E_{\rm CM}(t)$ The common mode template is constructed from the ensemble of raw M dwarf light curves from all telescopes and accounts for photometric trends that are shared in all MEarth M dwarf photometry (due to PWV variations, see Figures 4.1 and 4.2). The effect is stronger for redder stars; for MEarth targets, the best fit values of the coefficient $s_{\rm CM}$ correlates with stellar $r - J$ color.

$E_{\rm merid}(t)$ To account for stars sampling different regions of the detector when observing at positive or negative hour angles with MEarth's German Equatorial mounts, we include a "meridian flip" template. This template is simply defined as 0 for observations taken in one orientation and 1 for observations in the other, thus allowing light curves on two sides of the meridian flip to have different baselines.[4]

[4]In practice, we also allow additional offsets corresponding to each time a camera is

$E_{x,i}(t)$ and $E_{y,i}(t)$ for i=0,1 The pixel position templates are simply the x and y centroids of the target star on the detector, with their medians subtracted. Two sets are required, one for each side of the meridian. Correlations with these templates could arise as pointing errors allow a star to drift over uncorrected small-scale features in the sensitivity of the detector (e.g., transient dust donuts).

Additional external variables may also be used as systematics templates, such as FWHM or airmass. With MEarth, we find these variables are correlated with the photometry for only a few fields, and are usually excluded.

Variability Model

The $V(t)$ term in Eq. 4.2 describes the variability of the star throughout one night, independent of the presence of a transiting planet. Such variability includes fluctuations due to rotating spots (smoothly varying on the 0.1 to 100 day timescale of the star's rotation period) and flares (impulsively appearing, with a decay timescale typically of hours). The morphology of this variability can be quite complicated; we use a simplified model to capture its key features, writing

$$\begin{aligned} V(t) = \quad & v_{\rm night} + \\ & v_{\rm sin} \sin\left(\frac{2\pi t}{v_P}\right) + v_{\rm cos} \cos\left(\frac{2\pi t}{v_P}\right) + \\ & \sum_{j=1}^{N_{\rm flares}} f_j(t). \end{aligned} \tag{4.4}$$

taken off of its telescope. This is implemented as a simple extension of the $E_{\rm merid}(t)$ term described here.

The first term $v_{\rm night}$ allows each night to have its own baseline flux level. By itself, this term can capture most of the variability from stars with long rotation periods, where the flux modulation from starspots smoothly varies over timescales much longer than one night. By fitting for a different $v_{\rm night}$ for each night we can piece together the variability of the star on timescales > 1 day as a series of scaled step functions. The harmonic $v_{\rm sin}$ and $v_{\rm cos}$ terms capture variability with period of v_P and become especially important for stars with shorter rotation periods. Because we fit a separate $v_{\rm night}$ for each night, there can be substantial degeneracy between the harmonic terms and the nightly offsets, especially for slowly rotating stars. We discuss this issue, as well as how we estimate v_P in Section 4.4.2. Although we only include one harmonic of the fundamental period v_P in these sinusoidal terms, additional harmonics could be included if the data warranted them.

The final term in Eq. 4.4 includes contributions from $N_{\rm flares}$ hypothetical stellar flares $f_j(t)$ that may or may not be present within the night. Flares are suppressed in MEarth's relatively red bandpass, but not completely eliminated (see Tofflemire et al. 2012). The main purpose of the flare term is to identify those nights of photometry that may be corrupted due to the presence of flares. While we could in principle model a night that contained both flares and a transit, we find this to be very difficult in practice, due to the morphological complexity flares sometimes exhibit (e.g. Kowalski et al. 2010; Schmidt et al. 2012). Rather, on each night we model simple hypothetical flares as fast-rising and exponentially decaying, and perform a grid search over the start time and decay timescale. If any flares have amplitudes that are detected at $> 4\sigma$, we excise that night of data from our planet search. These cuts dramatically reduces planetary false positives due to flaring activity (e.g. confusing the start of the flare with the egress of a

transit), with the meager cost of ignoring 3.6% of MEarth's observations. Because both planetary transits and flares are rare in MEarth data, and their overlap even moreso, the losses from this strategy are small.

Planetary Eclipse Model

The last term in Eq. 4.2, $P(t)$, includes the signal of a hypothetical transiting planet. We model transits as having infinitely short ingress/egress times and ignore the effects of limb-darkening on the host star, so transits appear as simple boxcars. In this section, we are interested only in assessing the significance of a single transit event falling within a single night, not a periodic train of transits. With these simplifications, a lone planetary eclipse signal is completely described by a transit epoch p_E, a transit duration p_T, and a transit depth D. The signal is then simply

$$P(t) = \begin{cases} D & \text{if } |t - p_E| < p_T/2 \\ 0 & \text{otherwise} \end{cases} \tag{4.5}$$

This model includes only one eclipse event per night. We discuss combining these lone eclipses into periodic transit candidates in Section 4.4.3.

Photometric Uncertainty Model

A crucial component of the model is $\sigma(t)$, the photometric uncertainty of each observation. Our theoretical uncertainty estimate for a given datapoint $\sigma_{\rm the}(t)$ is a lower limit on the true uncertainty. To express this fact, we introduce a noise rescaling parameter $r_{\sigma,w}$ such that

$$\sigma(t) = r_{\sigma,w}\sigma_{\rm the}(t). \tag{4.6}$$

where $r_{\sigma,w} \geq 1$. The subscript w emphasizes that this is a *white* noise rescaling parameter that does not account for correlations between nearby data points. If left unmodelled, such *red* noise could substantially bias a transit's detection significance (Pont et al. 2006); we discuss a correction for red noise in §4.1.

4.4.2 The Posterior Probability

For a reasonable choice of parameters, the model in Eq. 4.2 could generate a fake light curve that would have most of the features of single night of a real MEarth light curve. But how do we pick a reasonable choice of parameters? In this section, we write down their probability distribution and show how to solve for its peak, which turns out to be a linear minimization process with slight iterative refinement.

Considering a single night of observations, we write the shape of the probability distribution of these parameters as

$$P(\mathbb{M}|\mathbb{D}) \propto P(\mathbb{D}|\mathbb{M})P(\mathbb{M}), \tag{4.7}$$

where $P(\mathbb{M}|\mathbb{D})$ is the posterior probability of the model $\mathbb{M}$ given the data $\mathbb{D}$, $P(\mathbb{D}|\mathbb{M})$ is the likelihood of the data given the model, and $P(\mathbb{M})$ is the prior probability of the model. These functions describe probability density distributions that live in an n-dimensional hyperspace with as many dimensions as there are parameters in the model.

The Likelihood = $P(\mathbb{D}|\mathbb{M})$

We describe each of the $N_{\rm obs}$ photometric observations $d(t_i)$ within a particular night as being drawn from a Gaussian distribution centered on $m(t_i)$ and with a variance of

$\sigma(t_i)^2$. Assuming the observations to be independent, the likelihood can be written as

$$P(\mathbb{D}|\mathbb{M}) = \prod_{i=1}^{N_{\rm obs}} \frac{1}{\sqrt{2\pi}\sigma(t_i)} \exp\left[-\frac{1}{2}\left(\frac{d(t_i) - m(t_i)}{\sigma(t_i)}\right)^2\right].$$

Taking the logarithm, substituting Eq. 4.6, and defining

$$\chi^2 = \sum_{i=1}^{N_{\rm obs}} \left[\frac{d(t_i) - m(t_i)}{\sigma_{\rm the}(t_i)}\right]^2, \tag{4.8}$$

we find that the (log) likelihood simplifies to

$$\ln P(\mathbb{D}|\mathbb{M}) = -N_{\rm obs} \ln r_{\sigma,w} - \frac{\chi^2}{2r_{\sigma,w}^2} + \text{constant} \tag{4.9}$$

where we have only explicitly included terms that depend on the parameters of the model. For fixed $r_{\sigma,w}$, maximizing Eq. 4.9 is equivalent to minimizing the commonly used χ^2 figure of merit.

The Prior = $P(\mathbb{M})$

For any one star, a particular night of MEarth photometry may contain roughly as many light curve points as there are parameters in our model. As such, the likelihood $P(\mathbb{D}|\mathbb{M})$ from one night of data only very weakly constrains the parameter space. But of course, each night of MEarth observations is just one of many nights spanning an entire season, and we should use this season-long information when investigating a single night. To implement this holistic awareness of the context provided by a large pool of observations, we generate probability distributions for various parameters by looking at the whole season of data. We then apply them as priors $P(\mathbb{M})$ on the parameters for an individual night.

By construction, the most important parameters of our model are linear parameters. The conditional likelihood of linear parameters (a slice through $P(\mathbb{D}|\mathbb{M})$ with other

parameters fixed) has a Gaussian form. For marginalization, it proves quite useful for the priors to be conjugate to this shape – that is, also take on a Gaussian form. Referring to these linear parameters with the vector $\mathbf{c} = \{D, v_{\rm night}, v_{\rm sin}, v_{\rm cos}, s_{\rm CM}, s_{\rm merid}, s_{x,i}, s_{y,i}, s_{\rm other?}\}$, we parameterize the prior $P(\mathbb{M})$ as being proportional to a Gaussian distribution in c_j that is centered on an expectation value $\overline{c_j}$ and with a variance of $\pi_{c_j}^2$. Multiplying the independent priors for the $N_{\rm coef}$ coefficients and defining

$$\Phi^2 = \sum_{j=1}^{N_{\rm coef}} \left(\frac{c_j - \overline{c_j}}{\pi_{c_j}} \right)^2 \tag{4.10}$$

leads to a term in the prior that looks like

$$\ln P(\mathbb{M}) = -\frac{1}{2}\Phi^2 + \cdots \tag{4.11}$$

The similarity in form of Φ^2 to χ^2 is the reason that the use of conjugate Gaussian priors is often described along the lines of "adding a prior as an extra data point in the χ^2 sum," because the effect is identical in the overall posterior. In this framework, the smaller values of π_{c_j} provide tighter constraints on the parameter; we could express a flat, non-informative prior for a particular c_j by choosing a large value of π_{c_j}. We set $\pi_D = \infty$, giving a flat prior on the transit depth. Note that we allow negative transit depths (i.e. "anti-transits") to avoid skewing the null distribution of transit depths away from 0.

For most of the remaining linear parameters, we take the values of $\overline{c_j}$ and π_{c_j} directly from the results of a simultaneous fit to the star's entire season of observations. In this prior-generating season-long fit, we fit the season-long light curve with a modified version of Eq. 4.2 that excludes both the $f_j(t)$ term from flares and the $P(t)$ term from hypothetical planets. To immunize against these unmodelled flares and eclipses, we perform the fit with 4σ clipping. We prefer to explain as much of the long-term

variability as possible with the harmonic terms, so we fit first including only these terms in $V(t)$. Then, fixing the values of $v_{\rm sin}$ and $v_{\rm cos}$, we fit again with one $v_{{\rm night},j}$ free parameter for each night represented within the season. Thus, the values of $v_{{\rm night},j}$ then represent the deviation of the nightly flux level from a baseline sinusoidal model.

Now, for the single night flux baseline parameter $v_{\rm night}$, we set $\pi_{v_{\rm night}}$ equal to $1.48 \times$ MAD (median absolute deviation) of the ensemble of $v_{{\rm night},j}$ values from the season fit. Stars that vary unpredictably from night to night will have a broad prior for $v_{\rm night}$, thus requiring more data within a night to determine its baseline level. Conversely, stars that remain constant from night to night or have variability that is well described by a sinusoid will have a very tight prior.

To understand the impact of $\pi_{v_{\rm night}}$, imagine the following hypothetical scenario: a night in which MEarth gathered only one observation of a star, and that observation happened to fall in the middle of a transit with a 0.01 magnitude depth. With what significance could we detect this transit? If $\pi_{v_{\rm night}} = 0.01$ magnitudes, then the detection significance would be at most 1σ. But if $\pi_{v_{\rm night}} = 0.001$, then the transit could in principle be detected at high significance with only the single data point, provided the photon noise limit for the observation was sufficiently precise.

We note that $s_{x,i}$ and $s_{y,i}$, the coefficients for the x and y pixel position templates, would not be expected to be constant throughout a season. These terms are designed to account for flat-fielding errors, which could easily change from week to week or month to month. As such, we do not take c_j and π_j from the season-wide fit for these parameters. Rather, we fix $c_j = 0$ and $\pi_j = 0.001$ for all four of these parameters. This has the desired effect that an apparent 0.005 magnitude transit event that is associated with

simultaneous 5 pixel shift away from the star's mean position on the detector would not be considered as a significant event.

The most significant non-linear parameter is the white noise rescaling parameter $r_{\sigma,w}$. In the season-long fit, Eq. 4.9 indicates that $P(\mathbb{M}|\mathbb{D})$ would have a shape of

$$\ln P(r_{\sigma,w}|\mathbb{D}) = -N_{\rm sea} \ln r_{\sigma,w} - \frac{\chi^2_{\rm sea}}{2r^2_{\sigma,w}} \tag{4.12}$$

where $\chi^2_{\rm sea}$ is the season-long χ^2 from the $N_{\rm sea}$ observations in the ensemble fit. This is maximized when $\overline{r} = \sqrt{\chi^2_{\rm sea}/N_{\rm sea}}$. We want the nightly prior on $r_{\sigma,w}$ to push it toward $\overline{r}$, but we also want to provide enough flexibility that nights that are substantially better or worse than typical can be identified as such. To implement this, we mimic the shape of the season-long probability distribution but artificially broaden it with an effective weighting coefficient $N_{\rm eff}$. Propagating this loose prior

$$\ln P(\mathbb{M}) = \cdots - N_{\rm eff} \ln r_{\sigma,w} - \frac{N_{\rm eff}\overline{r}^2}{2r^2_{\sigma,w}} + \cdots \tag{4.13}$$

into the posterior for an individual night, the Maximum A Posteriori (MAP) value of $r_{\sigma,w}$ will be

$$r_{\sigma,w} = \sqrt{\frac{\chi^2 + N_{\rm eff}\overline{r}^2}{N_{\rm obs} + N_{\rm eff}}}. \tag{4.14}$$

We artificially set $N_{\rm eff} = 4$, so on nights with fewer than 4 observations, the MAP value of $r_{\sigma,w}$ will be weighted most toward what the rest of the season says. On nights with more than 4 observations, the data from the night itself will more strongly drive the MAP value.

We use a modified periodogram (Irwin et al. 2011a) as part of the season-wide fit to identify the best value of $\overline{v_P}$, the period of the harmonic terms in $V(t)$. We fix v_P to this value in all later analysis. While this effectively places an infinitely tight prior on this

parameter, the degeneracy between it and the other variability parameters, especially on the timescale of a single night of data, means that its uncertainty is usually accounted for by those terms.

Maximizing and Marginalizing

The shape of $P(\mathbb{M})$ offers a big advantage to our goal of estimating the marginalized transit depth probability distribution. Accounting for all the terms in $P(\mathbb{M}|\mathbb{D})$ (Eq. 4.7, 4.9, 4.11, and 4.13), we find that the posterior $P(\mathbb{M}|\mathbb{D})$ can indeed be maximized and marginalized analytically. For fixed p_E, and p_T, the system of equations

$$\frac{\partial}{\partial c_j} \ln P(\mathbb{M}|\mathbb{D}, p_E, p_T) = \frac{\partial}{\partial c_j} \left(\frac{\chi^2}{r_{\sigma,w}^2} + \Phi^2 \right) = 0 \tag{4.15}$$

can be solved exactly for the MAP vector of values $\mathbf{c}_{\rm MAP}$ using only simple matrix operations. The procedure is directly analogous to the problem of weighted linear least squares fitting; see Sivia & Skilling (2006, ch. 8) for details of this solution. While not strictly necessary because the priors prevent unconstrained degeneracies in the solution, we use singular value decomposition (SVD) to avoid catastrophic errors in the matrix inversions (Press 2002).

The value of $r_{\sigma,w}$ sets the relative weighting between the likelihood and the prior. Thus it is important to estimate $r_{\sigma,w}$ accurately. We solve for it by iterating between Eq. 4.14 and Eq. 4.15; the solution typically converges to the MAP value within only a couple of iterations. We forego marginalizing over $r_{\sigma,w}$, instead fixing it to its MAP value. Solving for $r_{\sigma,w}$ independently on each night is a better approximation than blindly assuming a global value.

Importantly, the matrix solution to this problem gives not only the MAP values,

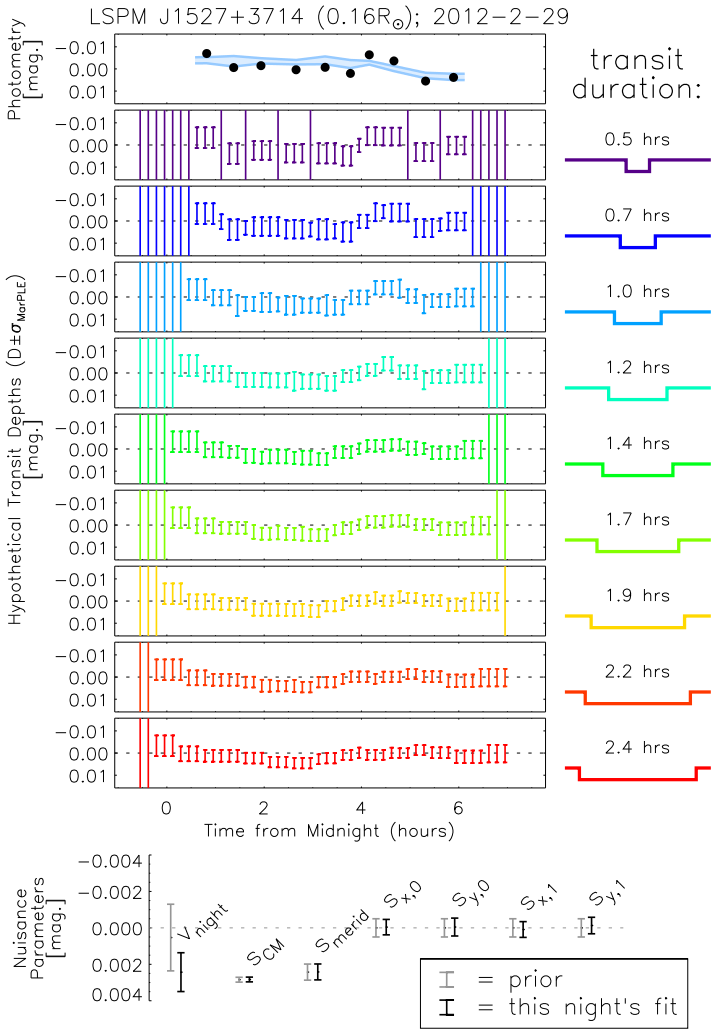

Figure 4.4: A demonstration of MISS MarPLE applied to one night of MEarth observations, with the star's identifier, inferred stellar radius, and the date indicated above. We show the original MEarth photometry (*top panel*, black points), a variability and systematics model with no transits included (*same panel*, $\pm 1\sigma$; blue swath), and a visualization (*next 9 panels*) of the probability distribution of hypothetical transit depths $P(D|p_E, p_T)$. In this visualization, error bars represent the central $\pm 1\sigma$ confidence regions of the marginalized, Gaussian-shaped $P(D|p_E, p_T)$ for the denoted values of the eclipse epoch p_E (along the time axis) and transit duration p_T (in separate panels). Some of the nuisance variability and systematics parameters over which $\sigma_{\rm MarPLE}$ has been marginalized are also shown (*bottom panel*), with the season-long priors in gray and the fits from this night in black. Note that the dip in photometry at the end of the night is not seen as a significant, because it can be explained as a systematic, in this case a variation in the common mode (see the first panel of Figure 4.1, from the same night).

it also gives the covariance matrix of the parameters in the fit, which is an exact representation of the shape of $P(\mathbb{M}|\mathbb{D}, p_E, p_T)$, which is a multidimensional Gaussian. The diagonal elements of this covariance matrix give the uncertainty in each parameter *marginalized over all the other linear parameters*. Because we have constructed our model in such a way that the parameters that most strongly influence estimates of the transit depth D are linear, we can use this analytical solution as a robust estimator the shape of the Marginalized Probability of a Lone Eclipse. It gives us both the maximum a posteriori transit depth $\overline{D}$ and the Gaussian width of the distribution σ_{MarPLE}.

Are the Priors Really Priors?

As the priors we use to regularize our model fits are themselves derived from MEarth data, one might object that the division between the likelihood and the prior is set somewhat arbitrarily. We include data only from a single night in the likelihood and group all the information from the rest of the nights into the prior. Indeed, we could have instead organized the entire season of data into the likelihood and left the priors uninformative. The division is arbitrary, but useful.

The advantages of treating nights other than that on which a candidate transit falls as external to likelihood are two-fold. First, it is more computationally efficient: instead of recalculating the likelihood of an entire season's data when investigating individual events, we only need to calculate the likelihood over the relevant night's data points. The information provided by the entire season changes little from candidate transit to candidate transit; thus it is best to store that information as a pre-computed prior.

Second, this organization scheme allows the flexibility for individual nights to behave

differently. For example, consider the pixel position $E_{x,i}(t)$ and $E_{y,i}(t)$ terms in the systematics model, which capture the influence of stars wandering across the detector. As the detector flat-field can change from night to night, it would be foolish to try to fit an entire season's light curve with one set of coefficients for $E_{x,i}(t)$ and $E_{y,i}(t)$; allowing those coefficients to vary from night to night, within a tightly constrained prior, is a more useful approach. Furthermore, dividing the weight of the likelihood and priors as we do provides a helpful degree of outlier resistance, by not forcing the model on any one night to account for strange behavior on one weird night from months before.

MarPLE in Practice

We calculate $\overline{D}$ and σ_{MarPLE} on a grid of single transit epochs p_E and durations p_T. We construct this grid for all nights with usable MEarth data. The epochs in this grid are evenly spaced by $\Delta p_E = 10$ minutes, thus subsampling the typical MEarth observational cadence. The durations are evenly spaced from 0.02 to 0.1 days, spanning the likely durations for the orbital periods to which MEarth has substantial sensitivity. We extend the grid of p_E before the first and after last observation of each night by half the maximum transit duration, thus probing partial transits.

We demonstrate this calculation graphically in Fig. 4.4. First, we show one night of a typical MEarth light curve. To give a sense of a baseline systematics and variability model, we show it the $\pm 1\sigma$ span of model light curves arising from a fit that contains no eclipses. Next, we show a visualization of the MarPLE, the probability distribution of hypothetical eclipse depths $P(D|p_E, p_T)$. For any chosen value of eclipse epoch and duration, the MarPLE is Gaussian-shaped; error bars in Fig. 4.4 show its central $\pm 1\sigma$

width over the entire grid of p_E and p_T. The width of $\sigma_{\rm MarPLE}$ can be seen to decrease for longer durations p_T, as more data points are included in each transit window. For epochs and durations with no in-transit points, the transit depth is unconstrained and $\sigma_{\rm MarPLE} \to \infty$. The transit-like dip in photometry at the end of the night does not register as significant anywhere in the MarPLE. The dip can be explained by MEarth's precipitable water vapor systematic (see first panel of Figure 4.1, corresponding to the same night).

At the bottom of Fig. 4.4 we also include a subset of the variability and systematics parameters, the "nuisance parameters" over which we marginalize. We show error bars representing the Gaussian widths of both the prior $P(\mathbb{M})$, established from the entire season of data, and the results of a fit to this one night of data, $P(\mathbb{M}|\mathbb{D})$. For $v_{\rm night}$, the nightly out-of-transit baseline level parameter, one night's data are more influential than the relatively weak prior, so the fit is notably offset from and tighter than the prior. In contrast, for the remaining nuisance parameters, the influence of one night's data is very weak, so the fit essentially reverts to the input priors.

In this example, only 10 data points are contributing to the likelihood. As the model contains almost as many parameters, one might be concerned that we are "over-fitting" the data. The bottom of Fig. 4.4 provide an initial step to allay this concern, emphasizing that except for the transit depth, each parameter in the fit has its associated prior that provides its own independent constraint on the parameter. In a pseudo least squares formalism, the presence of these informative priors act as (pseudo) data points, ensuring there are always more "data" than parameters.

Furthermore and perhaps more importantly, we could indeed be in severe danger of

over-fitting if we were interested in the exact values of the cleaned residuals from some single estimate of a best-fit systematics and variability model, but we are safe because we care instead about the marginalized probability of only one particular parameter (the transit depth). Marginalization ignores irrelevant information, so we can include an arbitrary number of nuisance parameters in the fit (see Hogg et al. 2010, for discussion). If (and only if) the inferred transit depth at any particular p_E and p_T happens to be strongly covariant with one of these nuisance parameters, then σ_{MarPLE} will include a contribution from that parameter. Without the priors degeneracies could potentially inflate σ_{MarPLE} to ∞, but with the informative season-long priors the nuisance parameters can only vary within the range shown in Fig. 4.4, limiting the degree to which they can in turn contribute to σ_{MarPLE}. In the extreme example, if we had a single data point on a night, the cleaned residuals might easily be identically zero (i.e. "over-fit") but the MarPLE would accurately express what the night told us about the presence or absence of transits.

Estimating σ_{MarPLE} across the whole grid of p_E for an entire season can be performed very quickly. Each grid point requires only several SVD's of a matrix whose dimension is the sum of the number of data points within the night and the number of linear parameters being fit. For a MEarth light curve containing 1000 points and spanning 100 days, the whole grid of calculations requires several seconds on a typical desktop workstation.

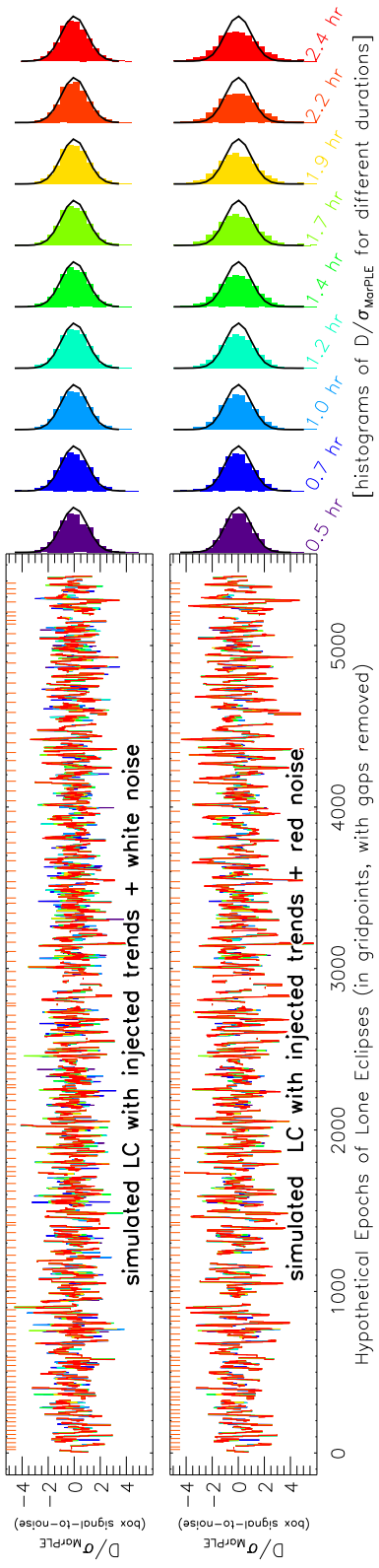

Figure 4.5: A demonstration of the effect of red noise on the inferred significance of transits, showing the signal-to-noise ratio of hypothetical transits with all possible epochs (along the x-axis) and durations (denoted by color). D, $\sigma_{\rm MarPLE}$, and colors are the same as in Figure 4.4. Results are shown for fake light curves generated from the time-stamps of a real MEarth target assuming either white Gaussian noise (*top*) or correlated red noise (*bottom*), before applying the red noise correction described in 4.1. Histograms (*right*) indicate that uncorrelated white noise leads to $D/\sigma_{\rm MarPLE}$ following a unit-variance Gaussian distribution (*black curves*) for all durations, whereas red noise in the light curve broadens the distribution, especially for long duration transits. If uncorrected, this would cause us to overestimate the significance of candidate transits.

Ad Hoc Red Noise Correction

The likelihood in Eq. 4.9 assumed that adjacent light curve data points were statistically independent. If our method fails to completely correct for systematics or stellar variability, this assumption will be violated. Time-correlated noise slows the $\sqrt{N}$ improvement that would be gained by obtaining N independent Gaussian measurements. So, if we were to ignore the temporal correlations between data points, we could substantially bias our estimates of $\sigma_{\rm MarPLE}$.

Specifically, correlated noise would cause us to overestimate the significance of transits that spanned multiple data points. We demonstrate this phenomenon in Fig. 4.5, which shows the MarPLE results for two simulated light curves (generated from the real time stamps of a typical MEarth target) with different levels of correlated noise. One light curve consists of pure white Gaussian noise. The other consists of the white light curve averaged with a smoothed version itself, scaled so both light curves have an identical RMS, roughly approximating a finite red noise contribution. We then inject common mode and meridian flip trends are injected into both light curves. The toy-model simulation of time-correlated noise is very coarse but is only meant to serve an illustrative purpose.

Because each estimate of $\overline{D}$ is drawn from a Gaussian distribution with a width $\sigma_{\rm MarPLE}$, the quantity $\overline{D}/\sigma_{\rm MarPLE}$ should ideally be Gaussian-distributed around 0 with a variance of 1, except when real transits are present. For the light curve with pure white noise, this is true for all transit durations in Fig. 4.5 (see the histograms at right). For the light curve with significant correlated noise, we underestimate $\sigma_{\rm MarPLE}$ and the distribution of $\overline{D}/\sigma_{\rm MarPLE}$ appears broadened for some durations. The effect is most

pronounced at longer durations, where more data points fall within each transit. For the shorter durations, typically only one or two light curve points fall within a transit so the red noise does not substantially affect our estimate of $\sigma_{\rm MarPLE}$. This general behavior, of overestimating the significance of longer duration transits, is common among MEarth targets whose light curves show features that are poorly matched by the input model.

To account for the problem, we posit that each light curve has some additional red noise source that can be expressed as a fixed fraction of the white noise, defining $r_{\sigma,r}$ as the ratio of red noise to white noise in a light curve. With this parameterization, the transit depth uncertainty associated with a transit that contains $N_{\rm tra}$ data points becomes

$$\sigma_{\rm MarPLE} = \sigma_{\rm MarPLE,w} \times \sqrt{1 + N_{\rm tra} r_{\sigma,r}^2} \tag{4.16}$$

where $\sigma_{\rm MarPLE,w}$ is the estimate of $\sigma_{\rm MarPLE}$ that accounted only for white noise. To determine its optimum value, we scale $r_{\sigma,r}$ until the distribution of $\overline{D}/\sigma_{\rm MarPLE}$ has a MAD of 1/1.48 (i.e. the distribution has a Gaussian width of unity). This correction is similar to the $\mathcal{V}(n)$ formalism described by Pont et al. (2006). Henceforth, when we use the term $\sigma_{\rm MarPLE}$, we are referring to its red-noise corrected value.

A more ideal solution would account for time-correlated noise directly in the likelihood (Eq. 4.9), but doing so would substantially decrease MISS MarPLE's computational efficiency. As such, we settle on Eq. 4.16 as a useful ad hoc solution. Fig. 4.6 shows the amplitude of $r_{\sigma,r}$ for all stars in the MEarth survey, indicating that most stars have low red noise contributions, after accounting for our stellar variability and systematics.

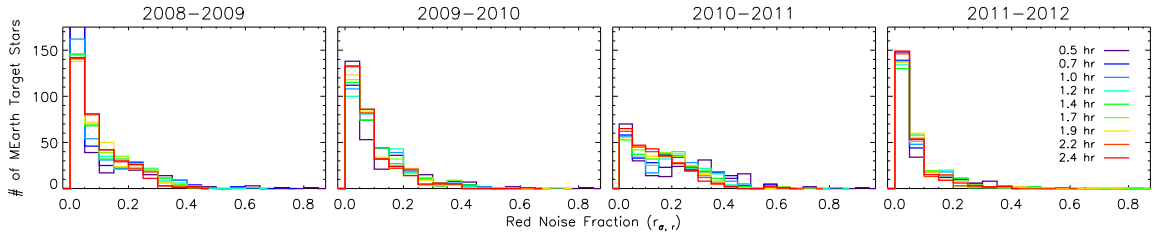

Figure 4.6: Histogram of the red noise rescaling factor $r_{\sigma,r}$ (see Eq. 4.16) in each of the four MEarth seasons of observations, estimated on different transit duration timescales. We experimented with a narrower filter in the 2010-2011 season in the hopes of alleviating our precipitable water vapor systematic; we found its long wavelength cutoff to be sensitive to humidity and temperature, exacerbating the problem and resulting in increased red noise for this year.

4.4.3 Phasing Multiple MarPLE's Together

MISS MarPLE, as just described, investigates the significance of a single transit event. The method can be straightforwardly extended to search for periodic transit candidates as well. Once $\overline{D}$ and σ_{MarPLE} have been calculated over a grid of p_E and p_T, characterizing periodic candidates is simply a matter of combining all precomputed lone eclipses from this grid that match the appropriate period p_P and starting epoch p_{E0}. In Kepler parlance, this is the step where Single Event Statistics are combined into Multiple Event Statistics (see Tenenbaum et al. 2012).

Given p_P and p_{E0}, we identify those values of p_E that fall within 5 minutes of this linear ephemeris and that have finite values of σ_{MarPLE}. Each lone eclipse carries its own Gaussian distribution in D. Multiplying these independent Gaussians together leads to the standard inverse-variance weighted average:

$$\overline{D_{\mathrm{phased}}} = \frac{\sum \overline{D_i}/\sigma^2_{\mathrm{MarPLE,i}}}{\sum 1/\sigma^2_{\mathrm{MarPLE,i}}} \tag{4.17}$$

$$\overline{\sigma_{\rm phased}}^2 = \frac{1}{\sum 1/\sigma^2_{\rm MarPLE,i}} \tag{4.18}$$

where the sums are performed over the $N_{\rm epoch}$ epochs that were observed for a given candidate period and starting epoch. If $\chi^2_{\rm phased} = \sum(\overline{D_i} - \overline{D_{\rm phased}})^2/\sigma^2_{\rm MarPLE,i}$ is greater than $N_{\rm epoch}$, we take it as an indication that the uncertainties would have to be underestimated if that candidate ephemeris were real. In this case, we rescale $\overline{\sigma_{\rm phased}}^2$ up by a factor of $\chi^2_{\rm phased}/N_{\rm epoch}$. In other words, we enforce that the independently measured transit depths that go into each phased candidate must agree to within their errors.

Because we evaluate the MarPLE on grid of epochs that is super-sampled with respect to both our observational cadence (20 minutes) and typical transit durations (0.5 to 2 hours), adjacent values of p_E will have highly correlated transit depth estimates. This simply reflects that a (complicated) binning over the transit duration p_T has already gone into these estimates. Whereas the above sums would be over all in-transit light curve points in a traditional BLS, with MISS MarPLE we include only one term in the sum for each independent event.

To perform a full search, we repeat this procedure on a grid of periods. Because we hope to identify planets with potentially very few events, it is absolutely crucial that we explore a fine enough grid in periods that we not miss any peaks in the probability distribution. We set Δp_P so that when moving from one period to the next, the first and last data points of a season move by 5 minutes with respect to each other in phase (leading to exponentially spaced candidate periods). MEarth target star mass and radius estimates are reliable to 30-35% (or better for those stars with parallaxes, see Nutzman & Charbonneau 2008); we use this information to search only up to the transit duration of a planet in a circular orbit with 0 impact parameter for each period.

This search is the most computationally intensive step in the process. Searching a typical MEarth season requires roughly 10^5 candidate periods and 10 minutes on a desktop workstation, using vectorized IDL code. Searching multiple seasons requires up to 10^6 periods, thus needing correspondingly more time.

A brief side note: cleaning methods like TFA or EPD can be run in a "reconstructive mode," in which they fit away systematics (Kovács et al. 2005) and/or variability (Kovacs & Bakos 2008) towards a known signal present in the data. Such reconstructive techniques have generally not been applied when running period searches for planets, because the computational cost of rerunning them for all possible transit periods, durations, and epochs is untenable. When calculating $P(D|p_E, p_T)$ with MISS MarPLE, we are performing an analysis that is in someways similar to a reconstructive TFA/EPD (i.e. fitting systematics and variability in the presence of a candidate transit). But in the case of MISS MarPLE, we first perform this analysis on individual transits using data from individual nights, and phase up the results to candidate periods second. Thus we postpone the combinatorics of the period search until after the costly matrix inversions.

The form of the weighted sums in Eq. 4.17 and 4.18 highlights an important feature. Events with few observations in a night, events that fall on nights with poor weather, events that correlate with the star's position on the detector, events at high airmass, events on nights where a star is acting weirdly – namely, bad events – will have large σ_{MarPLE}'s and be naturally down-weighted in the sum. In contrast, good events falling on well-sampled, well-behaved nights will get the credit they deserve, exactly as we want. The advantages extend even further, in that this sum can span beyond a single telescope or a single season, enabling the straightforward combination of data from multiple sources with multiple systematics and even at multiple wavelengths into a

coherent whole.

4.5 Results

We apply MISS MarPLE to real MEarth light curves for which we have at least 100 observations in a season, and discuss two aspects of the results here. First, we investigate the properties of simulated transits injected into MEarth light curves, in order to provide concrete examples and compare MISS MarPLE with other methods. Second, we show that the method behaves well when applied to the ensemble of real MEarth light curves and does not generate an overabundance of false positives.

Throughout this section, we occasionally point to $D/\sigma_{\mathrm{MarPLE}} = 3$ as a characteristic value of interest. MEarth light curves span typically a few thousand independent transit durations, so we expect to find several 3σ events by chance in each. However, a single candidate event identified at $D/\sigma_{\mathrm{MarPLE}} > 3$ significance would be sufficient to set off MEarth's real-time trigger, which would immediately gather new observations to confirm or deny the event. Triggered observations could potentially magnify the significance of the single transit until the chance of it being a false alarm is low: a single transit at 5σ should formally be expected by chance about once per 3.5×10^6 independent epochs tested, roughly comparable to the number of epochs probed across all the stars in the MEarth survey to date. These thresholds for single events are much lower than that required to eliminate false positives from a phased search for periodic candidates, which as we discuss in Section 4.5.2 is closer to 7 or 8σ.

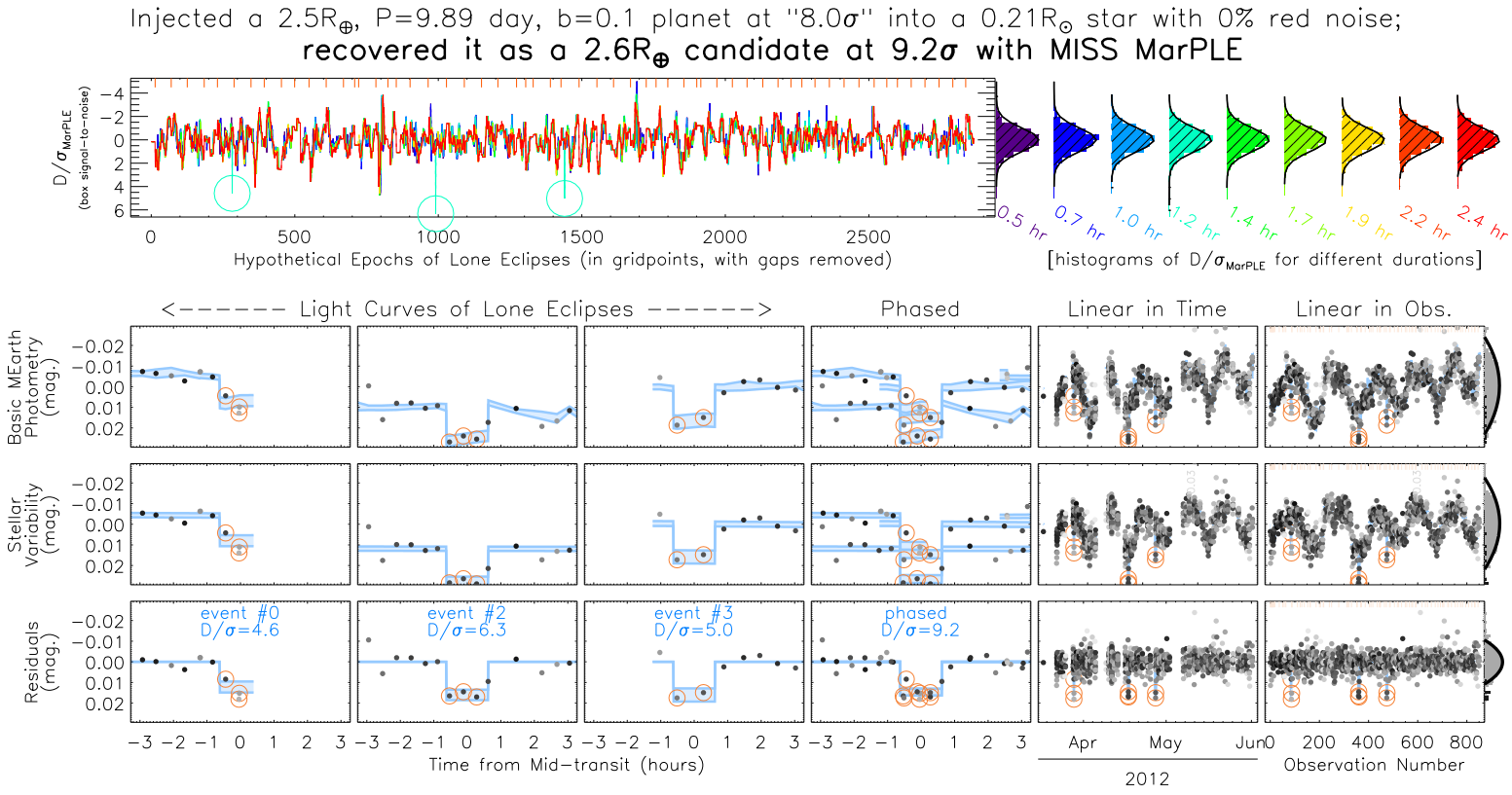

Figure 4.7: An end-to-end demonstration of MISS MarPLE applied to simulated transits injected into a real MEarth light curve. For this candidate, we show $D/\sigma_{\rm MarPLE}$, or the marginalized S/N, for all possible transit epochs and durations (*top*), both as an ordered timeseries (*left*) and as histograms at fixed duration (*right*). We also show MEarth photometry (*bottom*, filled circles, with grayscale proportional to $1/\sigma^2$) centered on the individual transit events (1^{st}-3^{rd} *columns*), phased to the injected planetary period (4^{th} *column*), linearly arranged in time (5^{th} *column*), and linearly arranged in observation number (6^{th} *column*, with nightly gaps denoted). Light curves are shown for basic MEarth photometry (1^{st} *row*), after subtracting the systematics model to show stellar variability (2^{nd} *row*), and after subtracting all aspects of the model except for planetary transits (3^{rd} *row*), along with samples from the probability distribution from our light curve model in each panel (blue swaths). Points in-transit are marked throughout this figure.

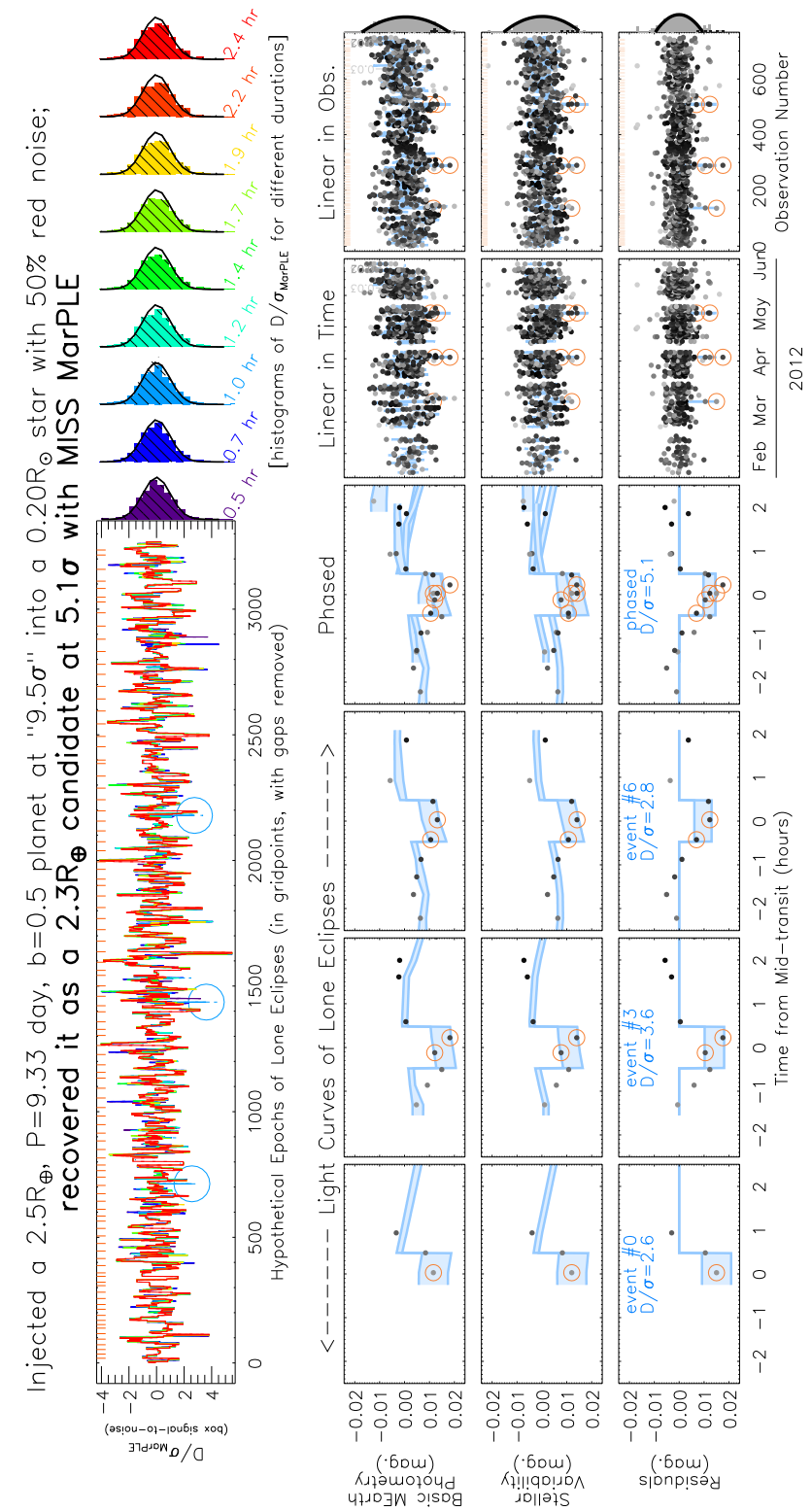

Figure 4.8: Another demonstration as in Figure 4.7, but for a more challenging star. In this case, a high residual red noise fraction and strong covariance between the systematics/variability model and the transit depth limit the recovery significance of this injected candidate. In a phased search, this candidate would not stand out as strong.

4.5.1 Injected Transits

To show how known transits appear through the lens of MISS MarPLE, we inject simulated transits into each of our raw light curves. Then we apply MISS MarPLE, and compare the significance of the recovered signals to those we injected. For the simulations, we inject 50,000 fake 2-4$R_\oplus$ planets into each MEarth target star, with periods from 0.5 to 20 days, random phases, and impact parameters between 0 and 1. The transits are limb-darkened, using quadratic coefficients for an M4 dwarf (Claret 2004).

We characterize each simulation by an "injected S/N": the injected transit depth $D_{\rm injected} = (R_p/R_\star)^2$ divided by $\sigma_{\rm injected}$. We calculate $\sigma_{\rm injected}$ by a $(\sum 1/\sigma^2)^{-1/2}$ estimator, using data points between 2^{nd} and 3^{rd} contact of the injected transit, with a global rescaling to match the RMS of the star's MAP-cleaned light curve. In the context of other transit detection algorithms that pair BLS with a pre-search cleaning step, this $D_{\rm injected}/\sigma_{\rm injected}$ has an important meaning. It would be the detection significance BLS would recover for a transit candidate if the pre-search data cleaning perfectly removed variability without influencing the depth of any transit events. Under the assumptions of this idealized BLS, the quantity $D_{\rm injected}/\sigma_{\rm injected}$ is directly linked (see Burke et al. 2006) to the "signal residue" detection statistic in the BLS paper (Kovács et al. 2002).

Individual Examples

We present a couple of illustrative simulations, to give a sense of how MISS MarPLE works. In each case, we use fake planets with three observed transits and periods near ten days. While such long periods would realistically offer this many transits only rarely,

we use these hand-picked examples as a convenient way to show both what individual transits of 10 day periods planets look like, and what phasing these transits into periodic candidates looks like. For simplicity's sake, we left MEarth's real-time trigger out of these simulations, showing what individual transits and phased candidates look like in low-cadence data. In reality, most of the injected transits above 3σ would have been detected by the real-time trigger, and their egresses' populated with additional high-cadence observations.

Figure 4.7 shows one example, a $2.5\mathrm{R}_\oplus$ radius planet with a $P = 9.89$ day period and $b = 0.1$ impact parameter injected into the raw MEarth light curve of a $0.21\mathrm{R}_\odot$ star. In this case, the 8.0σ injected S/N of the transit is well recovered by MISS MarPLE at 9.2σ, as is the inferred planet radius. Three transits fell during times of MEarth observations; they are marked in the plot of $D/\sigma_{\mathrm{MarPLE}}$, the eclipse S/N. This star exhibits 0% residual red noise and the transits all fall within well sampled nights; it is thanks to these favorable conditions that the injected and recovered S/N's are so similar.

For contrast, Figure 4.8 shows another example with a different star but broadly similar planetary parameters. Here, the recovered signal's 5.1σ significance is considerably lower than its injected 9.5σ strength. One reason for the difference is that the timescale of the intrinsic stellar variability of this star is short enough that the inferred transit depths are substantially correlated with it, thus making a larger contribution to σ_{MarPLE}. Additionally, our model does not completely remove all the structured features in this light curve so it exhibits a large red noise fraction ($r_{\sigma,r} = 0.5$), further suppressing the detection significance.

We also show in Figures 4.7 and 4.8 the photometry from MEarth, before and after

using the MAP values of our model parameters to subtract off systematics and stellar variability from the light curves. To emphasize that the result of MISS MarPLE is not simply one best-fit model of the systematics and variability, but rather an inferred probability distribution, we plot the swaths of light curve space that are spanned at $\pm 1\sigma$ by this distribution of models. We note that the probability distribution $P(\mathbb{M}|p_P, p_{E0}, p_T)$, is conditional on transit period, epoch and duration, so when we visualize the models with the light curves, we have fixed these parameters to their best values (as found in the grid search in Section 4.4.3). Because the transit search is entangled with the cleaning process, the models and appearance of the MAP-cleaned light curve would be different for different choices of p_P, p_{E0}, and p_T.

Relationship to BLS

By itself, a search with BLS will give the significance of a candidate transit that is conditional on the assumption that the out-of-transit baseline flux is constant and that its noise properties are globally known. If preceded by a light curve cleaning step, the transit significance is also conditional on the assumption that the aspects of the cleaning are correct. An important question is how much the marginalized significance of candidate transits found with MISS MarPLE differs from this conditional significance. Generally, the answer to this question will depend on the time sampling of the observations; for a very well-sampled and well-behaved light curve, the BLS and MarPLE results should converge to the same answer. But for the case of the real MEarth data, with its large gaps and fickle systematics, we approach this question with simulations.

Figure 4.9 shows the results of a head-to-head comparison of the significance with

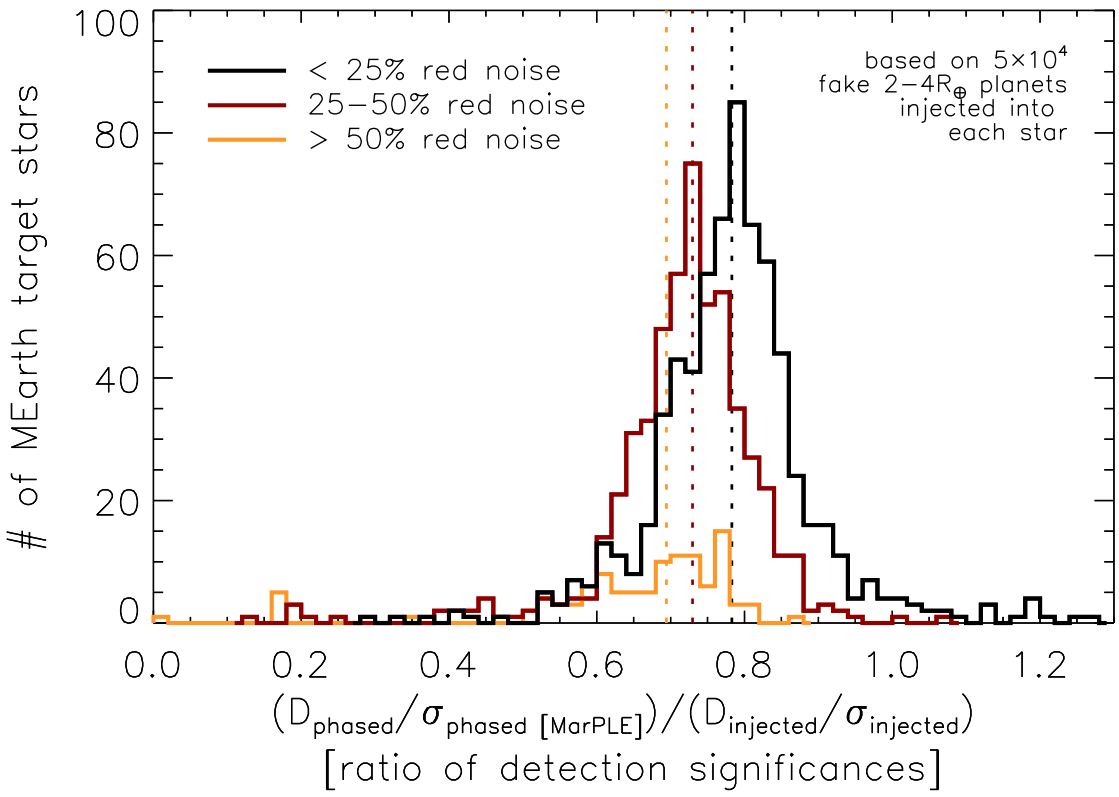

Figure 4.9: A comparison of the significance achieved in a phased search with MISS MarPLE ($\overline{D_{\rm phased}}/\sigma_{\rm phased}$) vs. an idealized BLS ($D_{\rm injected}/\sigma_{\rm injected}$). $\overline{D_{\rm phased}}$ and $\sigma_{\rm phased}$ represent the phase-folded combination of in-transit MarPLE's, as in Eq. 4.17 and 4.18. The definition of $D_{\rm injected}/\sigma_{\rm injected}$ is such that it represents a hypothetical in which any pre-BLS cleaning proceeded perfectly and without influencing the injected transit depth (see text). Each MEarth target star is represented once in this plot by the median of 4×10^4 simulations of planets with random periods, phases, impact parameters, and radii. The average significance ratio for each group of residual red noise factors $r_{\sigma,r}$ is shown (dashed lines); as most transits in these simulations contain only 1–2 points the impact of the red noise is relatively muted.

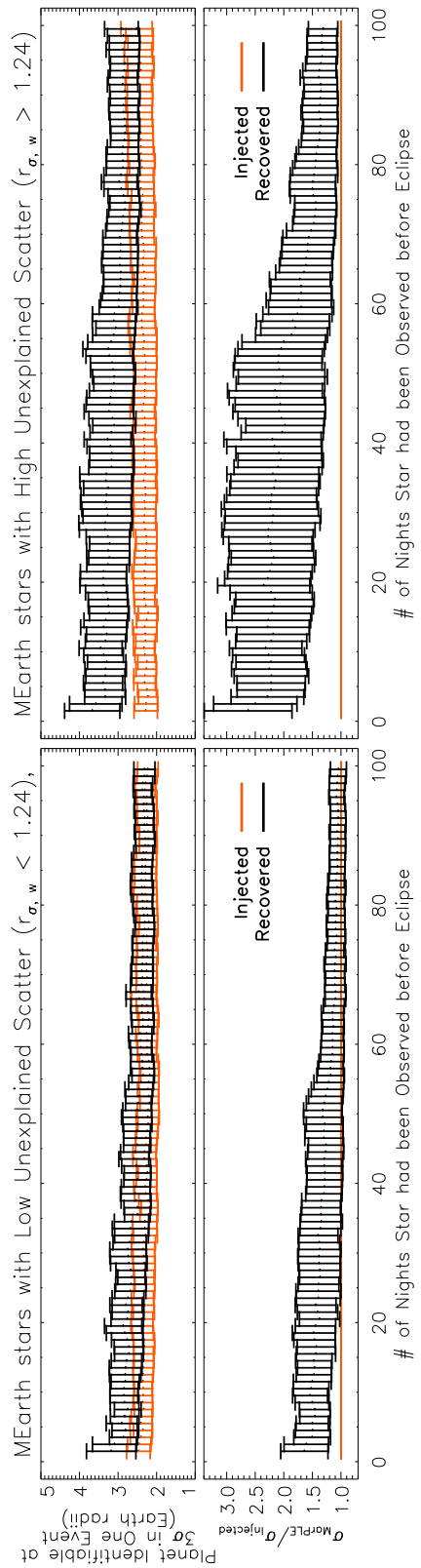

Figure 4.10: The results of "time-machine" simulations, in which we inject single transits into MEarth light curves and attempt to recover them, using only data *up to and including* the transit. As a function of how many nights the target was observed before the candidate transit (and thus how tight the priors can be), we show the smallest planet that could be identified at $> 3\sigma$ in a single event (*top*) and the ratio of the recovered (σ_{MarPLE}) to injected ($\sigma_{\mathrm{injected}}$) transit depth uncertainties (*bottom*). Each panel shows results from 50,000 injected transits in each of 100 random stars, with error bars representing the 25% and 75% quartiles of the distribution. We show the best (*left*) and worst (*right*) halves of the MEarth sample, based on how their average white noise rescaling parameter compares to the median of the sample ($r_{\sigma,w} = 1.24$). The lower envelope of each distribution typically corresponds to transits that fall in the middle of well-sampled nights; it converges as soon as tight priors can be established for the systematics coefficients. The upper envelope corresponds more to transits at the starts of nights or in poorly sampled nights; it converges more slowly, depending strongly on the priors for both the systematics and the variability coefficients.

which MISS MarPLE views phased (multiple-event) candidates with the significance that would go into a BLS calculation, based on ensemble of injected transits. A full period search was not run as part of these simulations; we calculated the detection statistics in both cases assuming the period was known. This is in line with our goal with MEarth, that we wish merely to identify whether a signal of a given significance is present, not accurately determine the period of that signal from the existing data.

For MEarth's best behaved stars (with $r_{\sigma,r} < 0.25$), the marginalized significance estimated by MISS MarPLE is typically 80% of that estimated by our idealized BLS. For these stars, properly accounting for all of the uncertainties in the cleaning process gets us to within 20% of the significance we could achieve in the unrealistic hypothetical that there were no uncertainties in the cleaning process. That MISS MarPLE tends to be more conservative than other methods is very important to our ultimate goal of using candidates identified by MISS MarPLE to invest limited period-finding follow-up observations. The 20% factor suppression of transit significance is comparable to the degree to which global filtering methods such as TFA suppress estimated transit depths (e.g. HATNet, see Bakos et al. 2013). However, the advantage of MISS MarPLE is more than simply knowing how much light curve cleaning suppresses transit significance *on average*; it is knowing what the cleaning's relative influence is *on individual events* and which events are more, or less, reliable. MISS MarPLE can give good events on good nights appropriately higher weight, unlike more global methods.

Figure 4.9 also shows that the penalty imposed by the red noise correction for those stars with $r_{\sigma,r} > 0.25$ is significant but not always debilitating. Because MEarth's cadence is so low that typically only 1–2 points fall within any given transit window, the influence of red noise on most transits is relatively small. However, in cases where the

cadence is much higher, such as a triggered event observed in real-time with MEarth, the red noise penalty could be much steeper. Also, as $D_{\rm injected}/\sigma_{\rm injected}$ is the best we could hope to achieve for each candidate, it is an important check that very few stars show significance ratios > 1.

Evolution of Priors

As more nights of observations are gathered, the priors on the systematics and variability parameters associated with a particular star will tighten. As these priors tighten, the significance with which a given transit can be detected will improve. We demonstrate this phenomenon graphically in Figure 4.10, which shows how $\sigma_{\rm MarPLE}$ for single events evolves as more observations are gathered as well as the impact of this evolution on the planet detection.

We injected transits as before but calculated the MarPLE for every individual event using only the data up to and including the event, excluding all data after 3^{rd} contact. These "time-machine" simulations are an approximation to the information available to the MEarth real-time trigger system when deciding whether to gather high-cadence followup of a candidate transit. We show the results for stars in the best and worst halves of the MEarth sample, as judged by how their white noise rescaling factors $r_{\sigma,w}$ compare to the median of the sample $\overline{r_{\sigma,w}} = 1.24$. Note that transits have a distribution of injected transit depth uncertainties ($\sigma_{\rm injected}$), based on the number of points in transit and the points' relative predicted uncertainties $\sigma_{\rm the}(t)$.

We highlight in Figure 4.10 the smallest planet that could be detected at 3σ confidence in a single low-cadence event, and how this quantity evolves a function of

the number of nights a star is observed before the event. Imagine a light curve contains 99 event-less nights and one event on the 100th night; Figure 4.10 indicates how much the information in the event-less nights improved the reliability of the single event's detection. In each panel, we show the 25 and 75% quartiles of the distribution (spanning both multiple stars and multiple random transits). For the stars with low $r_{\sigma,w}$, initially only planets larger than 2.5-3.8$R_\oplus$ exhibit deep enough transits to be detectable. But as more nights of observations tighten the priors, 2.0-2.6$R_\oplus$ planets become detectable, approaching the injected distribution. Stars with high $r_{\sigma,w}$ behave very differently, presumably because our model captures fewer of the features present in the light curves. For these stars, the minimum detectable planet sizes initially span 3.0-4.4$R_\oplus$ and never converge to the injected values.

We also show in Figure 4.10 the distribution of the ratio $\sigma_{\rm MarPLE}/\sigma_{\rm injected}$ for the simulated transits. The ratio starts off well in excess of unity, but approaches it as more prior-establishing observations are gathered. The range of values it spans corresponds to transits falling at more or less opportune moments. Values of $\sigma_{\rm MarPLE}/\sigma_{\rm injected}$ closer to 1 are usually associated with transits that fall in the middle of a well-behaved night. Higher values correspond to events that fall at the start of a night, events in a night with high excess scatter, or events that coincide with transit-like features in the systematics or variability models. By the end of a season, the distribution of $\sigma_{\rm MarPLE}/\sigma_{\rm injected}$ for single events in Figure 4.10 roughly approaches that for phased candidates in Figure 4.9. This makes sense, as the phased S/N ratios in Figure 4.9 use priors established from all the nights.

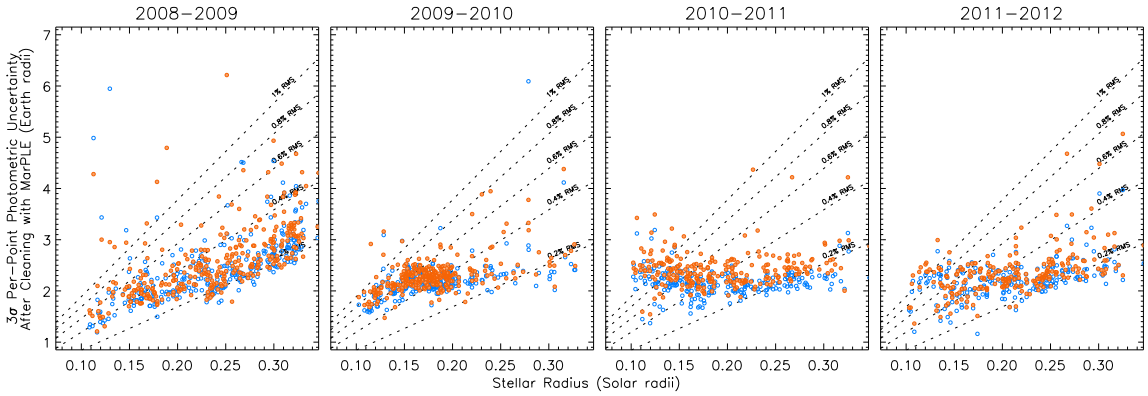

Figure 4.11: The per-point RMS photometric uncertainty as predicted from a CCD noise model (*open circles*) and that ultimately achieved in MarPLE-cleaned photometry, after subtracting off MAP models for systematics and stellar variability (*filled circles*). In each case, flares and in-transit points for each star's best candidate have been excluded from the calculation of the post-cleaning achieved RMS. Note that the improvement in the RMS relative to Figure 4.3 is achieved without blind suppression of planetary transits, as the MISS MarPLE cleaning occurs in tandem with the search for transits.

4.5.2 Application to MEarth Data

Finally, we employ MISS MarPLE to analyze all the MEarth target stars with no transits injected into them. Figure 4.11 gives one summary of the method's effectiveness. Here, we plot the achieved RMS in MEarth light curves after using MAP models of the systematics and variability to clean the light curves. Comparison to Figure 4.3 shows a dramatic improvement, moving the achieved RMS for all the stars much closer to their theoretical minima. However, the achieved RMS values still lie on a locus with a slight upward offset, indicating that our cleaning does not quite reach the photon noise limit. Indeed, this is a reflection of our finding that the median white noise rescaling parameter is $r_{\sigma,w} = 1.24$. Figure 4.11 also shows no evidence that we are over-fitting, in that we never achieve an RMS lower than predicted.

In Figure 4.12, we show the period and detection significance of the best phased candidate that we identify for each MEarth star using our MarPLE-based search. Here we have searched only one season of photometry at a time, so the same star may appear in multiple panels if we had multi-year observations of it. MEarth has published two systems with planet-sized eclipses: the planet GJ1214b (Charbonneau et al. 2009) and the brown-dwarf NLTT41135 (Irwin et al. 2010). The latter system is in a visual binary that was unresolved in the MEarth discovery data, so its eclipse depth was diluted to a planet-like 2% depth. While these systems were discovered by using an iterative median-filter (Aigrain & Irwin 2004) paired with traditional BLS (Kovács et al. 2002), we recalculate their detection significances using MISS MarPLE and indicate them in Figure 4.12. We also indicate the long-period low-mass eclipsing binary LSPM J1112+7626 (Irwin et al. 2011b), which was detected at very high significance (30σ) with the real-time detection trigger. Not shown is the short-period eclipsing binary GJ 3236 (Irwin et al. 2009a), as it was identified by eye in MEarth's commissioning data before the 2008-2009 season.

Several new candidates were initially identified above 8σ significance in Figure 4.12, but upon inspection the signals were found to be associated with bad raw images. The candidates evaporated after we removed these bad images from consideration. While we are actively investigating the most promising remaining candidates in Figure 4.12, none are as convincing as were our original confirmed systems in their discovery data.

The morphology of the plots in Figure 4.12 is roughly what we expect. Due to geometry, most of our stars will not host exoplanets that transit. Initially, one might think then that the cloud of candidates hovering around $5 - 6\sigma$ must mean we are substantially overestimating the significance for all of our stars. However, we must

consider what makes a reasonable detection threshold for a phased planet search. As discussed in detail by Jenkins et al. (2002), each phased search for planets constitutes an enormous number of effective hypotheses being tested against the data. Jenkins et al. (2002) found that a phased search of a Kepler light curve, with continuous cadence and a 4-year baseline, corresponded to an estimated number of equivalent independent tests (N_{EIT}) of $N_{\mathrm{EIT}} = 1.7 \times 10^7$. That is, the detection statistic expected from searching a transit-free Kepler light curve would be the same as asking for the maximum value achieved in 1.7×10^7 draws from a unit-variance Gaussian; the median null detection statistic should be above 5σ. It is this consideration that leads to the 7.1σ detection threshold for the nominal Kepler mission.

Although the relationship is complicated, generally N_{EIT} increases with the number of observations gathered, the number of periods, and the number of independent phases searched. Because 1-hour transits of M dwarfs are much shorter than the 10-hour transits typical for Kepler, we search many more phases for any given period. While the gap-filled, single-season MEarth light curves going into Figure 4.12 have very different properties than Kepler's, an estimate of N_{EIT} on the order of 10^7 is still a decent estimate. Indeed, using the Jenkins et al. (2002) bootstrap simulation method, we estimated for a MEarth light curve with 10^3 data points in which we searched 10^5 periods that $N_{\mathrm{EIT}} \approx 5 \times 10^6$. Null detection statistics above 5σ should be a regular occurrence in phased searches of MEarth targets. The position of MEarth's confirmed targets in Figure 4.12 suggests a $7 - 8\sigma$ threshold is probably appropriate for MEarth. Thresholds could safely be much lower for detecting single events, closer to 5σ, without the brutal combinatorics of a phased search.

Although it is too computationally intensive to calculate N_{EIT} for the different

observational coverages represented by all of the MEarth targets, we try graphically to demonstrate the effect of $N_{\rm EIT}$ in Figure 4.12. We fill the symbols with an intensity proportional to the number of independent eclipse epochs (p_E) that the light curve covers, using this as a very rough proxy for $N_{\rm EIT}$. This coloring scheme yields a vertical color gradient in all panels, reflecting the fact that targets with more observations have generally higher $N_{\rm EIT}$ and are more likely to generate high null detection statistics by chance.

Time-correlated noise can also disturb the frequency stability of the phased search, if correlations exist over timescales comparable to planetary periods being searched. For example, some uncorrected effect with a 1 day^{-1} frequency could build up over subsequent nights into what might look like a periodic planet signal. Our ad hoc correction in Section 4.1 does not account for that aspect of time correlated noise. It is likely that such extra uncorrected trends in the 2010-2011 season (which exhibited excess correlated noise, see Table 4.1) leads to excess of 1 day period candidates in Figure 4.12.

4.6 Future Directions

MISS MarPLE could be applied to other ground-based surveys for transiting exoplanets. Its advantages will be greatest for other pointed surveys like MEarth, where individual observations of individual stars are costly enough that it is worth the effort of optimally characterizing the information that each contributes. Aspects of MISS MarPLE be potentially useful to other surveys specifically targeting M dwarfs, such as PTF/M-dwarfs, APACHE, or RoPACS, where the variability and/or systematics are similar to those we described here.

Additionally, ground-based photometric followup to find transits of radial velocity planets (e.g. Kane et al. 2009) faces similar challenges. Typically looking for shallow transits in light curves of bright stars, such efforts require careful consideration of the systematic uncertainties associated with candidate events. For example, the RV-detected super-Earth HD97658b, initially announced to transit from ground-based photoelectric photometry at its predicted time and with 5.7σ confidence (Henry et al. 2011), was found not to transit in followup space-based photometry (Dragomir et al. 2012). This contradiction led to a reevaluation of the systematics in the ground-based observations, which were taken at high airmass. As the most exciting planet discoveries will often be those made very close to the detection threshold, it is important to accurately assess the uncertainties associated with the measured depths of putative transits. Some aspects of a method like the one we proposed here could be useful to marginalize over systematic uncertainties and thus give more confidence in the significance of transit detections in future followup efforts.

Many improvements could be made on our current implementation of MISS MarPLE. For one, the Gaussian likelihood we use to describe our data (Eq. 4.9) is an approximation. It is decent, but it could be elaborated by including a mixture of probability distributions for each data point (to account for junk outliers; e.g. Hogg et al. 2010; Sivia & Skilling 2006) or by directly modeling the correlations among data points (see, for example, Carter & Winn 2009). Also, the variability aspect of our generative light curve model is extremely simplistic (Eq. 4.4). By replacing our crude sinusoid + nightly offset model with a more sophisticated basis, one might be able to better capture all the variability features in real light curves, thus minimizing the uncertainty its correction injects into the marginalized probability of lone eclipses. In

particular, a variability model based on Gaussian processes (see Gibson et al. 2012, for an introduction) may be a promising route for setting dynamically evolving priors for the astrophysical behavior of a star on any given night.

In an upcoming paper, we intend to apply the MISS MarPLE framework to the task of estimating MEarth's sensitivity to 2-4$R_\oplus$ planets over the last four years. Given our single planet detection of GJ1214b, we will use this survey sensitivity estimate to place limits on the occurrence rate of short-period planets around nearby mid-to-late M dwarfs. Such limits would be complementary to results both from Kepler (Howard et al. 2012) and from the HARPS M dwarf radial velocity program (Bonfils et al. 2013).

Finally, our ultimate goal with MISS MarPLE is to identify promising candidates with MEarth and make follow-up observations to determine their periods. With this new well-tested method, we plan to focus our efforts in this direction in the years to come. Determining how to schedule the most useful observations for period-finding is a difficult task, but the "adaptive scheduling" algorithm proposed by Dzigan & Zucker (2011) may prove a very fruitful route.

4.7 Conclusions

In this work, we have proposed a new method for detecting planetary transits in wiggly, gap-filled light curves. A method such as this is necessary to eke the optimal sensitivity out of the MEarth Project, our survey for transiting 2-4$R_\oplus$ exoplanets around nearby mid-to-late M dwarfs. MEarth's unique observing strategy gives rise to new challenges (for example, Figure 4.1), thus inspiring our efforts to improve on existing transit

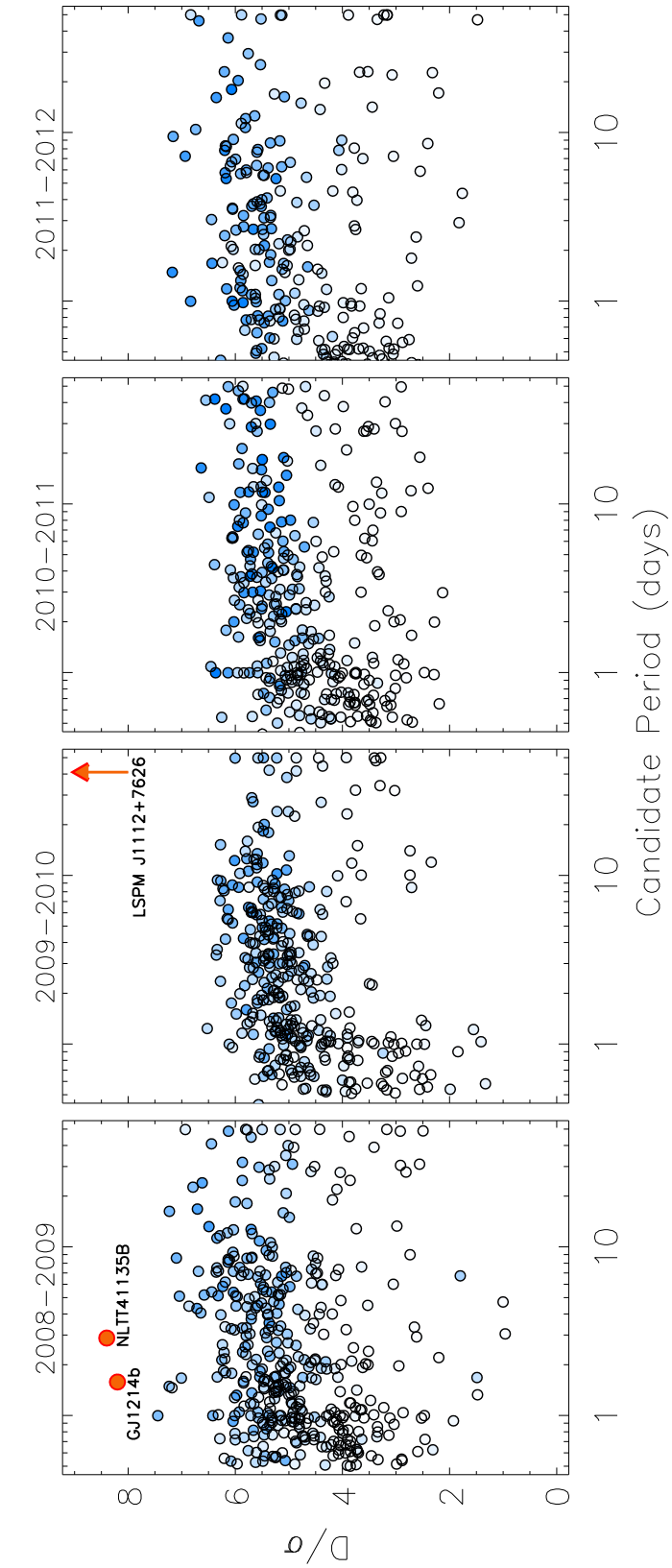

Figure 4.12: A summary of the application of MISS MarPLE to four individual seasons of MEarth data. Each point represents the best periodic candidate identified from a phased search of one star. As each phased transit search effectively performs many effective independent tests on the data, the position of the dense locus of candidates between 5 and 6σ is broadly consistent with the null hypothesis, of most of our stars not exhibiting planetary transits. For each star, the intensity of the symbol's color fill is proportional to the number of lone-eclipse epochs (p_E) for which observations exist. We also show the detection significance of published MEarth systems, based on the discovery data for each.

detection techniques.

One idea lies at the core of our new method: that when assessing the significance of any individual planetary transit, we want to marginalize over *all* the uncertainties, including those associated with cleaning systematics and intrinsic variability from the star's light curve. Our Method for Including Starspots and Systematics in the Marginalized Probability of a Lone Eclipse (MISS MarPLE) can investigate transits within the context of individual nights of observations (see Figure 4.4), sensibly accounting for various kinds of trends, occasionally messy observational cadences, and the vagaries of photometric conditions common to ground-based observatories. MISS MarPLE uses an analytic, semi-Bayesian approach to include information from an entire season of observations as priors to constrain the expected behavior of a star on any given night.

We applied MISS MarPLE to four seasons of MEarth photometry, showing that it improves our sensitivity to transiting exoplanets (Figures 4.3 and 4.11). By injecting simulated transiting planets into real MEarth light curves (Figures 4.7 and 4.8), we compare MISS MarPLE to the popular Box-fitting Least Squares (BLS) method (Kovács et al. 2002) and find that even for the best behaved MEarth targets, BLS underestimates the true transit depth uncertainties typically by 20% (Figure 4.9). That is, the covariance of hypothetical transit depths with systematics and variability corrections, on average, increases the true transit depth uncertainty by 20% for MEarth survey data. Simulations also show that 2-3$R_\oplus$ planets that are undetectable in the first few weeks a target is observed become detectable, either in archival data or in incoming data, later in the season as the behavior of the star is better constrained (Figure 4.10).

The "MarPLE," the probability distribution of hypothetical transit depths for any given transit duration and epoch, is a useful concept. Because this probability distribution is designed to be marginalized over all the complicated factors associated with the telescope or the night on which the observations were taken, it can be treated as a rigorous statistical summary for the presence or absence of a transit at any moment. Thus, we can straightforwardly combine these portable MarPLEs estimated from different telescopes using different filters at different observatories into coherent planet candidates. By properly accounting for so many transit detection uncertainties, the MarPLE should also save precious followup resources by not wasting time on too many false alarms. A framework such as MISS MarPLE could be a useful tool for any collaborative, global followup of long-period transiting exoplanet candidates that may be identified by MEarth or other observatories.

As the search for transiting planets around nearby stars pushes to radii smaller than $2R_\oplus$, properly accounting for systematics and variability will be become ever more important. MISS MarPLE may prove to be a valuable asset in the hunt for transiting exoplanets around bright M dwarfs in the years to come.

Acknowledgments

We thank Philip Nutzman for inspirational conversations regarding this work; John Johnson, Diana Dragomir, Scott Gaudi, and Elisabeth Newton for discussions regarding the method and the paper; and the referee whose careful reading improved the manuscript considerably. We gratefully acknowledge funding for the MEarth Project from the David and Lucile Packard Fellowship for Science and Engineering and from the National

Science Foundation (NSF) under grant number AST-0807690. The MEarth team is greatly indebted to the staff at the Fred Lawrence Whipple Observatory for their efforts in construction and maintenance of the facility and would like to thank Wayne Peters, Ted Groner, Karen Erdman-Myres, Grace Alegria, Rodger Harris, Bob Hutchins, Dave Martina, Dennis Jankovsky, Tom Welsh, Robert Hyne, Mike Calkins, Perry Berlind, and Gil Esquerdo for their support. This research has made use of NASA's Astrophysics Data System.

Chapter 5

Constraints on Planet Occurrence around Nearby Mid-to-Late M Dwarfs from the MEarth Project

This thesis chapter has recently been submitted as

Z. K. Berta, J. Irwin, D. Charbonneau, submitted to *The Astrophysical Journal* on 20 March 2013

Abstract

The MEarth Project is a ground-based photometric survey to find planets transiting the closest and smallest main-sequence stars. In its first four years, MEarth discovered one transiting exoplanet, the $2.7R_{\oplus}$planet GJ1214b. Here, we answer an outstanding question: in light of the bounty of small planets transiting small stars uncovered by the

Kepler mission, should MEarth have found more than just one planet so far? We estimate MEarth's ensemble sensitivity to exoplanets by performing end-to-end simulations of 1.25×10^6 observations of 988 nearby mid-to-late M dwarfs, gathered by MEarth between October 2008 and June 2012. For $2 - 4\mathrm{R}_\oplus$ planets, we compare this sensitivity to results from *Kepler* and find that MEarth should have found planets at a rate of $0.05 - 0.36$ planets/year in its first four years. When extrapolating between *Kepler*'s early M dwarfs and MEarth's mid-to-late M dwarfs, we find that assuming the planet occurrence distribution stays fixed with respect to planetary equilibrium temperature provides a good match to our detection of a planet with GJ1214b's observed properties. For larger planets, we find that the warm ($600 - 700$K), Neptune-sized ($4\mathrm{R}_\oplus$) exoplanets that transit early M dwarfs like Gl436 and GJ3470 occur at a rate of < 0.15/star (at 95% confidence) around MEarth's targets. We describe a strategy with which MEarth can increase its expected planet yield by 2.5× without new telescopes, by shifting its sensitivity toward the smaller and cooler exoplanets that *Kepler* has demonstrated to be abundant.

5.1 Introduction

An overarching goal of exoplanetary science is to understand the physical processes that shape exoplanets smaller than Neptune. Empirical measurements of the masses, radii, and atmospheres of such planets can serve as crucial inputs to developing this understanding. These measurements are most feasible for transiting planetary systems that exhibit favorable planet-to-star mass, radius, and temperature ratios (for amplifying signal strengths) and that are bright (for suppressing photon noise). Planets can satisfy

these criteria by transiting nearby small stars: the local M dwarfs. Observations from the *Kepler* mission indicate that early M dwarf stars host, on average, $0.90^{+0.04}_{-0.03}$ planets per star, for planetary radii between $0.5 - 4\mathrm{R}_\oplus$ and orbital periods shorter than 50 days (Dressing & Charbonneau 2013). This finding that small planets are so abundant around *Kepler*'s distant M dwarfs bodes well for our prospects of finding such planets transiting *nearby* M dwarfs, the close exemplars for which detailed characterization studies will be most rewarding.

The MEarth Project is an ongoing ground-based survey to identify planets transiting nearby mid-to-late M dwarfs (Nutzman & Charbonneau 2008; Berta et al. 2012b). With MEarth, we discovered the $2.7\mathrm{R}_\oplus$, $6.2\mathrm{M}_\oplus$ sub-Neptune GJ1214b (Charbonneau et al. 2009). GJ1214b transits a very small ($0.2\mathrm{R}_\odot$) M dwarf not far from the Sun (14.5 pc away; Anglada-Escudé et al. 2013), enabling characterization studies to probe the planet's atmosphere. Transmission spectroscopy from the ground (Bean et al. 2010a, 2011; Croll et al. 2011; Crossfield et al. 2011; de Mooij et al. 2012; Murgas et al. 2012; Narita et al. 2012; Teske et al. 2013) and from space (Désert et al. 2011a; Berta et al. 2012a; Fraine et al. 2013) is starting to provide boundary conditions for possible compositions of GJ1214b's gaseous outer envelope, constraints that may prove useful in our interpretation of the large population of sub-Neptunes unveiled both by Doppler surveys (Howard et al. 2010) and by *Kepler* (Batalha et al. 2013; Fressin et al. 2013)

Since the start of the MEarth survey in 2008, GJ1214b is the only planet we have found. If *Kepler* indicates planets are so common around M dwarfs, why has MEarth found only one? The shape of the occurrence distribution of the *Kepler* M dwarf planets plays strongly into the answer. Dressing & Charbonneau (2013) found that while small planets are common around M dwarfs, a planetary radius of $2\mathrm{R}_\oplus$ marks the start of a

precipitous decline in the rate of planet occurrence toward larger planets, particularly at short orbital periods. Yet MEarth was specifically designed (Nutzman & Charbonneau 2008) to be sensitive to planets larger than $2R_\oplus$. This mismatch between MEarth's achieved sensitivity and *Kepler*'s implied planet occurrence qualitatively explains MEarth's lack of further detections. In this work, we simulate MEarth's ensemble sensitivity during its first four years, in order to make this statement more quantitative.

Our aim is to provide perspective on the challenges and opportunities for finding planets transiting nearby, very small stars. Due to observational biases, *Kepler*'s M dwarf planet hosts are typically very distant ($100 - 300$ pc) and are dominated by earlier, larger ($0.3 - 0.6R_\odot$) spectral types. The *Kepler*-42 trio of sub-Earths transiting a $0.17R_\odot$ twin of Barnard's star 40 pc away is a notable but rather rare exception (Muirhead et al. 2012b). By and large, *Kepler* will not find the Earths, super-Earths, and sub-Neptunes that transit very nearby, very small, mid-to-late M dwarfs, those planets whose atmospheres could be studied in the near future. Finding these planets requires an all-sky approach, and *Kepler*'s 100 square degree field of view is only 0.3% of the entire sky. Focused searches will be required to find these planets, either targeted photometric surveys like MEarth or APACHE (Giacobbe et al. 2012), or radial velocity surveys followed by transit monitoring (Gillon et al. 2007b; Bonfils et al. 2012). An all-sky space telescope such as the proposed Transiting Exoplanet Survey Satellite (TESS) would make significant progress toward finding planets whose atmospheres can be studied (Deming et al. 2009), but is not yet available. The lessons we learn from MEarth and *Kepler* can guide these efforts in the years to come.

In addition to finding individual objects to use as observational laboratories, understanding the statistical properties of the planet population around M dwarfs is

important in its own right. Main sequence hydrogen-burning stars with spectral types later than M0 (roughly $0.6M_{\odot}$) outnumber G dwarfs like the Sun a dozen to one in the solar neighborhood (Henry et al. 2006)[1]. Estimates from SDSS indicate that the Galaxy's stellar mass function peaks at $0.25M_{\odot}$ (Covey et al. 2008; Bochanski et al. 2010), corresponding roughly to M4 in spectral type. The *Kepler* M dwarfs probe planet statistics down to this peak, but not much below it. In the SDSS mass function, 40% of stars fall below $0.25M_{\odot}$. If we want a complete census of exoplanets in the Milky Way, we need to understand how planets populate these smallest of stars.

Observations of how planet populations change with host star mass inform theories of planet formation and evolution. For example, Laughlin et al. (2004) argued that, in the core accretion paradigm, M dwarfs' less massive protoplanetary disks (see Andrews & Williams 2005) would inefficiently form Jupiter-mass planets. This hypothesis was later verified observationally as a deficit of Jupiters detected in radial velocity surveys (Johnson et al. 2007). By spanning a wide range of stellar masses, we can hope to provide a long lever arm to probe how the birth, growth, and survival of planets depends on various environmental factors (see Raymond et al. 2007; Barnes et al. 2008; Montgomery & Laughlin 2009; Ogihara & Ida 2009).

In this work, we use MEarth's single planet detection and lack of additional detections to explore the statistics of planets at the bottom of the main sequence. The paper is organized as follows. In §5.2, we describe four years of observations gathered by the MEarth Project. In §5.3, we outline our numerical simulations of MEarth's sensitivity during this time. In §5.4, we discuss the results and compare them directly to

[1] `Seealsohttp://www.recons.org/census.posted.htm`

the M dwarf planet statistics from *Kepler*. In §5.5, we demonstrate how to harness these findings to improve MEarth's planet yield in the years to come. We conclude in §5.6.

5.2 Observations

MEarth employs eight robotic 40-cm telescopes at the Fred Lawrence Whipple Observatory atop Mt. Hopkins in Arizona. The cameras are sensitive to 715–1000nm wavelengths, with the bandpass shaped by a fixed Schott RG715 filter[2] and the quantum efficiency of our back-illuminated e2v CCD42-40 detectors. The light curves analyzed in this paper were gathered between 2 October 2008 and 23 June 2012. Except for the annual monsoon season typically spanning July to September and occasional instrumental failures, MEarth observed every clear night during these four years. Nutzman & Charbonneau (2008) outline MEarth's design strategy, and descriptions of the light curve processing and properties can be found in Irwin et al. (2011a), Berta et al. (2012b), and on the MEarth website[3].

MEarth differs from wide-field transit searches. Targeting the brightest mid-to-late M dwarfs spread across the sky, MEarth observes each star one-by-one in a pointed fashion. As such, MEarth can tailor exposure times to individual targets; we observe each target long enough in each visit that its expected RMS from photon and scintillation noise is 1/3 the transit depth of a $2R_\oplus$ planet. If necessary (to avoid saturation), we gather multiple exposures within each visit to meet this goal. For the estimated

[2]A more narrow filter with an interference cutoff (715-895m) was used in the 2010–2011 season; see Berta et al. (2012b).

[3]`http://www.cfa.harvard.edu/MEarth/`

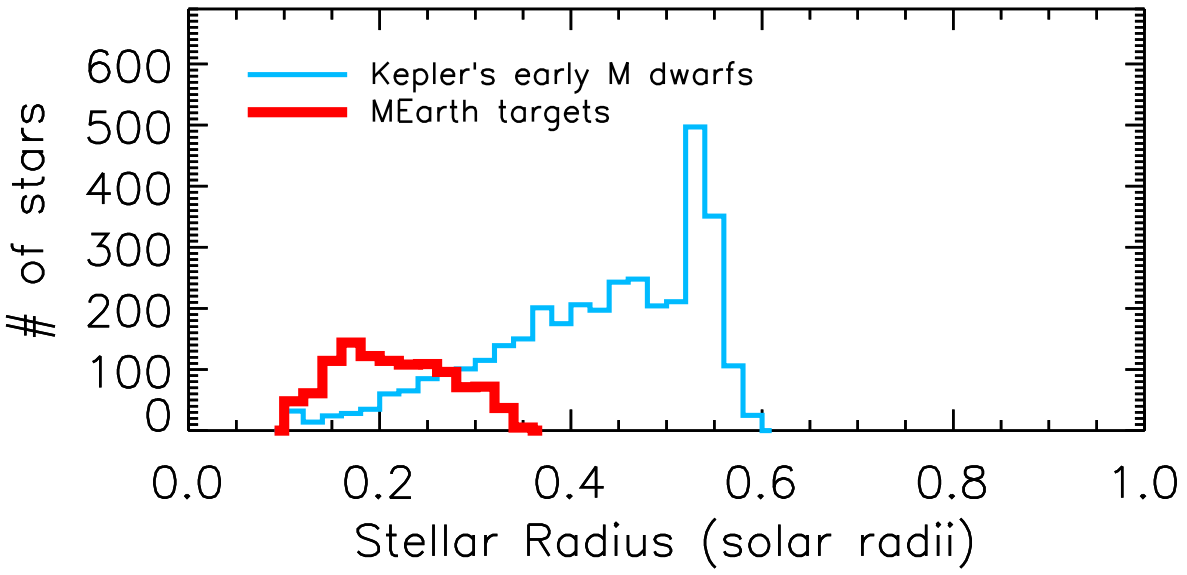

Figure 5.1: Comparison of MEarth targets (*red*) with the *Kepler* early M dwarf sample (*blue*). Stellar radius estimates are taken from Nutzman & Charbonneau (2008) for MEarth and from Dressing & Charbonneau (2013) for the *Kepler* M dwarfs. Only MEarth targets analyzed in this work are shown, not the entire Nutzman & Charbonneau (2008) sample.

$0.1 - 0.35 R_\odot$ radii of MEarth targets (Nutzman & Charbonneau 2008), these correspond to photometric precisions of $1.0 - 0.2\%$ per visit. Thanks to a comparatively small pixel scale (0.76"/pixel) and the large proper motion on which our targets were selected ($\mu > 0.15$"/year; from Lépine & Shara 2005), blended background eclipsing binaries are not a significant source of false positives (see Latham et al. 2009; Triaud et al. 2013a, for contrast).

MEarth's cadence is also unique. Transits of these stars have typical durations of $0.5 - 2$ hours. MEarth observes each star roughly once every 20–30 minutes, reduces the data in realtime, and can trigger immediate high-cadence monitoring of interesting in-progress transit events. By gathering many observations of a transit's egress with this

realtime trigger mode, MEarth can confirm or deny the presence of a planet from a single transit. As we discuss in §5.5, this mode can substantially improve MEarth's sensitivity to long period planets. The realtime trigger was operational for most of $2009 - 2012$, but was under continual development to ramp up its sensitivity during this time.

Figure 5.1 shows the estimated radii of the MEarth stars analyzed in this paper. For comparison, the *Kepler* early M dwarf sample analyzed by Dressing & Charbonneau (2013) is also included. The *Kepler* M dwarfs are strongly biased toward earlier spectral types, near the K/M boundary. The radius estimates for the MEarth sample are drawn from Nutzman & Charbonneau (2008). They are highly uncertain, but we address this issue in our analysis in §5.3.4.

All M dwarf light curves from MEarth are made publicly available for the community to use. Data are released on a yearly schedule, and can be found online at `http://www.cfa.harvard.edu/MEarth/`. The online release contains extensive notes on the data processing methods, and on the properties of the survey as a whole. We eagerly encourage those interested in the photometric variability of nearby mid-to-late M dwarfs, or the possible presence of planets around them, to make broad use of MEarth's public light curves.

5.2.1 Data Quality Cuts

For the analyses presented in this work, only M dwarf light curves consisting of more than 100 good visits are included. For the four seasons since 2008, there were $[349, 306, 256, 240]$ such light curves. We excluded from the statistical analysis all light curves where genuine astrophysical eclipse signals were detected during the course of

the survey, as they are dominated by high cadence follow-up observations and do not reflect our blind survey sensitivity. These objects are GJ1214 (Charbonneau et al. 2009), GJ 3236 (Irwin et al. 2009a), NLTT 41135 (Irwin et al. 2010), and LSPM J1112+7626 (Irwin et al. 2011b).

Most stars were observed for only a single season, but not all. Some M dwarfs (140 in total) were observed in two, three, or four seasons. One season typically dominates the sensitivity to planets in these cases; for simplicity, we include only the most sensitive season-long light curve for each star in the statistical analysis, thus slightly underestimating our true sensitivity. In total, 988 unique M dwarfs are represented here.

We make strict data quality cuts. We flag and remove any exposures in which thick clouds were present (worse than 0.25 magnitudes of losses), the target's position was offset from its median on the detector by more than 5σ, images show extremely large FWHM (3σ above its median) or high ellipticity (> 0.25), $> 10\%$ of comparison stars show 4σ outliers in their light curves, or 50% of the other exposures within a 4 hour window were excluded for any other reason. Some exposures, originally identified as contributing to candidate eclipse signals, were found to contain anomalies (inopportune glints or satellite tracks, shutter blade failures, other defects) and were also excluded. A total of 1.25×10^6 exposures distributed over 8.26×10^5 independent visits survive these cuts and contribute to the analysis.

5.3 Method

We aim to interpret MEarth's one planet detection and many non-detections in the light of the statistical rate of occurrence around these small stars. Doing so requires an estimate of the ensemble sensitivity of the survey to hypothetical planets during the past four seasons.

5.3.1 MEarth's Ensemble Sensitivity

We define MEarth's ensemble sensitivity, $S(R, P)$, as the number of planets with a particular radius R and orbital period P that MEarth should have found if all stars in our sample had one planet with that radius and period. We calculate the ensemble sensitivity as

$$S(R,P) = \sum_{i=1}^{N_\star} \eta_{\mathrm{tra},i}(R,P) \times \eta_{\mathrm{det},i}(R,P) \tag{5.1}$$

where the functions $\eta_{\mathrm{tra},i}(R,P)$ and $\eta_{\mathrm{det},i}(R,P)$ are the per-star transit probability and the per-star detection efficiency, respectively. Our targets span a factor of 3 in stellar radius, and each star has its own observational coverage and light curve properties. As such, we calculate each of these functions individually on a star-by-star basis. In addition to $S(R,P)$, we also calculate our sensitivity as a function of the equilibrium temperature T a planet would have assuming zero albedo and efficient heat redistribution[4]. We refer to this sensitivity as $S(R,T)$ and will use the two "period-like" parameters P and T somewhat interchangeably hereafter.

[4]We quote temperatures for a Bond albedo of $A = 0$ simply for ease of scaling. The relation that $T \propto (1-A)^{1/4}$ can be used to translate to a chosen planetary albedo.

5.3.2 Transit Probabilities = $\eta_{\mathrm{tra},i}(R,P)$

The first factor, $\eta_{\mathrm{tra},i}(R,P)$, is the probability that a planet is geometrically aligned such that transits occur in the system. We take $\eta_{\mathrm{tra}} = R_\star/a$, where $R_\star$ is the stellar radius and a is the semi-major axis; this is the probability that a planet in a circular orbit would have an impact parameter $b < 1$. From Kepler's third law and assuming the planet's mass is negligible, $R_\star/a = [3\pi/(GP^2\rho_\star)]^{1/3}$ where G is Newton's constant. We estimate the mean stellar density $\rho_\star$ from archival observations of each target (see Nutzman & Charbonneau 2008). We discuss the impact of uncertainties in $\rho_\star$ on our $S(R,P)$ in §5.3.4.

Non-zero eccentricity can increase (or decrease) transit probabilities for planets with Earth-ward periastrons (or apoastrons), but the net effect for blind transit surveys is small on average. For a reasonable underlying distribution of eccentricities, assuming circular orbits underestimates the total sensitivity of a photon-limited survey by only 4% and of a completely red-noise dominated survey by 25% (Burke 2008). With our sparse sampling, our ability to correct for known systematic effects, and the short $1-2$ hr transit durations we are targeting, MEarth's sensitivity is influenced but not dominated by red noise (Berta et al. 2012b). As such, we expect the eccentricity bias to be less than about 10% and ignore it.

5.3.3 Transit Detection Efficiencies = $\eta_{\mathrm{det},i}(R,P)$

The second factor, $\eta_{\mathrm{det},i}(R,P)$, is the probability that a planet aligned to transit would exhibit one or more such transits while MEarth was observing and that the signal would be strong enough cross an appropriate detection threshold. For idealized surveys, $\eta_{\mathrm{det},i}$

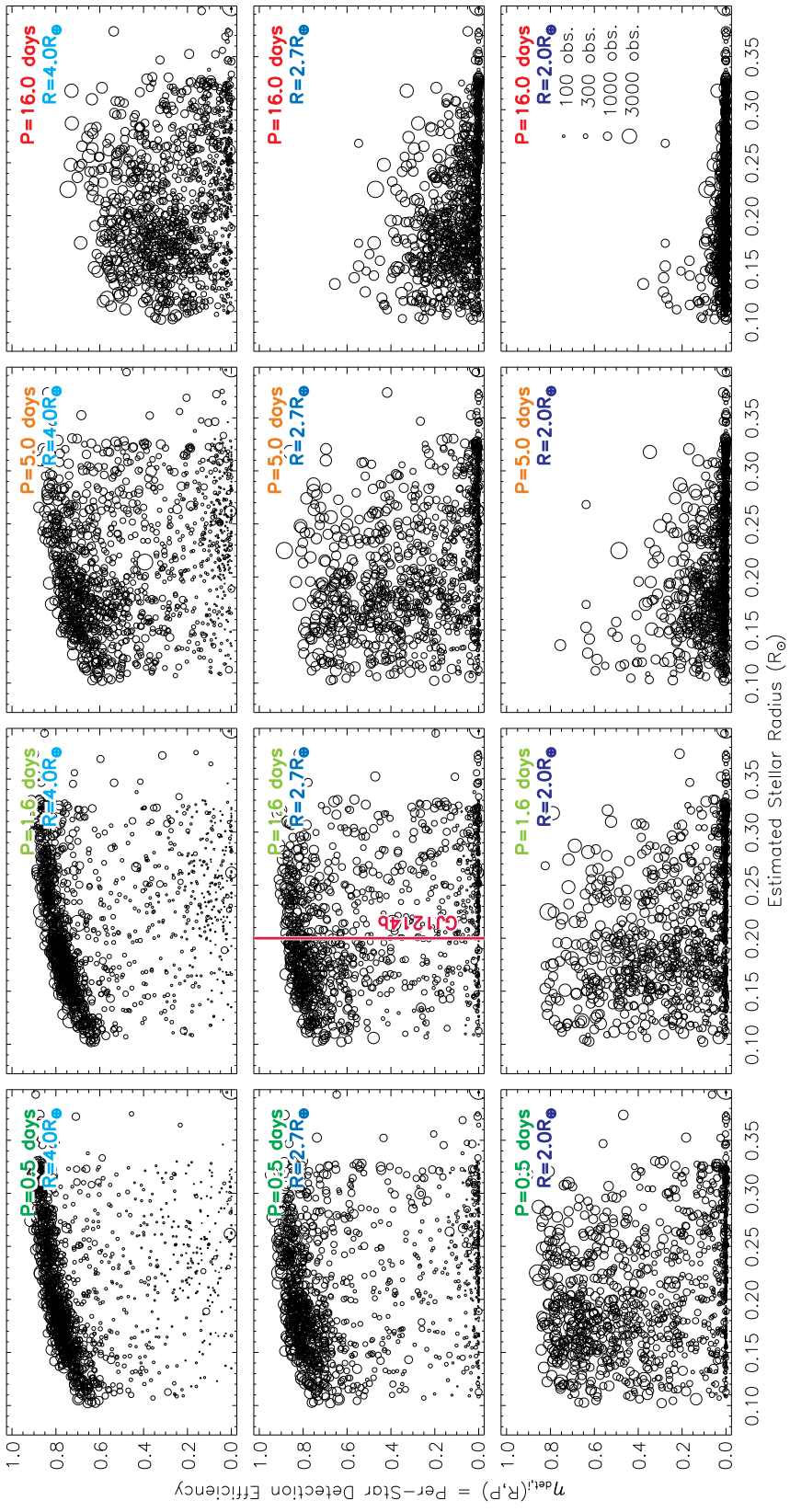

Figure 5.2: Results from the simulation described in §5.3.3, showing the per-star discovery efficiency $\eta_{\mathrm{det,i}}$ in a phase-folded search, for planets of decreasing radii (*rows*, [4.0, 2.7, 2.0]$\mathrm{R}_\oplus$ top to bottom) and increasing periods (*columns*, [0.5, 1.6, 5.0, 16] days left to right). Each star is represented by a single circle appearing in each of the 12 panels; the area of each circle is proportional to the number of observations MEarth gathered of that star. Transits become more difficult to detect for smaller planets and longer periods.

can be calculated analytically (Pepper et al. 2003; Gaudi et al. 2005), given a typical photometric precision and number of observations. Kepler's detection efficiency appears to match these predictions quite well (Youdin 2011; Howard et al. 2012). Without the benefit of *Kepler*'s uniform cadence, coverage, and noise properties, we turn to simulations to estimate $\eta_{\mathrm{det},i}$ for MEarth.

Into each MEarth light curve, we inject simulated, limb-darkened transit signals (Mandel & Agol 2002). The fake planets are generated on a grid of planetary radii spanning 1.5–4$R_\oplus$, and have random periods between 0.25 and 20 days, random epochs, and random impact parameters between $b = 0$ and $b = 1$ (all drawn from uniform distributions). In Berta et al. (2012b), we develop a framework for assessing the significance of transit signals in MEarth data. This framework, the Method to Include Starspots and Systematics in the Marginalized Probability of a Lone Eclipse (MISS MarPLE), is designed to provide robust detection statistics for transit signals amidst systematic noise sources arising from instrumental/telluric effects or from stellar variability. We inject the signals into our basic light curves[5], process those light curves with MISS MarPLE, and estimate the signal-to-noise ratio at which the injected signal would be recovered. We iterate this process 50,000 times for each star, to map out the planetary parameter space.

We estimate the transit significance at the known period and epoch of the injected signal; we do not perform a phased period search for each injected signal. Because we

[5]By "basic light curves", we mean MEarth light curves after the differential photometry correction has been applied. This differential photometry correction is estimated solely from comparison stars, not the target M dwarf itself, so any suppression of transits at this early stage of the processing should be minimal.

skip the computationally costly period search, these simulations know nothing about whether or not the existing MEarth light curve would be enough to recover the correct period for the planet. This is okay; we are interested merely in whether or not a strong enough signal would exist in the data that we would track it down with targeted follow-up observations.

A crucial component is the decision of a detection significance threshold. MISS MarPLE both corrects for stellar variability and instrumental systematics and keeps track of the excess uncertainty these corrections introduce to the marginalized transit depth uncertainties ($\sigma_{\rm MarPLE}$) of candidate transit signals. In that work, we found that all period-phased MEarth candidates above a transit depth signal-to-noise of $D/\sigma_{\rm MarPLE} > 7.5$ were either catastrophic instrumental problems, obvious stellar variability artifacts, or bonafide eclipsing systems (Charbonneau et al. 2009; Irwin et al. 2009a, 2010, 2011b). GJ1214b was identified in 2009 before the development of the full MISS MarPLE framework, using the Box-fitting Least Squares search algorithm (Kovács et al. 2002) and pre-cleaning with an iterative median filter (Aigrain & Irwin 2004) and Trend Filtering Algorithm (Kovács et al. 2005). Reanalyzing the data available up to the time of GJ1214b's original identification, we find that MISS MarPLE would have recovered the signal with $D/\sigma_{\rm MarPLE} = 8.2$.

We take $D/\sigma_{\rm MarPLE} > 7.5$ as our detection threshold for the following analysis. We emphasize that this is a *phased* transit significance, meaning that multiple transits can folded together to build up $D/\sigma = 7.5$. For Neptune-sized planets, a single transit observed at 25 minute cadence can often cross this threshold by itself. We impose no lower limit on the number of transits required for a detection, so such planets would be considered as recovered. For the main results in §5.4, we do not explicitly model

detecting planets from single transits with MEarth's realtime trigger mode. Because the realtime trigger was continually being improved throughout the survey, it is difficult to quantify the extent to which it boosted our sensitivity. In §5.4, we choose to include data that were gathered as part of high-cadence follow-up, but we do not simulate the generation of new triggers. As such, MEarth's actual sensitivity may have been higher than we quote. We explore the effect that the realtime trigger can have on our sensitivity in §5.5.

Note that in this context, "detected" at 7.5σ means that a candidate would have high enough significance that, after a thorough visual vetting, we would begin follow-up observations to confirm the transits and to measure the system's period. Using a detection threshold lower than 7.5σ would make MEarth appear more sensitive, but drawing statistical inferences would require that we prove the candidates above that threshold were either real or statistical flukes. We are in the process of vetting the most promising $6.5 - 7.5\sigma$ candidates, but have not yet determined whether they are bona fide planets or not. For simplicity, we simply adopt the more conservative 7.5σ threshold. We explored changing the assumed detection threshold down to 7σ and up to 8σ. We found that the quantitative results in §5.4 and §5.5 do change, but the qualitative conclusions do not.

Figure 5.2 visualizes the results of our simulations of the phased-search per-star detection efficiency $\eta_{\rm det,i}$. Each panel corresponds to a different slice of planetary radius and period, and contains one circle for every star in the sample. As expected, MEarth's ability to discover planets in a phased search falls off toward longer periods and smaller radii. The symbol area scales linearly with the number of observations MEarth gathered of that star, demonstrating how $\eta_{\rm det,i}$ grows with the number of

measurements gathered, and saturates for large radii and short periods. The simulations have effectively integrated over impact parameters $0 < b < 1$. Because $\eta_{\rm det,i}$ accounts for grazing eclipses being more difficult to detect, it never reaches unity. The downturn in the upper envelope of $\eta_{\rm det,i}$ toward the smaller stars for short-period, Neptune-sized planets is caused by the breakdown of MISS MarPLE's assumption that transits will be box-shaped. The mismatch in shape would be treated as excess nightly noise in the MarPLE framework, and these transits' significance down-weighted. A more flexible trapezoidal assumption for the light curve shape (Aigrain et al. 2007) would be able to improve sensitivity in this corner of parameter space.

5.3.4 Uncertainties in Stellar Parameters and Binarity

MEarth's actual sensitivity depends on how well we understand the physical masses and radii of the stars in our sample. For this work, we adopt the stellar parameters as derived in Nutzman & Charbonneau (2008). These parameter estimates are not necessarily the best estimates currently possible, and introduce uncertainties into our statistical analysis. In this section, we describe those uncertainties and quantify their overall effects.

First, the masses and radii are imprecise. Although some stellar parameter estimates were based on literature geometric or spectroscopic parallaxes, many are based on photographic plate measurements with substantial uncertainties. High proper motions ensure the sample is free of giants, but the roughly 30% uncertainties in stellar radii directly translate into comparably large uncertainties in the planetary radius to which a given transit depth corresponds.

Second, some stars may be contaminated with unresolved binaries. We made small

efforts to remove some known binaries from our sample, such as those that Lépine (2005) identified as visual binaries resolved on input photographic plates or those indicated as spectroscopic binaries in the literature (e.g. Shkolnik et al. 2010). However, these efforts took place concurrently while we were already gathering data and are still far from complete. MEarth definitely observed some contaminating binaries as part of the sample analyzed here. The additional flux from a binary companion dilutes transit depths, as well as perturbing the stellar parameters that we infer from (spectro)photometry.

These two effects have impacts on both $\eta_{\rm tra,i}$ and $\eta_{\rm det,i}$. To prevent them from biasing our ensemble sensitivity estimate, we simulate their influence in a Monte Carlo fashion and apply a correction to our final estimate of $S(R, P)$. In these simulations, we perturb the assumed mass of each star by a 30% Gaussian uncertainty combined with a prior so that the star is drawn from the local stellar mass function (Bochanski et al. 2010). Including the stellar mass function tends to increase the masses of those stars that were originally inferred to be the least massive, those below the mass function peak. We use the mass-radius-luminosity relations of Boyajian et al. (2012b) to generate new radii and luminosities from these masses. We randomly assign binary companions to each star, assuming a 34% multiplicity rate (Fischer & Marcy 1992; Janson et al. 2012) and uniform mass-ratio distribution (treating the original star as the primary). We shift the values of $\eta_{\rm tra,i}$ and $\eta_{\rm det,i}$ according to the simulated system's updated properties. We apply this process to all the stars included in the sample, calculate the updated $S(R, P)$ and $S(R, T)$, and repeat 20 times for the ensemble. We take the averages of these distributions as our best-guess sensitivity estimates.

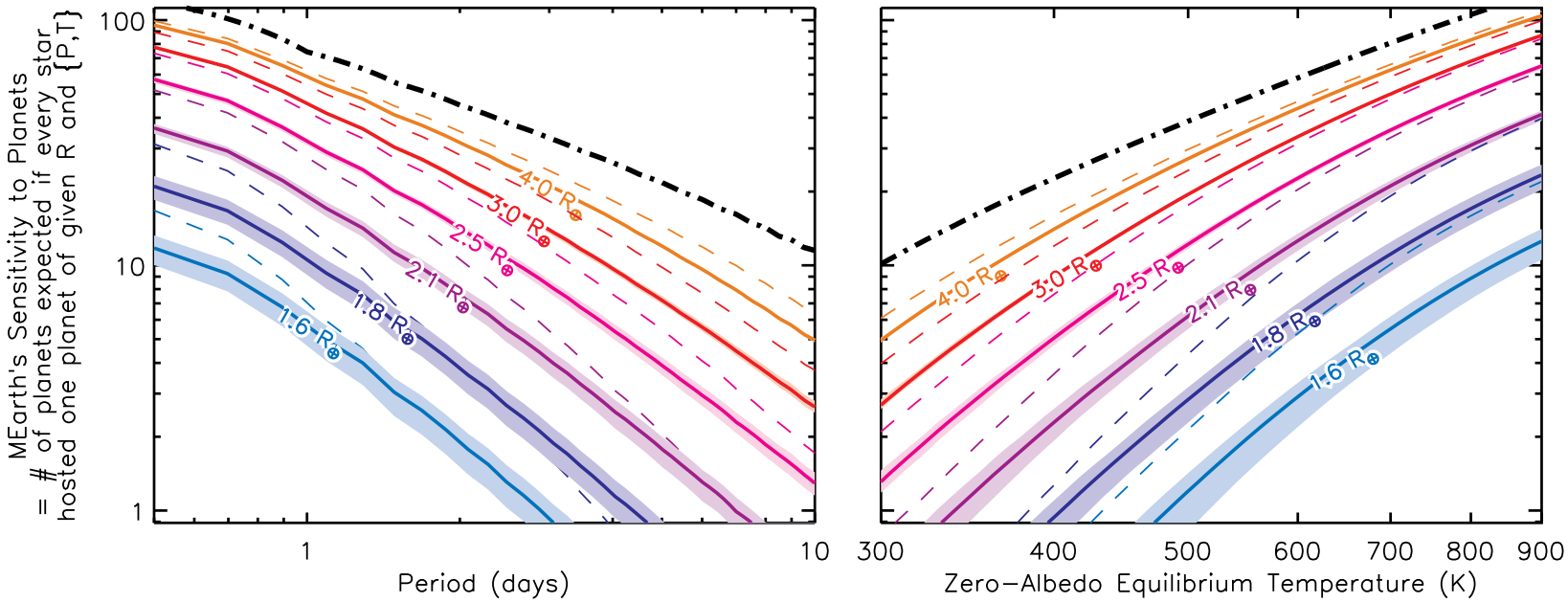

Figure 5.3: MEarth's ensemble sensitivity, as a function of planet radius (*colors*), orbital period (*left panel*), and bolometric incident flux (*right panel*), as given by Eq. 5.1. These sensitivity curves include both the geometrical probability that a planet would transit and the probability that we would detect transits that do occur. We show MEarth's sensitivity after accounting for uncertainties in the inferred stellar parameters and individual targets' unknown binarity (*solid lines*, with the range of values spanned in our perturbation analysis represented by the shaded swaths). For comparison, we show the naive estimate of what the sensitivity would be if we assumed all our targets were single and had perfectly known stellar parameters (*dashed lines*). The window function (*black dash-dotted line*) shows how sensitive MEarth would be if all that was required to detect a planet was a single in-transit point.

5.4 Results

The final sensitivity estimates are shown in Figure 5.3, both as a function of orbital period $[S(R, P)]$ and as a function of planet's zero-albedo equilibrium temperature $[S(R, T)]$. Because the MEarth stars span a range of effective temperatures, the mapping between these two parameter spaces is not necessarily one-to-one.

The detectability of planets near or below $2R_\oplus$ depends strongly on the inferred stellar radius of each star, so the above perturbation analysis introduces substantial uncertainties into the sensitivity for these smaller planets. In contrast, $4R_\oplus$ planets are often far above the detection threshold, so sensitivity depends weakly on the exact stellar radius. In most cases, the post-perturbation sensitivity decreased relative to the naive estimates. However, for planets smaller than $2R_\oplus$, the sensitivity increased relative to the naive estimate, thanks to some stars stars having radii smaller than originally assumed, so that MEarth's exposure times designed to detect $2R_\oplus$ planets could actually pick up even smaller planets.

Throughout this section, we test a variety of hypotheses against these sensitivity curves. Sometimes, we consider planets of particular periods P, other times planets with particular equilibrium temperatures T. We explore both parameters because we do not know the answer to the following question: When comparing samples of planets orbiting stars of different masses, what parameters define a planet population? Are two planets orbiting different stars comparable if they share the same radius? Or should they have the same mass, the same planet-to-star mass ratio, the same ratio to the isolation mass (e.g Ida & Lin 2008), or some other shared trait? Are two planets similar if they have the same period, the same semimajor axis, the same $a/R_\star$, the same proximity to the snow

line, or something else? Theory points to the importance of various of these parameters in formation, migration, and evolution scenarios (for a small selection of examples, see Fortney et al. 2007; Kennedy & Kenyon 2008; Hansen & Murray 2012; Kretke & Lin 2012; Chiang & Laughlin 2013; Swift et al. 2013), but the likelihood seems small that any simple answer for this big question could emerge. In the context of this uncertainty, we explore two representative examples. We consider period P as a population-defining parameter simply because it is closely tied to what we can detect; it may not be the most fundamental physical parameter. We suspect the bolometric flux the planet receives from the host star (as parameterized by T) may be important for us because MEarth is most sensitive to planets larger than $2R_\oplus$ and that are very close-in to their stars. In the *Kepler* context, Rogers et al. (2011) and Lopez et al. (2012) have argued that close-in $> 2R_\oplus$ planets, most of which will have gaseous outer layers, have been sculpted in large part by thermal evolution and atmospheric mass loss. The dependence of these processes on the amount of bolometric (and UV) flux that a planet receives motivates us to explore T as a population-defining parameter in the following analyses.

5.4.1 MEarth found no Jupiter-sized Exoplanets

We do not explicitly simulate our sensitivity to $> 4R_\oplus$ exoplanets. However, Figure 5.3 still encodes MEarth's ability to detect gas giants. Transits of Jupiter-sized ($11R_\oplus$) exoplanets are 8% deep for MEarth's largest targeted M dwarfs ($0.35R_\odot$), and could be as deep as 100% for the smallest M dwarfs. Planets this big could be detected at extremely high significance with a single MEarth observation. Our Jupiter sensitivity is probably very close to the survey window function in Figure 5.3, and is definitely

bounded between the window function and the Neptune-size ($4R_\oplus$) sensitivity curve. Conservatively, we'll assume the Jupiter sensitivity lies halfway between these two curves, in log space.

Figure 5.3 then indicates that if every M dwarf hosted exactly one Jupiter-sized planet in a 1.0 day orbit, MEarth should have found 70 such planets. By inverting a Poisson distribution, we can translate this statement into a 95% confidence upper limit that the occurrence of such 1-day Jupiters is < 0.042 planets/star in our sample. This upper limit is not surprising. Radial velocity surveys show that about 1% of solar-type stars host of hot Jupiters (Wright et al. 2012) and that giant planet occurrence (across all periods) decreases toward less massive stellar hosts (Johnson et al. 2010). Only two warm Jupiters are known to transit M dwarfs (Johnson et al. 2012; Triaud et al. 2013a), and for both systems the $0.6M_\odot$ host stars are so massive that they sit at the boundary of the K and M spectral types.

MEarth can also place an upper limit on the presence of Jupiters in orbits approaching M dwarfs' habitable zones. From the right panel of Figure 5.3, if every star hosted a Jupiter-sized planet with a zero-albedo equilibrium temperature of 300K, MEarth should have found 7. Our non-detection of any 300K Jupiters translates to < 0.44 planets/star for such planets at 95% confidence, again showing contradiction neither with M dwarf radial velocity surveys (Endl et al. 2003; Johnson et al. 2010; Bonfils et al. 2013) nor with microlensing surveys (Sumi et al. 2010; Gould et al. 2010).

Our findings from MEarth agree with other surveys that Jupiter-sized planets are uncommon around M dwarfs. We do not placer deeper limits than other surveys, although we do probe a less massive population of potential stellar hosts than they do.

To be sensitive to small planets, MEarth observes a small number of stars with high photometric precision. Placing deeper constraints on the occurrence of Jupiter-sized planets can be better achieved by those surveys that monitor a larger number of M dwarfs, such as PTF/M-Dwarfs (Law et al. 2012) or the WFCAM Transit Survey (Nefs et al. 2012).

5.4.2 MEarth found no Neptune-sized Exoplanets

MEarth is very sensitive to transits of planets the size of Neptune; the $4R_{\oplus}$ sensitivity curve is within a factor of two of the observational window function (the best we could possibly do) across the period and temperature ranges plotted in Figure 5.3. In light of this sensitivity, how does our lack of detections of transiting Neptunes compare with known populations?

In particular, the two Neptune-sized exoplanets that transit nearby early M dwarfs, Gl436b and GJ3470b, appear to define a family that we should address. Gl436b (Butler et al. 2004; Gillon et al. 2007b) is a $23M_{\oplus}$, $4.2R_{\oplus}$ planet transiting a $0.45M_{\odot}$, $0.46R_{\odot}$ M dwarf, with a 2.6 day orbital period that corresponds to a zero-albedo equilibrium temperature of 640K (Maness et al. 2007; Torres 2007). GJ3470b (Bonfils et al. 2012) is a $14M_{\oplus}$, $4.8R_{\oplus}$ planet transiting a $0.54M_{\odot}$, $0.57R_{\odot}$ M dwarf, with a 3.3 day orbital period that corresponds to a zero-albedo equilibrium temperature of 680K (Demory et al. 2013). Both planets were detected initially with radial velocities and later found to transit.

Figure 5.3 indicates that MEarth should have found 20 Neptune-sized planets in 3 day periods if every one of our targets hosted such a planet, or 60 Neptune-sized planets

with 650K temperatures, if every star had one. The corresponding 95% confidence upper limits are <0.15 planets/star or <0.06 planets/star, depending on whether orbital period or the flux received from the star more genuinely define the population. For comparison, on the basis of two radial velocity planet detections from HARPS, Bonfils et al. (2013) report that 3^{+4}_{-1}% of stars in their M dwarf sample host planets with minimum masses ($m \sin i$) between 10 and 100$M_\oplus$ and orbital periods of 1 to 10 days. On the basis of one transit detection, Dressing & Charbonneau (2013) find that the *Kepler* early M dwarfs host $0.004^{+0.0062}_{-0.002}$ planets per star, for planets with radii of $4.0 - 5.7 R_\oplus$ and periods <10 days. Determining whether the difference between these two numbers can be attributed to features in the densities for planets in this regime (see Wolfgang & Laughlin 2012) will require a more thorough analysis than we wish to present here and a reduction in the Poisson uncertainties in the measurements. The upper limits from MEarth do not contradict either of these measurements.

For a consistency check, it is interesting to ask whether we would have expected the two transiting radial velocity Neptunes Gl436b and GJ3470b to have been discovered, given the number of radial velocity planets that are not known to transit. Bonfils et al. (2013) compiled a list of all known planets orbiting M dwarfs. This list contains 10 non-transiting planets detected in radial velocity surveys with minimum masses in the range of $5 < m \sin i < 40 M_\oplus$. We supplement this list with the two transiting systems and add up the values of the *a priori* transit probability ($= R_\star/a$) for the 12 published radial velocity planets in this mass range. We find that the total expected number of radial velocity planets that should transit is 0.5. An analysis using the planets and parameters listed in the `exoplanets.org` database (Wright et al. 2011) yields the same result. The Poisson probability of having ≥ 2 detections when 0.5 are expected is only

10%; Gl436b and GJ3470b may have been lucky transit detections. Another possibility is that the list of published planets does not represent all the known radial velocity detections, and that M dwarf radial velocity surveys have a pool of unpublished planet candidates still being searched for transits.

5.4.3 MEarth found GJ1214b, a warm sub-Neptune

In §5.4.1 and §5.4.2, we made rough statements based on visual inspection of the sensitivity curves in Figure 5.3. These point estimates effectively treat the underlying planet populations as δ-functions in radius-period-temperature space. Directly integrating MEarth's sensitivity against hypothesized occurrence rate distributions is a more satisfying approach, and one that we adopt in this section to interpret our single $2.7 R_\oplus$, 1.6 day, 570K planet detection GJ1214b. We directly compare our single planet to how many planets we should have found assuming certain extrapolations from *Kepler*.

Throughout, we assume GJ1214b's orbital eccentricity is $e = 0$. A finite eccentricity could both alter GJ1214b's size (Carter et al. 2011) and (modestly) bias its transit probability, but the effect on our statistical inferences remain small compared to the Poisson uncertainty of a single detection.

Following Howard et al. (2012), we define a planet occurrence distribution function $d^2N/(d\log R\, d\log P)$ as the number of planets per star per logarithmic interval in radius and period. With this, we can calculate the number of planets we should have expected to find as

$$N_{\rm expected} = \int\int \frac{S(R,P) \times d^2N}{d\log R\; d\log P} d\log R\; d\log P \tag{5.2}$$

where the integrals are carried out over the radius and period ranges of interest. The

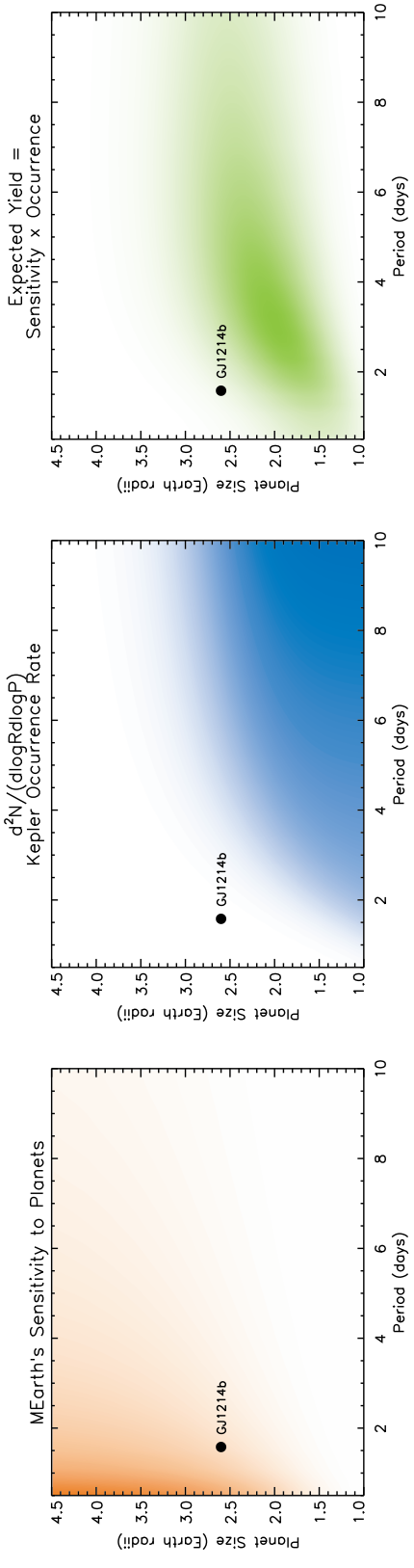

Figure 5.4: *Left:* MEarth's sensitivity (in orange), as a function of planet radius and orbital period. This panel is a 2-dimensional representation of exactly the same information as in the left plot of Figure 5.3. *Middle:* The occurrence rate (in blue) of planets from *Kepler* for early M dwarf hosts and showing an analytic fit for $d^2N/(d\log R\ d\log P)$ from the data of Dressing & Charbonneau (2013). *Right:* The product of the other two panels (in green), showing the probability distribution for MEarth's expected yield (and the integrand in Eq. 5.2). The integral over the parameter space shown in this panel indicates MEarth should have found 0.052 ± 0.015 planets/year. The color scale is linear in all three panels.

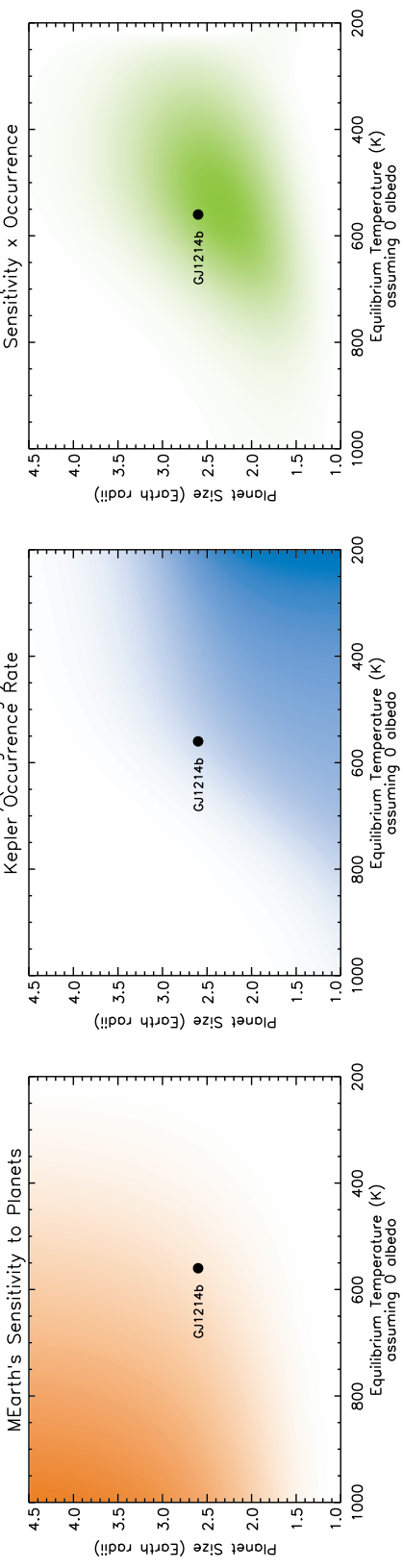

Figure 5.5: The same as Figure 5.4, except treating the planet occurrence distribution as a function of equilibrium temperature, $d^2N/(d\log R\ d\log T)$, as fixed between *Kepler*'s early M dwarfs and MEarth's mid-to-late M dwarfs. The integral over the parameter space shown indicates MEarth should have found 0.36 ± 0.08 planets/year. The color scale is linear in all three panels, but not the same as in Figure 5.4.

expression is identical as a function of equilibrium temperature, simply exchanging T for P.

Dressing & Charbonneau (2013) quote occurrence rates in bins of radius and period, but the bins are large compared to the scales on which MEarth's sensitivity changes substantially. To allow smooth interpolation over the parameter space and improve the accuracy of our calculation of $N_{\rm expected}$, we were motivated to fit analytic functions to the distribution. We do so in the Appendix, using the non-binning maximum likelihood formalism of Youdin (2011) and the updated M dwarf and planetary parameters for the cool *Kepler* Objects of Interest from Dressing & Charbonneau (2013). The results of these fits are analytic expressions for $d^2N/(d\log R\ d\log P)$ and $d^2N/(d\log R\ d\log T)$, along with MCMC samples that can be used to propagate the uncertainties in these distributions forward.

Figure 5.4 shows the steps leading to $N_{\rm expected}$, if we assume planetary occurrence as a function of planetary radius and orbital period – $d^2N/(d\log R\ d\log P)$ – is exactly the same (in both shape and normalization) between *Kepler*'s M dwarfs and the MEarth sample. The left panel shows $S(R, P)$, and the middle panel shows our fit to the *Kepler* M dwarf occurrence distribution. The right panel shows the product of the first two panels (the integrand in Equation 5.2), which we refer to as MEarth's expected yield. Expressed as an average over the first four years of the MEarth survey, the integral over the plotted parameter space predicts a total yield of $N_{\rm expected} = 0.052 \pm 0.015$ planets/year, where the error bar represents the central 1σ range of an ensemble of $N_{\rm expected}$ estimates calculated by randomly sampling both from our perturbed $S(R, P)$ curves (§5.3.4) and from the Markov chain for the occurrence rate fit.

This overall $N_{\rm expected}$ is technically consistent with our detection of GJ1214b; the Poisson probability of finding 1 or more planets when $4 \times 0.052 = 0.21$ planets are expected is 19%. However, the shape of the yield distribution also indicates what kind of planets would be the most probable for MEarth to find. In Figure 5.4, GJ1214b is set apart from the bulk of the expected yield distribution; $2.7{\rm R}_\oplus$, 1.6 day planets are exceedingly rare in the *Kepler* sample. If the hypothesis that $d^2N/(d\log R\ d\log P)$ stays the same between MEarth and *Kepler* holds, then the one planet that MEarth found should probably have had a longer period than GJ1214b, or been a bit smaller.

The agreement is much better if we assume that planet occurrence as function of radius and incident stellar flux – $d^2N/(d\log R\ d\log T)$ – is the same across stellar samples. Such a scaling is motivated by the dominant role thermal evolution and mass loss likely play in the histories of $2-4{\rm R}_\oplus$ planets (Rogers et al. 2011; Lopez et al. 2012). Figure 5.5 shows the results if we focus on equilibrium temperature T instead of period P. Because MEarth stars are on average cooler and smaller than the *Kepler* M dwarfs, this has the effect of bringing planets closer in to their host stars, making them easier for MEarth to detect. Integrating $S(R,T)$ against $d^2N/(d\log R\ d\log T)$ achieves a total $N_{\rm expected} = 0.36 \pm 0.08$ planets/year, in broad agreement with our single discovery in four years. The peak of the expected yield probability distribution closely matches GJ1214b's $2.7{\rm R}_\oplus$ radius and 570K equilibrium temperature, unlike the case in Figure 5.4.

Of the two hypotheses, keeping the planet occurrence distribution fixed as a function of T appears to be the better match. However, both are consistent with MEarth finding one planet in our first four years. As surveys like MEarth become more sensitive, it will be possible to make more quantitative statements about the differences that exist between the planet populations around mid-to-late M dwarfs and early M dwarfs. On the

main sequence, the M spectral class spans a factor of 5 in stellar mass (compared to the factor of 2.5 spanned by FGK dwarfs combined). Understanding the differences across this range would provide a strong lever arm to probe the environmental dependence of planet formation. Dressing & Charbonneau (2013) may have seen hints of such differences within the *Kepler* sample itself, although their sample of very late M dwarfs is small.

5.5 Discussion: Modifying MEarth

MEarth was never designed to be a statistical survey; it was designed to detect individual systems favorable for follow-up. We can use the machinery developed in this work to address how to better achieve that goal, how to increase the "planets/year" that MEarth could find. *Kepler* has told us planets are bountiful; how do we reap the harvest? Inspection of Figures 5.4 and 5.5 suggests that marginal improvement in MEarth's ability to detect smaller planets or cooler planets could dramatically increase the survey's yield. In this section, we describe changes to the MEarth strategy that could exploit this opportunity. Our conclusions may have broad impact for current and future surveys targeting M dwarfs, particularly those such as APACHE (Giacobbe et al. 2012) designs and strategies similar to MEarth's.

How do we increase sensitivity to smaller planets? Both the equipment and the observing strategy of MEarth were tailored toward detecting $> 2R_\oplus$ planets. To find smaller planets, we need to gather more photons during transit. Working with the existing MEarth observatory, we could achieve this goal either by exposing longer with each telescope during each pointing or by using more telescopes to monitor each

star. The former is dangerous, as systematic noise sources below MEarth's current photon noise limit could potentially counteract the benefits of the extra exposure time. The latter strategy is more compelling. With the exception of a common mode effect introduced by precipitable water vapor (Irwin et al. 2011a; Berta et al. 2011, 2012b), many noise sources should be uncorrelated between telescopes, allowing our noise to bin down with the number of telescopes (see Charbonneau et al. 2009). Increasing the number of telescopes observing each star ($n_{\rm tel}$) decreases the total number of stars that can be observed (by a factor of $n_{\rm tel}$) but also shrinks the minimum detectable planet radius for each (by $n_{\rm tel}^{-0.25}$). We vary $n_{\rm tel}$ and repeat the calculations in §5.4.3. We find $n_{\rm tel} = 2$ optimizes the overall $N_{\rm expected}$, for which a $1.7{\rm R}_\oplus$ planet becomes as detectable as a $2{\rm R}_\oplus$ planet was originally. MEarth's old sensitivity so closely skirts the cliff in occurrence to make the tradeoff of observing fewer stars more deeply worthwhile. In reality, the $n_{\rm tel}^{-0.25}$ scaling is unlikely to hold beyond 2–3 telescopes. When pushing to planets smaller than $1.7{\rm R}_\oplus$with MEarth's instrumentation, our ability to correct for the precipitable water vapor systematic would become a major concern (Berta et al. 2012b). Scintillation would also impose a limit for bright stars and shallow depths, although scintillation may be suppressed by the extra baseline of distributing aperture over multiple telescopes (see Young 1967). For $1.7{\rm R}_\oplus$ planets we consider here, transit are deep enough (4 millimags for a $0.25{\rm R}_\odot$ star) that we probably do not yet reach the multiple-telescope scintillation limit.

How do we increase sensitivity to cooler, longer period planets? The probability that multiple transits will be observable from the same site within one season shrinks dramatically with increasing periods. MEarth's best strategy for longer period planets is to reliably detect planets from single transit events, by using our realtime trigger.

This mode is fully functional; we have blindly rediscovered single transits of GJ1214b at high significance with it. We did not include simulations of the realtime trigger in the sensitivity estimate in §5.3 because we have neither yet confirmed a new transiting planet found through the realtime trigger nor followed up on candidate triggered events sufficiently to know they were not real.

Here, we consider in more detail what the realtime trigger means for MEarth's sensitivity. We repeat the transit injection simulations of §5.3 in a "time-machine" fashion, assessing the significance of a particular transit using only the light curve data up to and including the event. In the MISS MarPLE framework, the significance of a transit depends on how well-constrained the priors describing the star's behavior within a night are; these tighten as more nights of data are gathered (see Berta et al. 2012b). Also, transits that occur early within a night are less significant than those that occur later, after observations have been gathered to constrain the out-of-transit baseline and the instrumental parameters for the night. In the simulations, if a low-cadence transit exceeds a 3σ threshold we pretend the trigger would be activated, populate the remaining time left in the transit with high cadence data points, and assess the post-trigger significance. This is a rough approximation to the actual process that occurs on-sky (a 2.5σ dip triggers the immediate start of a feedback loop to observe and continually reassess the significance of the candidate event; if the significance reaches 5σ, the star is then observed continuously and unconditionally until an egress is detected or two hours elapse). For a threshold, we say that if the post-trigger significance of a single transit exceeds 5.5σ in these simulations, the planet is detected. This threshold is meant to be conservative, and it set by visual inspection of simulated events and of the histogram of single event statistics within our sample; it can be lower than the phased 7.5σ threshold

because fewer independent hypotheses are being tested against the data (see Jenkins et al. 2002). The simulations show that such a realtime trigger dramatically improves sensitivity to 10 day periods, especially for planets that are signal-to-noise limited in single transits (essentially everything smaller than Neptune).

For a planet detected with a single transit, we will have essentially no knowledge of its orbital period. One way to measure the period would be to continue to monitor the star with MEarth. This is an inefficient strategy, reintroducing the problematic bottleneck that MEarth would have to see multiple transits from the same site. External telescopes providing either continuous photometric monitoring or precise radial velocity observations could much more efficiently measure the period and confirm the planet. The photometric option could be achieved either with a longitudinally-distributed array of telescopes such as the Las Cumbres Observatory Global Telescope (LCOGT) network (Pickles et al. 2012) or with a coordinated effort across different observatories. For most planets MEarth could find, the radial velocity option will require roughly m/s precision. This precision is currently possible on mid-to-late M dwarfs only with spectrographs like HARPS, HARPS-N, or HIRES, but will soon be made easier by the forthcoming coming near-IR radial velocity instruments CARMENES (Quirrenbach et al. 2012) and HZPF (Mahadevan et al. 2012).

The combination of all of these improvements makes MEarth more sensitive to smaller, cooler planets. Figure 5.6 compares MEarth's yield distributions for two scenarios, both assuming that $d^2N/(d\log R\ d\log T)$ is fixed between early and late M dwarfs. The top is the original phased search yield, the same as in Figure 5.5. The bottom is the predicted yield if MEarth allocates two telescopes per star and makes full use of the realtime trigger to detect planets with single transits, as described above.

Because fewer stars would be observed with the multiple telescope strategy, we chose those stars wisely: we explicitly biased the target list toward closer, cooler stars (those within 20 pc and smaller than $0.25R_{\odot}$). In both panels of Figure 5.6, we include iso-density contours drawn so that the integral over each contour level will enclose a given number of "planets/year." These contour express the overall normalization of the yield distribution, allowing for direct comparison between the panels. In the original scenario (top panel), the integral over the entire plotted area for the original scenario is 0.36 ± 0.08 planets/year, as it was in §5.4.3. In contrast, with the updated scenario (bottom panel), the integral over this parameter space is 0.89 ± 0.13. Simply by revising our operational strategy, we can increase the expected yield of the survey by a factor of $2.5\times$.

5.6 Conclusions

In this paper, we simulated the sensitivity of the MEarth Project to exoplanets during its first four years. We found MEarth's detection of one exoplanet to be consistent with the planet occurrence distribution measured around M dwarfs by the *Kepler* mission (Dressing & Charbonneau 2013), given reasonable assumptions about how this distribution scales between *Kepler*'s early M dwarfs and MEarth's cooler mid-to-late M dwarfs. *Kepler* results indicate that planets are common around M dwarfs, but these planets are concentrated at orbital separations and sizes just beyond MEarth's sensitivity. Our non-detection of planets the size of Neptune or Jupiter is consistent with literature estimates for early M dwarfs, and provides new upper limits on the occurrence rates of these planets for mid-to-late M dwarfs.

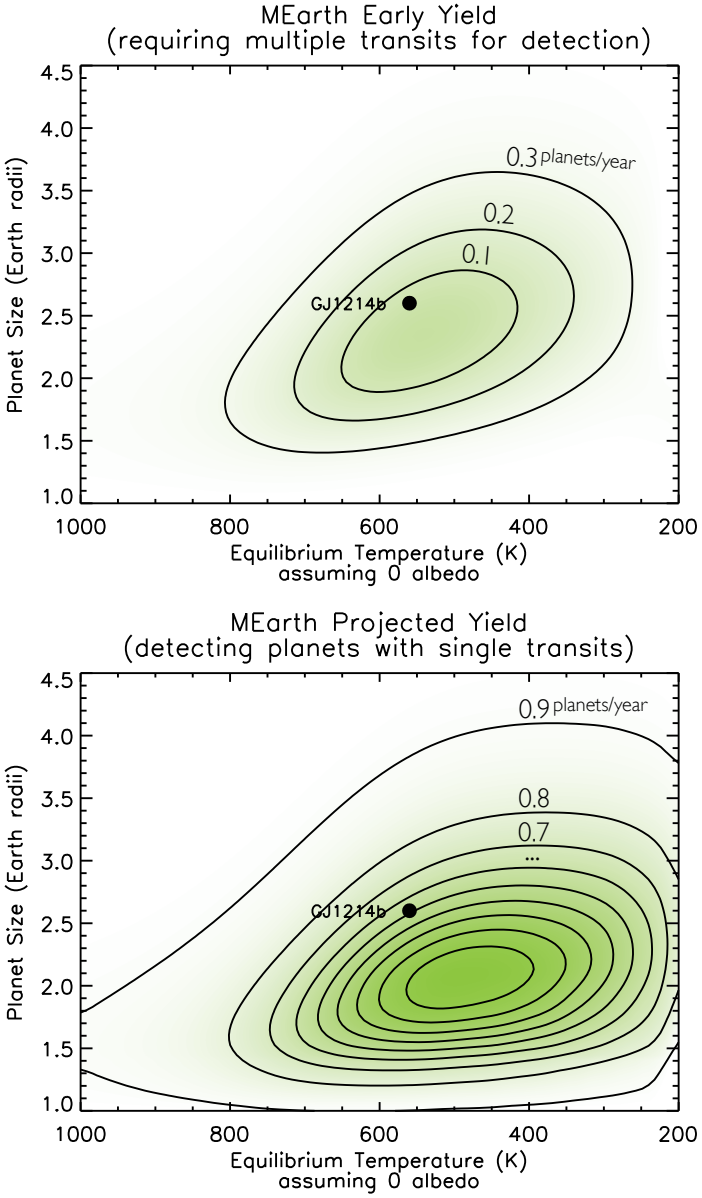

Figure 5.6: MEarth's planet yield, as shown in Figure 5.5 but including iso-density contours for quantitative comparison between plots. Two scenarios are shown: that achieved from phase-folded searches of the first four years of observations (*top*) and that which could be achieved if multiple MEarth telescopes observe the same target, planets are reliably detected with single transits, and external facilities can be used to measure their periods. Each contour is drawn such that the integral over the enclosed region corresponds to a fixed number of "planets/year" that MEarth should be finding. The contours increase outward in intervals of 0.1 planets/year.

Future work will more tightly constrain how the planet occurrence distribution changes over stellar masses. As a first hypothesis, we found here that we could satisifactorilty explain our discovery of GJ1214b by assuming that planet occurrence stays fixed with respect to a planet's radius and the flux it receives from its host star, when extrapolating between *Kepler*'s earlier M dwarfs to MEarth's later M dwarfs.

We illustrated how targeted changes to MEarth's operational program could boost its overall expected yield, by sculpting its sensitivity to match the planet distribution observed by *Kepler*. By sacrificing some sensitivity to large planets that we know to be rare, we can improve our sensitivity to small, cool planets that we now know to be common. Having implemented these modifications in the fall of 2012, we eagerly await the planet discoveries they may bear. The analyses we present in this paper may also help in the design of larger, future surveys to find smaller, Earth-sized planets in the habitable zones of nearby M dwarfs.

Acknowledgements

We thank Jason Dittmann and Elisabeth Newton for detailed conversations regarding the MEarth stellar sample, and Courtney Dressing for discussing *Kepler* M planet dwarf statistics. The MEarth Project gratefully acknowledges funding from the David and Lucile Packard Fellowship for Science and Engineering and from the National Science Foundation (NSF) under grant number AST-0807690. We are greatly indebted to the staff at the Fred Lawrence Whipple Observatory for their efforts in construction and maintenance of the MEarth facility and would like to thank Emilio Falco, Wayne Peters, Ted Groner, Karen Erdman-Myres, Grace Alegria, Rodger Harris, Bob Hutchins, Dave

Martina, Dennis Jankovsky, Tom Welsh, Robert Hyne, Mike Calkins, Perry Berlind, and Gil Esquerdo for their support. This research has made use of NASA's Astrophysics Data System.

Appendix to Chapter 5: Estimating Analytic Planet Occurrence Distributions from Kepler's Early M Dwarfs

Here, we estimate analytic functions that match the occurrence of planets implied by *Kepler*'s sample of planet candidates transiting early M dwarfs. We use the maximum likelihood approach outlined in Youdin (2011). We assume both that all the planet candidates listed in Dressing & Charbonneau (2013) are real planets (Morton & Johnson 2011), and that the *Kepler* pipeline is complete above the stated 7.1σ detection threshold (Tenenbaum et al. 2012).

We estimate *Kepler*'s sensitivity across the parameter space by calculating representative $\eta_{\rm tra}$ and $\eta_{\rm det}$ curves from the data in Dressing & Charbonneau (2013). We use the median stellar densities and luminosities of the M dwarf planet candidate hosts to calculate the relation between $\eta_{\rm tra}$ and P and T. Dressing & Charbonneau (2013) estimate the fraction of their 3897 *Kepler* M dwarfs around which each planet candidate could have been detected; as in Youdin (2011), we fit joint power-laws for $r_{\rm det} = \eta_{\rm det}/(1-\eta_{\rm det})$ as functions of (R, P) and (R, T) to those per-star estimates (provided by Courtney Dressing, private communication). The resulting analytic sensitivities $S_{Kepler}(R, P)$ and $S_{Kepler}(R, T)$ are shown in the left panels of Figures 5.7

and 5.8.

In the middle panels of Figures 5.7 and 5.8, we plot the *Kepler* M dwarf planet candidates as black circles. The symbol area is inversely proportional to $\eta_{\rm tra} \times \eta_{\rm det}$; with this scaling, the amount of black ink on the page should reflect the occurrence rate in that region of parameter space. We restrict our analysis to planet candidates in the plotted parameter ranges, with $0.8{\rm R}_\oplus < R < 4.5{\rm R}_\oplus$, 0.25 days $< P <$ 50 days, and $1500K > T > 200K$. The candidate KOI531.01 ($4.8{\rm R}_\oplus$, 3.7 days) sits just outside this region and is not included, as we are primarily interested in planets smaller than Neptune.

We write an analytic function for the occurrence rate distribution that is purely descriptive, meant to provide a smooth approximation over the parameter space. The functional form we choose is

$$\frac{d^2N}{d\log R\, d\log X} = \frac{(X/a)^\alpha}{1+\exp\left[\frac{R-R_0(X)}{\sigma}\right]} \text{ where } R_0(X) = \frac{b}{1+(X_0/X)^\beta} \tag{5.3}$$

where X represents either P in days or T in degrees Kelvin. This expression tries to capture three qualitative features of the *Kepler* M dwarf planet candidates. The power law in the period-like parameter $(X/a)^\alpha$ captures the steady rise toward longer periods and cooler temperatures. The sigmoid factor $1/[1+\exp(...)]$ describes a flat occurrence below a certain planet radius and steep dropoff above that radius (see Figure 16 of Dressing & Charbonneau 2013). Allowing the critical radius $R_0(X)$ to increase away from the star reproduces the absence of big planets and presence of small planets at small orbital distances.

Combining these components, we write down the Poisson likelihood *Kepler* would have found the planets that it did. We maximize this likelihood using an MCMC

method (Berta et al. 2012a), and quote the mostly likely coefficients in Table 5.1. The right panels of Figures 5.7 and 5.8 show the expected yield with these coefficients. The black dots should trace this maximum likelihood yield distribution; this is the space in which the data and our model are compared. The middle panels show the inferred $d^2N/(d\log R\ d\log P)$ and $d^2N/(d\log R\ d\log T)$, along with the candidates weighted as described above.

The uncertainties on the coefficients are large and strongly correlated. We do not quote these uncertainties here, but we use the MCMC samples themselves to propagate the uncertainties into the predictions in §5.4 and §5.5. In Figure 5.8, the planet temperature occurrence distribution is particularly uncertain coolward of 300K. At these temperatures, the demographics of the stars contributing to the sensitivity shifts toward the cooler ones, breaking our implicit assumption of uniform sample demographics in $\eta_{\rm tra}$ and $\eta_{\rm det}$. Many of the planets at these large separations also have very large (of order unity) uncertainties in their estimated radii (see Figure 17 of Dressing & Charbonneau 2013), which we did not account for here.

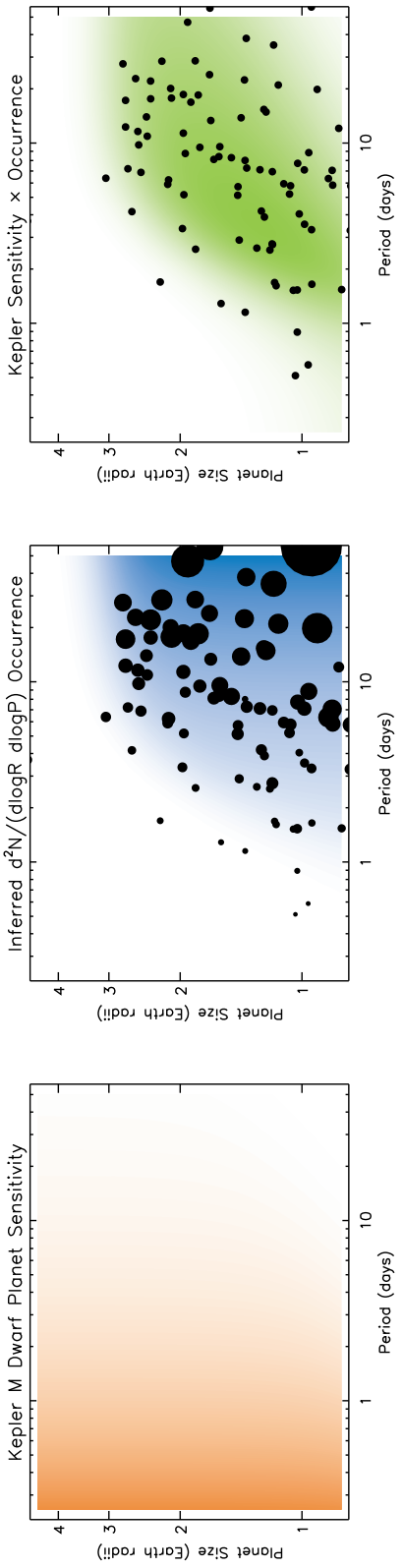

Figure 5.7: Demonstration of our calculation of an analytic $d^2N/(d\log R\, d\log P)$ planet distribution for *Kepler*'s M dwarfs, using the Dressing & Charbonneau (2013) stellar/planetary parameters and sensitivity estimates. *Left:* Estimate of *Kepler*'s sensitivity in orange. *Middle:* The inferred *Kepler* occurrence rate of planets around early M dwarfs in blue. *Kepler* planet candidates are shown as black circles whose areas scale with the number of missing planets they would represent after accounting for sensitivity effects. *Right:* The expected planet yield distribution from the analytic model in green; by design, this matches the distribution of the *Kepler* planet candidates (black circles).

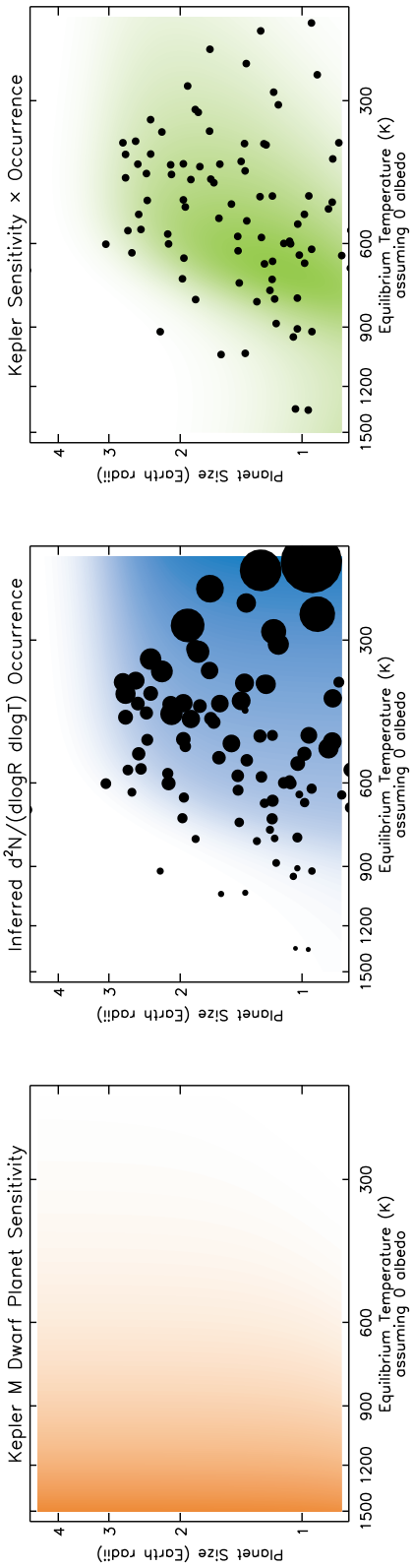

Figure 5.8: Exactly as in Figure 5.7, except for $d^2N/(d\log R\, d\log T)$.

Table 5.1. The maximum likelihood coefficients for Equation 5.3, for either $X = P$ or $X = T$.

Coefficient	Period (P)	Temperature (T)
a	11.07 days	1780K
α	0.51	-0.74
σ	$0.29R_\oplus$	$0.40R_\oplus$
b	$2.9R_\oplus$	$2.8R_\oplus$
β	1.6	-6.4
X_0	2.06 days	744K

Note. — No uncertainty estimates are provided here; these coefficients are meant to be used for rough interpolation only. They should not be used outside the radius, period, or temperature ranges on which they were fitted.

References

Aarnio, A. N., Matt, S. P., & Stassun, K. G. 2012, ApJ, 760, 9

Adams, E. R., Seager, S., & Elkins-Tanton, L. 2008, ApJ, 673, 1160

Agol, E. 2011, ApJ, 731, L31

Agol, E., Cowan, N. B., Knutson, H. A., et al. 2010, ApJ, 721, 1861

Aigrain, S., & Favata, F. 2002, A&A, 395, 625

Aigrain, S., Favata, F., & Gilmore, G. 2004, A&A, 414, 1139

Aigrain, S., Hodgkin, S., Irwin, J., et al. 2007, MNRAS, 375, 29

Aigrain, S., & Irwin, M. 2004, MNRAS, 350, 331

Aigrain, S., Pont, F., Fressin, F., et al. 2009, A&A, 506, 425

Aigrain, S., Pont, F., & Zucker, S. 2012, MNRAS, 419, 3147

Albrecht, S., Winn, J. N., Johnson, J. A., et al. 2012, ApJ, 757, 18

Alonso, R., Brown, T. M., Torres, G., et al. 2004, ApJ, 613, L153

Andrews, S. M., & Williams, J. P. 2005, ApJ, 631, 1134

Anglada-Escudé, G., Rojas-Ayala, B., Boss, A. P., Weinberger, A. J., & Lloyd, J. P. 2013, A&A, 551, A48

Anglada-Escudé, G., Shkolnik, E. L., Weinberger, A. J., et al. 2010, ApJ, 711, L24

Appenzeller, I., Fricke, K., Fürtig, W., et al. 1998, The Messenger, 94, 1

Apps, K., Clubb, K. I., Fischer, D. A., et al. 2010, PASP, 122, 156

Bailer-Jones, C. A. L., & Lamm, M. 2003, MNRAS, 339, 477

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

Bakos, G. Á., Csubry, Z., Penev, K., et al. 2013, PASP, 125, 154

Bakos, G. Á., Torres, G., Pál, A., et al. 2010, ApJ, 710, 1724

Ballard, S., Christiansen, J. L., Charbonneau, D., et al. 2010, ApJ, 716, 1047

Baraffe, I., Chabrier, G., Allard, F., & Hauschildt, P. H. 1998, A&A, 337, 403

Barnes, J. R., Jeffers, S. V., & Jones, H. R. A. 2011, MNRAS, 412, 1599

Barnes, R., Jackson, B., Greenberg, R., & Raymond, S. N. 2009, ApJ, 700, L30

Barnes, R., Raymond, S. N., Jackson, B., & Greenberg, R. 2008, Astrobiology, 8, 557

Batalha, N. M., Borucki, W. J., Bryson, S. T., et al. 2011, ApJ, 729, 27

Batalha, N. M., Rowe, J. F., Bryson, S. T., et al. 2013, ApJS, 204, 24

Bean, J. L. 2009, A&A, 506, 369

Bean, J. L., Désert, J.-M., Kabath, P., et al. 2011, ApJ, 743, 92

Bean, J. L., Miller-Ricci Kempton, E., & Homeier, D. 2010a, Nature, 468, 669

Bean, J. L., Seifahrt, A., Hartman, H., et al. 2010b, ApJ, 711, L19

Beatty, T. G., Pepper, J., Siverd, R. J., et al. 2012, ApJ, 756, L39

Beatty, T. G., & Seager, S. 2010, ApJ, 712, 1433

Benedict, G. F., McArthur, B., Nelan, E., et al. 1998, AJ, 116, 429

Benneke, B., & Seager, S. 2012, ApJ, 753, 100

Berta, Z. K., Charbonneau, D., Bean, J., et al. 2011, ApJ, 736, 12

Berta, Z. K., Charbonneau, D., Désert, J.-M., et al. 2012a, ApJ, 747, 35

Berta, Z. K., Irwin, J., Charbonneau, D., Burke, C. J., & Falco, E. E. 2012b, AJ, 144, 145

Bessell, M. S. 1990, PASP, 102, 1181

Blake, C. H., Bloom, J. S., Latham, D. W., et al. 2008, PASP, 120, 860

Blake, C. H., & Shaw, M. M. 2011, PASP, 123, 1302

Bochanski, J. J., Hawley, S. L., Covey, K. R., et al. 2010, AJ, 139, 2679

Bonfils, X., Delfosse, X., Udry, S., et al. 2005a, A&A, 442, 635

—. 2013, A&A, 549, A109

Bonfils, X., Forveille, T., Delfosse, X., et al. 2005b, A&A, 443, L15

Bonfils, X., Gillon, M., Udry, S., et al. 2012, A&A, 546, A27

Bonomo, A. S., & Lanza, A. F. 2008, A&A, 482, 341

Borucki, W. J., Koch, D., Basri, G., et al. 2010, Science, 327, 977

Borucki, W. J., Koch, D. G., Batalha, N., et al. 2012, ApJ, 745, 120

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

Bouchy, F., Mayor, M., Lovis, C., et al. 2009, A&A, 496, 527

Bouchy, F., Udry, S., Mayor, M., et al. 2005, A&A, 444, L15

Bowler, B. P., Liu, M. C., Shkolnik, E. L., & Tamura, M. 2012, ApJ, 756, 69

Boyajian, T. S., McAlister, H. A., van Belle, G., et al. 2012a, ApJ, 746, 101

Boyajian, T. S., von Braun, K., van Belle, G., et al. 2012b, ApJ, 757, 112

Bramich, D. M., Horne, K., Bond, I. A., et al. 2005, MNRAS, 359, 1096

Brown, T. M. 2001, ApJ, 553, 1006

Brown, T. M., Charbonneau, D., Gilliland, R. L., Noyes, R. W., & Burrows, A. 2001, ApJ, 552, 699

Browning, M. K., Basri, G., Marcy, G. W., West, A. A., & Zhang, J. 2010, AJ, 139, 504

Burke, C. J. 2008, ApJ, 679, 1566

Burke, C. J., Gaudi, B. S., DePoy, D. L., & Pogge, R. W. 2006, AJ, 132, 210

Burke, C. J., McCullough, P. R., Bergeron, L. E., et al. 2010, ApJ, 719, 1796

Burke, C. J., McCullough, P. R., Valenti, J. A., et al. 2007, ApJ, 671, 2115

Burrows, A., Rauscher, E., Spiegel, D. S., & Menou, K. 2010, ApJ, 719, 341

Butler, R. P., Vogt, S. S., Marcy, G. W., et al. 2004, ApJ, 617, 580

Carpano, S., Aigrain, S., & Favata, F. 2003, A&A, 401, 743

Carter, J. A., Agol, E., Chaplin, W. J., et al. 2012, Science, 337, 556

Carter, J. A., & Winn, J. N. 2009, ApJ, 704, 51

Carter, J. A., Winn, J. N., Gilliland, R., & Holman, M. J. 2009, ApJ, 696, 241

Carter, J. A., Winn, J. N., Holman, M. J., et al. 2011, ApJ, 730, 82

Carter, J. A., Yee, J. C., Eastman, J., Gaudi, B. S., & Winn, J. N. 2008, ApJ, 689, 499

Casagrande, L., Flynn, C., & Bessell, M. 2008, MNRAS, 389, 585

Chabrier, G., Gallardo, J., & Baraffe, I. 2007, A&A, 472, L17

Charbonneau, D., Allen, L. E., Megeath, S. T., et al. 2005, ApJ, 626, 523

Charbonneau, D., Berta, Z. K., Irwin, J., et al. 2009, Nature, 462, 891

Charbonneau, D., Brown, T. M., Latham, D. W., & Mayor, M. 2000, ApJ, 529, L45

Charbonneau, D., Brown, T. M., Noyes, R. W., & Gilliland, R. L. 2002, ApJ, 568, 377

Charbonneau, D., Knutson, H. A., Barman, T., et al. 2008, ApJ, 686, 1341

Chiang, E., & Laughlin, G. 2013, MNRAS

Choi, J., McCarthy, C., Marcy, G. W., et al. 2013, ApJ, 764, 131

Claret, A. 1998, A&A, 335, 647

—. 2000, A&A, 363, 1081

—. 2004, A&A, 428, 1001

Claret, A., & Hauschildt, P. H. 2003, A&A, 412, 241

Clay, R. W., Wild, N. R., Bird, D. J., et al. 1998, PASA, 15, 332

Cosentino, R., Lovis, C., Pepe, F., et al. 2012, in SPIE Conference Series, Vol. 8446

Covey, K. R., Hawley, S. L., Bochanski, J. J., et al. 2008, AJ, 136, 1778

Cowan, N. B., & Agol, E. 2011, ApJ, 729, 54

Croll, B., Albert, L., Jayawardhana, R., et al. 2011, ApJ, 736, 78

Croll, B., Matthews, J. M., Rowe, J. F., et al. 2007, ApJ, 671, 2129

Crossfield, I. J. M., Barman, T., & Hansen, B. M. S. 2011, ApJ, 736, 132

Czesla, S., Huber, K. F., Wolter, U., Schröter, S., & Schmitt, J. H. M. M. 2009, A&A, 505, 1277

de Mooij, E. J. W., Brogi, M., de Kok, R. J., et al. 2012, A&A, 538, A46

Defaÿ, C., Deleuil, M., & Barge, P. 2001, A&A, 365, 330

Delfosse, X., Bonfils, X., Forveille, T., et al. 2013, A&A, 553, A8

Delfosse, X., Forveille, T., Beuzit, J.-L., et al. 1999, A&A, 344, 897

Delfosse, X., Forveille, T., Perrier, C., & Mayor, M. 1998, A&A, 331, 581

Delfosse, X., Forveille, T., Ségransan, D., et al. 2000, A&A, 364, 217

Deming, D., Harrington, J., Seager, S., & Richardson, L. J. 2006, ApJ, 644, 560

Deming, D., Seager, S., Richardson, L. J., & Harrington, J. 2005, Nature, 434, 740

Deming, D., Seager, S., Winn, J., et al. 2009, PASP, 121, 952

Deming, D., Wilkins, A., McCullough, P., et al. 2013, arXiv:1302.1141, submitted to ApJ

Demory, B., Ségransan, D., Forveille, T., et al. 2009, A&A, 505, 205

Demory, B.-O., Torres, G., Neves, V., et al. 2013, ApJ, 768, 154

Désert, J.-M., Bean, J., Miller-Ricci Kempton, E., et al. 2011a, ApJ, 731, L40+

Désert, J.-M., Lecavelier des Etangs, A., Hébrard, G., et al. 2009, ApJ, 699, 478

Désert, J.-M., Sing, D., Vidal-Madjar, A., et al. 2011b, A&A, 526, A12

Desidera, S., Sozzetti, A., Bonomo, A. S., et al. 2013, arXiv:1302.3093, acccepted to A&A

Dittmann, J., Irwin, J. M., Berta, Z. K., Charbonneau, D., & the MEarth Team. 2012, in AAS Meeting Abstracts #219, Vol. 219, #345.09

Dittmann, J. A., Close, L. M., Green, E. M., & Fenwick, M. 2009, ApJ, 701, 756

Dorren, J. D. 1987, ApJ, 320, 756

Dotter, A., Chaboyer, B., Jevremović, D., et al. 2008, ApJS, 178, 89

Doyle, L. R., Deeg, H. J., Kozhevnikov, V. P., et al. 2000, ApJ, 535, 338

Dragomir, D., Matthews, J. M., Howard, A. W., et al. 2012, ApJ, 759, L41

Dravins, D., Lindegren, L., Mezey, E., & Young, A. T. 1998, PASP, 110, 610

Dressel, L., Wong, M., Pavlovsky, C., Long, K., et al. 2010, "Wide Field Camera 3 Instrument Handbook", Version 2.1, (Baltimore: STScI), `http://www.stsci.edu/hst/wfc3/`

Dressing, C. D., & Charbonneau, D. 2013, ApJ, 767, 95

Dumusque, X., Pepe, F., Lovis, C., et al. 2012, Nature, 491, 207

Dunkley, J., Bucher, M., Ferreira, P. G., Moodley, K., & Skordis, C. 2005, MNRAS, 356, 925

Dupuy, T. J., & Liu, M. C. 2009, ApJ, 704, 1519

Dzigan, Y., & Zucker, S. 2011, MNRAS, 415, 2513

—. 2012, ApJ, 753, L1

Eastman, J., Siverd, R., & Gaudi, B. S. 2010, PASP, 122, 935

Ehrenreich, D., Bourrier, V., Bonfils, X., et al. 2012, A&A, 547, A18

Ehrenreich, D., Tinetti, G., Lecavelier Des Etangs, A., Vidal-Madjar, A., & Selsis, F. 2006, A&A, 448, 379

Encrenaz, T. 2005, Space Sci. Rev., 116, 99

Endl, M., Cochran, W. D., Tull, R. G., & MacQueen, P. J. 2003, AJ, 126, 3099

Evans, D. W., Irwin, M. J., & Helmer, L. 2002, A&A, 395, 347

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

Fang, J., & Margot, J.-L. 2012, ApJ, 761, 92

Ferraz-Mello, S., Tadeu Dos Santos, M., Beaugé, C., Michtchenko, T. A., & Rodríguez, A. 2011, A&A, 531, A161

Fischer, D. A., & Marcy, G. W. 1992, ApJ, 396, 178

Fischer, D. A., & Valenti, J. 2005, ApJ, 622, 1102

Ford, E. B. 2005, AJ, 129, 1706

Fortney, J. J. 2005, MNRAS, 364, 649

Fortney, J. J., Marley, M. S., & Barnes, J. W. 2007, ApJ, 659, 1661

Fortney, J. J., Shabram, M., Showman, A. P., et al. 2010, ApJ, 709, 1396

Fraine, J. D., Deming, D., Gillon, M., et al. 2013, ApJ, 765, 127

France, K., Froning, C. S., Linsky, J. L., et al. 2013, ApJ, 763, 149

Fressin, F., Torres, G., Charbonneau, D., et al. 2013, ApJ, 766, 81

Fressin, F., Torres, G., Rowe, J. F., et al. 2012, Nature, 482, 195

Fukugita, M., Ichikawa, T., Gunn, J. E., et al. 1996, AJ, 111, 1748

Gaudi, B. S., Seager, S., & Mallen-Ornelas, G. 2005, ApJ, 623, 472

Gautier, D., Conrath, B. J., Owen, T., de Pater, I., & Atreya, S. K. 1995, in Neptune and Triton, ed. D. P. Cruikshank, M. S. Matthews, & A. M. Schumann, 547–611

Giacobbe, P., Damasso, M., Sozzetti, A., et al. 2012, MNRAS, 424, 3101

Gibson, N. P., Aigrain, S., Roberts, S., et al. 2012, MNRAS, 419, 2683

Gibson, N. P., Pont, F., & Aigrain, S. 2011, MNRAS, 411, 2199

Gillon, M., Bonfils, X., Demory, B.-O., et al. 2011, A&A, 525, A32+

Gillon, M., Demory, B.-O., Barman, T., et al. 2007a, A&A, 471, L51

Gillon, M., Pont, F., Demory, B., et al. 2007b, A&A, 472, L13

Gliese, W., & Jahreiß, H. 1979, A&AS, 38, 423

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

Gregory, P. C. 2005, Bayesian Logical Data Analysis for the Physical Sciences (Cambridge University Press)

Guillot, T., & Gautier, D. 2009, Treatise of Geophysics, vol. 10, Planets and Moons, Schubert G., Spohn T. (Ed.) (2007) 439-464, available as arXiv:0912.2019

Hansen, B. M. S., & Murray, N. 2012, ApJ, 751, 158

Harrington, R. S., & Dahn, C. C. 1980, AJ, 85, 454

Hartman, J. D., Bakos, G. Á., Kovács, G., & Noyes, R. W. 2010, MNRAS, 408, 475

Hatzes, A. P., Dvorak, R., Wuchterl, G., et al. 2010, A&A, 520, A93+

Hauschildt, P. H., Allard, F., & Baron, E. 1999, ApJ, 512, 377

Hawley, S. L., Allred, J. C., Johns-Krull, C. M., et al. 2003, ApJ, 597, 535

Helling, C., Ackerman, A., Allard, F., et al. 2008, MNRAS, 391, 1854

Heng, K., & Kopparla, P. 2012, ApJ, 754, 60

Henry, G. W., Howard, A. W., Marcy, G. W., Fischer, D. A., & Johnson, J. A. 2011, retracted article, arXiv:1109.2549

Henry, G. W., Marcy, G. W., Butler, R. P., & Vogt, S. S. 2000, ApJ, 529, L41

Henry, G. W., & Winn, J. N. 2008, AJ, 135, 68

Henry, T. J., Jao, W.-C., Subasavage, J. P., et al. 2006, AJ, 132, 2360

Hilbert, B., & McCullough, P. 2011, WFC3 ISR, 2011-10, `http://www.stsci.edu/hst/wfc3/`

Hogg, D. W., Bovy, J., & Lang, D. 2010, arXiv:1008.4686

Holland, H. D. 2006, Philos Trans R Soc Lond B Biol Sci, 361, 903

Holman, M. J., Winn, J. N., Latham, D. W., et al. 2006, ApJ, 652, 1715

Horne, K. 1986, PASP, 98, 609

Howard, A. W., Marcy, G. W., Bryson, S. T., et al. 2012, ApJS, 201, 15

Howard, A. W., Marcy, G. W., Johnson, J. A., et al. 2010, Science, 330, 653

Howe, A. R., & Burrows, A. S. 2012, ApJ, 756, 176

Hubbard, W. B., Fortney, J. J., Lunine, J. I., et al. 2001, ApJ, 560, 413

Huber, K. F., Czesla, S., Wolter, U., & Schmitt, J. H. M. M. 2010, A&A, 514, A39+

Huitson, C. M., Sing, D. K., Vidal-Madjar, A., et al. 2012, MNRAS, 422, 2477

Ida, S., & Lin, D. N. C. 2008, ApJ, 685, 584

Irwin, J., Berta, Z. K., Burke, C. J., et al. 2011a, ApJ, 727, 56

Irwin, J., Buchhave, L., Berta, Z. K., et al. 2010, ApJ, 718, 1353

Irwin, J., Charbonneau, D., Berta, Z. K., et al. 2009a, ApJ, 701, 1436

Irwin, J., Charbonneau, D., Nutzman, P., & Falco, E. 2009b, in IAU Symposium, Vol. 253, 37–43

Irwin, J., Irwin, M., Aigrain, S., et al. 2007, MNRAS, 375, 1449

Irwin, J. M., Quinn, S. N., Berta, Z. K., et al. 2011b, ApJ, 742, 123

Janson, M., Hormuth, F., Bergfors, C., et al. 2012, ApJ, 754, 44

Jehin, E., Gillon, M., Queloz, D., et al. 2011, The Messenger, 145, 2

Jenkins, J. M. 2002, ApJ, 575, 493

Jenkins, J. M., Caldwell, D. A., & Borucki, W. J. 2002, ApJ, 564, 495

Jenkins, J. M., Caldwell, D. A., Chandrasekaran, H., et al. 2010, ApJ, 713, L120

Jenkins, J. M., Doyle, L. R., & Cullers, D. K. 1996, Icarus, 119, 244

Johnson, D. R. H., & Soderblom, D. R. 1987, AJ, 93, 864

Johnson, J. A., Aller, K. M., Howard, A. W., & Crepp, J. R. 2010, PASP, 122, 905

Johnson, J. A., & Apps, K. 2009, ApJ, 699, 933

Johnson, J. A., Butler, R. P., Marcy, G. W., et al. 2007, ApJ, 670, 833

Johnson, J. A., Gazak, J. Z., Apps, K., et al. 2012, AJ, 143, 111

Joshi, M. 2003, Astrobiology, 3, 415

Joshi, M. M., Haberle, R. M., & Reynolds, R. T. 1997, Icarus, 129, 450

Kaltenegger, L., & Traub, W. A. 2009, ApJ, 698, 519

Kane, S. R., Mahadevan, S., von Braun, K., Laughlin, G., & Ciardi, D. R. 2009, PASP, 121, 1386

Kasting, J. F., Whitmire, D. P., & Reynolds, R. T. 1993, Icarus, 101, 108

Kennedy, G. M., & Kenyon, S. J. 2008, ApJ, 673, 502

Kim, D.-W., Protopapas, P., Alcock, C., Byun, Y.-I., & Bianco, F. B. 2009, MNRAS, 397, 558

Kim Quijano, J., et al. 2009, "WFC3 Mini-Data Handbook", Version 1.0, (Baltimore: STScI), `http://www.stsci.edu/hst/wfc3/`

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

—. 2010, MNRAS, 407, 301

—. 2011, MNRAS, 416, 689

Kipping, D. M., & Spiegel, D. S. 2011, MNRAS, 417, L88

Kiraga, M., & Stepien, K. 2007, Acta Astron., 57, 149

Knutson, H. A., Charbonneau, D., Allen, L. E., et al. 2007a, Nature, 447, 183

Knutson, H. A., Charbonneau, D., Cowan, N. B., et al. 2009a, ApJ, 690, 822

—. 2009b, ApJ, 703, 769

Knutson, H. A., Charbonneau, D., Noyes, R. W., et al. 2007b, ApJ, 655, 564

Knutson, H. A., Lewis, N., Fortney, J. J., et al. 2012, ApJ, 754, 22

Knutson, H. A., Madhusudhan, N., Cowan, N. B., et al. 2011, ApJ, 735, 27

Koch, D. G., Borucki, W. J., Basri, G., et al. 2010, ApJ, 713, L79

Kopparapu, R. K., Ramirez, R., Kasting, J. F., et al. 2013, ApJ, 765, 131

Kovács, G., Bakos, G., & Noyes, R. W. 2005, MNRAS, 356, 557

Kovacs, G., & Bakos, G. A. 2008, Communications in Asteroseismology, 157, 82

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

Kowalski, A. F., Hawley, S. L., Holtzman, J. A., Wisniewski, J. P., & Hilton, E. J. 2010, ApJ, 714, L98

Kretke, K. A., & Lin, D. N. C. 2012, ApJ, 755, 74

Kümmel, M., Kuntschner, H., Walsh, J. R., & Bushouse, H. 2011, ST-ECF ISR WFC3, 2011-01, `http://www.stsci.edu/hst/wfc3/`

Kümmel, M., Walsh, J. R., & Kuntschner, H. 2010, "aXe User Manual", Version 2.1, (Baltimore: STScI), `http://www.stsci.edu/hst/wfc3/`

Kümmel, M., Walsh, J. R., Pirzkal, N., Kuntschner, H., & Pasquali, A. 2009, PASP, 121, 59

Kundurthy, P., Agol, E., Becker, A. C., et al. 2011, ApJ, 731, 123

Kuntschner, H., Bushouse, H., Kümmel, M., & Walsh, J. R. 2009, ST-ECF ISR WFC3, 2009-17, `http://www.stsci.edu/hst/wfc3/`

Kuntschner, H., Bushouse, H., Walsh, J. R., & Kümmel, M. 2008, ST-ECF ISR WFC3, 2008-16, `http://www.stsci.edu/hst/wfc3/`

Kuntschner, H., Kümmel, M., Walsh, J. R., & Bushouse, H. 2011, ST-ECF ISR WFC3, 2011-05, `http://www.stsci.edu/hst/wfc3/`

Landolt, A. U. 1992, AJ, 104, 340

Lanza, A. F., Aigrain, S., Messina, S., et al. 2009, A&A, 506, 255

Lanza, A. F., Bonomo, A. S., Moutou, C., et al. 2010, A&A, 520, A53+

Latham, D. W., Bakos, G. Á., Torres, G., et al. 2009, ApJ, 704, 1107

Latham, D. W., Stefanik, R. P., Mazeh, T., Mayor, M., & Burki, G. 1989, Nature, 339, 38

Laughlin, G., Bodenheimer, P., & Adams, F. C. 2004, ApJ, 612, L73

Law, N. M., Kraus, A. L., Street, R., et al. 2012, ApJ, 757, 133

Lecavelier Des Etangs, A., Ehrenreich, D., Vidal-Madjar, A., et al. 2010, A&A, 514, A72+

Lecavelier Des Etangs, A., Pont, F., Vidal-Madjar, A., & Sing, D. 2008, A&A, 481, L83

Léger, A., Rouan, D., Schneider, J., et al. 2009, A&A, 506, 287

Leggett, S. K. 1992, ApJS, 82, 351

Leggett, S. K., Allard, F., Dahn, C., et al. 2000, ApJ, 535, 965

Lépine, S. 2005, AJ, 130, 1680

Lépine, S., & Shara, M. M. 2005, AJ, 129, 1483

Lewis, N. K., Showman, A. P., Fortney, J. J., et al. 2010, ApJ, 720, 344

Lissauer, J. J., Fabrycky, D. C., Ford, E. B., et al. 2011, Nature, 470, 53

Lodders, K., & Fegley, Jr., B. 2006, Chemistry of Low Mass Substellar Objects (Springer Verlag), 1

Loeb, A., & Maoz, D. 2013, MNRAS

Lomb, N. R. 1976, Ap&SS, 39, 447

Long, K. S., Baggett, S., Deustua, S., & Riess, A. 2010, WFC ISR, 2010-17, `http://www.stsci.edu/hst/wfc3/`

Long, K. S., Wheeler, T., & Bushouse, H. 2011, WFC ISR, 2011-09, `http://www.stsci.edu/hst/wfc3/`

Lopez, E. D., Fortney, J. J., & Miller, N. 2012, ApJ, 761, 59

López-Morales, M. 2007, ApJ, 660, 732

Madhusudhan, N., & Seager, S. 2009, ApJ, 707, 24

Mahadevan, S., Ramsey, L., Bender, C., et al. 2012, in SPIE Conference Series, Vol. 8446

Mamajek, E. E., & Hillenbrand, L. A. 2008, ApJ, 687, 1264

Mandel, K., & Agol, E. 2002, ApJ, 580, L171

Maness, H. L., Marcy, G. W., Ford, E. B., et al. 2007, PASP, 119, 90

Marcy, G. W., & Benitz, K. J. 1989, ApJ, 344, 441

Markwardt, C. B. 2009, in ASPC Series, Vol. 411, ADASS XVIII, ed. D. A. Bohlender, D. Durand, & P. Dowler, 251–+

Mayor, M., Marmier, M., Lovis, C., et al. 2011, arXiv:1109.2497, submitted to A&A

Mayor, M., Pepe, F., Queloz, D., et al. 2003, The Messenger, 114, 20

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

McCarthy, D. D. 1998, Journal for the History of Astronomy, 29, 327

McCullough, P. 2008, WFC3 ISR, 2008-26, `http://www.stsci.edu/hst/wfc3/`

McCullough, P., & Deustua, S. 2008, WFC3 ISR, 2008-33, `http://www.stsci.edu/hst/wfc3/`

McCullough, P. R., & MacKenty, J. W. 2011, WFC STAN 6, `http://www.stsci.edu/hst/wfc3/documents/newsletters/`

McCullough, P. R., Stys, J. E., Valenti, J. A., et al. 2005, PASP, 117, 783

—. 2006, ApJ, 648, 1228

McLaughlin, D. B. 1924, ApJ, 60, 22

Menou, K. 2012, ApJ, 744, L16

Miller, A. A., Irwin, J., Aigrain, S., Hodgkin, S., & Hebb, L. 2008, MNRAS, 387, 349

Miller-Ricci, E., & Fortney, J. J. 2010, ApJ, 716, L74

Miller-Ricci, E., Rowe, J. F., Sasselov, D., et al. 2008, ApJ, 682, 593

Miller-Ricci, E., Seager, S., & Sasselov, D. 2009, ApJ, 690, 1056

Miller-Ricci Kempton, E., & Rauscher, E. 2012, ApJ, 751, 117

Miller-Ricci Kempton, E., Zahnle, K., & Fortney, J. J. 2012, ApJ, 745, 3

Moehler, S., Freudling, W., Møller, P., et al. 2010, PASP, 122, 93

Montgomery, R., & Laughlin, G. 2009, Icarus, 202, 1

Morales, J. C., Ribas, I., Jordi, C., et al. 2009, ApJ, 691, 1400

Morley, C., Fortney, J. J., Kempton, E., Visscher, C. W., & Marley, M. S. 2013, in AAS Meeting Abstracts, Vol. 221, #126.05

Morton, T. D., & Johnson, J. A. 2011, ApJ, 738, 170

Moutou, C., Donati, J., Savalle, R., et al. 2007, A&A, 473, 651

Moutou, C., Pont, F., Barge, P., et al. 2005, A&A, 437, 355

Moutou, C., Pont, F., Bouchy, F., & Mayor, M. 2004, A&A, 424, L31

Muirhead, P. S., Hamren, K., Schlawin, E., et al. 2012a, ApJ, 750, L37

Muirhead, P. S., Johnson, J. A., Apps, K., et al. 2012b, ApJ, 747, 144

Murgas, F., Pallé, E., Cabrera-Lavers, A., et al. 2012, A&A, 544, A41

Murray-Clay, R. A., Chiang, E. I., & Murray, N. 2009, ApJ, 693, 23

Narita, N., Nagayama, T., Suenaga, T., et al. 2012, arXiv:1210.3169, accepted to PASJ

Naylor, T., Totten, E. J., Jeffries, R. D., et al. 2002, MNRAS, 335, 291

Nefs, S. V., Birkby, J. L., Snellen, I. A. G., et al. 2012, MNRAS, 3514

Nettelmann, N., Fortney, J. J., Kramm, U., & Redmer, R. 2011, ApJ, 733, 2

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

Nutzman, P., Gilliland, R. L., McCullough, P. R., et al. 2011a, ApJ, 726, 3

Nutzman, P. A., Fabrycky, D. C., & Fortney, J. J. 2011b, ApJ, 740, L10

Ofir, A., Alonso, R., Bonomo, A. S., et al. 2010, MNRAS, 404, L99

Ogihara, M., & Ida, S. 2009, ApJ, 699, 824

Pepper, J., & Gaudi, B. S. 2005, ApJ, 631, 581

Pepper, J., Gould, A., & Depoy, D. L. 2003, Acta Astronomica, 53, 213

Pickles, A., Rosing, W., Martinez, J., et al. 2012, in SPIE

Pirzkal, N., Mack, J., Dahlen, T., & Sabbi, E. 2011, WFC3 ISR, 2011-11, `http://www.stsci.edu/hst/wfc3/`

Poe, C. H., & Eaton, J. A. 1985, ApJ, 289, 644

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

Pont, F., Aigrain, S., & Zucker, S. 2011, MNRAS, 411, 1953

Pont, F., Gilliland, R. L., Moutou, C., et al. 2007, A&A, 476, 1347

Pont, F., Knutson, H., Gilliland, R. L., Moutou, C., & Charbonneau, D. 2008, MNRAS, 385, 109

Pont, F., Zucker, S., & Queloz, D. 2006, MNRAS, 373, 231

Press, W. H. 2002, Numerical recipes in C++ : the art of scientific computing (Cambridge University Press)

Queloz, D., Bouchy, F., Moutou, C., et al. 2009, A&A, 506, 303

Quirrenbach, A., Amado, P. J., Seifert, W., et al. 2012, in SPIE Conference Series, Vol. 8446

Rabus, M., Alonso, R., Belmonte, J. A., et al. 2009, A&A, 494, 391

Rauer, H., Gebauer, S., Paris, P. V., et al. 2011, A&A, 529, A8+

Raymond, S. N., Barnes, R., & Mandell, A. M. 2008, MNRAS, 384, 663

Raymond, S. N., Scalo, J., & Meadows, V. S. 2007, ApJ, 669, 606

Régulo, C., Almenara, J. M., Alonso, R., Deeg, H., & Roca Cortés, T. 2007, A&A, 467, 1345

Reiners, A. 2007, A&A, 467, 259

Reiners, A., & Basri, G. 2007, ApJ, 656, 1121

Reiners, A., Bean, J. L., Huber, K. F., et al. 2010, ApJ, 710, 432

Rivera, E. J., Laughlin, G., Butler, R. P., et al. 2010, ApJ, 719, 890

Rogers, L. A., Bodenheimer, P., Lissauer, J. J., & Seager, S. 2011, ApJ, 738, 59

Rogers, L. A., & Seager, S. 2010a, ApJ, 712, 974

—. 2010b, ApJ, 716, 1208

Rojas-Ayala, B., Covey, K. R., Muirhead, P. S., & Lloyd, J. P. 2010, ApJ, 720, L113

—. 2012, ApJ, 748, 93

Rossiter, R. A. 1924, ApJ, 60, 15

Sada, P. V., Deming, D., Jackson, B., et al. 2010, ApJ, 720, L215

Sagan, C., & Khare, B. N. 1979, Nature, 277, 102

Sanchis-Ojeda, R., Fabrycky, D. C., Winn, J. N., et al. 2012, Nature, 487, 449

Sanchis-Ojeda, R., Winn, J. N., Holman, M. J., et al. 2011, ApJ, 733, 127

Scalo, J., Kaltenegger, L., Segura, A. G., et al. 2007, Astrobiology, 7, 85

Scargle, J. D. 1982, ApJ, 263, 835

Schlaufman, K. C., & Laughlin, G. 2010, A&A, 519, A105

Schmidt, S. J., Kowalski, A. F., Hawley, S. L., et al. 2012, ApJ, 745, 14

Seager, S., & Deming, D. 2010, ARA&A, 48, 631

Seager, S., Kuchner, M., Hier-Majumder, C. A., & Militzer, B. 2007, ApJ, 669, 1279

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

Seager, S., & Sasselov, D. D. 2000, ApJ, 537, 916

Segura, A., Walkowicz, L. M., Meadows, V., Kasting, J., & Hawley, S. 2010, Astrobiology, 10, 751

Shkolnik, E., Liu, M. C., & Reid, I. N. 2009, ApJ, 699, 649

Shkolnik, E. L., Anglada-Escudé, G., Liu, M. C., et al. 2012, ApJ, 758, 56

Shkolnik, E. L., Hebb, L., Liu, M. C., Reid, I. N., & Collier Cameron, A. 2010, ApJ, 716, 1522

Sing, D. K., Désert, J.-M., Lecavelier Des Etangs, A., et al. 2009, A&A, 505, 891

Sing, D. K., Pont, F., Aigrain, S., et al. 2011, MNRAS, 416, 1443

Siverd, R. J., Beatty, T. G., Pepper, J., et al. 2012, ApJ, 761, 123

Sivia, D., & Skilling, J. 2006, Data Analysis: A Bayesian Tutorial, Oxford Science Publications (Oxford University Press)

Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163

Smith, J. C., Stumpe, M. C., Van Cleve, J. E., et al. 2012, PASP, 124, 1000

Smith, R. M., Zavodny, M., Rahmer, G., & Bonati, M. 2008a, in SPIE Conference Series, Vol. 7021

Smith, R. M., Zavodny, M., Rahmer, G., & Bonati, M. 2008b, in SPIE Conference Series, Vol. 7021

Snellen, I. A. G., de Kok, R. J., de Mooij, E. J. W., & Albrecht, S. 2010, Nature, 465, 1049

Sousa, S. G., Santos, N. C., Mayor, M., et al. 2008, A&A, 487, 373

Southworth, J. 2008, MNRAS, 386, 1644

Strassmeier, K. G. 2009, A&A Rev., 17, 251

Street, R. A., Horne, K., Lister, T. A., et al. 2003, MNRAS, 340, 1287

Stumpe, M. C., Smith, J. C., Van Cleve, J. E., et al. 2012, PASP, 124, 985

Sumi, T., Bennett, D. P., Bond, I. A., et al. 2010, ApJ, 710, 1641

Swain, M. R., Vasisht, G., & Tinetti, G. 2008, Nature, 452, 329

Swift, J. J., Johnson, J. A., Morton, T. D., et al. 2013, ApJ, 764, 105

Szentgyorgyi, A., Frebel, A., Furesz, G., et al. 2012, in SPIE Conference Series, Vol. 8446

Tamuz, O., Mazeh, T., & Zucker, S. 2005, MNRAS, 356, 1466

Tarter, J. C., Backus, P. R., Mancinelli, R. L., et al. 2007, Astrobiology, 7, 30

Tenenbaum, P., Christiansen, J. L., Jenkins, J. M., et al. 2012, ApJS, 199, 24

Teske, J. K., Turner, J. D., Mueller, M., & Griffith, C. A. 2013, MNRAS

Tingley, B. 2003, A&A, 408, L5

—. 2011, A&A, 529, A6

Tofflemire, B. M., Wisniewski, J. P., Kowalski, A. F., et al. 2012, AJ, 143, 12

Torres, G. 2007, ApJ, 671, L65

—. 2013, Astronomische Nachrichten, 334, 4

Torres, G., Andersen, J., & Giménez, A. 2010, A&A Rev., 18, 67

Torres, G., Fischer, D. A., Sozzetti, A., et al. 2012, ApJ, 757, 161

Triaud, A. H. M. J., Anderson, D. R., Collier Cameron, A., et al. 2013a, A&A, 551, A80

Triaud, A. H. M. J., Hebb, L., Anderson, D. R., et al. 2013b, A&A, 549, A18

Tyson, N. D., & Gal, R. R. 1993, AJ, 105, 1206

van Altena, W. F., Lee, J. T., & Hoffleit, E. D. 1995, The general catalogue of trigonometric [stellar] parallaxes (Yale University Observatory)

van Dokkum, P. G., & Conroy, C. 2010, Nature, 468, 940

van Eyken, J. C., Ciardi, D. R., von Braun, K., et al. 2012, ApJ, 755, 42

van Hamme, W. 1993, AJ, 106, 2096

Viana, A. C., & Baggett, S. 2010, WFC3 ISR, 2010-02, `http://www.stsci.edu/hst/wfc3/`

Vidal-Madjar, A., Lecavelier des Etangs, A., Désert, J.-M., et al. 2003, Nature, 422, 143

von Braun, K., Boyajian, T. S., ten Brummelaar, T. A., et al. 2011, ApJ, 740, 49

von Braun, K., Kane, S. R., & Ciardi, D. R. 2009, ApJ, 702, 779

von Paris, P., Cabrera, J., Godolt, M., et al. 2011, A&A, 534, A26

Walkowicz, L. M., Johns-Krull, C. M., & Hawley, S. L. 2008, ApJ, 677, 593

West, A. A., & Basri, G. 2009, ApJ, 693, 1283

West, A. A., Hawley, S. L., Bochanski, J. J., et al. 2008, AJ, 135, 785

Wheatley, P. J., Pollacco, D. L., Queloz, D., et al. 2013, arXiv:1302.6592

Winn, J. N. 2010, arXiv:1001.2010

Winn, J. N., Fabrycky, D., Albrecht, S., & Johnson, J. A. 2010, ApJ, 718, L145

Winn, J. N., Matthews, J. M., Dawson, R. I., et al. 2011, ApJ, 737, L18+

Wolfgang, A., & Laughlin, G. 2012, ApJ, 750, 148

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

Wolter, U., Schmitt, J. H. M. M., Huber, K. F., et al. 2009, A&A, 504, 561

Wright, J. T., Fakhouri, O., Marcy, G. W., et al. 2011, PASP, 123, 412

Wright, J. T., Marcy, G. W., Howard, A. W., et al. 2012, ApJ, 753, 160

Wright, J. T., Upadhyay, S., Marcy, G. W., et al. 2009, ApJ, 693, 1084

Yi, S., Demarque, P., Kim, Y.-C., et al. 2001, ApJS, 136, 417

Youdin, A. N. 2011, ApJ, 742, 38

Young, A. T. 1967, AJ, 72, 747