No Conclusive Evidence for Transits of Proxima b in *MOST* Photometry

David M. Kipping[1], Chris Cameron[2], Joel D. Hartman[3], James R. A. Davenport[4,18], Jaymie M. Matthews[5], Dimitar Sasselov[6], Jason Rowe[7], Robert J. Siverd[8], Jingjing Chen[1], Emily Sandford[1], Gáspár Á. Bakos[3,19,20], Andrés Jordán[9,10], Daniel Bayliss[11], Thomas Henning[12], Luigi Mancini[12], Kaloyan Penev[3], Zoltan Csubry[3], Waqas Bhatti[3], Joao Da Silva Bento[13], David B. Guenther[14], Rainer Kuschnig[15], Anthony F. J. Moffat[7], Slavek M. Rucinski[16], and Werner W. Weiss[17]

[1] Department of Astronomy, Columbia University, 550 W. 120th Street, New York, NY 10027, USA; dkipping@astro.columbia.edu
[2] Department of Mathematics, Physics & Geology, Cape Breton University, 1250 Grand Lake Road, Sydney, NS B1P 6L2, Canada
[3] Department of Astrophysical Sciences, 4 Ivy Lane, Princeton, NJ 08544, USA
[4] Department of Physics & Astronomy, Western Washington University, 516 High Street, Bellingham, WA 98225, USA
[5] Department of Physics & Astronomy, University of British Columbia, 6224 Agricultural Road, Vancouver, BC V6T 1Z1, Canada
[6] Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA
[7] Observatoire Astronomque du Mont Mégantic, Départment de Physique, Université de Montréal C. P. 6128, Succursale, Centre-Ville, Montréal, QC H3C 3J7, Canada
[8] Las Cumbres Observatory Global Telescope Network, 6740 Cortona Drive, Suite 102, Santa Barbara, CA 93117, USA
[9] Instituto de Astrofísica, Universidad Católica de Chile, Av. Vicuña Mackenna 4860, 7820436 Macul, Santiago, Chile
[10] Millennium Institute of Astrophysics, Av. Vicuña Mackenna 4860, 7820436, Santiago, Chile
[11] Observatoire Astronomique de Universite de Genéve, 51 ch. des Maillettes, 1290 Versoix, Switzerland
[12] Max Plank Institute for Astronomy, Königstuhl 17, D-69117 Heidelberg, Germany
[13] Research School of Astronomy & Astrophysics, Mount Stromlo Observatory, Australian National University, Cotter Road, Weston, ACT 2611, Australia
[14] Institute for Computational Astrophysics, Department of Astronomy and Physics, Saint Mary's University, Halifax, NS B3H 3C3, Canada
[15] Institut für Astronomie, Universität Wien Türkenschanzstrasse 17, A1180 Wien, Austria
[16] Department of Astronomy and Astrophysics, University of Toronto, Toronto, ON M5S 3H4, Canada
[17] Institute for Astrophysics, University of Vienna, Universitätsring 1, 1010 Vienna, Austria

Abstract

The analysis of Proxima Centauri's radial velocities recently led Anglada-Escudé et al. to claim the presence of a low-mass planet orbiting the Sun's nearest star once every 11.2 days. Although the a priori probability that Proxima b transits its parent star is just 1.5%, the potential impact of such a discovery would be considerable. Independent of recent radial velocity efforts, we observed Proxima Centauri for 12.5 days in 2014 and 31 days in 2015 with the *Microwave and Oscillations of Stars* space telescope. We report here that we cannot make a compelling case that Proxima b transits in our precise photometric time series. Imposing an informative prior on the period and phase, we do detect a candidate signal with the expected depth. However, perturbing the phase prior across 100 evenly spaced intervals reveals one strong false positive and one weaker instance. We estimate a false-positive rate of at least a few percent and a much higher false-negative rate of 20%–40%, likely caused by the very high flare rate of Proxima Centauri. Comparing our candidate signal to HATSouth ground-based photometry reveals that the signal is somewhat, but not conclusively, disfavored (1σ–2σ), leading us to argue that the signal is most likely spurious. We expect that infrared photometric follow-up could more conclusively test the existence of this candidate signal, owing to the suppression of flare activity and the impressive infrared brightness of the parent star.

Key words: planetary systems – stars: individual (Proxima Centauri) – techniques: photometric

Supporting material: machine-readable tables

1. Introduction

Proxima Centauri is the Sun's nearest stellar neighbor at a distance of 1.295 pc (van Leeuwen 2007). Despite this, Proxima's late spectral type (M5.5; Bessell 1991) makes it too faint to be seen by the naked eye ($V = 11.1$; Jao et al. 2014), elucidating why this is not the easiest target in the search for extrasolar planets. This challenge is exacerbated by the activity of Proxima itself, being a classic flare star (Christian et al. 2004).

Nevertheless, Proxima is one of the best-studied low-mass stars, and its diminutive mass offers an enhanced radial velocity semiamplitude, K, scaling as $M_\star^{-2/3}$. Accordingly, translating the same planet from the Sun to Proxima would cause K to

increase by a factor of four. Early radial velocity campaigns, such as those of Endl & Kürster (2008) and Zechmeister et al. (2009), found no signals at the few m s^{-1} level, ruling out super-Earths in the habitable zone.

The hunt for planets around our nearest star fell to the sidelines in the following years, notably during the era of NASA's *Kepler* mission. With thousands of planetary candidate detections pouring in (Batalha et al. 2013), the exoplanet community reasonably focused on these immediate discoveries. Although only a few thousand M dwarfs were observed by *Kepler* (out of ∼200,000 targets), the *Kepler* results ultimately rekindled our team's interest in the prospect of planets around Proxima.

First, with the discovery of a planetary system around one of *Kepler*'s lowest-mass stars, Kepler-42 (M5 dwarf), Muirhead et al. (2012) illustrated a putative template for what a potential planetary system around Proxima could resemble. Notably, the

---

planets were all sub-Earth sized and would thus have eluded the radial velocity search efforts of Endl & Kürster (2008) and Zechmeister et al. (2009), should similar planets orbit Proxima. Moreover, the planets were at extreme proximity to the star, with periods ranging from 0.45 to 1.86 days, leading to sizable geometric transit probabilities. Indeed, should Proxima harbor a Kepler-42-like system, the transit probability would be ∼10%.

Second, *Kepler* occurrence rate statistics showed that planets around early M dwarfs are very common, with an average of 2.5 ± 0.2 planets per star (Dressing & Charbonneau 2015). Radial velocity campaigns come to similar conclusions, finding evidence for at least one planet per M dwarf (Tuomi et al. 2014). Together then, this implies not only that Proxima has an excellent chance of harboring a planetary system, but also that such planets have a reasonable probability of transiting and producing millimagnitude-level signals. These arguments inspired our team to conduct a transit survey of Proxima starting in 2014 with the *Microwave and Oscillations of Stars* (*MOST*) telescope.

MOST is a 53 kg satellite in low Earth orbit with a 15 cm aperture visible band camera (35–750 nm). *MOST* is able to deliver millimagnitude-level photometry (for $V \lesssim 12$) at high cadence over several-week baselines, although observations are typically interrupted once per 101-minute orbit as the spacecraft passes behind Earth. *MOST* has been successful in discovering several new transiting systems, such as 55 Cnc e (Winn et al. 2011), HD 97658 b (Dragomir et al. 2013), and most recently HIP 116454 b (Vanderburg et al. 2015). For Proxima Centauri, we estimated that *MOST* should deliver ∼0.3 mmag precision photometry on an hour timescale, making it well suited for detecting the 4.2 mmag transit expected to be caused by an Earth-sized planet, and thus two seasons of observations were undertaken in 2014 and 2015.

Evidently, our team was not alone in returning to Proxima, with the Pale Red Dot (PRD) campaign conducting their own intensive search using radial velocities in 2016. By combining the PRD data with previous radial velocities, Anglada-Escudé et al. (2016) recently announced the detection of a 11.2-day planetary candidate, Proxima b. Since radial velocities do not reveal the inclination of the planetary orbit, only the minimum mass of Proxima b is currently known at $M_P \sin i = 1.27^{+0.19}_{-0.17} \ M_\oplus$. Since the transition from Terran (solid-like) to Neptunian worlds occurs at $(2.1 \pm 0.6) \ M_\oplus$ (Chen & Kipping 2016), the compositional nature of Proxima b is currently ambiguous. If transits of Proxima b were observed, the inclination could be resolved, as well as offering the opportunity to further characterize this remarkable world.

In this work, we present the results of our search for transiting planets around Proxima Centauri with *MOST* photometry. We describe the observations and data treatment stages in Section 2 and our photometric model in Section 3. In Section 4, we present the results of a localized search using the reported Proxima b ephemeris, followed by two sets of tests in Sections 5 and 6. Finally, we discuss the constraints our data place on Proxima b in Section 7.

2. Observations

2.1. MOST Observations

MOST observed Proxima Centauri in 2014 May (beginning on HJD(2000) 2,456,793.18) for about 12.5 days. Proxima Centauri falls outside of the continuous viewing zone of *MOST* ($-19°$ to $+36°$ in decl.; see Walker et al. 2003) and can only be observed for a fraction of the satellite's 101-minute orbit. For this reason, and for other science queue considerations, data were collected for about 30% of each *MOST* orbit and were sampled at an average rate of 63.4 s. *MOST* again observed Proxima Centauri in 2015 May (starting on HJD(2000) 2,457,148.54), this time for a total of 31 days with extended coverage to almost 50% of every *MOST* orbit. Data were again sampled at an average rate of about 63 s.

Flux measurements were extracted from each image using aperture photometry techniques outlined by Rowe et al. (2006). Background counts, interpixel correlations, and pointing drifts were accounted for by subtracting polynomials fitted through correlations between the measured target flux and those parameters. Removal of stray earthshine onto the CCD was done by folding the time series at the orbital period of *MOST* and subtracting a running mean through 30 orbital bins. Any remaining statistical outliers were removed, resulting in ∼2600 individual time series measurements from the 2014 data set and ∼13,000 data points from 2015.

The time series was then inspected for flare-like events using v1.3.11 of the flare-finding suite FBEYE from Davenport et al. (2014). The results of this exercise are discussed in detail in the accompanying paper of Davenport et al. (2016), but for the purposes of this work these points are removed in all subsequent analyses of the photometry. The locations of these events are highlighted in Figure 1.

2.2. Time-correlated Structure and Trends

After correcting the photometry and removing the flares, it is clear that our *MOST* data exhibit time-correlated structure in both seasons (see Figure 1 and Table 1). Long-term trends are pronounced in both seasons, with 2014 displaying a slow brightness increase along with a sinusoidal-like few-millimagnitude variation on the timescale of a week. In 2015, the structure appears more complex and exhibits a slow brightness decrease. These trends are not seen in any of the comparison stars, and thus we identify them as being astrophysical in nature.

The slow brightness trends may be associated with the claimed 83-day rotation period of Proxima Centauri (Benedict et al. 1998). The remaining, and quite pronounced, structure may be a result of frequent flaring (Davenport et al. 2016) and associated coronal mass ejections, as well as magnetic activity such as evolving spots, plages, and networks.

Interpreting the origin of the observed structure is beyond the scope of this work, for which this structure represents an impediment in our ability to detect putative transits of Proxima.

2.3. Gaussian Process (GP) Regression

In order to search for transits, the structure and trends present in our data require modeling. For reasons described in what follows, we elected to use GP regression to model out this structure. Here, one assumes that the data are distributed around the transit model as a multivariate Gaussian including off-diagonal elements within its covariance matrix, Σ. This eliminates the assumption of independent uncertainties and allows each point in the time series to have some degree of correlation with every other point. The log-likelihood function,
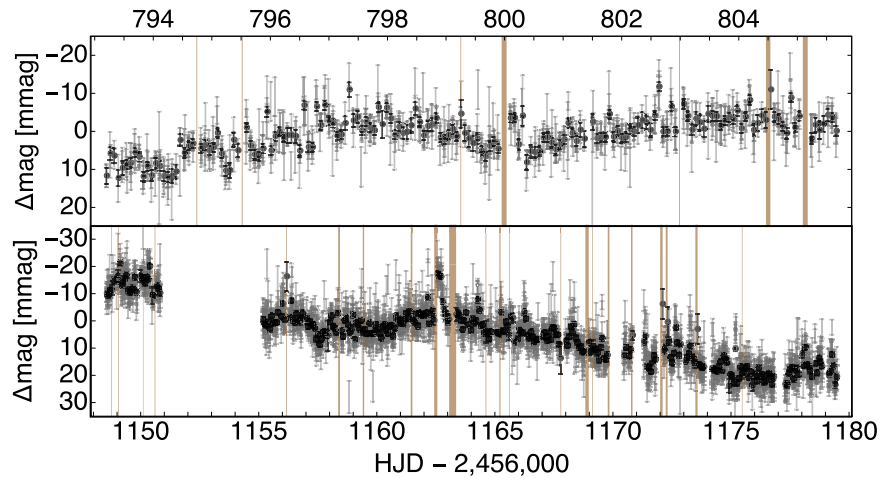
Figure 1. Corrected photometric observations of Proxima with *MOST* in 2014 (top) and 2015 (bottom) shown with 5× cadence binning (~5.3 minutes). Each orbital visit of *MOST* is binned together in black. The vertical colored regions denote regions ignored in this work, since flares were identified by Davenport et al. (2016).

Table 1
Reduced *MOST* Photometry Used in This Work,
Excluding Times Afflicted by Large Flares

HJD$_{\rm UTC}$−2,451,545	Δmag	Uncertainty
5248.197851471566	0.0118	0.0030
5248.200048126498	0.0116	0.0030
5248.259628134210	0.0028	0.0026
5248.262194952687	0.0092	0.0030
5248.270990357013	0.0065	0.0030
5248.329765768074	0.0072	0.0026
5248.332326375948	0.0054	0.0030
5248.399228360851	0.0155	0.0030

(This table is available in its entirety in machine-readable form.)

used for subsequent regression, may be written as

$$\log \mathcal{L} = -\tfrac{1}{2} \boldsymbol{r}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{r} - \tfrac{1}{2} \log \det \boldsymbol{\Sigma} - \tfrac{n}{2} \log 2\pi, \quad (1)$$

where $\boldsymbol{r}$ is a vector of the residuals between the transit model and the data. The very large number of covariance matrix elements are a priori unknown to us, but GPs model the covariance matrix with some assumed smooth, functional form, known as the kernel. The kernel, $\boldsymbol{K}$, is described by one or more hyperparameters, $\boldsymbol{\theta}_{\rm hyper}$, which are freely explored along with the usual model parameters, $\boldsymbol{\theta}_{\rm transit}$, during the fitting procedure.

GPs have emerged as one of the most popular and successful methods of modeling time-correlated noise in the analysis of transit photometry (Gibson et al. 2012; Berta-Thompson et al. 2015; Evans et al. 2015) and are appealing for their ability to model complex structure with relatively few new regression parameters. However, inverting the covariance matrix at each realization is computationally expensive, and typically GPs are computationally prohibitive for $n \gtrsim \mathcal{O}[10^3]$ data points, which is particularly relevant in this work given that we have over 10^4 photometric measurements.

2.4. Binning and Kernel Selection

To overcome the computational challenge of inverting the covariance matrix, one may first apply modest binning to the time series. Ideally, the relevant correlation timescale(s) should be significantly greater than the timescale used for binning, such that correlations are preserved.

The native cadence of our photometric measurements is 63.5 s, and after removing outliers and flares, we have 2461 data points in the 2014 season and 11,473 points in 2015. We first assume that the kernel parameters for each season are wholly independent, given the large change in time. We then focus on the null model of a transit-free case, where the data are solely described by an offset parameter, a, and the GP. Accordingly, the 2014 season can be treated independently of 2015.

We found that a total of 2461 data points was not a computationally prohibitive number of points for GP regression, which allows us to directly compare the inferred GP kernel parameters between the binned and unbinned data. We set a binning timescale equal to 317.5 s (equivalent to five consecutive cadences), or approximately 5 minutes, and employ temporal windows for the binning rather than N-point binning, due to the considerable number of data gaps present. For the GP kernel, we adopted the popular Matérn $3/2$ kernel given by

$$\boldsymbol{K}_{i,j}(\alpha, l) = \alpha^2 \left(1 + \frac{\sqrt{3}\, |t_i - t_j|}{l} \right) \exp\left(-\frac{\sqrt{3}\, |t_i - t_j|}{l} \right), \quad (2)$$

where l controls the timescale of correlations and α controls the magnitude. The full covariance matrix, $\boldsymbol{\Sigma}$, is the sum of the matrix $\boldsymbol{K}$ and a diagonal matrix of the square of the measurement uncertainties. We then regressed both versions of the 2014 time series using MULTINEST (Feroz & Hobson 2008; Feroz et al. 2009) and computed parameter posteriors. The agreement between the two is excellent, with $a = 0.79^{+0.59}_{-0.57}$ mmag from the unbinned data versus $a = 0.78^{+0.58}_{-0.58}$ mmag using the binned data. Similarly, the l correlation timescale is almost identical—$l = 130^{+16}_{-15}$ minutes in the unbinned data versus $l = 130^{+17}_{-15}$ minutes in the binned data. Further, this timescale is much greater than the 5-minute binning timescale adopted, ensuring that key correlations are not affected by the binning procedure.

For the 2015 data, we are unable to repeat this test given the much larger number of unbinned points. However, regressing
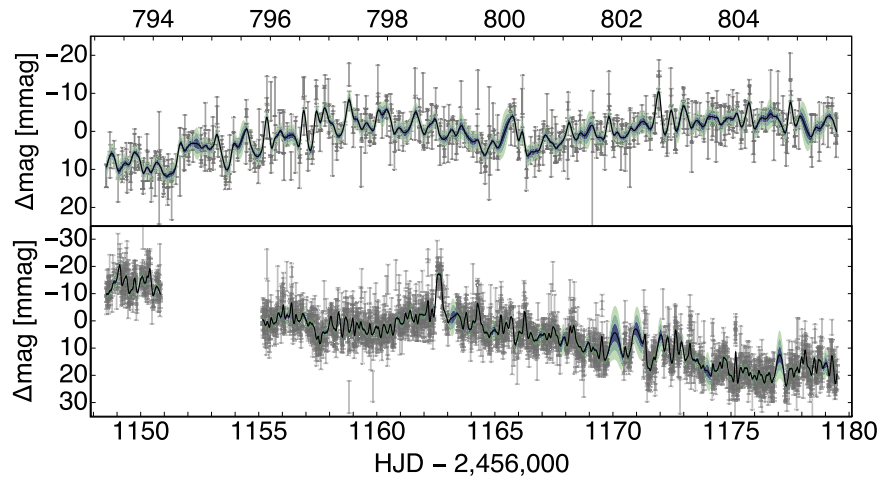
Figure 2. Corrected photometric observations of Proxima with *MOST* in 2014 (top) and 2015 (bottom) shown with 5× cadence binning (~5.3 minutes) with large flares excluded. The black line is the maximum likelihood fit of a GP using a Matérn 3/2 kernel, with the 0.5σ, 1.0σ, and 1.5σ confidence regions denoted by the gray regions (GP shown is from model $\mathcal{M}_2$).

Table 2
Comparison of Bayesian Evidences, $\mathcal{Z}$, for Four
Commonly Used Kernel Choices

Kernel	$\log \mathcal{Z}$
2014 unbinned	
Squared-exponential	9856.304 ± 0.088
Matérn 3/2	**9865.807 ± 0.083**
Matérn 5/2	9862.166 ± 0.085
Quasi-periodic Matérn 3/2	9865.513 ± 0.084
2014 binned	
Squared-exponential	2989.686 ± 0.088
Matérn 3/2	**2998.977 ± 0.084**
Matérn 5/2	2995.398 ± 0.085
Quasi-periodic Matérn 3/2	2998.795 ± 0.084
2015 binned	
Squared-exponential	11192.194 ± 0.109
Matérn 3/2	**11337.855 ± 0.100**
Matérn 5/2	11284.873 ± 0.102
Quasi-periodic Matérn 3/2	11337.101 ± 0.103

Note. The preferred model for each data set is shown in bold.

the same model and GP on the 2015 binned data reveals $l = 269^{+15}_{-15}$ minutes, which is again much greater than the binning timescale (in fact, even more so than before). Accordingly, we conclude that the 5-minute binning procedure does not affect the GP inference, nor should it affect our ability to detect transits, given that such events occur on significantly longer timescales too. Our final binned time series includes 709 points in the 2014 season and 2850 points in 2015, which are the points plotted in gray in Figure 1.

Although we assumed a Matérn 3/2 kernel in these tests, several other commonly used kernels are investigated before continuing. We compared the Bayesian evidence (or marginal likelihood) resulting from refitting both seasons of data for a Matérn 3/2 kernel, a Matérn 3/2 quasi-periodic kernel, a Matérn 5/2 kernel, and a squared-exponential kernel. After

conducting these fits on the 2014 unbinned, 2014 binned, and 2015 binned data, with identical priors, we find in all cases that the Matérn 3/2 kernel is favored, as shown in Table 2. We therefore adopt the Matérn 3/2 kernel in all subsequent photometric analysis of Proxima Centauri and show this favored GP overplotted with the data in Figure 2.

3. Priors, Models, and Tests

3.1. Predicting the Transit Ephemeris

The radial velocity solution of Anglada-Escudé et al. (2016) provides joint posterior distributions for numerous parameters, including the orbital period, P, and the mean anomaly at a reference time t_0, M_0. We first converted the M_0 column into a time of inferior conjunction, $t_{\rm IC}$, via Kepler's equation. For low-eccentricity orbits, such as that of Proxima b ($e < 0.3$ to 95% confidence; Anglada-Escudé et al. 2016), the time of transit minimum, τ, is equal to the time of inferior conjunction (Kipping 2011).

The time of inferior conjunction can be computed at any epoch of our choosing by adding on some integer number of periods. We therefore elected to calculate $t_{\rm IC}$ at every possible epoch from -200 to $+200$ orbital periods. For each realization, we computed the standard deviation of the resulting posterior and also the correlation with respect to the P posterior samples. We found that both are minimized for the -65th epoch, for which $t_{\rm IC} = 2,456,678.78 \pm 0.56$ HJD. Both this term and the orbital period of $P = (11.1856 \pm 0.0013)$ days are well approximated as two independent normal distributions.

3.2. Predicting the Radius

In order to guide our targeted search for transits of Proxima b, we first estimate the amplitude of the transit signal expected. If the mass of a planet is known, the radius can be predicted using an empirical mass–radius relation. In this work, we use the relation of Chen & Kipping (2016), which is probabilistic, includes freely inferred transitional regions, and was calibrated on the widest range of data available.

If the planet is transiting, this imposes the condition that $b < 1 + p$. Given that Proxima b is an Earth-mass planet, we expect $p \simeq 0.06$ and thus $p \ll 1$. This allows us to write that a

Table 3
Priors Used in the Targeted Transit Search of Proxima b, Spanning Four Different Models

Parameter	$\mathcal{M}_0$	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$
a_{2014} (mmag)	$\mathcal{U}[-10^{-2}, 10^{-2}]$	$\mathcal{U}[-10^{-2}, 10^{-2}]$	$\mathcal{U}[-10^{-2}, 10^{-2}]$	$\mathcal{U}[-10^{-2}, 10^{-2}]$
α_{2014}	$\mathcal{J}[-10^{-1}, 10^{1}]$	$\mathcal{J}[-10^{-1}, 10^{1}]$	$\mathcal{J}[-10^{-1}, 10^{1}]$	$\mathcal{J}[-10^{-1}, 10^{1}]$
l_{2014}	$\mathcal{J}[-10^{-2}, 10^{0}]$	$\mathcal{J}[-10^{-2}, 10^{0}]$	$\mathcal{J}[-10^{-2}, 10^{0}]$	$\mathcal{J}[-10^{-2}, 10^{0}]$
a_{2015} (mmag)	$\mathcal{U}[-10^{-2}, 10^{-2}]$	$\mathcal{U}[-10^{-2}, 10^{-2}]$	$\mathcal{U}[-10^{-2}, 10^{-2}]$	$\mathcal{U}[-10^{-2}, 10^{-2}]$
α_{2015}	$\mathcal{J}[-10^{-1}, 10^{1}]$	$\mathcal{J}[-10^{-1}, 10^{1}]$	$\mathcal{J}[-10^{-1}, 10^{1}]$	$\mathcal{J}[-10^{-1}, 10^{1}]$
l_{2015}	$\mathcal{J}[-10^{-2}, 10^{0}]$	$\mathcal{J}[-10^{-2}, 10^{0}]$	$\mathcal{J}[-10^{-2}, 10^{0}]$	$\mathcal{J}[-10^{-2}, 10^{0}]$
$p = (R_P/R_\star)$	⋯	$\mathcal{J}[10^{-3}, 10^{-0.70}]$	$\mathcal{J}[10^{-3}, 10^{-0.70}]$	$\mathcal{X}[0.069, 0.0050]$
b	⋯	$\mathcal{U}[0, 1.2]$	$\mathcal{U}[0, 1.2]$	$\mathcal{U}[0, 1.2]$
τ (HJD $-2{,}456{,}000$)	⋯	$\mathcal{W}[t_{ref}, t_{ref} + P]$	$\mathcal{N}[678.78, 0.59]$	$\mathcal{N}[678.78, 0.59]$
P (days)	⋯	$\mathcal{N}[11.1856, 0.0013]$	$\mathcal{N}[11.1856, 0.0013]$	$\mathcal{N}[11.1856, 0.0013]$
$\rho_\star$ (kg m^{-3})	⋯	$\delta[10^{4.792}]$	$\delta[10^{4.792}]$	$\delta[10^{4.792}]$
e	⋯	$\delta[0]$	$\delta[0]$	$\delta[0]$
$\log \mathcal{Z} - 14{,}300$	36.763 ± 0.065	46.286 ± 0.089	38.290 ± 0.079	41.336 ± 0.077

Note. $\mathcal{U}$ denotes a uniform prior, $\mathcal{W}$ a wraparound uniform, $\mathcal{J}$ a log-uniform, δ a delta function prior, and $\mathcal{X}$ a logistical distribution prior. We set the reference time to $t_{ref} = 2{,}457{,}165.385$ HJD.

transiting Proxima b must satisfy

$$\cos i < \frac{1}{a_R} \frac{1 - e^2}{1 + e \sin \omega}. \tag{3}$$

Using the joint posterior distribution samples for P, e, and ω from Anglada-Escudé et al. (2016), this condition requires $i < 89°.1$ to 95.45% confidence. For any transiting planet then, the $\sin i$ effect on the true mass is much smaller than the present measurement uncertainty on $M_P \sin i$. Accordingly, we may simply adopt $M_P \simeq M_P \sin i$ in estimating the radius of a transiting Proxima b.

We now use the posterior samples of $M_P \sin i$ to estimate a probabilistic range for R_P using the Forecaster code of Chen & Kipping (2016). This estimate accounts for the measurement uncertainty of Anglada-Escudé et al. (2016), the measurement uncertainties in the calibration of Chen & Kipping (2016), and the intrinsic dispersion in radii observed as a function of mass (Chen & Kipping 2016). Under the assumption that Proxima b is transiting, we estimate that $R_P = 1.06^{+0.13}_{-0.11}\ R_\oplus$. Normalizing by the radius of the star (Demory et al. 2009), we predict $p = R_P/R_\star = 0.0693^{+0.0095}_{-0.0083}$, which is well fit by a logistic distribution with shape parameters $\mu = 0.069$ and $s = 0.0051$. We also find that 99% of the posterior samples for p satisfy $p < 0.1$, and we conservatively double this limit to $p < 0.2$ as a truncation point to our prior. This also provides a cutoff for the impact parameter of $b < 1.2$.

3.3. Models Considered and Associated Priors

We considered three different transit models in our targeted search for transits of Proxima b, where we varied the degree of prior information we used from the Anglada-Escudé et al. (2016) discovery. We label the models as $\mathcal{M}_1$, $\mathcal{M}_2$, and $\mathcal{M}_3$, where the subscript increases with increased use of prior information. These fits may be compared directly to the null fit of $\mathcal{M}_0$, where no transit is included and only the GP hyperparameters, θ_{hyper}, are fitted.

We list the priors used in Table 3, where it can be seen that the GP hyperpriors are identical in all fits. This allows us to compare the Bayesian evidences between each model to aid model selection. All three transit models used the normal prior on orbital period; otherwise, the search would be blind rather than targeted. $\mathcal{M}_2$ uses an informative prior on the time of transit minimum, unlike $\mathcal{M}_1$. The final model, $\mathcal{M}_3$, also uses these period and phase informative priors plus an informative prior on the radius of the planet, as computed earlier in Section 3.2.

Limb-darkened transits are generated using the Mandel & Agol (2002) algorithm. Two quadratic limb-darkening coefficients are kept fixed at $u_1 = 0.7948$ and $u_2 = 0.0825$, estimated by finding nearest-neighbor interpolation of the PHOENIX model grids for *MOST* generated in Claret et al. (2014) (using $\log g = 5.25$ and $T_{eff} = 3050$ K). Eccentricity is kept fixed at zero, since Proxima b has a low eccentricity (Anglada-Escudé et al. 2016). Similarly, we fix the mean stellar density of the star, which is very well constrained given that Proxima is one of the most well-studied M dwarfs. These fixed parameters significantly reduce the number of parameters to explore, making the calculation of the Bayesian evidence of models using GP with several thousand data points computationally feasible. By fixing these terms, transit parameter inferences may slightly underestimate the true uncertainties, but since our primary objective is signal detection, the ability to feasibly compute evidences outweighs this cost, in our view.

3.4. Mis-specified Likelihood Function

As discussed earlier, the likelihood function used in this work is that of a GP, described in Section 2.3 and given by Equation 1. Additionally, we are computing evidences using MULTINEST, in order to conduct model comparison. Note that the computational expense of this work prohibited us from computing evidences with several different methods, and thus we adopt those from MULTINEST only in what follows.

A common perception of GPs is that they are extremely flexible models, seemingly able to model out just about any observed correlated noise structure, particularly when one regresses the GP kernel hyperparameters simultaneously with the model. Indeed, Feng et al. (2016) go as far as to actively caution against using GPs since they lead to frequently missing true signals, due their overzealous ability to fit out time series structure. This logic suggests that a GP-only model ($\mathcal{M}_0$) would generally be favored over a GP+transit model ($\mathcal{M}_2$ and
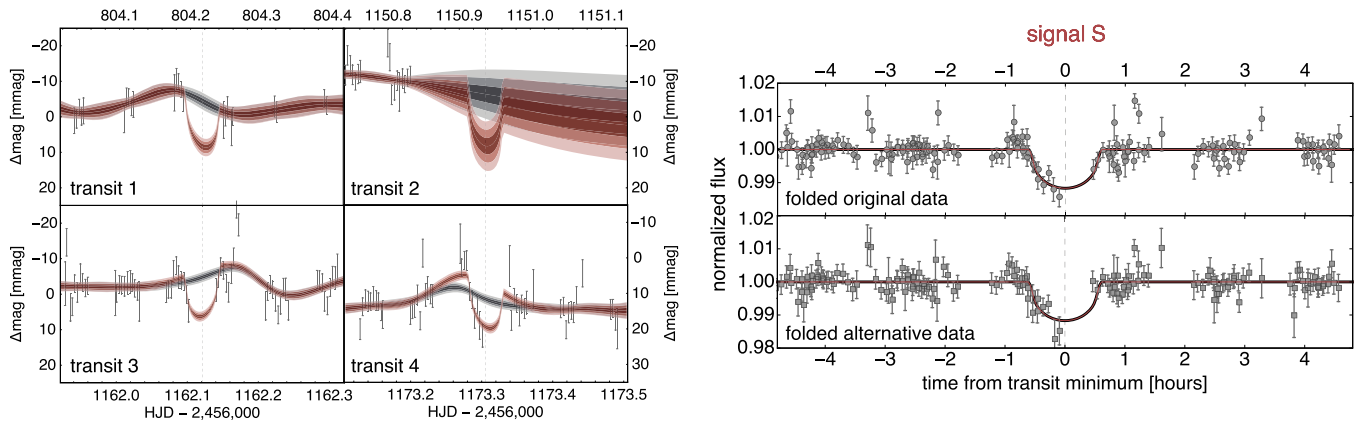
Figure 3. Left: zoom-in of the Proxima b transits for the spurious signal (signal S) from models $\mathcal{M}_1$. The GP+transit model 0.5σ, 1.0σ, and 1.5σ confidence regions are shown by the colored regions, whereas the GP-only (model $\mathcal{M}_0$) is depicted in gray. Right: phase-folded light curve of the transit signal S after removing the GP component of model $\mathcal{M}_1$ using the nominal *MOST* data (top) and an independent reprocessing (bottom).

$\mathcal{M}_3$) even when real signals are present. Or, equivalently, it implies that our evidences may be conservative and the actual weight of evidence for the transit models may be higher than that formally calculated.

This logic can be flawed if our likelihood function is mis-specified, which means that the marginal likelihood would be inaccurate. This can occur if the assumed functional form of the GP kernel is a poor approximation of the true (and unknown) covariance matrix. In this study, the high flare activity of Proxima makes this a plausible scenario. By extrapolating the rates of large flares, Davenport et al. (2016) estimate that Proxima exhibits a 0.5% brightness increase once every $\sim$20 minutes (on average). Note that this issue is not limited to just *MOST* data but implies that any visible light photometry of Proxima will be affected by ostensibly constant stochastic deviations at the level of 5 mmag, as a result of flares. A Matérn 3/2 kernel is not designed to describe a superposition of flare events, and thus formally we expect our likelihood function to be mis-specified for this reason.

In conclusion, the marginal likelihoods from our fits will not be accurate. It is not clear what alternative kernel or likelihood function could deal with this kind of noise structure either, and thus we still favor GPs over any alternatives. Although our evidences will be inaccurate, they may still be useful in guiding which models are preferred. Even if the Bayes factor is inaccurate, this does not mean it cannot be used to rank models in order of preference, since, after all, the majority of the residuals are indeed normally distributed. However, any candidate solutions from this process must be treated with great caution and subject to higher scrutiny and skepticism than usual.

4. Results

4.1. Signal S

We first discuss the results of model $\mathcal{M}_1$, where the transit phase is described by an uninformative prior. The marginal likelihood indicates a strong preference for $\mathcal{M}_1$ over $\mathcal{M}_0$, with $\Delta \log \mathcal{Z} = 9.52 \pm 0.11$. Hereafter, we refer to this solution as signal S.

The $\mathcal{M}_1$ ephemeris yields four transit epochs within our *MOST* time series, although one of these occurs during a data gap, as shown in Figure 3. We note that signal S is primarily driven by a

large feature in the fourth epoch, at HJD 2,457,173.3. To ensure that the recovered signal was not an artifact of our data processing method, one of us (J. Rowe) reprocessed the data independently. As shown by the squares in Figure 3, the signal appears coherent in both data products.

The time of transit minimum, τ, has a non-Gaussian but narrow marginal posterior with a 1σ credible interval of $\tau = 1150.9348^{+0.0019}_{-0.0026}$ (HJD $-2,456,000$), which deviates substantially from the predicted time based on the radial velocity fits of Anglada-Escudé et al. (2016). For reference, all of the model parameter credible intervals, from all four models, are listed in Table 4.

To quantify the ephemeris disagreement, we used the posterior samples of the radial-velocity-predicted time of transit minimum, computed earlier in Section 3.1, and propagated the joint posterior of P and τ to the equivalent epoch, giving $\tau = 1148.59^{+0.59}_{-0.59}$ (HJD $-2,456,000$). The p-value of signal S's τ posterior exceeds 4σ and is difficult to reconcile with the radial velocity solution. Note that the radial velocity posteriors include a floating eccentricity, and thus this effect is accounted for here. This point is formally established in the Bayesian framework by the fact that models $\mathcal{M}_2$ and $\mathcal{M}_3$ do not recover signal S when using the radial-velocity-derived τ prior.

We also note that the inferred planetary radius is at the 2σ upper limit of the Forecaster prediction, which, although not concerning in isolation, does compound upon these earlier concerns. Further, visual inspection of the transits (Figure 3) shows that the signal is primarily driven by an apparent flux increase around the times of transit, rather than a flux decrease, which raises additional skepticism.

While one could, in principle, refit the radial velocities imposing this transit phase as a prior, that model would be implicitly assuming that signal S is real—an assumption that is not warranted given the challenging noise structure of our data set and the arguments made above. The incompatibility of the transit phase, and to a lesser degree the poor phase coverage and inflated radius, leads us to conclude that signal S is unlikely associated with Proxima b and is either spurious, as a result of flare-induced likelihood mis-specification, or potentially an additional transiting planet, driven by a single event within our data.

Table 4
A Posteriori Median and 68.3% Credible Intervals of Each Model Parameter for the Four Models Regressed to the *MOST* Photometry

Parameter	$\mathcal{M}_0$	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$
a_{2014} (mmag)	$0.78_{-0.59}^{-0.59}$	$0.78_{-0.54}^{+0.55}$	$0.77_{-0.57}^{+0.57}$	$0.78_{-0.58}^{+0.58}$
α_{2014}	$1.44_{-0.10}^{+0.12}$	$1.44_{-0.09}^{+0.10}$	$1.44_{-0.10}^{+0.11}$	$1.45_{-0.10}^{+0.11}$
l_{2014} (minutes)	131_{-15}^{+16}	131_{-14}^{+15}	130_{-14}^{+15}	130_{-15}^{+16}
a_{2015} (mmag)	$5.41_{-0.87}^{+0.87}$	$5.38_{-0.80}^{+0.82}$	$5.40_{-0.85}^{+0.86}$	$5.40_{-0.89}^{+0.89}$
α_{2015}	$2.30_{-0.10}^{+0.11}$	$2.30_{-0.09}^{+0.11}$	$2.34_{-0.10}^{+0.11}$	$2.34_{-0.10}^{+0.11}$
l_{2015} (minutes)	269_{-14}^{+16}	268_{-13}^{+15}	274_{-14}^{+16}	274_{-15}^{+16}
R_P $(R_\oplus)$[a]	…	$1.38_{-0.12}^{+0.11}$	$1.23_{-0.15}^{+0.13}$	$1.14_{-0.10}^{+0.10}$
b	…	$0.22_{-0.14}^{+0.19}$	$0.28_{-0.19}^{+0.24}$	$0.25_{-0.18}^{+0.24}$
τ (HJD −2,456,000)	…	$983.1656_{-0.0330}^{+0.0064}$	$980.0554_{-0.0023}^{+0.0027}$	$980.0552_{-0.0026}^{+0.0029}$
P (days)	…	$11.18467_{-0.00039}^{+0.00200}$	$11.18725_{-0.00016}^{+0.00012}$	$11.18723_{-0.00019}^{+0.00014}$

Note.
[a] Assuming a fixed stellar radius of $0.123\ R_\odot$.
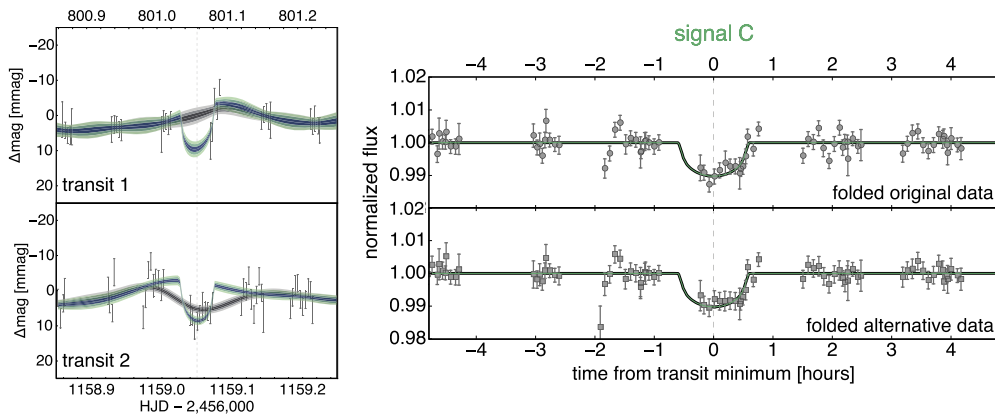
Figure 4. Same as Figure 3, except for the candidate transit signal of Proxima b, signal C. Right panels reproduce the left panels, except including the independently processed *MOST* data in lilac for comparison.

4.2. Signal C

We next consider the results of models $\mathcal{M}_2$ and $\mathcal{M}_3$, both of which recover the same transit signal, which we hereafter refer to as signal C. Three modes are recovered by the fit of $\mathcal{M}_2$, but the dominant mode is strongly favored at a Bayes factor of 20.1. Of the three modes, only the dominant is preferred over the null model of $\mathcal{M}_0$, and it is this mode that is compatible with the signal recovered by model $\mathcal{M}_3$. We therefore ignore the other two modes in what follows.

While the two models recover the same signal, $\mathcal{M}_2$ is favored over the null hypothesis with a Bayes factor of 4.6, whereas $\mathcal{M}_3$ is much stronger at 96.8. This can be understood by the fact the two models recover very similar p posteriors ($R_P(\mathcal{M}_2) = 1.23_{-0.15}^{+0.13}\ R_\oplus$ versus $R_P(\mathcal{M}_2) = 1.14_{-0.10}^{+0.10}\ R_\oplus$) but $\mathcal{M}_2$ used a broad, uninformative prior, over which the average likelihood is lower, thus penalizing the model for effectively being more complicated.

The signal is shown in Figure 4, where we highlight how once again the independent reprocessing of the *MOST* data produces a consistent signal.

The phase from both models is not incompatible with the radial velocity constraints, giving a p-value of 1.56σ, although this is to be expected since it was imposed as an informative prior. A more useful test is that the freely fitted radius from model $\mathcal{M}_2$ is compatible with the radius prediction from `Forecaster`.

We also note that the impact parameter of the signal is nongrazing (see Figure 5). This is important because observational bias of the transit method, given the `Forecaster` size prediction, makes it less likely that a detected signal would be caught on the limb. Integrating the conditional probability distribution of Kipping & Sandford (2016), we are able to estimate that it is in fact 35 times more likely that a real signal would be nongrazing than grazing.

In conclusion, analysis of signal C shows it to be compatible with that expected if Proxima b were observed to transit. While promising, this in itself does not prove that the signal is genuinely the transit of Proxima b, however, for reasons discussed earlier in Section 3.4.

5. Type I and II Error Rates

5.1. Evaluating the False-positive (Type I) Rate

In Section 3.4, it was established that the Bayesian evidence may not be a fully reliable tool for model selection in our case. Accordingly, we seek alternative methods to interpret the significance of signal C. While cross-validation would be a powerful alternative, Proxima b's period means that only two transits of signal C occur in our *MOST* photometry, and ignoring some fraction of the data is undesirable. Instead, we elected to perform a bootstrapping procedure to emulate our detection approach in the presence/absence of an injected signal.
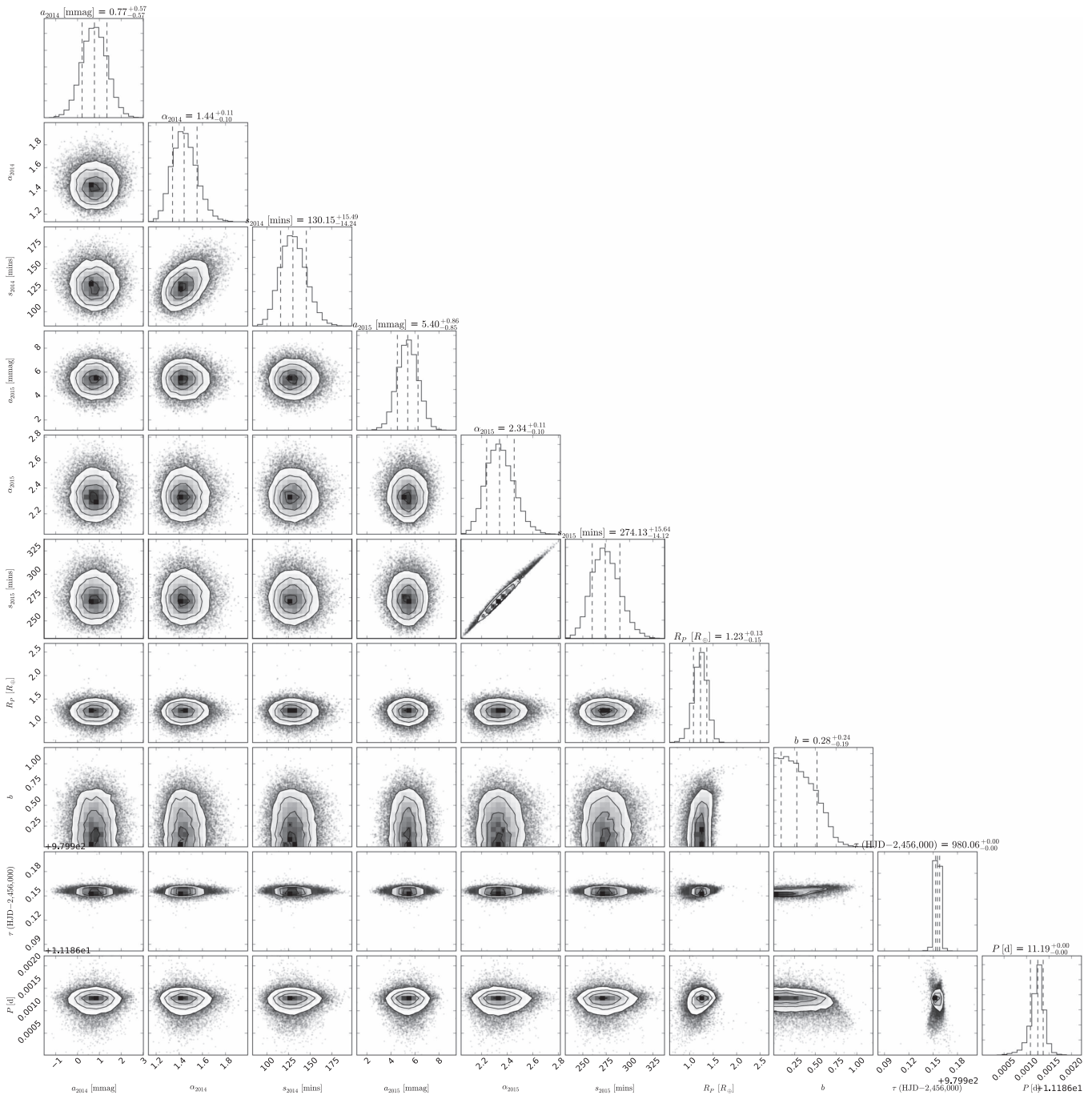
Figure 5. Corner plot of the joint posterior distribution of the fitted parameters retrieved for model $\mathcal{M}_2$. The parameters are consistent with those returned by $\mathcal{M}_3$.

We first consider the case of type I errors, which is the most critical term in assessing the credibility of any signals inferred by our approach. We specifically considered evaluating the type I error of model $\mathcal{M}_2$, which uses an informatively priored ephemeris but uninformative priors on the radius of Proxima b (see Table 3). To do this, we need a set of representative, synthetic fits of null data.

Since the GP model is argued to not represent a complete noise model (see Section 3.4), we cannot use the GP to generate synthetic, representative data sets. We also cannot randomly scramble the original data to create fake data, as this would remove the time-correlated noise structure clearly seen in our data. Performing a search for inverted transits is also not

useful, since flares are asymmetric flux increases mimicking such events. The best option remaining is to move along the data in a rolling-window style.

Accordingly, we use the original, unmodified *MOST* time series and simply modify the priors used. Specifically, in 100 fits, we iteratively translate the prior on τ by $0.01P$ until we loop back around to the original phase in the 100th trial. The disadvantage of this approach is that we cannot ensure that the null data are actually absent of signal (in fact, we already have established the presence of a spurious signal in the form of signal S; see Section 4.1). Consequently, our false-positive rate estimate may be an overestimate if latent but genuine transit signals reside in our *MOST* photometry.
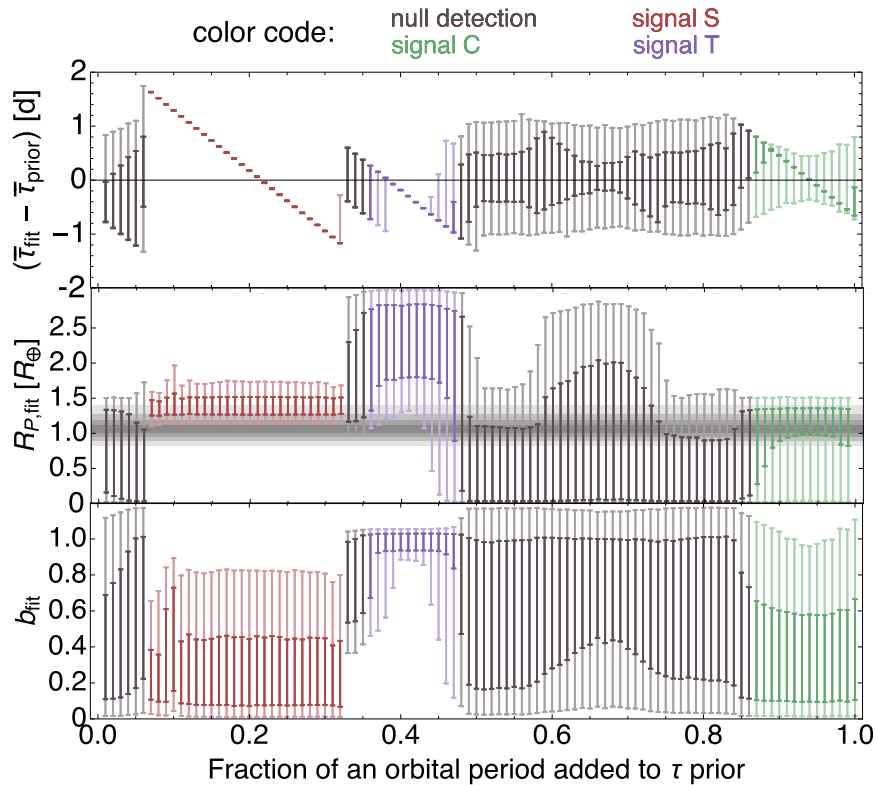
Figure 6. The 1σ (dark bars) and 2σ (light bars) credible intervals from 100 model $\mathcal{M}_2$ fits on the original *MOST* data, but slightly perturbing the informative prior on τ (*x*-axis). The color-code key is at the top. The horizontal region in the second panel denotes the `Forecaster` prediction for the planetary radius (Chen & Kipping 2016).

In each fit, we rerun an identical fit as before, using MULTINEST to explore the transit parameters and GP hyperparameters using otherwise identical priors. As before, we also compute the marginal likelihood. These tests, as well as the others needed to evaluate the type II error rate, demanded significant computational resources of tens of thousands of core hours on the NASA PLEIADES supercomputer, and this is why we are practically limited to only running 100 such tests.

In order to calculate the false-positive rate, we need to define what constitutes a "detection." A useful metric is to inspect the convergence of the time of transit minimum posterior. Detections will have a narrow posterior with most of the density located on a single mode. In contrast, the τ posterior of null detections will broadly reproduce the prior or display a broad, unconverged and structured form.

To more explicitly define what we mean by a "narrow" posterior, we demand that the range of the central 50% quantile, thereby constituting the majority of the samples, is less than one-half of the characteristic transit duration, $\mathbb{T}$. Assuming a circular orbit, adopting a representative impact parameter[21] of $b = 0.5$, and using the mode of our stellar density prior (see Table 3) give a characteristic transit duration of $\mathbb{T} = 65.3$ minutes. Further, we require that the Bayes factor between the fit and the null model ($\mathcal{M}_0$) exceeds e. We therefore require the following:

(i) The interval from the 25% quantile to 75% quantile of the τ posterior should be less than one-half of the characteristic transit duration, $\mathbb{T}/2$.

(ii) The evidence ratio should satisfy $\Delta \log \mathcal{Z} > 1$.

Turning to the results, we first note that from inspecting the τ posteriors of the 100 fits ran, it was immediately obvious that a considerable fraction of the fits recovered the signals S and C discussed in Sections 4.1 and 4.2. This can be seen from the top panel of Figure 6, where the τ posteriors latch onto a single solution and exhibit a linear-like trend on signals S and C. Since the τ prior (*x*-axis) is shifted each time but the best-fitting solution is the same, this creates the linear structure observed.

As a result of this behavior, it is necessary to establish a criterion to identify fits that recovered previously recognized signals, namely, signals S and C. We define such fits as those for which the median posterior τ sample lies within ± 0.5 transit durations of signal S/C's median posterior τ sample. Note that here it is unnecessary to use the characteristic duration, but instead we can use the actual duration measured from our earlier model fits. We thus define spurious signals as those satisfying the following criterion as well as criteria (i) and (ii) (to remove unconverged cases):

(iii) The median posterior τ sample of the fit should be within ± 0.5 transit durations of the median posterior τ sample of either signal S or signal C.

Using this criterion, we found that 26 of the 100 fits converged to signal S and another 14 converged to signal C. Inspection of Figure 6 reveals that a third signal appears to exist in the data, located around a phase of 0.4, causing 12 additional tests to converge to the same solution. We label this new signal as signal T. We show the credible intervals of the fitted parameters τ, p (converted to planetary radii), and b in Figure 6, with each realization color-coded to the aforementioned

---

[21] We use the median of the probability distribution of b after accounting for observational bias derived in Kipping & Sandford (2016).
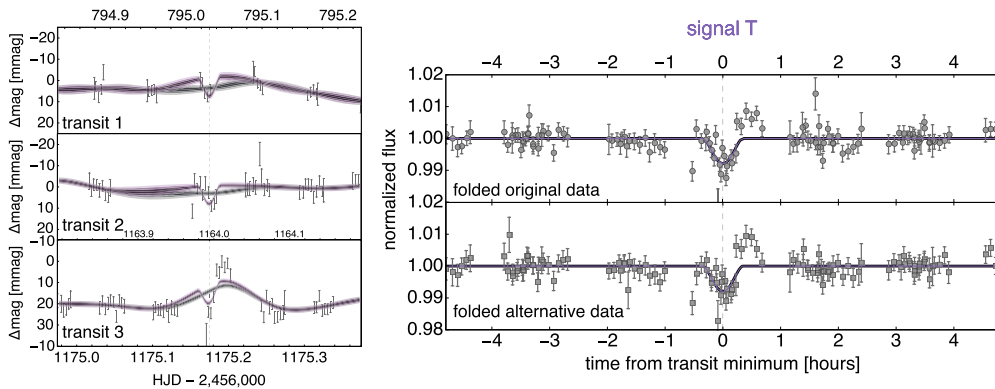
Figure 7. Same as Figure 3, except for the third transit signal found, signal T. The signal is far less convincing than signals S and C.

identities. We also plot the maximum a posteriori solution for signal T in Figure 7.

Signal T immediately raises skepticism about its validity. Over 95% of the posterior trials correspond to a grazing geometry, as evident from the shape of signal T in Figure 7. As discussed earlier in Section 4.2, observational biases mean that it is 35 times more likely that a detected transit would be nongrazing than grazing. Moreover, it is generally easier for sharp data artifacts to mimic a V-shaped event than a full transit morphology, by virtue of the former's simpler shape. The ratio-of-radii posterior is pushed up against the upper bounds of the prior, favoring a planet of $2.35^{+0.48}_{-0.55}\,R_\oplus$, which is highly incompatible with the `Forecaster` prediction. Finally, inspection of the data themselves reveals a far less convincing signal than signals C and T. We assert that signal T would never be genuinely considered a candidate transit signal of Proxima b, even if the phase had matched with the radial velocity prediction.

Because signal T would not be considered a detection if its phase had been compatible, we do not consider it to be a false-positive signal in what follows and simply discount it from the false-positive evaluations. In contrast, signal S shows no features that would have caused us to dismiss it as a false positive, had it landed at the correct phase. Therefore, of the 74 realizations not affected by signals C and T, 26 converge to a single spurious solution. Counting these cases as unique false positives would imply a false-positive rate of FPR = $(35 \pm 7)\%$, whereas counting them as belonging to a single false positive would give FPR = $(2 \pm 2)\%$. In conclusion, it is unclear precisely how to define the false-positive rate from these tests, but certainly the false-positive rate is nonzero and at least a few percent.

5.2. Evaluating the False-negative (Type II) Rate

To evaluate the false-negative rate, we used the same setup as for the type I tests, except we inject a $1.06\,R_\oplus$ planet (see Section 3.2) with $b = 0.5$ into the time series at each phase point. In each of the 100 tests we ran, the model is seeking an injected Proxima-b-like transit signal located within the specified phase prior. Since the data have been perturbed, it was necessary to rerun the null model, $\mathcal{M}_0$, on each of these synthetic data sets, in addition to model $\mathcal{M}_2$.

We classify null detections (i.e., the false negatives) as being any case for which criteria (i) and (ii) are not both satisfied,

which occurs for 23 of our 100 simulations. This sets a minimum limit on the false-negative rate of FNR = $(23 \pm 5)\%$.

We classify successful recoveries as cases where criteria (i) and (ii) are both satisfied and the median of the fit's τ posterior is less than one-half of the injected transit duration from the injected transit time. However, we exclude cases where signal C, S, or T is recovered using criterion (iii) (and extended now to include signal T). Defined in this way, we count 40 successful recoveries.

In addition to 21 redetections of signal S, 6 redetections of signal T, and 8 redetections of signal C, we find 2 additional "detections" that do not correspond to the injected signal, i.e., false positives. We plot the credible intervals of these simulations for three key transit parameters in Figure 8, where we color-code each of these cases.

Ignoring the previous redetections, we are left with 65 simulations, of which 2 are false positives and 40 are successful recoveries. First, this supports the previous argument for a false-positive rate of a few percent (here $(3 \pm 3)\%$). Second, it implies that the false-negative rate may be as high as FNR = $(39 \pm 8)\%$.

5.3. Summary

Since we lack a complete model to describe the noise structure of our data, new synthetic, representative data sets cannot be generated to evaluate the false-positive and false-negative rates of our putative Proxima b transit (signal C). This limits our options to using the original data themselves, which are unfortunately contaminated by three signals—the putative signal itself and two additional, likely spurious, signals.

Although these signals severely impede our ability to investigate the error rates of signal C, we estimate that the false-positive rate is at least a few percent, whereas the false-negative rate is considerably higher at $\sim$20%–40%. Both of these numbers are sufficiently high to warrant serious skepticism regarding the reality of signal C.

At this point, we concluded that the *MOST* data alone were unable to conclusively confirm or reject this candidate signal. While *MOST* data have some unique challenges, due to the orbital motion and data sparsity, we consider that the most likely reason why this analysis is so challenging is not associated with *MOST* itself but rather with Proxima's high flare activity, which leads to likelihood mis-specification.

Cross-validation is perhaps the model selection tool least likely to suffer from the effects of likelihood mis-specification.
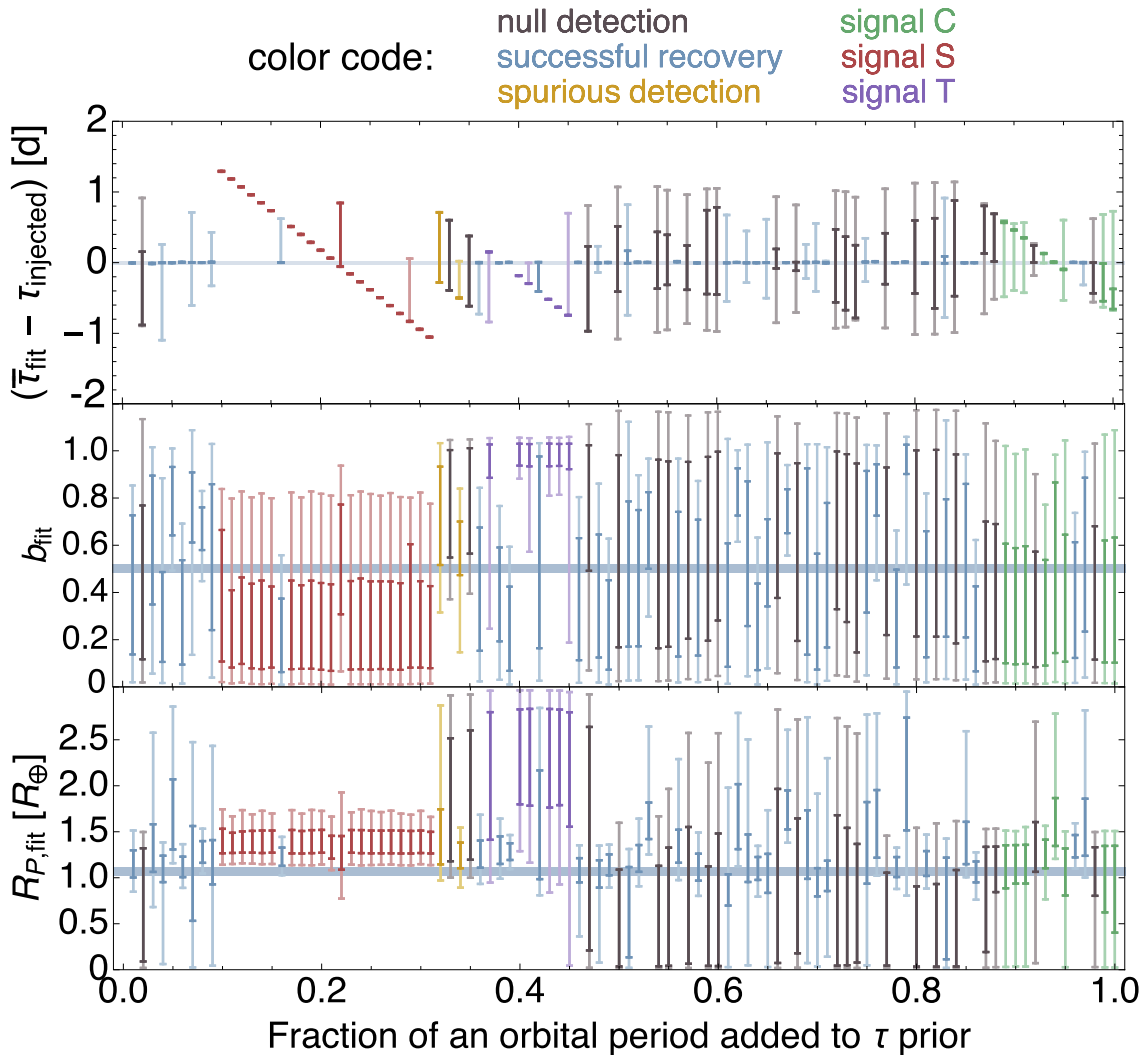
Figure 8. The 1σ (dark bars) and 2σ (light bars) credible intervals from 100 model $\mathcal{M}_2$ fits on the original *MOST* data with a Proxima-b-like transit injected at different phases (x-axis). The color-code key is at the top. The horizontal bars on each plot denote the value of the injected signal.

As mentioned earlier, this is impractical with just two transits observed by *MOST* for signal C. However, cross-validating the signal against other data sets would be a viable and robust way to establish the reality of signal C. Accordingly, we discuss such a test in the next section.

6. Cross-validating with Hatsouth Data

6.1. Observations

Independent of the MOST observations, Proxima Cen was also monitored by the HATSouth ground-based telescope network (Bakos et al. 2013). The network consists of six wide-field photometric instruments located at three observatories in the Southern Hemisphere (Las Campanas Observatory [LCO] in Chile, the High Energy Stereoscopic System [HESS] site in Namibia, and Siding Spring Observatory [SSO] in Australia), with two instruments per site. Each instrument consists of four 18 cm diameter astrographs and associated 4K × 4K backside-illuminated CCD cameras and Sloan *r*-band filters, placed on a common robotic mount. The four astrographs and cameras together cover a 8°.2 × 8°.2 mosaic field of view at a pixel scale of 3″.7 pixel^{-1}.

Observations of a field containing Proxima Cen were collected as part of the general HATSouth transit survey, with a total of 11,071[22] composite 3 × 80 s exposures gathered between 2012 June 14 and 2014 September 20. These include 3430 observations made with the HS-2 unit at LCO, 4630 observations made with the HS-4 unit at the HESS site, and 3011 observations made with the HS-6 unit at the SSO site. Due to weather and other factors, the cadence was nonuniform. The median time difference between consecutive observations in the full time series is 368 s.

The data were reduced to trend-filtered light curves using the aperture photometry pipeline described by Penev et al. (2013) and making use of the external parameter decorrelation (EPD) procedure described by Bakos et al. (2010) and the trend filtering algorithm (TFA) from Kovács et al. (2005). One notable change with respect to the procedure described by Penev et al. (2013) is that we made use of the proper-motion-corrected positions of celestial sources from the UCAC4 catalog (Zacharias et al. 2013) to determine the astrometric

---

[22] This number does not count observations that were rejected as not useful for high-precision photometry, or those that produced large-amplitude outliers in the Proxima Cen light curve.
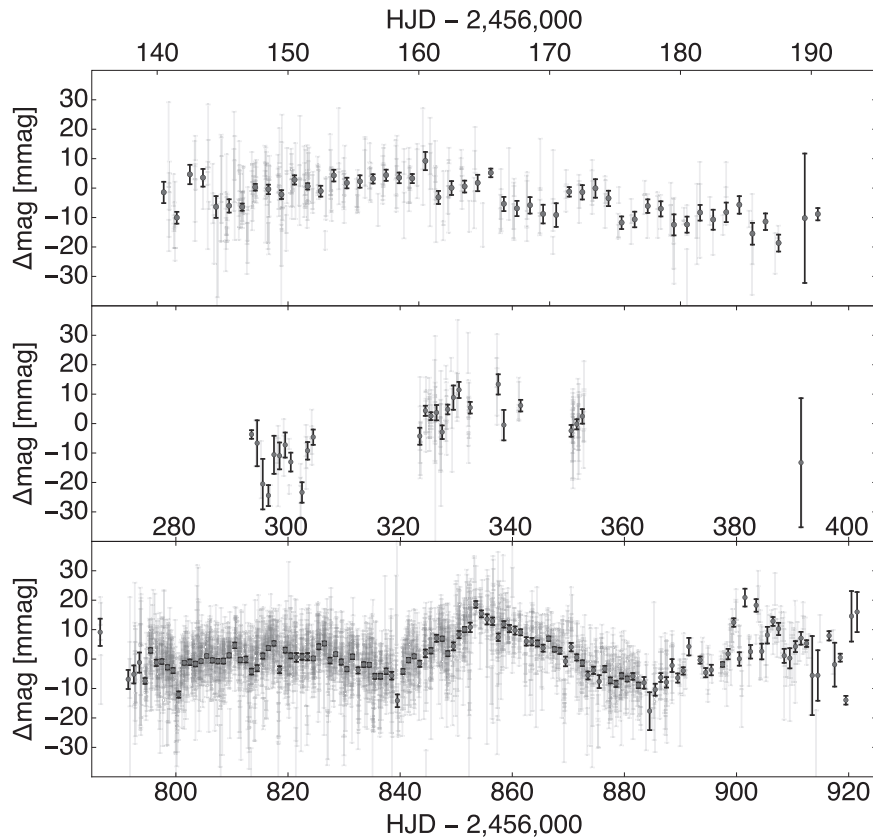
Figure 9. TFA-corrected photometric observations of Proxima with HATSouth from three observing seasons shown at native 240 s cadence (gray bars) and nightly averages (circles). The TFA procedure suppresses real low-frequency variability in addition to removing instrumental trends. The variation seen at HJD −2,456,850 is the residual, after TFA, of a much larger 79 mmag low-frequency astrophysical variation caused by the rotation of starspots.

Table 5
Reduced HATSouth Photometry Used in This Work,
After Correction for Systematic Trends by TFA

HJD$_{UTC}$−2,451,545	TFA Magnitude	Uncertainty
4547.6443629	7.17666	0.0018
4547.6501446	7.18168	0.0018
4595.5309943	7.22295	0.0016
4598.5210661	7.20900	0.0017
4598.5250688	7.22170	0.0014
4598.5290578	7.23077	0.0014
4598.5333811	7.22926	0.0015
4598.5360740	7.23682	0.0027

(This table is available in its entirety in machine-readable form.)

solution for each image and to position the photometric apertures. This modification was essential for Proxima Cen, which, at the start of the HATSouth observations, was displaced by 48″ (13 pixels) from its J2000.0 location and moved a total of 8.″7 (2.4 pixels) over the 828 days spanned by the observations. In this paper we use these data solely to cross-validate the candidate transit signals seen in the MOST light curve. The stellar variability of Proxima Cen as revealed by HATSouth, as well as a general search for transits in its light curve, will be discussed elsewhere. The EPS and TFA processed data are made available in Table 5.

We passed the TFA light curve through a median filter to remove 4σ outliers, and the final photometric light curve

contains 10,869 data points spanning 206 nights (see Figure 9). We find that the formal measurement uncertainties greatly underestimate the observed scatter with $\sqrt{\chi^2/n} = 7.3$, and thus, in what follows, we treat these uncertainties as relative weights but not as reliable estimates of the uncertainty on each point.

We note that the light curve of Proxima Cen shows higher scatter at short timescales than most stars observed by HATSouth in the same field with comparable brightness. Stars within 0.1 mag of Proxima Cen have a median post-TFA rms at the same cadence of 5.3 mmag. For comparison, Proxima Cen has an rms of 13.4 mmag without the additional 4σ clipping and median filtering, or 10.1 mmag after these additional cleaning procedures are applied. This high scatter is likely astrophysical in nature and may be attributable to rapid low-amplitude variations in brightness due to magnetic activity such as flaring.

6.2. Detrending

In order to look for the signal C transit, the TFA light curve requires detrending. We initially attempted to use a GP, as was used for the MOST data, and experimentation with different kernels again favored the Matérn 3/2 with a characteristic covariance timescale of $s = 5.6 \pm 0.4$ hr.

Unlike with the analysis of the MOST data, however, we are not attempting to conduct a joint fit of a GP+transit model, but rather simply wish to detrend the light curve using a GP-only model. This difference is important because GPs are highly
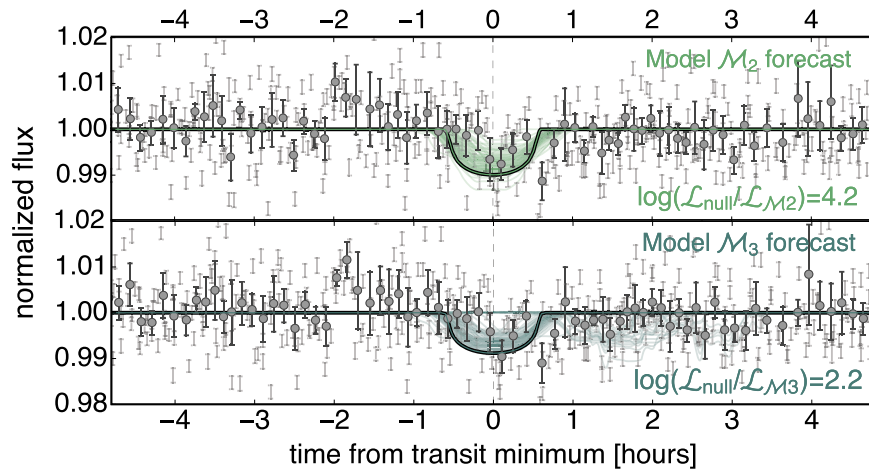
Figure 10. Phase-folded light curve of the detrended HATSouth data on the maximum a posteriori ephemeris from models $\mathcal{M}_2$ (top) and $\mathcal{M}_3$ (bottom), corresponding to signal C. The HATSouth data do not favor a model with signal C included.

flexible and as a detrending tool can actually remove the signals we seek.

In our case, we wish to fold the detrended data on signal C's ephemeris and look for a coherent signal. The GP, particularly with a covariance timescale of just a few hours, is sufficiently flexible to detrend both the long-term changes and potential transits themselves. Indeed, we can verify that this is true since we initially attempted to perform the cross-validation using the GP detrended curves but were unable to recover any injected signals in 100 attempts.

Instead, we used a more tried and tested method for HATSouth data: median filtering. Based on previous experience with HATSouth data, a 2-day window was selected. By injecting fake transits into the HATSouth data and applying the median filtering, we found that 2 days did not remove the injected events. In contrast, when we set the median filtering timescale to 6–12 hr, we observed that many injected signals were not recovered. For these reasons, we ultimately settled on the 2-day window.

6.3. Cross-validation

Phase-folding the detrended HATSouth light curve on the maximum a posteriori ephemeris of signal C reveals visually evident disagreement between the expected model and the observations, as shown in Figure 10. Note that the phase-folded light curve modifies the limb-darkening coefficients to account for the fact that HATSouth is nearly a Sloan r' bandpass (we used the nearest $T_{\rm eff} - \log g$ grid entry from the tabulated list of limb-darkening coefficients in Claret [2004] for the PHOENIX stellar atmosphere results).

We binned the phase-folded light curve of 10,869 points to 6-point bins in phase and computed the weighted mean and weighted standard deviation on each bin. This process enables us to compute realistic uncertainties on each binned point, to overcome the fact the formal uncertainties on the unbinned data are underestimates. Against a simple flat-line model, though, the 1806 binned points have a $\chi^2 = 2966$, indicating that these weighted standard deviations still somewhat underestimate the true uncertainty. Accordingly, we scale them up by a factor of 1.645 such that $\chi^2 = n$.

The χ^2 of the maximum a posteriori solution of model $\mathcal{M}_2$ should be lower than a simple flat line, if the transit were real, but instead it is slightly higher at $\chi^2_{\mathcal{M}2} - \chi^2_0 = 8.4$ for 1803
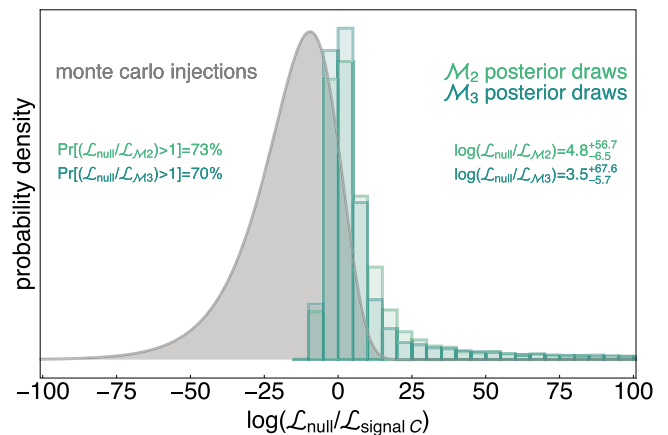
Figure 11. Histogram of the likelihood ratio between a null model and the phase-folded transit of signal C conditioned on the HATSouth data using the posterior samples from models $\mathcal{M}_2$ (light green) and $\mathcal{M}_3$ (dark green). The gray smoothed histogram shows the distribution from 100 injections of a signal-C-like planet phase-folded on the injected ephemeris.

binned points. As a likelihood ratio, this corresponds to a 2.4σ preference for the null model. Repeating for $\mathcal{M}_3$ reveals a similar situation, with $\chi^2_{\mathcal{M}3} - \chi^2_0 = 4.4$, or a 1.6σ preference for the null model. The slight dip seen around the time of transit minimum in Figure 10 can be understood as due to autocorrelation, which detects a very significant p-value for the phase-folded data.

However, this single realization does not capture the range of plausible models allowed by models $\mathcal{M}_2$ and $\mathcal{M}_3$. To assess this, we took each posterior sample, performed a phase fold and rebinning of the unbinned data, and then calculated the χ^2 of each model draw relative to the phase-binned HATSouth data. At each trial, we renormalized the errors such that a flat line through the data yielded a χ^2 equal to the number of data points, in order to fairly reproduce the procedure described earlier. The resulting distributions in $\Delta \log \mathcal{L} = -\Delta\chi^2/2$ for models $\mathcal{M}_2$ and $\mathcal{M}_3$ are shown in Figure 11.

Figure 11 reveals that signal C is clearly not confirmed by this cross-validation exercise. Indeed, for both $\mathcal{M}_2$ and $\mathcal{M}_3$, the majority of the draws favor a null model over the transit model (73% of samples for $\mathcal{M}_2$; 70% of samples for $\mathcal{M}_3$).

To provide some context, we injected 100 signal-C-like transits into the HATSouth data at equally spaced phases and found greater $\Delta \log \mathcal{L}$ values than those observed even in the tail of signal C's predictions.

We conclude that the HATSouth data do not provide compelling supporting evidence for the existence of signal C and moderately disfavor its existence at the 1σ–2σ level. Excluding the signal to greater confidence will likely require far-red, near-infrared, or infrared photometry, such as that which could be provided by *MEarth* or *Spitzer*. Given the sizable false-positive/negative rate of the *MOST* data, the high rate and amplitude of flares produced by Proxima, the a priori low transit probability (1.5%), and the lack of support from independent ground-based photometry, we conclude that there is not a compelling case to be made that our best candidate signal for a transit of Proxima b is genuine.

7. Discussion and Conclusions

In this work, we have searched for the transits of the recently announced planetary candidate, Proxima b (Anglada-Escudé et al. 2016), in two seasons (2014 and 2015) of space-based photometry obtained using the *MOST* satellite. Proxima b has an a priori transit probability of 1.5% and an expected depth of 5 mmag lasting up to an hour (Anglada-Escudé et al. 2016). Accordingly, the signal, should it exist, was expected to be quite detectable with our photometry, which after detrending exhibits an rms of 2–3 mmag per minute, with over 15,000 photometric points spanning 43.5 days at a duty cycle of ~30%.

Proxima Centauri exhibits a few-percent-level photometric variability in the *MOST* bandpass and displays dozens of detectable flares (Davenport et al. 2016). After removing obvious flares detected with the FBEYE approach (Davenport et al. 2014), we still expect a much greater number of smaller flares to exist in our data. Indeed, Davenport et al. (2016) predict 5 mmag flares (the expected transit depth of Proxima b) to occur every 20 minutes, on average. The sheer volume of such large flare events greatly complicates our analysis, and we argue in this work that even our preferred model for the time-correlated variability, a GP with a Matérn 3/2 kernel, is unlikely to be an accurate description of the star's true behavior.

We conduct Bayesian model selection of a GP-only versus GP +transit model using the multimodal nested sampling algorithm MULTINEST (Feroz & Hobson 2008; Feroz et al. 2009). However, the mis-specified likelihood function formally invalidates the Bayes factors that result. Nevertheless, if we assume that the sign of the Bayes factor is correct and impose an informative prior on the period and transit mid-time based of the radial velocity fits of Anglada-Escudé et al. (2016), then we do find a candidate transit signal for Proxima b in our data, which we label as signal C.

Signal C's freely fitted planetary radius is consistent with that expected based on the empirical, probabilistic mass–radius relation of Chen & Kipping (2016) using the Forecaster package, at $1.23^{+0.13}_{-0.15} \, R_\oplus$. As expected, repeating the fits but imposing an extra informative prior on the radius using Forecaster recovers the same signal. However, when we relax the transit mid-time (or phase) informative prior, a stronger transit signal is detected at a phase incompatible (at 4σ) with that expected from the radial velocities, which we label as signal S.

Since our noise model is argued to be inaccurate, we are unable to generate synthetic data sets for false-positive/ negative tests. Instead, we use the original data and perturb the original fit's phase prior 100 times and repeat the fit. These tests reveal a false-positive rate of at least a few percent and a much higher false-negative rate of 20%–40%. This process is complicated by the presence of signals C and S in the data set, though, as well an additional signal detected during these tests dubbed signal T, which is likely spurious based on it's V-shaped morphology.

To resolve the validity of signal C, we phase-folded HATSouth photometry onto the best-fitted ephemeris of our model, and a flat line provides a slightly preferred description at 1σ–2σ significance. Repeating for the posterior draws of our model fit reveals that ~75% of our model predictions give a worse likelihood than a simple flat-line model through the HATSouth data. This final test leads us to conclude that our signal C is unlikely to be associated with Proxima b and is most likely a spurious detection driven by the time-correlated noise structure of our data. As a result of the high false-positive and false-negative rates, even when the period of the signal is known, we did not pursue a blind-period search on this data set, since the reliability of any "detections" would be highly doubtful.

The high flare activity of Proxima Centauri poses a serious challenge for any photometric follow-up of Proxima b. Even if Proxima b is detected to transit, we predict that precise transmission spectroscopy of its atmosphere would be impacted by the flares. The most promising avenue to photometrically follow up Proxima will likely be in the red end of the visible spectrum, or with infrared measurements, where not only will the star be brighter ($K = 4.4$ versus $V = 11.1$) but also the influence of hot flares should be attenuated. Indeed, far-red/ near-infrared/infrared follow-up of the candidate transit signal reported here is recommended to conclusively exclude its existence.

---

[23] github.com/dfm/corner.py

References

Anglada-Escudé, G., Amado, P., Barnes, J., et al. 2016, Natur, 536, 437
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
Batalha, N. M., Rowe, J. F., Bryson, S. T., et al. 2013, ApJS, 204, 24
Benedict, G. F., McArthur, B., Nelan, E., et al. 1998, AJ, 116, 429
Berta-Thompson, Z. K., Irwin, J., Charbonneau, D., et al. 2015, Natur, 527, 204
Bessell, M. S. 1991, AJ, 101, 662
Chen, J., & Kipping, D. M. 2016, arXiv:1603.08614
Christian, D. J., Mathioudakis, M., Bloomfield, D. S., Dupuis, J., & Keenan, F. P. 2004, ApJ, 612, 1140
Claret, A. 2004, A&A, 428, 1001
Claret, A., Dragomir, D., & Matthews, J. M. 2014, A&A, 567, A3
Davenport, J. R. A., Hawley, S. L., Hebb, L., et al. 2014, ApJ, 797, 122
Davenport, J. R. A., Kipping, D. M., Sasselov, D., Matthews, J. M., & Cameron, C. 2016, arXiv:1608.06672
Demory, B.-O., Ségransan, D., Forveille, T., et al. 2009, A&A, 505, 205
Dragomir, D., Matthews, J. M., Eastman, J. D., et al. 2013, ApJL, 772, L2
Dressing, C. D., & Charbonneau, D. 2015, ApJ, 807, 45
Endl, M., & Kürster, M. 2008, A&A, 488, 1149
Evans, T. M., Aigrain, S., Gibson, N., et al. 2015, MNRAS, 451, 680
Feng, F., Tuomi, M., Jones, H. R. A., Butler, R. P., & Vogt, S. 2016, MNRAS, 461, 2440
Feroz, F., & Hobson, M. P. 2008, MNRAS, 384, 449
Feroz, F., Hobson, M. P., & Bridges, M. 2009, MNRAS, 398, 1601
Foreman-Mackey, D. 2016, JOSS, 1, 2
Gibson, N. P., Aigrain, S., Roberts, S., et al. 2012, MNRAS, 419, 2683
Jao, W.-C., Henry, T. J., Subasavage, J. P., et al. 2014, AJ, 147, 21
Kipping, D. M. 2011, PhD thesis, Univ. College London
Kipping, D. M., & Sandford, E. 2016, arXiv:1603.05662
Kovács, G., Bakos, G., & Noyes, R. W. 2005, MNRAS, 356, 557
Mandel, K., & Agol, E. 2002, ApJL, 580, L171
Muirhead, P. S., Johnson, J. A., Apps, K., et al. 2012, ApJ, 747, 144
Penev, K., Bakos, G. Á., Bayliss, D., et al. 2013, AJ, 145, 5
Rowe, J. F., Matthews, J. M., Kuschnig, R., et al. 2006, MmSAI, 77, 282
Tuomi, M., Jones, H. R. A., Barnes, J. R., Anglada-Escudé, G., & Jenkins, J. S. 2014, MNRAS, 441, 1545
Vanderburg, A., Montet, B. T., Johnson, J. A., et al. 2015, ApJ, 800, 59
van Leeuwen, F. 2007, A&A, 474, 653
Walker, G., Matthews, J. M., Kuschnig, R., et al. 2003, PASP, 115, 1023
Winn, J. N., Matthews, J. M., Dawson, R. I., et al. 2011, ApJL, 737, L18
Zacharias, N., Finch, C. T., Girard, T. M., et al. 2013, AJ, 145, 44
Zechmeister, M., Kürster, M., & Endl, M. 2009, A&A, 505, 859