The DRAKE mission: finding the frequency of life in the Cosmos

Subhajit Sarkar,[1]*

[1]*School of Physics and Astronomy, Cardiff University, Cardiff CF24 3AA UK*

ABSTRACT

In the search for life in the Universe, exoplanets represent numerous natural experiments in planet formation, evolution, and the emergence of life. This raises the fascinating prospect of evaluating cosmic life on a statistical basis. One key statistic is the occurrence rate of life-bearing worlds, f_L, the 'frequency of life' term in the famous Drake Equation. Measuring f_L would give profound insight into how common life is and may help to constrain origin-of-life theories. I propose f_L as the goal for the DRAKE mission (Dedicated Research for Advancing Knowledge of Exobiology): a transit spectroscopy survey of M-dwarf habitable zone terrestrial planets. I investigate how the uncertainty on the observed value of f_L scales with sample size. I determine that sampling error dominates over observational error and that the uncertainty is a function of the observed f_L value. I show that even small sample sizes can provide significant constraints on f_L, boding well for the transit spectroscopy approach. I perform a feasibility study of the DRAKE mission using a nominal instrument design and mission plan. Due to low observing efficiencies, DRAKE may need to be incorporated into a wider-ranging deep-space or lunar observatory. A 50-planet survey could constrain f_L to ≤ 0.06 (at 95% confidence) if the sample $f_L = 0$, or 0.03-0.2 if the sample $f_L = 0.1$. This can be achieved (on average) in 10 years using a 17-m telescope with an unrestricted field-of-regard. DRAKE is a viable approach to attempting the first experimental measurement of f_L.

Key words: astrobiology – planets and satellites: terrestrial planets – techniques: spectroscopic – space vehicles: instruments

1 INTRODUCTION

The search for life is one of the major goals of the field of exoplanet science. The past decade has seen great progress made in establishing the demographics of the exoplanet population with results that are encouraging for this search. What has emerged is that small rocky planets greatly outnumber gas giants and that such planets appear to be common in the circumstellar habitable zones of a range of stellar types (Petigura et al. 2013). Planets of this type are the focus in the search for water-based life akin to that of the Earth. The frequency of such planets per star, $\eta_\oplus$, has been estimated for Sun-like and M-dwarf stars based on Kepler statistics. Petigura et al. (2013) considered Earth-sized planets, defined as between 1-2 $R_\oplus$, finding $\eta_\oplus$ of 22% assuming that the habitable zone (HZ) is defined as between 0.25-4 $F_\oplus$ (where $F_\oplus$ is Earth-equivalent stellar irradiance). Using the more conservative definition of the HZ by Kopparapu et al. (2013), based on the moist greenhouse inner limit and maximum greenhouse outer limit, they found $\eta_\oplus$ falls to 8.6%. By comparison, for M-dwarf stars, using this same conservative HZ definition, Dressing & Charbonneau (2015) (see their table 8) found an $\eta_\oplus$ of 27.36% for planets between 1-2 $R_\oplus$ (15.85% for 1-1.5 $R_\oplus$ + 11.54% for 1.5-2 $R_\oplus$), indicating a higher HZ occurrence rate for terrestrial planets around M-dwarfs than around Sun-like stars.

These planets represent innumerable natural experiments in the formation of Earth-like worlds and possible origins of life. As such, astrobiological questions of a statistical nature can start to be ad-

dressed. A fundamental question is how common is life in the Universe? One paramaterisation of this is represented in the famous Drake Equation (Drake 2015) by the term f_L, the occurrence rate of life-bearing worlds. In this paper, I discuss f_L as a potential observable for future spectroscopic surveys of habitable zone exoplanets. I then present results from a bootstrap Monte Carlo simulation elucidating the relationship between the sample size and experimental uncertainty on observed f_L values. I then discuss a dedicated transit spectroscopy space mission called DRAKE (Dedicated Research for Advancing Knowledge of Exobiology) and present results of a feasibility study. The DRAKE mission is designed to perform a survey of M-dwarf habitable zone terrestrial planets with the primary goal of elucidating f_L for this subset.

2 THE FREQUENCY OF LIFE

The Drake Equation gives a sequence of physical factors that in combination give an estimate for the number of communicative civilisations, N, in the Galaxy:

$$N = R_* \cdot f_p \cdot \eta_\oplus \cdot f_L \cdot f_i \cdot f_c \cdot L. \tag{1}$$

The latter terms: f_i (the fraction of life-bearing planets developing intelligence), f_c (the fraction of these developing communication) and L (lifetime of such civilisations) relate to development of intelligence and technology once life has emerged and are unlikely to have experimental verification in this era. However, the first four terms each represent a potentially measurable observable. The first three of

these terms already have experimental measurements: R_* (the rate of star formation) [e.g. 1.65 $M_\odot$/yr (Licquia & Newman 2015)], f_p (the fraction of stars with planets) [~ 1 (Cassan et al. 2012; Batalha 2014)], and $\eta_\oplus$ (the number of planets per stellar system with an environment suitable for life, interpreted here as the occurrence rate of terrestrial planets in the habitable zone).

A survey of atmospheric spectra of terrestrial planets in the habitable zones of stars could potentially, for the first time, provide experimental measurements for the fourth term, f_L, the fraction of planets where life emerges. I term this the 'frequency of life' and define it here as the ratio of the number of terrestrial planets in the habitable zones of stars where life has emerged to the total number of such planets. As such f_L depends on the definition of the HZ boundaries as wider boundaries will fold in more 'candidate' planets, diluting the f_L value. In addition, such a survey will only be sensitive to life 'as we know it', which amounts to carbon-based biochemistries producing identifiable biosignature gases. Life which might arise outside the habitable-zone would not be detected in such a survey.

The inner and outer boundaries of the HZ define a region within which liquid water could exist on the surface of a planet. Frequently used definitions are those based on 1-D cloud-free planet models by Kasting et al. (1993) and later modified by Kopparapu et al. (2013). These studies define the inner 'moist greenhouse' and outer 'maximum greenhouse' limits, referred to as the 'conservative' HZ, while inner 'recent Venus' and outer 'late Mars' limits define the 'optimistic' HZ. However, there exists much debate on the extent of the habitable zone. The effects of clouds, reduced planetary rotation, or desert-like worlds can bring the inner limit closer to the star (e.g., Kitzmann et al. 2010; Zsom et al. 2013; Yang et al. 2014). The effects of a H/He dominated atmosphere could extend the outer limit (Pierrehumbert & Gaidos 2011), potentially into interstellar space when combined with internal heat sources (Stevenson 1999). A future survey of potential habitable zone planets could itself provide experimental constraints on the habitable zone boundaries that would provide a firmer determination of $\eta_\oplus$ and thus f_L. For the purposes of this paper, we assume the optimistic HZ boundaries of Kopparapu et al. (2013) as the basis for selecting candidate planets from which to measure f_L.

Measuring f_L will give insight into how common life is and how life correlates with different planetary conditions. The latter could potentially help to constrain origin-of-life theories. The f_L factor also features in the 'Biosignature Drake Equation' or Seager Equation (Seager 2018). The Seager Equation calculates the number of planets, N, with detectable signs of life based on biosignature gases:

$$N = N_* \cdot f_Q \cdot \eta_\oplus \cdot f_O \cdot f_L \cdot f_S, \tag{2}$$

where N_* is the number of stars considered, f_Q is the fraction of these stars that are suitable for planet-finding (quiescent, non-variable or non-binary), $\eta_\oplus$ is the occurrence rate of HZ terrestrial planets, f_O is the fraction of such planets that are observable (by way of planet orbital geometry), f_L is the fraction of planets that have life, and f_S is the fraction for which detectable biosignature gases are produced. We can rearrange this equation to solve for f_L:

$$f_L = \left(\frac{N}{N_* \cdot \eta_\oplus \cdot f_O} \right) \left(\frac{1}{f_Q \cdot f_S} \right) \tag{3}$$

Here the first bracketed term represents the ratio of the number of planets in a sample positive for biosignatures in the atmospheric spectrum to the total sample of planets observed. Assuming this sample is representative of the larger population, this ratio will give an observed sample value for f_L. This value does not account for

possible false negatives captured in the second bracketed term, e.g. planets orbiting stars that impact the detection of biosignatures or planets where life exists but detectable biosignature gases are not produced. Constraining f_S is highly challenging in this era as we would need to detect life without the presence of biosignature gases. Theoretical work has been done looking at the impact of UV radiation on M-dwarf Earth-like planet spectra (Rugheimer et al. 2015) which may go some way to constraining f_Q, however a definitive value for this term is not yet available. As such, in this paper I consider f_L only for planets where life is detectable through biosignature gases and without accounting for the potential impact of the host star on the biosignature detectability.

The false negative rate might also include life that arises from alternate non-carbon-based biochemistries (e.g. silicon-based) (Petkowski et al. 2021; Bains 2004) that might generate unrecognisable biosignatures. The habitable zone concept is based around liquid water as the medium for biochemical reactions and carbon-based organic chemistry works well in water. This together with favorable physico-chemical properties of water (Pohorille & Pratt 2012) for mediating macromolecular interactions and the high cosmic abundance of carbon together with its ability to easily form complex molecules (Schwieterman et al. 2017) may strongly favour the paradigm of water- and carbon-based life in the habitable zone planets we are surveying. It may be the case that such alternate biochemistries become more important under different physico-chemical conditions with non-water-based solvents (Bains 2004) as might be present in planets or moons outside the habitable zone. If so, then the false negative rate from this effect for planets within the habitable zone might be small, however this and the other false negative considerations might lead to an unquantified bias in estimating the true f_L which is an unavoidable limitation of this study.

3 THE DRAKE MISSION

A survey of terrestrial planets in the habitable zones of stars may not only provide the first constraints on f_L but could address other fundamental questions. What is the range of planetary and atmospheric compositions in the HZ? What initial and boundary conditions determine if a planet evolves onto an 'Earth-like' track as opposed to 'Venus-like' or 'Mars-like' tracks? Are these clearly distinct evolutionary paths or more of a continuum of outcomes? Can the presence of surface liquid water be detected and thus experimental constraints to the habitable zone determined?

While the Hubble WFC3 has detected water vapour around super-Earth and sub-Neptune planets, its wavelength coverage is narrow (e.g. the G141 grism extends from 1.1-1.7 μm), which limits its capability to detect biosignatures. The first realistic prospects for biosignature detection will come through transit spectroscopy with JWST. However, JWST will still require many tens to hundreds of stacked transit observations to characterise a small number of the highest signal-to-noise habitable zone rocky planets (Lustig-Yaeger et al. 2019; Barstow et al. 2015; Barstow & Irwin 2016; Greene et al. 2016; Seager et al. 2009). Ground-based high-dispersion spectroscopy (HDS) is another technique proposed to detect molecular oxygen biosignatures in terrestrial planets. Simulations of HDS on the E-ELT have indicated the oxygen 0.76 μm 'A-band' could be detectable on Earth-like planets around nearby M-dwarf stars (Snellen et al. 2013; Rodler & López-Morales 2014). However, the number of transit observations required would take many years to complete in each case and would be sensitive to only close-by and later M-dwarf hosts. Combining HDS with high contrast imaging (HCI) has

also been proposed to further improve detectability of both Earth-like planets and of oxygen in their atmospheres. Snellen et al. (2015) simulated HDS/HCI with a hypothetical optical integral field spectrograph for the ELT METIS instrument finding that an Earth-sized planet in the habitable zone of Proxima Centauri could be detectable in one night of observing. Lovis et al. (2017) simulated HDS/HCI coupling the SPHERE high-contrast imager and the ESPRESSO spectrograph on the VLT, finding that oxygen could be detected at 3.6σ in 60 nights of observation spread over 3 years. Thus detecting biosignatures with HDS, with or without HCI, remains challenging, requiring many years of observing time and may be limited to nearby M-dwarfs and thus a small overall sample. Current capabilities to to perform a large survey of Earth-like planet spectra are therefore limited. For such a survey, we must ideally develop a new facility dedicated to achieving that specific goal.

Previous mission concepts to obtain large numbers of Earth-like planet spectra have focused on direct-imaging techniques with space-based formation-flying nulling interferometers such as Darwin (Cockell et al. 2009), Terrestrial Planet Finder (Beichman et al. 1999) and LIFE (Quanz et al. 2018). While this approach has many advantages (e.g. a wide range of possible stellar host stars, and large potential sample sizes), significant technological development is still needed.

Instead, I propose the DRAKE mission concept based on the well-established technique of transit spectroscopy (Seager & Sasselov 2000; Charbonneau et al. 2002). The purpose of this mission is to survey a sample of habitable zone terrestrial planets, obtaining atmospheric spectra through transmission spectroscopy at sufficient signal-to-noise ratio (SNR) and wide enough wavelength coverage to detect and characterise biosignatures, with the prime goal of obtaining the first constraints on f_L. The technological development required would be significantly less than for direct imaging methods, and it would benefit from on-going community efforts to optimise the data from existing transit spectroscopy campaigns.

There are several reasons why transit spectroscopic approaches might not have been previously favoured. Transit spectroscopy of terrestrial HZ planet atmospheres is essentially limited to M-dwarf stars. Due to their high molecular weight atmospheres, low temperatures and small radii, the spectral amplitudes observed in transmission spectra from such planets are much smaller than for hot Jupiters. These amplitudes are enhanced as the stellar radius falls, making their detection over noise more feasible around M-dwarf stars than Sun-like stars. This is also true in emission, where the planet-star flux ratio will be greater around M-dwarfs than Sun-like stars. Even so, co-addition of several transit or eclipse observations is required to reduce the noise, and this becomes impractical except around M-dwarfs where the periods in the HZ are in the order of days not years. This well-known 'M-dwarf advantage' essentially rules out study of Earth-like planets around Sun-like stars using transit spectroscopy.

The habitability or biosignature detectability of M-dwarf planets may be affected by stellar flares as well as X-ray and UV radiation from such stars (e.g. Wheatley et al. 2016; Rugheimer et al. 2015). In additional star spots can impact the accurate interpretation of a transmission spectrum. Both occulted and unocculted star spots can bias transit depth measurements if not adequately corrected for (Pont et al. 2013). This may require a correction at the level of the light curve for occulted spots or a wavelength-dependent correction to the transmission spectrum for unocculted spots (e.g. Sing et al. 2011). The spot filling factor is needed for the latter correction and can be estimated through monitoring of the stellar variability. In M-dwarfs, the contamination signal from unocculted spots may be significant and corrections may be complicated by the presence of faculae and

potential uncertainties in determining the filling factor from variability measurements (Rackham et al. 2018). Spot correction efforts remain an area of active research and could potentially utilise additional techniques such as doppler imaging to constrain the filling factor. Stellar pulsations and granulation produce a correlated noise component on the light curve. This may require decorrelation if the photon noise levels are very low especially at shorter wavelengths where it is more prominent, although the absolute level of this noise may be lower in M-dwarfs than Sun-like stars (Sarkar et al. 2018).

The evolutionary development of planets and their atmospheres around M-dwarfs may be different than around Sun-like stars, both due to the longer lifetimes of the stars and the different spectral energy distribution, which could result in more abiotic oxygen false positives (e.g. Luger & Barnes 2015; Domagal-Goldman et al. 2014). There are several other considerations (e.g. IR photosynthesis, stellar variability, tidal locking) (Heath et al. 1999; Tarter et al. 2007) which make it clear that life around an M-dwarf must be considered a particular subset of all possible life, and so extrapolation of results to the wider galactic basis must be carefully evaluated. However, an alternate argument is that M-dwarf HZs are actually the galactic norm, with M-dwarfs hosting the majority of planetary systems, and having a higher occurrence of rocky planets in their HZs than Sun-like stars. As such, they could be considered a priority for such studies.

Another limitation is the geometric transit probability of habitable zone planets which greatly reduces the available sample size. This is increased for M-dwarfs compared to Sun-like stars but remains of the order of about 1%. In addition such M-dwarf systems would need to be bright (thus close), to minimise the fractional photon noise.

Until recently, there was a paucity of known terrestrial planets ($\leq 2\,R_\oplus$) in the habitable zones of M-dwarfs, and even now the numbers of such planets around close M-dwarfs, amenable to spectroscopic follow-up, remain low. At this time 59 planets of radius $\leq 2.5\,R_\oplus$ are known within the optimistic habitable zones boundaries of Kopparapu et al. (2013) [1]. Of these, 14 are $< 2\,R_\oplus$ in size and transiting M-dwarfs. Only 9 of these 14 are within 100 pc. The five closest of these, TRAPPIST-1 planets d,e,f and g, and LHS 1140 b, have had transmission spectra obtained using the Hubble Wide-Field Camera 3 (de Wit et al. 2018; Edwards et al. 2021), with a possible detection of water vapour on LHS 1140 b. The number of known HZ planets orbiting bright M-dwarfs will continue to increase in the coming years as a result of the Transiting Exoplanet Survey Satellite (TESS) mission, together with the CHEOPS satellite, the future PLATO mission and ground-based surveys. In terms of the TESS yield for planets of $< 2\,R_\oplus$ in the optimistic HZ of M-dwarf stars, simulations by Barclay et al. (2018) predicted nine such planets, while Sullivan et al. (2015) predicted 14±4. The extended TESS mission is likely to increase these predicted yields (Bouma et al. 2017). However, these projections indicate that the total number of known transiting Earth-sized planets within the optimistic HZ boundaries of M-dwarfs may remain in the order of tens rather than hundreds for the foreseeable future. A more liberal definition of the habitable zone brings a larger yield. If the habitable zone is defined as 0.2-2 $F_\oplus$, Sullivan et al. (2015) predict that 48±7 planets of $< 2\,R_\oplus$ will be discovered. If it can be shown that small sample sizes of the order of a few tens can provide significant constraints on f_L, then the DRAKE mission's goal of finding f_L becomes a viable prospect.

---

[1] http://phl.upr.edu/projects/habitable-exoplanets-catalog

4 SAMPLE SIZE

It is thus important to establish the relationship between sample size and experimental uncertainty on any observed value for f_L. In this section, I attempt to establish this relationship using a bootstrap Monte Carlo simulation. This is also used to predict the minimum sample size for a single detection for any given population f_L.

The overall uncertainty in the observed value of f_L will result from the uncertainty in positive biosignature identification, observational error (due to noise on the detection), and sampling error due to the sample size. The uncertainty in identification is a complex issue that requires delineating all possible gaseous biosignatures, the subject of much current research and debate. Earth-like biosignatures include the products of metabolism such as oxygen (O_2) and its by-product ozone (O_3), methane (CH_4) and nitrous oxide (N_2O) (discussed further in Section 5.2) (e.g. Marais et al. 2002; Segura et al. 2003; Kaltenegger & Traub 2009; Seager 2014; Schwieterman et al. 2017). However, biosignatures may present differently at earlier times in a planet's evolution (e.g. Kaltenegger et al. 2007) or if non-Earth-like biochemistries have evolved. False positives for oxygen in particular must be accounted for, as discussed later.

For this study, I have assumed that, in the absence of any noise on the spectrum, positive or negative biosignature identification can be made with certainty on each spectrum. While this simplification might not be justified with our current knowledge, the database of spectra collected on any survey could be re-analysed repeatedly into the future and, as our understanding of biosignatures improves, such identification can be made with growing certitude. I therefore examine only the effects of observational error and sampling error on the overall observed f_L uncertainty in relation to sample size.

In the following discussion I term the observed sample value of f_L as $f_L[\text{obs}]$, i.e. the measured experimental value of f_L found from a survey of habitable zone terrestrial planets of sample size N_{samp}. The true population value of f_L is termed $f_L[\text{true}]$, i.e. the actual value of f_L in the population of such planets. The primary goal of the simulation is to find how the uncertainty or experimental error on $f_L[\text{obs}]$ scales with sample size, N_{samp}. This is parameterised as the 95% confidence interval (CI), which is defined here as the range of most probable values of $f_L[\text{true}]$ that could give that value of $f_L[\text{obs}]$ 95% of the time: a measure of the likely range of the true, unknown parameter.

4.1 Monte Carlo simulation

The simulation is run under two conditions. In the first condition, we ignore any observational error in assigning biosignature positivity due to noise on the final spectrum, so that the uncertainty on f_L is purely due to sampling error. In the second, we incorporate observational error as well as sampling error. As mentioned above, I assume that there is no uncertainty in identifying a given pattern of gaseous signatures in a noiseless spectrum as either positive or negative for life.

The simulation is initiated by setting up a total population of transiting habitable zone terrestrial planets of size N_{pop}. If we consider M-dwarfs only, these make up ~70% of the ~10^9 stars in the Galaxy. If we assume about half of these have a planet of 2 $R_\oplus$ or smaller in the habitable zone, then assuming a transit probability of about 1%, this gives an order of magnitude for N_{pop} of ~ 10^6. I thus adopt N_{pop} = 10^6 in these simulations.

A given sample size, N_{samp}, is chosen, where N_{samp} = 10, 20, 30, 40, 50, 100, 200, 500 or 1000 planets. A value for $f_L[\text{true}]$ is chosen, ranging from 0 to 1 in steps of 0.001. For each $f_L[\text{true}]$ the

following algorithm is implemented. Each planet in the population is randomly assigned to be either positive or negative for spectral biosignatures, while ensuring that the the total number of positive planets is exactly $N_{\text{pop}} \times f_L[\text{true}]$. A sample of planets of size N_{samp} is randomly selected (without replacement) from this population repeatedly 1000 times. In each sample, the $f_L[\text{obs}]$ is recorded by taking the ratio of positive planets in the sample to the total sample size. This way, 1000 measurements of $f_L[\text{obs}]$ are taken for each $f_L[\text{true}]$. When accounting for the observational error, I assume that the spectrum from each planet is identified as positive or negative at 3σ significance. As such, I assume that there is a 99.73% chance that the planet is correctly identified and a 0.27% chance it is incorrectly identified. Based on these probabilities the assignment of each planet to positive or negative for life is either retained (correctly identified as a true positive or true negative) or flipped (to give a false positive or false negative due to noise on the observation). The simulation is repeated 10 times and an average set of results used for the subsequent analysis.

I use the occurrence rates of $f_L[\text{obs}]$ for each $f_L[\text{true}]$ (averaged over the 10 simulations) to estimate the minimum sample size required to obtain at least one positive planet at 95% confidence. This is a function of $f_L[\text{true}]$. At each $f_L[\text{true}]$, the fraction of outcomes where $f_L[\text{obs}]$ is non-zero to the total number of outcomes, gives the 'positive fraction', the probability that at least one planet in the sample gives a positive result. The positive fraction is found at each N_{samp} and then a line fit is used to determine the N_{samp} value that corresponds to a positive fraction of 0.95, i.e. the sample size where there is a 95% chance of at least one positive result. The line fit is achieved using a polynomial in log-log space[2]. In some cases, a line fit is not needed, e.g. if the positive fraction is already >0.95 with N_{samp} = 10 or if the positive fraction does not reach 0.95 even at N_{samp} = 1000 (in which case no result is recorded). The latter occurs when $f_L[\text{true}]$ reaches very low values.

Next, we invert the occurrence rates to find the number of occurrences of $f_L[\text{true}]$ for each $f_L[\text{obs}]$. The relative occurrence rates reflect the probability distribution of $f_L[\text{true}]$ for a given $f_L[\text{obs}]$ and N_{samp}. I use these distributions to find 95% confidence intervals for each $f_L[\text{obs}]$ at each N_{samp}. The range of possible $f_L[\text{obs}]$ values is a function of N_{samp}, varying from 0 to 1 in steps of $1/N_{\text{samp}}$. Since the distributions are in most cases skewed, the 'shortest interval' definition of the confidence interval (Hall 1988) is used, where there is a combined tail-end probability of α, where, in this case, $\alpha = 0.05$. This means that if $f_L[\text{obs}]$ is 0 or 1, the confidence interval (CI) will be entirely one-sided, which is appropriate since $f_L[\text{true}]$ cannot be < 0 or > 1. In each case, the shortest confidence interval is found by determining the shortest distance parallel to the x-axis between two points on the curve (of number of occurrences vs $f_L[\text{true}]$) that encloses 95% of the total area under the curve[3] as shown in Figure 1.

4.2 Results

Figure 2 shows the minimum sample size that will give at least one positive result vs $f_L[\text{true}]$. Adding the observational error reduces the

---

[2] I find that the following degrees of polynomial, r, give good fits to the region around the 0.95 positive fraction in the following $f_L[\text{true}]$ ranges: $r = 5$ if $f_L[\text{true}] < 0.01$, $r = 6$ for $0.01 \leq f_L[\text{true}] < 0.1$, and $r = 7$ for $f_L[\text{true}] \geq 0.1$.
[3] This is achieved using a custom code utilising the scipy functions: 'interpolate' and 'integrate'.
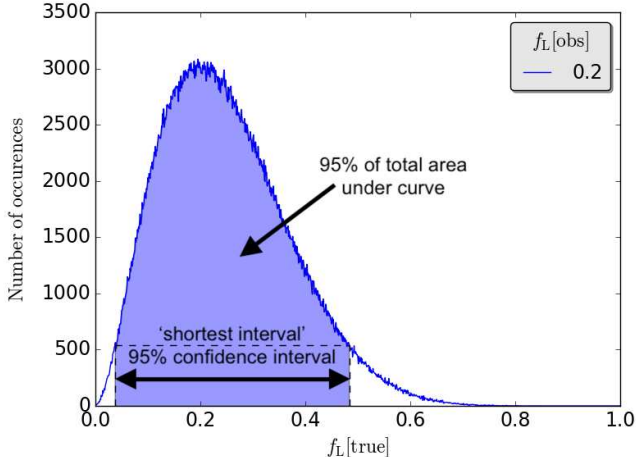
Figure 1. Shortest 95% confidence interval. This is the shortest distance parallel to the x-axis that contains 95% of the area under the curve, with a combined tail probability of 5% asymmetrically distributed between the two tails. This example shows the 95% CI for an observed value $f_L[\text{obs}]$ = 0.2 when N_{samp} = 10 (based on simulations that included sampling and observational error and averaged over 10 simulations). The curve plots the number of occurrences of a given $f_L[\text{true}]$ that resulted in an observed f_L of 0.2. This distribution is proportional to the probability distribution and hence used to find the confidence interval as shown.

minimum sample size (the effect being greater at smaller $f_L[\text{true}]$) due to the inclusion of false positives. As would be expected, larger $f_L[\text{true}]$ corresponds to smaller minimum sample sizes. If $f_L[\text{true}]$ were 5% or more, then a sample size of 58 would give at least one positive result that was not a false positive due to observational error (Figure 2, inset). This is similar to the prediction from The LUVOIR Team (2019): 54 planets for one detection if the frequency of habitable planets is 5%. The problem with using this type of information is that we have no a priori information on $f_L[\text{true}]$ and, as such, this kind of analysis is of limited use in designing a mission where the available number of planets might be small.

The analysis of the uncertainty in $f_L[\text{obs}]$ might be more useful in this regard, as we make no assumptions on the underlying value of $f_L[\text{true}]$. Due to the low probability assumed for false positive and false negative detections, which I have taken to constitute the observational error, the final results for $f_L[\text{obs}]$ uncertainty are dominated by sampling error, and as such no significant differences are found between the simulation that included observational error and the one that omits it. Therefore I present and discuss here only the results that included the observational error. Figure 3 (left) shows the upper and lower boundaries of the 95% CI on $f_L[\text{obs}]$ vs $f_L[\text{obs}]$ for different sample sizes, N_{samp}. These are asymmetric around the observed value except if $f_L[\text{obs}]$ = 0.5, and completely one-sided at $f_L[\text{obs}]$ of 0 or 1. For example, if a survey had a sample of 20 planets, and the experimental result was an $f_L[\text{obs}]$ of 0.2, $f_L[\text{true}]$ ranges around the observed value from -0.13 to +0.2, and the overall size of the CI is 0.33. Figure 3 (right) shows the absolute size of the CI vs $f_L[\text{obs}]$ for different N_{samp} values. We see that the CI size is a function of $f_L[\text{obs}]$ and is maximal when $f_L[\text{obs}]$ = 0.5 and minimal at $f_L[\text{obs}]$ of 0 or 1. Larger N_{samp} values result in smaller uncertainties, however, the incremental improvement in uncertainty is greatest at the smaller values of N_{samp}, so that pushing the sample size up has less and less impact on the final precision.

To better visualise the relationship between N_{samp} and uncertainty,

in Figure 4 the CI is plotted vs N_{samp} for different $f_L[\text{obs}]$ with the interval filled between the upper and lower limits. We again see the value of increasing the sample size at small sample sizes and the reduced impact at large sample sizes. An interesting outcome of this study is the result for $f_L[\text{obs}]$ = 0, i.e. a completely negative sample. Even in such a situation we can determine an upper limit to f_L at 95% confidence. The upper limit depends on the sample sized used, falling with N_{samp}. If we consider a sample with just 10 planets returning an $f_L[\text{obs}]$ of 0, we can still constrain the upper limit of $f_L[\text{true}]$ to 0.24 at 95% confidence (Figure 4, left, inset), i.e. there is a 5% chance it would be higher than this. While this is still a large uncertainty, it might still give a profound insight into the occurrence rate of life in the Universe. If a sample size of 50 planets returned $f_L[\text{obs}]$ of 0, the upper limit to $f_L[\text{true}]$ becomes 0.06, thus constraining the frequency of life to 6 percent or less at 95% confidence (again not accounting for false negatives). The same sample size would constrain $f_L[\text{true}]$ to between 0.36-0.63 at 95% confidence if an $f_L[\text{obs}]$ of 0.5 was measured.

Formulae such as Cochran's formula (Cochran 1963) estimate the sample size, n_0, required to measure the fraction of a parameter, p, (e.g. f_L) in a large population to obtain a given margin of error, e, (due to sampling error) at a given confidence level (represented by the Z-score): $n_0 = Z^2 p(1-p)/e^2$. One reason it is not directly applicable here is that it returns the error on a measurement assuming knowledge of the underlying true value of p. For example, the Cochran formula would return an error of zero for p = 0 or 1. However, clearly the error on an observed value of f_L that is 0 or 1 will not be zero, so Cochran's formula cannot be directly applied in these cases, since we have no a priori knowledge of the underlying true population f_L. If the true fraction of the parameter were 0.5, then the Cochran formula gives the maximum possible margin of error. We find that for large values of N_{samp}, the maximum CI size on $f_L[\text{obs}]$ matches the prediction from the Cochran formula well (Figure 3, right). However, there is a deviation from the formula at low N_{samp} values where our method returns smaller CI sizes to the formula. This may be related to progressive deviation of $f_L[\text{true}]$ occurrence rates from the normal distribution as the sample size falls.

We therefore find that even small sample sizes can return interesting results, with diminishing return on the improvement in uncertainty as the sample size is pushed up. A high-quality survey of about 50 planets could provide significant results for f_L. It is plausible that in the next decade we may have such an available sample amenable to transit spectroscopy.

5 OBSERVATORY DESIGN

The DRAKE mission concept is centred around a transit spectroscopy instrument on a space-based visible-IR observatory that conducts a scheduled sequence of spectroscopy transit observations of the exoplanet sample. Space is favoured over ground for transit spectroscopy: the noise is much lower due to the absence of atmospheric scintillation and turbulence, together with reduced observatory emission at longer wavelengths. Furthermore, ground-based continuous wide-wavelength coverage is curtailed due to telluric absorption bands.

I assume here an observatory design specially tailored for such a mission, however the goals of the DRAKE mission could equally be achieved as a specific component of a general purpose observatory project. Possible locations include deep space, such as the L2 Lagrange point or a location on the Moon, such as the Shackleton crater at the lunar South Pole (Schneider et al. 2021; Eads & Angel 2021). Each location has advantages and disadvantages. Low-Earth orbit,
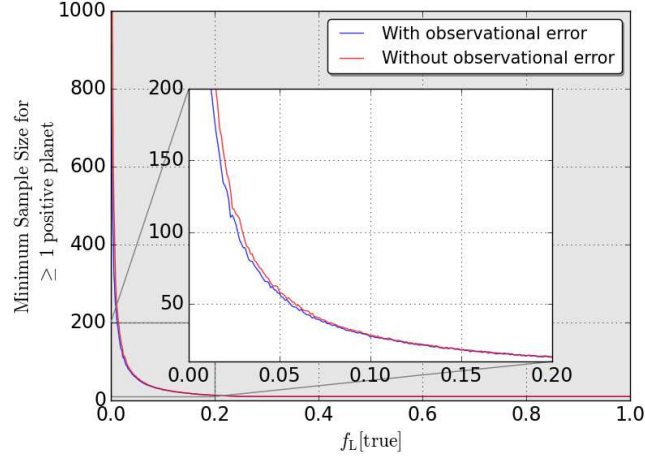
Figure 2. Minimum sample size that returns at least one planet positive for biosignatures vs $f_L[\text{true}]$.
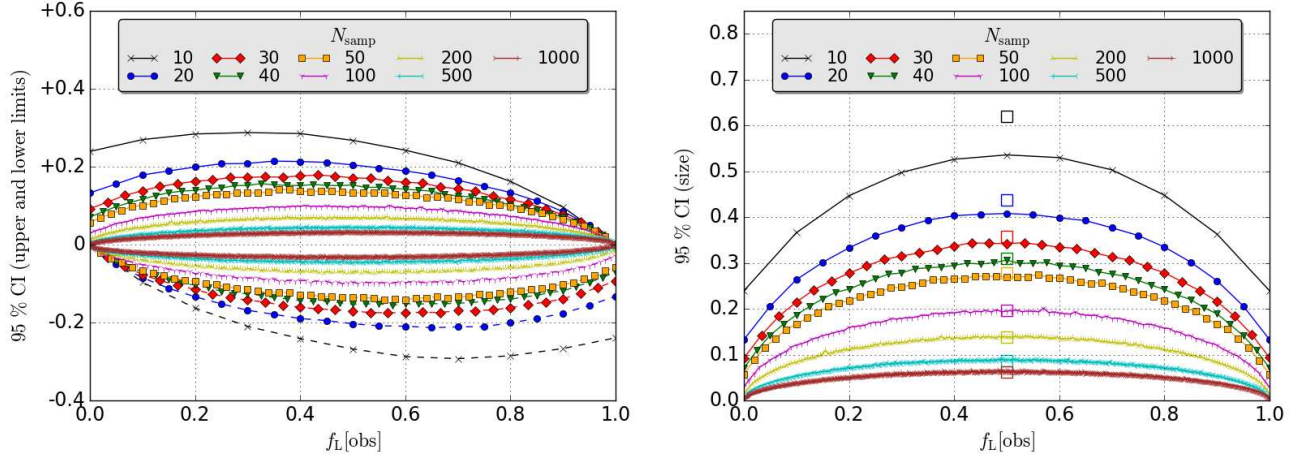

Figure 3. 95% confidence intervals for $f_L[\text{obs}]$ for different sample sizes, N_{samp}. Left: CI upper (solid lines) and lower (dashed lines) limits. Right: CI absolute size. Also shown on the right plot as squares are the predicted maximum 95% CI sizes using the Cochran formula (see text for details).
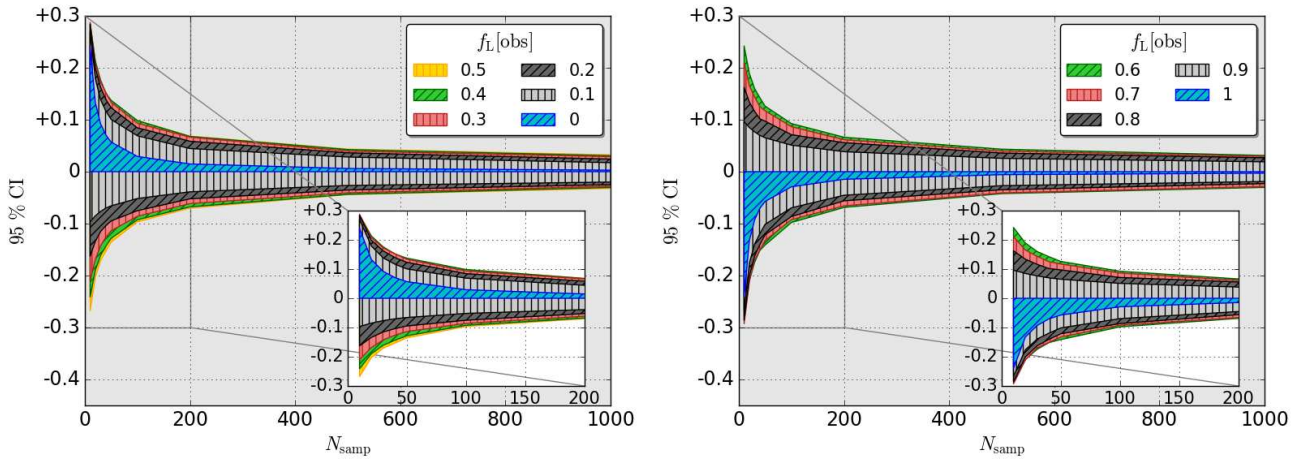

Figure 4. 95% confidence interval vs N_{samp} for different $f_L[\text{obs}]$ values. Left: $f_L[\text{obs}]$ 0-0.5. Right: $f_L[\text{obs}]$ 0.6-1.

while allowing for the possibility of maintenance missions, would be subject to interruptions by diurnal cycles, and thus is not ideal for this mission where long out-of-transit baseline observations are needed to reduce the noise on the transit measurement and allow for better characterisation of host star activity and variability.

Designing a free-flying observatory at L2 would benefit from the heritage of previous observatories, such as Ariel and JWST, but would also have a limited life-span as consumables such as coolant and propellant run out. I consider 10 years to be a likely upper limit for the mission lifetime without maintenance. Maintenance with human missions would be unlikely at this location so that minimising mission duration is essential. For missions at L2, thermal constraints limit the allowable range of solar aspect angle, with an instantaneous field-of-regard (FOR) that is restricted and changing with time. This means that different planets would be observable at different times as the observatory orbits the Sun. Since the mission relies on conducting sequential observations on exoplanets that may be located anywhere in the sky, this will impact on the efficient scheduling of these observations, increasing mission duration. Novel designs could be considered to increase the instantaneous FOR and thus reduce the mission duration. For example, if a free-floating Sun-shade is used, the FOR would be almost unrestricted except for the region occulted by the shield. An alternate power supply to solar-power would however be needed or a design with solar panels that could extend beyond the shaded zone.

A polar lunar crater could provide a continuously-shaded thermally-stable location where solar power could be transferred from illuminated areas at a remote distance. The FOR would be permanently restricted to one half of the sky so that roughly only half the possible number of candidate planets could be observed. However the observable sample would be be always in the FOR so that scheduling efficiency would be close to that for an unrestricted FOR. A possible second telescope at the other pole could simultaneously cover the planets in the other hemisphere, further increasing efficiency. Lunar telescopes however would require huge investments in infrastructure, and while plans exist for humans to return to the Moon in the coming decade, it is unclear how this will develop with time. A lunar telescope could however be maintained indefinitely, potentially allowing for a mission duration extending beyond 10 years.

The nominal mission assumes obtaining transmission spectra in primary transit, rather than emission spectra from secondary eclipse. I assume that signals in transmission will generally be stronger than in emission. For example, examining the model transmission spectrum for an Earth-like planet orbiting the M3 star AD Leo (3390 K, 0.39 $R_\odot$) from Meadows (2017) (their Fig. 2), which includes photochemistry effects, we find that a typical spectral amplitude is about 4 ppm. For the same planet and star, a dayside emission spectrum (assuming the spectral amplitude $\approx F_p(\lambda)/F_s(\lambda)$, where F_p is the flux from the planet and F_s is the flux from the star arriving at the telescope, and modelling both planet and star as blackbodies) gives a maximum contrast ratio of 2.8 ppm at 11 μm (the longest wavelength proposed in the prototype design), and below 8.5 μm, the signal is always below 1 ppm.

I next describe the key features of a prototype design which forms the basis of the instrument model in the subsequent feasibility study.

5.1 Telescope and common optics

The telescope and common optics (TCO) describe all optical elements in the light path preceding the DRAKE spectrometer instrument. The key elements are the primary mirror and the sequence of smaller mirrors in the light path which affect the overall transmission

and will contribute to the optical emission background. The primary mirror diameter, D_{tel}, is the key design parameter which is investigated in the feasibility study. D_{tel} is varied from 10 to 50 m. The total transmission (TCO transmission combined with the spectrometer transmission), ν, is assumed to be 0.6.

5.2 Spectrometer

The baseline concept is for a visible to infrared low resolution dispersive spectrometer providing wide wavelength coverage to maximise the identification of biosignatures in the pattern of spectral features from an exoplanet atmosphere. The wavelength range is divided into two channels. This allows each channel to be optimised for its particular wavelength range, e.g. detector characteristics, spectral dispersion, spectral resolving power etc.

5.2.1 Channel A (0.6-5 μm)

A spectrometer channel ranging from visible to near infrared (NIR) wavelengths (0.6-5 μm) would permit detection of key biosignature molecules as well as the characterisation of clouds, hazes, and Rayleigh scattering, and monitoring for stellar activity. Meadows et al. (2018) present a possible flow chart (their Fig. 11) for spectroscopic identification of a photosynthetic biosphere or 'Archaen Earth' exoplanet based on the presence or absence of water (H_2O), O_2, O_4, CH_4, carbon dioxide (CO_2) and carbon monoxide (CO). This aims mainly to interpret any detection of O_2 in the context of other gases that may or may not support its biological origin. Abiotic origin of O_2 could result from photolysis of water, especially in ocean-loss scenarios, which might present with extremely high O_2 levels marked by the presence of O_4 (Luger & Barnes 2015; Schwieterman et al. 2016). It could also arise from photolysis of CO_2, correlating with large CO_2 abundances and the presence of CO as a photolytic by-product (Hu et al. 2012; Domagal-Goldman et al. 2014).

The wavelength coverage of Channel A includes features from all these gases. There is strong O_2 absorption at 0.76 μm ('A-band'), with weaker features at 0.69 μm ('B-band') and 0.63 μm ('γ band'), and O_4 features at 1.06 and 1.27 μm (Meadows et al. 2018). Strong CH_4 features occur at 1.65, 2.4 and 3.3 μm (Schwieterman et al. 2017), and large amounts of CH_4 could be supportive of biogenic rather than photolytic origin of O_2 as CH_4 is a sink for photochemically generated O_2 (Meadows et al. 2018). Strong CO_2 absorption occurs at ∼ 1.65, 2 and 4.2 μm and CO at 2.35 and 4.6 μm (Meadows et al. 2018). The presence of water vapour would be a necessary but not sufficient marker for life as we know it. In this range there are water absorption bands at 0.65, 0.7, 0.73, 0.8, 0.95, 1.1, 1.4, and 1.8–2.0 μm (Meadows et al. 2018). Organic sulphide gases are produced by bacteria and fresh water green and blue-green algae (Rasmussen 1974; Cooper et al. 1987; Pilcher 2003). The sulphide gases methanethiol (CH_3SH), dimethyl sulphide (CH_3SCH_3) and dimethyl disulphide (CH_3SSCH_3), together with methyl chloride (CH_3Cl) (another potential biosignature associated with numerous sources such as algae, plants and fungi (Khalil & Rasmussen 1999; Harper 1985; Yokouchi et al. 2002)), all have an absorption band between 3-4 μm (Schwieterman et al. 2017, Fig. 6). Nitrous oxide (N_2O) is a biosignature gas generated by nitrogen-fixing bacteria and algae (Sagan et al. 1993). In this channel, N_2O has significant features at 3.7 and 4.5 μm (Schwieterman et al. 2017). Trivalent phosphorus compounds have recently been proposed as possible biomarkers of anerobic organisms, in particular phosphine (Sousa-Silva et al. 2020)

for which there has been a claim of detection on Venus (Greaves et al. 2020). Phosphine has strong bands around 2.7-3.6 μm and 4.0-4.8 μm (Sousa-Silva et al. 2020).

5.2.2 Channel B (5-11 μm)

Wavelength coverage extending into the mid-infrared (MIR) may be desirable to resolve degeneracies resulting from overlapping molecular features as well as between gas abundances and temperature structure (in eclipse) and gas abundances and cloud coverage (in transit) (Barstow et al. 2015), although eclipse observations are not included in the baseline design. Channel B covers the wavelength range between 5-11 μm. This range includes a water band at 6.3 μm and a very strong O_3 feature at 9.6 μm. O_3 is a photochemical byproduct of O_2 (Meadows et al. 2018). The upper bound of the wavelength range is chosen to include the full O_3 feature but is not extended to longer wavelengths as the benefit of covering longer wavelengths is outweighed by the negative impact on SNR. N_2O absorbs at 7.8 and 8.6 μm and CH_4 at ~7-8 μm (Schwieterman et al. 2017). In addition, organic sulphide gases have several absorption features in this range, e.g. dimethyl sulphide at 6–7 and 10 μm (Schwieterman et al. 2017). The relevance of organic sulphide gases as biosignatures is determined in relation to ethane (C_2H_6), which has an absorption feature in this channel at 6–7 μm (Schwieterman et al. 2017). CH_3Cl also has features in this range at 7 and 9.7 μm (Schwieterman et al. 2017). Phosphine has absorption at 7.8-11.5 μm (Sousa-Silva et al. 2020).

5.2.3 Detectors

Different optimal detector technologies may be needed for each channel, e.g. mercury-cadmium-telluride (MCT) for Channel A or arsenic-doped silicon impurity band conduction for Channel B. At this stage I do not specify detailed characteristics of the detectors, such as dark current, read noise, pixel full-well capacity or flatfield variations, since the feasibility study will assume a photon-noise-limited instrument. For both channels the detector quantum efficiency, QE, is assumed to be 0.8[4], and the quantum yield[5] is assumed to be unity. The final photon-converting efficiency, $\kappa = QE \times \nu$, is then 0.48 (simplified here to be non-wavelength dependent).

5.2.4 Spectral resolving power

The spectral resolving power R (where $R = \lambda/\Delta\lambda$) of the dispersive element sets a limit to the width of spectral features that can be detected, with narrower features needing higher R. So each channel must have a minimum R power that permits the detection of key spectral features in its wavelength range.

The detectability of a given spectral feature may depend both on its SNR per spectral bin (which falls with R) and also on the number of points (spectral bins) sampled across the feature (which increases with R). The complex balance between these two effects will influence the optimal R power but this will require more detailed studies with both more defined instrumental modelling and spectral retrieval studies. While the relationship between SNR and

R is straightforward for a photon-noise-limited instrument, when instrumental noise sources such as read noise are included, it becomes more complex. For example, changing the R power affects the linear dispersion and the overall length of the spectral trace, which in turn affects the number of pixels per resolution element and hence the read noise and dark current noise. Other considerations include the point-spread-function (PSF), which will also affect the R power. The PSF size should be such as to be Nyquist-sampled by the detector but it is also affected by wavefront-error aberration that is related to the surface quality of the primary mirror. The requirements on the latter may strongly impact feasibility, particularly through overall cost. PSF size will also influence the size of the extraction aperture and hence the amount of instrumental and background noise folded into each spectral element. The linear dispersion and PSF shape will also impact the saturation time on the detector for a given pixel full-well capacity, limiting the maximum brightness of the target that can be observed. Future trade-off studies will be needed to find the optimal instrumental design, which maximises the detectability of the final spectrum. For now the R power per channel is based on previous studies.

Robinson et al. (2016) modeled a coronographic instrument and found the oxygen A-band detectable at at Earth-like abundances with $R = 70$. An analysis by Brandt & Spiegel (2014) found that the optimal resolution for an A-band O_2 detection varied from $R \sim 70$ to many hundreds, depending on the instrumental noise model adopted. Thus an R of 100 is adopted for Channel A and it is assumed that this is likely to be adequate for detecting the O_2 A-band and other features in Channel A.

In the MIR, ro-vibrational spectral bands for biosignature gases tend to be broader than at shorter wavelengths (Schwieterman et al. 2017, Fig. 6) so that lower R powers can be considered. Meadows et al. (2018) suggest an R of 10 could be used to scan for broad features such as water or O_3. However, to to give some margin to detect narrower features at the shorter wavelengths, an R of 30 is adopted for Channel B. This is consistent with the minimum R power of the Ariel long-wave channel (AIRS channel 1) operating between 3.9 and 7.8 μm (Tinetti et al. 2018).

5.3 Spacecraft

Assuming a deep-space location, the TCO and spectrometer will interface with the spacecraft service module which controls vital functions such as power, telecommunication, data handling, thermal control and attitude control. A lunar location would require a modification of this interface suitable for a ground-based observatory.

6 FEASIBILITY STUDY

In this section, I simulate the nominal DRAKE mission and assess its capability to fulfil the primary goal of obtaining an experimental value for f_L. The mission consists of a sequence of transit spectroscopic observations of a sample of N_{samp} planets producing a transmission spectrum for each planet extending from 0.6 to 11 μm.

The feasiblity study has the following steps, which are described in more detail below. A population of Earth-sized planets around M-dwarf stars is simulated using a model planet population generator, filtered and then ordered by photometric detectability. 100 such randomised realisations are generated. In each instance, for each candidate planet, the SNR to detect a 'typical' spectral feature is found at each wavelength in each channel for a single transit, $SNR_1(\lambda)$. The minimum SNR_1 is then used to calculate the the required number

---

[4] This is consistent with quantum efficiency measurements for HxRG detectors (e.g. Blank et al. 2011; Mosby et al. 2020) and for experimental LWIR MCT detectors (Cabrera et al. 2020).

[5] The internal quantum yield is the number of electron-hole pairs produced per photon absorbed.

of transit observations, N_t. This is done using an instrument model of the DRAKE observatory with a primary mirror size, D_{tel}, of 30 m. The SNR results (and thus N_t) are then scaled for different values of D_{tel} ranging from 10-50 m. Samples sizes, N_{samp}, are considered ranging from 10-100 planets A scheduling algorithm is used to find the total mission time, T_{miss}, to complete observations for all N_{samp} planets. This is done by first selecting an initial pool of planets, N_{pool}, which is $1.2 \times N_{samp}$ and chosen in order of photometric detectability. From the 100 realisations for each case, average results are obtained. We thus obtain relationships between T_{miss}, N_{samp} and D_{tel} from which feasibility is assessed.

6.1 Instrument model

The instrument model is based on the design presented in Section 5 and assumes a photon-noise-limited regime with no other influences on photometric stability. As such, we assume that other noise sources and systematics can be mitigated either through design or in data reduction to levels well below the photon noise. These include read noise, dark current noise, pointing jitter, detector non-linearity, persistence and 1/f noise, as well as astrophysical factors including stellar activity (Rackham et al. 2018) and stellar pulsation and granulation (Sarkar et al. 2018). Background noise contributions from optical surface emissions and zodiacal light are assumed to be negligible. For the former, this should be a reasonable assumption with passive or active cooling of optical elements to about 50-70 K. However, at the extreme long wavelength end of Channel B background contributions such as zodiacal light noise could become significant (Sarkar et al. 2020).

6.2 Model planet population

We set up a model planet population for this study as follows. The space around the Earth is divided into consecutive shells of width 5 pc out to a distance of 100 pc. The TESS Input Catalogue Candidate Target List (CTL) (Stassun et al. 2018) is mined to find the number of known M-dwarfs in each subclass within each shell[6]. The search looks for stars included in the 'cool dwarfs' special list with TESS magnitude of 16 or less. Stellar subclass is assumed based on temperature, as categorised in Kaltenegger & Traub (2009). Using these criteria no stars of M6 or later are found. Figure 5 shows the numbers of stars of each subclass in each shell.

These distributions are used to set up a randomised model stellar population. In each shell, stars of the same subclass are given the same radius (R_s), mass (M_s) and temperature (T_s), based on Kaltenegger & Traub (2009). However, in each realisation their distance, d, is randomised within the shell. Each star is then randomly allocated a habitable zone planet or no planet. If allocated a planet, it is specified as a 'sub-Earth' ($0.8R_\oplus < R_p < 1R_\oplus$), an Earth-sized planet ($1R_\oplus < R_p < 1.5R_\oplus$), or a 'super-Earth' ($1.5R_\oplus < R_p < 2R_\oplus$) based on the occurrence rates in Dressing & Charbonneau (2015) for optimistic (Early Mars - Recent Venus) HZ limits. These are 13.09% for sub-Earths, 24.28% for Earth-sized planets, and 20.69% for super-Earths. Each planet is randomly assigned a semi-major axis, a, within the HZ boundaries for that star. The boundaries of the HZ are calculated for each star based on effective stellar flux values for Early Mars and Recent Venus boundaries adjusted according to Eq. 2 in Kopparapu et al. (2013). A transit probability, P_t, is assigned for each planet, where $P_t = R_s/a$. For each planet, if a random number between
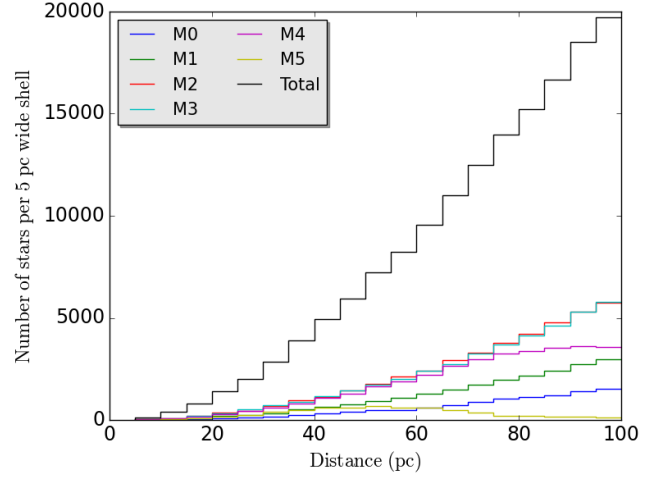

Figure 5. M-dwarfs per subclass per 5 pc wide shell versus distance obtained from the TESS Input Catalogue (Stassun et al. 2018) Candidate Target List 'cool dwarfs' special list that have TESS magnitude $\leq$ 16.

0 and 1 falls below the planet's transit probability it is assumed to be transiting and is included for further consideration. The best-fit spline model of planet size distribution from Fulton et al. (2017) (their Fig. 7) is used as a probability distribution for planet radius. For each transiting planet its radius, R_p, is randomly assigned within the range of radii for its class (sub-Earth, Earth-sized or super-Earth), weighted by this probability distribution. The mass of each planet, M_p, is assigned from mass-radius relation of Otegi et al. (2020) where $M_p = 0.9R_p^{3.45}$. The equilibrium temperature for each planet, T_{eq}, is calculated assuming it is not tidally-locked and has an albedo of 0.3.

Each planet is assumed to have an atmosphere, with a mean molecular weight, μ, which is assigned in two different ways. In the first (type 1) μ is fixed to 29 amu[7] for all planets. In the second (type 2) μ is randomly assigned between 15-45 amu[8]. The scale height of the atmosphere, H, is then given by $kT_{eq}/(\mu g)$, where g is the calculated surface gravity and k is Boltzmann's constant. The period, P, for each planet is calculated from Kepler's third law and the transit duration, $T_{14} = 2(R_p + R_s)/v$, is calculated, where v is the orbital velocity. Circular orbits are assumed, together with inclination angles of 90° in all cases.

6.3 Planet detectability

Next, we must consider the fact that not all transiting planets that exist will be known. The list of known planets at the time of the mission will depend on their previous detection by transit photometric observing campaigns. The following model is used to determine an order of relative detectability for each transiting planet in the model population.

A hypothetical transit photometry survey is considered that scans the entire sky such that each star is observed continuously for the same total observing time, $T_{obs}{}^p$ (where the superscript p denotes photometer). This is a simplification of how all-sky transit surveys actually perform where, in reality, different stars may be observed for

---

[6] Data is obtained from https://filtergraph.com/tess_ctl

[7] The approximate value for the Earth.
[8] Effectively there are two sets of simulations performed: one for fixed μ and one for randomised μ.

different total observing times. The SNR for detection, SNR_{det}, is defined to be the ratio of the photometric fractional transit depth due to the planet, $\tau^P = (R_p/R_s)^2$, to the noise on this transit depth, σ_{τ^P}, and will differ for each planet. σ_{τ^P} will decrease with the square root of the number of transits observed (N_t^P) within the total observing time. I assume $N_t^P = T_{obs}^P/P$. Each planetary transit observation is modeled as a 'box-car'[9] with an in-transit duration, T_{14}, and an out-of-transit (OOT) duration, P-T_{14}. A 100% efficient duty cycle is assumed. If the photometric observatory has a collecting area, A_{tel}^P, and photon conversion efficiency, κ^P, then assuming photon noise only (and $\tau^P << 1$), we can estimate σ_{τ^P} as:

$$\sigma_{\tau^P} \approx \left[T_{obs}^P \kappa^P A_{tel}^P \int_{\lambda_0}^{\lambda_1} F_{tel}(\lambda) \frac{\lambda}{hc} d\lambda \left(\frac{T_{14}[P - T_{14}]}{P^2} \right) \right]^{-1/2} \quad (4)$$

where h and c are Planck's constant and the speed of light respectively, and $F_{tel}(\lambda)$ is the stellar flux density received at the telescope, which is integrated over a wavelength passband ($\lambda_0 - \lambda_1$). SNR_{det} can then be estimated as the product of two factors: α, a factor that is instrument- and observation-dependent, and β, a factor that depends on the planet and star parameters alone:

$$SNR_{det} \approx \alpha\beta \quad (5)$$

$$\alpha = \sqrt{\kappa^P A_{tel}^P T_{obs}^P} \quad (6)$$

$$\beta = \left(\frac{R_p}{R_s} \right)^2 \left[\int_{\lambda_0}^{\lambda_1} F_{tel}(\lambda) \frac{\lambda}{hc} d\lambda \left(\frac{T_{14}[P - T_{14}]}{P^2} \right) \right]^{1/2} \quad (7)$$

If we assume α is a constant for all planets for a given survey, then the relative detectability is given by β, which can be calculated for each planet. I assume here that the wavelength passband ($\lambda_0 - \lambda_1$) is between 0.5-1.0 μm (since most photometric surveys use the optical range) and I approximate F_{tel} with a blackbody function so that $F_{tel}(\lambda) = \pi B_\lambda(T_s)(R_s/d)^2$. The relative photometric detectability is used to select planets for inclusion in the mission schedule in order of their detectability. For example, if 100 planets are to be included in the mission schedule, the 100 most detectable planets are chosen for inclusion.

6.4 Mission duration

For N_{samp} = 10, 20, 30, 40, 50, 60, 70, 80, 90 and 100 planets and D_{tel} = 10, 20, 30, 40 and 50 m, T_{miss} is found using the following methodology.

6.4.1 Atmospheric spectral feature detection

For each planet we must first calculate the wavelength-dependent SNR for one transit, $SNR_1(\lambda)$, in each channel for detection of a 'typical' atmospheric spectral feature targeted in each channel. The height and width of a typical spectral feature is estimated as follows.

I utilise the publicly-available model transmission spectra produced by Wunderlich et al. (2019) for an Earth-analogue planet orbiting stars of different M-dwarf subclasses[10]. These spectra include the impact of photochemistry and cover the full wavelength range of the DRAKE instrument. They are also provided in units of effective height (in km) which allows conversion to dimensionless units

of scale height. These spectra were downloaded for the following M-dwarf subclasses: M1 (GJ 832), M2 (GJ 176, GJ 581,) M3 (GJ 436, GJ 644, AD Leo, GJ 667C), M4 (GJ 876, GJ 1214) and M5 (Proxima Centauri). Each spectrum was resampled to standardised wavelength grid and divided into the Channel A and Channel B wavelength ranges. In the Channel A range the following ro-vibrational spectral features were used to derive height and width: 0.9 μm CH_4, 1.2 μm CH_4/H_2O, 1.4 μm CH_4/H_2O, 1.7 μm CH_4, 2.3 μm CH_4, 2.7 μm CO_2/H_2O, 3.3 μm CH_4, 4.3 μm CO_2 and 4.7 μm CO. In the Channel B range the following features were used: 5.9 μm H_2O, 7.7 μm CH_4 and 9.5 μm O_3. The central wavelengths given above are the weighted averages from the estimated extent of each feature.

The average peak effective height, h_{peak}, for the features in each channel were determined for each spectrum per host star. These were further averaged for stars within each M-dwarf subclass, to give an average h_{peak} for each subclass (Table 1).

Next, for each feature in each spectrum, the full-width half-maximum (FWHM) was estimated as follows. In general, the boundary wavelengths of the feature at half-maximum (λ_1 and λ_2) were found and this span taken to be the FWHM. In some cases (due to overlapping features) only one side of the spectral feature is able to yield a measurable boundary wavelength (e.g. λ_1 or λ_2 but not both). In such cases, I take the difference between the above central wavelength and the measurable boundary wavelength and double it to get a FWHM. For each channel, the average FWHM is obtained for each star, and then averaged again for each M-dwarf subclass. Since the averaged FWHM estimates were similar across the different subclasses, these were further averaged over all subclasses to give a single value for each channel: 0.3 μm for Channel A and 1.2 μm for Channel B[11].

Next, I assume that the typical spectral feature is sampled at least twice over its FWHM to properly sample its shape, giving a maximum spectral bin size, $\Delta\lambda_A$, of 0.5 FWHM. For each channel and M-dwarf subclass, the feature is modelled as a Gaussian function of height h_{peak} (Table 1) and standard deviation = FWHM/2.355 (where the average FWHM estimates for Channel A and Channel B are used). Assuming the above spectral bin is centred over the peak, the average height within the bin, h_{bin}, is then calculated (and will be slightly lower than the peak value). Since the planets in Wunderlich et al. (2019) are modelled as Earth analogues with N_2-O_2-dominated atmospheres, I make the assumption that the scale heights are the same as for the Earth at 8.5 km. Dividing h_{bin} in km by 8.5 km, we obtain n_H, the amplitude in units of the scale height (Table 1). This method considers only a single spectral feature in isolation and does quantify the impact of adjacent spectral features, any gaseous continuum, clouds or haze.

Next, I assume the fractional transit depth A caused by the target feature is given by:

$$A(\lambda) \approx \frac{2n_H(\lambda)HR_p}{R_s^2} \quad (8)$$

where n_H varies according the host star spectral subclass (for M0 we use the M1 value) and channel (Table 1) and H varies for each planet according to its T_{eq}, g and μ. I define the SNR for detection of the target feature in one transit observation, $SNR_1(\lambda)$, at any wavelength

Table 1. Typical spectral feature estimates.

Subclass	Channel A 0.6-5 µm			Channel B 5-11 µm		
	h_{peak} (km)	h_{bin} (km)	n_H (H)	h_{peak} (km)	h_{bin} (km)	n_H (H)
M1	35	33	3.9	36	34	4.0
M2	41	39	4.6	42	39	4.6
M3	38	36	4.2	41	38	4.5
M4	47	45	5.2	46	43	5.1
M5	50	48	5.5	47	46	5.4

as:

$$SNR_1(\lambda) = \frac{A(\lambda)}{\sigma_A(\lambda)} \qquad (9)$$

where σ_A is the noise on the measurement of A. I estimate σ_A as follows.

I adopt a simplified representation of the the spectral feature, modelling it as 'box car' of height $A(\lambda)$ and width $\Delta\lambda_A(\lambda)$ (where $\Delta\lambda_A$ = 0.5 FWHM, i.e. 0.15 µm in Channel A and 0.6 µm in Channel B), rising above a flat baseline. If sampled at the intrinsic resolving power of the instrument, R, the number of such samples across the spectral bin of width $\Delta\lambda_A$ is:

$$n_A(\lambda) = \frac{R\Delta\lambda_A(\lambda)}{\lambda} \qquad (10)$$

If the fractional transit depth (due to both atmosphere and planet) at a given wavelength is $\tau(\lambda)$, then the noise on this at the intrinsic resolving power of the instrument is $\sigma_\tau(\lambda)$. Assuming negligible uncertainty on the spectral baseline, the uncertainty on A is then approximately:

$$\sigma_A(\lambda) \approx \frac{\sigma_\tau(\lambda)}{\sqrt{n_A(\lambda)}} \qquad (11)$$

To find σ_τ I assume that each transit observation consists of an in-transit period, T_{14}, and an out-of-transit period equal to xT_{14}. As in the previous detection model, I assume a 'box car' model of the transit, with 100% duty cycle efficiency and only photon noise included. The photon-conversion efficiency, κ, is set to 0.48. The total collecting area is given by $A_{\text{tel}} = \pi(D_{\text{tel}}/2)^2$. For these calculations D_{tel} was set to 30 m, and the SNR results subsequently scaled for different values of D_{tel} as described later. The total number of photoelectrons per spectral resolution element per transit observation, $S(\lambda)$, is then given by:

$$S(\lambda) = \kappa A_{\text{tel}} F_{\text{tel}}(\lambda) \frac{\lambda}{hc} \frac{\lambda}{R} ([1 - \tau(\lambda)]T14 + xT14) \qquad (12)$$

where $F_{\text{tel}}(\lambda)$ is the flux density received at the telescope calculated using a blackbody function, so that $F_{\text{tel}}(\lambda) = \pi B_\lambda(T_s)(R_s/d)^2$, and λ/R is the width of the spectral resolution element. Since $\tau << 1$ we can simplify this to:

$$S(\lambda) \approx \kappa A_{\text{tel}} F_{\text{tel}}(\lambda) \frac{\lambda}{hc} \frac{\lambda}{R} (1 + x)T14 \qquad (13)$$

The noise on the transit depth, σ_τ, is then given by:

$$\sigma_\tau(\lambda) = \frac{1}{\sqrt{S(\lambda)}} \frac{1 + x}{\sqrt{x}} \qquad (14)$$

In these simulations I adopt $x = 4$. This gives a long baseline per transit to reduce photon noise (assuming a very low noise floor), and thus

the uncertainty on the transit depth, and also for the characterisation and correction of any stellar variability.

For each planet the minimum SNR_1 across both channels is obtained, SNR_{1min}, and used to calculate the number of transits, N_t. Assuming that $\sigma_\tau(\lambda)$ falls with the square root of the number of transit observations and a 3σ detection threshold [as used in Rauer et al. (2011)]:

$$N_t = \left(\frac{3}{SNR_{1min}}\right)^2 \qquad (15)$$

Thus when this number of transit observations have been co-added for a given planet, the target spectral feature will be detectable at an SNR of 3 or more in every spectral bin in boths channels. Once N_t has been calculated for each transiting planet we proceed to a scheduling algorithm.

6.4.2 Scheduling

While each planet needs to complete its required number of transit observations (N_t) the scheduling of these must take into account the transit ephemerides of other planets in the sample. The schedule must avoid any overlaps in observation times between different planets and must minimize the the total time taken to complete all required transit observations in the sample.

The scheduling is performed for two different conditions. The first considers an 'unrestricted' FOR so that all planets in the sample are observable at all times. This would approximate the situation for a lunar polar telescope or novel deep-space observatory design at L2 operating in the shadow of a detached Sun-shield as described previously. The second considers a 'restricted' FOR and applies for an observatory at L2 with an attached Sun-shield and an architecture that permits observations from 90-180° in solar elongation, giving 50% sky instantaneous sky coverage, centred on the anti-Sun axis. This compares to 85-135° solar elongation for JWST, which gives it 39% instantaneous sky coverage. The direction of the centre of the FOR changes with time and has a period of 365.25 days. As the observatory orbits the Sun, different planets in the sample will be inside or outside the the FOR.

For each realisation the scheduling algorithm functions as follows. For a given planet, the total duration of one transit observation equals $(1 + x)T_{14}$. For each planet, the start time for the first transit observation is chosen randomly from within its period, the first central transit time being $0.5(1 + x)T14$ longer than this. Each planet is also randomly assigned an ecliptic longitude. For the restricted FOR case, at any time, the observatory has an ecliptic longitude, θ, i.e. the direction of the centre of the FOR. Starting with $\theta = 0°$, the rate of change of θ is 360/365.25 degrees per day. Only planets with ecliptic longitudes within $\pm90°$ of θ will be viewable at any time.

The algorithm proceeds as follows for the unrestricted FOR case. An initial timeline is constructed of length = 5 × the maximum value of $N_t \times P$ in the sample. For each planet, all its potential transit observations are initially marked on this timeline (start and end times of each transit observation). The planet with the longest period is then chosen as the first 'reference' planet. The reference planet transits are truncated at N_t transits (i.e. all transits > N_t are removed from the timeline). The reference planet's remaining transit observation times are compared with each of the other planets in order of decreasing period. Any overlapping transit observations are deleted from the timelines of the shorter period planets, such that the end of this process the reference planet should have no observing clashes with any of the other planets. The planet with the next longest period now becomes the reference planet and the process is repeated,

truncating its transits to N_t, comparing its timeline to those of shorter period planets and removing clashes. This process is repeated with the reference planet chosen in order of period until the shortest period planet becomes the reference planet. At that stage there should be no clashing observations and all planets will have N_t transit observations in their timelines.

For the restricted FOR case the same algorithm is used with the following modification. The periods where the planet will be out of the instantaneous FOR are marked on the timeline for each planet and any transit observations that are in or cross into the out-of-view periods are deleted from the timeline. The algorithm then proceeds as for the unrestricted FOR case, so that the final schedule has the required number of transit observations for each planet, N_t, there are no clashes between planets, and all scheduled observations will be within the instantaneous FOR of the observatory.

Once the algorithm is completed, the mission is fully scheduled, and the mission duration, T_{miss}, is given by the time when all planets have completed their transit observations. The total mission time does not include any time for commissioning, and it is assumed that slewing and housekeeping functions are accommodated in the time between active observations.

6.4.3 Type 1 and type 2 simulations

We perform scheduling and obtain T_{miss} results independently for the cases where μ is fixed (type 1) and for where μ is random (type 2). In the calculation of N_t (and thus the scheduling and calculation of mission duration) μ is the only factor that cannot be reasonably estimated a priori. Hence the type 1 case reflects a mission scheduled without a priori knowledge of μ. The type 2 simulations produce a schedule based on the actual μ of each planet and is thus closer to the 'ground truth'. Despite this, the type 2 schedule would not be easy to produce in reality (i.e. in planning the actual mission) relying as it does on knowing the actual μ for each planet. We could envisage a situation however where, starting with a schedule that has an indefinite number of transits per planet, as the mission proceeds the atmospheric SNR for each planet is determined experimentally, and when the goal SNR is reached it is deemed to have completed N_t transits and its remaining transits removed from the schedule. The schedule is then recalculated to account for the removed observations. N_t is effectively found 'experimentally' as the spectra are analysed. The final schedule thus obtained should then approximate that simulated here. We can compare the results for both type 1 and type 2 schedules in this study to see if the more practical type 1 schedule gives results close to the more realistic type 2 schedule.

6.4.4 N_{pool}

It was found that simply scheduling a sample of N_{samp} planets resulted in the small number planets of the sample pushing up the mission duration disproportionately. This effect is shown in Figure 6 (upper plot) for $N_{samp} = 50$ and $N_{samp} = 100$ using $D_{tel} = 20$ m with an unrestricted FOR and type 2 simulations. In each of the 100 realisations of the final schedule, the planets are ordered by the mission elapsed time at which they complete their required number of transits, N_t. The mean time for the nth planet in the sequence is then obtained and plotted as shown in Figure 6 (upper plot). The time for the final planet gives the mission duration, T_{miss}. Initially the mission elapsed time increases with planet in a near linear way, however for the final few planets in the sequence the mission elapsed times are disproportionately higher leading to an 'uptick' in the curve

that pushes up the final T_{miss}. This effect occurs also in the restricted FOR schedule simulations.

Considering the unrestricted FOR case, the minimum time to complete all of a planets observations is $N_t \times P$, although this may be increased by the scheduling algorithm if there are clashing transit observations with other planets. In any sample there is a range of $N_t \times P$. It is the small number of high $N_t \times P$ planets (toward the end of the sequences shown) that tend to push up the mission time disproportionately in the context of the scheduling algorithm.

If the sample size is expanded, we bring in some new planets which have lower $N_t \times P$ than the highest $N_t \times P$ planets in the smaller sample. Therefore in the expanded sample the nth planet may be completed sooner than than in the smaller sample, since the nth planet has a lower $N_t \times P$ in the expanded sample. This is particularly noticeable when n is close to N_{samp} of the smaller sample. This can be seen in 6 (upper plot) where the 50th planet is completed earlier in the $N_{samp} = 100$ sample than the $N_{samp} = 50$ sample. This 'uptick' effect thus has the potential to significantly increase T_{miss} to the detrimental of overall mission feasibility.

To mitigate this, we can instead initially schedule more planets than N_{samp} itself and cut off the mission when N_{samp} planets have completed their observations. This way the best performing planets are selected from a slightly larger 'pool' of planets of size N_{pool}. In Figure 6 (lower plot) we used $N_{pool} = 1.2 \times N_{samp}$. For example, in a mission with $N_{samp} = 50$, we initially schedule $N_{pool} = 60$ planets. We can see that N_{samp} planets are completed just before the 'uptick' effect kicks in, largely mitigating it. This significantly improves T_{miss} (at the expense of requiring more planets to be known at the start). Given this benefit to mission times, the final scheduling was performed with $N_{pool} = 1.2 \times N_{samp}$ in all cases, and T_{miss} obtained when the nth ($n = N_{samp}$) planet in the sequence completes all its N_t observations.

6.4.5 Varying D_{tel}

To investigate how T_{miss} varies with telescope aperture for different N_{samp} values, I take the SNR_1 values obtained using the simulation with $D_{tel} = 30$ m, for each of the 100 realisations, and scale these to different D_{tel} values ($D_{tel} = 10, 15, 20, 30, 40$ and 50 m) by multiplying the results for the baseline case by $D_{tel}/30$. SNR_{1min} and N_t are obtained for each case, the planets ordered by detectability and then run through the scheduling algorithm for different N_{samp} values as above. For each D_{tel} and N_{samp} combination, 100 values for T_{miss} are thus obtained and the mean and standard deviation found for each case.

6.5 Results

Figure 7 shows T_{miss} vs N_{samp} for different D_{tel} values. The mean results over 100 realisations and 1σ confidence region are displayed in each case. Results are shown for fixed μ (type 1) and randomised μ (type 2) simulations, for both the restricted and unrestricted FOR cases. We can see that type 1 results give slightly shorter T_{miss} values than type 2. As may be expected the restricted FOR results in much longer T_{miss} than the corresponding unrestricted FOR case.

In Figure 8 these T_{miss} results are plotted vs D_{tel} for different sample sizes, N_{samp}. The dots and error bars give the mean and 1σ range respectively from 100 realisations. 3rd order polynomials are fitted in log-log space to the mean values and also to the maximum and minimum values of each error bar. The resulting equations allow us to infer the mean and 1σ range of T_{miss} for D_{tel} values between
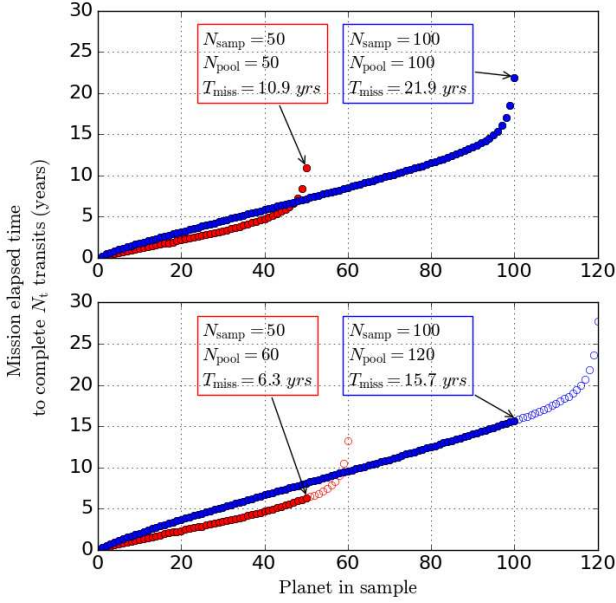
Figure 6. The 'uptick' effect on the mission elapsed time. Results shown for $D_{tel} = 20$ m with a type 2 simulation and an unrestricted FOR. Dots give the average mission elapsed time for the nth planet in the time-ordered sequence of a scheduled sample to complete its N_t transits. A small number of planets at the end of the ordered sequence take up a disproportionate amount of mission time. Upper plot: The effect shown for $N_{samp} = 50$ and $N_{samp} = 100$. Lower plot: Mitigation of this effect by including N_{pool} planets in the initial schedule but completing mission when N_{samp} is achieved, reducing mission duration. In this case $N_{pool} = 1.2 \times N_{samp}$.

and beyond those obtained directly through the simulations. We use these to obtain average T_{miss} values over a finer D_{tel} grid (with a gradient of 0.1 m) giving the dotted lines shown. The corresponding 1σ ranges are shown by the shaded areas. The mean and 1σ range lines are also extended to smaller D_{tel} values down to 8 m as shown. These lines using the finer D_{tel} grid are then used to generate the contour plots[12] in Figures 9 (unrestricted FOR) and 10 (restricted FOR). These allow T_{miss} to be estimated for any combination of N_{samp} and D_{tel}. The regions for $T_{miss} < 10$ years and > 10 years are demarcated. I adopt 10 years as maximum viable mission duration for a mission at L2 (assuming no maintenance of the observatory).

Results for selected cases are summarised in Table 2, combined with CI estimates from Section 4.2. Four observatory types are considered: 1) an L2 observatory with restricted FOR, 2) an L2 observatory with unrestricted FOR, 3) a lunar observatory at one pole, 4) two lunar observatories, one at each pole. The results obtained here for the restricted FOR and unrestricted FOR are directly applicable to the first and second L2 types respectively. For two lunar telescopes we can also use the unrestricted FOR results. In Table 2 the minimum number of planets needed, N_{min}, takes into account the fact that both the type 1 and type 2 schedules are based on selecting the best performing planets from a pool of planets, N_{pool}, of size $1.2 \times N_{samp}$. In most of these cases $N_{min} = N_{pool}$. For a single lunar telescope we can use the unrestricted FOR results with the caveat that since only half the sky will be observable $N_{min} = 2 \times N_{pool}$ assuming an even distribution over the whole sky.

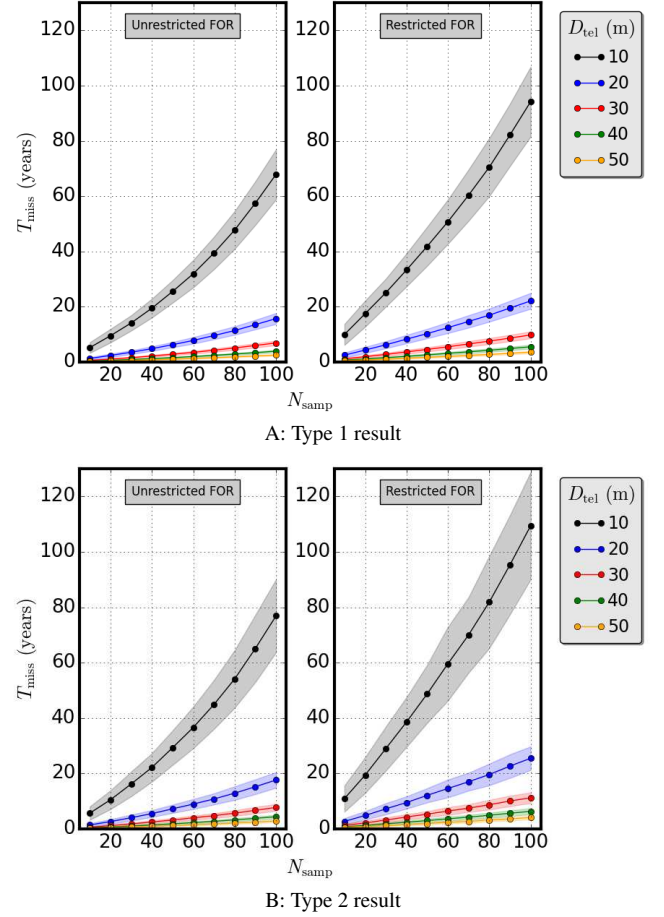[12] The matplotlib 'contour' function is used.

Figure 7. Mission duration, T_{miss}, vs sample size, N_{samp}, for different primary mirror sizes, D_{tel}. Dots show mean results obtained over 100 realizations, with 1σ range shown as shaded area. A: Type 1 simulations. B: Type 2 simulations. Left-sided plots: Unrestricted field-of-regard. Right-sided plots: restricted field-of-regard.

These results of course reflect the instrument design chosen, the various assumptions made for the the modeling, the metrics for detection (including the need to achieve adequate SNR out to 11 μm), and the efficiency of scheduling. There is quite a bit of variability around the mean cases as shown by the right-hand plots in Figures 9 and 10. Thus depending on the exact details of the actual planet population that will be available at the time of the mission, the final mission times may be greater or less than the predicted mean results.

6.5.1 10-m telescope

For the unrestricted FOR, if we look at the lower 1σ limit results (the optimistic limit in terms of mission duration) we find that it is unlikely that any combination of $N_{samp} > 27$ and $D_{tel} < 10$ m will result in a mission of < 10 years duration using the type 1 results (Figure 9 A lower right). On average a 10-m telescope delivers a 19 (type 1 result) to 21 (type 2 result) planet survey in 10 years (Figure 9 A left and B left).

For the restricted FOR, with increased mission times, the yields are lower for the same duration. Taking the most optimistic lower 1σ limit, both the type 1 and type 2 results indicate that it is unlikely that any combination of of $N_{samp} > 16$ and $D_{tel} < 10$ m will result in a mission of < 10 years duration (Figure 10 A lower right and
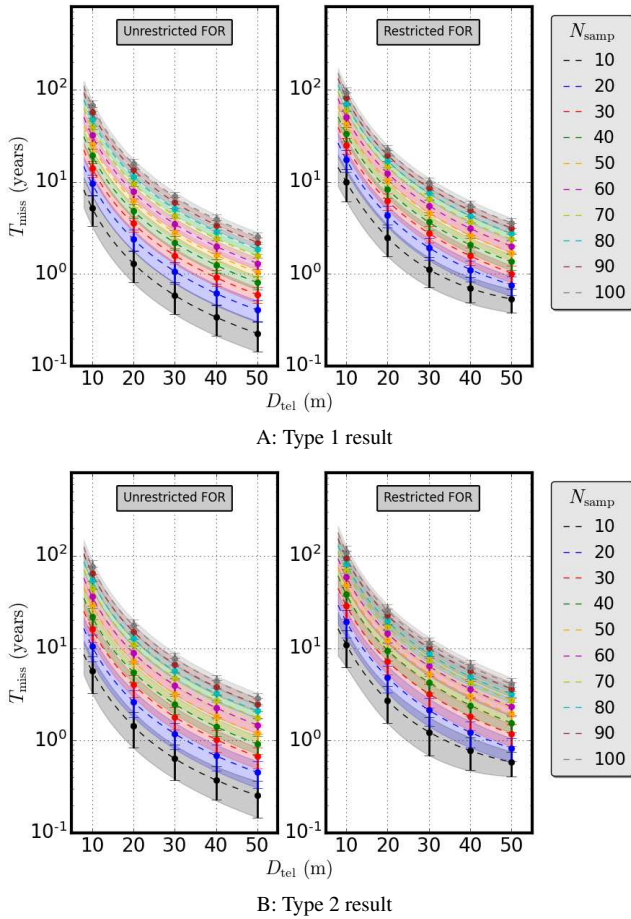
Figure 8. Mission duration, T_{miss}, vs primary mirror size, D_{tel}, for different sample sizes, N_{samp}. Dots show the mean result over 100 realisations and error bars show 1σ range. Polynomial fits to the mean values are shown with the dotted lines. Polynomial fits to the maximum and minimum error bar values are shown by the edges of the shaded regions. A: Type 1 simulations. B: Type 2 simulations. Left-sided plots: Unrestricted field-of-regard. Right-sided plots: restricted field-of-regard.

B lower right). The average results indicate that a 10-m telescope would deliver a 10-planet survey in 10 years going by the type 1 result (Figure 10 A left) or an 8-planet survey by the type 2 result. The latter value was obtained by fitting a 2nd order polynomial to points along the 10-year demarcation line (Figure 10 B left) and using this to predict the result falling just outside the boundaries of the chart.

6.5.2 20-planet survey

Considering $N_{samp} = 20$, we can see from Table 2 that 15-m class telescopes are very likely to be able to complete a 20-planet survey within 10 years. Taking the worst case result, the L2 observatory with restricted FOR and type 2 simulation, we find that a 10-year mission is achieved with a 14.0 (+2.1/-2.4) m telescope. 15-m telescopes could complete the 20-planet survey in anything from 4.3 (±1.1) years to 8.7 (±2.7) years depending on the observatory type. 10-m class telescopes might be able to achieve a 20-planet survey in 10 years in the best case scenario, using the unrestricted FOR L2 observatory or the lunar telescope cases. A survey of this size could constrain the f_L 95% CI size from 0.13 to 0.4 depending on the observed value.

6.5.3 15-m telescope

If we consider what a 15-m telescope can achieve in a 10 year mission, we find that using the slightly more pessimistic type 2 results it can complete an N_{samp} of 23 (+8/-6) planets for the restricted FOR L2 observatory, increasing to 40 (+9/-7) for the unrestricted FOR cases. If we consider an f_L[obs] of 0.05, i.e. one planet in the survey appears positive for life, the 23 planet survey result would constrain the upper limit of f_L to to ~ 0.19 at 95% confidence, while 40 planets would constrain this to ~ 0.15. The lower bounds in each of the previous cases are <0.01. Thus there is only a modest improvement in f_L precision between the two cases, so that a 15-m telescope in a 'traditional' L2 configuration with restricted FOR performs close to one with unrestricted FOR in this regard.

6.5.4 50-planet survey

However, when we increase the sample size to 50, 15-m class telescopes are much less likely to succeed within 10 years. Table 2 shows that D_{tel} would need to be 15.9 m (+1.1/-1.3) in the best case scenario, type 1 with an unrestricted FOR. If we go by the more pessimistic type 2 results of $D_{tel} = 17.1(+1.6/-1.8)$ m, we can say that an N_{samp} of 50 probably calls for at least a 17-m telescope. Similarly looking at the type 2 result for the restricted FOR case, $D_{tel} = 22.0$ (+2.1/-2.4) m, a 22-m telescope is probably required for $N_{samp} = 50$ in this situation. A 50-planet survey could constrain the upper limit of f_L to ~ 0.06 at 95% confidence if the sample was completely negative, or between ~ 0.03-0.2 if f_L[obs] was 0.1 (i.e. 5 positive planets in the sample).

6.5.5 100-planet survey

Finally, Table 2 summarises the results for a 100-planet survey. This sample size demands D_{tel} of between ~ 25 to 32 m for 10 year missions depending on the simulation type and the field-of-regard. However, given the predictions of TESS yields (Barclay et al. 2018; Sullivan et al. 2015), it is possible that very large sample sizes (>> 50) may simply not exist in the coming decade. Also, as shown earlier, the incremental improvement on the CI of f_L[obs] falls with sample size. Table 2 shows only a modest reduction of CI size for $N_{samp} = 100$ compared to $N_{samp} = 50$. For these reasons planning a mission around a 50-planet survey may be optimal.

6.5.6 Total observing time

The total observing time, T_{obs}, is the sum of all the time spent observing all of the planets in the sample. For each planet this equates to $(1 + x)T_{14} \times N_t$.

In Figure 11 we plot the ratio of mean T_{obs} over 100 realisations to the mean T_{miss} over 100 realisations vs D_{tel} for different N_{samp} values. This serves as a measure of the efficiency of the mission in terms of observing time. The unrestricted FOR case is more efficient since T_{miss} values are smaller compared to the restricted FOR. The type 2 simulations return slightly less efficient missions than type 1 which may be related to the longer T_{miss} times for the type 2 simulations. Efficiency improves with N_{samp} and generally with larger D_{tel} (although this trend is less apparent in the restricted FOR cases for $N_{samp} = 10$). A wide range of efficiency is seen ranging from ~ 5% (for $N_{samp} = 10$ in the type 2 restricted FOR case) to ~ 41% (for $N_{samp} = 100$ in the type 1 unrestricted FOR case with a D_{tel} of 50 m). If we consider a mission with a goal N_{samp} of 50, then if we use a 17-m telescope with an unrestricted FOR, e.g. the novel
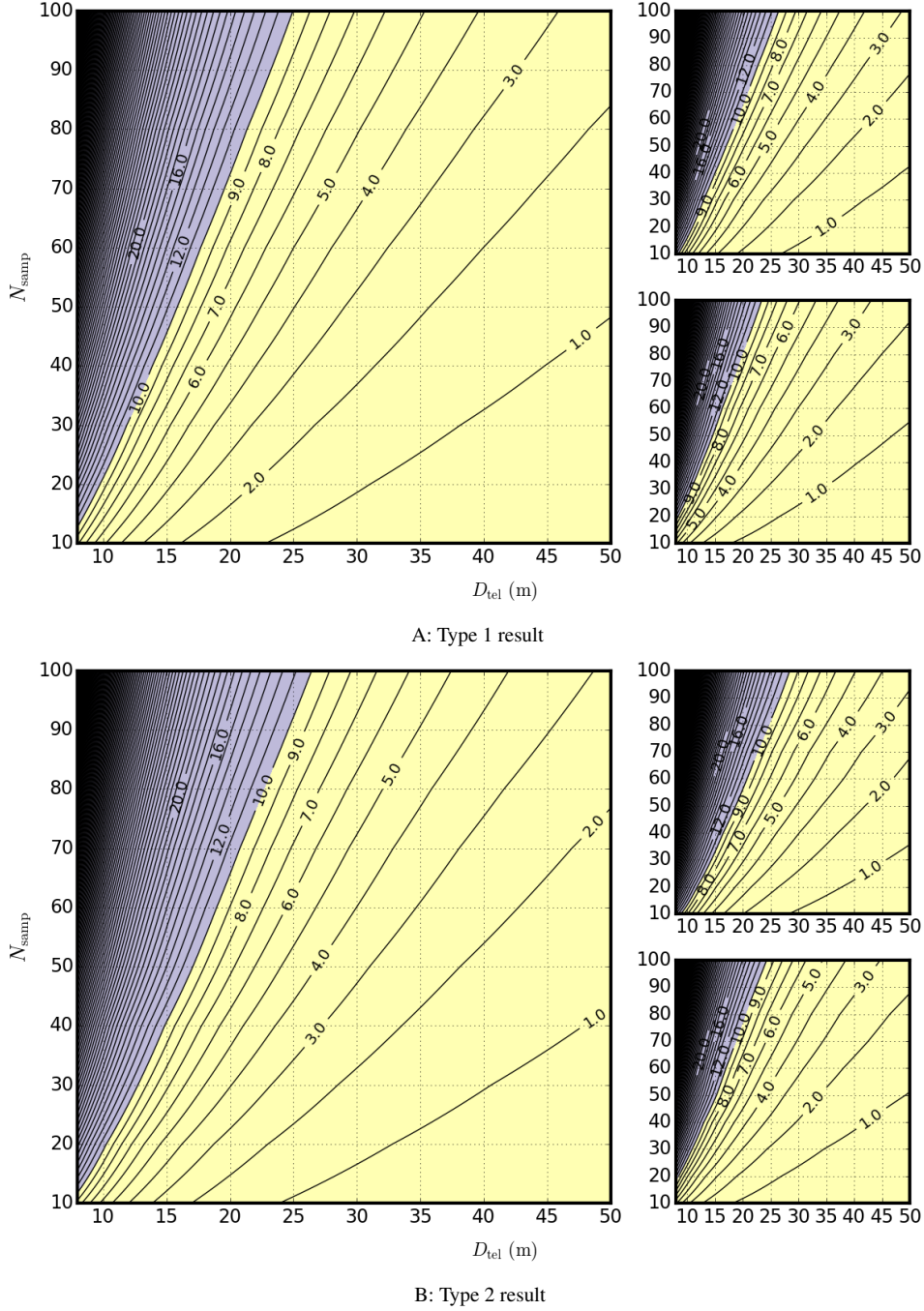
Figure 9. Contour plots of mission duration, T_{miss}, (in years) vs primary mirror size, D_{tel}, and sample size, N_{samp}, for the unrestricted field-of-regard case. A: Type 1 simulation results. B: Type 2 simulation results. The yellow-shaded region indicates a mission duration < 10 years and the purple-shaded region indicates a mission duration > 10 years. Left-sided plots: mean result. Smaller upper right plots: upper 1σ limit. Smaller lower right plots: lower 1σ limit.

L2 design or the lunar telescope cases, the efficiency is around 28% from type 1 simulations or 24% from type 2 simulations. If we consider a more traditional L2 setup with a restricted FOR then using a 22-m telescope the efficiency for N_{samp} = 50 is 18% from the type 1 simulation or 15% from the type 2 simulation. Also if the amount of OOT time per observation xT_{14} (where $x = 4$ for these simulation) is reduced then the efficiency will drop further. While some of the off-target time will be needed for housekeeping, calibration and slewing time, these ratios of T_{obs} to T_{miss} indicate a low efficiency in terms

of utlising the full mission time. This is an inevitable limitation of using the transit spectroscopy approach since long stretches of time may elapse while waiting for the next available transit, and the active observing time is a small fraction of a planet's period. The time between observations could of course be used for alternative science goals. Therefore these results indicate that the DRAKE mission should probably be an integral part of a more wide-ranging observatory similar to NASA-sponsored Decadal Survey Mission Concept Studies such as the Origins Space Telescope (Battersby et al. 2018).
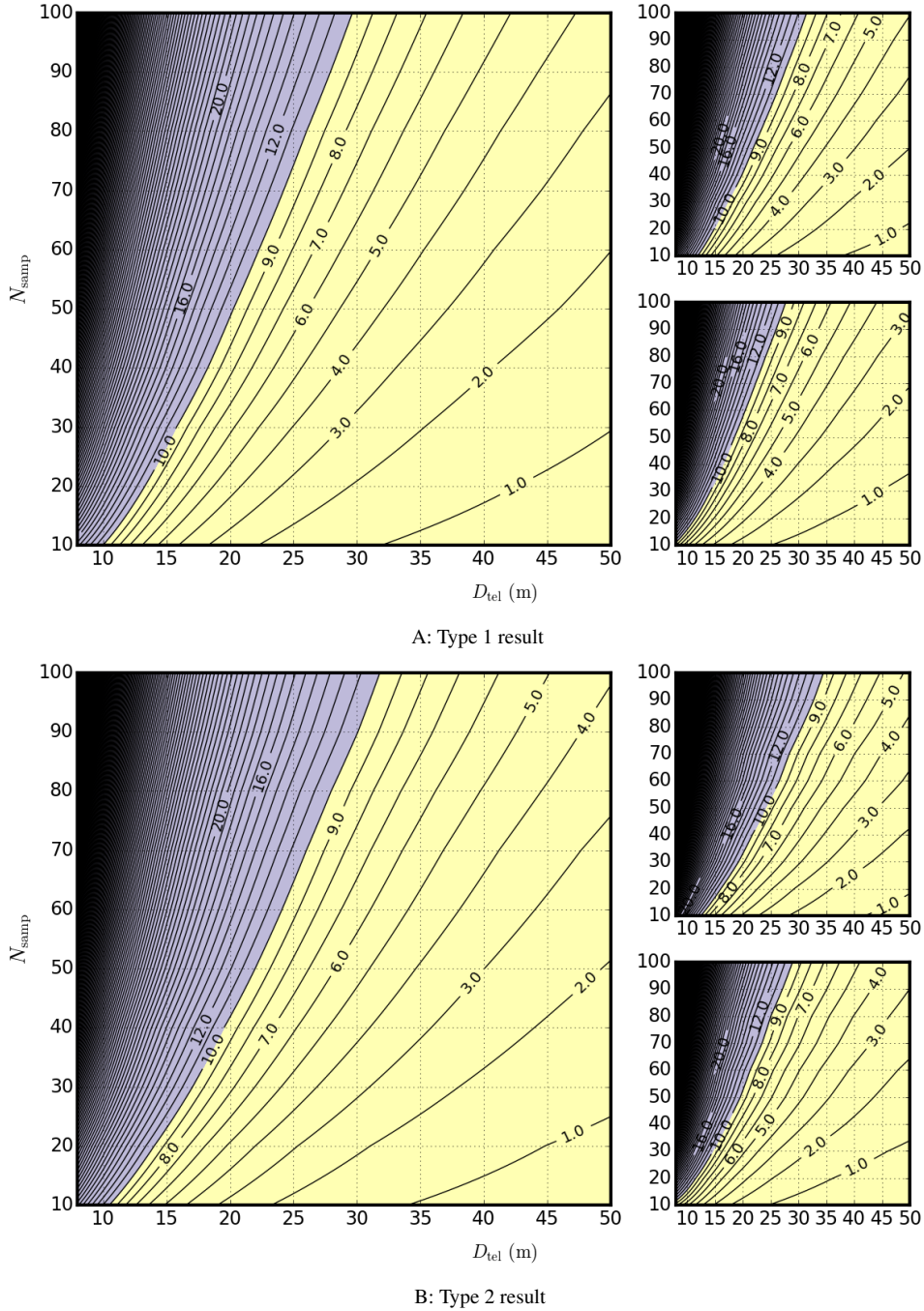
Figure 10. Contour plots of mission duration, T_{miss}, (in years) vs primary mirror size, D_{tel}, and sample size, N_{samp}, for the restricted field-of-regard case. A: Type 1 simulation results. B: Type 2 simulation results. The yellow-shaded region indicates a mission duration < 10 years and the purple-shaded region indicates a mission duration > 10 years. Left-sided plots: mean result. Smaller upper right plots: upper 1σ limit. Smaller lower right plots: lower 1σ limit.

It would also be consistent with the themes of ESA Voyage 2050 (Favata et al. 2021) and could be a science driver for a new space- or Moon-based observatory within that programme.

7 CONCLUSIONS

In this paper I discussed the occurrence rate of life-bearing planets, termed 'the frequency of life': the goal of the DRAKE mission. This is parameterised as f_{L} in the Drake and Seager equations. It depends

in turn on how the habitable zone itself is defined to give a subset of 'candidate' planets. In this study we have chosen the optimistic habitable zone limits of Kopparapu et al. (2013) but with other definitions, different results will be obtained. There might also be planets where biosignature gases are not observable and yet life exists (e.g. in subsurface oceans). Since there is no information on the rate of such false negatives, these have been ignored in this study, as has the impact of the star on falsifying the detection of a true biosignature. Thus the observable f_{L} considered here may underestimate the true

Table 2. Summary of findings for selected cases. For each N_{samp} the corresponding 95% CI is given for three different observed f_L values. The minimum number of known candidate planets, N_{min}, needed is given for each observatory type. The minimum telescope primary mirror diameter, D_{tel}, for a 10-year mission is given for each case, as well as mission duration, T_{miss}, for 15-m, 22-m and 30-m telescopes. Bracketed values give the 1σ range.

N_{samp}	f_L[obs] 0.0	95% CI f_L[obs] 0.1	f_L[obs] 0.5	Observatory type	N_{min}	D_{tel} (10 yr) (m) Type 1 Type 2	T_{miss} (15-m) (yrs) Type 1 Type 2	T_{miss} (22-m) (yrs) Type 1 Type 2	T_{miss} (30-m) (yrs) Type 1 Type 2
20	0-0.13	0.02-0.28	0.3-0.7	L2 (Unrestricted FOR)	24	9.8 (+1.2/-1.4) 10.3 (+1.5/-1.8)	4.3(±1.1) 4.7(±1.5)	2.0(±0.5) 2.2(±0.7)	1.1(±0.3) 1.2(±0.4)
				L2 (Restricted FOR)	24	13.3 (+1.6/-1.9) 14.0 (+2.1/-2.4)	7.8(±2.0) 8.7(±2.7)	3.6(±0.9) 4.0(±1.3)	1.9(±0.5) 2.1(±0.7)
				Lunar (1 pole)	48	9.8 (+1.2/-1.4) 10.3 (+1.5/-1.8)	4.3(±1.1) 4.7(±1.5)	2.0(±0.5) 2.2(±0.7)	1.1(±0.3) 1.2(±0.4)
				Lunar (2 pole)	24	9.8 (+1.2/-1.4) 10.3 (+1.5/-1.8)	4.3(±1.1) 4.7(±1.5)	2.0(±0.5) 2.2(±0.7)	1.1(±0.3) 1.2(±0.4)
50	0-0.06	0.03-0.20	0.36-0.63	L2 (Unrestricted FOR)	60	15.9 (+1.1/-1.3) 17.1 (+1.6/-1.8)	11.2(±1.7) 13.0(±2.6)	5.2(±0.8) 6.0(±1.3)	2.8(±0.4) 3.2(±0.7)
				L2 (Restricted FOR)	60	20.3 (+1.6/-1.8) 22.0 (+2.1/-2.4)	18.4(±3.2) 21.7(±4.7)	8.5(±1.4) 10.0(±2.1)	4.6(±0.7) 5.3(±1.1)
				Lunar (1 pole)	120	15.9 (+1.1/-1.3) 17.1 (+1.6/-1.8)	11.2(±1.7) 13.0(±2.6)	5.2(±0.8) 6.0(±1.3)	2.8(±0.4) 3.2(±0.7)
				Lunar (2 pole)	60	15.9 (+1.1/-1.3) 17.1 (+1.6/-1.8)	11.2(±1.7) 13.0(±2.6)	5.2(±0.8) 6.0(±1.3)	2.8(±0.4) 3.2(±0.7)
100	0-0.03	0.05-0.17	0.40-0.60	L2 (Unrestricted FOR)	120	25.0 (+1.4/-1.6) 26.4 (+2.0/-2.1)	28.5(±3.7) 32.2(±5.2)	12.9(±1.6) 14.5(±2.3)	6.9(±0.8) 7.8(±1.2)
				L2 (Restricted FOR)	120	29.6 (+1.9/-1.9) 31.8 (+2.7/-2.9)	40.1(±5.3) 46.2(±7.6)	18.2(±2.3) 21.0(±3.5)	9.8(±1.2) 11.2(±1.9)
				Lunar (1 pole)	240	25.0 (+1.4/-1.6) 26.4 (+2.0/-2.1)	28.5(±3.7) 32.2(±5.2)	12.9(±1.6) 14.5(±2.3)	6.9(±0.8) 7.8(±1.2)
				Lunar (2 pole)	120	25.0 (+1.4/-1.6) 26.4 (+2.0/-2.1)	28.5(±3.7) 32.2(±5.2)	12.9(±1.6) 14.5(±2.3)	6.9(±0.8) 7.8(±1.2)

frequency of life, but to what extent is unknown. Even with these limitations on the definition of f_L, an initial result from a spectroscopic survey would still deliver significant insights into how common life is and possibly the planetary basis for abiogenesis.

Using a bootstrap Monte Carlo simulation I explored how the uncertainty (95% CI) on f_L[obs] scales with number of planets in the sample. Given the assumption made here that a biosignature pattern can be correctly identified through observation at 3σ significance, the simulation indicates that sampling error will dominate over this observational error. The uncertainty is a function not only of the sample size, but also of f_L[obs] itself: maximal at f_L[obs] = 0.5, and minimal at f_L[obs] = 0 or 1. Even a completely negative sample (i.e. f_L[obs] = 0) can constrain the upper limit of f_L[true] at 95% confidence. Due to diminishing return in terms of reducing the CI size with N_{samp}, a 50-planet sample size might be optimal as a target. If f_L[obs] = 0, this would constrain f_L[true] to no more than 0.06 at 95% confidence, and between 0.03-0.2 if f_L[obs] = 0.1. The statistical modeling approach presented in this paper can be further developed and refined in future studies.

The DRAKE mission is built around using transit spectroscopy to obtain atmospheric spectra and as such has a number of limitations compared to direct imaging approaches. The first is that any sample will be inevitably biased in terms of stellar type. This is because only M-dwarf HZ planets will be viable for this technique, so that extending the conclusions to Sun-like stars may not be possible. Even among M-dwarfs, the sample may be biased towards early M-

dwarfs with later types possibly under-represented due to their low brightness. Compared to direct imaging approaches, the available sample sizes will be small and the efficiency of observing time to total mission time low. However, transit spectroscopy is a mature technique that would require much less technological development compared to direct imaging in order to observe HZ planets. Although the sample is confined to M-dwarfs, these could be considered a galactic norm representing 70% of all stars, and therefore even if the mission is limited to finding f_L for this subset of systems, its conclusions will still be highly significant. We have shown that small sample sizes can still provide constraints on the 95% CI for f_L and so using transit spectroscopy to find a first experimental estimate for f_L is a viable prospect.

I presented a baseline observatory design and mission plan for DRAKE, and performed a feasibility study. The final results are very sensitive to the assumptions and detection criteria used in this study. N_t for each planet is highly sensitive to any factors that affect the calculation of SNR_1, as well as the choice of goal SNR for detection. These include the choices made in instrument design, as well as star, planet and noise modelling. I find that such a mission is feasible under certain conditions and is therefore a viable alternative to direct imaging approaches. The concept warrants further study with more advanced instrumental design and noise modelling, together with performance metrics based on spectral retrievals.

In the current study, I conclude that for a 50-planet survey, 17-to 22m-class telescopes will most probably be required to achieve
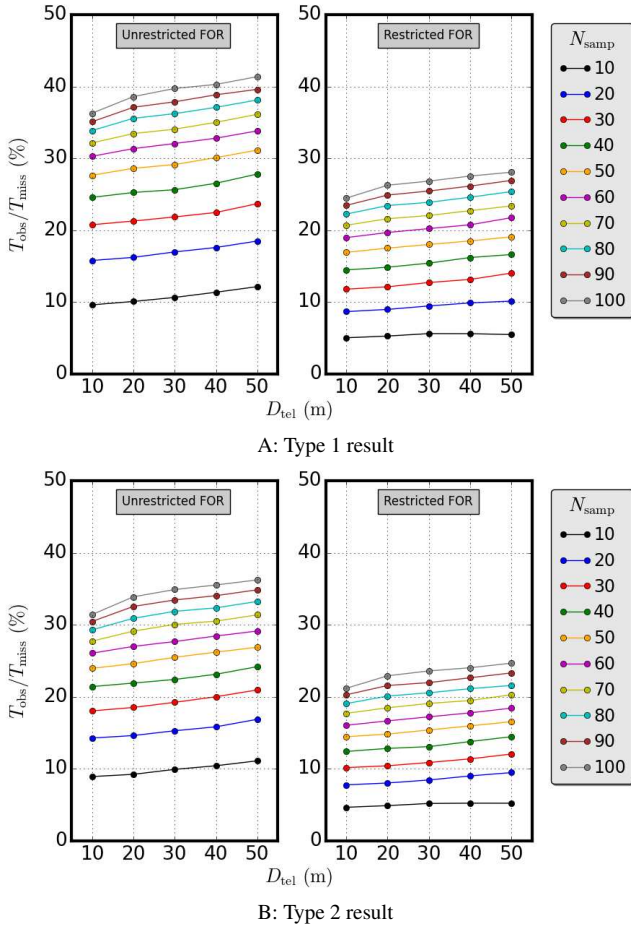
Figure 11. Ratio of total observing time, T_{obs}, to total mission time, T_{miss}, vs primary mirror size, D_{tel}, for different sample sizes, N_{samp}. Dots give the mean result over 100 realisations. A: Type 1 simulations. B: Type 2 simulations. Left-sided plots: Unrestricted field-of-regard. Right-sided plots: restricted field-of-regard.

a mission of less than 10 years duration. The requirements depend on whether the putative observatory has an unrestricted or restricted FOR. For a space-based design with a 50% FOR, e.g. an observatory based at L2, I conclude that a 22-m telescope is likely to successfully achieve complete a 50-planet survey within 10 years. If a novel design permits almost unrestricted FOR at L2 or if we consider lunar polar telescopes, then the size of telescope can be reduced. A 17-m telescope is likely to achieve the mission under such circumstances.

15-m class telescopes are unlikely to achieve a 50-planet survey in 10 years, however we show that on average a 23-planet survey can be achieved in the restricted FOR case, rising to 40 planets in the unrestricted FOR cases. Such surveys have the potential to constrain f_L to < 0.2 at 95% confidence assuming 1 planet is positive in the survey. Such a size of primary mirror has already been proposed for the Luvoir-A mission (The LUVOIR Team 2019). The Luvoir-A High Definition Imager (HDI) has the capability to perform transit spectroscopy in the wavelength range 200 nm to 2.5 μm. This wavelength range is shifted to shorter wavelengths than proposed for DRAKE, and so the feasibility study results obtained in this paper are not directly transferable to Luvoir-A. However, the results obtained here pertaining to sample size and uncertainty on f_L are not exclusively applicable to transit spectroscopy of M-dwarfs or DRAKE, and could be more generally applied to other HZ surveys based either on transit

spectroscopy or direct imaging (so long as definitive biosignature identification can be made).

Lunar-based polar telescopes could be considered for the DRAKE mission, and could utilise 17-m telescopes to achieve a 50-planet survey within 10 years. A single polar telescope would have an unchanging FOR, but to only half the sky. Thus the unrestricted FOR simulations can be applied, but it must be assumed that at least twice as many planets must be known as in the survey. The simulations also assumed that the initial pool of available planets for given survey, N_{pool} was 1.2× N_{samp}. For the single lunar telescope this would mean at least 120 candidate planets (Earth-sized in the HZ of M-dwarfs) must be known. Given the likely yields for TESS and other missions in the coming decade, it seems unlikely at this time that 120 such planets will become known in the next few years. From that standpoint the two-pole lunar telescope might be more feasible since that configuration has an $N_{min} = 60$.

These results are predictions based on Monte Carlo realisations of possible model planet populations. In reality much will depend on the actual final planet sample available at the time of the mission. Ultimately, the DRAKE mission, if given the go ahead, would adapt to the planet population known at the time. This will shape the final expected mission time or the exact diameter of telescope needed. We find here that simulations that assume an Earth-like mean molecular weight, μ, for all planets (type 1) tend to underestimate the mission time slightly compared to simulations where μ was varied randomly (type 2), although in Table 2 most of the type 1 results are within 15% of the type 2 results, and all are within 20%. It may be necessary to have some kind of adaptive scheduling during the real mission based on the likely cumulative SNR achieved for a given planet which will come to approximate the type 2 schedule simulated here.

Due to the low observing efficiency, I conclude that the DRAKE mission might be best incorporated into plans for a wider-ranging observatory. It could be a strong scientific goal for a future L2 or lunar observatory. While the primary goal of DRAKE is to find f_L, it can also be used to survey other habitable zone planet characteristics on a statistical basis, e.g. water and CO_2 abundances as suggested by Bean et al. (2017), providing experimental constraints to the habitable zone itself.

The discovery of the first life on another planet will have major scientific and cultural impacts. However, elucidating how common life is, and the planetary conditions under which it arises, will allow us to probe even deeper questions into the occurrence and origin of life. The DRAKE mission is a viable approach to attempt the first measurement of the frequency of life-bearing planets in the Cosmos.

DATA AVAILABILITY STATEMENT

The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

Bains W., 2004, Astrobiology, 4, 137

Barclay T., Pepper J., Quintana E. V., 2018, The Astrophysical Journal Supplement Series, 239, 2

Barstow J. K., Irwin P. G. J., 2016, Monthly Notices of the Royal Astronomical Society: Letters, 461, L92

Barstow J. K., Aigrain S., Irwin P. G. J., Kendrew S., Fletcher L. N., 2015, MNRAS, 448, 2546

Batalha N. M., 2014, Proceedings of the National Academy of Sciences, 111, 12647

Battersby C., et al., 2018, Nature Astronomy, 2, 596

Bean J. L., Abbot D. S., Kempton E. M. R., 2017, ApJ, 841, L24

Beichman C. A., Woolf N. J., Lindensmith C. A., 1999, The Terrestrial Planet Finder (TPF) : a NASA Origins Program to search for habitable planets

Blank R., et al., 2011, in Kuhn J. R., Harrington D. M., Lin H., Berdyugina S. V., Trujillo-Bueno J., Keil S. L., Rimmele T., eds, Astronomical Society of the Pacific Conference Series Vol. 437, Solar Polarization 6. p. 383

Bouma L. G., Winn J. N., Kosiarek J., McCullough P. R., 2017, Planet Detection Simulations for Several Possible TESS Extended Missions (arXiv:1705.08891)

Brandt T. D., Spiegel D. S., 2014, Proceedings of the National Academy of Science, 111, 13278

Cabrera M. S., McMurtry C. W., Forrest W. J., Pipher J. L., Dorn M. L., Lee D., 2020, Journal of Astronomical Telescopes, Instruments, and Systems, 6, 011004

Cassan A., et al., 2012, Nature, 481, 167

Charbonneau D., Brown T. M., Noyes R. W., Gilliland R. L., 2002, The Astrophysical Journal, 568, 377

Cochran W. G., 1963, Sampling Techniques. John Wiley

Cockell C., et al., 2009, Astrobiology, 9, 1

Cooper W. J., Cooper D. J., Saltzman E. S., Mello W. Z. d., Savoie D. L., Zika R. G., Prospero J. M., 1987, Atmospheric Environment, 21, 1491

Domagal-Goldman S. D., Segura A., Claire M. W., Robinson T. D., Meadows V. S., 2014, ApJ, 792, 90

Drake F., 2015, in Vakoch D. A., Dowd M. F., eds, The Drake Equation: Estimating the Prevalence of Extraterrestrial Life through the Ages. Cambridge University Press, Cambridge

Dressing C. D., Charbonneau D., 2015, The Astrophysical Journal, 807, 45

Eads R. W., Angel J. R. P., 2021, Philosophical Transactions of the Royal Society of London Series A, 379, 20200141

Edwards B., et al., 2021, AJ, 161, 44

Favata F., Hasinger G., Tacconi L. J., Arridge C. S., O'Flaherty K. S., 2021, Experimental Astronomy, 51, 551

Fulton B. J., et al., 2017, AJ, 154, 109

Greaves J. S., et al., 2020, Nature Astronomy,

Greene T. P., Line M. R., Montero C., Fortney J. J., Lustig-Yaeger J., Luther K., 2016, The Astrophysical Journal, 817, 17

Hall P., 1988, The Annals of Statistics, 16, 927

Harper D. B., 1985, Nature, 315, 55

Harris C. R., et al., 2020, Nature, 585, 357–362

Heath M. J., Doyle L. R., Joshi M. M., Haberle R. M., 1999, Origins of life and evolution of the biosphere, 29, 405

Hu R., Seager S., Bains W., 2012, ApJ, 761, 166

Hunter J. D., 2007, Computing in Science Engineering, 9, 90

Kaltenegger L., Traub W. A., 2009, ApJ, 698, 519

Kaltenegger L., Traub W. A., Jucks K. W., 2007, The Astrophysical Journal, 658, 598

Kasting J. F., Whitmire D. P., Reynolds R. T., 1993, Icarus, 101, 108

Khalil M. A. K., Rasmussen R. A., 1999, Atmospheric Environment, 33, 1305

Kitzmann D., Patzer A. B. C., von Paris P., Godolt M., Stracke B., Gebauer S., Grenfell J. L., Rauer H., 2010, A&A, 511, A66

Kopparapu R. K., et al., 2013, ApJ, 765, 131

Licquia T. C., Newman J. A., 2015, ApJ, 806, 96

Lovis C., et al., 2017, A&A, 599, A16

Luger R., Barnes R., 2015, Astrobiology, 15, 119

Lustig-Yaeger J., Meadows V. S., Lincowski A. P., 2019, AJ, 158, 27

Marais D. J. D., et al., 2002, Astrobiology, 2, 153

Meadows V. S., 2017, Astrobiology, 17, 1022

Meadows V. S., et al., 2018, Astrobiology, 18, 630

Mosby G., et al., 2020, Journal of Astronomical Telescopes, Instruments, and Systems, 6, 046001

Otegi J. F., Bouchy F., Helled R., 2020, A&A, 634, A43

Petigura E. A., Howard A. W., Marcy G. W., 2013, Proceedings of the National Academy of Sciences, 110, 19273

Petkowski J., Bains W., Seager S., 2021, in Bulletin of the American Astronomical Society. p. 1146

Pierrehumbert R., Gaidos E., 2011, ApJ, 734, L13

Pilcher C. B., 2003, Astrobiology, 3, 471

Pohorille A., Pratt L. R., 2012, Origins of Life and Evolution of Biospheres, 42, 405

Pont F., Sing D. K., Gibson N. P., Aigrain S., Henry G., Husnoo N., 2013, MNRAS, 432, 2917

Quanz S. P., Kammerer J., Defrère D., Absil O., Glauser A. M., Kitzmann D., 2018, in Proc.SPIE. , doi:10.1117/12.2312051, https://doi.org/10.1117/12.2312051

Rackham B. V., Apai D., Giampapa M. S., 2018, ApJ, 853, 122

Rasmussen R. A., 1974, Tellus, 26, 254

Rauer H., et al., 2011, A&A, 529, A8

Robinson T. D., Stapelfeldt K. R., Marley M. S., 2016, PASP, 128, 025003

Rodler F., López-Morales M., 2014, ApJ, 781, 54

Rugheimer S., Kaltenegger L., Segura A., Linsky J., Mohanty S., 2015, The Astrophysical Journal, 809, 57

Sagan C., Thompson W. R., Carlson R., Gurnett D., Hord C., 1993, Nature, 365, 715

Sarkar S., Argyriou I., Vandenbussche B., Papageorgiou A., Pascale E., 2018, MNRAS, 481, 2871

Sarkar S., Madhusudhan N., Papageorgiou A., 2020, MNRAS, 491, 378

Schneider J., Silk J., Vakili F., 2021, Philosophical Transactions of the Royal Society of London Series A, 379, 20200187

Schwieterman E. W., et al., 2016, ApJ, 819, L13

Schwieterman E. W., et al., 2017, Astrobiology, 18, 663

Seager S., 2014, Proceedings of the National Academy of Sciences, 111, 12634

Seager S., 2018, International Journal of Astrobiology, 17, 294

Seager S., Sasselov D. D., 2000, ApJ, 537, 916

Seager S., Deming D., Valenti J., 2009, in Thronson H. A., Stiavelli M., Tielens A., eds, Astrophysics in the Next Decade. Springer Netherlands, Dordrecht, pp 123–145

Segura A., Krelove K., Kasting J. F., Sommerlatt D., Meadows V., Crisp D., Cohen M., Mlawer E., 2003, Astrobiology, 3, 689

Sing D. K., et al., 2011, MNRAS, 416, 1443

Snellen I. A. G., de Kok R. J., le Poole R., Brogi M., Birkby J., 2013, ApJ, 764, 182

Snellen I., et al., 2015, A&A, 576, A59

Sousa-Silva C., Seager S., Ranjan S., Petkowski J. J., Zhan Z., Hu R., Bains W., 2020, Astrobiology, 20, 235

Stassun K. G., et al., 2018, The Astronomical Journal, 156, 102

Stevenson D. J., 1999, Nature, 400, 32

Sullivan P. W., et al., 2015, ApJ, 809, 77

Tarter J. C., et al., 2007, Astrobiology, 7, 30

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

Tinetti G., et al., 2018, Experimental Astronomy, 46, 135

Virtanen P., et al., 2020, Nature Methods, 17, 261

Wheatley P. J., Louden T., Bourrier V., Ehrenreich D., Gillon M., 2016, Monthly Notices of the Royal Astronomical Society: Letters, 465, L74

Wunderlich F., et al., 2019, A&A, 624, A49

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

Yokouchi Y., Ikeda M., Inuzuka Y., Yukawa T., 2002, Nature, 416, 163

Zsom A., Seager S., de Wit J., Stamenković V., 2013, ApJ, 778, 109

de Wit J., et al., 2018, Nature Astronomy, 2, 214

This paper has been typeset from a TEX/LATEX file prepared by the author.