



Biosignature surveys to exoplanet yields and beyond

Mccullen Sandora, Joseph Silk

► To cite this version:

Mccullen Sandora, Joseph Silk. Biosignature surveys to exoplanet yields and beyond. Monthly Notices of the Royal Astronomical Society, 2020, 495, pp.1000-1015. 10.1093/mnras/staa1284 . insu-03748053

HAL Id: insu-03748053

<https://insu.hal.science/insu-03748053>

Submitted on 7 Jul 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Biosignature surveys to exoplanet yields and beyond

McCullen Sandora^{1,2★} and Joseph Silk^{2,3,4}

¹*Center for Particle Cosmology, University of Pennsylvania, Philadelphia, PA 19104, USA*

²*Institut d'Astrophysique de Paris, 98 bis Bd. Arago, F-75014 Paris, France*

³*Department of Physics and Astronomy, Johns Hopkins University, Baltimore, MD 21218, USA*

⁴*BIPAC, Department of Physics, University of Oxford, Keble Road, Oxford OX1 3RH, UK*

Accepted 2020 May 5. Received 2020 April 9; in original form 2019 December 15

ABSTRACT

Upcoming biosignature searches focus on indirect indicators to infer the presence of life on other worlds. Aside from just signalling the presence of life, however, some biosignatures can contain information about the state that a planet's biosphere has achieved. This additional information can be used to measure what fractions of planets achieve certain key stages, corresponding to the advent of life, photosynthesis, multicellularity, and technological civilization. We forecast the uncertainties of each measurement for upcoming surveys, and outline the key factors that determine these uncertainties. Our approach is probabilistic and relies on large numbers of candidates rather than detailed examination of individual exoplanet spectra. The dependence on survey size, likeliness of the transition, and several measures of degrees of confidence are discussed, including discussion of geological false positives in biosignatures as well as how combining data from different missions can affect the inference. Our analysis should influence policy recommendations for future mission design and strategy to minimize the impact of measurement uncertainties.

Key words: astrobiology – methods: statistical – planets and satellites: detection.

1 INTRODUCTION

The era of large exo-atmosphere surveys will quickly be upon us. Near future experiments aim to measure the atmospheres of several Earth-like planets, and in a few decades, large-scale surveys will collect data on dozens of Earth-like worlds (Kiang et al. 2018). This is exciting because life can potentially have a large impact on a planet's atmosphere, and so the measurement of the gas content of an exoplanet atmosphere can serve as an indirect detection of the presence of any life (Lovelock 1965). We address the following issue via a probabilistic framework: what is the optimal number of exoplanet targets? Our arguments are simple: we do not present detailed spectral analyses of exoplanet atmospheres, undoubtedly a crucial ingredient, but we focus on giving reasonable odds for success based on having a sufficient number of targets and making plausible assumptions about the nature of biosignatures.

The stated goal of many missions has been to clearly detect atmospheric gases in as many systems as possible. However, biosignatures may carry more information than just the presence or absence of life: it is also possible to deduce the level of sophistication that life on a planet has attained, relying on the fact that different signatures signal alternative biochemical processes. The life history of our own planet can be seen as a sequence of transitions wrought

by evolutionary innovations, from biogenesis to the evolution of photosynthesis, multicellularity, and technological civilization (Szathmáry & Smith 1995). As far as these transitions can be expected to be generic, they can each be sought for independently through their characteristic atmospheric imprints. The question we address here is, what fraction of planets undergoes each transition, and more importantly, which can be measured with upcoming surveys? By quantifying the uncertainty in measurements of each of these quantities, we provide a framework for understanding how they depend on proposed mission designs as well as on atmospheric modelling. If the goal is to maximize the information return to the extent possible, this also provides a means of policy recommendation for which aspects of future missions most effort should be devoted.

We begin in Section 2 with an overview of each of the different transitions that we think are important, and some ideas for the associated signatures that can be measured. Not all of these are likely to be constrained in the immediate future, but enunciating our ultimate desires for science return can then serve to focus future research into some of the more challenging measurements. In Section 3 we review current and future telescope missions, and their expected return on each of these measurements. Then, we outline the formalism for parameter estimation, as a function of number of surveyed planets and detections: we begin in Section 4 with the simple case of detecting a single signal, and outline how this is affected by priors. In Section 5, we discuss the

* E-mail: mccullen.sandora@gmail.com

detection of multiple signals in a single data set, and find that most effort should be spent measuring signals that correspond to likely transitions. Section 5.1 is devoted to the subtleties that arise when combining multiple, possibly incomplete data sets, and we find recommendations for which mission yields to optimize. In Section 6 we outline how false positives and negatives influence the variance of our likelihoods, and find that generically the desired confidence in our signal interpretation should be roughly the same magnitude as the likelihood of the signal. Finally, in Section 7, we provide a general prescription for maximizing science return.

2 SIGNATURES

We focus here on what we expect to be a nested sequence of biosignatures, corresponding to the major evolutionary transitions that occur on Earth, as outlined in Carter (1983), Szathmáry & Smith (1995). Many of the biosignatures reorganize the energy flow of the entire biosphere, leaving a large imprint on the planet, most notably via the vast increase in available energy that accompanies these transitions.

The first step we take will be the presence of life in any form. The next step is photosynthesis, the harvesting of energy from the planet's host star thereby providing a mechanism to harvest the dominant source of free energy on a planet. After this, we discuss the evolution of multicellularity on other worlds. Lastly, and most speculatively, the detection of technological civilizations, both of the level Earth has achieved and beyond, is considered. Though most work has been done on detecting photosynthesis and technosignatures, and these are likely the only ones which will be unambiguously measurable with the next generation of telescopes, our goal is to call attention to the importance of distinguishing the other two, with the hopes of spurring the creativity needed to make measuring these a reality.

2.1 Life

Any form of life is expected to be accompanied by chemosynthesis, the rearrangement of chemical matter. Several simple gaseous byproducts have been identified that are indicative of life. It is important to note that most of the biosignatures that we discuss do not unambiguously signal the presence of life, however: several abiotic sources of these gases have been identified that could give rise to false positives. Identifying combinations of observables that would strengthen our confidence that life is the only possible source (type III biosignatures in the terminology of Seager, Bains & Hu 2013) is an important industry, but here we wish to merely include a sampling of the types of signatures that have been proposed. Similarly, detecting the presence of one of these signatures will not guarantee that none of the other, later stages have not been achieved, only that this one has.

The most studied byproduct of simple life is methane (CH_4) (Guzmán-Marmolejo, Segura & Escobar-Briones 2013). Methanogenesis evolved very early on Earth (Battistuzzi, Feijao & Hedges 2004), and is best detected in the 1.0–1.7 μm wavelength regime. Several of its spectral peaks coincide with those of water, but observing the 0.84 and 0.92 μm features at $\text{SNR} > 5$ will unambiguously signal detection (Arney et al. 2017). Additionally, it has a feature at 7.7 μm that would be suitable for measuring with eclipse spectroscopy on planets whose blackbody spectrum peaks near this location (Fujii et al. 2018).

Though the methane on Earth is overwhelmingly produced by life (Etiope & Sherwood Lollar 2013), there are several abiotic pro-

duction mechanisms, such as the serpentinization of rocks, so that it may not automatically be considered a biosignature, especially if the atmosphere is reducing, as on Titan. Ways to disambiguate these two sources are to look for chemical disequilibria: for example, the simultaneous presence of CH_4 , CO_2 , N_2 , and liquid water would signal some prodigious methane source, most easily explained by biology (Krissansen-Totton, Olson & Catling 2018b).

Methane was probably not detectable throughout all of Earth's history; oxygenation of the Proterozoic caused methane levels to drop to undetectable levels (Reinhard et al. 2016). It is likely that methane can only be detected when primary productivity is very high, which requires an energy source as ubiquitous as oxygenic photosynthesis (Wolstencroft & Raven 2002). However, its associated byproduct, haze, can serve as an important indirect indicator of its presence (Arney et al. 2017). Haze is only formed (around sunlike stars) when biological levels of methane are present, and can be detected fairly easily from its continuum NUV signature, most relevant for the Archean eon, when methane levels were 2–3 orders of magnitude higher than present (Pavlov et al. 2000).

It was found in Wang et al. (2018) that methane is detectable with *HabEx* after 850 h of exposure time with favourable cloud conditions, and after 25 h with LUVOIR under similar conditions.

An additional chemical signature of life is nitrous oxide N_2O (Des Marais et al. 2002). This gas is produced when other biogenic nitrogen compounds react with atmospheric oxygen, and so requires the presence of oxygenic photosynthesis, in the absence of any other oxygen sources. In contrast with CH_4 , this biosignature was enhanced during the mid-Proterozoic as a consequence of enhanced oxygen levels (Buick 2007) and can be observed from its several peaks in the 1.5–2.0 μm range. Abiotic sources that would need to be taken into account include lightning (Harman et al. 2018) and flares (Airapetian et al. 2016). Additional biosignature gases have been proposed as well, such as methyl chloride (CH_3Cl) and dimethyl sulfide ($\text{C}_2\text{H}_6\text{S}$) (Seager, Bains & Petkowski 2016), as well as the generic strategy of searching for chemical disequilibrium. Another generic feature of life is biological homochirality, the preference for one molecular handedness over the other, which we discuss in the next subsection.

2.2 Photosynthesis

The advent of oxygenic photosynthesis, the capability of using sunlight to rip electrons off water molecules, was one of the key innovations in the history of life on Earth. It likely only appeared on Earth sometime between 3.7 Ga (Rosing & Frei 2004) and 2.4 Ga (Kirschvink & Kopp 2008 and see Lyons, Reinhard & Planavsky 2014 for a review), though the date of its appearance is somewhat controversial. This innovation had major consequences not only for the size of the Earth's biosphere and types of possible organisms, but also for the entire chemistry of the planet itself: it was a key contributor to the oxygenation of the atmosphere.

Molecular oxygen (O_2) is best detected by a strong spectral line at 0.76 μm , observation of which will necessitate having a spectral resolution of greater than 70 (Gaudi et al. 2018). Detection would require near-present day levels, and if Proterozoic levels were 0.1 per cent of present atmospheric level as some reconstructions indicate, O_2 would not have been directly observable during this time (Planavsky et al. 2014).

However, it is much easier to infer the presence of O_2 indirectly through its photolytic byproduct ozone (O_3). This has many features in the range 0.3–0.7 μm , and on modern Earth manifests itself as

a sharp cutoff at $0.33\ \mu\text{m}$ (Gaudi et al. 2018). Ozone would have been readily observable during the Proterozoic, independently of atmospheric oxygen level (Team et al. 2019). Further, the production of NO and OH can be used to infer the presence of an oxygen atmosphere by their signatures at 5.3 and $1.7\ \mu\text{m}$, respectively (Airapetian et al. 2017b).

Several abiotic sources of oxygen have been detailed recently, among them the photodissociation of water and subsequent loss of hydrogen to space (Luger & Barnes 2015; Airapetian et al. 2017a). Most known abiotic production mechanisms produce a very high atmospheric O/H ratio, which leads to a lack of water vapour (Gao et al. 2015) and clouds (Wordsworth & Pierrehumbert 2014), and so can be ruled out if atmospheric water content is accurately measured. However, there may be certain planets that flaunt this rule, and so a full contextual analysis must be undertaken if oxygen is detected on any exoplanet (Meadows et al. 2018). Determining the water abundance requires observing multiple spectral features throughout the $0.8\text{--}2\ \mu\text{m}$ range, which will be achievable with a spectral resolution of 70 (Team et al. 2019). Many of the abiotic production mechanisms also lead to a much higher concentration of oxygen than observed on Earth, which would lead to O_4 absorption features in the $0.3\text{--}0.8$ and $1.0\text{--}1.4\ \mu\text{m}$ bands that can be searched for (Luger & Barnes 2015). LUVOIR aims to take advantage of these signals to be able to rule out all known oxygen false positives. Additionally, seasonal O_3 differences may be sought, but these will only occur for low O_2 levels, where the ozone layer has not saturated to its maximal value (Mason 2008).

An additional signature of photosynthesis exploits the fact that chlorophyll and other known photosynthetic pigments have optimized their frequency response profile to maximize the number of photons collected, while simultaneously screening out higher energies that would cause overheating (Ford, Seager & Turner 2001). This leads to what is known as the red edge, a sharp drop-off in the absorption properties at $0.7\ \mu\text{m}$, giving rise to the characteristic green appearance of our planet, which is spectrally detectable from space (Arnold et al. 2002). This was proposed as a promising biosignature for exoplanets in Seager et al. (2005), and has no known false positives in the wavelength region where we may observe it. It is important to note, however, that since photosynthesis is optimized for the incident light spectrum, the edge may occur in a different part of the spectrum for planets orbiting different mass stars (Wolstencroft & Raven 2002). If so, then the reflectance profiles of certain minerals, namely cinnabar and sulphur, which have edges at 0.6 and $0.45\ \mu\text{m}$, can mimic the edge expected from biology (Schwieterman et al. 2018). Being a broad-band optical feature, high resolution is not required for detection of this.

Homochirality manifests itself as a 0.01 per cent circular polarization near the red edge (Sparks et al. 2009). Remote detection of polarization from light scattered off vegetation has recently been demonstrated in both linear (Berdyugina et al. 2016) and circular (Patty et al. 2019) configurations, demonstrating the feasibility of this search strategy. LUVOIR will contain a spectropolarimeter, but only in the range of $0.1\text{--}0.4\ \mu\text{m}$. However, it was suggested in Kiang et al. (2007b, a) that the location of the spectral edge is dictated by the stellar spectrum, so that around low mass stars ($\sim 0.3\text{ M}_\odot$) photosynthesis may be optimized in this range. Polarization data are a core science goal of the ELF mission. A final biosignature to note is biofluorescence (O'Malley-James & Kaltenegger 2016), which would signal the downregulation of harmful light to lower frequencies, and may present itself as a detectable afterglow accompanying flares.

2.3 Multicellularity

Multicellularity represents an enormous reorganization of the biosphere, and is easily argued to be a prerequisite for intelligent life. All known multicellular organisms are eukaryotes, and it has been argued that eukaryogenesis is a necessary precondition for multicellularity to occur (Lane & Martin 2010). As opposed to prokaryotes, eukaryotes have a tightly controlled internal structure which is capable of selectively expressing genes when certain conditions are met (Bains 2016). This, more than anything else, is what enabled the multicellular cooperation necessary for microscopic organisms to reorganize into a macroscopic creature with a single germline. Important as eukaryogenesis was for life on Earth, however, it is relatively invisible in terms of remote detection, with no proposed strategies that the authors are aware of. Given the almost immediate corollary of multicellularity, however, this latter transition can serve as a proxy for the former, if we can find ways to search it out.

Many environments on Earth are only capable of being inhabited by extremophiles, which are uniformly unicellular (Grant & Horikoshi 1998). Additionally, it was recently pointed out that the habitable zone for unicellular life is expected to be much broader than the habitable zone for complex life (Schwieterman et al. 2019), the latter being only 20 – 28 per cent the width of the former by their estimates. Indeed, the complexity that we witness today only began in the Cambrian era 540 Mya, so that the majority of Earth's history was exclusively populated by simple organisms (Marshall 2006).

Detection of multicellularity is much more difficult than the previous steps, and consequently not as much attention has yet been paid to this. One possible avenue of inference is the detection of large forests: on Earth, the majority of land plant biomass is in the form of multicellular organisms, as opposed to the mostly unicellular sea plants (Sigman & Hain 2012). This is argued to be a generic requirement of water and mineral collection, which are not as available on land as in the ocean (Niklas 1997). The phase dependence of broad-band properties of scattered light (Fujii et al. 2010) would provide a method of detecting land forests, as would be multipixel imaging of target planets. Several proposed mission designs aim to do just this: ELF can infer subcontinent resolution of several nearby systems in the near future by inverting the time-resolved photometry (Berdyugina et al. 2018), and further afield, hypertelescopes would be able to obtain 30×30 pixel pictures of Earth-like planets at or beyond 3 pc (Labeyrie 2016).

However, the presence of widespread lichens may confound efforts to distinguish multicellular from unicellular life. Indeed, evidence from weathering rates testifies for widespread land life as early as 2.8 Ga on Earth (Stüeken, Catling & Buick 2012) and that it covered a significant portion of the Earth's surface at least as far back as 1.1 Ga (Kenny & Knauth 2001). Work has been done on detecting lichens remotely via satellite imagery in the visible to mid-infrared bands (Gilichinsky et al. 2011) and, while subdominant on Earth, lichens may potentially grow to cover the majority of a planet's land surface in the absence of multicellular competitors.

A method to distinguish tall forests from surface lichens was developed in Doughty & Wolf (2010): by measuring the dependence of reflectance on phase angle, the presence of tall shadows can be inferred, even when averaged over the entire planet. These methods have further been developed and applied to data collected from Earth by the Galileo space probe (Doughty & Wolf 2016) and with the POLDER satellite (Doughty et al. 2020).

It may be possible to infer that a photosynthetic signal is land-based spectrally: for this, it is important to bear in mind that in the

ocean, plankton occur throughout the euphotic zone, where light can penetrate. In fact, the chlorophyll maximum occurs 80 m below the ocean surface (Sigman & Hain 2012). This shifts the spectrum of light collected by these cells, and their chlorophyll pigments have evolved to harvest lower frequencies because of this (Kiang et al. 2007a). While this avenue would be fraught with several stages of indirect inference, it represents a possible method to determine if an observed signal may possibly imply multicellularity.

2.4 Technosignatures

Ultimately, we would like to determine how many other planets host civilizations like our own. There has been ongoing effort to detect evidence of radio communication through SETI (Tarter 2001), and there is currently a push to search for more general signatures of technological civilization (Technosignatures Workshop Participants 2018). In fact, it may be possible to find traces of technosignatures in exoplanet spectra. In Lin, Abad & Loeb (2014), the possibility of detecting the industrial pollutants CF_4 and CCl_3F were outlined, where they found it would be possible around white dwarfs with JWST for 1–2 d exposure time for $10\times$ our current terrestrial levels. These pollutants, along with all other chlorofluorocarbons, have no known non-technological source (Seager et al. 2016), and have residence times of 50 000 yr. They can be observed in the 7.8 and 11.6–12.0 μm bands, respectively.

An additional technosignature includes a solar power panel analogue to the red edge found in plants that can indicate stellar energy harvesting (Lingam & Loeb 2017).

With multipixel resolution, it can become feasible to search for signatures of metropolises on exoplanets, either through their waste heat or artificial light (Berdyugina et al. 2018).

There is no reason to restrict our biosignature searches to developments that the Earth has so far attained. If civilization's heat output rivals that of the host planet, this defines a Kardashev type I civilization (Mullan & Haqq-Misra 2018). Kuhn & Berdyugina (2015) find that planetary infrared anomalies can be detected with a contrast two orders of magnitude greater than could be detected in the visible spectrum.

Similarly, it is possible to search for Kardashev type II and III civilizations by looking for anomalous luminosities of stars and galaxies, respectively. Stars were searched in RAVE and *GAIA* data in Zackrisson et al. (2018), but of the some 8000 suitable stellar targets, no clear detection occurred. Galactic sized civilizations were searched for in *WISE* data in Wright et al. (2014), but again, none were detected.

3 TELESCOPES

It is important to determine the total number of systems that can be observed with a given technology, in order to determine the mission parameters that will maximize the scientific return. This has been treated in many places (Agol 2007; Stark et al. 2014; Stark et al. 2016). Here, we provide a simple analytic approximation of the total exoplanet yield, which will facilitate comparison between the different missions we consider. For definiteness, we define the yield as the number of Earth-like planets around main sequence stars.

The time it takes a telescope to make a measurement with a signal-to-noise ratio of 10 can be summarized from Stark et al. (2014) as

$$\tau = 10^3 \frac{(24 r_p + 4.2 \zeta) D_i^2 + 431 r_z \lambda^2}{\phi_* r_p^2 D_i^4 \Delta \lambda}. \quad (1)$$

Here r_p is the ratio of planet to star flux, ζ is the starlight suppression factor, r_z is the ratio of zodiacal light to star flux, ϕ_* is the stellar photon flux per unit wavelength, D_i is the telescope diameter, and $\Delta \lambda$ is the width of the passband. The first term in the numerator is the noise arising from photon count number, and the second from the background of the target stars. The third term represents a combined contribution of solar and extrasolar zodiacal light, though this depends on assumptions on its prevalence around other star systems. This will become better measured with future missions (Weinberger et al. 2015). The numerical coefficients will depend on design efficiencies and the particular signal measured, but what will be more important to us are the scalings with the parameters.

We recap the various mission designs, and how they affect the required integration time: first, upcoming telescopes are planned that are both space-based and ground-based. Ground-based telescopes can be larger, but the Earth's atmosphere sets a lower limit on the contrast that can be achieved at $\zeta = 10^{-8}$ (Wang et al. 2018). There, the authors find that the second term in equation (1) dominates the first unless $\zeta < 10^{-8}$ for M dwarfs and 10^{-10} for sunlike stars (Wang et al. 2018), and so red dwarfs will be the only suitable targets for upcoming ground-based experiments.

Telescopes can detect exoplanets through the transit method or direct imaging. Most missions we discuss use a coronagraph, or a starshade. The main advantage of the latter is a significantly decreased inner working angle. Starshade missions will also face constraints from fuel and repositioning time (Stark et al. 2016), but this is not taken into consideration in our analysis.

The first two terms in equation (1) scale with the distance to the target d as d^2 . If these are dominant, then we can compute the number of systems observed if a telescope operates for a total time T . To do so, we use the continuum limit, where observable systems are uniformly distributed throughout space with density n , equal to the fraction of stars that possess Earth-like planets multiplied by the density of stars in our local neighbourhood (further multiplied by the probability for transit alignment for telescopes using the transit method). The total number of signals able to be processed in that time will be

$$N_{\text{tot}} = 2 \times 10^{-4} \left(\frac{r_p^2}{92 r_p + 16 \zeta} n^{2/3} \ell_* \Delta \lambda D_i^4 T \right)^{3/5}. \quad (2)$$

Here $\ell_* = 4\pi d^2 \phi_*$ is the stellar luminosity per frequency. However, if instead, the noise is dominated by the zodiacal contribution in equation (1), it will scale as d^4 , and¹

$$N_{\text{tot}} = 6.6 \times 10^{-6} \left(\frac{n^{4/3} \ell_*^2 r_p^2 \Delta \lambda D_i^4 T}{\phi_z \lambda^2} \right)^{3/7}. \quad (3)$$

This can be compared to the scaling $N_{\text{tot}} \sim 17.29(T/\text{yr})^{.41} - 1.79$ found in Stark et al. (2014), which uses a sophisticated target selection algorithm and actual star catalogues: notice the exponent closely matches $3/7 = 0.429$. They also find that the number scales with telescope diameter as $N \sim 0.39(D_i/\text{m})^{1.80} - 0.9$, which corresponds nicely with our value of $12/7 = 1.71$.

So, while the continuum approximation is not perfect for missions aiming at sample sizes of 10–100, it provides a worthwhile approximation to bear in mind. It suggests that to increase return, telescope area, bandwidth, and mission lifetime should be maximized, and

¹If both contributions are important, a septic equation must be solved that interpolates between these two behaviours along the lines of Agol (2007), but this will not be undertaken here.

Table 1. The number of Earth-like planets around main sequence stars for which each transition could be measured with different proposed telescope technologies.

telescope	year	life/photo	tech
<i>JWST</i>	2021	2	–
<i>WFIRST</i>	2020s	4	–
GMT	2024	5	–
TMT	2027	5	–
E-ELT	2025	10	–
<i>HabEx</i>	2030s	12	12
OST (6 m)	2035	10	10
OST (9 m)	2035	20	20
LUVIR (8 m)	2038	56	56
LUVIR (15 m)	2038	108	108
RAVE/GAIA	2018	–	8365
ELF (20 m)	–	12	–
ELF (50 m)	–	100	100
OWL-MOON	–	1000	1000
Hypertelescope	–	50 000	50 000
FOCAL	–	4×10^6	4×10^6

noise minimized, but that the returns for all but telescope diameter will be sublinear. Additionally, as is already well known, it suggests to look in directions of higher stellar density if the survey is not to be full-sky, and to focus on intrinsically brighter planets.

3.1 Future missions

Now, we comment on future exoplanet missions, estimate their yields, and comment on their various targets and constraints. The result of this is summarized in Table 1.

JWST will be a 6.5 m space telescope with wavelength range 0.6–29 μm (Gardner et al. 2006). It has the coronagraphic sensitivity to detect Jupiter at 30 pc, but its rather high noise floor of 10 ppm will restrict its exoplanet targets to bright red dwarfs hosting large planets (Greene et al. 2016). Though it will have to be incredibly lucky to detect oxygen, *JWST* will be able to detect CH_4 and CO_2 around Trappist-1 planets with 10 transits (Krissansen-Totton et al. 2018a) and around GJ876 (Arney et al. 2017) after 65 h, so the expected yield is at least 2. Its mid-IR capabilities make detecting larger molecules possible, such as are produced by technosignatures, though the sensitivity of these may only be feasible around white dwarfs (Lin et al. 2014).

WFIRST (Spergel et al. 2015) is a future 2.4 m infrared space telescope equipped with a coronagraph. Its noise floor will be a few ppb, and though it will detect thousands of exoplanets down to Mars mass, these will primarily be outside of the snow line of their system. In Seager (2018) it was estimated that *WFIRST* will be able to detect atmospheric gases around four Earth-like exoplanets.

ELTs are extremely large ground-based telescopes slated for the 2020s. Their contrast is limited by Earth’s atmosphere to be 10^{-8} in the near-infrared, and so red dwarf stars will be their primary exoplanet targets. GMT (25 m) (Johns 2006) and TMT (30 m) (Skidmore et al. 2015) are expected to yield 5–10 Earth-like planets, and can detect oxygen on an Earth-like planet orbiting an M4 star 5 pc away in 70 h. The E-ELT (Gilmozzi & Spyromilio 2007) will be 39 m and cover the 0.39–2.5 μm range at high spectral resolution. It will be able to target 10–20 rocky habitable zone planets that will have been flagged for follow-up by TESS, GMT, and TMT (López-Morales et al. 2019). In our analyses, we quote the pessimistic values of these numbers.

HabEx (Gaudi et al. 2018) is a proposed 4 m space telescope designed to directly image Earth-like planets around sunlike stars out to a distance of 8 pc. It includes a UV spectrograph, coronagraph, and is equipped with a starshade that can be used for the most interesting 50–100 systems. The integration time needed to detect the Earth-sun system at 7 pc is 1 month, and will spend 3.5 yr detecting exo-Earth candidates, with an expected return of 12. Because its frequency bands extend into the visible range, it will be capable of characterizing technosignatures from city lights at night.

OST (OST 2018) is a proposed 9.1 m space telescope with active cooling down to 4 K that can resolve terrestrial planets from 5 to 660 μm , primarily around M dwarfs. Though most of this range has a contrast of 1 ppm, the 25–200 μm range will be 1–2 orders of magnitude better than *JWST*. Over its 5–10 yr lifetime goal, it will be able to detect ice features, NH_3 , and, thanks to its coronagraph with 10^{-7} contrast sensitivity in the infrared, potentially technosignatures. It is projected to measure CO_2 and O_3 on 30 and 20 rocky planets, respectively. An alternative design consists of a 5.9 m primary mirror with the same science goals but correspondingly decreased planet yield, which we infer using equation (3) (Battersby et al. 2018).

LUVIR (Team et al. 2019) is a proposed mission that will be capable of directly imaging the Earth, Venus, and Jupiter at a distance of 13 pc. The diameter will be either 8 or 15 m, depending on budget choices, and will carry multiple instruments capable of observing in the 0.1–2.5 μm range. It would have a prime mission lifetime of 5 yr, with a lifetime goal of 25 yr. The 15 m design has an expected return of 54 Earth-like planets around AFGKM stars after its initial survey, and the 8 m design anticipates 28. Using our scaling from equation (3), this translates into an expected 108 and 56 Earth-like planets over the full 25 yr lifespan, respectively.

The RAVE DR5 and GAIA DR1 data sets were searched for civilizations using a significant fraction of stellar light in Zackrisson et al. (2018), which could in principle alter the star’s spectrum and luminosity profile. There, 8365 stars in both catalogues were selected by comparing parallax distance with IR spectra. Though six potentially anomalous stars were found, these can all be explained by measurement error and, in one case, a binary companion.

ELF (Berdyugina et al. 2018) is a potential ground-based circular array of 9–25 4–8 m telescopes surveying in the 0.3–5 μm range, and with polarization capabilities. It will be able to directly image nearby planets with continent-scale resolution, and the larger version would be able to detect waste heat from a civilization that uses 25 times our current energy output. It will be able to detect CFCs in about a day’s exposure time. The proposal estimates that at least a dozen Earth-like planets will be characterizable with a 20 m design, and over 100 with a 50 m design.

OWL-MOON (Schneider et al. 2019) is a potential design that would place an overwhelmingly large telescope in a crater on the south pole of the moon. It could potentially be 50–100 m large (or even larger), and would sidestep conventional ground-based telescope challenges such as atmospheric noise, thermal noise, and wind stresses. It would be able to resolve emission lines from an Earth–Sun system at 40 pc in 3 h, and the total estimated yield for the size given above is quoted as 1000.

The hypertelescope (Labeyrie 2016) is a potential ‘flotilla’ array of interferometers potentially spanning 100 km across. This gives it an effective collecting area comparable to a 39 m telescope. With this setup, a 30 min exposure would be able to resolve the spectrum of an Earth-like planet at 3 pc for a 30×30 pixel grid. This would enable detection of continents and cities at night, should any be present in this distance range. Extrapolating the yields of OWL-

MOON to a 1 km design using equation (3), we find that even a modest hypertelescope would yield 50 000 Earth-like planet spectra.

FOCAL (Turyshev et al. 2018) is a potential plan to put a telescope to 550 AU and beyond in order to use the sun’s gravitational lens to vastly enhance the collecting area. With this, a 1 m telescope would be able to reposition itself to scan over the image of a planet, giving it an effective diameter of 12 Earth radii. Over a 7 week integration time for an Earth-like planet at 30 pc, it would be able to create a megapixel image, which, to reiterate, corresponds to 1000×1000 pixels. Observing multiple targets with this technique would be challenging, as repositioning would be prohibitive. To estimate the total number of exo-Earths that could be detected with this technology, we tally the total number within a 300 pc radius of the sun.

Having overviewed the telescopes slated for deployment in the near (and not so near) future, we now outline the general formalism for how well each quantity can be measured for a given survey size. We make our analysis as analytic as possible, in order to track the dependence of the uncertainties on the various experimental parameters as clearly as possible. We illustrate our formalism with several idealized test cases first, in order of increasing complexity: in Section 4, we begin with analysing the case where only one biosignature is measured, and find that maximizing science return in this case is indeed equivalent to maximizing the total number of observed systems, as is so often stated in the literature. In Section 5 we extend our analysis to the measurement of multiple biosignatures, and explore how to combine two different data sets in our framework. Depending on the precise setup, we find recommendations for optimum survey size. In Section 6 we incorporate false positives and false negatives, and study how these affect our formalism. We find that the desired degree of confidence will generically be of the order of the observed fraction of systems that possess a given biosignature, in a variety of particular setups.

4 SINGLE BIOSIGNATURE

To begin, we make several simplifications: first, we treat the stellar population as identical, having no characteristics that would affect its probability of hosting a biosphere, or our ability to detect it. Secondly, we assume all biosphere detections are unambiguous. If we have the ability to survey N_{tot} planets and the fraction of planets with detectable biospheres is f_{bio} , then the number of detections N_{det} is given by a binomial distribution:

$$p(\text{data}|f_{\text{bio}}) = \binom{N_{\text{tot}}}{N_{\text{det}}} f_{\text{bio}}^{N_{\text{det}}} (1 - f_{\text{bio}})^{N_{\text{tot}} - N_{\text{det}}}. \quad (4)$$

When needed, this distribution will be referred to as $B(N_{\text{tot}}, N_{\text{det}}, f)$ below, and data refers generically to the numbers of total and detected systems relevant to the consideration at hand – here, N_{det} and N_{tot} .

We will be more interested in determining the ratio f_{bio} , given the number of surveyed planets and detected biospheres. This is related to the above through Bayes’ Theorem,

$$p(f_{\text{bio}}|\text{data}) \propto p(\text{data}|f_{\text{bio}}) p_{\text{prior}}(f_{\text{bio}}). \quad (5)$$

As an example, let us take the case where $p_{\text{prior}}(f_{\text{bio}})$ is given by a uniform distribution. Then the probability distribution for f_{bio} will be described by a beta distribution,

$$p(f_{\text{bio}}|\text{data}) = \beta(N_{\text{tot}}, N_{\text{det}}, f_{\text{bio}}), \quad (6)$$

where we define $\beta(N_{\text{tot}}, N_{\text{det}}, f_{\text{bio}}) = (N_{\text{tot}} + 1)B(N_{\text{tot}}, N_{\text{det}}, f_{\text{bio}})$. The expected value and variance of f_{bio} are given by (see e.g. Johnson,

Kotz & Balakrishnan 1995):

$$\langle f_{\text{bio}} \rangle = \frac{N_{\text{det}} + 1}{N_{\text{tot}} + 2}, \quad \sigma_{f_{\text{bio}}}^2 = \frac{(N_{\text{det}} + 1)(N_{\text{tot}} - N_{\text{det}} + 1)}{(N_{\text{tot}} + 3)(N_{\text{tot}} + 2)^2}. \quad (7)$$

In the limit where both N_{det} and N_{tot} are large we have, defining the observed fraction as $r_{\text{det}} = N_{\text{det}}/N_{\text{tot}}$,

$$f_{\text{bio}} \sim r_{\text{det}} \pm \sqrt{\frac{r_{\text{det}}(1 - r_{\text{det}})}{N_{\text{tot}}}}. \quad (8)$$

As expected from generic properties of the beta distribution. This informs us that to get a precise measure of the probability of hosting a biosphere, the total number of surveyed planets should be larger than $1/r_{\text{det}}$.

If we fail to detect any biospheres after our survey, then the probability distribution for f_{bio} reduces to

$$p(f_{\text{bio}}|0) = (N_{\text{tot}} + 1)(1 - f_{\text{bio}})^{N_{\text{tot}} + 1}. \quad (9)$$

This tends to 0 for $f_{\text{bio}} \gtrsim 1/N_{\text{tot}}$. Then, we could infer that $f_{\text{bio}} \sim (1 \pm 1)/N_{\text{tot}}$.

If we want to ensure that we have probability p_0 of detecting at least one biosphere, we will need to design an experiment capable of surveying at least

$$N_{\text{tot}} > \frac{\log(1 - p_0)}{\log(1 - f_{\text{bio}})} \quad (10)$$

different planets. In the limit $f_{\text{bio}} \ll 1$, this reduces to $N_{\text{tot}} \gtrsim \mathcal{O}(1)/f_{\text{bio}}$, where the coefficient depends on the desired confidence.

4.1 Priors

We now address the question of what effect the prior distribution $p_{\text{prior}}(f_{\text{bio}})$ has on the final inference. As usual, if enough samples are taken, the form of the prior is diminished; however, the expected return will likely not be quite large enough to completely extinguish the prior’s influence. Here, we investigate several alternatives to a uniform prior, and the effects they introduce.

As a first example, let us suppose that the prior distribution for f_{bio} is log-uniform, $p_{\text{prior}}(f_{\text{bio}}) \propto 1/f_{\text{bio}}$. For the most part, this does not substantially alter the analysis: the result is simply a shift in the parameters of the beta distribution, so that $p(f_{\text{bio}}|\text{data}) = \beta(N_{\text{tot}} - 1, N_{\text{det}} - 1, f_{\text{bio}})$. The expected value of f_{bio} and variance in this case are given by

$$\langle f_{\text{bio}} \rangle = \frac{N_{\text{det}}}{N_{\text{tot}} + 1}, \quad \sigma_{f_{\text{bio}}}^2 = \frac{N_{\text{det}}(N_{\text{tot}} - N_{\text{det}} + 1)}{(N_{\text{tot}} + 2)(N_{\text{tot}} + 1)^2}. \quad (11)$$

These numbers can be seen to be shifts of those in equation (7). These results hold as long as at least one detection is made. If $N_{\text{det}} = 0$, however, it is necessary to introduce the smallest conceivable probability of a planet hosting a biosphere p_{min} in order to regulate the otherwise divergent expressions. This gives

$$f_{\text{bio}} \sim \frac{-1 \pm 1}{N_{\text{tot}}(\gamma_E + \log(p_{\text{min}} N_{\text{tot}}))}, \quad (12)$$

where γ_E is the Euler-Mascheroni constant, and we have taken the large N_{tot} , small f_{bio} limit prior to displaying these formulas. For this to be positive, $p_{\text{min}} < 0.56/N_{\text{tot}}$, which is a sensible condition anyway if no life had been detected by that point. Similarly, using the Jeffreys prior, $p_{\text{prior}}(f_{\text{bio}}) \propto 1/\sqrt{f_{\text{bio}}(1 - f_{\text{bio}})}$, yields $p(f_{\text{bio}}|\text{data}) = \beta(N_{\text{tot}} - 1, N_{\text{det}} - 1/2, f_{\text{bio}})$, again tantamount to simply shifting the survey size and number of detected signals by small numbers.

Let us also note an important factor when measuring more advanced biosignatures: because what we often measure will be

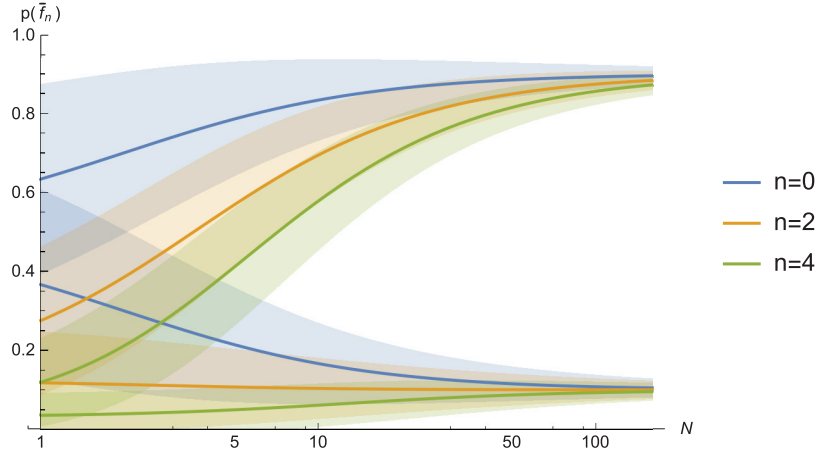


Figure 1. Inferred value of $\bar{f}_n$ (with one standard deviation width) for various values of number of compounded steps n , as given by equation (14). Two curves are displayed for each: one with the fraction of systems that display the signal equal to 0.1, the other with 0.9.

products of sequential probabilities, each of which we take to be uniform, the prior we should assign to the measured quantity will not in fact be uniform. If we measure $\bar{f}_n = \prod_{k=1}^n f_k$, then the prior distribution for this can be found by successively integrating over the latent variables, assuming a uniform distribution for each:

$$p_{\text{prior}}(\bar{f}_n) = \frac{1}{(n-1)!} \log(\bar{f}_n^{-1})^{n-1}. \quad (13)$$

So that, not only will the inferred distribution depend on the prior, but also on the number of steps we choose to include in our counting scheme. As the number of steps to include is far from clear, this makes the desire to collect enough samples to circumvent this issue quite strong. But how many is enough? For this, we compute the mean and variance by integrating equation (4) with the above prior, and take the large N_{tot} limit:

$$\langle \bar{f}_n \rangle \rightarrow r + \frac{1-2r}{N_{\text{tot}}} + \frac{(n-1)(1-r)}{\log(r)N_{\text{tot}}}, \quad (14)$$

where $r = N_{\text{det}}/N_{\text{tot}}$. The first term in the $\mathcal{O}(1/N_{\text{tot}})$ correction is present even for a uniform prior, and simply results from an expansion of equation (7). The term proportional to n represents a systematic negative shift of the average value of $\bar{f}_n$, arising from the preference for smaller compound probabilities. The variance, however, is unchanged to leading order in N_{tot} : $\sigma^2 \rightarrow r(1-r)/N_{\text{tot}}$. The number of samples needed for the negative shift to be within a standard error is then

$$N_{\text{tot}} > n^2 \frac{1-r}{r \log(r)^2}. \quad (15)$$

This diverges for r close to 0 or 1, and attains a minimum of $1.54n^2$ at $r = 0.20$. The dependence of the mean and standard deviation on n and N_{det} are displayed in Fig. 1.

For the remainder of this paper, we will only report on the leading behaviour of these quantities.

5 MULTIPLE NESTED BIOSIGNATURES

Here we extend the results of the previous section, which dealt with the detection efficiency of a single biosignature, to the case where a number of different biosignatures are observed. Our chief concern will be to determine how well we will be able to measure the states of various exoplanet biospheres: for the various transitions discussed in Section 2, we wish to determine the transition rates

for each level, the extent to which they can be measured with a given technology, and a means to inform policy recommendations for instrument design in order to maximize the scientific profit towards this goal. If the returns for the various biosignatures are forecast for a specific mission, as is collated for several upcoming telescopes in Table 1, the final numbers may be easily imported into the expressions we derive to find the achievable measurement accuracies.

The quantities we wish to determine with as much precision as possible are f_i , the fraction of planets that, having achieved level $i-1$, also achieve level i . As before, the setup will be to assume that we have the ability to survey a number N_{tot} of systems. Then, the total number of observed planets of each types will be given by a conditional binomial distribution:

$$p(\text{data}|\{f_i\}) = \prod_i B(N_{i-1}, N_i, f_i), \quad (16)$$

where $N_0 = N_{\text{tot}}$, $\{f_i\}$ refers to the full collection of inferred fractions, and we have made the simplification that the detection efficiency of each biosignature is perfect (to be discussed further in Section 6). The crucial feature of this distribution is that aside from the nested dependence on the number of available systems at each level, the distribution of each variable functions independently of the others. Because of this, inverting this to yield a likelihood function for the variables f_i given the observational yield N_i , we have simply

$$p(\{f_i\}|\text{data}) = \prod_i \beta(N_{i-1}, N_i, f_i). \quad (17)$$

The normalization assumes that the f_i are distributed uniformly, and equation (7) may be used to compute

$$\langle f_i \rangle = \frac{N_i}{N_{i-1}}, \quad \sigma_{f_i}^2 = \frac{\frac{N_i}{N_{i-1}} \left(1 - \frac{N_i}{N_{i-1}}\right)}{N_{i-1}}. \quad (18)$$

The most important feature of these expressions is that the variance of each quantity is controlled by the number of systems displaying the previous signal. A consequence is that, as long as the probability of attaining each successive transition is not 1, the confidence in measuring each successive variable will diminish. This is illustrated in Fig. 2 for the case where all $f_i = 1/2$, and for a total sample of 100 stars, as would be appropriate for LUVOIR or ELF, from Table 1. Here, while the measurement of f_{life} can be determined

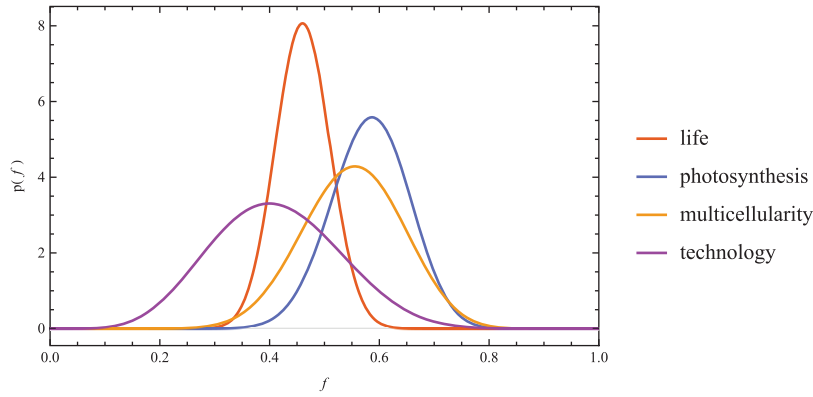


Figure 2. Inferable distributions for f_i s under a random realization of 100 total samples. Here, the distributions are given by equation (17), with expected values equal to $1/2$ for each, and the number of systems possessing each successive quality are 46, 27, 15, and 6. Note that the uncertainties grow for each successive biosignature fraction. It is worth stressing that in this figure, it is assumed that the detection of the biosignature would unambiguously signal the presence of the stage of life under consideration. This assumption will be relaxed in Section 6.

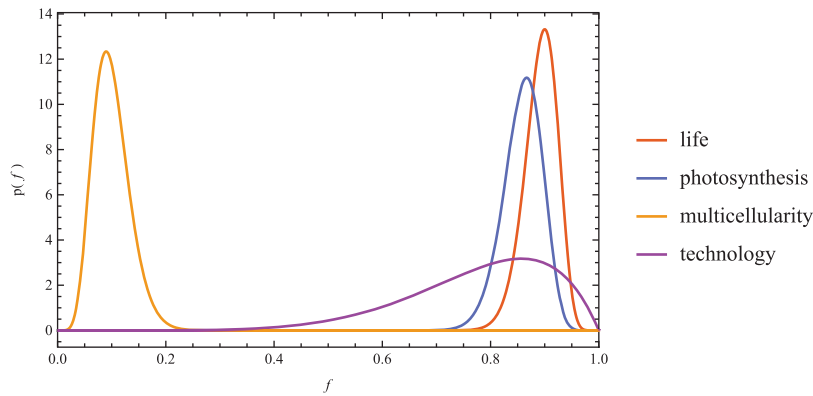


Figure 3. The same as above, except taking all $f_i = 0.9$ except for $f_{\text{mult}} = 0.1$. The number of systems with each quality in this sample are 90, 78, 7, and 6. Here, the uncertainties for the first three biosignatures are roughly the same, but the uncertainty for the last is very large, as it is dictated by the small number that precedes it.

with reasonable accuracy, the latest stages of innovation are highly dominated by sample noise.

This may be contrasted to another possible case displayed in Fig. 3, where all but one transition is taken to occur rather frequently, $f_i = 0.9$, leading to one bottleneck in the road of progress which we have chosen for purely illustrative purposes to be $f_{\text{mult}} = 0.1$. From here, it can be seen that for all transitions up to and including the bottleneck, the underlying fraction of systems can be measured quite accurately. The transitions which occur after the bottleneck, however, are rather poorly constrained.

However, as the initial sample size is increased, the remainder after the bottleneck transition can attain statistically significant values. This suggests that if we do expect one of the transitions to be a bottleneck, we should not invest much effort in designing a mission to measure the transition probabilities that occur after this until we are able to pool from a large enough number of systems to beat the sample noise. From above, the number required will be

$$N_{\text{tot}} \geq \frac{1}{\prod_{i=1}^b f_i}. \quad (19)$$

Then we will not do well to invest in measuring f_b until enough systems have been harvested so that this equality is satisfied. For any value of total sample given in Table 1, this defines values for the f_i for which this measurement will be worthwhile.

Alternatively, one strategy will be to design a supplemental mission capable of measuring the effects of some transition, but not the previous targets, provided that the survey is much larger than the one originally under discussion. Such a mission would only be able to measure the compound probability $\prod_{i=1}^b f_i$, but supplementing with information gleaned from the original mission on each individual transition rate before that will enable us to disentangle the separate effects.

As it stands now, though, we do not know the individual f_i s, which were needed in our heuristic for the threshold number of systems to observe. It would serve us well, then, to determine the most primitive of these values first, and then work our way through the succession as more information becomes available. Only through this method will we be able to accurately determine the expected yields of future missions with any sort of certainty.

We also display a 2D joint pdf in Fig. 4, so as to give a feel for the correlations between variables.

5.1 Combining two data sets

We wish to discuss the scenario where incomplete data on each of the desired signals is gained, and the limits on what we can infer for each of the f_i s. Our treatment will be far from comprehensive, but we will illustrate several examples of situations where we are

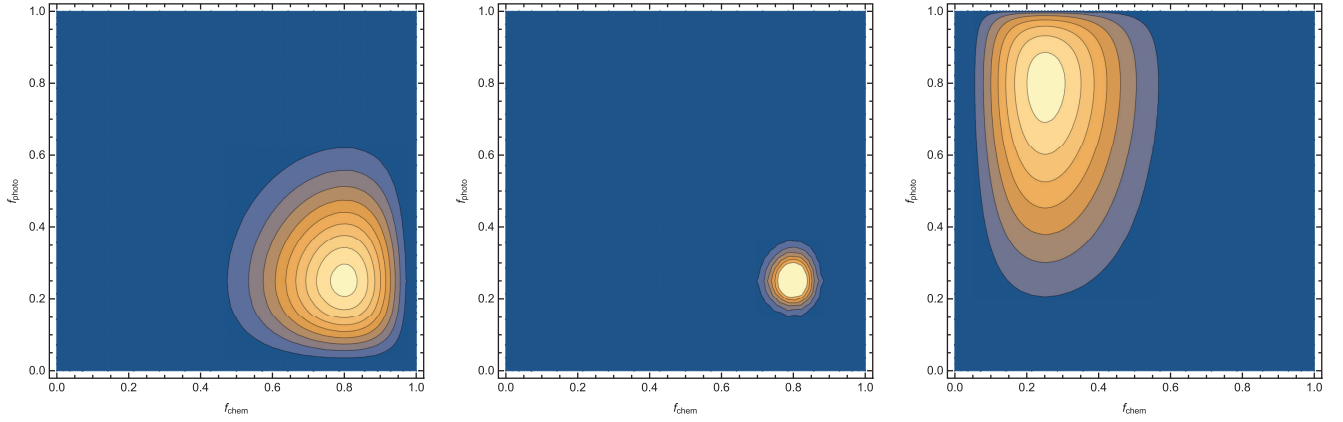


Figure 4. Three different likelihood functions, superposed, illustrating the range of uncertainties on these parameters that will be possible with the upcoming surveys. The distributions are given by equation (17), with values taken to be $(f_{\text{life}} = 0.8, f_{\text{photo}} = 0.25, N_{\text{tot}} = 10)$, $(f_{\text{life}} = 0.8, f_{\text{photo}} = 0.25, N_{\text{tot}} = 100)$, and $(f_{\text{life}} = 0.25, f_{\text{photo}} = 0.8, N_{\text{tot}} = 10)$. Though the fractions are simply swapped between the first and third plots, the uncertainties are magnified in the third, reflecting the diminution of successive samples in the latter case.

likely to find ourselves in the future. For simplicity, we restrict ourselves here to just two signals to be measured, f_1 and f_2 , where transition 1 precedes number 2 in our ordering. For example, we could be measuring the fraction of planets which attain life and photosynthesis.

Suppose we take a population N_{tot} with N_1 instances of signal 1 and N_2 instances of signal 2, and similarly a separate population M_{tot} with measurements M_1 and M_2 . This may come about from data from two different telescopes, for instance, or else from data from the same telescope where we expect not to be able to measure some of the signals in each population. Let us go through several different possible cases, which are illustrated in Fig. 5.

5.1.1 Case 1: All present

This case is quite easy to handle: one can simply combine the two data sets into one larger one. The inferred distributions of the parameters is then

$$p(f_1|\text{data}) = \beta(M_{\text{tot}} + N_{\text{tot}}, M_1 + N_1, f_1),$$

$$p(f_2|\text{data}) = \beta(M_1 + N_1, M_2 + N_2, f_2) \quad (20)$$

the analysis of the previous section can be straightforwardly applied to derive the mean and variance of these two quantities. The means are $\langle f_1 \rangle = \hat{r}_1$ and $\langle f_2 \rangle = \hat{r}_2/\hat{r}_1$, and, more importantly for our analysis, the variances are

$$\sigma_1^2 = \frac{\hat{r}_1(1 - \hat{r}_1)}{M_{\text{tot}} + N_{\text{tot}}}, \quad \sigma_2^2 = \frac{\hat{r}_2}{\hat{r}_1^2} \left(1 - \frac{\hat{r}_2}{\hat{r}_1}\right), \quad (21)$$

where $\hat{r}_i = (M_i + N_i)/(M_{\text{tot}} + N_{\text{tot}})$. As can be seen, M_{tot} and N_{tot} here are on an equal footing, so that in order to maximize the desired signal, one could choose to increase either quantity, giving preference to whichever would be cheapest.

5.1.2 Case 2: $N_2 = 0$

In this case, we have information about the more primitive signal in both data sets, and the more advanced signal only in one. This scenario could arise, for instance, when combining data from two different telescopes, such as *JWST* and *WFIRST*, as displayed in Table 1. Alternatively, this situation could be relevant

if measurement of the second signal is impossible for part of the sample from a single telescope, due to faintness or other factors. For this, the probability distributions are rather simple to write down:

$$p(f_1|\text{data}) = \beta(M_{\text{tot}} + N_{\text{tot}}, M_1 + N_1, f_1),$$

$$p(f_2|\text{data}) = \beta(M_1, M_2, f_2). \quad (22)$$

The statistics for f_1 are the same as in case 1, but now $\langle f_2 \rangle = r_2/r_1$, and

$$\sigma_2^2 = \frac{r_2}{r_1^2} \left(1 - \frac{r_2}{r_1}\right). \quad (23)$$

This is usually going to be larger than the variance of f_1 unless $\langle f_2 \rangle$ is either very close to 0 or 1. As such, in this case priority should be given to increasing M_{tot} rather than N_{tot} : this will, after all, improve measurements of both quantities, rather than just one.

5.1.3 Case 3: $N_2 = 0, M_1 = 0$

In this case, we have information about the first transition from the first data set, but the second data set only gives us information about the product $f_1 f_2$. This scenario could come about with the combination of any of the first five entries delineated in Table 1 with the technosignature search results of RAVE/GAIA. In this case, we have

$$p_1(f_1|\text{data}) = \beta(N_{\text{tot}}, N_1, f_1),$$

$$p_{12}(f_1 f_2|\text{data}) = \beta(M_{\text{tot}}, M_2, f_1 f_2). \quad (24)$$

In order to reconstruct the desired distribution $p_2(f_2|\text{data})$, we integrate over f_1 :

$$p_2(f_2|\text{data}) = \int_0^1 df_1 p_1(f_1|\text{data}) p_{12}(f_1 f_2|\text{data}). \quad (25)$$

This integral, along with most of the subsequent ones we will encounter, can be expressed exactly in terms of hypergeometric functions, essentially by definition. However, these are rarely illuminating and often difficult to manipulate, so we choose not to display them explicitly unless they are unavoidable for our conclusions. Note that we do not include the extra Jacobian factor of f_2 , and instead utilize the standard measure $df_1 df_2$. This is the only

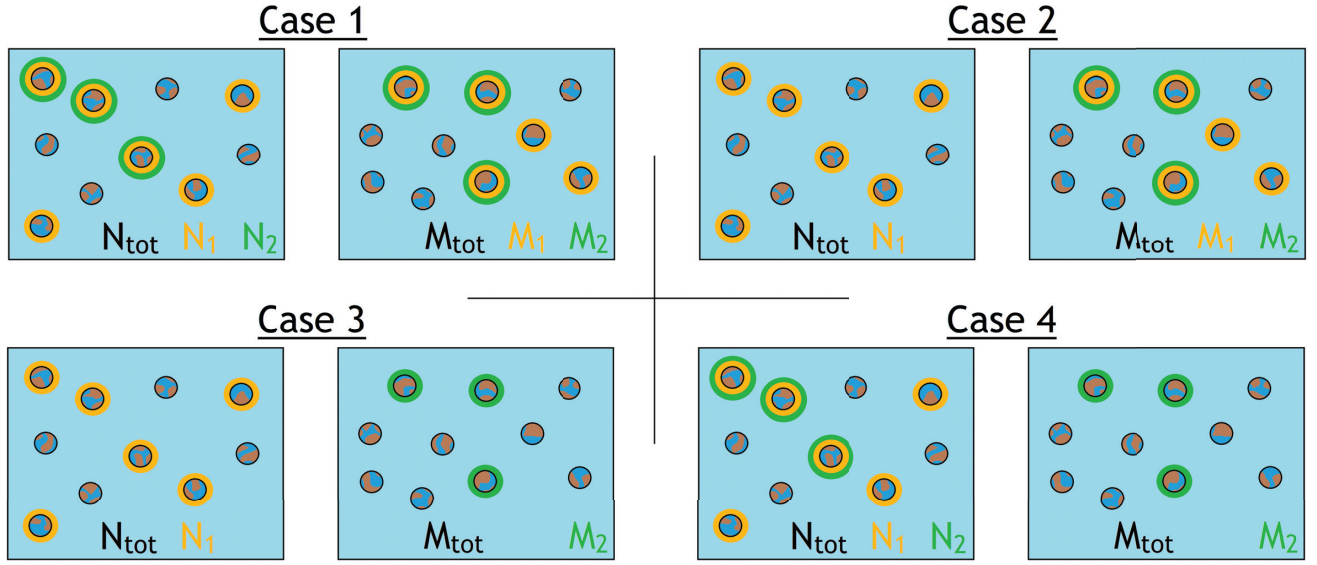


Figure 5. Different possible cases when combining two data sets. Here, the orange glow indicates the presence of a more primitive signal, and the green a more advanced.

appropriate prescription, as the other would break the degeneracy between f_1 and f_2 in the $N_{\text{tot}} = 0$ case.

In order to estimate the total variance for this distribution, it is useful to note that in the large sample limit, beta distributions can be well approximated by Gaussians:

$$\beta(N_{\text{tot}}, N_1, f) \rightarrow \mathcal{N}(r_1, \sigma_1), \quad (26)$$

where $r_1 = N_1/N_{\text{tot}}$ and $\sigma_1^2 = r_1(1-r_1)/N_{\text{tot}}$. Then equation (25) becomes (assuming $p_1(f_1)$ is well localized away from 0 or 1 and that $r_{12} = M_2/M_{\text{tot}} < r_1$ for compatibility):

$$p(f_2|\text{data}) \rightarrow \frac{e^{-\frac{(r_1 f_2 - r_{12})^2}{2(f_2^2 \sigma_1^2 + \sigma_{12}^2)}}}{\sqrt{2\pi (f_2^2 \sigma_1^2 + \sigma_{12}^2)}}. \quad (27)$$

This is plotted for several different survey totals in Fig. 6. The variance of this distribution is approximately

$$\sigma_2^2 \rightarrow \frac{r_{12}^2(1-r_1)}{r_1^3 N_{\text{tot}}} + \frac{r_{12}(1-r_{12})}{r_1^2 M_{\text{tot}}}. \quad (28)$$

If $r_{12} = r_1$, corresponding to $f_2 \approx 1$, then the two contributions to the variance will be equal when the two surveys are the same size. Otherwise, the second term will dominate. Correspondingly, one should prefer to make M_{tot} much larger than N_{tot} . The two contributions are equal when

$$\frac{N_{\text{tot}}}{M_{\text{tot}}} = \frac{r_{12}(1-r_1)}{r_1(1-r_{12})}. \quad (29)$$

An ideal mission would be designed so that the survey sizes match this ratio. This advice is only useful if there is some prior information about r_1 and r_{12} beforehand; otherwise, the mission needs to be completed before it can be optimized. In the absence of this information, one may as well use the expectation value of this quantity: if r_1 and $r_2 = r_{12}/r_1$ are both uniform, then the mission should be designed so that

$$\frac{N_{\text{tot}}}{M_{\text{tot}}} = 2 - \frac{\pi^2}{6} = 0.36 \quad (30)$$

Let us also note that, specifically when combining exoatmosphere searches with technosignature surveys, the number yields of the latter will vastly outnumber the former, at least for the foreseeable future. In this case, $M_{\text{tot}} \gg N_{\text{tot}}$, and the first term in equation (28) will dominate as long as $f_{\text{tech}} > N_{\text{tot}}/M_{\text{tot}}$. Taking the projection for the E-ELT yields, this threshold is 0.001.

The above assumed that the probability distributions are well approximated by a Gaussian, which holds true if the observed values are well separated from 0. In the current situation, with the absence of technosignature detections, this approximation does not hold, and instead we have $p_{12}(f_1 f_2) \sim \mathcal{U}(0, 1/M_2)$, where $\mathcal{U}(0, t)$ is a uniform distribution. When the integral in equation (25) is performed it results in $p_2(f_2) \sim M_2 c_1 (\min(1, 1/(f_2 M_2))) \sim \mathcal{U}(0, 1/M_2)$, irrespective of the distribution of f_1 . In this case, we recover no additional information about the fraction of life that develops into technological societies by measuring the prevalence of life.

5.1.4 Case 4: $M_1 = 0$

The most difficult case to handle is when the more primitive signal is unable to be measured in one data set. Again, this situation may arise from one telescope if the signal is unavailable for part of the sample, or else by the combination of data from two different telescopes. In this case, the data do not allow for the distributions of the two fractions to be factorized, so the full joint distribution of f_1 and f_2 must be employed:

$$p(f_1, f_2|\text{data}) \propto B(N_{\text{tot}}, N_1, f_1) B(N_1, N_2, f_2) B(M_{\text{tot}}, M_2, f_1 f_2). \quad (31)$$

As in case 3, the variance can be estimated in the large survey size limit, when all binomial distributions can be well approximated by Gaussians. For the variance we find

$$\sigma_2^2 \rightarrow \left(\frac{1}{\frac{r_{12}^2(1-r_1)}{r_1^3 N_{\text{tot}}} + \frac{r_{12}(1-r_{12})}{r_1^2 M_{\text{tot}}}} + \frac{1}{\frac{r_2(1-r_2)}{r_1 N_{\text{tot}}}} \right)^{-1}. \quad (32)$$

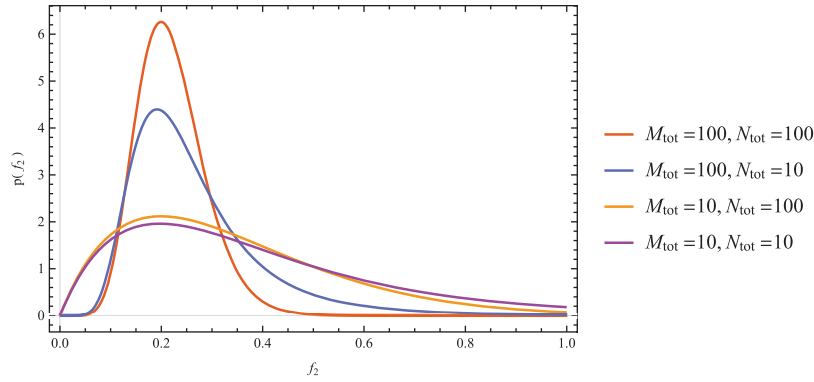


Figure 6. The distribution for f_2 in case 3, as given by equation (25). Here, we have set $f_1 = 1/2$ and $f_2 = 1/5$. Notice that the variance is rather insensitive to N_{tot} , but is rather sensitive to M_{tot} .

And a similar expression for σ_f^2 , with labels interchanged $1 \leftrightarrow 2$. This combines the variance from case 3 with a new term ‘in parallel’, and so the total variance will resemble the smallest of the terms. Usually, this will be most readily given by the new term, and so the recommendation for this scenario is to increase N_{tot} as much as possible. Again, the intuitive reasoning behind this is that increasing this gives information on both signals, whereas increasing M_{tot} only gives information about f_2 .

The rest of the possible cases omit even more measurements from the survey and so are trivial. Having gained some general insights into how to maximize return when combining two data sets, we now turn our attention to the effects of imperfect surveys.

6 FALSE POSITIVES, FALSE NEGATIVES, AND FALSE SAMPLES

Up to this point, we have operated under the simplification that any biosignature detection will be able to be unambiguously interpreted as the presence of life. This is a drastic oversimplification, as false positives are expected to be pernicious obstacles that need to be overcome in order to assess the reliability of any feature. Common methods are to search for a degree of redundancy with correlated signatures that may boost our confidence that the signal we measure is biotic (see Harman & Domagal-Goldman 2018 for a recent review). But what is the desired degree of confidence when it comes to inferring the f s? Obviously, the higher confidence the better, but in this section we show that there is a diminishing return beyond a certain point, and delineate the conditions for when effort should be expended in raising the confidence levels rather than optimizing some other part of the mission.

To address this, we again restrict our attention to a single transition. As an example, this could be the advent of photosynthesis. As discussed above, the primary signature of this is detection of oxygen in a planet’s atmosphere, but myriad abiotic processes are capable of oxygen production as well. Searching for correlated signatures such as water vapour and other gases would reduce this ambiguity but would be more difficult, require a more expensive mission, and would take more time which may potentially be used to explore other systems. Let us say, then, that for the measurement of some abstract biosignature, the confidence that can be ascribed to it being biotic in origin can be determined to be c . We also take the opportunity to define the diffidence $d = 1 - c$. In this case, if we detect N_{det} systems with this signature in our mission, we would only expect $N_{\text{bio}} \approx cN_{\text{det}}$ to be biotic. Though this will hold on average, imperfect confidence introduces additional uncertainty

in the number of systems which have attained the biogenic state we wish to measure. The main problem here is that we observe N_{det} , but wish to infer N_{bio} , which determines f . To do this, we must average over each individual case, in which the sample we observe can have any number due to biotic and abiotic effects, weighted by the probability for each. The distribution for f will be given by

$$p(f|\text{data}, c) = \sum_{N_{\text{bio}}=0}^{N_{\text{det}}} B(N_{\text{det}}, N_{\text{bio}}, c) \beta(N_{\text{tot}}, N_{\text{bio}}, f). \quad (33)$$

This is simply the marginalization over the unobservable number N_{bio} , which sums over all possible values that are consistent with the number of detections N_{det} . This summation can be performed by explicitly using equation (4) for each distribution, leading to the pdf being expressible in terms of hypergeometric functions:

$$p(f|\text{data}, c) = (N_{\text{tot}} + 1)(1 - c)^{N_{\text{det}}} (1 - f)^{N_{\text{tot}}} \times {}_2F_1 \left(-N_{\text{det}}, -N_{\text{tot}}, 1, \frac{cf}{(1-c)(1-f)} \right). \quad (34)$$

It will not really be necessary to use this full form in this section, but it will be used in later sections.

Then, the average can be determined from

$$\langle f \rangle = \sum_{N_{\text{bio}}=0}^{N_{\text{det}}} B(N_{\text{det}}, N_{\text{bio}}, c) \langle f | N_{\text{bio}} \rangle = \frac{c N_{\text{det}} + 1}{N_{\text{tot}} + 2}, \quad (35)$$

where we have used equation (7) for $\langle f | N_{\text{bio}} \rangle$, the average given N_{bio} systems that have life. This indeed goes to $\langle f \rangle \rightarrow c r_{\text{det}}$ in the large survey limit, in agreement with our expectations.

A similar calculation may be carried out for the variance, yielding

$$\sigma_f^2 = \frac{\langle f \rangle (1 - \langle f \rangle + d)}{N_{\text{tot}}}, \quad (36)$$

where we have taken the large N_{tot} limit. This reduces to the previous expression (7) when the confidence is exactly 1, but includes an additional contribution to the variance from the diffidence of our measurements. If the goal is to minimize this, $\langle f \rangle$ should be treated as a fixed number, but either N_{tot} can be increased, as usual, or the diffidence can be decreased (confidence increased). This gives conditions for which should be preferred: since the diffidence contributes to the variance additively with respect to the usual $1 - f$ term, the gain in precision from increasing the confidence only occurs while $c < \langle f \rangle$. Expending effort to increase confidence beyond this value will lead to diminishing returns. If one takes

the ratio of maximally confident to maximally diffident standard deviations, for instance, one finds it to be $\sigma_f(d=0)/\sigma_f(d=1) = \sqrt{(1-\langle f \rangle)/(2-\langle f \rangle)}$: this yields a factor of 1.4 for $f \rightarrow 0$, but can be arbitrarily small for $f \rightarrow 1$, a consequence of the fact that the ordinary contribution to the variance vanishes in this limit. This suggests that we ought to invest more time in raising our confidence of measurements of those transitions which have a high probability of occurring, and not bother as much with the relatively rare transitions.

6.1 Confidence of the confidence

Before, we were treating the confidence c as if it were a known quantity with which we could assess the number of signals in our sample which were produced biotically. However, this is likely to not be the case. To account for this, we must marginalize the expression (33) over all possible values, weighted according to the probability we assign to any given value of the confidence. In this instance, we have

$$p(f|\text{data}) = \int_0^1 dc p_c(c) \sum_{N_{\text{bio}}=0}^{N_{\text{det}}} B(N_{\text{det}}, N_{\text{bio}}, c) \beta(N_{\text{tot}}, N_{\text{bio}}, f). \quad (37)$$

For instance, if the confidence is completely unknown, it can be treated as a uniform random variable $c \sim \mathcal{U}(0, 1)$. This expression acts as an approximate step function, making any value of f less than $N_{\text{det}}/N_{\text{bio}}$ equally probable, and excluding values above this threshold, so that $p(f) \approx \mathcal{U}(0, N_{\text{det}}/N_{\text{tot}})$. This is plotted in Fig. 7, where it can be seen that the width of the transition regime decreases with increasing N_{tot} . While a measurement like this would provide us with an upper bound on f_{bio} , this situation is far from ideal, and so it is recommended that at least some effort be spent on determining the value of c before the measurement takes place.

From the figure, we can see that this information erasure persists in the large sample limit, being a property of marginalizing over uncertainties in the setup rather than arising from measurement error. For a more general $p_c(c)$, the resultant distribution for f will take the same form, only scaled to vanish for $f > r_{\text{det}}$. This can be seen by noting that in the large sample limit, $B(N_{\text{det}}, N_{\text{bio}}, c) \rightarrow \delta_{N_{\text{bio}}, cN_{\text{det}}}$ and $\beta(N_{\text{tot}}, N_{\text{bio}}, f) \rightarrow \delta(N_{\text{bio}} - fN_{\text{tot}})$, so that equation (37) gives $p(f|\text{data}) \rightarrow p_c(f/r_{\text{det}})/r_{\text{det}}$. From this we conclude that in order to maximize information gain we would want to make $p_c(c)$ as close to a delta function as possible.

6.2 Conditional confidence

A further complication is introduced when the confidence depends on the fraction of systems which possess life, which is unknown before the measurement is made. In this case, the confidence that our signal is biotic is related to f_{bio} , the quantity we are trying to measure! This can be summed up in the expression

$$c = \frac{f_{\text{bio}}}{f_{\text{bio}} + f_{\text{abio}}}, \quad (38)$$

where f_{bio} is the fraction of systems that produce the signal in question biotically, and f_{abio} is the fraction of systems that produce the signal abiotically. This was discussed in Catling et al. (2018) and Walker et al. (2018), who use a Bayesian framework for inferring f_{bio} . From here, we can see that if enough care is taken to remove all false positives, by searching for enough independent lines of evidence that we can be certain the signal could not have been produced abiotically, then $c \rightarrow 1$. Apart from that, then our confidence in the signal depends on the value f_{bio} itself. This subtlety is actually rather straightforward to deal with: one may simply

substitute this expression for c into the probability density for f given by equation (34) (taking care to set the normalization of the distribution so that it integrates to 1). However, when this is done, an alarming conclusion is reached: the distribution actually peaks at two places: one at $f \sim r_{\text{det}} = N_{\text{det}}/N_{\text{tot}}$, and the other at $f=0$! This effect persists even for large N_{tot} , and can be understood as follows: imagine the simplified scenario where we have detected N_{det} stars from a sample of N_{tot} , but are unable to determine between the case $c=0$ and $c=1$. Then, our inferred value of f_{bio} will be given by

$$p(f_{\text{bio}}|N_{\text{det}}) = \alpha \binom{N_{\text{tot}}}{N_{\text{det}}} f_{\text{bio}}^{N_{\text{det}}} (1-f_{\text{bio}})^{N_{\text{tot}}-N_{\text{det}}} + (1-\alpha)(1-f_{\text{bio}})^{N_{\text{tot}}}, \quad (39)$$

where α is our confidence that $c=1$. Even in this simple setting, two peaks are evident, at $f_{\text{bio}} = r_{\text{det}} - f_{\text{abio}}$ and $f_{\text{bio}} = 0$. Furthermore, since the second peak is a stronger function, it can carry more weight than the other unless α is close enough to 1. This tells us that unless we are sure enough in our systematic account for the signal's origin, the conclusion of our measurement is that it almost certainly occurred abiotically. In this situation, it would not be necessary to invoke the presence of life to explain any of the detections we make.

In order to quantify the certainty required to avoid this situation, let us return to the full analysis: several example distributions are depicted in Fig. 8.

In order to estimate the relative likelihood that $f_{\text{bio}} \approx 0$, we can make a crude linear approximation around $f_{\text{bio}} = 0$ to determine the total probability that the inferred fraction is within this regime:

$$p(f_{\text{bio}} \sim 0|\text{data}) \approx (N_{\text{tot}} + 1) \left(1 - \left(N_{\text{tot}} + \frac{N_{\text{det}}}{f_{\text{abio}}} \right) f_{\text{bio}} \right) + \mathcal{O}(f_{\text{bio}}^2), \quad (40)$$

which gives

$$p(f_{\text{bio}} \sim 0|\text{data}) \approx \frac{1}{2} \frac{N_{\text{tot}} + 1}{N_{\text{tot}} + \frac{N_{\text{det}}}{f_{\text{abio}}}} \rightarrow \begin{cases} \frac{1}{2} & f_{\text{abio}} \gg r_{\text{det}} \\ \frac{f_{\text{abio}}}{r_{\text{det}}} & f_{\text{abio}} \ll r_{\text{det}} \end{cases}. \quad (41)$$

From here, we can see that in order to avoid any contribution from the $f_{\text{bio}} \sim 0$ peak, we would like to be in the $f_{\text{abio}} \ll r_{\text{det}}$ regime. This is intuitively clear, as since the peaks of the pdf occur at 0 and $f_{\text{bio}} \approx r_{\text{det}} - f_{\text{abio}}$, they will merge if the fraction of abiotic signals is close to the detected fraction into a single peak at $f_{\text{bio}} \approx 0$.

6.3 False negatives

Just as there may be false positive biosignatures, there may be false negatives as well. These can come about if an existing biosphere does not produce the signal we expect of it, in which case it is termed cryptic (Cockell 2014), or if the signal is produced but obscured, say by clouds or haze (Zugger et al. 2011). It is worthwhile to bear in mind that due to haze, many of the proposed biosignatures would have been unobservable on Earth during the Proterozoic eon (Reinhard et al. 2016), which represented a substantial fraction of Earth's history. For this reason it will be important to understand obscuration as well as we possibly can, and efforts to understand the diversity of biological and planetary environments will mitigate some of these uncertainties. However, even if these factors are understood perfectly, there will still be considerable uncertainty in the observed number of exoplanet biospheres: if a seemingly lifeless planet is mostly obscured, how sure can we be that it actually does not possess life? We outline the statistical framework for how to handle this here.

If we survey N_{tot} planets and make N_{det} detections of life, this leaves $N_{\text{tot}} - N_{\text{det}}$ where no life has been detected. If our confidence

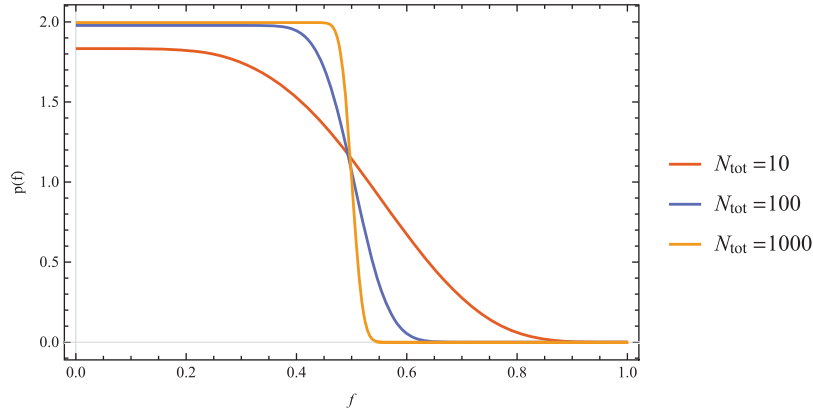


Figure 7. The inferred pdf of f_{bio} as given by equation (37) when marginalizing over a uniform $p_c(c)$. Here $r_{\text{det}} = 1/2$. As can be seen, the full distribution approximately mimics the confidence prior, scaled to be between 0 and r_{det} . This effect persists, and indeed becomes more exact, for large N_{tot} .

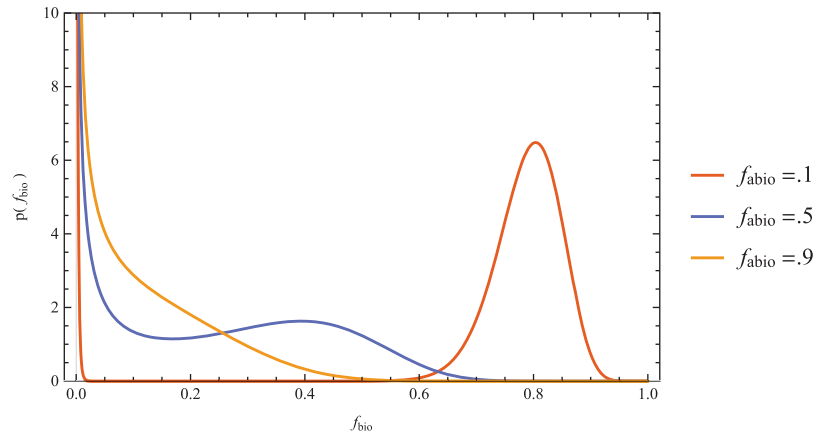


Figure 8. The pdf of f_{bio} when the confidence is related to f_{bio} , found by inserting expression (38) into the distribution (34). All graphs here have $N_{\text{tot}} = 100$ and $r_{\text{det}} = 0.9$. Notice that the peak at 0 still persists, even for the $f_{\text{abio}} = 0.1$ case, and that the main peak is offset from the observed by the fraction of abiotic signals. The desired situation is when these two peaks are well separated and most of the weight is in the one centred on the non-zero value, which occurs when $f_{\text{abio}} \ll r_{\text{det}}$.

that life is absent on each of these is given by c_a , then the number of inhabited planets in our sample is between $N_{\text{det}} \leq N_{\text{bio}} \leq N_{\text{tot}}$. The distribution for the fraction of inhabited planets is given by

$$p(f_{\text{bio}}|\text{data}, c_a) = \sum_{N_{\text{bio}}=N_{\text{det}}}^{N_{\text{tot}}} B(N_{\text{tot}} - N_{\text{det}}, N_{\text{tot}} - N_{\text{bio}}, c_a) \times \beta(N_{\text{tot}}, N_{\text{bio}}, f_{\text{bio}}). \quad (42)$$

This resembles the result we had for false positives in equation (33), and reduces to the standard beta distribution in the limit that $c_a \rightarrow 1$.

The average and variance can be computed for this distribution: the average is

$$\langle f_{\text{bio}} \rangle = r_{\text{det}} + (1 - c_a)(1 - r_{\text{det}}) \quad (43)$$

and the variance

$$\sigma^2 \rightarrow \frac{(\langle f_{\text{bio}} \rangle + d_a)(1 - \langle f_{\text{bio}} \rangle)}{N_{\text{tot}}}. \quad (44)$$

If we wish to minimize the uncertainty of our measurement, we should minimize the diffidence $d_a = 1 - c_a$, but only to the point where it is the same order as $\langle f_{\text{bio}} \rangle$. Notice a key difference in interpretation here: if $c_a \rightarrow 0$, $\langle f_{\text{bio}} \rangle \rightarrow 1$, independent of the

number of biosignatures measured. In this limit, our complete lack of confidence implies that we are actually certain that the lack of signal from a given planet is due to life being present there, but the signal being obscured. This is a somewhat pathological limit, and does not express ignorance as well as the $c \rightarrow 1/2$ limit, for example.

As before, the confidence in the interpretation of the absence of a biosignature depends on the likelihood of life occurring, which is the quantity we set out to measure. If the confidence is given by

$$c_a = \frac{f_{\text{absent}}}{f_{\text{absent}} + f_{\text{hidden}} f_{\text{bio}}} = \frac{1 - f_{\text{bio}}}{1 - (1 - f_{\text{hidden}}) f_{\text{bio}}} \quad (45)$$

then the distribution for f_{bio} can be found with this substitution into equation (42) (again, by altering the normalization to ensure that it integrates to 1). This is plotted for several values of f_{absent} in Fig. 9. Here, a peak at $f_{\text{bio}} \approx 1$ is prominent unless $f_{\text{hidden}} < f_{\text{bio}}$. The explanation is the same as for the false positive case.

6.4 Total number unknown

Another scenario that may occur is that the number of detections could be precisely measured, yet the total number in the sample may be unknown. An example of this would be if the mass of

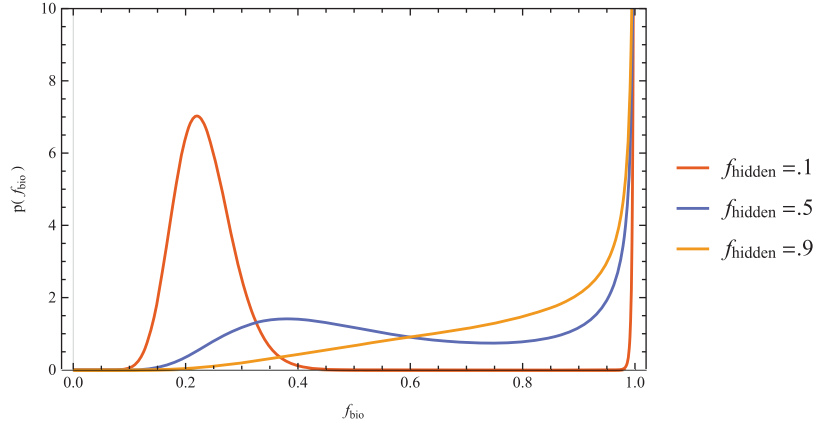


Figure 9. Probability distributions for f_{bio} with false negatives of various degrees of confidence taken into account, as found by inserting equation (45) into the distribution (42). Here, $N_{\text{tot}} = 100$ and $N_{\text{det}} = 20$ throughout. As with the false positive case, peak around the detected value dominates only when $f_{\text{hidden}} \ll f_{\text{bio}}$.

the planet were determined through radial velocity techniques only in the combination $m \sin i$, with i the inclination angle, it would not be clear whether the planets in the sample are Earth-like or not. Additionally, any false positives or negatives in surveys of primitive biosignatures would manifest as an unknown total number in surveys of more advanced ones. Here, we outline the effects of this type of uncertainty, and how best to mitigate it.

For our setup, we will take N_{det} as the number of detections, c_T as the confidence that a non-detection should be included in the total count, and N_{top} as the upper limit of possible systems in the sample. The inferred distribution for f_{bio} will then be

$$p(f_{\text{bio}} | \text{data}, c_T) \propto \sum_{N_{\text{tot}}=N_{\text{det}}}^{N_{\text{top}}} B(N_{\text{top}}, N_{\text{tot}}, c_T) B(N_{\text{tot}}, N_{\text{det}}, f) \\ = B(N_{\text{top}}, N_{\text{det}}, c_T f). \quad (46)$$

Here, we have taken the lower limit to be the number of detections in the sample: in general, this could be an arbitrary number instead, but the analysis is complicated considerably, and the insights we glean from this simpler exercise hold in the more general case. Now, the average can be given in terms of incomplete beta functions:

$$\langle f_{\text{bio}} \rangle = \frac{\beta_{c_T}(N_{\text{det}} + 2, N_{\text{top}} - N_{\text{det}} + 1)}{c_T \beta_{c_T}(N_{\text{det}} + 1, N_{\text{top}} - N_{\text{det}} + 1)} \rightarrow \begin{cases} \frac{r_{\text{det}}}{c_T} & c_T \rightarrow 1 \\ 1 & c_T \rightarrow 0 \end{cases}, \quad (47)$$

where $r_{\text{det}} = N_{\text{det}}/N_{\text{top}}$. These limits are as expected. The variance is similarly given by

$$\sigma_{f_{\text{bio}}}^2 = \frac{\beta_{c_T}(N_{\text{det}} + 3, N_{\text{top}} - N_{\text{det}} + 1)}{c_T^2 \beta_{c_T}(N_{\text{det}} + 1, N_{\text{top}} - N_{\text{det}} + 1)} - \langle f_{\text{bio}} \rangle^2 \\ \rightarrow \begin{cases} \frac{r_{\text{det}}(1-r_{\text{det}})}{c_T^2 N_{\text{top}}} & c_T \rightarrow 1 \\ \frac{1}{N_{\text{top}}^2} & c_T \rightarrow 0 \end{cases}. \quad (48)$$

This latter limit is somewhat unusual, since here the total number in the sample is bound to be N_{det} . The real question, though, is what value of c_T sets the transition between these two different behaviours. This can be determined by finding the subleading corrections to the asymptotic expressions above. When this is done, it is found that these expressions lose their validity when $c_T \approx r_{\text{det}}$. As before, the desired confidence, even in this slightly different

setting, should be greater than the observed ratio in order to garner useful information from the measurement.

7 DISCUSSION

Let us summarize the lessons we've garnered from Sections 4–6. We first noted, somewhat obviously, that uncertainty in measurement of any biosignature decreases with sample size. An observation with somewhat more content is that when measuring multiple biosignatures, if a bottleneck transition is present, then the uncertainties in all the transitions that occur after that are doomed to be large, and so not much will be gained in trying to measure them. When combining data sets that both measure two biosignatures, the variances are proportional to the combined survey sizes. If one survey only measures the more primitive biosignature, then the uncertainty in that depends on a sum of the two survey sizes, while the second biosignature depends only on the size of the second survey. If the first survey only measures the more primitive and the second the more advanced biosignature, then uncertainty in the more advanced is a sum of two terms, and minimizing this will usually occur when the two are equal. When false positives are taken into account, the measurement uncertainty picks up an additional term that becomes important when the uncertainty in interpretation is of the order of the signal prevalence. This conclusion holds true when considering false negatives, false samples, and signal confidence which is conditional on the rate of occurrence: it behooves a survey planner to ensure that the confidence in interpretation is greater than the measurement, but not necessarily much more than that, if resources could instead be devoted to increasing the survey volume. These general lessons will hopefully bear useful to bear in mind when future missions are being designed.

The stated goal of many missions is to ‘maximize science return’. When attempting something that has never been done before, merely accomplishing the goal once, in the cheapest and most guaranteed manner possible, suffices. Once this is done, the method can be extended to larger data sets, with the ultimate goal of generating new knowledge. In the context of measuring atmospheres of Earth-like exoplanets, we are still at the stage where satisfying the task should be our top priority. With the next generation of experiments, it will be possible to obtain the spectra of dozens of Earth-like worlds, but we should not be too discriminate about which are worthy of observation time: any within reach should be targeted. Moving

beyond that, when the characterization of hundreds of worlds is possible, we should rather focus on studying those that we suspect can teach us something qualitatively new.

Is it possible to formalize this intuition? This is the task we are trying to accomplish above. Our main point is that data about various biosignatures can be used to infer values for the fraction of planets undergoing successive biosphere transitions. The uncertainties in these measurements serve as natural candidates for the quantities to be minimized.

Though we have illustrated how this approach may be used to given design and/or target selection recommendations in several idealized scenarios, our discussion to this point has made no pretense at providing a comprehensive framework. In particular, since there are actually multiple simultaneous quantities to be measured, which one should be prioritized? Here, we offer a proposal: the science return of a mission can, loosely, be equated with the information that is gained. We propose that this sentiment should be taken literally, in the technical sense. Then the science return can be defined as $S = -\int d^n f p(f_i|\text{data}) \log p(f_i|\text{data})$, where p is the joint pdf of all measured quantities. Though this is somewhat cumbersome to compute, an analytic expression for science return can be found. In the case of a single biosignature with confidence $c = 1$, equation (6) can be used to compute S as:

$$S = \log \left(\frac{(N_{\text{tot}} + 1)!}{N_{\text{det}}!(N_{\text{tot}} - N_{\text{det}})!} \right) + N_{\text{det}} H(N_{\text{det}}) - N_{\text{tot}} H(N_{\text{tot}} + 1) + (N_{\text{tot}} - N_{\text{det}}) H(N_{\text{tot}} - N_{\text{det}}), \quad (49)$$

where $H(n) = \sum_{k=1}^n 1/k$ is the n th harmonic number. In the large survey limit, the full science return can be approximated as

$$S \approx -\frac{1}{2} \sum_i \log \sigma_i^2, \quad (50)$$

where the variance of each quantity is summed. This definition has a number of attractive features. First, in the standard case of an unambiguous measurement of a single quantity, the variance is given by that of a binomial distribution, equation (7), and maximizing the science return corresponds simply to increasing the number of samples as much as possible. When there are multiple measured quantities, if we can abstractly parametrize how each depends on a single quantity e , denoting effort, then, if the derivative of each is denoted by $R_i = d\sigma_i/de$, we have $dS/de = -\sum_i R_i/\sigma_i$. To maximize, effort should be invested into whichever term in this sum is the largest, corresponding to whichever has a large R_i or a small σ_i where improvement is easy and precision is possible.

In Table 2, we display the science return for the upcoming missions we referred to in Section 3. This table treats biosignature detection as having absolute certainty, but folding additional analyses such as confidence levels into this expression would merely shift the numbers, leaving the overall trends unaltered. Note that in the first half of the table, the information gain is greater for smaller f_i ; this is because, when measuring the rate of a rare event and a common event to the same precision, the measurement of the rare event yields more information. In the lower half of the table, however, more information is gained if the fractions are more common; this is because these missions will measure multiple biosignatures, which will probably be non-existent in the samples if the fractions are small. This should give some indication of the merit of these missions, though again we stress that these numbers are only in terms of the particular question we pose, and are meant as a rough heuristic only. Taking this quantity too seriously can run afoul of Goodhart's law of perverse incentives, but as a heuristic it

Table 2. The expected science return for different proposed telescope technologies, as given by equation (49). Different columns correspond to different values for the actual fractions of each type of biosignature, and we have used log base 2 so as to assign an interpretation as the number of bits of information that a given mission will return.

telescope	$S(f_i = 1/2)$	$S(f_i = 1/10)$
<i>JWST</i>	0.18	0.44
<i>WFIRST</i>	0.39	0.80
GMT	0.47	0.93
TMT	0.47	0.93
E-ELT	0.81	1.39
<i>HabEx</i>	1.47	1.76
OST (6 m)	1.29	1.59
OST (9 m)	2.03	2.31
LUVUOIR (8 m)	3.33	3.61
LUVUOIR (15 m)	4.22	4.51
RAVE/GAIA	4.97	4.54
ELF (20 m)	0.91	1.51
ELF (70 m)	4.11	4.40
OWL-MOON	7.38	7.69
Hypertelescope	13.02	13.33
FOCAL	19.34	19.65

can serve to quickly clarify which features of a complex mission should be improved.

Though our discussion was highly idealized, we were capable of gleaning several lessons. We were able to quantify how large our sample should be to avoid prior bias, which signals will be too weak to bother measuring, which survey to spend time maximizing when combining two, and how well we should establish the confidence of our measurements. Though our recommendations are always phrased in terms of desired relative sample sizes or degrees of confidence, if this is coupled with a model for how these scale with price for a particular telescope, more concrete recommendations can be made. Extensions to our work could be equally profitable. Perhaps the most urgent task is to relax the assumption that all exoplanet systems are identical, and allow the fractions we discuss to depend on environmental variables. Extending our analysis to cover this scenario should be capable of yielding recommendations on which hypothesized trends to focus on measuring in upcoming data sets.

ACKNOWLEDGEMENTS

We would like to thank Jean-Philippe Beaulieu, Ashley Baker, Cullen Blake, Marc Postman, and Ileana Pérez-Rodríguez for helpful discussions. MS is grateful to the Balzan Centre for Cosmological Studies and the Institut d'Astrophysique de Paris for their support during the completion of this work.

REFERENCES

- Agol E., 2007, *MNRAS*, 374, 1271
- Airapetian V., Glocer A., Gronoff G., Hébrard E., Danchi W., 2016, *Nat. Geosci.*, 9, 452
- Airapetian V. S., Glocer A., Khazanov G. V., Loyd R. O. P., France K., Sojka J., Danchi W. C., Liemohn M. W., 2017a, *ApJ*, 836, L3
- Airapetian V. S., Jackman C. H., Mlynczak M., Danchi W., Hunt L., 2017b, *Sci. Rep.*, 7, 14141

- Arney G. N., Meadows V. S., Domagal-Goldman S. D., Deming D., Robinson T. D., Tovar G., Wolf E. T., Schwieterman E., 2017, *ApJ*, 836, 49
- Arnold L., Gillet S., Lardi re O., Riaud P., Schneider J., 2002, *A&A*, 392, 231
- Bains W., 2016, in Pontarotti P., ed., *Evolutionary Biology*. Springer, Berlin, p. 255
- Battersby C. et al., 2018, *Nat. Astron.*, 2, 596
- Battistuzzi F. U., Feij o A., Hedges S. B., 2004, *BMC Evolutionary Biol.*, 4, 44
- Berdyugina S. V., Kuhn J. R., Harrington D. M.,  antl-Temkiv T., Messersmith E. J., 2016, *Int. J. Astrobiol.*, 15, 45
- Berdyugina S. et al., 2018, in Marshall H. K., Spyromilio J., eds, *Proc. SPIE Conf. Ser. Vol. 10700, Ground-based and Airborne Telescopes VII*. SPIE, Bellingham, p. 107004I
- Buick R., 2007, *Geobiology*, 5, 97
- Carter B., 1983, *Phil. Trans. R. Soc. Lond. A*, 310, 347
- Catling D. C. et al., 2018, *Astrobiology*, 18, 709
- Cockell C. S., 2014, *Phil. Trans. R. Soc. Lond. A*, 372, 20130082
- Des Marais D. J. et al., 2002, *Astrobiology*, 2, 153
- Doughty C. E., Wolf A., 2010, *Astrobiology*, 10, 869
- Doughty C. E., Wolf A., 2016, *PLoS one*, 11, e0167188
- Doughty C. E., Abraham A., Windsor J., Mommert M., Gowenlock M., Robinson T., Trilling D., 2020, preprint (arXiv:2002.10368)
- Eti pe G., Sherwood Lollar B., 2013, *Rev. Geophys.*, 51, 276
- Ford E. B., Seager S., Turner E. L., 2001, *Nature*, 412, 885
- Fujii Y., Kawahara H., Suto Y., Taruya A., Fukuda S., Nakajima T., Turner E. L., 2010, *ApJ*, 715, 866
- Fujii Y. et al., 2018, *Astrobiology*, 18, 739
- Gao P., Hu R., Robinson T. D., Li C., Yung Y. L., 2015, *ApJ*, 806, 249
- Gardner J. P. et al., 2006, *Space Sci. Rev.*, 123, 485
- Gaudi B. S. et al., 2018, preprint (arXiv:1809.09674)
- Gilichinsky M., Sandstr m P., Reese H., Kivinen S., Moen J., Nilsson M., 2011, *Int. J. Remote Sens.*, 32, 455
- Gilmozzi R., Spyromilio J., 2007, *Messenger*, 127, 3
- Grant W. D., Horikoshi K., 1998, *Extremophiles: Microbial Life in Extreme Environments*, Vol. 28. Wiley-Liss, New York
- Greene T. P., Line M. R., Montero C., Fortney J. J., Lustig-Yaeger J., Luther K., 2016, *ApJ*, 817, 17
- Guzm n-Marmolejo A., Segura A., Escobar-Briones E., 2013, *Astrobiology*, 13, 550
- Harman C. E., Domagal-Goldman S., 2018, in Deeg H. J., Belmonte J. A., eds, *Handbook of Exoplanets*. Springer International, Berlin, p. 1
- Harman C., Felton R., Hu R., Domagal-Goldman S. D., Segura A., Tian F., Kasting J., 2018, *ApJ*, 866, 56
- Johns M., 2006, in Stepp L. M., ed., in *Proc. SPIE Conf. Ser. Vol. 6267, Ground-based and Airborne Telescopes*. SPIE, Bellingham, p. 626729
- Johnson N. L., Kotz S., Balakrishnan N., 1995, *Continuous Univariate Distributions*, vol. 2. Wiley, New York
- Kenny R., Knauth L. P., 2001, *Geol. Soc. Am. Bull.*, 113, 650
- Kiang N. Y. et al., 2007a, *Astrobiology*, 7, 252
- Kiang N. Y., Siefert J., Govindjee, Blankenship R. E., 2007b, *Astrobiology*, 7, 222
- Kiang N. Y., Domagal-Goldman S., Parenteau M. N., Catling D. C., Fujii Y., Meadows V. S., Schwieterman E. W., Walker S. I., 2018, *Astrobiology*, 18, 619
- Kirschvink J. L., Kopp R. E., 2008, *Phil. Trans. R. Soc. B*, 363, 2755
- Krissansen-Totton J., Garland R., Irwin P., Catling D. C., 2018a, *AJ*, 156, 114
- Krissansen-Totton J., Olson S., Catling D. C., 2018b, *Sci. Adv.*, 4, eaao5747
- Kuhn J. R., Berdyugina S. V., 2015, *Int. J. Astrobiol.*, 14, 401
- Labeyrie A., 2016, *EAS Publ. Ser.*, 78, 45
- Lane N., Martin W., 2010, *Nature*, 467, 929
- Lin H. W., Abad G. G., Loeb A., 2014, *ApJ*, 792, L7
- Lingam M., Loeb A., 2017, *MNRAS*, 470, L82
- L pez-Morales M. et al., 2019, preprint (arXiv:1903.09523)
- Lovelock J. E., 1965, *Nature*, 207, 568
- Luger R., Barnes R., 2015, *Astrobiology*, 15, 119
- Lyons T. W., Reinhard C. T., Planavsky N. J., 2014, *Nature*, 506, 307
- Marshall C. R., 2006, *Annu. Rev. Earth Planet. Sci.*, 34, 355
- Mason J. W., 2008, *Exoplanets*. Springer, Berlin
- Meadows V. S. et al., 2018, *Astrobiology*, 18, 630
- Mullan B., Haqq-Misra J., 2018, *Futures*, 106, 4
- Niklas K. J., 1997, *The Evolutionary Biology of Plants*. University of Chicago Press, Chicago
- O'Malley-James J. T., Kaltenegger L., 2016, 488(4), 4530
- OST mission concept study team Interim Report, 2018, preprint (arXiv:1809.09702).
- Patty C., Kate I. L. t., Buma W. J., van Spanning R. J., Steinbach G., Ariese F., Snik F., 2019, *Astrobiology*, 19(10), 1221
- Pavlov A. A., Kasting J. F., Brown L. L., Rages K. A., Freedman R., 2000, *J. Geophys. Res.*, 105, 11981
- Planavsky N. J. et al., 2014, *Science*, 346, 635
- Reinhard C. T., Planavsky N. J., Olson S. L., Lyons T. W., Erwin D. H., 2016, *Proc. Natl. Acad. Sci.*, 113, 8933
- Rosing M. T., Frei R., 2004, *Earth Planet. Sci. Lett.*, 217, 237
- Schneider J., Silk J., Vakili F., Crawford I., Elvis M., Jacquemoud S., Kaiser R., Sandora M., 2019, preprint (arXiv:1908.02080)
- Schwieterman E. W. et al., 2018, *Astrobiology*, 18, 663
- Schwieterman E. W., Reinhard C. T., Olson S. L., Harman C. E., Lyons T. W., 2019, *The Astrophysical Journal*, 878(1), 19
- Seager S., 2018, *Int. J. Astrobiol.*, 17, 294
- Seager S., Turner E. L., Schafer J., Ford E. B., 2005, *Astrobiology*, 5, 372
- Seager S., Bains W., Hu R., 2013, *ApJ*, 777, 95
- Seager S., Bains W., Petkowski J., 2016, *Astrobiology*, 16, 465
- Sigman D. M., Hain M. P., 2012, *Nat. Educ. Knowl.*, 3, 21
- Skidmore W., et al., 2015, *Res. Astron. Astrophys.*, 15, 1945
- Sparks W. B. et al., 2009, *Proc. Natl. Acad. Sci.*, 106, 7816
- Spergel D. et al., 2015, preprint (arXiv:1503.03757)
- Stark C. C., Roberge A., Mandell A., Robinson T. D., 2014, *ApJ*, 795, 122
- Stark C. C., Shaklan S. B., Lisman P. D., Cady E. J., Savransky D., Roberge A., Mandell A. M., 2016, *J. Astron. Telesc. Instrum. Syst.*, 2, 041204
- St eken E. E., Catling D. C., Buick R., 2012, *Nat. Geosci.*, 5, 722
- Szathm ry E., Smith J. M., 1995, *Nature*, 374, 227
- Tarter J., 2001, *ARA&A*, 39, 511
- LUVUOIR Team, 2019, preprint (arXiv:1912.06219)
- Technosignatures Workshop Participants N., 2018, preprint (arXiv:1812.08681)
- Turyshev S. G. et al., 2018, *COSPAR Scientific Assembly*, 42
- Walker S. I. et al., 2018, *Astrobiology*, 18, 779
- Wang J., Mawet D., Hu R., Ruane G., Delorme J.-R., Klimovich N., 2018, *J. Astron. Telesc. Instrum. Syst.*, 4, 035001
- Weinberger A. J. et al., 2015, *ApJSS*, 216, 24
- Wolstencroft R. D., Raven J. A., 2002, *Icarus*, 157, 535
- Wordsworth R., Pierrehumbert R., 2014, *ApJ*, 785, L20
- Wright J., Griffith R., Sigurdsson S., Povich M., Mullan B., 2014, *ApJ*, 792, 27
- Zackrisson E., Korn A. J., Wehrhahn A., Reiter J., 2018, *ApJ*, 862, 21
- Zugger M., Kasting J., Williams D., Kane T. J., Philbrick C., 2011, *ApJ*, 739, 12

This paper has been typeset from a $\text{\LaTeX}$ file prepared by the author.