The Astronomical Journal, 161:150 (19pp), 2021 March https://doi.org/10.3847/1538-3881/abdb36

Probing the Capability of Future Direct-imaging Missions to Spectrally Constrain the Frequency of Earth-like Planets

Jade H. Checlair[1], Geronimo L. Villanueva[2], Benjamin P. C. Hayworth[3], Stephanie L. Olson[1,4], Thaddeus D. Komacek[1], Tyler D. Robinson[5], Predrag Popović[1,6], Huanzhou Yang[1], and Dorian S. Abbot[1]
[1] Department of the Geophysical Sciences, University of Chicago, Chicago, IL 60637, USA; jadecheclair@uchicago.edu
[2] NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA
[3] Department of Geosciences, Pennsylvania State University, University Park, PA 16802, USA
[4] Department of Earth, Atmospheric, and Planetary Sciences, Purdue University, West Lafayette, IN 47907, USA
[5] Department of the Astronomy and Planetary Science, Northern Arizona University, Flagstaff, AZ 86011, USA
[6] Institut de Physique du Globe de Paris, Paris, France
Received 2020 August 12; revised 2020 December 10; accepted 2021 January 7; published 2021 February 25

Abstract

A critical question in astrobiology is whether exo-Earth candidates (EECs) are Earth-like, in that they originate life that progressively oxygenates their atmospheres similarly to Earth. We propose answering this question statistically by searching for O_2 and O_3 on EECs with missions such as HabEx or LUVOIR. We explore the ability of these missions to constrain the fraction, f_E, of EECs that are Earth-like in the event of a null detection of O_2 or O_3 on all observed EECs. We use the Planetary Spectrum Generator to simulate observations of EECs with O_2 and O_3 levels based on Earth's history. We consider four instrument designs—LUVOIR-A (15 m), LUVOIR-B (8 m), HabEx with a starshade (4 m, "HabEx/SS"), and HabEx without a starshade (4 m, "HabEx/no-SS")—as well as three estimates of the occurrence rate of EECs (η_{earth}): 24%, 5%, and 0.5%. In the case of a null detection, we find that for $\eta_{earth} = 24\%$, LUVOIR-A, LUVOIR-B, and HabEx/SS would constrain f_E to $\leqslant 0.094$, $\leqslant 0.18$, and $\leqslant 0.56$, respectively. This also indicates that if f_E is greater than these upper limits, we are likely to detect O_3 on at least one EEC. Conversely, we find that HabEx/no-SS cannot constrain f_E, due to the lack of a coronagraph ultraviolet channel. For $\eta_{earth} = 5\%$, only LUVOIR-A and LUVOIR-B would be able to constrain f_E, to $\leqslant 0.45$ and $\leqslant 0.85$, respectively. For $\eta_{earth} = 0.5\%$, none of the missions would allow us to constrain f_E, due to the low number of detectable EECs. We conclude that the ability to constrain f_E is more robust to uncertainties in η_{earth} for missions with larger aperture mirrors. However, all missions are susceptible to an inconclusive null detection if η_{earth} is sufficiently low.

Unified Astronomy Thesaurus concepts: Astrobiology (74); Exoplanet atmospheres (487); Direct imaging (387); Biosignatures (2018)

1. Introduction

The past decade has been incredibly productive in exoplanet research, with more than 2500 exoplanets confirmed by NASA's Kepler and K2 missions. Of these, at least 30 planets with radii less than twice Earth's radius were found orbiting in the habitable zone of their star (e.g., Burke et al. 2015; Dressing & Charbonneau 2015). More recently, a number of Earth-sized exoplanets have been found orbiting in the habitable zone of nearby stars (e.g., Dittmann et al. 2017; Gillon et al. 2017). These are prime targets for future instruments to study in more detail and determine whether they are truly habitable, or even inhabited (e.g., Kreidberg & Loeb 2016; Meadows et al. 2018a). NASA's upcoming James Webb Space Telescope should dramatically increase our ability to find and characterize terrestrial exoplanets, but all will be orbiting M stars and they will mostly be too hot for habitability (e.g., Deming et al. 2009; Cowan et al. 2015). The next generation of proposed instruments for the NASA Astronomy and Astrophysics Decadal Survey includes two space-based telescopes: HabEx (Gaudi et al. 2020) and LUVOIR (The LUVOIR Team et al. 2019). These instruments will allow us to characterize the atmospheres of habitable-zone planets orbiting Sun-like stars via direct imaging.

Life can have a measurable impact on the composition of its host planet's atmosphere. A long-standing goal in astrobiology is to spectrally determine the presence of life via biosignatures in the atmosphere (Schwieterman et al. 2018). Several potential biosignatures have been proposed, such as detecting trace amounts of biologically derived molecules (e.g., Seager et al. 2005, 2016; Meadows 2008; Seager & Deming 2010; Sousa-Silva et al. 2019), measuring thermodynamic chemical disequilibrium between atmospheric species (Lovelock 1965; Krissansen-Totton et al. 2018), observing a "red edge" in the atmospheric spectrum (Seager et al. 2005), and detecting seasonal variation (Olson et al. 2018a). Here, we focus on perhaps the most robust biosignatures for G-star planets, O_2 and O_3 (e.g., Owen 1980; Sagan et al. 1993; Des Marais et al. 2002; Meadows 2008, 2017; Schwieterman et al. 2018). We find that O_3 is always easier to detect than O_2 at UV-VIS wavelengths, so we expect that O_3 is the main signal for an oxygenated atmosphere that will be used in future space telescope missions.

Though extensive and ongoing work is being done to determine abiotic sources of O_2 (Tian et al. 2014; Harman et al. 2015; Meadows 2017), most potential "false-positive" scenarios explored in the literature so far occur for either M-dwarf planets or planets that have since gone through a moist or runaway greenhouse phase, rendering them outside the classical habitable zone. False-positive scenarios for Sun-like star planets with an Earth-like inventory of noncondensing gases have not yet been identified (Wordsworth & Pierrehumbert 2014; Harman et al. 2018; Meadows et al. 2018b);

therefore, we expect that we will be able to interpret O_2 and O_3 detections confidently within their broader chemical and planetary context. Regardless, false positives are not considered in this work as our primary concern is to determine what conclusions could be drawn from a potential null detection of O_2 and/or O_3 on all of the detected planets.

As future instruments are being launched and developed, the consensus is that planets that are Earth-sized, terrestrial, and orbiting in the habitable zone of their stars are great targets for the search for life. We will refer to Earth-sized habitable-zone planets as "exo-Earth candidates" (EECs), although this name does not indicate that they are "Earth-like." We will call an "Earth-like planet" an EEC that develops Earth-like O_2-producing life that oxygenates its atmosphere roughly following Earth's oxygenation history. Following this statistical definition, Earth-like EECs would start at negligible levels of atmospheric O_2 such as Earth did during its Hadean and Archean eras, and progressively develop an oxygenated atmosphere. A number of Earth-like EECs would be expected to be found in a Hadean- or Archean-like era and so to lack remotely detectable levels of O_2 and O_3m as Earth did during its Hadean and Archean eras. This definition of Earth-like EECs is therefore not meant to be used for individual planets but instead is purely statistical. If we do not detect O_2 or O_3 on an individual planet, we cannot know whether it resembles Hadean or Archean Earth and will eventually develop an oxygenated atmosphere, or if it is sterile and never will. Instead, we will simply know that at this stage in its history, it does not currently have remotely detectable levels of oxygen in its atmosphere.

We will refer to the fraction of EECs that are Earth-like as f_E. If EECs are generally unlikely to be Earth-like (low f_E), it could either mean that they usually do not originate Earth-like life in the first place, or that although they do originate life, O_2 levels tend to never increase past Archean-like levels (either because oxygenic photosynthesis is rare or because oxygenic photosynthesis does not always manifest as planetary oxygenation). In both scenarios, EECs would not statistically be considered Earth-like, and we will not detect O_2 or O_3. The question we are trying to answer is: could future direct-imaging instruments, such as HabEx and LUVOIR, constrain the frequency of Earth-like EECs, even if they do not detect any?

A critical consideration in this work is the possibility of mission-level false negatives: cases where it is common for EECs to be Earth-like but we do not detect O_2 or O_3 on any of them. We will be adopting a statistical approach to address this problem (Bean et al. 2017; Checlair et al. 2019; Bixel & Apai 2020). This means that we will not investigate false negatives on particular planets. Instead, we will try to determine whether there might be mission-level false negatives for a particular mission based on all of the information that we can gain from all of the EECs that this mission can be expected to observe. Even with a large sample of EECs, we may not detect O_2 or O_3 with either LUVOIR or HabEx if the origination of life is uncommon or if EECs do not generally develop oxygenic photosynthesis. The frequency of life origination on habitable planets is highly uncertain (e.g., Sandberg et al. 2018), so this is a scenario that should be seriously considered. Because it is possible that Earth-like EECs are very rare, we should design an instrument that would discern this, rather than possibly being a mission-level false-negative scenario.

Using a statistical methodology with future direct-imaging instruments will maximize the scientific return of these missions by allowing us to test theories of planetary habitability. This article focuses on statistically testing whether EECs are generally Earth-like, which necessitates a large-enough sample of EECs so that mission-level false-negative scenarios are unlikely and that we may constrain f_E. In previous work, Bixel & Apai (2020) proposed statistically testing the "age–oxygen" correlation with future observatories to determine whether we could place constraints on the amount of O_2 on EECs given the system age. This would then allow us to prioritize systems of a certain age for characterization when a large sample of EECs is available. Bixel & Apai (2020) found that testing this hypothesis would require a large number of EECs (~200 EECs if 10% of detectable EECs have detectable O_2 or O_3, ~20–40 EECs if 50% of detectable EECs have detectable O_2 or O_3) and may therefore only be possible with a LUVOIR-A-like instrument, which is expected to detect more EECs, assuming most EECs have detectable O_2 or O_3.

The occurrence rate of EECs ($\eta_\oplus$) is difficult to constrain as there have not yet been any Earth-sized exoplanets detected in the habitable zone of G-stars. Estimating it therefore requires extrapolation from Kepler data based on the population of small short-period planets. Based on this sample, many different estimates have been published in the literature that vary significantly (e.g., Catanzarite & Shao 2011; Petigura et al. 2013; Burke et al. 2015; Mulders et al. 2018). To come to a community consensus, the NASA-funded Exoplanet Exploration Program Analysis Group (ExoPAG) led Study Analysis Group 13 (SAG13) to compile published occurrence rates from the literature and proposed an average value of $\eta_\oplus$ with uncertainties (Belikov et al. 2017; Kopparapu et al. 2018). Based on the SAG13 study, Stark et al. (2019) adopted boundaries for planet radii of $8((a/1\ \mathrm{au}))^{-0.5}\ R_\oplus \leqslant R \leqslant 1.4\ R_\oplus$, and for semimajor axes of 0.95–1.67 au, and integrated the SAG13 occurrence rates over these boundaries to estimate $\eta_\oplus$ $\sim 24\ ^{+46}_{-16}$ %. However, recent studies from Pascucci et al. (2019) and Neil & Rogers (2020) showed that extrapolating from that Kepler sample is problematic as it is contaminated by stripped sub-Neptune cores, and this sample can therefore not be reliably used to estimate $\eta_\oplus$. This suggests that the estimate used by Stark et al. (2019) may be overly optimistic. Pascucci et al. (2019) re-evaluated $\eta_\oplus$ using exoplanets at larger separations and excluding short-period planets, which may be contaminated by stripped cores, and estimated $\eta_\oplus$ to be ~5%–10%. More recently, Neil & Rogers (2020) used Bayesian models fit to the Kepler data to calculate occurrence rates of planets in different regimes and found that using models with envelope mass loss predicts an order of magnitude drop in $\eta_\oplus$, down to ~0.5%. In this work, we consider three estimates of $\eta_\oplus$: 24% (Stark et al. 2019), 5% (Pascucci et al. 2019), and 0.5% (Neil & Rogers 2020), and we explore how these estimates affect the ability of future direct-imaging missions to constrain the fraction of Earth-like EECs in the case of a null detection of O_2 and O_3 on all of the observed EECs.

This paper is organized as follows. In Section 2, we outline how we determine the number of EECs on which LUVOIR and HabEx could detect O_2 and/or O_3 and the constraints that could be placed on f_E in the case of a null detection. In Section 3 we first theoretically solve for the probability of a mission-level false-negative scenario for any mean EEC yield. We then present the integration times necessary for LUVOIR

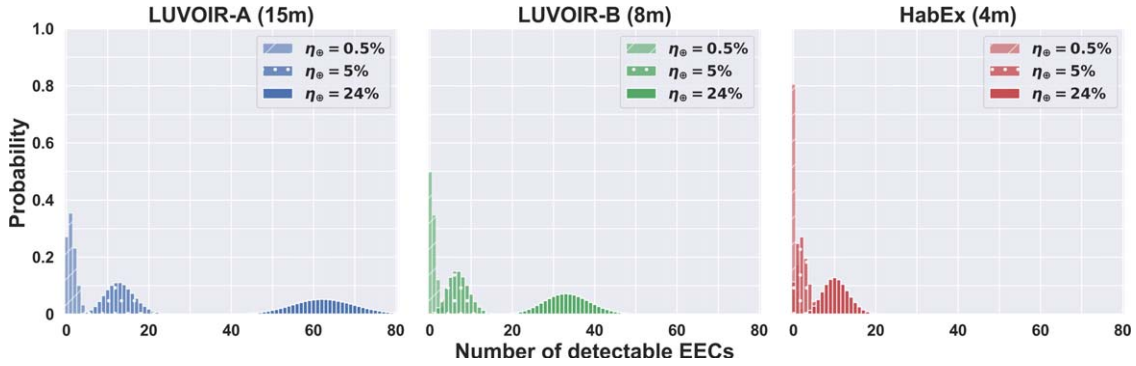

Figure 1. Probability distributions of the number of EECs that would be detected by 15 m segmented on-axis LUVOIR-A (left), 8 m segmented off-axis LUVOIR-B (center), and 4 m monolith off-axis HabEx (right). For $\eta_\oplus = 5\%$, there is an 11% chance of not detecting a single EEC with HabEx. For $\eta_\oplus = 0.5\%$, there is a 27%, 50%, and 81% chance of not detecting any EECs with LUVOIR-A, LUVOIR-B, and HabEx, respectively. All distributions assume a high instrumental throughput. Histograms are based on data from Stark et al. (2019). Solid histograms assume $\eta_\oplus = 24\%$, following Stark et al. (2019); dotted histograms assume $\eta_\oplus = 5\%$, following Pascucci et al. (2019); and barred histograms assume $\eta_\oplus = 0.5\%$, following Neil & Rogers (2020).

and HabEx to detect O_2 and O_3 at 5σ. Next, we present the number of EECs on which LUVOIR and HabEx could detect O_2 and/or O_3 for different values of $\eta_\oplus$. We end that section by discussing how we could use these observations to constrain the fraction, f_E, of EECs that are Earth-like and how this may be affected by the adopted estimate of $\eta_\oplus$. In Section 4, we discuss some implications and caveats of our work, and we summarize results in Section 5.

2. Methods

2.1. Overview

To estimate the distribution of the number of detected Earth-like EECs, we perform Monte Carlo simulations where we consider an ensemble of many repeated HabEx and LUVOIR experiments. For each Monte Carlo realization, we draw the number of planets, N_p, detected by each instrument, and we assign to each detected planet a distance from Earth and an age that are used to determine whether O_2 and O_3 are detectable on the planet. Below, we first explain how we draw N_p, and then we explain how we draw age and distance.

We consider four different mission designs: 15 m segmented on-axis LUVOIR-A, 8 m segmented off-axis LUVOIR-B, 4 m monolith off-axis HabEx with a starshade ("HabEx/SS," where "SS" refers to HabEx's starshade, the complete proposed HabEx mission), and 4 m monolith off-axis HabEx without a starshade ("HabEx/no-SS," a proposed descoped version of HabEx). We choose these designs because they bracket the range of reasonably likely space-based direct-imaging missions over the next few decades. We use the results of Stark et al. (2019) for the expected number of detectable EECs, N_p, for each mission design. These expected yields assume $\eta_\oplus = 24\%$. To consider lower estimates of $\eta_\oplus$, we perform Monte Carlo simulations by resampling the yield estimate data from Stark et al. (2019) and weighting each EEC draw by a factor of (5/24) for $\eta_\oplus = 5\%$ (Pascucci et al. 2019) and by a factor of (0.5/24) for $\eta_\oplus = 0.5\%$ (Neil & Rogers 2020). The resulting distributions of N_p for LUVOIR-A, LUVOIR-B, and HabEx, for different values of $\eta_\oplus$, are shown in Figure 1. We note here that we draw values for the number of detectable EECs, N_p, for both HabEx/SS and HabEx/no-SS from the same HabEx distribution in the left panel of Figure 1. We set the exposure time to 1000 hr per planet, which is a realistic upper limit for observatories such as LUVOIR and HabEx. Observations may require a shorter integration time, while a longer one, although possible in principle, would likely hamper other mission science objectives.

For each Monte Carlo realization, we first draw a value for N_p from Figure 1 for each instrument. This is the number of planets that will be detected using either instrument for each Monte Carlo realization. We then draw an age for each planet from a uniform distribution between 4.54 Gyr old and 0 Gyr old, to cover Earth's history. The age we draw belongs to one of Earth's eras: Hadean (4.54–4 Gya, 11.9% probability of being drawn), Archean (4–2.5 Gya, 33.0% probability of being drawn), pre-Great Oxidation Event (GOE) Proterozoic (2.5–2.3 Gya, 4.4% probability of being drawn), post-GOE Proterozoic (2.3–0.5 Gya, 39.7% probability of being drawn), and Phanerozoic (0.5–0 Gya, 11.0% probability of being drawn). We then draw a distance in parsecs for each of these N_p planets. We draw these distances from a list of targets for each mission provided by Chris Stark (Stark et al. 2019), weighted by the habitable-zone yield estimates, $\eta_\oplus$, for each target. Considering the planet's era and distance, we calculate the signal-to-noise ratio (S/N) for O_2 and O_3 detections with each telescope design given an exposure time of 1000 hr. We define a detectability S/N threshold of 5.0 for O_2 and/or O_3 to be considered detectable, and we count the number of EECs on which this condition is met. We find that the S/N is always higher for O_3 for all the mission designs considered. We repeat this for 10^7 Monte Carlo realizations.

An important assumption we make is that EECs that are Earth-like develop life and atmospheric O_2 following the same trajectory as Earth: Hadean, Archean, Proterozoic, and Phanerozoic. This means that we assume $f_E = 1$ to determine whether there are any mission-level false-negative scenarios under this assumption. We view this as a starting assumption that is necessary to make progress, rather than the most likely scenario. We also consider a wide range of Proterozoic O_2 levels, which allows us to investigate a large range of O_2 history scenarios. We assume that O_2 and O_3 are undetectable at Hadean, Archean, and pre-GOE Proterozoic levels. We also assume EECs remain inhabited for the current inhabited history of Earth of ~3.8 Gyr (Schidlowski 1988; Dodd et al. 2017). This of course would depend on the planet's position in its star's habitable zone, as this will determine the length of time the planet remains habitable (Kopparapu et al. 2013). Earth will only remain in its own habitable zone for ~1.75 Gyr before entering a runaway greenhouse climate (Rushby et al. 2013). During this time, Earth's atmospheric CO_2 should decrease to

Table 1
Instrument Parameters We Used in PSG to Simulate Observations with LUVOIR-A, LUVOIR-B, HabEx/SS, and HabEx/no-SS

Parameter	LUVOIR-A	LUVOIR-B	HabEx/SS	HabEx/no-SS
Diameter	15 m	8 m	4 m	4 m
Wavelength range	UV: 0.2–0.515 μm	UV: 0.2–0.515 μm	UV: 0.2–0.45 μm	UV: 0.35–0.45 μm
	VIS: 0.515–1.0 μm	VIS: 0.515–1.0 μm	VIS: 0.45–0.975 μm	VIS: 0.45–0.975 μm
	NIR: 1.0–2.0 μm	NIR: 1.0–2.0 μm	NIR: 0.975–1.8 μm	NIR: 0.975–1.8 μm
Resolution	UV: 7	UV: 7	UV: 7	UV: 7
	VIS: 140	VIS: 140	VIS: 140	VIS: 140
	NIR: 70	NIR: 70	NIR: 40	NIR: 40
Exozodi level	4.5	4.5	4.5	4.5
(Relative to the solar system)				
Contrast	1×10^{-10}	1×10^{-10}	1×10^{-10}	2.5×10^{-10}
IWA [λ/D]	4	3.5	39 (UV), 58 (VIS), 104 (NIR) mas	2.5
Read noise [e−]	UV: 0	UV: 0	UV: 0.008	UV: 0.008
	VIS: 0	VIS: 0	VIS: 0.008	VIS: 0.008
	NIR: 2.5	NIR: 2.5	NIR: 0.32	NIR: 0.32
Dark noise [e−/s]	UV: 3e−5	UV: 3e−5	UV: 3e−5	UV: 3e−5
	VIS: 3e−5	VIS: 3e−5	VIS: 3e−5	VIS: 3e−5
	NIR: 0.002	NIR: 0.002	NIR: 0.005	NIR: 0.005
$T_{\text{coronagraph}}$	0.27	0.46	0.7	0.55
T_{opt}	UV: 0.13	UV: 0.13	UV: 0.38	UV: 0.09
	VIS: 0.21	VIS: 0.21	VIS: 0.27	VIS: 0.15
	NIR: 0.3	NIR: 0.3	NIR: 0.36	NIR: 0.15

Note. The throughput values for $T_{\text{coronagraph}}$ and T_{opt} are representative average values over the wavelength ranges. We include details about all assumed throughputs in Appendix E.

very low values as a result of the silicate-weathering feedback so that oxygenic photosynthesis by land plants will eventually fail (Caldeira & Kasting 1992). This would likely result in a major decrease in Earth's atmospheric O_2. Because the trajectory of Earth's future atmospheric O_2 levels is highly uncertain, here we simply draw O_2 levels from Earth's history but recognize this may be an optimistic assumption.

2.2. Planetary Atmosphere Simulations

We start by generating atmospheric profiles appropriate for Earth throughout its history. We assume a cloud-free atmosphere for all of the profiles. We do not generate profiles for the Hadean and Archean, as O_2 and O_3 concentrations remain below $\sim 10^{-5}$ present atmospheric level (PAL) for the Archean (Kasting et al. 1979; Pavlov & Kasting 2002). For the Phanerozoic, we use empirical atmospheric profiles for the Modern Earth provided by NASA's MERRA-2 data set (Gelaro et al. 2017; Villanueva et al. 2018). For the Proterozoic, we calculate the mixing ratio profiles using a one-dimensional, horizontally averaged photochemical model (Segura et al. 2007). The model has 35 long-lived chemical species, 16 short-lived chemical species, and 220 reactions. A two-stream approximation is used for radiative transfer, using a fixed zenith angle of 50°. The model solves for the steady-state solution at each altitude layer, accounting for chemical reactions, photolytic reactions, and vertical transport parameterized using Earth-like eddy diffusion profiles (Segura et al. 2007; Harman et al. 2015).

Proterozoic O_2 levels are poorly constrained. We repeat our calculations for several Proterozoic O_2 scenarios ranging from 10^{-5} to 10^{-1} PAL to survey the full range of estimates existing in the literature (Pavlov & Kasting 2002; Lyons et al. 2014; Planavsky et al. 2014; Reinhard et al. 2017; Olson et al. 2018b). Here we note that the lower end of that range (10^{-5} PAL) is difficult to explain in biogeochemical and photochemical models if oxygenic photosynthesis was occurring at or near modern rates (Ozaki et al. 2019). The most likely range of Proterozoic O_2 may therefore be 10^{-4}–10^{-1} PAL. We discuss the constraints on Proterozoic O_2 further in Section 4. Reinhard et al. (2017) and Olson et al. (2018b) also surveyed the existing literature for CO_2 and CH_4 estimates throughout Earth's history, and placed upper and lower bounds on their abundances during each era. They argued for stricter constraints on each of these species' mixing ratios, providing "preferred ranges" for the Proterozoic. We use the midpoint values of those "preferred ranges" for Proterozoic CO_2 ($\sim$2000 μbar) and CH_4 ($\sim$5 μbar). For our Proterozoic water vapor profile, we assume a moist adiabat with a fixed relative humidity of 0.8. We then calculate the O_3 profile based on our background atmosphere (assumed to be 1 bar with N_2 as the major constituent). We then calculate the temperature profile using CLIMA, a one-dimensional radiative-convective climate model (Kopparapu et al. 2013). As the Proterozoic has less O_3 than Modern Earth, there is a smaller temperature inversion in the stratosphere. We set the surface albedo to 0.3 for all of our atmospheric profiles.

2.3. Simulated Observations

We use the Planetary Spectrum Generator (PSG; Villanueva et al. 2018, https://psg.gsfc.nasa.gov) to simulate observations with the 15 m segmented on-axis LUVOIR-A, 8 m segmented off-axis LUVOIR-B, 4 m monolith off-axis HabEx with a starshade ("HabEx/SS"), and 4 m monolith off-axis HabEx without a starshade ("HabEx/no-SS"). For all instruments, we set the exposure time to 1000 hr. We choose the instrument parameters based on their reported values in their respective final reports (The LUVOIR Team et al. 2019; Gaudi et al. 2020). We summarize these parameters in Table 1.

2.4. S/N Calculations

To calculate the S/Ns, we simulate two spectra: one absorbing spectrum with all atmospheric species, and one continuum spectrum with all atmospheric species except the chosen absorber (either O_2 or O_3). We calculate the signal by taking the difference between the two spectra: the signal is higher where the absorber has a stronger absorption feature compared to the continuum. We then divide the signal by the noise that PSG simulates based on the chosen instrument to obtain the S/N as a function of wavelength, which is positive only at wavelengths where the absorbing gas absorbs. The total S/N for O_2 or O_3 is then the square root of the sum of the square of the S/Ns (Lustig-Yaeger et al. 2019):

$$S/N_{total} = \sqrt{\sum_{\lambda_i} (S_{\lambda_i}/N_{\lambda_i})^2}, \quad (1)$$

where S_{λ_i} and N_{λ_i} are the signal and the noise for each wavelength. We calculate the S/Ns at 5, 10, 15, and 20 pc, and interpolate between those values for other distances between 5 and 20 pc. For distances below 5 pc or above 20 pc, we scale the S/Ns as being inversely proportional to distance (Stark et al. 2014). To get a better idea of whether O_2 and O_3 would be detectable in real observations for our different assumptions, we first calculate the integration time necessary for the total S/N to be equal to 5.0 for an Earth-like planet at 5, 10, 15, and 20 pc (see Section 3.2). In our Monte Carlo realizations, we consider the observation of an EEC of a given age and at a given distance as a positive life detection if the S/N of either O_2 or O_3 is above the threshold of 5.0 (see Section 3.3), following Lustig-Yaeger et al. (2019).

3. Results

3.1. EEC Yield Required to Avoid Mission-level False-negative Scenarios

We calculate the probability of a mission-level false-negative scenario for an imaginary mission with any yield of EECs. Let us consider two cases: (1) O_3 is detectable on every EEC at every distance, but only at modern levels; and (2) O_3 is detectable on every EEC at every distance, at both modern and Proterozoic levels. In case (1), O_3 is then detectable on 11% of detectable EECs, as Earth has spent 11% of its history in the modern era. In case (2), O_3 is detectable on 50.7% of detectable EECs, as Earth has spent 50.7% of its history in the Proterozoic and modern eras.

We can then find the probability of a mission-level false-negative scenario using the percentage of detectable EECs with detectable O_3 for that imaginary mission and its mean EEC yield:

$$P = (1 - p)^N, \quad (2)$$

where P is the probability of a mission-level false0negative scenario, p is the percentage of EECs with detectable O_3, and N is the mean yield of EECs. We show the two theoretical curves for cases (1) and (2) in Figure 2. We have also superimposed the mean yields of LUVOIR-A, LUVOIR-B, and HabEx on top of the theoretical curves. Using these curves, we can determine the minimum number of EECs that would need to be detectable to have no chance of a mission-level false-negative scenario. To have less than a 1% chance of a mission-level false-negative scenario, in case (1), 39.5 EECs would need to be detectable. In case (2), 6.5 EECs would suffice as O_3 would be detectable on a greater percentage of them. Similarly, to have less than a 5% chance of a mission-level false-negative scenario, in case (1), 25.7 EECs would need to be detectable, while in case (2), we would need to detect 4.2 EECs.

For $\eta_\oplus = 24\%$, LUVOIR-A's large EEC yield allows it to have less than a 1% chance of a mission-level false-negative scenario even if it could only detect modern levels of O_3. Similarly, LUVOIR-B has less than a 5% chance of a mission-level false-negative scenario even if it could only detect modern levels of O_3. On the other hand, HabEx needs to be able to reliably detect O_3 at Proterozoic levels to avoid a mission-level false-negative scenario. That is due to its smaller EEC yield, which requires it to be able to detect lower levels of O_3.

For $\eta_\oplus = 5\%$, LUVOIR-A and LUVOIR-B need to be able to detect Proterozoic levels of O_3 to have less than a 1% chance of a mission-level false-negative scenario. HabEx on the other hand will have a >20% chance of a mission-level false-negative scenario even if it could detect O_3 at modern and Proterozoic levels on every EEC.

For $\eta_\oplus = 0.5\%$, LUVOIR-A, LUVOIR-B, and HabEx would have a ~40%, ~60%, and ~85% chance of a mission-level false-negative scenario, respectively, even if they could detect O_3 at modern and Proterozoic levels on every EEC. Therefore, mission-level false-negative scenarios are unavoidable for such a low estimate of $\eta_\oplus$.

3.2. Integration Times to Detect O_2 and O_3 with LUVOIR and HabEx

We present the integration times required for a 5σ detection of O_2 and O_3 at 5 pc for cloud-free atmospheres calculated using PSG (Villanueva et al. 2018) in Table 2 for LUVOIR-A, LUVOIR-B, HabEx/SS, and HabEx/no-SS. The integration times calculated at 10, 15, and 20 pc can also be found in Appendix C. To calculate these integration times, we first calculate the S/Ns for O_2 and O_3 detections using the method outlined in Section 2.4 with an exposure time of 1000 hr. We then calculate what the integration time would have to be for these S/Ns to be equal to 5.0:

$$\Delta t = 1000 \text{ (hr)} \times \left(\frac{5}{S/N}\right)^2. \quad (3)$$

We consider six different O_2 levels: modern (1 PAL), and five Proterozoic estimates (10^{-1}–10^{-5} PAL).

In all cases for LUVOIR and HabEx, O_3 is easier to detect than O_2 due to its deep and broad feature between 0.2 and 0.3 μm (Hartley bands) and between 0.5 and 0.7 μm (Chappuis bands). Most importantly, O_3 is detectable at 5σ with LUVOIR-A, LUVOIR-B, and HabEx/SS in under 500 hr even at very low estimates for Proterozoic O_2 levels (down to 10^{-5} PAL). For HabEx/no-SS, detecting O_3 without a starshade at 5 pc is difficult (>1000 hr) for Proterozoic O_2 levels below 10^{-2} PAL. We note here that the HabEx/SS integration times we calculated for O_2 are ~1–4 orders of magnitude larger than reported in the HabEx report (Figures 3.3–7 of Gaudi et al. 2020) depending on the O_2 level, due to the fact that our background atmosphere includes other absorbers such as O_3 and H_2O while that of the HabEx report only includes N_2. However, our calculations agree closely with

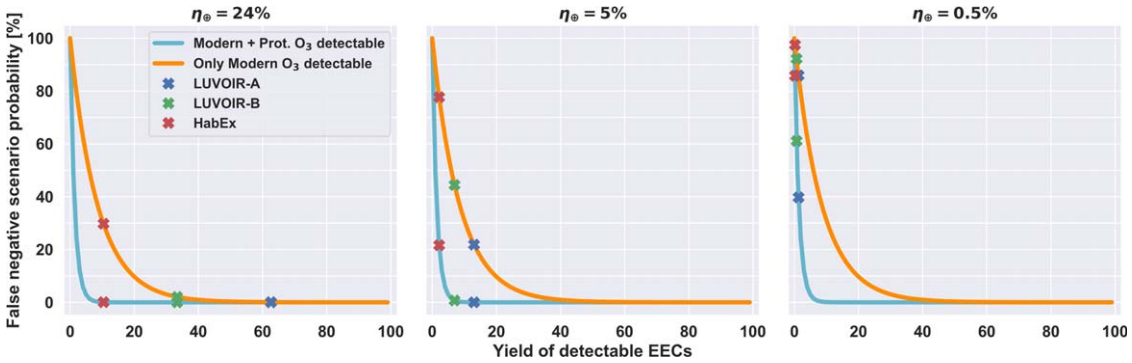

Figure 2. Probability of a mission-level false-negative scenario as a function of the yield of detectable EECs for three different values of $\eta_\oplus$. For $\eta_\oplus = 24\%$, LUVOIR-A and LUVOIR-B have a <1% and <5% chance of a mission-level false-negative scenario, respectively, even if they could only detect modern levels of O_3. On the other hand, HabEx needs to be able to detect Proterozoic levels of O_3 on every EEC to have a <1% chance of a mission-level false-negative scenario. For $\eta_\oplus = 5\%$, LUVOIR-A and LUVOIR-B must be able to detect Proterozoic levels of O_3 to have a<1% chance of a false-negative scenario, while HabEx has a >20% chance of one regardless of its ability to detect O_3. For $\eta_\oplus = 0.5\%$, all missions have probable mission-level false-negative scenarios: ⩾40%, ⩾60%, and ⩾85%, for LUVOIR-A, LUVOIR-B, and HabEx, respectively. Orange line: only modern levels of O_3 are detectable on every EEC; cyan line: modern and Proterozoic levels of O_3 are detectable on every EEC. Blue cross: LUVOIR-A's mean yield, green cross: LUVOIR-B's mean yield, red: HabEx's mean yield.

Table 2
Integration Times (hr) with LUVOIR-A (15 m), LUVOIR-B (8 m), HabEx/SS (4 m), and HabEx/no-SS (4 m) to Yield a 5σ Detection of O_2 and O_3 for an Earth-like Planet without Clouds at 5 pc for Six Different O_2 Levels

	15 m LUVOIR-A	8 m LUVOIR-B	4 m HabEx/SS	4 m HabEx/no-SS
$O_2 = 1$ PAL	O_2: 0.75 hr	O_2: 3.31 hr	O_2: 12.8 hr	O_2: 44.2 hr
	O_3: 0.19 hr	O_3: 0.56 hr	O_3: 0.45 hr	O_3: 17.1 hr
$O_2 = 10^{-1}$ PAL	O_2: 4.24 hr	O_2: 17.43 hr	O_2: 65.9 hr	O_2: 211.7 hr
	O_3: 0.33 hr	O_3: 0.89 hr	O_3: 0.41 hr	O_3: 62.8 hr
$O_2 = 10^{-2}$ PAL	O_2: 41.2 hr	O_2: 167.9 hr	O_2: 564.1 hr	O_2: 2002.1 hr
	O_3: 0.57 hr	O_3: 1.48 hr	O_3: 0.46 hr	O_3: 546.9 hr
$O_2 = 10^{-3}$ PAL	O_2: 658.0 hr	O_2: 2679.8 hr	O_2: 6772.4 hr	O_2: 3.2×10^4 hr
	O_3: 0.99 hr	O_3: 2.59 hr	O_3: 0.70 hr	O_3: 1463.9 hr
$O_2 = 10^{-4}$ PAL	O_2: 3.1×10^4 hr	O_2: 1.2×10^5 hr	O_2: 1.6×10^5 hr	O_2: 1.5×10^6 hr
	O_3: 4.15 hr	O_3: 11.1 hr	O_3: 2.22 hr	O_3: 9.5×10^4 hr
$O_2 = 10^{-5}$ PAL	O_2: 2.7×10^6 hr	O_2: 1.1×10^7 hr	O_2: 1.0×10^7 hr	O_2: 1.3×10^8 hr
	O_3: 125.7 hr	O_3: 338.2 hr	O_3: 58.94 hr	O_3: 2.0×10^7 hr

Note. Integration times calculated at 10, 15, and 20 pc can be found in Appendix C.

simulations made using the Robinson et al. (2016) model when including additional background gases in their model (see Appendix D for further details).

3.3. Likelihood of Detecting O_2 and/or O_3 with LUVOIR and HabEx

3.3.1. $\eta_\oplus = 24\%$

We present the probability distributions of total O_2 and/or O_3 detections for LUVOIR-A (leftmost), LUVOIR-B (center left), HabEx/SS (center right), and HabEx/no-SS (rightmost) in Figure 3, assuming $\eta_\oplus = 24\%$. As described in Section 2, we calculate the S/N of O_2 and O_3 for a given EEC at a certain distance and with a certain level of O_2 based on its age drawn from a uniform distribution. If either of these S/Ns is above the threshold of 5.0 (for an exposure time of 1000 hr), we consider that EEC a positive O_2/O_3 detection. We note here that because O_3 is easier to detect than O_2 in all cases we considered (see Table 2), O_3 detectability is the limiting factor in determining whether O_2 and/or O_3 is detectable on a given EEC. We assume that $f_E = 1$ (all EECs are Earth-like) and determine whether there are any mission-level false-negative scenarios even under this optimistic assumption. We vary the Proterozoic O_2 from 10^{-5} to 0.1 PAL.

For LUVOIR-A, with Proterozoic levels of O_2 equal to or larger than 10^{-4} PAL, O_3 is detectable for EECs at all target distances for both modern and Proterozoic levels. Because of this, the first four rows of Figure 3 show the same distribution. The reason the distribution is offset from 62 (peak of Figure 1, left panel) is that O_2 and O_3 are undetectable for EECs in a Hadean or Archean era and so all EECs drawn to be of Hadean or Archean age will correspond to a null detection of O_2 and O_3. For these four levels of Proterozoic O_2, we find that LUVOIR-A has a 95% chance of detecting 21–43 Earth-like EECs (for $f_E = 1$). At Proterozoic O_2 levels of 10^{-5} PAL, O_3 is only detectable on EECs at close distances (<11.5 pc), and therefore, the distribution shifts toward zero due to the EECs that are farther away and drawn to be of Proterozoic age. We find that LUVOIR-A has a 95% chance of detecting 7–21 Earth-like EECs for Proterozoic levels of 10^{-5} PAL (for $f_E = 1$). In the first four cases (Proterozoic O_2 between 0.1 and 10^{-4} PAL), the distribution of the number of EECs that have detectable O_3 is well above zero. For the case with Proterozoic levels of 10^{-5} PAL, the curve's end member is above zero, with five as the smallest number of detected Earth-like EECs. There is therefore no mission-level false-negative scenario for LUVOIR as a mission if $f_E = 1$: O_2 and/or O_3 will be detected on a number of EECs if they are generally Earth-like.

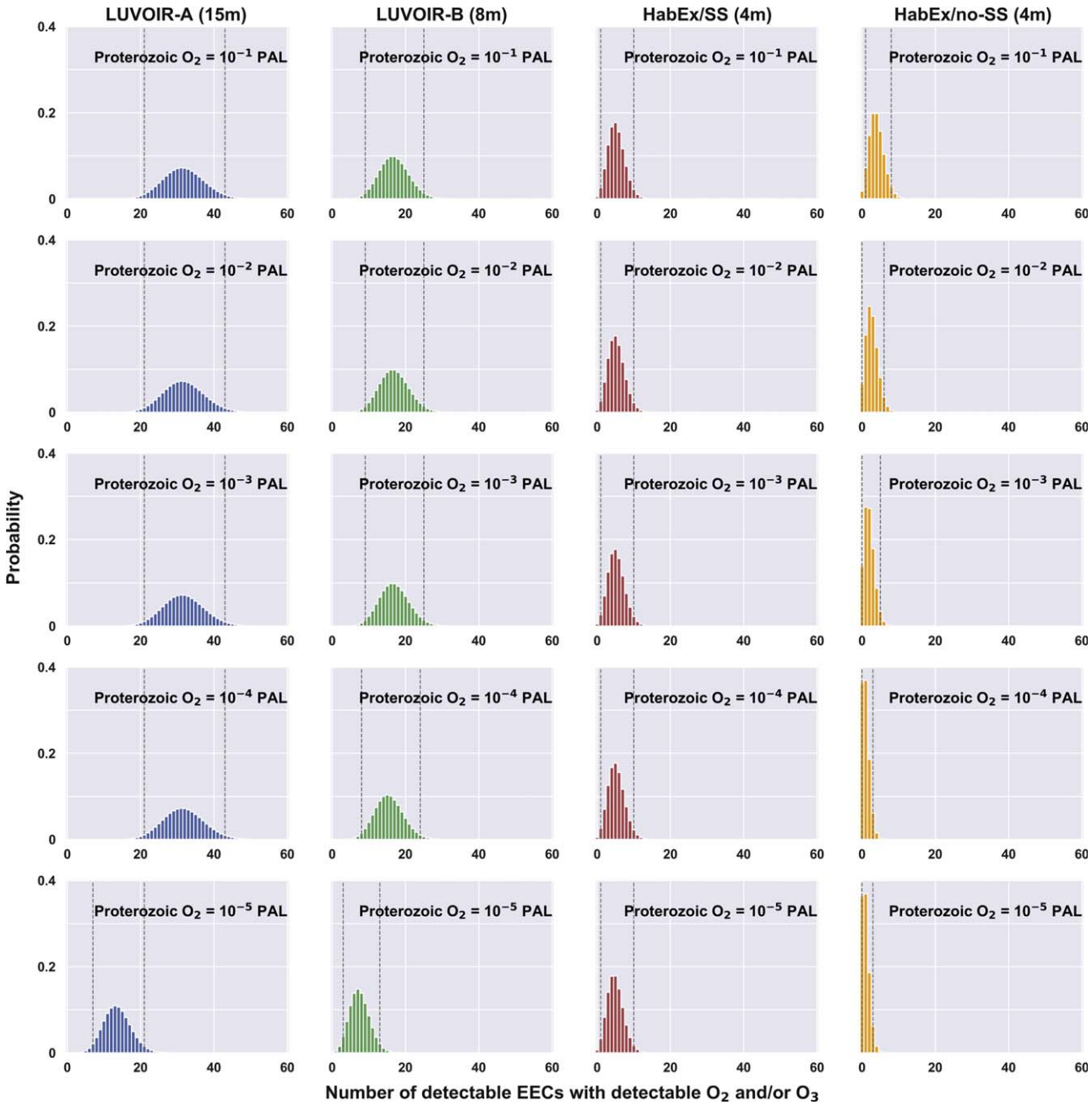

Figure 3. Probability distributions of the number of EECs on which an O_2 and/or O_3 signature could be detected, for $\eta_{\oplus} = 24\%$, assuming $f_E = 1$. For LUVOIR-A and LUVOIR-B, there are no mission-level false-negative scenarios and O_2 and/or O_3 will be detected on a number of EECs if they are all Earth-like. For HabEx/SS, there is a small mission-level false-negative scenario probability (0.5%–0.6%). For HabEx/no-SS, there are different mission-level false-negative scenarios depending on the level of O_2 we assume for the Proterozoic (probability of 1.8% for 10^{-1} PAL, 6.6% for 10^{-2} PAL, 13.8% for 10^{-3} PAL, and 36.4% for $\leqslant 10^{-4}$ PAL of Proterozoic O_2). Panels: 15 m segmented on-axis LUVOIR-A (leftmost), an 8 m segmented off-axis LUVOIR-B (center left), 4 m monolith off-axis HabEx/SS (starshade, center right), and 4 m monolith off-axis HabEx/no-SS (no starshade, rightmost). We assume that all EECs are Earth-like ($f_E = 1$). If the distribution reaches zero, there is a nonzero mission-level false-negative probability: O_2 and/or O_3 are not detected on any EECs despite all of them being Earth-like. We draw an age for EECs between Hadean, Archean, pre-GOE Proterozoic, post-GOE Proterozoic, and modern eras, and assume Hadean, Archean, and pre-GOE Proterozoic O_2 and O_3 levels are undetectable. Post-GOE Proterozoic O_2 concentrations range between 10^{-5} and 0.1 PAL. The vertical gray dotted lines represent the 95% confidence interval.

For LUVOIR-B, with Proterozoic levels of O_2 equal to or larger than 10^{-3} PAL, O_3 is detectable for EECs at all target distances for both modern and Proterozoic levels. At lower Proterozoic O_2 levels, O_3 is only detectable at <19.1 pc for 10^{-4} PAL and at <8.1 pc for 10^{-5} PAL. We find that LUVOIR-B has a 95% chance (for $f_E = 1$) of detecting 9–25 Earth-like EECs for Proterozoic O_2 levels of 0.1–10^{-3} PAL, 8–24 for 10^{-4} PAL, and 3–13 for 10^{-5} PAL. Similar to LUVOIR-A, for Proterozoic O_2 between 0.1 and 10^{-4} PAL, the distribution is well above zero, while for 10^{-5} PAL, the smallest number of detected Earth-like EECs is 1. Therefore, similarly to LUVOIR-A, there are no mission-level false-negative scenarios for LUVOIR-B as a mission if $f_E = 1$.

For HabEx/SS, with Proterozoic levels of O_2 equal to or larger than 10^{-4} PAL, O_3 is detectable for EECs at all target distances for both modern and Proterozoic levels. At 10^{-5}

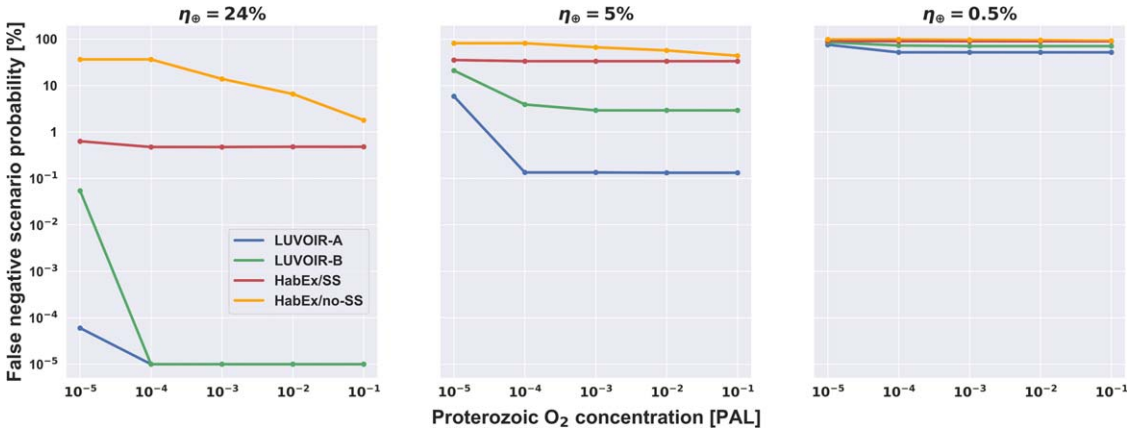

Figure 4. Probability of a mission-level false-negative scenario for different assumptions of Proterozoic O_2 concentrations, assuming $f_E = 1$, for $\eta_\oplus = 24\%$, 5%, and 0.5%. LUVOIR-A, given its large expected EEC yield, is most robust to lowering estimates of $\eta_\oplus$, allowing low probabilities of a false-negative scenario for an $\eta_\oplus$ estimate down to 5%. A mission-level false-negative scenario is defined as not detecting O_2 or O_3 on any of the observed EECs even though we assume they are all Earth-like. The lowest possible mission-level false-negative probability we can estimate is $<10^{-5}$% given that we perform 10^7 Monte Carlo simulations.

Table 3
Probability of a Mission-level False-negative Scenario for Different Assumptions of Proterozoic O_2 Concentrations, Assuming $f_E = 1$, for $\eta_\oplus = 24\%$

	15 m LUVOIR-A	8 m LUVOIR-B	4 m HabEx/ SS	4 m HabEx/ no-SS
$O_2 = 10^{-1}$ PAL	$<10^{-5}$%	$<10^{-5}$%	0.48%	1.8%
$O_2 = 10^{-2}$ PAL	$<10^{-5}$%	$<10^{-5}$%	0.48%	6.6%
$O_2 = 10^{-3}$ PAL	$<10^{-5}$%	$<10^{-5}$%	0.48%	13.8%
$O_2 = 10^{-4}$ PAL	$<10^{-5}$%	$<10^{-5}$%	0.48%	36.4%
$O_2 = 10^{-5}$ PAL	6.0×10^{-5}%	0.054%	0.63%	36.4%

Note. We draw a distance from the mission target list and an age for each EEC observed. A mission-level false-negative scenario is defined as not detecting O_2/O_3 on any of the observed EECs even though we assume they are all Earth-like. The lowest possible mission-level false-negative probability we can estimate is $<10^{-5}$% given that we perform 10^7 Monte Carlo simulations.

PAL, O_3 is only detectable on EECs at <13.9 pc. We find that HabEx/SS has a 95% chance of detecting 1–10 Earth-like EECs for all levels of Proterozoic O_2 considered (for $f_E = 1$). Because of the small number of EECs that HabEx can detect, the smallest possible number of Earth-like EECs for HabEx is 0. There is therefore a probability of a mission-level false-negative scenario with HabEx even if $f_E = 1$; however, it is only 0.5%–0.6%.

For HabEx/no-SS, O_3 is only detectable at certain distances for high-enough levels of Proterozoic O_2: <13.1 pc for 1 PAL, <9.8 pc for 10^{-1} PAL, <6.6 pc for 10^{-2} PAL, <4.1 pc for 10^{-3} PAL. At Proterozoic O_2 levels of 10^{-4} and 10^{-5} PAL, O_3 is not detectable at any distance. We find that HabEx/no-SS has a 95% chance (for $f_E = 1$) of detecting one to eight Earth-like EECs for Proterozoic levels of 0.1 PAL, zero to six for 10^{-2} PAL, zero to five for 10^{-3} PAL, and zero to three for 10^{-4} and 10^{-5} PAL. For all cases of Proterozoic O_2, HabEx/no-SS has a mission-level false-negative scenario where we do not detect O_2 or O_3 on any of the EECs we observe even though $f_E = 1$. That mission-level false-negative probability depends on the assumed Proterozoic O_2 level, and in the cases, we considered, it is 1.8% for 10^{-1}, 6.6% for 10^{-2}, 13.8% for 10^{-3}, and 36.4% for 10^{-4} and 10^{-5} PAL. We summarize these mission-level false-negative probabilities in Table 3 and Figure 4.

3.3.2. $\eta_\oplus = 5\%$

For the lower $\eta_\oplus$ estimate of 5%, Figure 5 shows that all distributions are shifted toward zero compared to Figure 3. The most important difference compared to Figure 3 is the existence and greater likelihood of mission-level false-negative scenarios for $f_E = 1$. LUVOIR-A still has negligible (~0.13%) mission-level false-negative scenarios for Proterozoic O_2 levels $\geqslant 10^{-4}$ PAL, but has a mission-level false-negative probability of 5.8% for Proterozoic O_2 of 10^{-5} PAL, for $f_E = 1$. LUVOIR-B has mission-level false-negative scenario probabilities for every Proterozoic O_2 case (2.9% for $\geqslant 10^{-3}$, 3.9% for 10^{-4}, and 21.0% for 10^{-5} PAL), for $f_E = 1$. HabEx/SS and HabEx/no-SS have significantly greater mission-level false-negative scenario probabilities. HabEx/SS has a 33.3% chance of a mission-level false-negative probability for Proterozoic $O_2 \geqslant 10^{-4}$ and 35.3% for 10^{-5} PAL, for $f_E = 1$. HabEx/no-SS has a 43.6% chance of a mission-level false-negative for Proterozoic O_2 of 10^{-1}, 56.9% for 10^{-2}, 66.3% for 10^{-3}, and 81.1% for $\leqslant 10^{-4}$ PAL, for $f_E = 1$. We summarize these mission-level false-negative probabilities in Table 4 and Figure 4.

3.3.3. $\eta_\oplus = 0.5\%$

The lowest estimate of $\eta_\oplus$ predicts that EECs are rare enough that it will be difficult to detect them in the first place (see Figure 1, right panel), as there is a 27%, 50%, and 81% chance of not detecting any of them for LUVOIR-A, LUVOIR-B, and HabEx, respectively. Because of this, for $f_E = 1$, there is a 51.6% chance of a mission-level false-negative scenario with LUVOIR-A for Proterozoic $O_2 \geqslant 10^{-4}$ PAL and 75.3% for Proterozoic O_2 of 10^{-5} PAL. LUVOIR-B has a 70.3% chance of a false negative for Proterozoic $O_2 \geqslant 10^{-3}$ PAL, 72.4% for 10^{-4} PAL, and 85.6% for 10^{-5} PAL, for $f_E = 1$. HabEx/SS has an 89.6% chance of a false negative for Proterozoic $O_2 \geqslant 10^{-4}$ PAL and a 90.1% chance for Proterozoic O_2 of 10^{-5} PAL, for $f_E = 1$. HabEx/no-SS has a 91.0%, 94.5%, 96.0%, and 97.9% chance of a false negative for Proterozoic O_2 of 10^{-1} PAL, 10^{-2} PAL, 10^{-3} PAL, and $\leqslant 10^{-4}$ PAL, for $f_E = 1$. To summarize, if $\eta_\oplus = 0.5\%$, all missions that we considered will

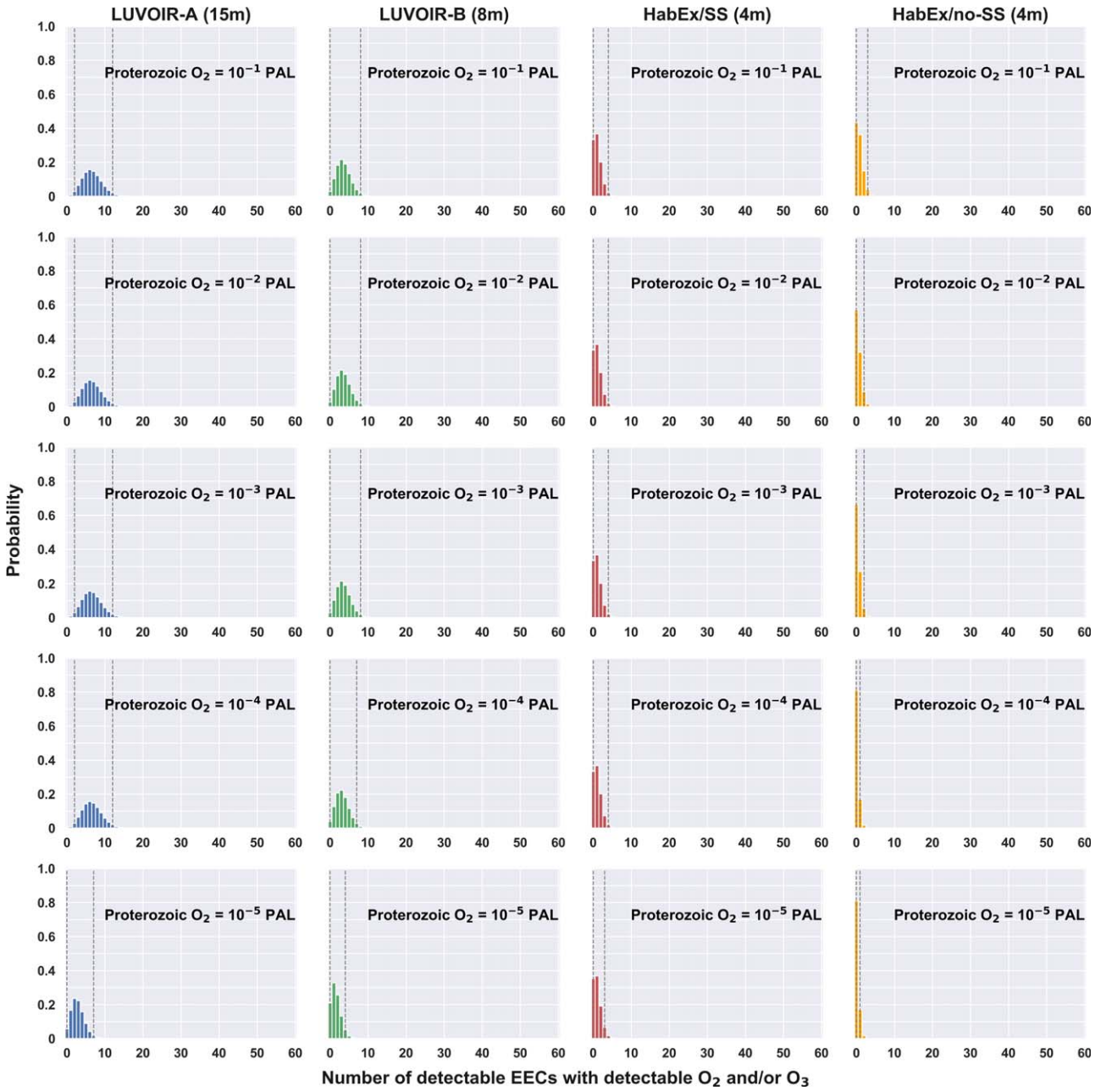

Figure 5. Same as Figure 3 but for $\eta_{\oplus} = 5\%$, assuming $f_E = 1$. For LUVOIR-A, there are no mission-level false-negative scenarios for Proterozoic $O_2 \geqslant 10^{-4}$ PAL. For Proterozoic O_2 of 10^{-5} PAL, there is a 5% chance of a mission-level false-negative scenario. For LUVOIR-B, there is a 1.4% chance of a mission-level false-negative scenario for Proterozoic $O_2 \geqslant 10^{-3}$ PAL, a 2.3% chance for 10^{-4} PAL, and a 20.2% chance for 10^{-5} PAL. For HabEx/SS, there is a 34.2% chance of a mission-level false-negative scenario for Proterozoic $O_2 \geqslant 10^{-4}$ PAL, and a 36.7% chance for 10^{-5} PAL. For HabEx/no-SS, the mission-level false-negative probabilities are as follows: 46.9% for 10^{-1}, 61.6% for 10^{-2}, 71% for 10^{-3}, and 84.5% for $\leqslant 10^{-4}$ PAL of Proterozoic O_2. Note that the scale of the *y*-axis is different than in Figure 3.

Table 4
Same as Table 3 but for $\eta_{\oplus} = 5\%$

	15 m LUVOIR-A	8 m LUVOIR-B	4 m HabEx/ SS	4 m HabEx/ no-SS
$O_2 = 10^{-1}$ PAL	0.13%	2.9%	33.3%	43.6%
$O_2 = 10^{-2}$ PAL	0.13%	2.9%	33.3%	56.9%
$O_2 = 10^{-3}$ PAL	0.13%	2.9%	33.3%	66.3%
$O_2 = 10^{-4}$ PAL	0.13%	3.9%	33.3%	81.1%
$O_2 = 10^{-5}$ PAL	5.8%	21.0%	35.3%	81.1%

Table 5
Same as Table 3 but for $\eta_{\oplus} = 0.5\%$

	15 m LUVOIR-A	8 m LUVOIR-B	4 m HabEx/ SS	4 m HabEx/ no-SS
$O_2 = 10^{-1}$ PAL	51.6%	70.3%	89.6%	91.0%
$O_2 = 10^{-2}$ PAL	51.6%	70.3%	89.6%	94.5%
$O_2 = 10^{-3}$ PAL	51.6%	70.3%	89.6%	96.0%
$O_2 = 10^{-4}$ PAL	51.6%	72.4%	89.6%	97.9%
$O_2 = 10^{-5}$ PAL	75.3%	85.6%	90.1%	97.9%

detect too few EECs to rule out false-negative scenarios even if $f_E = 1$. We do not include a reproduction of Figures 3 and 5 for $\eta_\oplus = 0.5\%$ as most panels simply show high peaks at zero, but we summarize the mission-level false-negative probabilities in Table 5 and Figure 4.

3.4. Using a Null Detection to Constrain the Fraction f_E of Earth-like EECs

In the highly likely scenario that only a fraction f_E of EECs are actually Earth-like, the number of planets on which we detect O_2/O_3 will be decreased by a factor f_E. For example, if only 10% of EECs are Earth-like ($f_E = 0.1$), assuming $\eta_\oplus = 24\%$, the peak in the first-row panel of Figure 3 for LUVOIR-A would shift from 31 to 3.1. For HabEx/SS, the peak of all panels would shift toward zero to create distributions with peaks at zero, making it unlikely that we will be able to detect O_2 or O_3 with HabEx/SS if f_E is low.

An instrument that lacks mission-level false negatives for $f_E = 1$, such as (for $\eta_\oplus = 24\%$) LUVOIR-A, LUVOIR-B, or HabEx/SS, will allow our observations to put a constraint on f_E even if we cannot detect O_2 or O_3. On the other hand, an instrument such as HabEx/no-SS will not allow us to make any inference about f_E in the event of a null detection, no matter the value of $\eta_\oplus$, as it could be caused by a mission-level false-negative scenario even when $f_E = 1$. For missions lacking mission-level false-negative scenarios when $f_E = 1$, how well we can constrain f_E depends on the mission and on the value of $\eta_\oplus$and is a function of the number of detectable EECs and of the ability to detect various levels of O_3.

For different values of f_E, we perform Monte Carlo simulations by resampling from the distributions in Figure 3 and weighing each draw by f_E, resulting in the number of EECs with detectable O_2 or O_3 for that value of f_E. We show the probability that at least one Earth-like EEC will have detectable O_2 or O_3 as a function of the fraction of Earth-like EECs, f_E, in Figure 6.

3.4.1. $\eta_\oplus = 24\%$

If we do not detect O_2 or O_3 on any EECs with LUVOIR-A, this null detection would mean that $f_E \leqslant 0.094$ for Proterozoic O_2 levels above 10^{-4} PAL, and $\leqslant 0.22$ for Proterozoic O_2 levels of 10^{-5} PAL. Therefore, a null detection with LUVOIR-A will constrain f_E with 95% confidence to $\leqslant 0.094$–0.22. Similarly, a null result with LUVOIR-B would mean that f_E is $\leqslant 0.18$ for Proterozoic O_2 levels above 10^{-3} PAL, $\leqslant 0.19$ for Proterozoic O_2 levels of 10^{-4} PAL, and $\leqslant 0.40$ for Proterozoic O_2 levels of 10^{-5} PAL. Therefore, a null detection with LUVOIR-B will constrain $f_E \leqslant 0.18$–0.40 with 95% confidence. HabEx/SS will also allow us to constrain $f_E \leqslant 0.56$ for Proterozoic O_2 levels above 10^{-4} PAL, and $\leqslant 0.59$ for Proterozoic O_2 levels of 10^{-5} PAL in the case of a null detection, all with 95% confidence. On the other hand, a null detection with HabEx/no-SS will not allow us to constrain f_E as there are mission-level false-negative scenarios where we do not detect O_2 or O_3 even if all EECs are Earth-like. These potential constraints on f_E are summarized in Figure 7.

This also implies that we have a 95% chance of detecting O_2 or O_3 on at least one EEC with LUVOIR-A for $f_E > 0.094$, with LUVOIR-B for $f_E > 0.18$, and with HabEx/SS for $f_E > 0.56$ (all for Proterozoic $O_2 \geqslant 10^{-4}$ PAL, which is the likely lower limit of the Proterozoic O_2 range). However, a null detection is possible with HabEx/no-SS for any value of f_E.

3.4.2. $\eta_\oplus = 5\%$

A lower $\eta_\oplus$ estimate of 5% would also impact the extent to which we could constrain f_E in the case of a null detection. If we do not detect O_2 or O_3 on any EECs with LUVOIR-A, we could constrain f_E with 95% confidence to $\leqslant 0.45$ for Proterozoic $O_2 \geqslant 10^{-4}$ PAL. However, for Proterozoic O_2 of 10^{-5} PAL, we will not be able to constrain f_E as there is a 6% chance of a mission-level false-negative scenario. Similarly, a null detection with LUVOIR-B will allow us to constrain f_E with 95% confidence to $\leqslant 0.85$ for Proterozoic $O_2 \geqslant 10^{-3}$ PAL, and to $\leqslant 0.92$ for Proterozoic O_2 of 10^{-4} PAL, but we will not be able to constrain it at lower Proterozoic levels as there is a 21% chance of a mission-level false negative for Proterozoic O_2 of 10^{-5} PAL. HabEx/SS and HabEx/no-SS have a 33%–35% and 44%–81% chance of a mission-level false negative, respectively, and therefore will not allow us to constrain f_E. These potential constraints on f_E are summarized in Figure 7.

This also implies that we have a 95% chance of detecting O_2 or O_3 on at least one EEC with LUVOIR-A for $f_E > 0.45$, with LUVOIR-B for $f_E > 0.85$. However, a null detection is possible with HabEx/SS and HabEx/no-SS for any value of f_E.

3.4.3. $\eta_\oplus = 0.5\%$

Using the lowest estimate of $\eta_\oplus = 0.5\%$ introduces the possibility of detecting no EECs in the first place, hence a null detection of O_2 and O_3 on all EECs is unlikely to help constrain f_E. In this scenario, none of the missions would inform us of the likely value of f_E in the case of a null detection, and it is likely we would not detect O_2 or O_3 on any EECs even if they are all Earth-like.

4. Discussion

If EECs are all Earth-like in that they develop Earth-like oxygenic photosynthesis that oxygenates their atmosphere, we should detect O_2 or O_3 with LUVOIR-A, LUVOIR-B, and HabEx/SS (with starshade) as long as $\eta_\oplus$ is sufficiently large. For $\eta_\oplus = 24\%$, LUVOIR-A and LUVOIR-B have a negligible chance of a mission-level false-negative scenario, where EECs are all Earth-like but we do not detect O_2 or O_3 on any of them. HabEx/SS similarly has only a very small chance of such a false negative. For $\eta_\oplus = 5\%$, for Proterozoic $O_2 \geqslant 10^{-4}$ PAL, LUVOIR-A similarly has a negligible chance of a mission-level false-negative scenario, and LUVOIR-B's chance of a mission-level false-negative scenario is <4%, while HabEx/SS has a 33% chance of one. For $\eta_\oplus = 0.5\%$, it is reasonably likely that no EEC will be detected in the first place: there is a 27%, 50%, and 81% chance that LUVOIR-A, LUVOIR-B, and HabEx do not detect any. As a result, all missions have a significant probability of a mission-level false-negative scenario even if all EECs are Earth-like ($f_E = 1$). For any value of $\eta_\oplus$, HabEx/no-SS has a significant chance of a mission-level false-negative scenario, showing therefore that a starshade is crucial to HabEx's design.

If we detect O_2 or O_3 on a number of EECs that is lower than the number we would expect if $f_E = 1$, we will be able to constrain the fraction of EECs that are actually Earth-like (f_E) with all instruments. In the event that we do not detect O_2 or O_3 on any EECs, we will also be able to constrain f_E to different extents based on the mission and based on the value of $\eta_\oplus$. For $\eta_\oplus = 24\%$, LUVOIR-A, LUVOIR-B, and HabEx/SS will

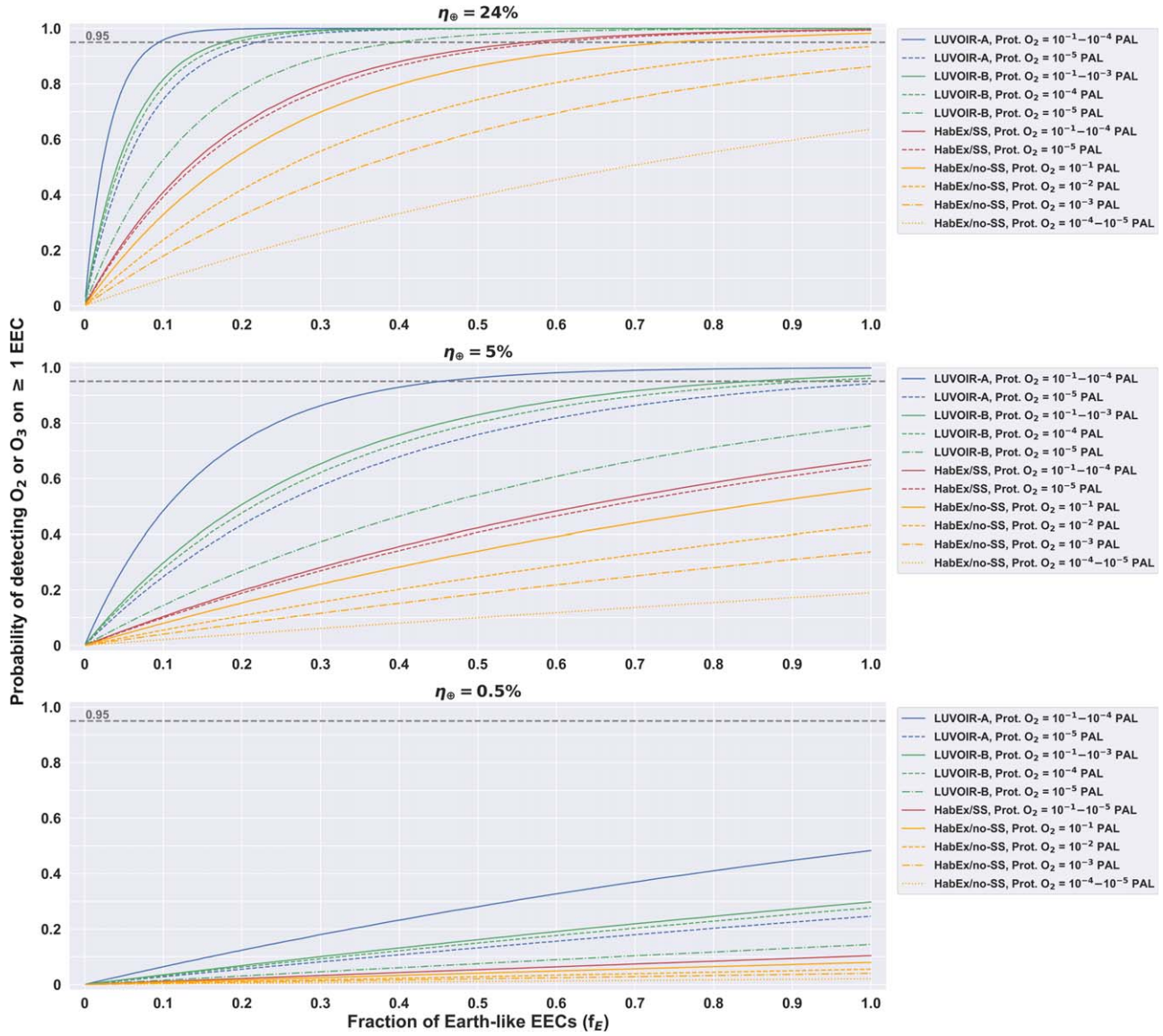

Figure 6. Probability of detecting O_2 or O_3 on at least one EEC as a function of the fraction of Earth-like EECs (f_E), for $\eta_\oplus = 24\%$, 5%, and 0.5%. The intersection of each curve with the 95% vertical gray line denotes the upper limit that will be placed on f_E at 95% confidence in the event of a null detection of O_2 and O_3 on all observed EECs. If the curve does not intersect the 95% gray line, a constraint cannot be placed on f_E in the event of a null detection due to the existence of mission-level false-negative scenarios. For a given Proterozoic O_2 level, the mission with the highest expected number of detectable EECs allows for the best constraint to be placed on f_E.

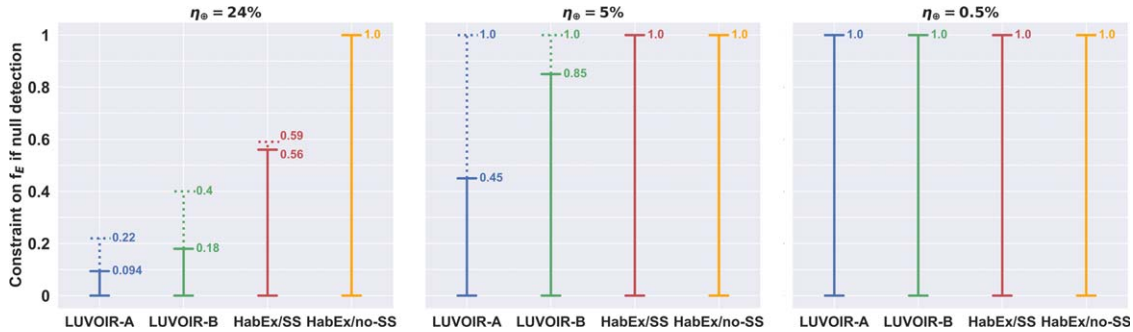

Figure 7. Constraint that we will be able to place on f_E, with 95% confidence, in the case of a null detection where we do not detect O_2 or O_3 on any EECs for $\eta_\oplus = 24\%$, 5%, and 0.5%. These constraints also imply that for f_E greater than the upper limits, we have a 95% chance to detect O_3 on at least 1 EEC. Solid lines: constraints for the Proterozoic O_2 level of 10^{-1}–10^{-4} PAL; dashed lines: constraints for the Proterozoic O_2 level of 10^{-5} PAL. An upper limit of 1.0 for the constraint on f_E means that there is a $\geqslant$5% chance of a mission-level false-negative scenario, where we do not detect O_2 or O_3 on any EECs even if they are all Earth-like, and so f_E cannot be constrained. Note that LUVOIR-B's constraint on f_E will be $\leqslant$19% for $\eta_\oplus = 24\%$ and Proterozoic O_2 level of 10^{-4} PAL, and $\leqslant$92% for $\eta_\oplus = 5\%$ and Proterozoic O_2 level of 10^{-4} PAL.

allow us to constrain $f_E \leqslant 0.094$, $f_E \leqslant 0.18$, and $f_E \leqslant 0.56$, respectively, with 95% confidence, for Proterozoic O_2 levels $\geqslant 10^{-4}$ PAL. On the other hand, HabEx/no-SS will not allow us to constrain f_E due to the existence of mission-level false-negative scenarios. For $\eta_\oplus = 5\%$, LUVOIR-A and LUVOIR-B will allow us to constrain $f_E \leqslant 0.45$ and $f_E \leqslant 0.85$, respectively, while HabEx/SS and HabEx/no-SS will not allow us to constrain f_E. For $\eta_\oplus = 0.5\%$, none of the missions will allow us to constrain f_E due to the high probability of mission-level false-negative scenarios. This illustrates the fact that a mission with a higher expected yield of EECs is more robust to the uncertainty in the value of $\eta_\oplus$, and that further constraining $\eta_\oplus$ should be a priority to allow us to make predictions of the outcome of future observations.

As HabEx and LUVOIR are still mission concepts, their specifications are not yet finalized. The number of EECs that we can detect with either instrument depends on the set of specificities we choose from Stark et al. (2019). Particularly, our analysis was performed using a high throughput scenario for both instruments, which is likely optimistic. We considered a pessimistic scenario where we use a low throughput and recreated Figure 1. For $\eta_\oplus = 24\%$, we found that the number of EECs that we could detect with LUVOIR-A decreases from 62 to 50, with LUVOIR-B from 33 to 27, and with HabEx from 10 to 8. In this low-throughput scenario, the peaks of Figure 3 are shifted to the left as well. For example, for $\eta_\oplus = 24\%$ and Proterozoic O_2 levels of 10^{-5} PAL, this increases the mission-level false-negative scenario probability from 0.6% (in the high throughput scenario) to 1.8% for HabEx/SS, and from 36% to 45% for HabEx/no-SS. On the other hand, assuming a low throughput does not introduce any mission-level false-negative scenario for LUVOIR-A and LUVOIR-B for this estimate of $\eta_\oplus$.

Exoplanet atmospheres are expected to be diverse in their composition, even if they are inhabited by Earth-like life. In this work, we considered an Earth-like background atmosphere in our S/N calculations that includes gases such as O_2, O_3, CO_2, H_2O, N_2, and CH_4. However, EEC atmospheres may contain different gases that could affect the S/Ns of O_2 and O_3 features. Observations would require a full atmospheric retrieval study to confirm O_2 and O_3 detections (Feng et al. 2018).

The HabEx/SS integration times we calculated for O_2 are greater than those reported in the HabEx report (Gaudi et al. 2020). We find that the integration times necessary to detect O_2 and O_3 with HabEx/SS are one to four orders of magnitude larger than in their study (see their Figures 3.3–7). The reason for this is that we included various gases such as O_3, CO_2, and H_2O in our background atmosphere, while the HabEx report calculations only considered N_2 as a background gas. Because of this, in their calculations, the difference between the continuum (N_2 only) and the absorption spectrum (N_2 and O_2) is large at $\sim$0.2 μm as N_2 does not absorb there. In Earth-like atmospheres, species such as O_3 and H_2O absorb at these short wavelengths. This minimizes the difference between the continuum spectra (without O_2) and the absorber spectra (with O_2) at 0.2 μm. Therefore, we do not expect the signal to be large at 0.2 μm for Earth-like planets. We note, however, that our calculations agree closely with simulations made using the Robinson et al. (2016) model when including additional background gases (see Appendix D for further details).

The level of O_2 during the Proterozoic is highly uncertain (Reinhard et al. 2017; Olson et al. 2018b), with lower and upper constraints from various geochemical records and modeling efforts that vary by four orders of magnitude from 10^{-5} to 10^{-1} PAL (Pavlov & Kasting 2002; Lyons et al. 2014; Planavsky et al. 2014; Olson et al. 2018b). The lower limit on Proterozoic O_2 is inferred from the end of mass-independent fractionation of S isotopes in the wake of the "Great Oxidation Event." The upper limit of 10^{-1} PAL comes from the observation that the deep ocean remained anoxic throughout the Proterozoic, implying that surface environments were only mildly oxygenated. More recent work by Planavsky et al. (2014) provides an upper limit of 10^{-3} PAL by leveraging the absence of Cr isotope fractionation in Proterozoic marine sediments. Although such geochemical data constrains the range of Proterozoic O_2 between 10^{-5} and 10^{-3} PAL, the low end of that range is difficult to explain in biogeochemical and photochemical models if oxygenic photosynthesis was occurring at or near modern rates (Ozaki et al. 2019). A number of previous studies that considered the detectability of Proterozoic O_2 remained above an O_2 threshold of approximately 10^{-3} PAL (Reinhard et al. 2017; Schwieterman et al. 2018; Gaudi et al. 2020). We considered Proterozoic O_2 levels as low as 10^{-5} PAL to span the full range of estimates existing in the literature and permissible by existing geochemical data, and to allow us to consider every possible scenario in the search for biosignatures with HabEx and LUVOIR, but we note that the low estimate of 10^{-5} PAL may be less likely than higher estimates.

In this work, we have based the definition of "Earth-like" on the premise that EECs would oxygenate roughly following Earth's oxygenation trajectory and have used Earth history as a prior. However, the timing and timescale of oxygenation on Earth are not well understood, and it is possible that EECs would follow a different oxygenation trajectory. If Earth is rare and EECs never oxygenate past Archean levels, then as per our statistical definition of "Earth-like," EECs would not be statistically Earth-like. Similarly, if only a small fraction of EECs oxygenate past Archean levels, then statistically, only a small fraction of EECs are Earth-like and f_E is small. This is something we will be able to test with a mission that includes a large-aperture telescope, as whether or not we detect O_3, it will allow us to infer a maximum value of f_E. We have also explored other possible oxygenation scenarios for $\eta_\oplus = 24\%$: (1) Proterozoic O_2 levels are as high as modern levels, and (2) modern levels of O_2 are as low as those of the Proterozoic. In scenario 1, the distributions of detectable EECs with detectable O_3 are similar to those of Figure 3's top panels (where Proterozoic O_2 is 0.1 PAL) for LUVOIR-A, LUVOIR-B, and HabEx/SS. That is because O_3 is detectable on every target EEC at every distance in the case where Proterozoic O_2 is 0.1 PAL, so assuming that Proterozoic levels are as high as modern levels will produce the same results. For HabEx/no-SS, the mission-level false-negative scenario probability decreases from 1.8%–36% to 1.5%. In scenario 2, the distributions of detectable EECs with detectable O_3 are the same as those in Figure 3 in most cases: for Proterozoic $O_2 \geqslant 10^{-4}$ PAL for LUVOIR-A, for Proterozoic $O_2 \geqslant 10^{-3}$ PAL for LUVOIR-B, and for Proterozoic $O_2 \geqslant 10^{-4}$ PAL for HabEx/SS. That is because, again, for Proterozoic O_2 above those levels, O_3 is detectable on every target EEC at every distance. For lower Proterozoic O_2, there is still no mission-level false-negative

scenario probability for the 10^{-5} PAL case for LUVOIR-A and for the 10^{-4} PAL case for LUVOIR-B, but it increases to 1.3% for 10^{-5} PAL for LUVOIR-B and to 1.8% for 10^{-5} PAL for HabEx/SS. For HabEx/no-SS, the mission-level false-negative scenario probability increases from 1.8%–36% to 3.3%–100%. Further oxygenation scenarios could be explored in future work.

Whether the origination of life is an extremely rare occurrence or common throughout the universe is a heavily debated topic. The frequency of life originating on habitable planets is highly uncertain (e.g., Sandberg et al. 2018), and the fact that life originated on Earth does not constrain this frequency very much (Spiegel & Turner 2012). However, the large number of exoplanets that we may be able to soon characterize with future missions offers an opportunity to test whether the origin of life is common. If we observe a number of EECs and detect clearly biogenic O_2 or O_3 on at least one of them, the origination of life on habitable planets must be common. Conversely, if we do not detect O_2 or O_3 with a LUVOIR-A or LUVOIR-B-like instrument, we will know that EECs are generally unlikely to be Earth-like. This could mean that either the origination of life is very rare, or that life rarely develops oxygenic photosynthesis. In this scenario of null-life detection, we may be able to improve our estimate of the probability of the origination of life using a Bayesian analysis similar to that of Spiegel & Turner (2012) and Kipping (2020). Future work could look at what constraints we can put on the origination of life using a Bayesian analysis for different observation scenarios.

5. Conclusions

In this article, we considered whether observations of EECs with HabEx and LUVOIR may inform us of the fraction of EECs that are Earth-like (f_E) in that they develop Earth-like, O_2-producing life (oxygenic photosynthesis) and become oxygenated roughly following Earth's oxygenation history. To do that, we first considered the probability that HabEx and LUVOIR will detect O_2 and/or O_3 on EECs. Then, we determined whether a null detection, where we do not detect O_2 or O_3 on any EEC, would allow us to constrain f_E. We adopted a statistical approach to this problem. Instead of investigating false negatives on particular planets, we determined whether there might be mission-level false negatives for missions such as LUVOIR and HabEx based on all of the information that we can gain from all of the EECs that these missions can be expected to observe. We considered four different telescope designs: 15 m segmented on-axis LUVOIR-A, 8 m segmented off-axis LUVOIR-B, 4 m monolith off-axis HabEx with a starshade ("HabEx/SS," where "SS" refers to HabEx's starshade), and 4 m monolith off-axis HabEx without a starshade ("HabEx/no-SS"). We also considered three different estimates for $\eta_\oplus$: 24%, following Stark et al. (2019); 5%, following Pascucci et al. (2019); and 0.5%, following Neil & Rogers (2020). In each case, we explore five different levels of Proterozoic O_2 (10^{-1}–10^{-5} PAL), but we note that the lowest end of this range (10^{-5} PAL) is less likely as it is difficult to explain if oxygenic photosynthesis was occurring at modern rates. Therefore, these conclusions report our results for Proterozoic O_2 levels between 10^{-1} and 10^{-4} PAL. The main conclusions of this article are:

1. First, we considered the possibility of not detecting any EECs. The probability of that occurring depends strongly on both $\eta_\oplus$ and the mirror diameter of the instrument. For the cases we considered, these probabilities are:
   (a) For $\eta_\oplus = 24\%$: 0% for all missions.
   (b) For $\eta_\oplus = 5\%$: 11% for HabEx, 0% for LUVOIR-A and LUVOIR-B.
   (c) For $\eta_\oplus = 0.5\%$: 27%, 50%, and 81%, for LUVOIR-A, LUVOIR-B, and HabEx.
2. Second, we considered the possibility of mission-level false-negative scenarios where we do not detect O_2 or O_3 on any of the EECs we observe, even if they were all Earth-like ($f_E = 1$). We find that the main factor determining whether such mission-level false-negative scenarios exist is the yield of detectable EECs, which mainly depends on the value of $\eta_\oplus$ and on the instrument's mirror diameter. We found that the level of Proterozoic O_2 we assume does not matter much for these missions, as long as it is above 10^{-4} PAL, as in that case, O_3 is detectable on all target EECs with LUVOIR-A, LUVOIR-B, and HabEx/SS. The probabilities of a mission-level false-negative scenario are as follows:
   (a) For $\eta_\oplus = 24\%$: 0% for LUVOIR-A and LUVOIR-B; 0.5% for HabEx/SS; 2%−36% for HabEx/no-SS.
   (b) For $\eta_\oplus = 5\%$: 0.1% for LUVOIR-A; 3%–4% for LUVOIR-B; 33% for HabEx/SS; 44%–81% for HabEx/no-SS.
   (c) For $\eta_\oplus = 0.5\%$: 52% for LUVOIR-A; 70%–72% for LUVOIR-B; 90% for HabEx/SS; 91%–98% for HabEx/no-SS.
3. Finally, we considered whether we could constrain the fraction of EECs that are Earth-like (f_E) even if we do not detect O_2 or O_3 on any EEC. The extent to which we may be able to constrain f_E in the case of a null detection depends on the mission and relies primarily on a sufficiently large number of detectable EECs as well as on the ability to detect low levels of O_3.
   (a) For $\eta_\oplus = 24\%$, a null detection with LUVOIR-A, LUVOIR-B, or HabEx/SS will allow us to constrain the fraction of Earth-like EECs: $f_E \leqslant 0.094$ with LUVOIR-A, $f_E \leqslant 0.18$ with LUVOIR-B, and $f_E \leqslant 0.56$ with HabEx/SS, all with 95% confidence and for Proterozoic O_2 levels $\geqslant 10^{-4}$ PAL. These constraints also imply that for f_E greater than these upper limits, we have a 95% chance of detecting O_3 on at least one EEC, and are therefore likely to do so with LUVOIR-A for $f_E \geqslant 0.094$, with LUVOIR-B for $f_E \geqslant 0.18$, and with HabEx for $f_E \geqslant 0.56$.
   (b) For $\eta_\oplus = 5\%$, a null detection with LUVOIR-A or LUVOIR-B would similarly allow us to constrain $f_E \leqslant 0.45$ with LUVOIR-A, and $f_E \leqslant 0.85$ with LUVOIR-B, with 95% confidence and for Proterozoic O_2 levels $\geqslant 10^{-4}$ PAL. HabEx/SS and HabEx/no-SS on the other hand will not allow us to constrain f_E due to the existence of mission-level false-negative scenarios even for $f_E = 1$. This also means that we may not detect O_3 on any EECs with HabEx/SS and HabEx/no-SS, while we are likely to do so with LUVOIR-A for $f_E \geqslant 0.45$, and with LUVOIR-B for $f_E \geqslant 0.85$.
   (c) For $\eta_\oplus = 0.5\%$, none of the missions would allow us to constrain f_E in the event of a null detection, as all

have mission-level false-negative scenarios with a probability greater than 5%. Therefore, despite the fact that missions with larger aperture mirrors are more robust to uncertainties in $\eta_{\oplus}$, all missions are vulnerable to inconclusive null detections if $\eta_{\oplus}$ is as low as 0.5%.

We thank our anonymous reviewer for a constructive and insightful review that helped us better this manuscript. We thank Chris C. Stark for sharing and guiding us through his target list database for HabEx and LUVOIR. We also thank him, as well as Giada Arney, for help in verifying assumptions made for the LUVOIR mission design. We thank Jacob L. Bean for insightful discussions on the detectability of O_2 and O_3. This work was supported by the NASA Astrobiology Program grant No. 80NSSC18K0829 and benefited from participation in the NASA Nexus for Exoplanet Systems Science research coordination network. S.L.O. acknowledges support from the T.C. Chamberlin Postdoctoral Fellowship in the Department of the Geophysical Sciences at the University of Chicago. S.L.O. additionally acknowledges support from the NASA Habitable Worlds Program. T.D.K. acknowledges funding from the 51 Pegasi b Fellowship in Planetary Astronomy sponsored by the Heising-Simons Foundation. P.P. acknowledges support from the James S. McDonnell Foundation.

Appendix A
Introduction to Appendices

Computing the integration times required for a detection of O_2 or O_3 requires two simulation components: a planetary spectrum model and a noise model. Molecular signatures for each atmospheric species of interest were computed by taking the difference between the continuum (all atmospheric species except the species of interest) and absorbance (all atmospheric species) spectra. The noise and S/N calculations involve the characterization of the different factors impacting the sensitivity. The signal refers to the planetary photons that are successfully counted by the instrument, after considering the telescope, coronagraph/starshade, optics, and detector efficiencies. The noise is a combination of effects that includes the Poisson noise introduced by the planetary photons, noise from residual photons from the star, noise from background sources (e.g., exozodi and local zodiacal fluxes), and the intrinsic noise introduced by the detector.

In order to further expand on these components, the appendix is therefore structured as follows: (1) we show geometric albedo spectra used to calculate the S/N of O_2 and O_3 detections with PSG, (2) we present integration times required for a 5σ detection of O_2 and O_3 at 10, 15, and 20 pc calculated using PSG, (3) we compare our calculated integration times to those calculated using the model of Robinson et al. (2016), and (4) we further detail the assumptions made in simulating observations with LUVOIR and HabEx using PSG.

Appendix B
Geometric Albedo Spectra

As described in Section 2.4, we calculate the S/N of O_2 and O_3 detections by first simulating two spectra: one absorbing spectrum with all atmospheric species, and one continuum spectrum with all atmospheric species except the chosen absorber (either O_2 or O_3). In Figure 8, we show both spectra for O_3 (left) and O_2 (right) at the six O_2 levels considered in this work.

O_3's strong feature at $\sim$0.25 μm allows for it to be detected even at very low levels. On the other hand, O_2's strongest feature at 0.76 μm becomes difficult to detect at Proterozoic O_2 levels. We note here that we added the collision-induced O_2–O_2 absorption bands in the UV (Wulf bands) between 0.24 and 0.3 μm as well as the Herzberg O_2 continuum bands (Fally et al. 2000), and the Herzberg O_2 band system (Jenouvrier et al. 1999; Mérienne et al. 2000, 2001), neither of which are included in the HITRAN database.

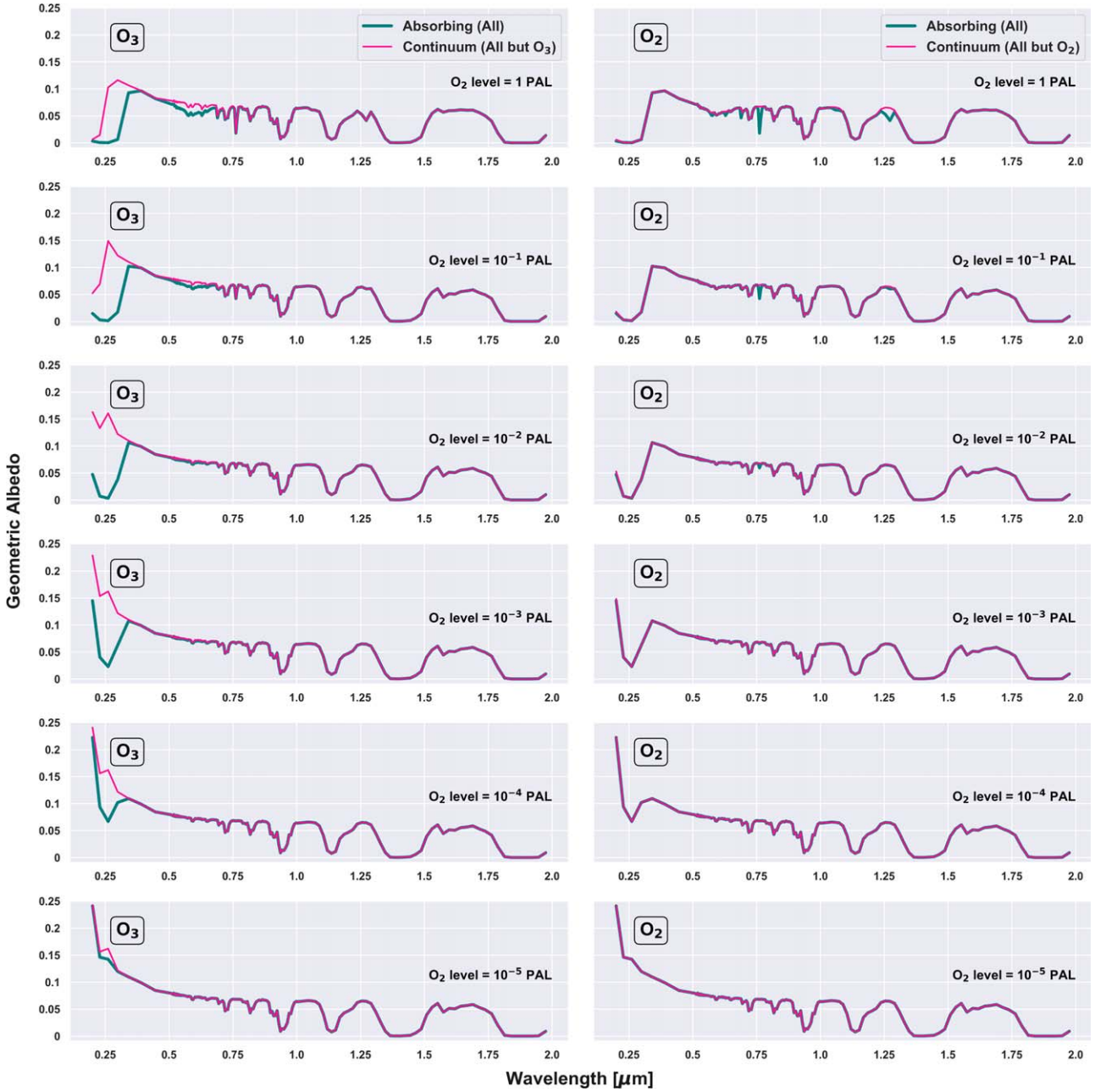

Figure 8. Continuum and absorption spectra for O_3 (left) and O_2 (right) detection for an Earth-like planet orbiting at 1 au around a Sun-like star at 5 pc with an exposure time of 1000 hr. We consider six different O_2 levels from 10^{-5} to 1 PAL. Background gases include gases such as N_2, H_2O, CH_4, and CO_2.

Appendix C
Integration Times at 10, 15, and 20 pc

Here we present the integration times required for a 5σ detection of O_2 and O_3 at 10, 15, and 20 pc calculated using PSG (Villanueva et al. 2018) in Tables 6–8 for LUVOIR-A, LUVOIR-B, HabEx/SS, and HabEx/no-SS (see Section 2.4 for details and for calculations at 5 pc).

Table 6
Integration Times (hr) with LUVOIR-A (15 m), LUVOIR-B (8 m), HabEx/SS (4 m), and HabEx/no-SS (4 m) to Yield a 5σ Detection of O_2 and O_3 for an Earth-like Planet without Clouds at 10 pc for Six Different O_2 Levels

	15 m LUVOIR-A	8 m LUVOIR-B	4 m HabEx/SS	4 m HabEx/no-SS
$O_2 = 1$ PAL	O_2: 5.93 hr	O_2: 47.7 hr	O_2: 182.3 hr	O_2: 1121.5 hr
	O_3: 1.04 hr	O_3: 4.6 hr	O_3: 3.33 hr	O_3: 332.8 hr
$O_2 = 10^{-1}$ PAL	O_2: 30.1 hr	O_2: 241.1 hr	O_2: 877.8 hr	O_2: 5376.6 hr
	O_3: 1.71 hr	O_3: 6.68 hr	O_3: 2.92 hr	O_3: 1217.6 hr
$O_2 = 10^{-2}$ PAL	O_2: 288.5 hr	O_2: 2310.8 hr	O_2: 7233.8 hr	O_2: 5.1×10^4 hr
	O_3: 2.94 hr	O_3: 10.8 hr	O_3: 3.12 hr	O_3: 1.1×10^4 hr
$O_2 = 10^{-3}$ PAL	O_2: 4603.8 hr	O_2: 3.7×10^4 hr	O_2: 7.5×10^4 hr	O_2: 8.1×10^5 hr
	O_3: 5.24 hr	O_3: 19.31 hr	O_3: 4.73 hr	O_3: 2.8×10^4 hr
$O_2 = 10^{-4}$ PAL	O_2: 2.1×10^5 hr	O_2: 1.7×10^6 hr	O_2: 1.4×10^6 hr	O_2: 3.8×10^7 hr
	O_3: 23.53 hr	O_3: 89.33 hr	O_3: 14.73 hr	O_3: 1.8×10^6 hr
$O_2 = 10^{-5}$ PAL	O_2: 1.5×10^8 hr	O_2: 1.9×10^7 hr	O_2: 7.9×10^7 hr	O_2: 3.3×10^9 hr
	O_3: 740.5 hr	O_3: 2884.0 hr	O_3: 387.9 hr	O_3: 3.8×10^8 hr

Table 7
Integration Times (hr) with LUVOIR-A (15 m), LUVOIR-B (8 m), HabEx/SS (4 m), and HabEx/no-SS (4 m) to Yield a 5σ Detection of O_2 and O_3 for an Earth-like Planet without Clouds at 15 pc for Six Different O_2 Levels

	15 m LUVOIR-A	8 m LUVOIR-B	4 m HabEx/SS	4 m HabEx/no-SS
$O_2 = 1$ PAL	O_2: 30.56 hr	O_2: 320.0 hr	O_2: 951.7 hr	O_2: 1.6×10^4 hr
	O_3: 3.34 hr	O_3: 20.7 hr	O_3: 13.2 hr	O_3: 2963.5 hr
$O_2 = 10^{-1}$ PAL	O_2: 154.8 hr	O_2: 1614.1 hr	O_2: 4411.1 hr	O_2: 8.0×10^4 hr
	O_3: 5.27 hr	O_3: 28.31 hr	O_3: 11.27 hr	O_3: 1.1×10^4 hr
$O_2 = 10^{-2}$ PAL	O_2: 1485.9 hr	O_2: 1.5×10^4 hr	O_2: 3.6×10^4 hr	O_2: 7.6×10^5 hr
	O_3: 8.93 hr	O_3: 44.8 hr	O_3: 11.88 hr	O_3: 9.5×10^4 hr
$O_2 = 10^{-3}$ PAL	O_2: 2.4×10^4 hr	O_2: 2.5×10^5 hr	O_2: 3.5×10^5 hr	O_2: 1.2×10^7 hr
	O_3: 16.2 hr	O_3: 81.2 hr	O_3: 17.85 hr	O_3: 2.5×10^5 hr
$O_2 = 10^{-4}$PAL	O_2: 1.1×10^6 hr	O_2: 1.1×10^7 hr	O_2: 5.8×10^6 hr	O_2: 5.7×10^8 hr
	O_3: 77.1 hr	O_3: 394.7 hr	O_3: 55.1 hr	O_3: 1.7×10^7 hr
$O_2 = 10^{-5}$ PAL	O_2: 9.8×10^7 hr	O_2: 1.0×10^9 hr	O_2: 3.3×10^8 hr	O_2: 5.0×10^{10} hr
	O_3: 2532.2 hr	O_3: 1.3×10^4 hr	O_3: 1442.8 hr	O_3: 3.4×10^9 hr

Table 8
Integration Times (hr) with LUVOIR-A (15 m), LUVOIR-B (8 m), HabEx/SS (4 m), and HabEx/no-SS (4 m) to Yield a 5σ Detection of O_2 and O_3 for an Earth-like Planet without Clouds at 20 pc for Six Different O_2 Levels

	15 m LUVOIR-A	8 m LUVOIR-B	4 m HabEx/SS	4 m HabEx/no-SS
$O_2 = 1$ PAL	O_2: 95.84 hr	O_2: 1564.0 hr	O_2: 1.4×10^4 hr	O_2: 1.6×10^5 hr
	O_3: 9.12 hr	O_3: 69.71 hr	O_3: 38.67 hr	O_3: 2.4×10^4 hr
$O_2 = 10^{-1}$ PAL	O_2: 465.8 hr	O_2: 8033.6 hr	O_2: 6.4×10^4 hr	O_2: 8.9×10^5 hr
	O_3: 13.2 hr	O_3: 91.1 hr	O_3: 31.9 hr	O_3: 8.9×10^4 hr
$O_2 = 10^{-2}$ PAL	O_2: 4441.8 hr	O_2: 7.7×10^4 hr	O_2: 3.4×10^5 hr	O_2: 8.6×10^6 hr
	O_3: 21.6 hr	O_3: 142.1 hr	O_3: 33.23 hr	O_3: 7.8×10^5 hr
$O_2 = 10^{-3}$ PAL	O_2: 7.1×10^4 hr	O_2: 1.2×10^6 hr	O_2: 1.7×10^6 hr	O_2: 1.4×10^8 hr
	O_3: 39.4 hr	O_3: 258.9 hr	O_3: 49.7 hr	O_3: 2.1×10^6 hr
$O_2 = 10^{-4}$ PAL	O_2: 3.3×10^6 hr	O_2: 5.8×10^7 hr	O_2: 1.9×10^7 hr	O_2: 6.5×10^9 hr
	O_3: 195.4 hr	O_3: 1292.4 hr	O_3: 152.3 hr	O_3: 1.4×10^8 hr
$O_2 = 10^{-5}$ PAL	O_2: 2.9×10^8 hr	O_2: 5.1×10^9 hr	O_2: 1.0×10^9 hr	O_2: 5.7×10^{11} hr
	O_3: 6625.7 hr	O_3: 4.4×10^4 hr	O_3: 3983.3 hr	O_3: 2.9×10^{10} hr

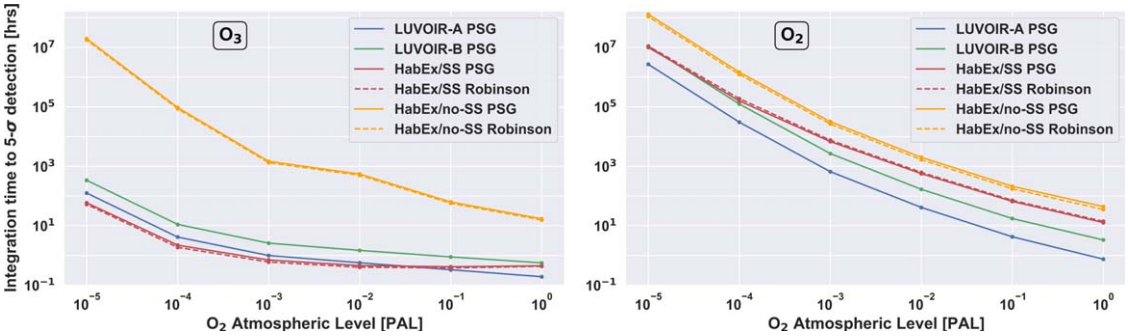

Figure 9. Integration times necessary to detect O_3 (left) and O_2 (right) at 5σ as a function of O_2 atmospheric level, for an Earth-like planet orbiting at 1 au around a Sun-like star at 5 pc. Solid lines: PSG calculations, dashed lines: Robinson et al.'s (2016) model. Blue: LUVOIR-A, green: LUVOIR-B, red: HabEx/SS, orange: HabEx/no-SS.

Appendix D
Comparison to Robinson et al. (2016) Model

In Figure 9, we compare the integration times we calculated using PSG to those calculated using the Robinson et al. (2016) model for an Earth-like planet at 5 pc. In both cases, we calculate the integration times following the method outlined in Section 2.4. We note here that Robinson et al.'s (2016) model was updated for these calculations according to the latest figures reported in the Final Reports (The LUVOIR Team et al. 2019; Gaudi et al. 2020), and we include Earth-like background gases such as O_2, O_3, N_2, H_2O, CO_2, and CH_4 in the atmosphere. Therefore, the Robinson et al. (2016) model calculations may not match those reported in Figures 3.3–7 of the HabEx report (Gaudi et al. 2020).

The integration times calculated using each model agree with each other very closely. This figure also exhibits the fact that HabEx/SS performs very well in the UV thanks to its starshade and despite its small size. Because of that, it outperforms LUVOIR-A and LUVOIR-B at low O_3 concentrations.

Appendix E
Simulation Parameters

The planetary photons being measured by the detector go through a series of optical systems that each can be assumed to have a specific efficiency or throughput. Background sources also go through the same optical path as the planetary photons. We define the end-to-end throughput for the planetary fluxes as $T_{\rm total} = T_{\rm Tele} \times T_{\rm coronagraph} \times T_{\rm opt} \times T_{\rm read} \times T_{\rm QE}$, where $T_{\rm Tele}$ accounts for light lost due to contamination and inefficiencies in the main collecting area, $T_{\rm coronagraph}$ is the coronagraphic throughput at this planet–star separation, $T_{\rm opt}$ is the optical throughput (the transmissivity of all optics), $T_{\rm QE}$ is the raw quantum efficiency (QE) of the detector, and $T_{\rm read}$ is the read-out efficiencies. A summary and representative value for each of these parameters can be found in Table 1, and we show the coronagraph throughput as a function of separation in Figure 10.

For $T_{\rm Tele}$, we adopt 0.95 for all wavelengths, on par with the particulate coverage fraction for JWST's mirrors. EMCCD detectors are expected to have $T_{\rm read}$ near 0.75 (Stark et al. 2019), while for NIR and other detectors, read-out inefficiencies and bad pixels may account for a similar value, and we simply adopt $T_{\rm read} = 0.75$ across all detectors as a conservative estimate. The reported QE of the different detectors ranges from 0.6 to 0.9 (LUVOIR and HabEx reports; The LUVOIR Team et al. 2019; Gaudi et al. 2020), yet technological improvements in several of these detectors could be expected in the near future, and we adopt a general $T_{\rm QE} = 0.9$ for all detectors, bands, and for both observatories.

Optical efficiencies ($T_{\rm opt}$) for HabEx were taken from the HabEx report (Gaudi et al. 2020), specifically from Figures 6.4–10 for HabEx/SS and from Figures 6.3–6 for HabEx/no-SS, considering the IFS mode for the visible and infrared channels. For LUVOIR, the optical efficiencies were taken from the Final Report (The LUVOIR Team et al. 2019) and from Figure 4 (IFS mode) of Stark et al. (2019). These are shown in Figure 11.

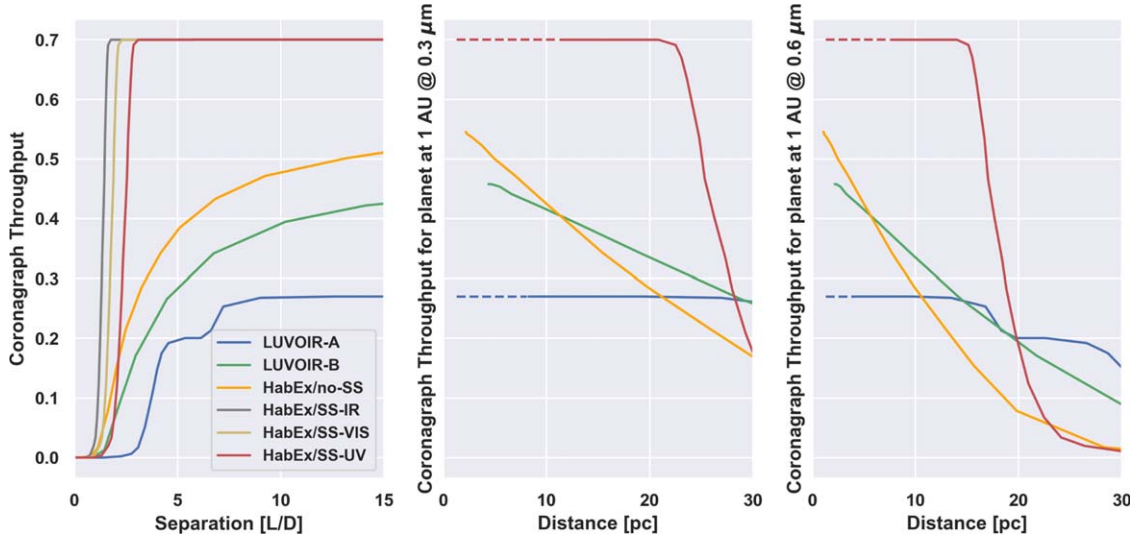

Figure 10. Coronagraph throughput of LUVOIR-A, LUVOIR-B, HabEx/SS, and HabEx/no-SS as a function of separation (left). Coronagraph throughput at 0.3 μm (center) and at 0.6 μm (right) as a function of distance. Following Stark et al. (2019), LUVOIR-A corresponds to the APLC coronagraph with a segmented on-axis OTA, LUVOIR-B corresponds to the DMVC coronagraph with a segmented on-axis OTA, and HabEx/no-SS corresponds to a vortex coronagraph with an off-axis monolithic OTA. The HabEx/SS throughput was taken from Figures 6.4–3 in the HabEx Final Report and integrates a 70% loss due to PSF losses.

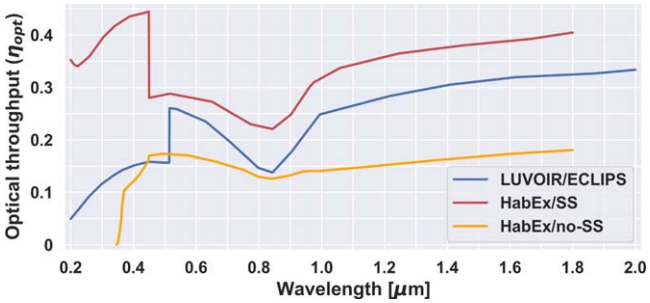

Figure 11. Optical throughput of LUVOIR, HabEx/SS, and HabEx/no-SS as a function of wavelength. Optical efficiencies for HabEx were taken from the final report (Gaudi et al. 2020), considering the IFS mode for the visible and infrared channels. For LUVOIR, the optical efficiencies were captured from the Final Report (The LUVOIR Team et al. 2019) and from Figure 4 (IFS mode) of Stark et al. (2019).

ORCID iDs

Jade H. Checlair https://orcid.org/0000-0001-8724-833X
Stephanie L. Olson https://orcid.org/0000-0002-3249-6739
Thaddeus D. Komacek https://orcid.org/0000-0002-9258-5311
Tyler D. Robinson https://orcid.org/0000-0002-3196-414X
Huanzhou Yang https://orcid.org/0000-0001-8693-7053
Dorian S. Abbot https://orcid.org/0000-0001-8335-6560

References

Bean, J. L., Abbot, D. S., & Kempton, E. M.-R. 2017, ApJL, 841, L24
Belikov, R., Stark, C., Angerhausen, D., et al. 2017, ExoPAG SAG-13 Final Report
Bixel, A., & Apai, D. 2020, ApJ, 896, 131
Burke, C. J., Christiansen, J. L., Mullally, F., et al. 2015, ApJ, 809, 8
Caldeira, K., & Kasting, J. F. 1992, Natur, 360, 721
Catanzarite, J., & Shao, M. 2011, ApJ, 738, 151
Checlair, J. H., Abbot, D. S., Webber, R. J., et al. 2019, arXiv:1903.05211
Cowan, N., Greene, T., Angerhausen, D., et al. 2015, PASP, 127, 311
Deming, D., Seager, S., Winn, J., et al. 2009, PASP, 121, 952
Des Marais, D. J., Harwit, M. O., Jucks, K. W., et al. 2002, AsBio, 2, 153
Dittmann, J. A., Irwin, J. M., Charbonneau, D., et al. 2017, Natur, 544, 333
Dodd, M. S., Papineau, D., Grenne, T., et al. 2017, Natur, 543, 60
Dressing, C. D., & Charbonneau, D. 2015, ApJ, 807, 45
Fally, S., Vandaele, A. C., Carleer, M., et al. 2000, JMoSp, 204, 10
Feng, Y. K., Robinson, T. D., Fortney, J. J., et al. 2018, AJ, 155, 200
Gaudi, B. S., Seager, S., Mennesson, B., et al. 2020, arXiv:2001.06683
Gelaro, R., McCarty, W., Suárez, M. J., et al. 2017, JCli, 30, 5419
Gillon, M., Triaud, A. H., Demory, B.-O., et al. 2017, Natur, 542, 456
Harman, C., Felton, R., Hu, R., et al. 2018, ApJ, 866, 56
Harman, C., Schwieterman, E., Schottelkotte, J. C., & Kasting, J. 2015, ApJ, 812, 137
Jenouvrier, A., Mérienne, M.-F., Coquart, B., et al. 1999, JMoSp, 198, 136
Kasting, J. F., Liu, S., & Donahue, T. 1979, JGR, 84, 3097
Kipping, D. 2020, PNAS, 117, 1
Kopparapu, R. K., Hébrard, E., Belikov, R., et al. 2018, ApJ, 856, 122
Kopparapu, R. K., Ramirez, R., Kasting, J. F., et al. 2013, ApJ, 765, 131
Kreidberg, L., & Loeb, A. 2016, ApJL, 832, L12
Krissansen-Totton, J., Olson, S., & Catling, D. C. 2018, SciA, 4, eaao5747
Lovelock, J. E. 1965, Natur, 207, 568
Lustig-Yaeger, J., Meadows, V. S., & Lincowski, A. P. 2019, AJ, 158, 27
Lyons, T. W., Reinhard, C. T., & Planavsky, N. J. 2014, Natur, 506, 307
Meadows, V. S. 2008, in Exoplanets, ed. J. W. Mason (Berlin: Springer), 259
Meadows, V. S. 2017, AsBio, 17, 1022
Meadows, V. S., Arney, G. N., Schwieterman, E. W., et al. 2018a, AsBio, 18, 133
Meadows, V. S., Reinhard, C. T., Arney, G. N., et al. 2018b, AsBio, 18, 630
Mérienne, M.-F., Jenouvrier, A., Coquart, B., et al. 2000, JMoSp, 202, 171
Mérienne, M.-F., Jenouvrier, A., Coquart, B., et al. 2001, JMoSp, 207, 120
Mulders, G. D., Pascucci, I., Apai, D., & Ciesla, F. J. 2018, AJ, 156, 24
Neil, A. R., & Rogers, L. A. 2020, ApJ, 891, 12
Olson, S. L., Schwieterman, E. W., Reinhard, C. T., et al. 2018a, ApJL, 858, L14
Olson, S. L., Schwieterman, E. W., Reinhard, C. T., & Lyons, T. W. 2018b, in Handbook of Exoplanets, ed. H. Deeg & J. Belmonte (Cham: Springer), 189
Owen, T. 1980, ASSL, 83, 177
Ozaki, K., Reinhard, C. T., & Tajika, E. 2019, Geobiology, 17, 3
Pascucci, I., Mulders, G. D., & Lopez, E. 2019, ApJL, 883, L15
Pavlov, A., & Kasting, J. 2002, AsBio, 2, 27
Petigura, E. A., Howard, A. W., & Marcy, G. W. 2013, PNAS, 110, 19273
Planavsky, N. J., Reinhard, C. T., Wang, X., et al. 2014, Sci, 346, 635
Reinhard, C. T., Olson, S. L., Schwieterman, E. W., & Lyons, T. W. 2017, AsBio, 17, 287
Robinson, T. D., Stapelfeldt, K. R., & Marley, M. S. 2016, PASP, 128, 025003
Rushby, A. J., Claire, M. W., Osborn, H., & Watson, A. J. 2013, AsBio, 13, 833
Sagan, C., Thompson, W. R., Carlson, R., Gurnett, D., & Hord, C. 1993, Natur, 365, 715
Sandberg, A., Drexler, E., & Ord, T. 2018, arXiv:1806.02404

Schidlowski, M. 1988, Natur, 333, 313
Schwieterman, E. W., Kiang, N. Y., Parenteau, M. N., et al. 2018, AsBio, 18, 663
Seager, S., Bains, W., & Petkowski, J. 2016, AsBio, 16, 465
Seager, S., & Deming, D. 2010, ARA&A, 48, 631
Seager, S., Turner, E. L., Schafer, J., & Ford, E. B. 2005, AsBio, 5, 372
Segura, A., Meadows, V., Kasting, J., Crisp, D., & Cohen, M. 2007, A&A, 472, 665
Sousa-Silva, C., Petkowski, J. J., & Seager, S. 2019, PCCP, 21, 18970
Spiegel, D. S., & Turner, E. L. 2012, PNAS, 109, 395
Stark, C. C., Belikov, R., Bolcar, M. R., et al. 2019, JATIS, 5, 024009
Stark, C. C., Roberge, A., Mandell, A., & Robinson, T. D. 2014, ApJ, 795, 122
The LUVOIR Team et al. 2019, arXiv:1912.06219
Tian, F., France, K., Linsky, J. L., Mauas, P. J., & Vieytes, M. C. 2014, E&PSL, 385, 22
Villanueva, G. L., Smith, M. D., Protopapa, S., Faggi, S., & Mandell, A. M. 2018, JQSRT, 217, 86
Wordsworth, R., & Pierrehumbert, R. 2014, ApJL, 785, L20