Variability due to climate and chemistry in observations of oxygenated Earth-analogue exoplanets

G. J. Cooke [1]★, D. R. Marsh,[1,2]★ C. Walsh,[1]★ S. Rugheimer[3,4] and G. L. Villanueva[5]

[1]*School of Physics and Astronomy, University of Leeds, Leeds LS2 9JT, UK*
[2]*National Center for Atmospheric Research, Boulder CO 80301, USA*
[3]*University Oxford, Atmospheric, Oceanic, and Planetary Physics Department, Clarendon Laboratory, Sherrington Road, Oxford OX1 3PU, UK*
[4]*Dept of Physics and Astronomy, York University, 4700 Keele Street, North York, Ontario 3MJ 1P3, Canada*
[5]*NASA Goddard Space Flight Center, Solar System Exploration Division, 8800 Greenbelt Road, Greenbelt, MD 20771, USA*

ABSTRACT

The Great Oxidation Event was a period during which Earth's atmospheric oxygen (O_2) concentrations increased from $\sim 10^{-5}$ times its present atmospheric level (PAL) to near modern levels, marking the start of the Proterozoic geological eon 2.4 billion years ago. Using WACCM6, an Earth System Model, we simulate the atmosphere of Earth-analogue exoplanets with O_2 mixing ratios between 0.1 and 150 per cent PAL. Using these simulations, we calculate the reflection spectra over multiple orbits using the Planetary Spectrum Generator. We highlight how observer angle, albedo, chemistry, and clouds affect the simulated observations. We show that inter-annual climate variations, as well short-term variations due to clouds, can be observed in our simulated atmospheres with a telescope concept such as LUVOIR or HabEx. Annual variability and seasonal variability can change the planet's reflected flux (including the reflected flux of key spectral features such as O_2 and H_2O) by up to factors of 5 and 20, respectively, for the same orbital phase. This variability is best observed with a high-throughput coronagraph. For example, HabEx (4 m) with a starshade performs up to a factor of two times better than a LUVOIR B (6 m) style telescope. The variability and signal-to-noise ratio of some spectral features depends non-linearly on atmospheric O_2 concentration. This is caused by temperature and chemical column depth variations, as well as generally increased liquid and ice cloud content for atmospheres with O_2 concentrations of <1 per cent PAL.

Key words: astrobiology – planets and satellites: atmospheres – planets and satellites: physical evolution – planets and satellites: terrestrial planets.

1 INTRODUCTION

During the current era of exoplanet discovery, the search for a temperate Earth-sized exoplanet orbiting around a G-type star continues. Data from telescopes and future observatories will add thousands more exoplanets to the list of over 5000 that have been discovered to date.[1,2] Upcoming transit and radial velocity surveys will make more precise measurements for the masses and radii of terrestrial exoplanets, thus enabling confident estimates of their bulk density and surface gravity. In the future, an Earth-sized exoplanet orbiting a G-type star at a semimajor axis of approximately 1 au may be found. This poses the question: will it be habitable?

Atmospheric information can be gained from measurements of emission, reflection, or transmission spectroscopy. Exoplanet atmospheres have been observed and characterized since the early 2000's (Charbonneau et al. 2002), with larger, Jupiter-sized planets being the least challenging to observe (Seager & Deming 2010). Detecting atmospheric properties and molecules for an Earth-sized exoplanet around a G-type star is much more challenging due to the comparatively weaker spectroscopic signal.

Any measurement of an exoplanet's atmosphere will be taken during a specific period of its geological evolution (Kaltenegger, Traub & Jucks 2007). The Earth itself has had a complicated geological and biological history which is still not wholly understood (Catling & Zahnle 2020; Cole et al. 2020; Steadman et al. 2020). At least three geological eons have harboured life on Earth: the Archean (4–2.4 Gyr ago), the Proterozoic (2.4–0.54 Gyr ago), and the Phanerozoic (0.54 Gyr ago – present; Lyons & Reinhard 2009; Dodd et al. 2017; Betts et al. 2018). The atmosphere during the Archean was anoxic, likely containing larger amounts of CO_2 and CH_4 compared to present-day (see Table 1; Haqq-Misra et al. 2008; Charnay et al. 2020). The emergence of cyanobacteria, in addition to hydrogen escape throughout the Archean (Catling, Zahnle & McKay 2001; Zahnle, Gacesa & Catling 2019), have been proposed to have led to the Great Oxidation Event (GOE) approximately 2.4 Gyr ago (Warke et al. 2020). The GOE enabled oxygen to increase to concentrations $\geq$0.1 per cent the present atmospheric level (PAL; Kasting & Siefert 2002), with possible Proterozoic eon

★ E-mail: pygjc@leeds.ac.uk (GJC); d.marsh@leeds.ac.uk (DRM); c.walsh1@leeds.ac.uk (CW)
[1]https://exoplanetarchive.ipac.caltech.edu/
[2]http://www.exoplanet.eu/

Table 1. A simplified chemical timeline of Earth is presented for O_2, CO_2, and CH_4, during the Archean, mid-Proterozoic, and Phanerozoic. Molecular abundances are constraints from years of research, including geological evidence and modelling studies. The 'inclusive' data from Olson et al. (2018a) is listed. Data for O_2 and CH_4 during the mid-Proterozoic is also from Large et al. (2019), Steadman et al. (2020), and Lyons et al. (2021). The mixing ratio for O_2 is given relative to its present atmospheric level (PAL) of 0.21. The mixing ratios for CO_2 and CH_4 are given in terms of parts per million by volume (ppmv). 1 Ga represents 1 billion years in the past.

Molecular abundance	Archean 4–2.4 Ga	mid-Proterozoic 1.8–0.8 Ga	Phanerozoic 0.54–0 Ga
O_2 (PAL)	$10^{-12}-10^{-5}$	$10^{-5}-1$	$10^{-1}-2$
CO_2 (ppmv)	$2500-40\,000$	$1400-28\,000$	$200-2800$
CH_4 (ppmv)	$100-35\,000$	$1-100$	$0.4-10$

concentrations lying between 1 PAL and 10^{-3} PAL (Steadman et al. 2020; Lyons et al. 2021). Phanerozoic O_2 concentrations are posited to have been between ∼0.1 PAL and ∼2 PAL (Olson et al. 2018a; Large et al. 2019). Table 1 presents a simplified overview of the composition of Earth's atmosphere for the past 4 billion years. These three geological eons represent important observational windows for Earth-analogue exoplanets (Rugheimer & Kaltenegger 2018) and the search for habitable and inhabited exoplanets.

Gases such as O_2, N_2O, and CH_4 are primarily produced by life on Earth. However, these gases can also be abiotically produced (Wordsworth & Pierrehumbert 2014; Kleinböhl et al. 2018) and the detection of such gases in an exoplanet atmosphere is not a conclusive detection of the presence of life (Selsis, Despois & Parisot 2002; Grenfell 2017). O_3 is the photochemical by-product of O_2 and therefore indicates the presence of O_2 without a detection of O_2 itself. Several modelling studies have shown that the amount of O_3 in an Earth-analogue atmosphere increases as the concentration of O_2 rises (Kasting & Donahue 1980; Segura et al. 2003; Way et al. 2017; Gregory, Claire & Rugheimer 2021; Cooke et al. 2022). Olson et al. (2018b) discussed how seasonal O_2 variations could give rise to seasonal changes in the reflectance spectra of O_3 in the ultraviolet (UV) region, noting however that separate hemispheres that experience opposite seasons mask the full seasonality signature.

Previous work has modelled modern Earth as an exoplanet (Kaltenegger & Traub 2009), as well as Earth at different geological time periods (Rugheimer et al. 2015; Arney et al. 2016; Rugheimer & Kaltenegger 2018; Kaltenegger, Lin & Madden 2020a; Kaltenegger, Lin & Rugheimer 2020b). Other studies investigated the photometric and spectroscopic variations of an Earth-analogue exoplanet (Ford, Seager & Turner 2001; Montañés-Rodríguez et al. 2006; Tinetti et al. 2006a, b). These investigations have used a variety of 1D and 3D models, and it is clear that numerous parameters, including geological eras with different atmospheric compositions, influence molecular detectability. The emission and transmission spectra of oxygenated exoplanets (varying through geological time) has been predicted with 1D models (e.g. Segura et al. 2003; Rugheimer & Kaltenegger 2018; Kaltenegger et al. 2020a). While most work in this area has focused on modelling, De Cock et al. (2022) performed a phase angle analysis of Earth using data from the EPOXI mission which observed Earth to test exoplanet observational predictions (Livengood et al. 2011).

Emission and reflectance spectra can be observed through high-contrast imaging (HCI). This does not require the planet to transit the star, although the observer's inclination to the plane of the orbit will affect the observations because the brightness amplitude is reduced as the orbital inclination decreases from 'edge-on' to 'face-on' (Maurin et al. 2012). As the planet orbits, it reflects light, particularly in the ultraviolet and visible wavelength regions. The Earth's emission peaks in the infrared, where the emission (both in terms of its strength and wavelength variance) depends on temperature, clouds, and chemistry (Parmentier & Crossfield 2018). The brightness of the host star (G-type) far outshines the brightness of an Earth-analogue exoplanet. Coronagraphs can be used to reduce the incoming light from the star, with a contrast of ∼10^{-10} required to reliably characterize an Earth-Sun analogue system (Lee 2018; Currie et al. 2022); hence, we explore the use of a coronograph in this work.

Recently, Checlair et al. (2021) evaluated the detectability of O_3 and O_2 on Earth-analogue (defined as possessing an oxygenated atmosphere due to life) exoplanets with four different telescope concepts: LUVOIR A (15 m diameter), LUVOIR B (8 m diameter), HabEx (4 m diameter) with and without a starshade. They simulated observations with these missions for O_2 concentrations of 10^{-5} PAL and upwards to 1 PAL, finding that larger diameter telescopes will better constrain the frequency of Earth-analogue worlds, noting that this depends on the occurrence rate of exo-Earth candidates.

In this work, we present thousands of synthetic high-contrast imaging spectra based on simulations of Earth-analogue exoplanet atmospheres over various atmospheric oxygen concentrations. This enables us to predict the variability in future observations, both within spectral features of interest, and in the broad-band. These atmospheric simulations utilize a 3D Earth System Model with fully coupled chemistry and physics – the Whole Atmosphere Community Climate Model 6 (WACCM6). Using the Planetary Spectrum Generator (PSG), we investigate the impact that changing oxygen concentrations has on simulated observations, considering climate, chemistry, temporal, annual, and seasonal variations. We determine the observability of such variations with next-generation telescope concepts.

2 METHODS

2.1 WACCM6

WACCM6 is an Earth System Model configuration of the Community Earth System Model (CESM) that couples together atmosphere, land, land-ice, ocean, and sea-ice sub models (Gettelman et al. 2019). The model release used in this work was the Coupled Earth System Model version 2.1.3 (CESM2.1.3[3]). The configuration that we used (BWma1850 'compset'[4]) has 98 chemical species, 208 chemical reactions, and 90 photolysis reactions. The model resolution was $1.875°$ in latitude by $2.5°$ in longitude.

We simulated five different scenarios that are presented in Table 2. The scenarios were a pre-industrial atmosphere simulation (hereafter PI), and simulations with atmospheres that have O_2 concentrations of 150 per cent PAL, 10 per cent PAL, 1 per cent PAL, and 0.1 per cent PAL. Each simulation had a 24 h rotation rate and the modern Earth's land and ocean configuration. Each simulation was stopped when atmospheric trends in temperature and chemistry were determined to have fluctuations that are expected from a climate model (e.g. year-to-year and diurnal cycle variations) for the last four model years,

---

[3] http://www.cesm.ucar.edu/models/cesm2/

[4] A list of compsets can be found here:
https://www.cesm.ucar.edu/models/cesm2/config/2.1.3/compsets.html

Table 2. Annual mean modelled temperatures and ozone columns under five different O_2 mixing ratio (f_{O_2}) scenarios. f_{O_2} is given in terms of present atmospheric level (PAL = 21 per cent by volume of the atmosphere). Listed are the global mean surface temperature, $\overline{T}_S$, the global mean tropopause temperature ($\overline{T}_{Tp}$), and the minimum ($T_{Tp,\,min}$) and maximum values ($T_{Tp,\,max}$) of the global tropopause temperature. Also listed is the global mean ozone column, $\overline{C}_{O_3}$, as well as the maximum $C_{O_3,max}$, and minimum, $C_{O_3,min}$, ozone columns in Dobson Units (DU), where 1 DU $= 2.687 \times 10^{20}$ molecules m^{-2}. All temperatures and O_3 columns are time-averaged over the final 4 yr of each simulation.

Simulation name	f_{O_2} (PAL)	$\overline{T}_S$ (K)	$\overline{T}_{Tp}$ (K)	$T_{Tp,\,min}$ (K)	$T_{Tp,\,max}$ (K)	$\overline{C}_{O_3}$ (DU)	$C_{O_3,min}$ (DU)	$C_{O_3,max}$ (DU)
150 per cent PAL	1.500	288.5	205.0	193.6	216.3	313	261	426
PI	1.000	288.5	204.4	193.3	216.0	279	236	375
10 per cent PAL	0.100	288.9	204.9	194.0	215.7	169	130	235
1 per cent PAL	0.010	288.5	202.4	191.6	213.5	66	41	83
0.1 per cent PAL	0.001	286.0	196.6	184.4	210.0	18	13	23

rather than ongoing trends (e.g. a decrease in surface temperature year on year).

The mixing ratios of the following species were held constant at the surface: O_2 (varied as in Table 2), CH_4 (0.8 ppmv), CO_2 (280 ppmv), N_2O (270 ppbv), and H_2 (500 ppbv). The fixed mixing ratio of other minor constituents (mixing ratio <1 ppbv) are listed on the Exo-CESM GitHub[5] page. The chemistry of all other species was left to evolve according to the chemical mechanism. All simulated atmospheres have a surface pressure of 1000 hPa. In each simulation in which the O_2 concentration was decreased, the N_2 concentration was increased to maintain a 1000 hPa surface pressure. Each simulation used the modern-day Sun as the solar input spectrum. This solar spectrum's irradiance was specified between 10.5 and 99975.0 nm.

Model data were output as mean average values (the sum of the instantaneous output at each time-step divided by the number of time steps considered) every 5 d, and mean average values every month. For the mean vertical profiles computed for Fig. 1, time and longitudinal means were taken, and latitudinal weights were averaged over (grid boxes near the poles have a smaller area than grid boxes near the equator), leaving just the vertical level dimension. Additionally, the instantaneous model state was output every 5 d (a 'snapshot' of the model). The snapshots were used for estimating reflection/emission spectra at a specific point in time. Time means presented in Fig. 1 and Table 2 were averaged over the last 4 yr of each simulation.

2.2 Planetary spectrum generator

The PSG (Villanueva et al. 2018) was used to compute theoretical relection/emission spectra for the WACCM6 atmospheric simulations. We used the GlobES 3D mapping tool[6] to upload 3D data from our simulated atmospheres. To reduce the memory footprint of the model input to be compatible with GlobES in PSG, the data were rebinned to a resolution of 10° in longitude only (with the latitudinal grid resolution unchanged at a value of 1.875°).

All of the synthetic spectra were calculated assuming an Earth-analogue exoplanet around a G2V star with a 1-yr orbit at a 90° inclination angle. Edge on systems are also likely to be best characterized in terms of mass and radius. All synthetic spectra were set to an exposure time of 60 s, with 1440 exposures, adding up to a total integration time of 24 h. Although the noise estimates do not simply scale with distance because of reduced coronagraphic

throughput at larger distances, to good approximation, the noise estimates do scale with total integration time, T_{int} ($\sigma_{noise} \propto T_{int}^{0.5}$). We show spectra for an integration time of 48 h, where we assume that the noise is reduced by a factor of $\sqrt{2}$ from the 24 h PSG results. Synthetic spectra from HCI observations were computed with WACCM6 snapshots every 5 d at 00:00 UTC, from January 1 until December 27. Consequently, the illuminated side of the planet is centred at 180° longitude, where it is predominantly ocean (i.e. it is mostly the Pacific ocean that is receiving sunlight at these times). Each snapshot was ingested into PSG by updating the orbital parameters for Earth (using the ephemeris data from the year 2020) every 5 d. This was then matched with the longitude and latitude of the Earth grid, to simulate the orbit in PSG to coincide with the orbit of the WACCM6 simulations.

The molecules used for the computation of reflection/emission spectra were N_2, O_2, CO_2, H_2O, O_3, CH_4, and N_2O. Each molecule used the default HITRAN opacity data (Gordon et al. 2017). All available collision induced absorption coefficients were utilized. The effects of scattering were included, as well as that of ice clouds and water clouds (Villanueva et al. 2011). The effective radius of the cloud particles was assumed to be 5 μm for water (stratocumulus) clouds and 100 μm for ice (cirrus) clouds, as is typical in the modern Earth's atmosphere. We also ran the same set of simulations with smaller diameter cloud particles, assuming 1 μm for water clouds and 1 μm for ice clouds – equivalent figures to Figs 3–7 are shown in the Supplementary Information.

Telescope templates are used for the Large Ultraviolet Optical Infrared Surveyor (LUVOIR; The LUVOIR Team 2019) concept and the Habitable Exoplanet Observatory (HabEx; Gaudi et al. 2020) concept. For LUVOIR, we used both the A (15 m primary mirror) and B (8 m primary mirror) templates. We changed the B template to a 6 m primary mirror based on the recent suggestions from the Decadal Survey (National Academies of Sciences, Engineering, and Medicine 2021). The LUVOIR A and B concepts have spectral resolutions of $R = 7$ (0.2–0.515 μm), $R = 140$ (0.515–1.0 μm), and $R = 70$ (1.01–2.0 μm) for the UV, VIS, and NIR channels, respectively (where $R = \lambda/\Delta\lambda$, λ is the wavelength, and $\Delta\lambda$ is the wavelength bin width). The HabEx concept had the same spectral resolutions in the UV (0.2–0.45 μm and VIS channels (0.45–0.975 μm) but instead with $R = 40$ in the NIR channel (0.975–1.8 μm).

Idealized spectra were computed with PSG (Fig. 2). However, to deduce whether atmospheric constituents in any of the WACCM6 simulations are physically observable, we also estimated the atmospheric signal-to-noise ratio (SNR) on the produced spectra. PSG calculated the noise, σ_{noise}, on an observation set. σ_{noise} included contributions from the source (the star and planet), from
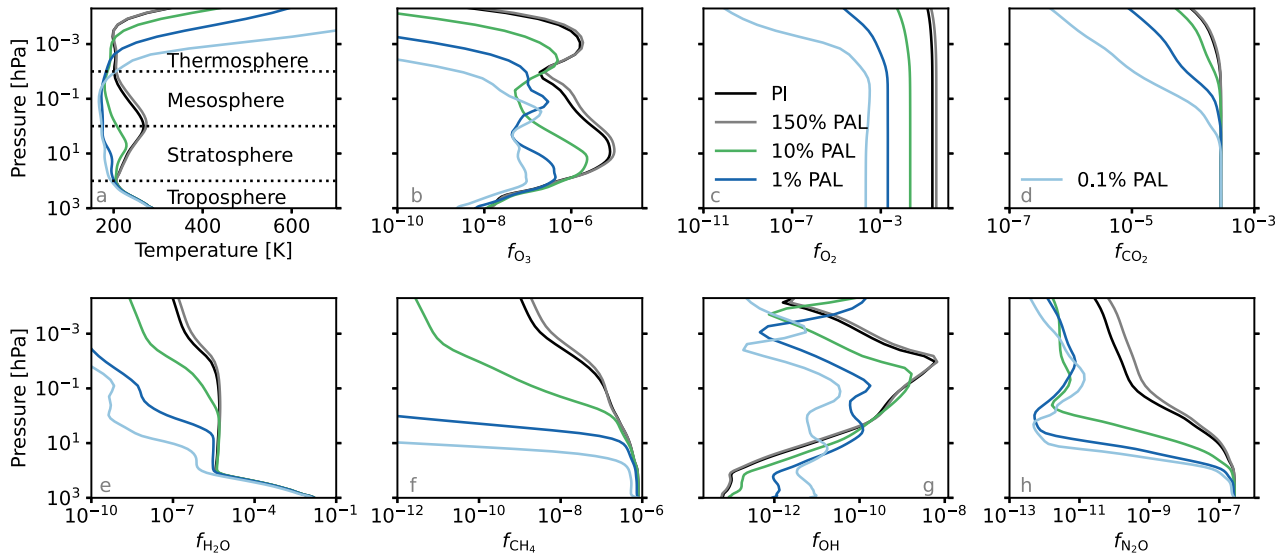
Figure 1. (a) The vertical profile of global mean temperatures for the following simulations: 150 per cent PAL (grey), PI (black), 10 per cent PAL (green), 1 per cent PAL (dark blue), and 0.1 per cent PAL (light blue) O_2 atmospheres. Black dashed lines indicate the transitions between the major temperature layers in the pre-industrial atmosphere. Mixing ratio vertical profiles, denoted as f_X, where X is the molecule shown, are plotted for O_3 (b), O_2 (c), CO_2 (d), H_2O (e), CH_4 (f), OH (g), and N_2O (h).

the background sky, and from the telescope itself (read noise and dark noise).

The SNR of a molecular detection also depends on a 'baseline' atmosphere that does not contain that molecule. This provides a continuum with which to estimate the absorption strength of the spectral features. We use a 1000 hPa N_2 atmosphere as our baseline which was created from each simulation by removing all other atmospheric molecules, but keeping the presence of liquid and ice clouds in the atmosphere. SNR were calculated at a specific wavelength with no rebinning performed on the simulated spectra.

The PYTHON code to produce configuration files to upload to PSG is included with the data associated with this manuscript. We note here that all simulations with PSG computed for this work are forward models. More details on the simulations completed are given in Appendix A.

3 RESULTS

3.1 WACCM6 simulations

Here, we summarize the globally averaged and time-averaged atmospheric chemistry of the simulations that were fed into PSG. See Cooke et al. (2022) for a more detailed discussion.

Fig. 1 shows the atmospheric mixing ratio vertical profiles from all five WACCM6 simulations for O_3, O_2, CO_2, H_2O, CH_4, OH, and N_2O. Also shown are the vertical profiles for temperature. When O_2 is reduced, there is less O produced from the photodissociation of O_2, such that fewer O_3 molecules can be produced from the reaction

$$O + O_2 + M \longrightarrow O_3 + M, \qquad (1)$$

where M is any third body.

When the O_2 concentration decreases (from 150 per cent PAL down to 0.1 per cent PAL), the O_3 column also decreases. While this effect has been reported in previous work, our WACCM6 simulations result in O_3 columns that are reduced by up to a factor of ∼5 compared to previous work for O_2 concentrations between 0.5 per cent and

50 per cent PAL (see Cooke et al. 2022). With less O_3 to absorb UV radiation, at ≤ 1 per cent PAL, there is reduced heating in the lower stratosphere and this cools the tropopause. Hence, more water is frozen out in the tropopause 'cold-trap' (Fueglistaler et al. 2009), generally resulting in more liquid and ice clouds.

Increased photolysis rates for CH_4, N_2O, and CO_2 result in lower mixing ratios. The mixing ratios for N_2O and CO_2 are reduced by a factor of 100 or more by the mesopause, and that for CH_4 is reduced by several orders of magnitude above the tropopause. Furthermore, as O_2 reduces, H_2O is increasingly photolysed above the tropopause and has mixing ratios reduced below 0.1 ppmv (parts per million by volume) in the stratosphere for the 1 per cent PAL and 0.1 per cent PAL cases, or in the upper mesosphere for the 10 per cent PAL case. This is in contrast to the PI atmosphere that retains at least 0.1 ppmv of H_2O in the lower thermosphere. Therefore, the five major greenhouse gases in Earth's atmosphere are all reduced above the tropopause when molecular oxygen is reduced. In addition to reduced heating from UV absorption by O_3, this results in cooling of the troposphere in the simulations with the lowest O_2 concentrations, resulting in a reduced water vapour column.

3.2 PSG reflection and emission spectra output

Fig. 2 shows reflection and emission spectra for the PI, 150 per cent PAL, 10 per cent PAL, 1 per cent PAL, and 0.1 per cent PAL atmospheres. These spectra are simulated on March 22, which is at $359.5°$ phase in our geometrical set-up, just before the exoplanet goes into secondary eclipse at $0°$ phase. In reality, this idealized spectrum will not be observable because of the small angular separation at this phase that results in low coronagraph throughput.

The emission from several spectral features are modified when oxygen is reduced. For example, the peak at 15 μm for CO_2, within an absorption trough, indicates the presence of a hot stratosphere, as has previously been reported (Selsis 2000; Kaltenegger et al. 2007; Rugheimer & Kaltenegger 2018). Once O_2 has reduced to ≤1 per cent PAL in our simulations, the stratosphere no longer
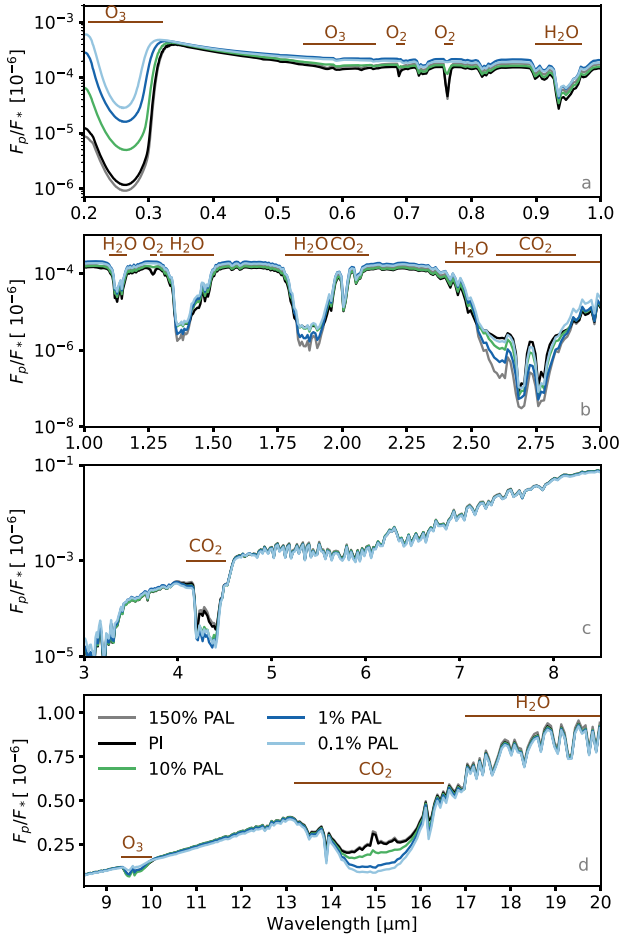
Figure 2. Idealized reflection/emission spectra in which a coronagraph is not used, given in terms of planet-to-star flux ratio (F_p/F_* [10^{-6}]), for the 150 per cent PAL (grey), PI (black), 10 per cent PAL (green), 1 per cent PAL (dark blue), and 0.1 per cent PAL (light blue) atmospheres between 0.1–1 μm (a), 0.1–3 μm (b), 3–8.5 μm (c), and 8.5–20 μm (d). Key spectral features are indicated. The spectra are computed at a phase of 395.5° for climate simulations at midnight on March 22. Note that each subplot has a different *y*-axis range and scaling. The spectra are binned to a spectral resolution of $R = 250$, where $R = \lambda/\Delta\lambda$, λ is the wavelength of light, and $\Delta\lambda$ is the wavelength bin width.

exists, so this feature is seen completely in absorption. O_2 features seen in absorption are at 0.69, 0.76, and 1.27 μm. An increase in oxygen increases the depth of these features relative to the continuum. There is a noticeable difference in magnitude between the PI and 150 per cent PAL cases; the 150 per cent PAL O_2 features are 13 per cent, 9 per cent, and 28 per cent deeper relative to the continuum at $R = 250$, respectively. When O_2 is reduced, there is less stratospheric O_3 resulting in a lower total O_3 column; hence, there is reduced atmospheric absorption by O_3. This is most noticeable at UV wavelengths, where the depth of the feature lies between $\approx 9 \times 10^{-7}$ and $\approx 3 \times 10^{-5}$ ppm. The depth of H_2O features mostly result from the climate at the chosen time stamp and orbital phase, although the time-averaged water column is reduced up to 20 per cent at 0.1 per cent PAL because of lower tropospheric temperatures.

The spectra in Fig. 2 are simulated at a single point in time. However, the spectra will change throughout the orbit due to seasonal, climatic, and chemical changes in the atmosphere. The presence and distribution of clouds also affects reflection and emission spectra.

Clouds can either mask molecular features, or boost them through reflectivity (Rugheimer & Kaltenegger 2018). We now simulate the spectra of these exoplanets as seen through HCI by sampling the climate system along the orbit. We then use these simulations to estimate the variations expected from future observations of oxygenated Earth-analogue exoplanets.

3.3 Spectra from high-contrast imaging

We present spectra generated using HCI for the PI and 0.1 per cent PAL atmosphere with the LUVOIR A telescope concept for an exoplanet at 10 pc, with 24 or 48 h of integration time (annual variability curves for the 10 per cent PAL and 1 per cent PAL simulations are included in the Supplementary Information). We recognize that this is an optimistic distance at which an Earth-analogue exoplanet could be found, so we also include calculations at 25 and 50 pc. Previous work has used 10 pc for a variety of rocky exoplanet predictions (e.g. Fujii et al. 2011; Checlair et al. 2021; Alei et al. 2022) and so this distance also allows meaningful comparisons with those works.

The PSG radiation output is in terms of planetary flux (F_p), stellar flux (F_*), and the total flux received by the detector ($F_T = F_p + F_*$). Note that F_p cannot be empirically detected in such coronagraphic observations, where some stellar light still reaches the detector. The error bars on F_T depend on several sources of noise, as described in Section 2.

3.3.1 Maximum SNR

We quantify the maximum SNR possible with the three telescope concepts in Fig. 3, at 10, 25, and 50 pc, after a total integration time of 24 h. Occurrence rate estimates for Earth-analogue planets around G-type stars (usually denoted as $\eta_\oplus$) in the literature have various definitions (Catanzarite & Shao 2011; Traub 2012; Fressin et al. 2013; Zink & Hansen 2019; Bryson et al. 2021), and the calculated rates typically vary between 0.01 and 2 per star (Traub 2012; Petigura, Howard & Marcy 2013; Burke et al. 2015; Bryson et al. 2021), with 2 being the optimistic end of estimates. According to the SIMBAD online data base (Wenger et al. 2000),[7] there are 13 G0V-G9V stars closer than 10 parsecs and 2830 G0V–G9V stars closer than 75 pc. This implies that an Earth-analogue exoplanet with a random orbital inclination should be found somewhere between 10 and 75 pc. This is a large range and should be kept in mind when considering the results from Fig. 3.

Assuming the threshold for detection of a molecule is an SNR of ≥ 5, then O_3 in the UV can be detected for all atmospheres and telescopes for an exoplanet at 10 pc. O_2 can be detected in the PI atmosphere by LUVOIR A at both 0.69 and 0.76 μm, although for the 0.1 per cent PAL atmosphere this would require an integration time of ≈ 50 and ≈ 1.2 d for the 0.69 and 0.76 μm features, respectively. H_2O can be detected at 0.94 μm with LUVOIR A but not at 1.4 μm, and the LUVOIR B and HabEx SS cannot detect it at 10 pc with a 24 h integration time. Rebinning of the two assessed H_2O features would improve their SNRs. At a distance of 25 pc, only LUVOIR A can detect any molecules (O_3). No molecules can be detected by any telescope in any atmosphere for an exoplanet at 50 pc, but O_3 in the PI and 10 per cent PAL atmospheres may be detectable for $T_{int} \approx 3.4$ d with LUVOIR A. LUVOIR A (15 m) is the best performing telescope overall, and HabEx with a starshade (4 m) performs better

---

[7] http://simbad.u-strasbg.fr/simbad/
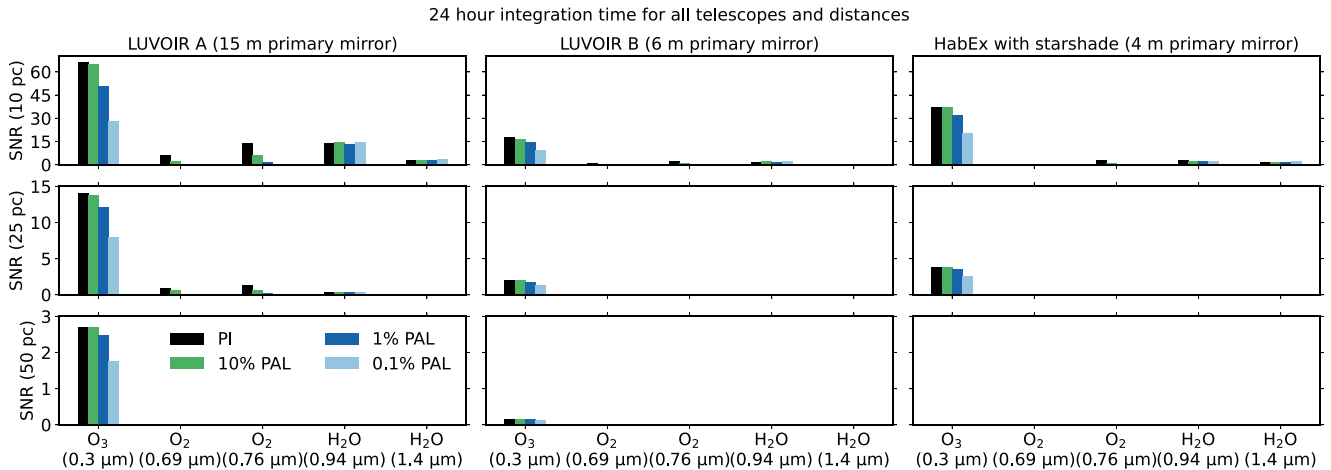
Figure 3. The signal-to-noise ratio for five molecular features are given for the PI (black), 10 per cent PAL (green), 1 per cent PAL (dark blue), and 0.1 per cent PAL (light blue) atmospheres, at a distance of 10 pc (*top row*), 25 pc (*middle row*), and 50 pc (*bottom row*). The LUVOIR A (*left column*), LUVOIR B (*middle column*), and HabEx with a starshade (*right column*) telescope concepts are evaluated, each observing for a total integration time of 24 h. Only the first year of the final 4 yr of each simulation is evaluated in this figure.

than LUVOIR B (6 m) despite the larger diameter telescope size of LUVOIR B. For example, the SNR for O_3 at 10 pc is up to a factor of 2.2 higher for HabEx with a starshade when compared to LUVOIR B.

Similar to Checlair et al. (2021), where a cloud-free atmosphere was assumed in their calculations, we also find that the detection of O_2 and O_3 becomes increasingly difficult as the O_2 concentration reduces. Because a different O_2 concentration affects atmospheric chemistry and photolysis rates, the energy budget for the atmosphere changes. This results in different tropospheric temperatures that are positively correlated with the total water column (see the Supplementary Information for how the water column varies with orbital phase) and thus influences the reflection and emission spectra by modulating absorption and emission. Therefore, we find that the SNR for detecting H_2O is also influenced by the O_2 concentration, although the change between atmospheres is small. The effect will be more significant for transmission spectra where the middle atmospheric concentration of H_2O reduces with reduced O_2.

Additionally, decreasing the diameter of cloud particle sizes from 5 and 100 μm to 1 and 1 μm (for water and ice clouds, respectively) increases the maximum SNR available during an orbit due to an increased amount of photons reaching the telescope detector. To give an example, the maximum SNR at 24 h integration time with LUVOIR A for O_3 at 0.3 μm in the PI atmosphere is 66 with the larger diameter particles, versus an SNR of 100 with the smaller diameter cloud particles.

3.3.2 Annual variations with LUVOIR A at 10 pc

In Fig. 4, we show how the same atmospheres exhibit interannual variability by modelling the spectra from HCI of four consecutive model years using the LUVOIR A telescope concept. This is the most optimistic future telescope concept we have assessed as discussed in Section 3.3.1. We choose the final 4 yr for each simulation as these data are how we have assessed the stability of the climate (see Section 2.1). Fig. 4 shows the measured total flux (F_T) as a function of orbital phase, where January 1 is at a phase of 279°.

For LUVOIR A observing an oxygenated Earth analogue exoplanet at 10 pc, the inter-annual variability for the broadband and individual spectral features year-on-year is larger than the noise in the VIS channel at multiple phases. Specific lines within the UV do not show observable year-to-year variations, although the broadband in this channel does. Because the inner working angle (IWA) of the coronagraph is proportional to wavelength, longer wavelengths are increasingly harder to observe due to reduced coronagraphic throughput – see Appendix B. This means that longer wavelengths, especially in the NIR channel, cannot be differentiated year on year. Additionally, certain phases cannot be differentiated, and this is either because the magnitude of the year-on-year variations are within the noise, or the coronagraph throughput is too low to detect any meaningful planetary flux variation.

The H_2O and O_2 features have larger variations compared to O_3 features. Variations for H_2O and O_2 at some phases are distinguishable from the noise for some years, but not necessarily all of the years. Therefore, multiple years ($\geq$3) of observations may be needed to witness variations at particular phases. Greater integration times will increase the amount of phases where inter-annual variation is observable. Almost all phases can be differentiated in broad-band observations.

In the UV channel, the same wavelength at the same phase between years can vary in brightness by up to 1.3 times and 2.4 times for F_T and F_p, respectively. In the VIS and NIR channels, these numbers can reach up to 1.5 and 2.5, and 1.7 and 5, respectively.

Fig. 5 shows the year 1–year 2 percentage differences in the total flux versus orbital phase for the 0.76 μm O_2 and 0.94 μm H_2O features. In addition, we swapped either clouds, composition, or surface albedo between the years to isolate the contribution of these planetary properties to the total year-to-year difference. The largest percentage difference for H_2O between these 2 yr is +22.4 per cent, occurring at an orbital phase of 316°, where the percentage difference from only swapping clouds is +21.9 per cent. On the other hand, a swap in chemistry never contributes a change of more than $\pm$4 per cent, and surface albedo never more than $\pm$1 per cent. We can summarize this by stating that in our simulations, the inter-annual variation is primarily influenced by clouds, secondarily by chemistry (composition), and surface albedo plays only a minor role. This trend is similar across oxygenation states and wavelengths.
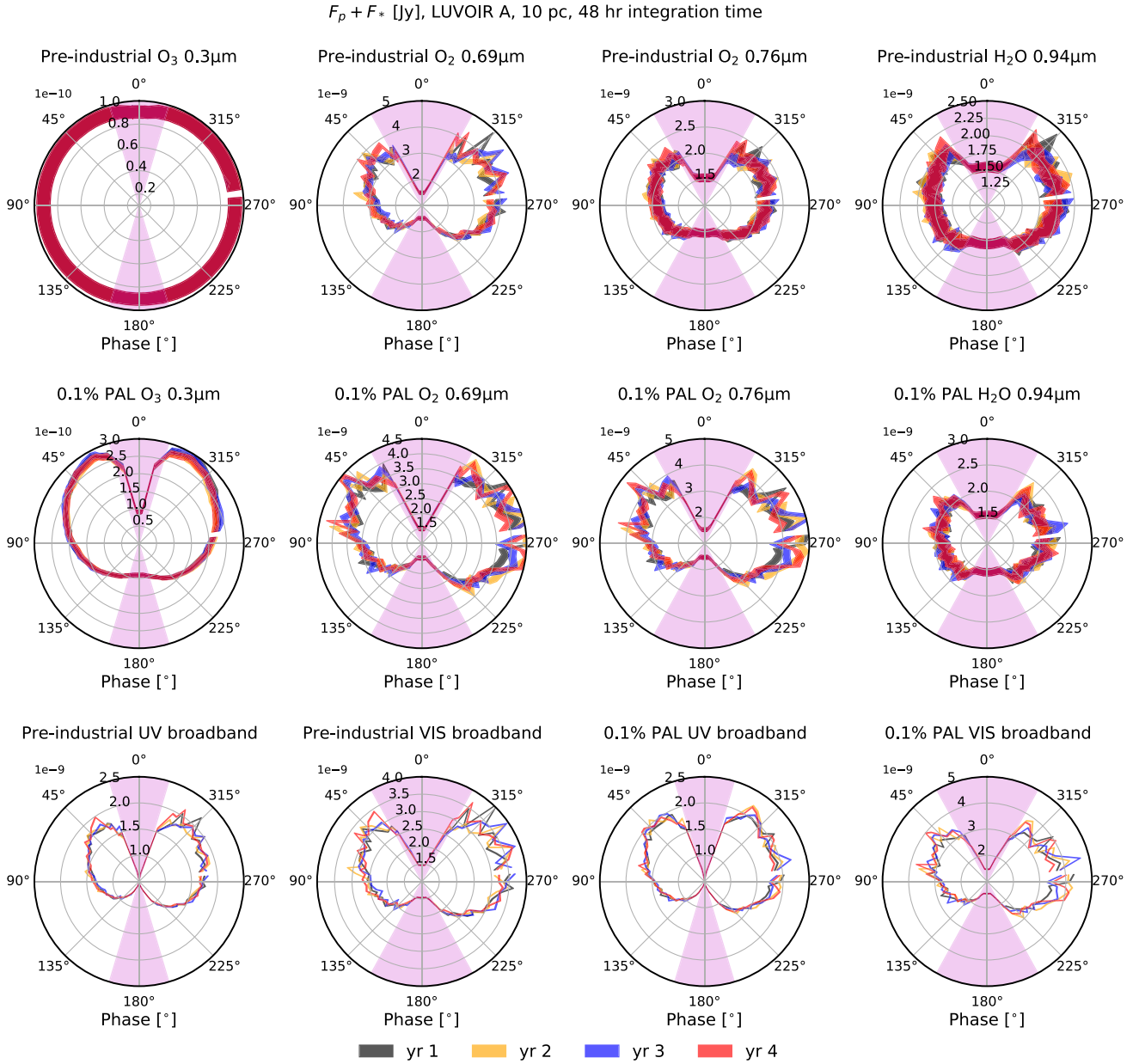
Figure 4. The total flux measured as a function of orbital phase for four consecutive years is plotted to demonstrate year-on-year variability. Plotted in the radial direction in units of Jy is the total flux $F_T = F_p + F_*$ from the planet and star as seen by the detector from the final 4 yr (1st year = black, 2nd year = yellow, 3rd year = blue, 4th year = red) of the PI and 0.1 per cent PAL simulated atmospheres. Note that the total flux never goes to zero because there is always some flux from the star. The total integration time is 48 h using LUVOIR A for an exoplanet at 10 pc. The discontinuity in the curves is between December 27 and January 1. *Top and Middle*: The top and middle rows show the variation with phase (in the azimuthal direction) at the 0.3 μm O_3 feature, at the 0.69 μm O_2 feature, at the 0.76 μm O_2 feature, and at the 0.94 μm H_2O feature for the PI and 0.1 per cent PAL atmospheres. The width of the lines indicate the noise for that particular measurement. *Bottom*: The bottom row shows the broad-band variation in the PI and 0.1 per cent PAL atmospheres for the UV and VIS channels. The broad-band noise is indicated by the width of the lines. The magenta shaded regions represent phases where the exoplanet would be inside the inner working angle of the LUVOIR A coronagraph, such that the coronagraph throughput significantly drops.

Total cloud coverage and cloud water path changes often for each atmosphere, such that clouds change the overall albedo of the planet. The O_2 column does not significantly vary throughout an orbit, whereas the H_2O column does due to atmospheric temperature fluctuations and resultant changes in water phase. This is why a change in composition influences the spectra more for H_2O features than for O_2 features. For a specific date and corresponding phase, the surface albedo is almost constant. In the cases

year-to-year, the surface albedo is almost constant. In the cases considered here, it has little effect on observed annual variations. The surface albedo does affect the seasonal variations, however, as we will show in the next section.

In Fig. 6, we show how each atmosphere varies over four consecutive years at quadrature (maximum planet–star separation, where the orbital phase is either 90° or 270°). The mean is plotted
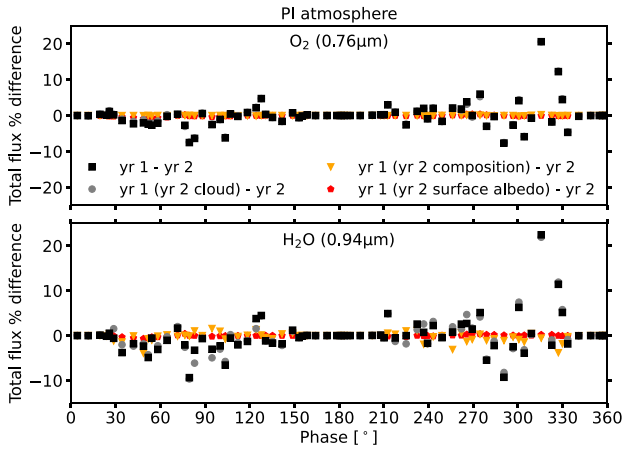
Figure 5. The total flux difference (in terms of percentage difference) versus orbital phase is plotted for the first and second year of WACCM6 data from the PI atmosphere for the 0.76 μm O_2 feature (*top*) and for the 0.94 μm H_2O feature (*bottom*). Black squares indicate the difference between the standard output from the atmosphere with no changes. Grey circles, orange triangles, and red pentagons indicate the flux difference when clouds, composition, and surface albedo, respectively, from the second year, are swapped into the first year atmosphere before being processed by PSG. Although these swaps do not represent a physical atmospheric prediction, they enable quantification of what is affecting the atmospheric variability between years.

negative percentage difference is -18, -23, -23, and -19 per cent for the PI, 10 per cent PAL, 1 per cent PAL, and 0.1 per cent PAL atmosphere, respectively.

We also show the total range between all phases and wavelengths (shaded background in Fig. 6). Although the SNR for O_3 is high in the UV, O_3 column variations due to climate may not be detectable because of the very small range at 0.3 μm. The total flux range varies with wavelength, and the largest total flux range occurs in the 0.1 per cent PAL case. The flux range of wavelengths that cover O_3 and O_2 features, including the continuum, generally increases with decreasing atmospheric O_2 concentration. This trend with O_2 concentration is not always linear for H_2O features. These results indicate that inter-annual variations in observations may be dependent on O_2 concentrations, but more than four orbits would be required to determine this as a statistically reliable result, rather than a consequence of chemistry-independent climate variability within the model.

3.3.3 Seasonal variations with LUVOIR A at 10 pc

Aside from interannual variations, observer geometry also affects the spectra as we will demonstrate in this section. We have varied the observer geometry simply by adding 90°, 180°, and 270° to the orbital phase to which each model date corresponds to. We do this for the synthetic spectra for two consecutive model years (year 2 and year 3 from the final 4 years of data) to demonstrate how seasonal variations affect exoplanet spectra in Fig. 7.

We find that seasonal variations are larger than inter-annual variations between the same phase, as would be expected for a planet like Earth where the seasons are set by Earth's obliquity, i.e. the climatic differences between July 1 and January 1 are more likely to

alongside the range. The largest positive percentage difference in flux from the mean flux at quadrature for each atmosphere is $+18$, $+21$, $+29$, and $+30$ per cent, for the PI, 10 per cent PAL, 1 per cent PAL, and 0.1 per cent PAL atmosphere, respectively. Likewise, The largest
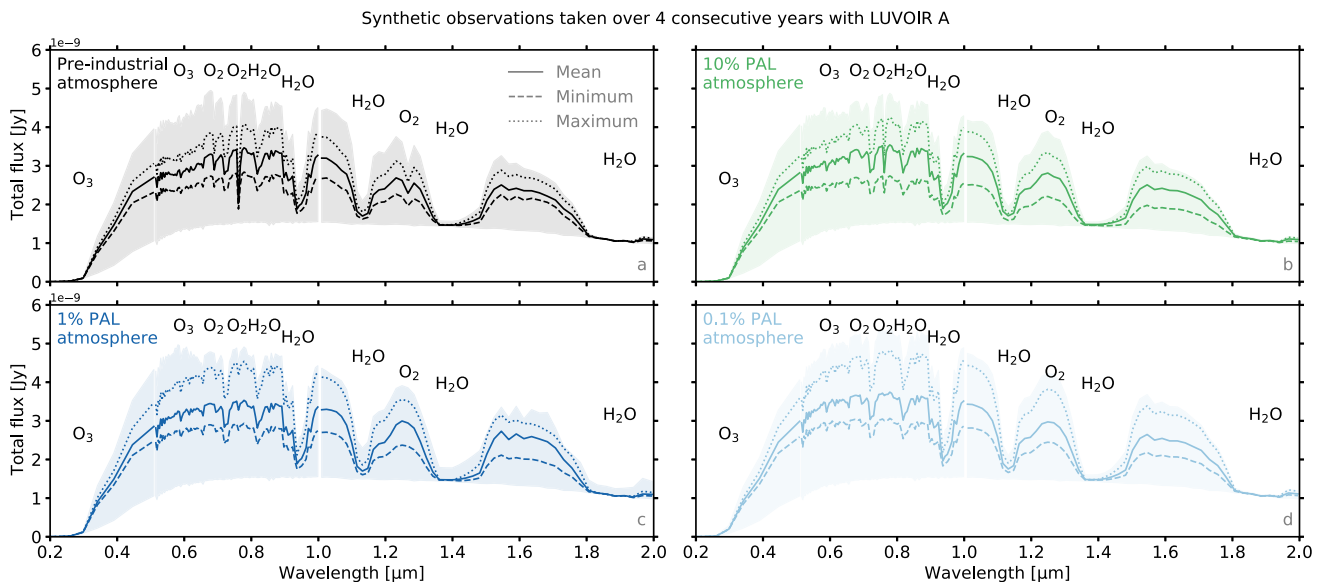

Figure 6. The mean (unbroken line), minimum (dashed line), and maximum (dotted line) for observations taken at maximum projected star-exoplanet separation (orbital phases closest to 90° and 270°) are plotted for LUVOIR A at 10 pc. The minimum and maximum observations at quadrature are determined to be the date at which the 0.88 μm line (often the wavelength with the highest flux) exhibits minimum and maximum flux for that particular year, respectively. The shaded regions show the range between the maximum and minimum total flux at each wavelength, now accounting for all phases. The results from the PI (a), 10 per cent PAL (b), 1 per cent PAL (c), and 0.1 per cent PAL (d) atmospheres are all plotted. Whilst maximum separation does not correspond to maximum planetary brightness, it does correspond to the most observable phase at different distances. Closer star-exoplanet separations may have reduced coronagraphic throughput.

$F_p + F_*$ [Jy], LUVOIR A, 10 pc, 48 hr integration time
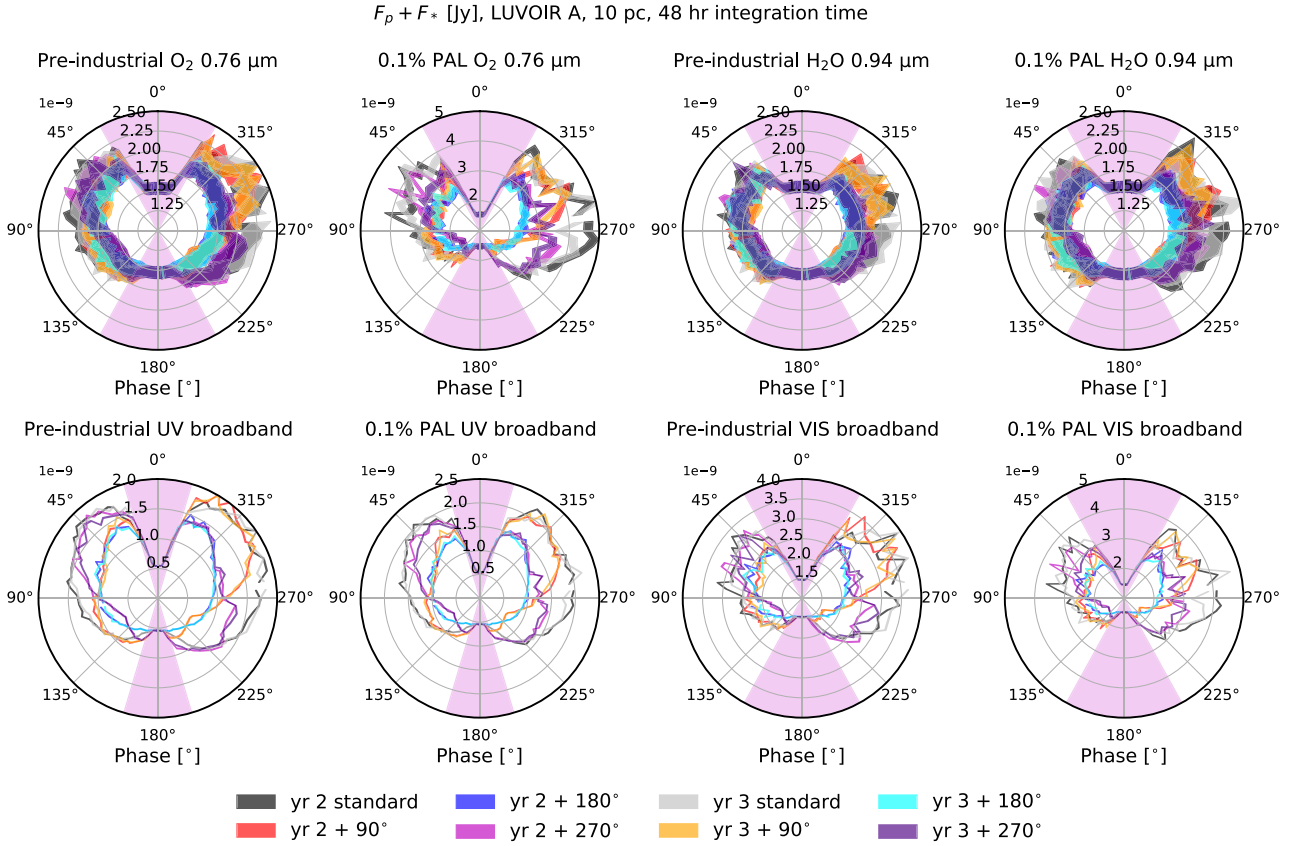
Figure 7. This figure demonstrates how seasons can impact exoplanet spectra. In units of Jy the total flux received by the telescope ($F_T = F_p + F_*$) is plotted in the radial direction against phase in the azimuthal direction. This is for a 48 h integration time using LUVOIR A for an exoplanet at 10 pc. A standard year is plotted, and then observer geometry is rotated $+90°$, $+180°$, and $+270°$. 'yr 2' and 'yr 3' refer to the second and third years of the final 4-yr data set, respectively. *Top*: Variation for the pre-industrial and 0.1 per cent PAL atmospheres are shown for O_2 at 0.76 μm and H_2O at 0.94 μm. The width of the lines represent the $\pm 1\sigma_{noise}$ uncertainty on the observations. *Bottom*: The broad-band variation is shown in the UV and VIS channels with phase. The discontinuity in the curves is between December 27 and January 1. The magenta shaded regions represent phases where the exoplanet would be inside the IWA of the LUVOIR A coronagraph. The width of the lines represent the $\pm 1\text{-}\sigma_{noise}$ broad-band uncertainty.

exhibit a greater variation in magnitude than two consecutive years on January 1. Because one would likely not know a priori which season is being observed, this can initially limit the observational information gained about a particular exoplanet's climate because of the potential for degenerate interpretations.

The reason why the phases between 180° and 360° in Fig. 4 are often (but not always) brighter than 0°–180° by up to ≈15 per cent is due to an increase in surface ice coverage. The H_2O column shows a seasonal cycle and this will affect the magnitude of the signal for H_2O features (e.g. more H_2O results in increased absorption). The influence from the cloud column is less obvious when comparing these two time periods because clouds do not vary with a seasonal cycle as clear as the ice seasonal cycle. 180°–360° roughly corresponds to September–March, which covers winter in the Northern hemisphere, when Arctic ice is most prevalent. The period between 0° and 180° roughly corresponds to March–September, with increased surface ice around Antarctica during this time, although a higher proportion of the Earth's surface in the Arctic is covered in ice during the Northern hemisphere's winter. More ice raises the albedo of Earth, such that the phases between 180°–360° are generally brighter.

By shifting the phases that the dates correspond to, the same wavelength at the same phase for shifted seasons can vary in

brightness by up to 2.8 and 10.5 times for F_T and F_p, respectively, in the UV channel. In the VIS and NIR channels, these numbers can reach up to approximately 2.3 and 17, and 2 and 20, respectively.

4 DISCUSSION

We have demonstrated that chemistry and clouds impact the variability of high contrast imaging spectra of oxygenated terrestrial exoplanets. The variability depends upon the exoplanet's climate and chemistry, but the ability to detect this variation will be influenced by the distance, orbital geometry, and telescope design. Naturally, this will dictate what scientific results are retrievable from specific targets.

4.1 Molecular detectability

We acknowledge that the SNR estimates in Section 3.3.1 depend on various parameters that we have assumed, such as a 90° inclination angle, as well as the underpinning simulations from the Earth System Model WACCM6. Other chemistry-climate models when used in conjunction with PSG as we have here would yield different results (Fauchez et al. 2022). The estimates also depend on telescope

concepts that have not been constructed. Therefore, the results are an indicative, not exhaustive, picture of possible future measurements based on the information available for each concept at the time of writing.

In the event that an Earth-like exoplanet is found even farther away (e.g. 75 pc) it would be much more challenging to detect any particular molecular feature. A longer integration time may not improve the performance because the coronagraph throughput significantly drops with distance for these concepts and the exoplanet would be observable only over a narrow range of phases and wavelengths, if any. The coronagraph throughput drops when the exoplanet passes inside the IWA of the coronagraph and is therefore partly or fully obscured by the coronagraph. The IWA for the LUVOIR concept is proportional to wavelength so longer wavelengths are cut off and cannot be detected (The LUVOIR Team 2019) – see Appendix B.

The SNR of each feature is set in part by its depth relative to the continuum. It also depends on the telescope's diameter, the telescope's coronagraph, the total integration time, and distance to the planetary system. For example, even though the H_2O feature at 1.4 μm is deeper than that at 0.94 μm, the SNR is lower because of reduced coronagraphic throughput, and higher noise, in the near-infrared. The HabEx with a starshade concept, when compared to a LUVOIR concept with an identically sized primary mirror, enables characterization at greater distance due to the enhanced coronagraphic throughput (Checlair et al. 2021).

4.2 Predicted spectra variations from a 3D chemistry-climate model

The O_2 concentrations that we calculate in the lower and middle atmosphere are identical to those predicted by previous work (Kasting & Donahue 1980; Segura et al. 2003; Gregory et al. 2021). However, we find that the lower and middle atmospheric concentrations of H_2O, O_3, CH_4, OH, and other molecules, differ considerably compared to prior 1D and 3D modelling (Cooke et al. 2022). In this work, we have considered the temporal variability of chemistry and climate and how this affects predicted reflection spectra.

Although the effect of chemistry on the spectra for different oxygenation states is evident in Fig. 2, chemistry has a smaller impact than when considering the time variability from the same simulation. The largest variation between years is caused by clouds, then chemistry, and then surface albedo. Nevertheless, surface albedo is important for seasonal cycles and is crucial for understanding the spectra presented in Section 3.3.3. 1D atmospheric models that run to steady state cannot access this temporal variability, nor can they assess the seasonal effect on exoplanet spectra. Because clouds in 3D global climate models show significant variability between models (Ceppi et al. 2017; Lauer & Hamilton 2013; Fauchez et al. 2020), future comparisons with 3D time-resolved simulations using a different global climate model will therefore be of interest to determine how the predicted magnitude of spectral variations from HCI differ between models and oxygenation states (or various other chemical regimes). A comparison has been done for transmission spectra from 3D models in the TRAPPIST Habitable Atmosphere Intercomparison (Fauchez et al. 2022). These comparisons could also be done with 1D models to evaluate the uncertainty across the entire range of models used in the exoplanet community, which would then be used to assign confidence to the interpretation of future observations.

Our simulated spectra were evaluated at 00:00 UTC, such that the Pacific ocean is illuminated, with a small proportion of the

Earth's land and ice illuminated. The albedos of land and ice are both greater than that of the ocean, so reflection spectra from HCI could be affected if observations were calculated at different times. Additionally, the model time step is 30 min in our simulations, but we assumed this instantaneous state at 00:00 UTC held for long enough to integrate 24 or 48 h of observations over. Our reasoning for this assumption is that saving multiple 3D variables every time-step for four model years and for multiple oxygenation states, then simulating each time-step snapshot with PSG (and with three telescope concepts), would require vastly more data storage and computational expense. As this assumption likely does not hold true (see the diurnal variations predicted in low precision photometry measurements by Ford et al. 2001), greater temporal variations may be introduced by further surface albedo and cloud variations. These variations during the time-scale of the observations are likely to be important, especially when the observer does not know the rotation rate of the exoplanet. For instance, Guzewich et al. (2020) predicted how different planetary rotation rates affected the absorption depth of H_2O, O_2, and O_3 features. Fortunately, the rotation rate may be known provided that approximately 5–15 rotations are observed, and the integration time is less than 1/6 to 1/4 of the planetary rotation period (Li et al. 2022).

The exoplanet is usually at its brightest just before secondary eclipse, but with the coronagraphs simulated here, such orbital phases are not observable as the planet is also being blocked by the coronagraph. Therefore, there is a trade-off between detectability and brightness. Other orbital phases show different ranges between minimum and maximum flux (see Fig. 6), so the observer may not be able to constrain certain parameters, including the amount of year-to-year variation across multiple phases, depending on the telescope used and the distance to the exoplanetary system.

With the mission concepts evaluated in this work, monitoring a nearby (<25 pc) Earth-analogue exoplanet for a single orbit is long enough to inform the observer about short-term climate variability and is ample time to detect if molecules of interest (e.g. O_2, O_3, and H_2O) are present. If inter-annual climate variability is of interest to future observers, then our simulations suggest it is observable provided conditions are favourable. Fully characterizing an exoplanet (determining temperature, chemistry, seasons, obliquity, cloud coverage, rotation rate, land and ocean coverage, etc.) will be aided by observations of multiple orbits. Nevertheless, there will remain significant challenges to interpreting those observations.

4.3 Challenges and future work

Our atmospheric scenarios are based on the current understanding of Earth's atmospheric history since the GOE, when Earth's atmospheric oxygen concentrations rose above 10^{-5} PAL. However, the abundance of gases throughout Earth's history is not well constrained. Even if better constraints are achieved in the future, there is no reason why the atmospheric evolution of exoplanets that are oxygenated must take a similar path to Earth (Krissansen-Totton et al. 2021). None the less, the Earth system provides the best-known template for an oxygenated atmosphere of a terrestrial planet. Large changes in the magnitude of the mixing ratios of other chemical species (e.g. CO_2, CH_4), decreased or increased atmospheric pressure, and the proportion of land versus ocean coverage (Macdonald et al. 2022), will all affect our predictions.

Changes in assumed cloud properties will also change the calculations presented in this work. Komacek et al. (2020) varied ice cloud particle radius between 20 and 200 μm and showed that this non-linearly affected the amount of transits required to detect water

vapour in transmission spectra of a terrestrial exoplanet atmosphere using the *James Webb Space Telescope*. Although we have not covered a large range of cloud parameters, our results show that assumed particle sizes are important in PSG, affecting both variability and SNRs. Thus, the effects of how clouds are modelled in global climate models and radiative transfer models, and their impacts on predictions, requires further investigation.

There are many unknowns for exoplanets that compound the difficulty of modelling and interpreting spectra from a directly-imaged exoplanet. We have investigated a temporal and 3D problem that depends on: (1) albedo; (2) seasonal climatic variability (Cowan, Voigt & Abbot 2012); (3) clouds (Parmentier & Crossfield 2018); (4) chemistry (Parmentier & Crossfield 2018); (5) inter-annual climatic variability; (6) eccentricity (Cowan et al. 2012); (7) obliquity (Gaidos & Williams 2004; Cowan et al. 2012); (8) observer inclination angle (Maurin et al. 2012; Boutle et al. 2017); and (9) other exoplanets in the system contaminating the signal (Parmentier & Crossfield 2018). Some of these variables interact with each other. Our high-contrast imaging spectra predictions account for some of these parameters (points 1–5) using a 3D fully coupled climate model for the first time, whilst others were not investigated in this work (6–9). Further work should investigate how these other parameters affect the predicted spectra of oxygenated exoplanets. At greater distances, where less orbital phases are available to observe, assessments of the year-to-year variability we demonstrate in our simulations will be less reliable. Additionally, our simulations assume that whilst F_* varies with wavelength, the flux in each wavelength bin is assumed to be constant over the duration of an orbit. If we instead modified our simulations to include realistic stellar flux variations, it is unclear how detectable the observed flux variations due to the exoplanet would be.

Several methods already exist to obtain information that is otherwise difficult to access. For example, exo-cartography has been studied as a process for retrieving the surface maps of exoplanets, despite the planetary surface being unresolved in observations (Cowan & Fujii 2018; Fan et al. 2019; Berdyugina & Kuhn 2019). Already it is clear that different surfaces produce similar light curves, and lifting degeneracies to accurately map the surface of these exoplanets will be difficult. Teinturier et al. (2022) used Earth observation data and found that for exoplanets with a similar climate to Earth, removing clouds to obtain a surface albedo map may produce unreliable results. They remarked that longer, continuous observing periods will help with mapping. Indeed, clouds may have the largest impact on spectra variation (Paradise et al. 2022), as we have also found in our simulations. Additionally, Schwartz et al. (2016) found that, at least theoretically, it should be possible to determine the obliquity of an exoplanet by observing at 2-4 different orbital phases. With known obliquity, the impact of seasons can be estimated. Recently, Pham & Kaltenegger (2022) used a machine learning algorithm – analysing more than 50 000 synthetic spectra of cold, Earth-like planets – to show that snow/clouds and water could be detected in 90 per cent and 70 per cent of cases, respectively. A similar evaluation using machine learning, and including both spectra produced from 1D and 3D atmospheric models, would be useful. Such a study would enable assessment of how reliably the community can use machine learning and 1D models to explore the parameter space that 3D models struggle to cover due to computational expense.

All of our simulations presume a biotic source of O_2. Even if a biogenic molecule is detected with high SNR, this may not be enough to prove that life exists on an exoplanet. Further information will likely be required for context to confirm that any potential biosignature that is detected has arisen from the presence of extraterrestrial life (Meadows 2017; Catling et al. 2018; Krissansen-Totton et al. 2021). This includes but is not limited to the detection of chemical disequilibrium (e.g. O_2 and CH_4; Lovelock 1965; Sagan et al. 1993; Krissansen-Totton, Bergsman & Catling 2016), or temporal changes in biogenic gases that indicate atmospheric modulation by life (Olson et al. 2018b) rather than by geological processes. However, detecting such temporal chemical variations will likely be complicated by temporal changes in planetary albedo. We note here that detecting seasonal O_3 variations in the UV as suggested by Olson et al. (2018b) may be complicated by clouds, as well as the seasons available to the observer, and because we have not explored all observing geometries, the effect of different inclination angles (Olson et al. 2018b).

Additionally, abiotic build-up of O_2 is physically feasible on exoplanets (and references therein Meadows 2017). Moreover, there is the possibility of detecting an exoplanet–exomoon system which has O_2 and CH_4 in the spectra, but not on the same celestial body (Rein, Fujii & Spiegel 2014). In the coming decades, observational data, computational procedures which identify the sources of biologically-relevant gases (Catling et al. 2018; Lisse et al. 2020), and simulations that utilize state-of-the art climate models, will all be needed to determine the most habitable exoplanets (Shields 2019), and indeed to determine if any are inhabited.

Finally, we have only evaluated telescope concepts which will have spectroscopic capabilities between the ultraviolet and near-infrared wavelength regions. To detect molecules such as O_3, H_2O, and CO_2 in the infrared, this will require a mission design different to LUVOIR or HabEx. Proposed candidates include the Large Interferometer For Exoplanets (LIFE; Quanz et al. 2018; Defrère et al. 2018) and the Origins Space telescope (OST; Wiedner et al. 2021) missions. These telescope concepts may be able to constrain planetary emission in the mid-infrared and detect molecules such as H_2O, O_3, and CH_4 (Alei et al. 2022), potentially enabling the confirmation of chemical disequilibrium. Certainly, multiwavelength observations which cover ultraviolet wavelengths to mid infrared wavelengths will enable a greater degree of atmospheric characterization. Evaluating the OST and LIFE missions with our simulated atmospheres and calculating any variability that may be detected is left as future work.

5 CONCLUSIONS

Using a state-of-the-art Earth System Model, WACCM6, we have simulated Earth-analogue exoplanets around a G2V star with varying concentrations of O_2 (0.1 per cent PAL → 150 per cent PAL). Based on the Earth System Model output for these hypothetical exoplanets, we used the PSG to simulate thousands of high-contrast imaging observations with three future telescope concepts: LUVOIR A, LUVOIR B, and HabEx with a starshade.

All telescopes are capable of detecting O_3 at 10 pc after 1 d of integration time. H_2O and O_2 detection may require longer integration times, or high O_2 concentrations for improved SNR for O_2 (e.g. ≥ 10 per cent PAL). At 25 pc, O_3 in the UV can still be detected with LUVOIR A, but integration times longer 13 d may be required for H_2O and O_2 detection.

Long-term interannual climate variations and short-term variations can theoretically be observed in our simulated atmospheres with a telescope concept such as LUVOIR or HabEx. Both annual and seasonal variability can affect the brightness of key chemical features such as O_2 and H_2O. The annual variability is primarily caused by clouds, and the variability appears to depend non-linearly on atmospheric O_2 concentration (although more than four orbits would enable this to be a statistically robust result), as well as depending on wavelength, distance, and telescope design. The seasons that can be

accessed through observations will affect the maximum magnitude of the incoming flux, with more ice coverage increasing this limit.

Future potential missions, such as the 6-m diameter UV/VIS/IR telescope that was recommended by the Decadal Survey, and the LIFE mission, will present exciting opportunities to characterize the atmospheres of possible Earth analogues. Confident interpretation of future observations that next-generation telescopes return will require sophisticated planetary modelling and observational analysis. It is important that future work continues to account for the complexity of chemistry and clouds, as well as long- and short-term climate variations.

DATA AVAILABILITY

The data underlying this article are available in the Dryad Digital Repository, at https://doi.org/10.5061/dryad.cz8w9gj6f.

REFERENCES

Alei E. et al., 2022, A&A, 665, A106
Arney G. et al., 2016, Astrobiology, 16, 873
Berdyugina S. V., Kuhn J. R., 2019, AJ, 158, 246
Betts H. C., Puttick M. N., Clark J. W., Williams T. A., Donoghue P. C., Pisani D., 2018, Nat. Ecol. Evol., 2, 1556
Boutle I. A., Mayne N. J., Drummond B., Manners J., Goyal J., Hugo Lambert F., Acreman D. M., Earnshaw P. D., 2017, A&A, 601, A120
Bryson S. et al., 2021, AJ, 161, 36
Burke C. J. et al., 2015, ApJ, 809, 8
Catanzarite J., Shao M., 2011, ApJ, 738, 151
Catling D. C. et al., 2018, Astrobiology, 18, 709
Catling D. C., Zahnle K. J., 2020, Sci. Adv., 6, eaax1420
Catling D. C., Zahnle K. J., McKay C. P., 2001, Science, 293, 839
Ceppi P., Brient F., Zelinka M. D., Hartmann D. L., 2017, Wiley Interdiscip. Rev. Climate Change, 8, e465
Charbonneau D., Brown T. M., Noyes R. W., Gilliland R. L., 2002, ApJ, 568, 377
Charnay B., Wolf E. T., Marty B., Forget F., 2020, Space Sci. Rev., 216, 90
Checlair J. H. et al., 2021, AJ, 161, 150
Cole D. B., Mills D. B., Erwin D. H., Sperling E. A., Porter S. M., Reinhard C. T., Planavsky N. J., 2020, Geobiology, 18, 260
Cooke G. J., Marsh D. R., Walsh C., Black B., Lamarque J. F., 2022, Royal Soc. Open Sci., 9, 211165
Cowan N. B., Fujii Y., 2018, Handbook of Exoplanets, Mapping Exoplanets. Springer International Publishing AG, Cham, New York, p. 147
Cowan N. B., Voigt A., Abbot D. S., 2012, ApJ, 757, 80
Currie T., Biller B., Lagrange A.-M., Marois C., Guyon O., Nielsen E., Bonnefoy M., De Rosa R., 2022, preprint (arXiv:2205.05696)
De Cock R., Livengood T. A., Stam D. M., Lisse C. M., Hewagama T., Deming L. D., 2022, AJ, 163, 5
Defrère D. et al., 2018, Exp. Astron., 46, 543
Dodd M. S., Papineau D., Grenne T., Slack J. F., Rittner M., Pirajno F., O'Neil J., Little C. T., 2017, Nature, 543, 60
Fan S., Li C., Li J.-Z., Bartlett S., Jiang J. H., Natraj V., Crisp D., Yung Y. L., 2019, ApJ, 882, L1
Fauchez T. J. et al., 2020, Geosci. Model Devel., 13, 707
Fauchez T. J. et al., 2022, The Planetary Science Journal, 3, 213
Ford E. B., Seager S., Turner E. L., 2001, Nature, 412, 885
Fressin F. et al., 2013, ApJ, 766, 81
Fueglistaler S., Dessler A. E., Dunkerton T. J., Folkins I., Fu Q., Mote P. W., 2009, Rev. Geophys., 47, RG1004
Fujii Y., Kawahara H., Suto Y., Fukuda S., Nakajima T., Livengood T. A., Turner E. L., 2011, ApJ, 738, 184
Gaidos E., Williams D. M., 2004, New Astron., 10, 67
Gaudi B. S. et al., 2020, preprint (arXiv:2001.06683)
Gettelman A. et al., 2019, J. Geophys. Res. (Atmospheres), 124, 12,380
Gordon I. E. et al., 2017, J. Quant. Spec. Radiat. Transf., 203, 3
Gregory B. S., Claire M. W., Rugheimer S., 2021, Earth Planetary Sci. Lett., 561, 116818
Grenfell J. L., 2017, Phys. Rep., 713, 1
Guzewich S. D., Lustig-Yaeger J., Davis C. E., Kopparapu R. K., Way M. J., Meadows V. S., 2020, ApJ, 893, 140
Haqq-Misra J. D., Domagal-Goldman S. D., Kasting P. J., Kasting J. F., 2008, Astrobiology, 8, 1127
Kaltenegger L., Lin Z., Madden J., 2020a, ApJ, 892, L17
Kaltenegger L., Lin Z., Rugheimer S., 2020b, ApJ, 904, 10
Kaltenegger L., Traub W. A., 2009, ApJ, 698, 519
Kaltenegger L., Traub W. A., Jucks K. W., 2007, ApJ, 658, 598
Kasting J. F., Donahue T. M., 1980, J. Geophys. Res., 85, 3255
Kasting J. F., Siefert J. L., 2002, Science, 296, 1066
Kleinböhl A., Willacy K., Friedson A. J., Chen P., Swain M. R., 2018, ApJ, 862, 92
Komacek T. D., Fauchez T. J., Wolf E. T., Abbot D. S., 2020, ApJ, 888, L20
Krissansen-Totton J., Bergsman D. S., Catling D. C., 2016, Astrobiology, 16, 39
Krissansen-Totton J., Fortney J. J., Nimmo F., Wogan N., 2021, AGU Adv., 2, e00294
Large R. R., Mukherjee I., Gregory D., Steadman J., Corkrey R., Danyushevsky L. V., 2019, Mineralium Deposita, 54, 485
Lauer A., Hamilton K., 2013, J. Climate, 26, 3823
Lee C.-H., 2018, Galaxies, 6, 51
Li J., Jiang J. H., Yang H., Abbot D. S., Hu R., Komacek T. D., Bartlett S. J., Yung Y. L., 2022, AJ, 163, 27
Lisse C. M. et al., 2020, ApJ, 898, L17
Livengood T. A. et al., 2011, Astrobiology, 11, 907
Lovelock J. E., 1965, Nature, 207, 568
Lyons T. W., Diamond C. W., Planavsky N. J., Reinhard C. T., Li C., 2021, Astrobiology, 21, 906
Lyons T. W., Reinhard C. T., 2009, Nature, 461, 179
Macdonald E., Paradise A., Menou K., Lee C., 2022, MNRAS, 513, 2761
Maurin A. S., Selsis F., Hersant F., Belu A., 2012, A&A, 538, A95
Meadows V. S., 2017, Astrobiology, 17, 1022
Montañés-Rodríguez P., Pallé E., Goode P. R., Martín-Torres F. J., 2006, ApJ, 651, 544
National Academies of Sciences, Engineering, and Medicine, 2021, Pathways to Discovery in Astronomy and Astrophysics for the 2020s. The National Academies Press, Washington, DC
Olson S. L., Schwieterman E. W., Reinhard C. T., Lyons T. W., 2018a, Handbook of Exoplanets, Earth: Atmospheric Evolution of a Habitable Planet. Springer Nature, New York, p. 189

Olson S. L., Schwieterman E. W., Reinhard C. T., Ridgwell A., Kane S. R., Meadows V. S., Lyons T. W., 2018b, ApJ, 858, L14

Paradise A., Menou K., Lee C., Fan B. L., 2022, MNRAS, 512, 3616

Parmentier V., Crossfield I. J. M., 2018, Exoplanet Phase Curves: Observations and Theory. Springer, Cham, p. 116

Petigura E. A., Howard A. W., Marcy G. W., 2013, Proc. Natl. Acad. Sci., 110, 19273

Pham D., Kaltenegger L., 2022, MNRAS, 513, L72

Quanz S. P., Kammerer J., Defrère D., Absil O., Glauser A. M., Kitzmann D., 2018, in Creech-Eakman M. J., Tuthill P. G., Mérand A., eds, Proc. SPIE Conf. Ser. Vol. 10701, Optical and Infrared Interferometry and Imaging VI. SPIE, Bellingham, p. 107011I

Rein H., Fujii Y., Spiegel D. S., 2014, Proc. Natl. Acad. Sci., 111, 6871

Rugheimer S., Kaltenegger L., 2018, ApJ, 854, 19

Rugheimer S., Segura A., Kaltenegger L., Sasselov D., 2015, ApJ, 806, 137

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

Schwartz J. C., Sekowski C., Haggard H. M., Pallé E., Cowan N. B., 2016, MNRAS, 457, 926

Seager S., Deming D., 2010, ARA&A, 48, 631

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

Selsis F., 2000, in Schürmann B., ed., ESA Special Publication, Vol. 451, Darwin and Astronomy: The Infrared Space Interferometer. ESA, Noordwijk, the Netherlands, p. 133

Selsis F., Despois D., Parisot J. P., 2002, A&A, 388, 985

Shields A. L., 2019, ApJS, 243, 30

Steadman J. A. et al., 2020, Precambrian Res., 343, 105722

Teinturier L., Vieira N., Jacquet E., Geoffrion J., Bestavros Y., Keating D., Cowan N. B., 2022, MNRAS, 511, 440

The LUVOIR Team, 2019, The LUVOIR Mission Concept Study Final Report, preprint (arXiv:1912.06219)

Tinetti G., Meadows V. S., Crisp D., Fong W., Fishbein E., Turnbull M., Bibring J.-P., 2006a, Astrobiology, 6, 34

Tinetti G., Meadows V. S., Crisp D., Kiang N. Y., Kahn B. H., Fishbein E., Velusamy T., Turnbull M., 2006b, Astrobiology, 6, 881

Traub W. A., 2012, ApJ, 745, 20

Villanueva G. L., Liuzzi G., Faggi S., Protopapa S., Kofman V., Fauchez T., Stone S. W., Mandell A. M., 2022, Fundamentals of the Planetary Spectrum Generator. Available at: https://psg.gsfc.nasa.gov/help.php

Villanueva G. L., Mumma M. J., DiSanti M. A., Bonev B. P., Gibb E. L., Magee-Sauer K., Blake G. A., Salyk C., 2011, Icarus, 216, 227

Villanueva G. L., Smith M. D., Protopapa S., Faggi S., Mandell A. M., 2018, J. Quant. Spec. Radiat. Transf., 217, 86

Warke M. R. et al., 2020, Proc. Natl. Acad. Sci., 117, 13314

Way M. J. et al., 2017, ApJS, 231, 12

Wenger M. et al., 2000, A&AS, 143, 9

Wiedner M. C. et al., 2021, Exp. Astron., 51, 595

Wordsworth R., Pierrehumbert R., 2014, ApJ, 785, L20

Zahnle K. J., Gacesa M., Catling D. C., 2019, Geochim. Cosmochim. Acta, 244, 56

Zink J. K., Hansen B. M. S., 2019, MNRAS, 487, 246

SUPPORTING INFORMATION

Supplementary data are available at *MNRAS* online.

Supplementary Information for Variability due to climate and chemistry in observations of oxygenated Earth-analogue exoplanets.

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

APPENDIX A: FURTHER DETAILS OF COMPLETED PSG SIMULATION

We have not performed an exhaustive list of simulations in PSG. The parameter space is vast and to cover every future telescope concept, planet distance, oxygenation state, etc., is beyond the scope of this paper. See Table A1 for the simulations that have been performed.

The PSG packages required for these calculations were BASE, PROGRAMS, and CORRKLOW. The versions used were last updated on 20/05/2022, 04/02/2022, and 01/03/2022, respectively.

Each simulation used NMAX = 3 and LMAX = 41. NMAX is defined as the number of stream pairs and the Legendre terms is given as LMAX – see the Fundamentals of the Planetary Spectrum Generator for more details (Villanueva et al. 2022). Calculation speed is proportional to LMAX but is approximately proportional to $NMAX^3$ (Villanueva et al. 2022). Multiple tests with different values of NMAX and LMAX were performed to ensure spectra contributions from scattering were accurate, while not compromising the speed in such a way that collating the simulations would be unachievable within a reasonable time frame. In a small proportion of cases (<2 per cent) where unusually large brightness occurred, likely due to an asymmetry calculation in PSG, the substellar point was slightly shifted horizontally (<3°) to produce consistent results.

Table A1. For each of the PI, 10 per cent PAL, 1 per cent PAL, and 0.1 per cent PAL atmospheres, the PSG simulations that have been run are listed below. A single checkmark means that LUVOIR A has been evaluated in all three channels (UV, VIS, NIR) for 73 phase points around the orbit (one snapshot every 5 d), for a 24 h integration time. Three checkmarks mean that simulated HCI observations for a full orbit for LUVOIR A, LUVOIR B, and HabEx with a starshade, have been evaluated. If they have also been evaluated at 25 and 50 pc, this is indicated in brackets. A dash means no telescope observations have been evaluated. A standard orbit is the where we have used PSG to calculate the total flux in Jy for each particular date and phase, using the ephemeris data from the year 2020. SNRs have been calculated where a full orbit has been simulated in PSG for a particular atmosphere which has had all molecules removed apart from N_2 and clouds. A phase shift orbit has a checkmark if spectra have been produced for the +90°, +180°, and +270° configurations. The water cloud particle size (WCPS) and the ice cloud particle size (ICPS) is indicated.

	Year 1	Year 2	Year 3	Year 4
Standard orbit (5 μm WCPS, 100 μm ICPS)	✓✓✓(10, 25, and 50 pc)	✓(10 pc)	✓(10 pc)	✓(10 pc)
Standard orbit SNR (5 μm WCPS, 100 μm ICPS)	✓✓✓(10, 25, and 50 pc)	–	–	–
Phase shift orbit (5 μm WCPS, 100 μm ICPS)	–	✓(10 pc)	✓(10 pc)	–
Standard orbit (1 μm WCPS, 1 μm ICPS)	✓✓✓(10, 25, and 50 pc)	✓(10 pc)	✓(10 pc)	✓(10 pc)
Standard orbit SNR (1 μm WCPS, 1 μm ICPS)	✓✓✓(10, 25, and 50 pc)	–	–	–
Phase shift orbit (1 μm WCPS, 1 μm ICPS)	–	✓(10 pc)	✓(10 pc)	–

APPENDIX B: OBSERVABILITY OF EXOPLANET USING CORONAGRAPH

As an exoplanet orbits a star, with an orbital phase denoted by ψ, the coronagraph throughput varies because the projected exoplanet-star angular separation changes for the observer. When using a coronagraph, the IWA is typically defined as

$$IWA = \frac{\lambda}{D},$$ (B1)

where λ is the wavelength of light and D is the diameter of the telescope. The telescope coronagraphs evaluated here are the LUVOIR ECLIPS instrument and HabEx with a starshade. For LUVOIR, the IWA is given in the LUVOIR Final Report (The LUVOIR Team 2019) as $4\lambda/D$ for the UV channel, and $3.5\lambda/D$ for the VIS and NIR channels. For HabEx with a starshade, an IWA of 58 mas at 0.3–1.0 μm is given in the HabEx Final Report (Gaudi et al. 2020). For the planet to be visible (and not blocked by the coronagraph), the IWA must be smaller than the angular separation of the star and planet (θ), which is given by

$$\theta \approx \frac{s}{d},$$ (B2)

where s is the projected separation and d is the distance to the planet–star system. The maximum projected separation for a circular orbit is the semimajor axis (a; 1 au for Earth). For circular orbits of inclination, i, the separation, s, takes the general form

$$s = a\sqrt{\sin^2(\psi) + \cos^2(i)\cos^2(\psi)}.$$ (B3)

In the simplified case of a system inclined at 90° from the perspective of an observer,

$$s = \pm a\sin(\psi).$$ (B4)

When the IWA $<\theta$, the exoplanet is fully observable and when the IWA $>\theta$, the coronagraph throughput is reduced. The equations can be rearranged to find the orbital phase at which the exoplanet is about to enter inside the IWA, such that

$$\psi = \sin^{-1}\left(\frac{\lambda d}{aD}\right) \text{ for } \frac{\lambda d}{aD} \leq 1.$$ (B5)

Thus, one can see that as the wavelength λ increases or the distance d increases, then a smaller proportion of the orbit will be observable with the coronagraph. On the other hand, if the semimajor axis or the telescope diameter increases, then a larger proportion of the total orbit will be accessible for higher SNR observations.

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