Astronomy & Astrophysics

Large Interferometer For Exoplanets (LIFE)

V. Diagnostic potential of a mid-infrared space interferometer for studying Earth analogs

Eleonora Alei[1,2] , Björn S. Konrad[1,2] , Daniel Angerhausen[1,2,3] , John Lee Grenfell[4], Paul Mollière[5],
Sascha P. Quanz[1,2] , Sarah Rugheimer[6] , Fabian Wunderlich[4], and the LIFE Collaboration[*]

[1] ETH Zurich, Institute for Particle Physics & Astrophysics, Wolfgang-Pauli-Str. 27, 8093 Zurich, Switzerland
  e-mail: `elalei@phys.ethz.ch`
[2] National Centre of Competence in Research PlanetS, Gesellschaftsstrasse 6, 3012 Bern, Switzerland
[3] Blue Marble Space Institute of Science, Seattle, USA
[4] Department of Extrasolar Planets and Atmospheres (EPA), Institute for Planetary Research (PF), German Aerospace Centre (DLR),
  Rutherfordstr. 2, 12489 Berlin, Germany
[5] Max-Planck-Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany
[6] Department of Physics, University of Oxford, Oxford OX1 3PU, UK

ABSTRACT

Context. An important future goal in exoplanetology is to detect and characterize potentially habitable planets. Concepts for future space missions have already been proposed: from a large UV-optical-infrared space mission for studies in reflected light, to the Large Interferometer for Exoplanets (LIFE) for analyzing the thermal portion of the planetary spectrum. Using nulling interferometry, LIFE will allow us to constrain the radius and effective temperature of (terrestrial) exoplanets, as well as provide unique information about their atmospheric structure and composition.
Aims. We explore the potential of LIFE for characterizing emission spectra of Earth at various stages of its evolution. This allows us (1) to test the robustness of Bayesian atmospheric retrieval frameworks when branching out from a modern Earth scenario while still remaining in the realm of habitable (and inhabited) exoplanets, and (2) to refine the science requirements for LIFE for the detection and characterization of habitable, terrestrial exoplanets.
Methods. We performed Bayesian retrievals on simulated spectra of eight different scenarios, which correspond to cloud-free and cloudy spectra of four different epochs of the evolution of the Earth. Assuming a distance of 10 pc and a Sun-like host star, we simulated observations obtained with LIFE using its simulator LIFEsim, considering all major astrophysical noise sources.
Results. With the nominal spectral resolution ($R = 50$) and signal-to-noise ratio (assumed to be S/N = 10 at 11.2 μm), we can identify the main spectral features of all the analyzed scenarios (most notably CO_2, H_2O, O_3, and CH_4). This allows us to distinguish between inhabited and lifeless scenarios. Results suggest that O_3 and CH_4 in particular yield an improved abundance estimate by doubling the S/N from 10 to 20. Neglecting clouds in the retrieval still allows for a correct characterization of the atmospheric composition. However, correct cloud modeling is necessary to avoid biases in the retrieval of the correct thermal structure.
Conclusions. From this analysis, we conclude that the baseline requirements for R and S/N are sufficient for LIFE to detect O_3 and CH_4 in the atmosphere of an Earth-like planet with an O_2 abundance of around 2% in volume mixing ratio. Doubling the S/N would allow a clearer detection of these species at lower abundances. This information is relevant in terms of the LIFE mission planning. We also conclude that cloud-free retrievals of cloudy planets can be used to characterize the atmospheric composition of terrestrial habitable planets, but not the thermal structure of the atmosphere. From the inter-model comparison performed, we deduce that differences in the opacity tables (caused by, e.g., a different line wing treatment) may be an important source of systematic errors.

Key words. methods: statistical – planets and satellites: terrestrial planets – planets and satellites: atmospheres

1. Introduction

Temperate terrestrial exoplanets are predicted to be very abundant in our galaxy (Bryson et al. 2021). These planets are ideal candidates when searching for life beyond our Solar System. A powerful way to characterize a terrestrial exoplanet in the context of its habitability is by detecting and studying its atmosphere with the goal of constraining its surface conditions. Atmospheric spectra are influenced by many parameters and processes, such as the chemical composition, the temperature structure of the

atmosphere, the presence of clouds, and emission and scattering from the surface.

The detection and characterization of potentially habitable, rocky exoplanets is challenging with current facilities. For this reason, there is widespread interest in the community to build new instruments for the search for life in the universe, as reported in the White Paper series in the context of the ESA "Voyage 2050" process[1], as well as the US Astro 2020 Decadal survey (National Academies of Sciences, Engineering, and Medicine 2021). Space missions that aim at characterizing terrestrial

---

[*] `www.life-space-mission.com`

[1] `https://www.cosmos.esa.int/web/voyage-2050`

exoplanets have been proposed, such as the Habitable Exoplanet Observatory (HabEx; Gaudi et al. 2020) and the Large Ultraviolet Optical Infrared Surveyor (LUVOIR; Peterson et al. 2017), which focus on the reflected (visible and near-infrared) portion of the planetary spectrum, as well as the Large Interferometer for Exoplanets (LIFE; Quanz et al. 2022, hereafter Paper I), which will characterize terrestrial planets in the thermal (mid-infrared) emitted portion of the planetary spectrum. Using nulling interferometry, LIFE will allow us to constrain the radius and effective temperature of (terrestrial) exoplanets, as well as provide unique information about their atmospheric structure and composition (Dannert et al. 2022; Konrad et al. 2022, hereafter Papers II and III, respectively).

Due to the current lack of high-quality observational data, we must rely for now on simulated observations of terrestrial planets to create and improve the analysis algorithms, but also to provide scientific and technical requirements when planning a mission. This effort is currently ongoing within the LIFE Initiative; in a previous study (Paper III), we built a Bayesian retrieval routine to estimate the planetary and atmospheric parameters of a simulated modern Earth twin at a distance of 10 pc as it would be observed by LIFE. In this work, we extend this exercise to other stages in the evolution of Earth's atmosphere.

Our planet has been habitable for about 4.4 billion years (see, e.g., Heller et al. 2021, and references therein). In this context, we define a planet as habitable if its physical and chemical conditions would allow water, if present, to be liquid on the surface.

In the prebiotic stage of Earth's evolution, the atmosphere lacked O_2 (currently about 21% of the atmospheric composition by volume). It was instead a CO_2–N_2–H_2O-rich atmosphere, with traces of CH_4 from volcanism. Early forms of life developed under a reducing environment and survived under anaerobic conditions (Olson et al. 2018, and references therein). Methanogenesis was thought to be a dominant metabolism at this stage (around 3.5 Ga), which would explain the increase in CH_4 in the atmosphere (see, e.g., Wolfe & Fournier 2018).

Around 3 Ga, life-forms that could use carbon dioxide to produce oxygen (via oxygenic photosynthesis) appeared (Marais 2000). This eventually led to a significant increase in O_2, maximally up to ~1% PAL[2] (see Gregory et al. 2021; Lyons et al. 2014, 2021, and references therein) around 2.33 Ga (Luo et al. 2016), during the so-called Great Oxygenation Event (GOE). There is also evidence pointing to a second increase in the O_2 abundance (up to ~10% PAL) that occurred around 0.8 Ga, during the "Neoproterozoic Oxygenation Event" (NOE; Shields-Zhou & Och 2011; Campbell & Squire 2010).

The high abundance of carbon dioxide in the early Earth's atmosphere would have enhanced the atmospheric greenhouse effect, allowing Earth to be habitable despite the fainter solar irradiation (see, e.g., Feulner 2012, and references therein). The positive feedback between the carbon-silicate cycle and the increase in irradiation would have then allowed temperatures conducive to liquid water to be maintained over the last 4 Ga. The increase in irradiation from the Sun over the eons has made the weathering of CO_2 more efficient, decreasing the amount of carbon dioxide in the atmosphere and thus dampening the atmospheric greenhouse effect (see, e.g., Graham 2021, and references therein). The appearance of photosynthetic life-forms and the onset of plate tectonics also contributed to the depletion of atmospheric CO_2.

Numerous processes, including biology and geology, have driven the wide-ranging evolution of Earth's atmosphere during the various epochs of its development. Our modern atmosphere represents only a small fraction of Earth's evolutionary states. It is therefore important to simulate a suitable range of different atmospheric epochs from Earth's history when investigating Earth-like atmospheres. For this study, we simulated observations obtained by LIFE starting from theoretical spectra of four distinct epochs of Earth's atmospheric evolution, produced from a self-consistent one-dimensional climate and photochemistry model coupled with a line-by-line radiative transfer model (Rugheimer & Kaltenegger 2018). The observed spectra were simulated using the LIFE noise simulator LIFEsim (for details on the simulator, see Paper II). We then used the Bayesian retrieval routine presented in Paper III to characterize the different atmospheres.

We aim to address a number of research questions. The science-driven questions are: (1) how well LIFE will be able to characterize atmospheres of habitable planets; (2) whether LIFE will be able to differentiate between different atmospheres, and with what confidence; (3) what the impact of clouds on this assessment will be; and (4) what the most promising (combinations of) detectable biosignatures are.

Technology- and computationally driven questions include: (1) whether the combination of the spectral resolution ($R = \lambda/\Delta\lambda$), signal-to-noise ratio (S/N), and wavelength range defined in Paper III are still adequate for this case study; (2) what the caveats and limitations of the Bayesian retrieval routine are; and (3) what systematics may arise when comparing two different models (e.g., in terms of differences in line lists, scattering treatment, and identification of biomarkers).

We discuss how we adapted the input spectra to simulate LIFE observations and describe the grid of scenarios in Sect. 2. We show and describe the results in Sect. 3. A thorough discussion of our findings and of the potential systematic uncertainties of the retrieval routine is provided in Sect. 4. In Sect. 5, we report the main takeaway points from this study, and in Sect. 6 we provide an overview of ongoing and future studies.

2. Methods

We start by discussing the details of the input spectra that were used in this study (Sect. 2.1). We then discuss the updates on the Bayesian retrieval routine (Sect. 2.2). We describe the assumptions that our model takes into account and the main potential source of systematic errors in the retrievals in Sect. 2.3.

2.1. Input spectra and scenarios

We considered spectra corresponding to four different evolutionary epochs of Earth: the prebiotic Earth (3.9 Ga), the Earth shortly after the GOE (2.0 Ga), the NOE (0.8 Ga), and modern Earth. All considered Earth spectra were produced by Rugheimer & Kaltenegger (2018). These self-consistent spectra were produced using a one-dimensional convective–radiative transfer model loosely coupled with a one-dimensional climate model and a one-dimensional photochemistry model. The authors accounted for the thermal chemistry and photochemistry of more than 55 species. The atmospheres were modeled up to 10^{-4} bar and split into 100 layers. The radiative forcing of clouds was included by adjusting the surface albedo of the planet.

The results of the photochemistry-climate-radiative model were then fed to a line-by-line radiative transfer model to produce emission spectra. The line lists and the pressure broadening coefficients were from the HITRAN 2016 database (Gordon et al. 2017). Surface scattering was included in the calculations,

---

[2] Present atmospheric level.

Table 1. Model description, identifiers, and colors.

Identifier	Color	Model description
MOD-CF		Modern Earth, cloud-free sky
MOD-C		Modern Earth, cloudy sky
NOE-CF		Neoproterozoic Oxygenation Event Earth, cloud-free sky
NOE-C		Neoproterozoic Oxygenation Event Earth, cloudy sky
GOE-CF		Great Oxygenation Event Earth, cloud-free sky
GOE-C		Great Oxygenation Event Earth, cloudy sky
PRE-CF		Prebiotic Earth, cloud-free sky
PRE-C		Prebiotic Earth, cloudy sky

Notes. The identifier "CF" denotes "cloud-free" (meaning "clear sky"). The identifier "C" denotes "cloudy." See main text for details.

assuming 70% ocean, 2% coast, and 28% land. In some scenarios, a partial cloud coverage was directly included in the calculation of the emission spectrum. In these cloudy cases, the authors assumed a 60% cloud coverage (split into 40% water clouds at 1 km altitude, 40% water clouds at 6 km altitude, and 20% ice clouds at 12 km altitude) consistent with an averaged Earth cloud model. Aerosol was not included in the calculation. The model has been validated from the visual to the mid-infrared (MIR) wavelength ranges with observations of Earth (Kaltenegger et al. 2007; Kaltenegger & Traub 2009; Rugheimer et al. 2013). For further details, we refer the reader to Rugheimer & Kaltenegger (2018).

For each epoch, we considered both clear sky and cloudy sky spectra, which yields a total of eight scenarios. We assigned every modeled scenario an identifier and a specific color, as listed in Table 1. We use these identifiers throughout the remainder of the paper.

We simulated observations with LIFE via the LIFEsim tool (see Paper II for a description of the simulator). LIFEsim estimates the wavelength-dependent S/N considering all major astrophysical noise sources (stellar leakage, local zodiacal dust emission, and exo-zodiacal dust emission). We considered an Earth-sized planet on a 1 AU orbit around a Sun-like star at a 10 pc distance. For our baseline analyses, we assumed the nominal simulation parameters for LIFEsim as summarized in Table 2 (see Paper I and Paper II for details). We considered an exozodi level of three times the local zodiacal dust density, based on the results from the HOSTS survey (Ertel et al. 2020). Similarly to Paper III, we assumed that the noise does not impact the flux. Rather, the noise calculated by LIFEsim represents the uncertainty of each spectral point. This might lead to optimistic results. However, running retrievals on non-randomized spectra can still provide useful information, since they approximate the average retrieval behavior on randomized spectra (see Appendix C of Paper III for a detailed discussion).

2.2. Updates on the Bayesian retrieval framework

We denote the input spectra from Sect. 2.1 as the "true spectra." To simulate a LIFE-like observation of these targets, we ran LIFEsim on the true spectra, thus obtaining simulated "observed spectra." We then performed a retrieval on the observed spectra using petitRADTRANS (Mollière et al. 2019) as "forward model" in the retrieval routine, and the Bayesian sampler model pyMultiNest (Buchner et al. 2014) as "parameter estimation routine" (cf. Paper III).

The theoretical one-dimensional atmospheric model petitRADTRANS (Mollière et al. 2019) applies the radiative transfer equation to calculate spectra corresponding to a set

Table 2. Simulation parameters used in LIFEsim for the baseline analyses.

Parameter	Value
Detector quantum efficiency	0.7
Total instrument throughput	0.05
Minimum wavelength	4 μm
Maximum wavelength	18.5 μm
Spectral resolution	50
S/N [a]	10
Interferometric baseline	10–100 m
Aperture diameter	2 m
Exozodi level	$3 \times$ local zodiacal dust
Planet radius	$1\,R_\oplus$
Distance to the system	10 pc

Notes. See Paper I and Paper II for details. [a]This S/N is fixed at a wavelength of 11.2 μm and the S/N of all other spectral bins is computed via LIFEsim.

of parameters. These parameters describe the bulk parameters (planetary mass and radius), the pressure–temperature (P–T) structure (approximated by a fourth-order polynomial), and the chemical composition of the atmosphere.

Our Bayesian retrieval framework recursively draws combinations of parameters from a set of "priors" that describe the "a priori" probability distribution of each parameter (listed in Table 3) and uses the forward model to compute the corresponding spectra. Then, the Bayesian framework tests how well these calculated spectra fit the observed one using a "likelihood" function (see Eq. (3) in Paper III). In order to sample the prior space efficiently, our retrieval relies on the parameter estimation routine pyMultiNest (Buchner et al. 2014), which is based on MultiNest (Feroz et al. 2009). This routine applies the Nested Sampling algorithm (Skilling 2006) to fit the theoretical spectral model to the observed spectrum and thereby yields estimates and uncertainties for the model parameters. These estimates are the "posterior probability distributions" (or "posteriors"). The posteriors contain the information on which combinations of model parameters best describe the observed spectrum. For more details about the Bayesian retrieval framework, we refer the reader to Paper III.

In our previous work, we argue that the effects of scattering on simulated MIR spectra are negligible at the considered resolutions and LIFEsim noise patterns. In that study, we had to be particularly mindful of the computing time. The version of petitRADTRANS used in Paper III only allowed spectra at

Table 3. Summary of the parameters used in the retrievals, their expected values, and their prior distributions.

Parameter	Description	Prior	Expected values (1)			
			Modern	NOE	GOE	Prebiotic
$\sqrt[4]{a_4}$	P–T parameter (degree 4)	$\mathcal{U}(0.5, 1.8)$	0.694	1.117	0.999	1.189
a_3	P–T parameter (degree 3)	$\mathcal{U}(0, 100)$	18.644	17.562	12.886	19.430
a_2	P–T parameter (degree 2)	$\mathcal{U}(0, 500)$	110.331	72.414	59.984	71.778
a_1	P–T parameter (degree 1)	$\mathcal{U}(0, 500)$	177.997	140.294	123.764	120.546
a_0	P–T parameter (degree 0)	$\mathcal{U}(0, 1000)$	293.601	307.747	290.306	280.615
$\log_{10}(P_0 \text{ [bar]})$	Surface pressure	$\mathcal{U}(-4, 3)$	0.007	0.017	0.008	0.008
$R_{pl}\ [R_\oplus]$	Planet radius (bulk value)	$\mathcal{G}(1.0, 0.2)$	1.000	1.000	1.000	1.000
$\log_{10}\left(M_{pl}\ [M_\oplus]\right)$	Planet mass (bulk value)	$\mathcal{G}(0.0, 0.4)$	0.000	0.000	0.000	0.000
$\log_{10}(N_2)$	N_2 mass fraction	$\mathcal{U}(-15, 0)$	–0.113	–0.016	–0.005	–0.073
$\log_{10}(O_2)$	O_2 mass fraction	$\mathcal{U}(-15, 0)$	–0.631	–1.623	–2.622	–6.002
$\log_{10}(H_2O)$	H_2O mass fraction	$\mathcal{U}(-15, 0)$	–2.607	–2.143	–2.498	–2.345
$\log_{10}(CO_2)$	CO_2 mass fraction	$\mathcal{U}(-15, 0)$	–3.265	–1.807	–1.806	–0.827
$\log_{10}(CH_4)$	CH_4 mass fraction	$\mathcal{U}(-15, 0)$	–6.028	-3.635	–3.035	–6.056
$\log_{10}(O_3)$	O_3 mass fraction	$\mathcal{U}(-15, 0)$	–6.436	–6.567	–7.412	–10.301
$\log_{10}(CO)$	CO mass fraction	$\mathcal{U}(-15, 0)$	–7.215	–6.072	–7.206	–4.915
$\log_{10}(N_2O)$	N_2O mass fraction	$\mathcal{U}(-15, 0)$	–6.343	–6.859	–7.962	–

Notes. $\mathcal{U}(x, y)$ denotes a boxcar prior with a lower threshold x and upper threshold y; $\mathcal{G}(\mu, \sigma)$ represents a Gaussian prior with mean μ and standard deviation σ. For a_4 we chose a prior on $\sqrt[4]{a_4}$, which allows us to sample small values more densely, typical of a fourth-order coefficient, and then take the fourth power to obtain a_4. The expected values of the P–T parameters are generated by fitting the input P–T profiles (calculated by Rugheimer & Kaltenegger 2018) with a fourth-degree polynomial. The expected values of the abundances are calculated considering the weighted mean by the pressure of the input abundance profiles (calculated by Rugheimer & Kaltenegger 2018).
References. (1) Rugheimer & Kaltenegger (2018).

$R = 1000$ to be calculated, which were subsequently binned down to the resolution of the input spectrum ($R = 35$–100 in Paper III). Calculating a spectrum at $R = 1000$ excluding scattering required ≈ 0.5 seconds, whereas including scattering required ≈ 18 seconds. Since millions of spectra have to be calculated for a single Bayesian retrieval, including scattering was prohibitive with respect to the computing time, especially when considering a large grid of retrieval runs. Recent updates to `petitRADTRANS` have enabled us to compute spectra at any resolution, provided a grid of correlated-k tables at that resolution is available. These tables can be produced with `petitRADTRANS` by binning down the $R = 1000$ opacity tables. We compiled a correlated-k opacity database for $R = 50$ and used it to produce spectra directly at this resolution. This reduces the computation time per spectrum significantly and allows us to compute emission spectra in ≈ 0.04 seconds when excluding scattering, and in ≈ 0.5 seconds when including scattering. This reduction in computing time allowed us to include scattering in the theoretical spectral model.

Other updates on `petitRADTRANS` were performed. The treatment of collision-induced absorption (CIA) was modified by updating the interpolation of the CIA tables, originally in FORTRAN, to Python. Minor variations in the interpolation options (log-linear compared to nearest neighbor in Paper III) make the new model not directly comparable with its previous version. However, under Earth-like conditions, the differences in the CIA signature on the spectrum remain negligible. Also, the CIA features impact mostly the short wavelengths ($\lambda < 6\ \mu m$), where the LIFESIM noise is large. Furthermore, the treatment of scattering was updated. Originally implemented in `petitRADTRANS` only for gaseous planets (see Mollière et al. 2019), it was adapted to include the scattering by a rocky surface. More information on the new implementation of scattering can be found in Appendix A, as well as the `petitRADTRANS` documentation[3].

This new version of `petitRADTRANS` is available in the main GitLab repository[4].

Using the updated retrieval framework, we ran retrievals for the eight different spectra introduced in Sect. 2.1. We retrieved the same parameters as in Paper III leaving most prior distributions unchanged. An exhaustive list of the parameters and priors used, and the corresponding expected values for the different epochs, are listed in Table 3. Most priors are represented by a boxcar function (hereafter "uniform" priors): every value is equally probable if within a certain range. Two notable exceptions are the priors for the planetary radius R_{pl} and the logarithm of the planetary mass $\log_{10}(M_{pl})$, whose priors are Gaussian. The prior we assumed on R_{pl} is based on the radius estimate expected from observing a terrestrial planet with LIFE during its search phase (see Paper II for details). The $\log_{10}(M_{pl})$ prior was inferred from R_{pl} using the statistical mass-radius relation presented in Chen & Kipping (2016; see Paper III for details).

2.3. Assumptions and discrepancies

When performing retrievals, we are limited by the maximum number of parameters retrieved in a reasonable computing time. For this reason, we need to make a few simplifications. First, as in Paper III, we parameterized the P–T profile in the retrieval by a fourth-order polynomial. Second, we assumed the abundances of the considered species to be independent of altitude. Third, our retrieval framework did not model clouds for any of the considered scenarios. This is a strong simplification for the cases where we retrieve cloudy input spectra. However, this retrieval approach allows us to investigate the biases in the results obtained when retrieving a cloudy spectrum assuming a cloud-free atmosphere. The addition of a cloud model to our retrieval framework will be tackled in a future study. Finally, a surface

---

[3] `https://petitradtrans.readthedocs.io`

[4] `https://gitlab.com/mauricemolli/petitRADTRANS`

Table 4. References for the molecular opacities used in the retrievals.

Species	Line list	Pressure broadening	Line cutoff
CO_2	HITEMP (1)	γ_{air}	(6)
O_3	HITRAN 2012 (2)	γ_{air}	(7)
CH_4	ExoMol 2014 (3)	Eq. (15), (5)	(7)
CO	HITEMP (1)	γ_{air}	(7)
H_2O	HITEMP (1)	γ_{air}	(7)
N_2O	ExoMol 2021 (4)	γ_H, γ_{He}	Eq. (6), (8)

Notes. Adapted from Mollière et al. (2019).
References. (1) Rothman et al. (2010); (2) Rothman et al. (2013); (3) Yurchenko & Tennyson (2014); (4) Chubb et al. (2021); (5) Sharp & Burrows (2007); (6) Burch et al. (1969); (7) Hartmann et al. (2002); (8) Chubb et al. (2021).

Table 5. References for the CIA and Rayleigh opacities used in the retrievals.

CIA	References	Rayleigh	References
N_2-N_2	HITRAN (1)	N_2	(2,3)
O_2-O_2	HITRAN (1)	O_2	(2,3)
O_2-N_2	HITRAN (1)	CO_2	(4)
CO_2-CO_2	HITRAN (1)	H_2O	(5)
		CH_4	(4)
		CO	(4)

Notes. Adapted from Mollière et al. (2019).
References. (1) Karman et al. (2019); (2) Thalman et al. (2014); (3) Thalman et al. (2017); (4) Sneep & Ubachs (2005); (5) Harvey et al. (1998).

reflectance of 0.1 was assumed for all the wavelength range. This is a common value for water-rich habitable terrestrial planets, dominated by the low infrared reflectance of oceans and ice.

In contrast, Rugheimer & Kaltenegger (2018) used P–T profiles that were self-consistently calculated by their climate-photochemistry model to generate the input spectra. The chemical abundance profiles were also altitude-dependent and calculated self-consistently by the climate-photochemistry model. In order to compare the results of the retrievals with the input, we approximated the input P–T and abundance profiles to calculate the expected values of the parameters considered in the retrievals. Regarding the thermal structure of the atmosphere, we determined the expected values of the polynomial coefficients (a_4, a_3, a_2, a_1, a_0) listed in Table 3 by fitting a fourth-order polynomial to the self-consistent P–T profiles. As for the abundances, we assumed the weighted means over the pressure grid of the altitude-dependent abundance profiles. This is sensible, since the denser layers of the atmosphere (which corresponds to higher pressures) are generally the ones that contribute more to the spectrum. The expected values for the considered species are listed in Table 3 as well. For what concerns the surface reflectance, Rugheimer & Kaltenegger (2018) considered a wavelength-dependent reflectance. However, the reflectance still averaged 0.1 in the wavelength range of interest. We would therefore not expect large variations due to this parameter.

There are also differences between the opacity tables that the two models use. We used the default set of opacities for `petitRADTRANS` as presented in Mollière et al. (2019). We added the N_2O opacity from the ExoMol database (Chubb et al. 2021). The CIA opacities were taken from the HITRAN database. Details and reference papers corresponding to the opacity line lists are shown in Tables 4 and 5. In contrast, the input spectra were calculated using HITRAN 2016 opacities (Rugheimer & Kaltenegger 2018). Differences in line lists and broadening coefficients are therefore to be expected and may cause biases in the results.

Due to these differences in the atmospheric models, we would imagine finding some small discrepancies between the spectra that were published in Rugheimer & Kaltenegger (2018) and the ones that our framework can calculate. We discuss this particular aspect in Sect. 4.4.

3. Results

In this section, we show the results from the retrievals on the grid of different input spectra (see Table 1) assuming the baseline

parameters listed in Table 2. We start by analyzing the retrieved spectra (Sect. 3.1) to offer a broad overview of the retrieval performance. Then, we study the retrieved P–T profiles (Sect. 3.2), the planetary parameters (Sect. 3.3) and abundances (Sect. 3.4). We also ran additional retrievals of the same scenarios (shown in Table 1) by varying R and S/N. We compare the results of these retrievals in Sect. 3.5.

3.1. Retrieved emission spectra

The main outputs of the Bayesian retrieval framework are the posterior distributions of the parameters, necessary to produce theoretical spectra, that best match the data. The posteriors can be visualized as an N-dimensional space that is a subset of the larger N-dimensional prior space, N being the number of parameters. Each point included in the posterior space has N coordinates and represents a combination of N parameters that, if fed to the theoretical spectral model, would produce a spectrum that was determined by the Bayesian framework to resemble the observed spectrum.

From the available sets of parameters within the posteriors that the routine has calculated, we can therefore produce "retrieved spectra." These are shown in Fig. 1. Each subplot presents the results for a specific model, compared to the input spectrum binned down to $R = 50$. Similarly, Fig. 2 shows the logarithm of the ratio between the retrieved emission spectrum and the input emission spectrum for each scenario.

The retrieved spectra are generally in good agreement with the input spectra (within 1σ) for all considered cases. This shows that our retrieval framework can reproduce the simulated input spectra, regardless of the complexity of the input model (in terms of thermal and abundance profiles, and cloud coverage). However, we notice regions with larger uncertainties, especially at wavelengths shorter than $\approx 8\,\mu m$. Here, the ratio between the retrieved spectra and the input spectrum (as shown in Fig. 2) reaches up to a few orders of magnitude. Such differences are, however, still within the noise uncertainty (gray-shaded areas). Smaller differences can be noticed in the main CO_2 band at $\approx 15\,\mu m$. Since the noise is not as high as it is in the short wavelength range, these differences are probably due to discrepancies in the opacity tables (see Sect. 4.4).

The parameter estimation routine included in the Bayesian retrieval framework has the task of minimizing the difference between model output and data. For this, no detailed parameterization of the relevant physical processes is required. Many of the relevant physical and chemical parameters are correlated (e.g.,
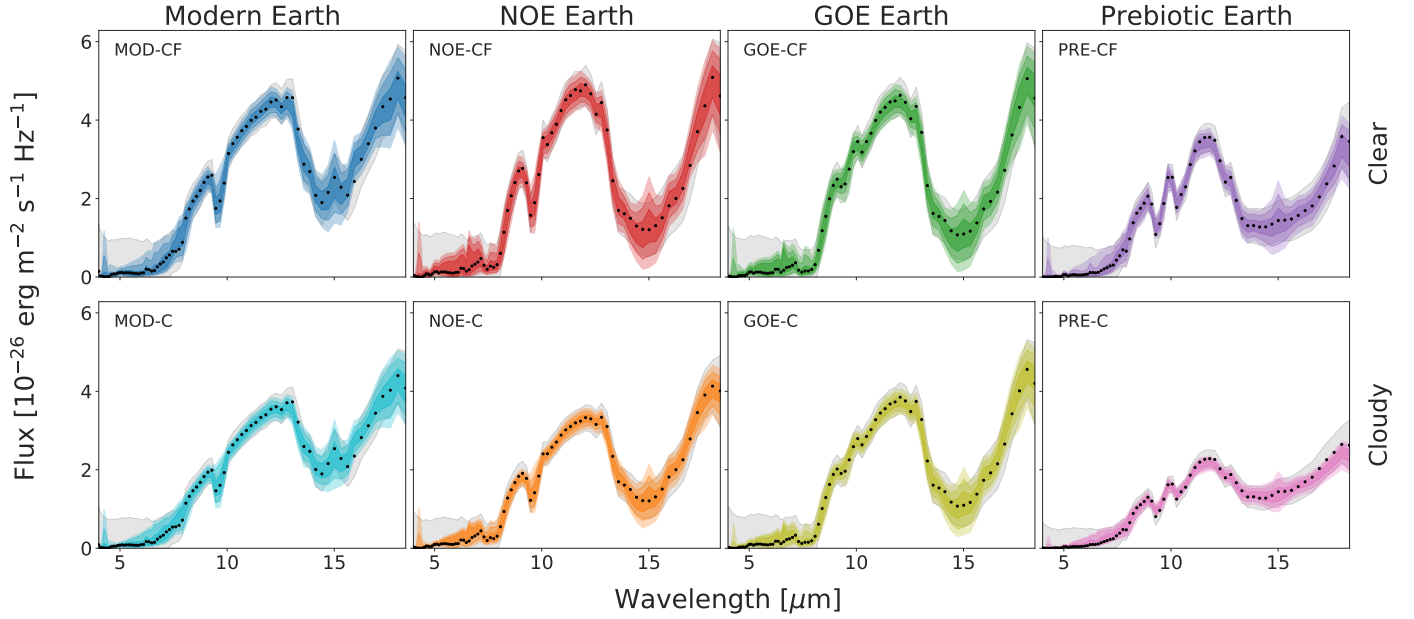
Fig. 1. Retrieved spectra compared to the input spectra (black dots) for the various scenarios, ordered by epoch (*columns*) and cloud coverage (*rows*). The gray-shaded area indicates the LIFEsim uncertainty. The color-shaded areas represent the confidence envelopes (darker shading corresponds to a higher confidence). The scenarios are color-coded according to Table 1.

Fig. 2. Ratios between the retrieved flux and the input flux (in logarithmic scale) for the various scenarios, ordered by epoch (*columns*) and cloud coverage (*rows*). The gray-shaded area indicates the LIFEsim uncertainty. The color-shaded areas represent the confidence envelopes (darker shading corresponds to a higher confidence). The scenarios are color-coded according to Table 1.
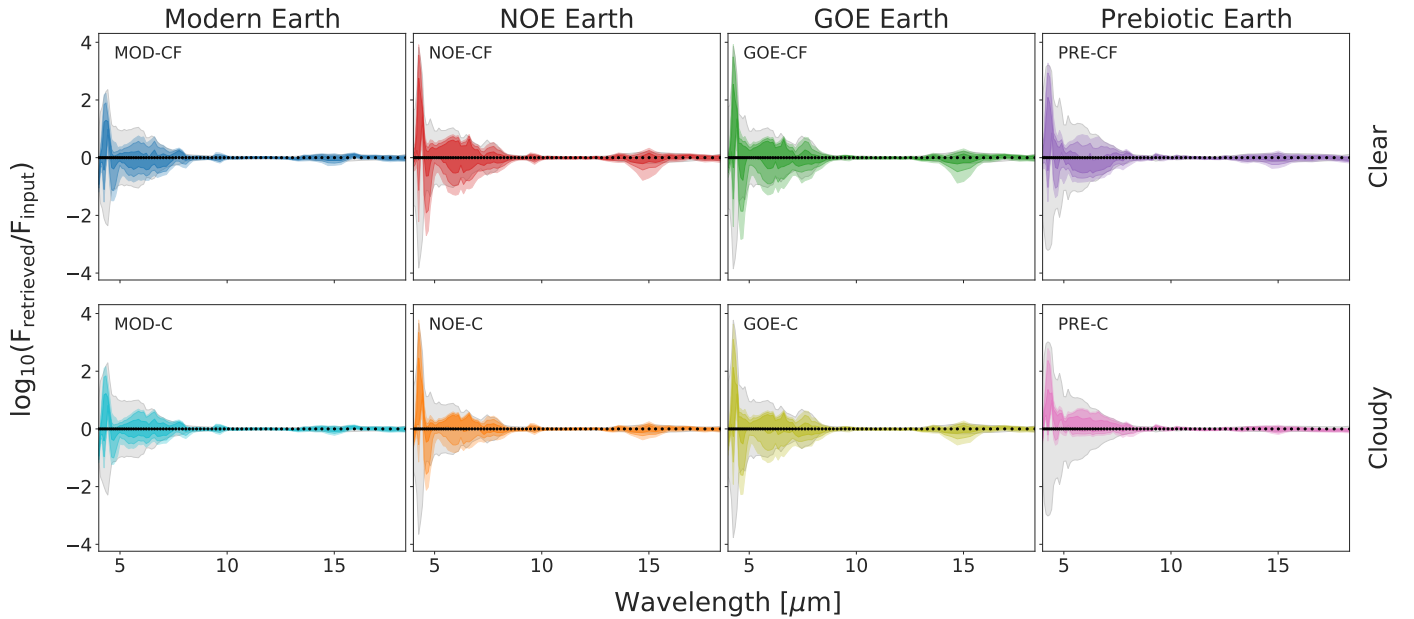
the planetary mass and the pressure; the pressure and the chemical abundances), often in a nonlinear way. It is therefore possible for the parameter estimation routine to produce similar spectra as the full (physical) input model over a diverse set of parameters, as a result of such correlations. Hence, it is appropriate to question whether the parameter estimation routine results are on the one hand physically representative, or whether degeneracies and systematics between the retrieved parameters could be influencing the results. The next sections will explore these issues in more detail.

3.2. Retrieved P–T profiles

In Fig. 3, we show the retrieved P–T profiles compared to the input profiles for all combinations of the four epochs (columns) and the two cloud coverages (rows).

The vertical shape of the retrieved P–T profiles in the lower atmosphere (pressures $\geq 10^{-2}$ bar) roughly follows that of the true P–T profiles. In most cases, the true profiles are contained within the 1σ uncertainty envelope. As in Paper III, the uncertainties grow larger at higher altitudes (pressures $\leq 10^{-2}$ bar). This indicates that, for the quality of the input spectra we consider for this
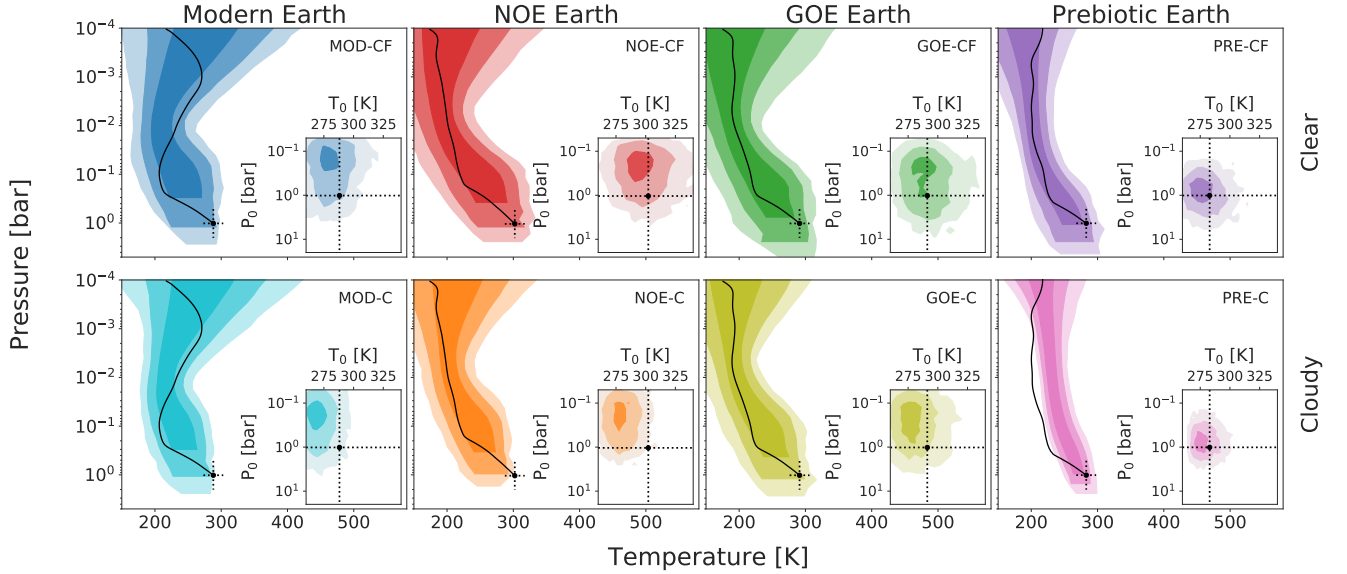
Fig. 3. Retrieved $P-T$ profiles compared to the input profiles (solid black line) for the various scenarios, ordered by epoch (*columns*) and cloud coverage (*rows*). The scenarios are color-coded according to Table 1. In each subplot, we also show an inset plot with the two-dimensional histogram of the retrieved surface $P-T$ values. The 1σ, 2σ, and 3σ confidence levels in the $P-T$ profiles and the two-dimensional histograms are indicated by the increasing intensity of the color fill (darker shading corresponds to a higher confidence).

study, it is not possible to distinguish atmospheres with a stratospheric temperature inversion (i.e., the modern Earth scenario) from those with an isothermal stratosphere (i.e., the NOE, GOE, and prebiotic scenarios). This retrieval limitation is a result of the small overall contribution of the upper atmospheric layers to the planet's MIR emission spectrum.

A few additional inconsistencies between the retrieved and input $P-T$ profiles are also apparent in the lower altitudes (high pressures). A general feature in the three biotic epochs (modern, NOE, and GOE Earth) is the retrieval of underestimated values for the ground pressure P_0 (~0.1 bar as opposed to the true value of ~1 bar). This occurs for both the cloud-free and cloudy spectra. Such an offset could be explained by systematic differences between the radiative transfer models used to produce and retrieve the simulated spectra. We discuss this in more detail in Sect. 4.4. The ground temperatures T_0 are on average well retrieved for all clear sky scenarios. In contrast, the retrievals performed for the cloudy spectra systematically underestimate T_0, with differences between the retrieved and true value $\lesssim 25$ K. These results have an impact on assessing the habitability of the simulated exoplanets, which will be discussed in more detail in Sect. 4.

For the prebiotic Earth input spectra, the retrievals provide estimates for P_0 and T_0 that are in agreement with the true parameter values. Furthermore, the overall uncertainties on the retrieved $P-T$ profile are generally smaller than for the other epochs. However, for the cloudy prebiotic Earth (PRE-C) spectrum, the retrieved $P-T$ profile is a few tens of kelvin warmer than the true value in the intermediate layers of the atmosphere (~10^{-1} to ~10^{-3} bar). This effect is likely related to the much weaker emission features in the PRE-C scenario compared to the other epochs. We discuss the impact of neglecting clouds in the retrievals in Sect. 4.2.

3.3. Retrieved planetary parameters

Figure 4 shows the posterior density distributions we retrieve for the planetary parameters (R_{pl}, M_{pl}) and surface conditions (P_0, T_0) for all considered input spectra. R_{pl}, M_{pl}, and P_0 are

directly retrieved by our framework, while the T_0 posterior is calculated from the $P-T$ parameters (polynomial coefficients a_i) and P_0 posteriors. The results are color-coded based on Table 1 and grouped by epoch (rows) and planetary parameters (column). The retrieved parameters for the clear and cloudy input spectrum of the same epoch are shown in the same subplot to facilitate comparison.

We obtain good estimates for all of the parameters considered, especially for the cloud-free scenarios. The posterior distributions roughly follow a Gaussian distribution and are typically centered on the true values.

By comparing the retrieved posteriors for R_{pl} to the corresponding Gaussian prior range (shown in Fig. 4 as a dotted line), we observe that we manage to better constrain the planet radius R_{pl} with respect to the prior distribution. For the GOE and modern Earth scenario, the retrieved posteriors do not significantly depend on the cloud coverage. For the NOE and prebiotic input we underestimate R_{pl} in the retrievals of the cloudy spectra (retrieved as ~0.8–0.9 $R_\oplus$ instead of 1 $R_\oplus$). This difference is more pronounced in the prebiotic scenario.

For what concerns M_{pl}, the retrieval analysis does not add further constraints on the estimates for the planet's mass. Also, there is no noticeable difference in the retrieval results for M_{pl} between the four epochs. This finding holds for both the clear and cloudy scenarios and is in agreement with the results we presented in Paper III. It is important to note that, in contrast to the prior assumption (see Sect. 2.2), M_{pl} is not linked to R_{pl} through a mass-radius relationship during the retrieval, but it is instead a free parameter. This means that both M_{pl} and R_{pl} are independently drawn from their respective prior. In the retrieval, we use both parameters to calculate the planet's surface gravity, which is required to compute the theoretical emission spectrum. In addition to the surface gravity calculation, we also use R_{pl} to scale the flux emitted per unit area at the top of the atmosphere (as calculated by `petitRADTRANS`) to the observed exoplanet flux at a distance d from the observer (generally well known, 10 pc in our study). We do so by multiplying the flux at the top of the atmosphere by the factor $(R_{pl}/d)^2$.
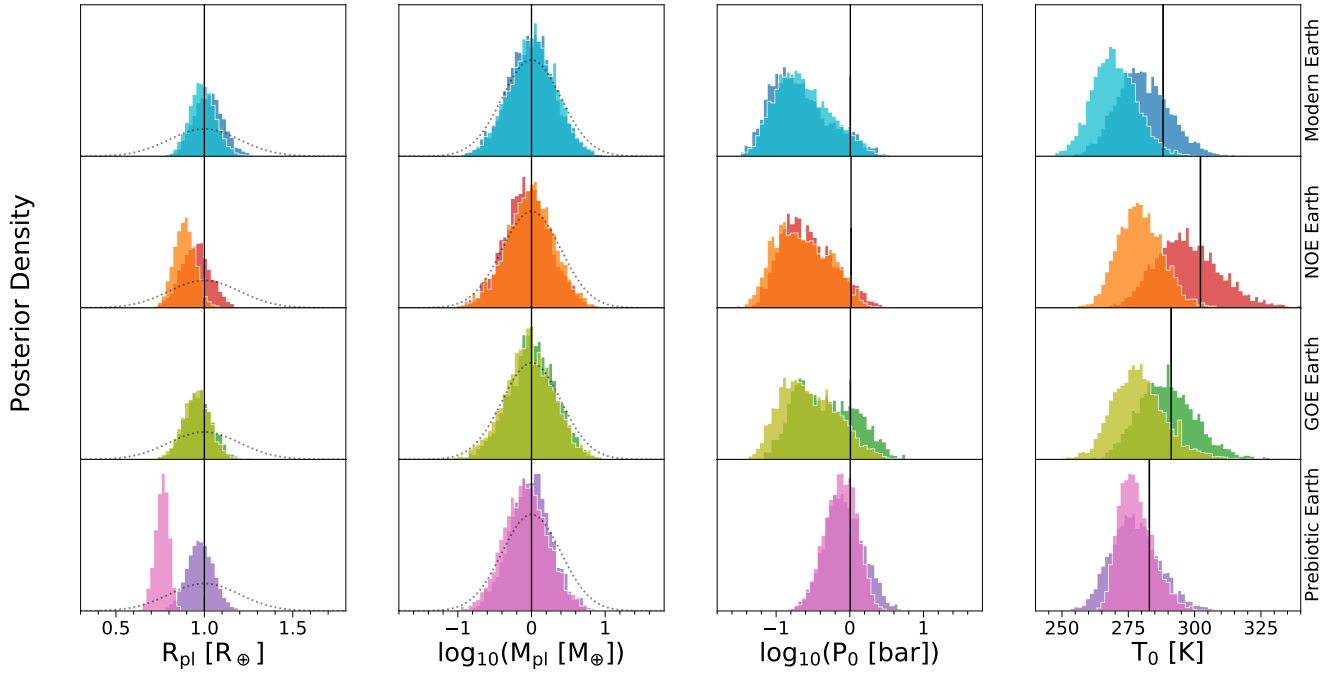
Fig. 4. Posterior density distributions for the retrieved exoplanet parameters (*columns*) for the different epochs (*rows*) and cloud coverages. We follow the color-coding listed in Table 1 to differentiate the different scenarios. The vertical, solid lines mark the true values for each parameter. The dotted lines in the R_{pl} and M_{pl} plots indicate the assumed Gaussian priors. For P_0 and T_0 we assume broad, flat priors, which are not plotted.

Since the surface gravity is typically not directly constrainable due to the gravity-abundance degeneracy (see Sect. 3.5), the retrieval struggles to further constrain M_{pl}. In contrast, the retrieval does manage to constrain R_{pl} further, as it does not only depend on the surface gravity, but also on the distance-scaling of the spectrum.

The retrieved posteriors for the surface pressure P_0 are significantly smaller than the assumed prior distribution (10^{-4} to 10^3 bar), meaning that we manage to pose strong constraints on P_0 with respect to the assumed prior knowledge on the parameter. However, the retrieval tends to underestimate the value of P_0 in all cases except for the prebiotic one. The retrieved posteriors in these cases are not well represented by a Gaussian, which indicates that the retrieval results for P_0 could be degenerate. We discuss this in more detail in Sect. 4.3.

As shown in Sect. 3.2, the retrieved posteriors for the surface temperature T_0 are centered on the true values for the cloud-free retrievals. For the cloudy input spectra, the retrievals tend to underestimate the surface temperature. The standard deviation of the retrieved T_0 posteriors is roughly ±20 K for all biotic scenarios. This is in agreement with the findings made in Paper III. The spread of retrieved posteriors could potentially be reduced by increasing the R or S/N of the input spectra (see Sect. 3.5). Observing strategies for the trade-off between R and S/N will be addressed in Sect. 4.3.

3.4. Retrieved chemical abundance parameters

Figure 5 shows the retrieved posterior distributions for the main atmospheric gases. We again arrange the various scenarios by epoch (row), and atmospheric species (column) and use the color-coding from Table 1. The results for the clear and the cloudy retrievals of one epoch are shown in the same subplot to facilitate comparison.

We plot our expected abundances (listed in Table 3), which are the weighted means (with respect to the pressure) of the

original abundance profiles, as black vertical lines. If no true value is plotted, the molecule is not present in the input spectrum. We further indicate the range of variability of the true, pressure-dependent abundance profiles (minimum to maximum) via the shaded gray area in each subplot.

We adopt the same posterior classification scheme that was introduced in Paper III, for an easier comparison of the results. This scheme divides the retrieved posteriors into the following four classes. The first is the constrained (C) class: the posterior is best described by a Gaussian distribution. This implies that abundances both significantly lower and higher than the true value can be ruled out. The second is the sensitivity limit (SL): the abundance is at the retrieval's detection limit for the species. The posterior exhibits a distinct peak. However, low abundances are not ruled out. The posterior is best described by the convolution of a soft-step function with a Gaussian. The third is the upper limit (UL): the posterior resembles a soft-step function. Large abundances can be excluded, low ones cannot. The last is unconstrained (UC): we cannot retrieve information on the atmospheric abundance. The posterior resembles a constant function over the full prior range. For further details on the specifics of the posterior classification, we refer the reader to Appendix B of Paper III.

We obtain UC posteriors for the abundances of N_2 and O_2 in all retrievals performed. In accordance with the findings presented in Paper III, these molecules are not detectable in any of the considered scenarios. This finding indicates that the corresponding CIA spectral signatures are too weak to be detectable in the considered input spectra with $R = 50$ and S/N = 10. To increase readability, we choose not to show the retrieval results for N_2 and O_2 in Fig. 5. The posterior distributions of these molecules can however be found in the corner plots in Appendix B. Similarly, the trace gases N_2O and CO are not detected in any of our retrievals, obtaining unconstrained posteriors for all epochs. The MIR absorption features of these molecules at the considered abundances are also too weak in the
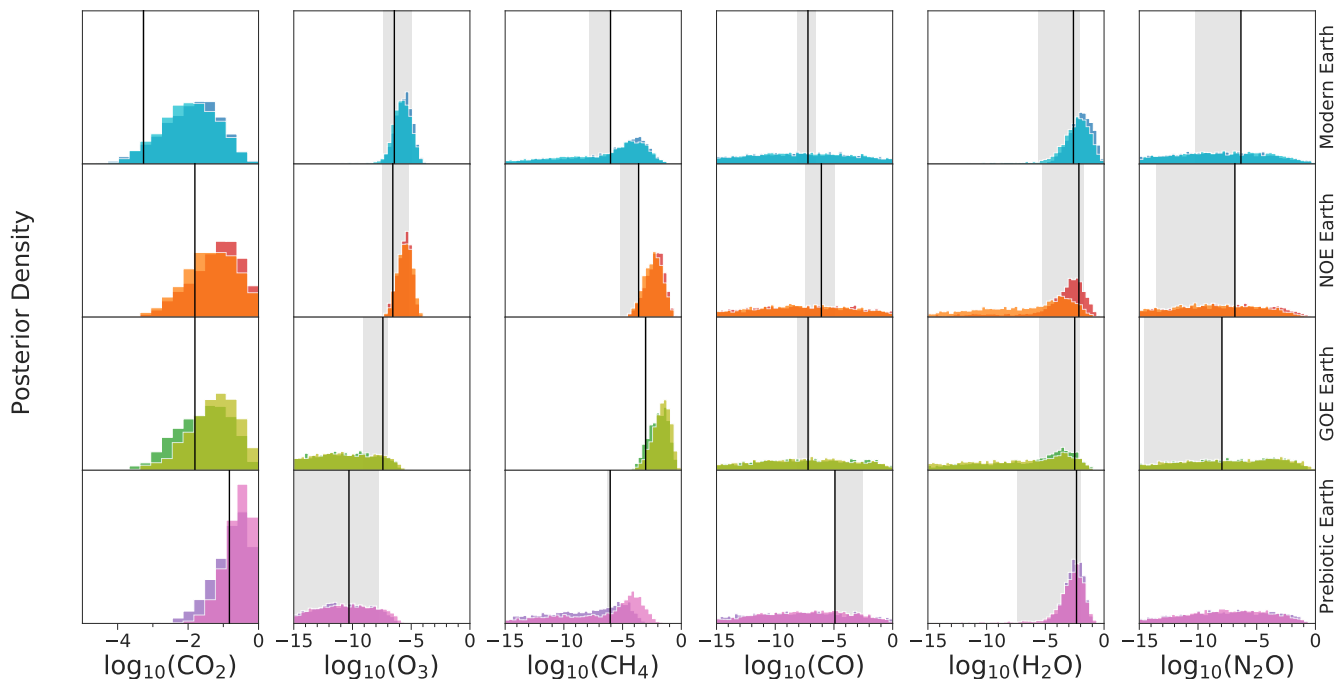
Fig. 5. Posterior density distributions for the retrieved species (*columns*), the different epochs (*rows*), and the different cloud coverage scenarios. Results from the various scenarios use the color-coding from Table 1. The solid black lines indicate the expected values for each species, which vary depending on the epoch. The gray-shaded area marks the range of values in the vertically nonconstant abundance profiles, which were used to compute the input spectra.

considered input spectra to be constrained in our retrievals. This can be seen by the flat posterior distributions for both species in all considered cases (see Fig. 5).

We detect CO_2 in all retrievals and the received posterior distributions are generally Gaussian-like (C-type posteriors). Our results suggest that the median abundances of the different posterior distributions are higher than the true value for all the epochs. However, in the prebiotic, GOE, and NOE scenarios the true abundances still lie within the 1σ envelope of the retrieved abundances. For the modern Earth scenarios, the true value lies within the 3σ range of the retrieved posterior. This is consistent with a "compensation effect" whereby the retrieval framework is correcting for the underestimated pressure. The degeneracy between chemical composition and atmospheric pressure is well known and it was already encountered in Paper III (see Sect. 4.1). All retrieved CO_2 posteriors span about three orders of magnitude (3 dex). They all appear very similar even though the expected values of CO_2 span from roughly 0.01% (modern Earth) to the order of 10% (prebiotic Earth). This forbids the use of CO_2, one of the major absorbers in the atmosphere, as a discriminator between the considered epochs. To reduce the variance in the retrieved abundances, an increase in R and/or S/N might be recommended (see Sect. 3.5).

For the remaining species (O_3, CH_4, and H_2O), the retrieval results depend on the considered epoch: O_3 is retrieved accurately in both the modern Earth and NOE Earth scenarios (C-type posteriors). These are the two cases where O_3 is more abundant ($\sim 10^{-6}$ in mass fraction). In contrast, for the prebiotic and GOE scenario, we only retrieve upper limits (UL-type posteriors) for the O_3 abundance. This means that we can rule out high abundances of O_3 ($\gtrsim 10^{-6}$ in mass fraction), but cannot exclude abundances below the retrieved upper limit.

We manage to detect CH_4 in the NOE and GOE Earth spectra (C-type posterior), which have a higher CH_4 abundance. For the

modern and prebiotic Earth spectra, where the abundances are lower, we retrieve an SL-type posterior, which is characterized by a peak in the distribution roughly at the true value and a non-negligible tail toward low abundances. Similar to CO_2, CH_4 is generally overestimated when detected. However, the true value still lies within the 2σ envelope of the retrieved posterior distribution. The conditional retrieval of O_3 and CH_4 is particularly significant when discussing the detectability of biosignatures in Earth-like planets with LIFE, which will be discussed in Sect. 4.5.

H_2O is constrained in both the modern and prebiotic scenarios, as well as the clear NOE Earth (NOE-CF) model, and the retrieved posteriors are centered on the true value. In contrast, we only detect SL-type posteriors for both the GOE Earth spectra and the cloudy NOE model. In this case, such a difference in retrieval performance cannot be explained by a difference in the abundance of the species, which is fairly constant throughout all the epochs (around $\sim 10^{-2}$ in mass fraction). This is likely to be related to the overestimation of CH_4, a species that is much more abundant in the GOE and NOE scenarios (10^{-3} in mass fraction compared to 10^{-6} for the modern and prebiotic epochs). The spectral signature of H_2O overlaps with that of CH_4 between 5 and 7 μm, which is where the noise level is very high and the flux levels are low. Therefore, the retrieval framework favors a higher abundance of CH_4 at the expense of a larger uncertainty on H_2O in these scenarios.

Finally, our results suggest only modest differences in the retrieved abundances obtained from the clear and cloudy input spectra. The presence of clouds in the spectrum does not seem to deteriorate the abundance estimation capabilities for most molecules. In contrast, as seen in the previous subsection, the characterization of a cloudy atmosphere with a cloud-free model will likely result in biases in the retrieved physical parameters. We discuss this topic further in Sect. 4.2.
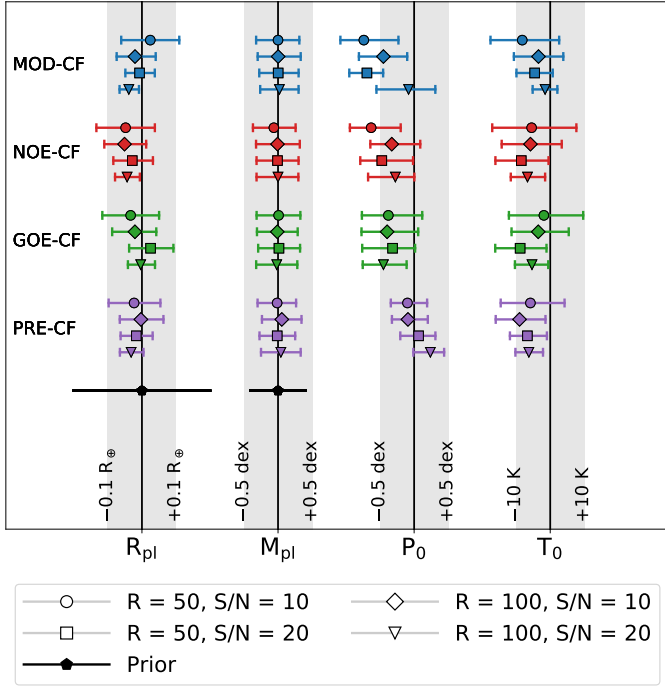
Fig. 6. Retrieved exoplanet parameters for the different scenarios with varying R and S/N values. The error bars denote the 68% confidence intervals. For M_{pl} and R_{pl}, we also plot the assumed prior distributions. For T_0 and P_0, we assumed flat, broad priors. The vertical lines mark the true parameter values.

3.5. Runs at higher resolution and/or signal-to-noise ratio

In this section, we investigate whether our retrieval results can be improved by increasing the quality (R and/or S/N) and thus the information content of the input spectra. We ran ancillary retrievals for the eight scenarios and chose the following combinations of R and S/N[5]: (1) $R = 50$ and S/N = 10 (the reference case); (2) $R = 100$ and S/N = 10; (3) $R = 50$ and S/N = 20; and (4) $R = 100$ and S/N = 20.

We provide a summary of the results obtained from these additional retrieval runs in Figs. 6 (planetary parameters) and 7 (abundances). Results from the different ancillary runs are represented using different markers. The results for the different epochs are color-coded according to Table 1. Here, we only show the results for the clear input spectra. The plots corresponding to the cloudy scenarios can be found in Appendix C.

We are particularly interested in significant increases in accuracy (i.e., the retrieved values agree better with the input "truth") or in precision (i.e., the posterior's variance is reduced). For higher S/N we expect more precise results since the uncertainty in the input spectrum is lower. This yields stronger constraints on the model parameters. An increase in R should allow for a more robust identification and characterization of the spectral features and thus more accurate retrieval results. Both increased accuracy and/or precision could allow us to differentiate between the different epochs (and generally between different planets) more clearly. This will be of great importance, especially when searching for signatures of life in exoplanetary atmospheres (see Sect. 4.5). We should point out that in our current simulation

setup, which ignores (systematic) instrumental noise terms, doubling R at a constant S/N means doubling the integration time, while doubling S/N at a constant R means, roughly, quadrupling it[6]. This information is crucial for the mission planning and will be further discussed in Sect. 4.3.

In Fig. 6, we notice that increasing the S/N to 20 while keeping $R = 50$ (square markers) generally results in a narrower posterior for R_{pl}. We observe a reduction in the variance of the R_{pl} posterior by up to a factor of 2 compared to the reference case ($R = 50$, S/N = 10; the circular markers). An increase in R ($R = 100$, S/N = 10; the diamond markers) causes the variance of the R_{pl} posterior to shrink to about 70% of the reference case variance. In contrast, we observe no noticeable gain in the accuracy of the retrieved value for R_{pl} when increasing S/N and R at the same time. On the other hand, the precision of the measurement at $R = 100$, S/N = 20 improves significantly, with the variance of the R_{pl} posterior shrinking up to three times compared to the reference case.

We further find that the retrieval of the planetary mass M_{pl} does not improve significantly when moving to higher R and S/N input spectra. We observe no significant increase in both accuracy and precision. This finding is consistent with the results shown in Paper III. The underlying reason for this observation is the degeneracy between the surface gravity (and thus also M_{pl}) and the abundances of trace gases (see, e.g., Paper III, Mollière et al. 2015; Feng et al. 2018; Madhusudhan 2018; Quanz et al. 2021). Since gravity and abundances are involved in the hydrostatic equilibrium, it is possible to reproduce the same spectral feature using different combinations of these parameters. This broadens the variance of the posteriors of M_{pl} and of the atmospheric species.

Increasing the quality of the input spectrum does improve the accuracy of the retrieval for P_0 in the clear modern Earth (MOD-CF) case. The results for the other epochs do not exhibit a similar trend with increasing input quality. This failure to retrieve accurate ground pressure estimates is likely rooted in differences between the opacity tables used by the retrieval framework and the ones used to calculate the input spectra (see Sect. 4.4 for more details). Additionally, no noticeable decrease in the variance of the retrieved P_0 estimate is present for higher values of R or S/N. This is likely a result of the pressure-abundance degeneracy, which has already been described in Sect. 3.4.

For the surface temperature T_0, we do not notice any substantial improvements in the accuracy of the retrieved values when increasing R or S/N. However, as for R_{pl}, we observe a significant reduction in the variance of the posteriors when increasing S/N and R. Compared to the reference case, the uncertainty in T_0 is reduced by a factor of 2 for the runs with S/N = 20 and to about 70% of the reference variance for the runs with $R = 100$. These improvements in temperature accuracy could be crucial when assessing the potential habitability of an observed exoplanet.

In Fig. 7, we summarize the retrieved posterior distributions in the abundances for the reference case ($R = 50$ and S/N = 10, circular markers) and all other R and S/N combinations. The abundance posteriors are classified according to our classification scheme (see Sect. 3.4 and Paper III).

Generally, we observe that increases in both S/N and R do not significantly improve the accuracy nor the precision of the retrieved posteriors for the majority of the scenarios. This is again the result of the pressure- and gravity-abundance

---

[5] We remind the reader that the S/N refers to the value at a reference wavelength of 11.2 μm.

[6] We refer the reader to the appendix of Paper I, where we show a breakdown of the typical noise contributions for planets detected around solar-type stars.
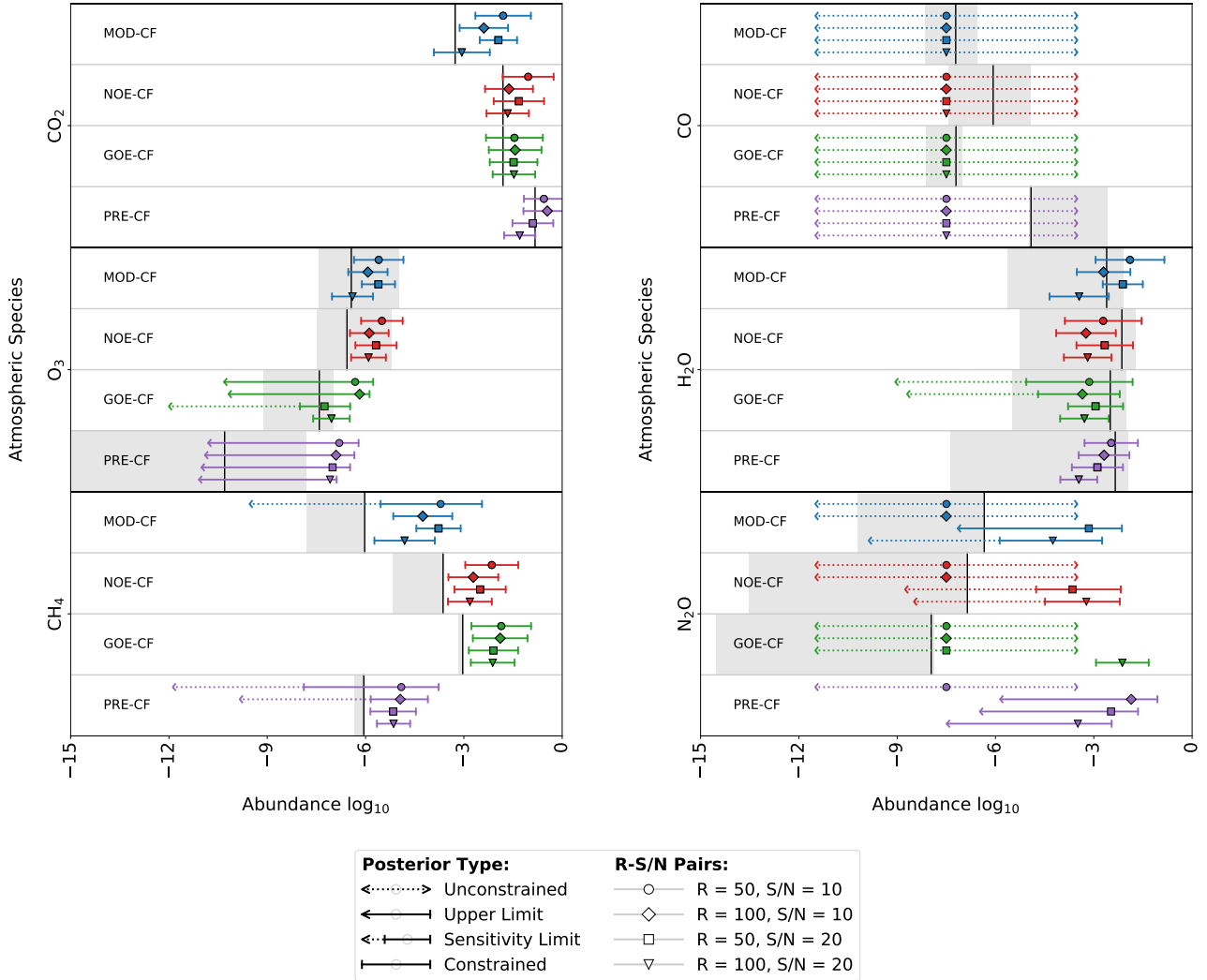
Posterior Type:
- ◄········► Unconstrained
- ◄─────┤ Upper Limit
- ◄··├───┤ Sensitivity Limit
- ├────┤ Constrained

R-S/N Pairs:
- ──○── R = 50, S/N = 10
- ──◇── R = 100, S/N = 10
- ──□── R = 50, S/N = 20
- ──▽── R = 100, S/N = 20

Fig. 7. Retrieved atmospheric abundances for the different ancillary runs. Results belonging to the various scenarios are provided using the color-coding from Table 1. We use different markers for the runs at different R–S/N (see legend). The solid lines indicate the expected values for each species, which vary depending on the epoch. The gray-shaded areas mark the range of values in the vertically nonconstant abundance profiles of the input spectra. The posterior distributions were classified using our posterior classification scheme (see Sect. 3.4 for details).

degeneracies. In particular, the pressure-abundance degeneracy is responsible for the shifts with respect to the true values, whereas the gravity-abundance degeneracy defines the variance of the abundance posteriors. The effects of the pressure-abundance degeneracy can be noticed for CO_2 in the MOD-CF scenario. For the reference case (circular marker) we strongly underestimated P_0, which is compensated by an overestimation in the CO_2 abundance. As we move to higher R and S/N input spectra, our estimate for P_0 improves, which results in better accuracies for the retrieved CO_2 abundance. The same connection between P_0 and the retrieved abundances can be seen for all other constrained species.

In contrast, the variance of the CO_2 posterior does not decrease significantly with increases in R and S/N since it is limited by the variance of the M_{pl} posterior (due to the gravity-abundance degeneracy), which is the same for all considered cases. While this behavior describes the results for most species well, there are some noteworthy exceptions that we will discuss here.

Firstly, there could be a tentative detection of O_3 (an SL posterior) in the clear GOE Earth (GOE-CF) epoch when increasing the S/N to 20 (square marker). If also the resolution is increased

to $R = 100$ (triangular marker), we could better constrain the O_3 abundance. Purely increasing R to 100 would not improve the accuracy or the precision of O_3 (diamond marker). Similarly, increasing the S/N would allow for detection of CH_4 in all four epochs, which was not possible for the reference case (circular marker). However, the retrieved CH_4 abundances are one to two orders of magnitude higher than the truths. Results suggest similar, but less pronounced, systematic offsets with respect to the true values for the other constrained species. These offsets are likely the result of a combination of the degeneracy between P_0 and the abundances and systematic errors, such as differences in the molecular line lists (see Sect. 4.4). Both O_3 and CH_4 are of particular interest for astrobiology, since they are indicative of disequilibrium chemistry in the atmosphere and could indicate the presence of biological activity on the planet. We discuss this in more detail in Sects. 4.3 and 4.5.

Furthermore, an increase in S/N would enable robust detection of H_2O in the GOE-CF epoch (a C- instead of an SL-type posterior). On the contrary, increasing the resolution alone does not have the same effect.

Finally, CO is unconstrained for all epochs and R–S/N pairs, which indicates that this species could not be detected in an

Earth-like atmosphere with LIFE. Similarly, none of the runs were able to fully constrain the N_2O abundance. The retrieval can only provide upper limits on the N_2O abundance, which often only manage to rule out atmospheric abundances greater than 1% in mass fraction. The retrieval is therefore not sensitive to these molecules – their spectral signatures are too small compared to the LIFESIM noise to be detected even in the best considered scenario. An exception would be the GOE-CF at $R = 100$ and S/N = 20 scenario, for which we retrieve a wrong estimate of N_2O (around 1% in mass fraction), about six orders of magnitude larger than the true value. The retrieval is most likely fitting the noise and/or spectral signatures of most of the other species at shorter wavelengths ($\lambda \lesssim 8\,\mu m$). Hence, when analyzing observations of potentially habitable terrestrial planets, we should be mindful not only of the false positive mechanisms that may be active in the atmosphere, but also of the false positives that the retrieval routines can infer. One could try to solve this issue by averaging over multiple retrieval runs, or by reducing the prior space with inferred knowledge from independent observations.

The retrievals of the ancillary cloudy input spectra (shown in Appendix C) do not show any noticeable improvement in either accuracy or precision for all scenarios with increased R and S/N. The values of R_{pl} and T_0 are still underestimated for all the scenarios. However, all considered R and S/N combinations still allow for an atmospheric abundance characterization for the cloudy input spectra. This analysis is subject to similar limitations as those already discussed for the subset of clear input spectra. The impact of clouds in retrievals will be discussed in more detail in Sect. 4.2.

4. Discussion

In Sect. 4.1, we compare the results we obtain for the cloud-free modern Earth twin with the results from a similar study performed in Paper III. As previously mentioned, we retrieved spectra of cloudy exoplanets, while neglecting clouds in the forward model of the retrieval framework. We describe this effect in Sect. 4.2. We discuss the impact of the quality of the data on the retrievals in Sect. 4.3 and the systematic effects of the retrieval runs in Sect. 4.4. Finally, we quantify the potential that LIFE has in differentiating the various epochs (Sect. 4.5).

For completeness, we mention that we also tested the impact of varying complexity of the theoretical spectral model on retrievals, by including and excluding scattering and/or CIA in the calculation. Through the analysis and comparison of ancillary retrieval grids, we confirm that including or neglecting scattering and CIA in the calculation does not influence the quality of the results. We show the results in Appendix D.

4.1. Comparison with Paper III

To allow for a proper comparison, we selected the model from Paper III that uses the same R, S/N, and wavelength range ($4-18.5\,\mu m$, $R = 50$, and S/N = 10). The major difference between these two retrieval studies is that in Paper III retrievals were performed using the same theoretical atmospheric model that was also used to generate the input spectra. In contrast, in this work we used an atmospheric model in the retrieval that is different than the one that was used to generate the spectrum. Further, in our previous study, we assumed abundance profiles that were vertically constant while the spectra calculated by Rugheimer & Kaltenegger (2018) were based on a self-consistent, altitude-dependent atmospheric composition. In Fig. 8 we compare the retrieval results for the constrained
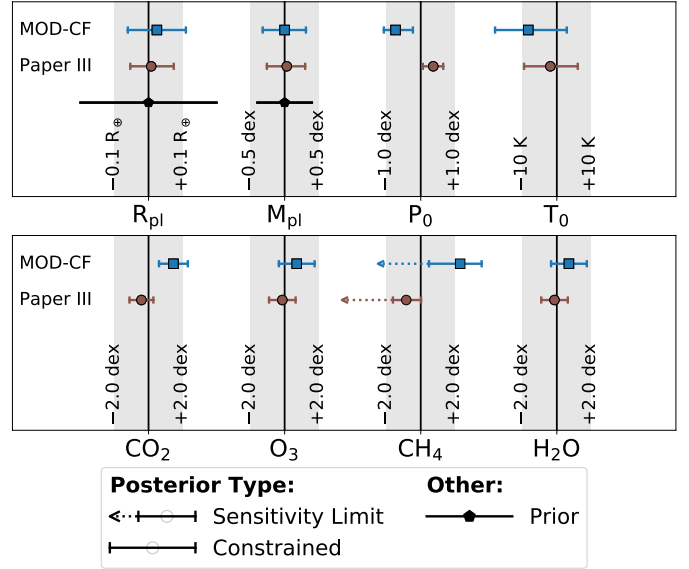

Fig. 8. Comparison of retrieval results for constrained planetary parameters and atmospheric abundances in the MOD-CF case (blue, square marker) with results from Paper III (brown, circular marker) for input spectra with the same properties (wavelength coverage $4-18.5\,\mu m$, $R = 50$, S/N = 10). The vertical black lines indicate the true values assumed for the parameters in each study. For parameters where we assumed a non-flat prior, we indicate the prior range (black, pentagonal marker). The error bars on the constrained posteriors denote the 68% confidence intervals.

planetary parameters and abundances from Paper III to our findings for the MOD-CF case.

In the upper panel of Fig. 8, we plot the retrieval results for the planetary parameters. The planetary radius R_{pl} is well constrained with respect to the assumed prior distribution and both posteriors are roughly centered on the corresponding truths. However, the spread of the MOD-CF R_{pl} posterior is larger than in Paper III, which indicates that the radius is slightly less well constrained. For M_{pl} our results are comparable to Paper III. Our results for the surface pressure P_0 and surface temperature T_0 agree less well with the results presented in Paper III. This is probably caused by small differences in the input $P-T$ profiles, as well as potential systematic errors, which we discuss in more detail in Sect. 4.4.

In the lower panel of Fig. 8, we show the results obtained for the abundances of the trace gases that are constrained (C- or SL-type posteriors) by our retrieval analysis. We observe that the MOD-CF retrieval tends to overestimate the true abundances, while the estimates from Paper III appear more accurate. The retrieved posterior types match for all of the atmospheric gases considered. Additionally, for CO_2, O_3, and H_2O, the spread of the posteriors for the MOD-CF runs is comparable to the results from Paper III. The larger spread in our CH_4 abundance is the result of a slightly reduced sensitivity, which is most likely evoked by differences in the atmospheric scenarios used to generate the input spectrum and in the retrievals. These differences will reduce the accuracy overall of the retrieval results.

4.2. Impact of clouds on retrieval results

As pointed out in Sect. 3, using a cloud-free atmospheric model in our retrievals will likely have introduced biases into our results for the cloudy scenarios. The presence of clouds in an atmosphere will reduce the MIR continuum emission of the observed

exoplanet. The emission spectra of terrestrial exoplanets are typically dominated by the lowest, nonopaque atmospheric layers. In cloudy exoplanets, part of this thermal emission is hidden below the clouds. Thus, the atmospheric layers above the clouds contribute more to the overall spectrum. Because the atmospheric temperature at the top of the clouds is typically lower than the surface temperature of the planet, cloud coverage will generally lead to a cooler retrieval temperature with reduced continuum flux. This reduction in continuum flux can be clearly seen when comparing the clear to the cloudy spectra in Fig. 1.

Since the theoretical spectral model we use for our retrievals assumes a cloud-free atmosphere, the continuum emission must be reduced in other ways to achieve a satisfactory fit to the input spectrum. This reduction can be obtained by reducing the radius (and thus the emitting surface) and/or the surface temperature of the exoplanet. Due to these compensation effects, most of our retrievals of cloudy input spectra yield smaller radii and/or cooler surface temperatures than the cloud-free inputs (see Fig. 4). This is also valid at higher spectral resolutions and signal-to-noise ratios, showing that the compensation effects are independent of the quality of the data (see Fig. C.1).

In addition to the surface conditions, the thermal structure of the layers and the chemical composition of the atmosphere also play a major role in shaping the emission spectrum, especially in the absorption and emission features. This could yield biased retrieval results for other parameters, such as the a_i coefficients of the polynomial P–T profile. In our results, the PRE-C model shows the clearest signature of this compensation effect (see Fig. 3). This degeneracy between cloud coverage and thermal structure in retrievals has also been found in other studies (see, e.g., Molline et al. 2020, and references therein).

Such biased results could be misleading, especially when trying to analyze the habitability of an observed exoplanet. If, by neglecting clouds in our theoretical spectral model, we underestimate the surface temperature, we could therefore misclassify habitable exoplanets. A clear example is the cloudy modern Earth (MOD-C) scenario: using our cloud-free forward model to retrieve this cloudy spectrum causes the retrieved ground temperature to be colder than 275 K. Such low temperatures suggest a potentially uninhabitable planet, which we know is not the correct interpretation for the MOD-C spectrum.

On the other hand, when looking at the retrieved chemical abundances (see Fig. 5), we observe only minor variations in the shape of the posteriors for all the major absorbing gases in the atmosphere. This indicates that, despite having a major impact on the retrieved physical parameters (R_{pl} and T_0), retrieving cloudy spectra with a cloud-free model does not significantly impact the chemical characterization of the atmosphere.

It is important to mention that the clouds included in Rugheimer & Kaltenegger (2018) act purely as continuum absorbers or emitters, with no specific cloud spectral features. Assuming a more complex treatment of clouds in the input spectra might still cause degenerate results, especially in the retrieval of a correct water abundance.

Therefore, including a cloud model in the theoretical spectral model that we use for retrievals could improve the quality of the results. However, this depends on the goal of the analysis. If we aim to characterize the chemical composition of the atmosphere, it may be sufficient to use cloud-free retrievals. This would be a smart strategy considering that including a somewhat realistic cloud treatment in the theoretical spectral model significantly increases the number of retrieved parameters and subsequently the running time. Performing retrievals on input spectra that include visible and near-infrared data in addition to the MIR

observations will likely provide additional information about the cloud composition and structure. In this sense, coupling with data acquired by HabEx or LUVOIR and LIFE may significantly improve retrieval results. We will compare the retrieval performance for different cloud models and discuss the capabilities of joint reflected light and thermal emission retrievals in future publications.

4.3. Increasing the quality of the input spectra

The results of the retrievals performed assuming other combinations of R and S/N, described in Sect. 3.5, show that increasing the S/N to 20 will allow us to detect both O_3 and CH_4 in more cases. Increasing to $R = 100$ would also improve our results, especially when combined with an increase in S/N. This is an interesting finding for multiple reasons.

From the scientific point of view, simultaneously detecting O_3 (which can provide an indirect estimate of O_2) and CH_4 would be a strong indicator of chemical disequilibrium in the atmosphere possibly hinting at the existence of biological activity. Such a detection would make the respective exoplanet a high-priority target for the search for life beyond the Solar System. This concept will be further explored in Sect. 4.5.

From the technical point of view, it would mean that one needs to consider longer integration times while maintaining a stable architecture of the interferometer array. For the assumptions in our baseline case (see Table 2), doubling the resolution would roughly correspond to a doubling of the integration time (from ~50 to ~100 days) while doubling the S/N would translate in integration times roughly four times longer (from ~50 to ~200 days). This poses challenges in terms of mission technical feasibility as well as mission scheduling. Increasing the instrument throughput, for which we assumed a conservative value (cf. Paper I), or the aperture size would bring the required integration times down. Also, the nearest rocky exoplanets orbiting within the habitable zone (HZ) of their solar-type host stars may not be 10 pc away. Bryson et al. (2021) estimate that with 95% confidence the nearest HZ planet around G and K dwarfs is ~6 pc away and they predict ~4 HZ rocky planets around G and K dwarfs within 10 pc of the Sun. Taking all this together, we would therefore recommend to stick to the baseline requirements for LIFE of $R = 50$ and S/N = 10, as proposed in Paper III, since they allow for a reliable and quantitative characterization of the most important physical and chemical properties of the considered atmospheres. The most promising targets could then be observed further to increase the S/N, thus allowing a more precise characterization of the atmosphere.

4.4. Systematics and current challenges

Thus far, we can confidently conclude that our Bayesian framework can retrieve consistent and robust results. This is not only valid for simulated observations generated with petitRADTRANS (see Paper III), but also for input spectra produced by other radiative transfer models (here by Rugheimer & Kaltenegger 2018). These results are highly promising in the context of analyzing real observational data in the future. However, as we mentioned in the previous sections, our work has identified some aspects that may lead to biased results. Some issues are linked to the intrinsic limitations of the Bayesian retrieval routine we described in Sect. 2.3. Ideally, these can be mitigated to improve the results, for example by choosing a different P–T profile parametrization, or by adding a cloud model to the retrieval. Further, we purposely chose to perform our
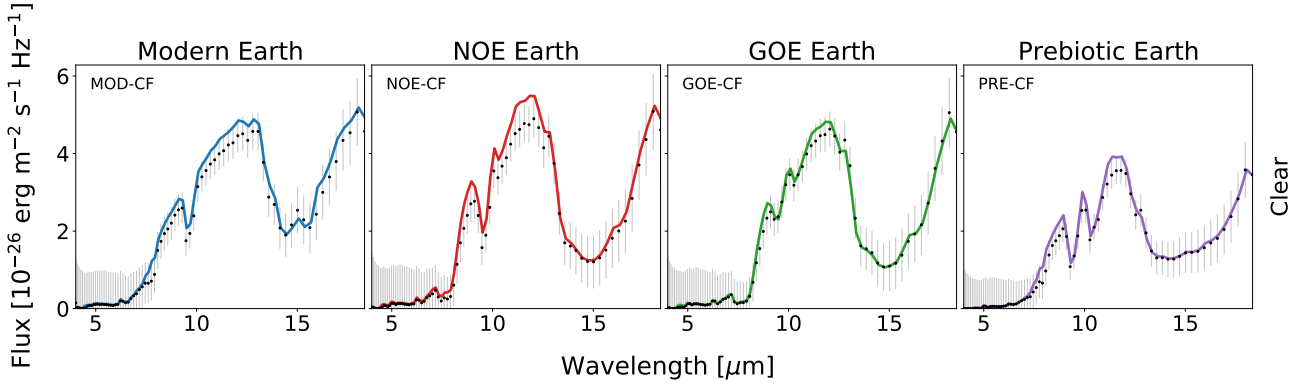
Fig. 9. Comparison of the $R = 50$ MIR spectra of the four clear-sky epochs (MOD-CF, NOE-CF, GOE-CF, and PRE-CF; solid lines following the color scheme of Table 1) produced with `petitRADTRANS` with the results from Rugheimer & Kaltenegger (2018) (black dots) assuming the same input parameters (i.e., $P–T$ profile, abundances, planetary dimensions). The error bars indicate the LIFESIM uncertainty assumed for the main grid of retrievals (S/N = 10 at 11.2 μm).

retrievals assuming uniform priors for most parameters where all values were possible if within a specified, wide range (see Sect. 2 for details). However, for future observations, the prior space might already be constrained (e.g., if one or more parameters are already measured by independent observations) and this would likely improve our retrieval results.

Despite these possibilities, we will eventually be limited by two factors. The first is the number of parameters that the Bayesian framework can handle within reasonable computing time. This limit on the number of parameters will remain unless novel parameter estimation algorithms emerge. An example would be the use of machine-learning retrieval routines (e.g., Waldmann 2016; Márquez-Neila et al. 2018; Cobb et al. 2019). Second, for a given resolution the information content of the spectrum is limited. Therefore, considering additional parameters in the retrieval framework could bias the results, for example causing a false positive inference of an atmospheric species.

However, the most relevant issues are independent of the parameter estimation routine. They are rooted in the intrinsic differences between individual radiative transfer models used to produce the MIR input spectra and the theoretical spectra in the retrievals. Such discrepancies may be caused by a slightly different treatment of physical or chemical processes, or differences in the assumed opacity tables. To investigate these issues, we computed the MIR spectra for the four clear-sky scenarios (MOD-CF, NOE-CF, GOE-CF, and PRE-CF) using `petitRADTRANS`. We assumed exactly the same input parameters (i.e., $P–T$ profile, abundances, planetary dimensions) that Rugheimer & Kaltenegger (2018) used to produce their spectra. We show the results for $R = 50$ in Fig. 9. The `petitRADTRANS` spectra are plotted as solid lines using the color scheme from Table 1. The input spectra from Rugheimer & Kaltenegger (2018) are shown as black dots. The error bars indicate the LIFESIM uncertainty used in the main grid of retrievals (S/N = 10 at 11.2 μm).

We observe that the `petitRADTRANS` spectra deviate (mostly within the LIFESIM uncertainty) from the spectra calculated by Rugheimer & Kaltenegger (2018), despite both models assuming the exact same input. While the absorption features are generally in agreement with each other, the spectra produced by `petitRADTRANS` show a higher continuum flux, especially around 8–12 μm. This discrepancy is likely linked to differences in the opacity tables used by the two radiative transfer models. As stated in Sect. 2.3, these differ with respect to the wing cutoff, the line list databases, and the pressure broadening coefficients.

To prevent the wings of the pressure-broadened lines from extending to infinity (nonphysical), it is necessary to introduce a wing cutoff. However, different radiative transfer models assume different cutoff thresholds (see the comparisons performed by, e.g., Lee et al. 2019; Baudino et al. 2017; Barstow et al. 2020). Rugheimer & Kaltenegger (2018) used a wing cutoff at $25\,\mathrm{cm}^{-1}$ from the line center. In contrast, the line cutoff used for the `petitRADTRANS` opacity tables assumes an exponential line wing decrease (for details, see Mollière et al. 2019). This may explain the higher continuum emission observed for all `petitRADTRANS` spectra.

Regarding the line list databases, the default opacity tables used by `petitRADTRANS` stem from different sources. They were calculated from the HITEMP, HITRAN 2012, or ExoMol line lists (see Table 4). In contrast, the spectra from Rugheimer & Kaltenegger (2018) were computed using only the HITRAN 2016 line lists, which in some cases are more recent than the ones adopted in our study. At the pressures and temperatures of interest in the study, we would not expect large variations in the line lists, provided all the databases are synchronous. The use of different versions of the same database (e.g., HITRAN 2012 versus HITRAN 2016) might cause variations in the opacities since databases more recently updated generally include more transition lines (see, e.g., Gordon et al. 2017). Furthermore, the default `petitRADTRANS` opacities only account for transitions of the main isotope, while the opacity tables used in Rugheimer & Kaltenegger (2018) can account for additional isotopes.

For the pressure broadening coefficients: to compute the line profiles, it is necessary to account for collision-driven line broadening. This depends on the pressure and composition of each atmospheric layer. For most molecules, both models assumed air broadening, which is based on a modern-Earth-like atmospheric composition. However, for CH_4, the `petitRADTRANS` opacity table assumed a theoretical broadening model based on Eq. (15) in Sharp & Burrows (2007), which was experimentally validated. Another exception is N_2O, for which H-He broadening was assumed (see Chubb et al. 2021). However, at the pressures and temperatures of interest, we do not expect large differences due to pressure broadening (Sharp & Burrows 2007; Mollière et al. 2019; Gharib-Nezhad & Line 2019; Chubb et al. 2021). We mention it here for completeness.

These differences likely also account for a substantial part of the offsets we find in the retrieved parameter values. Future intercomparison studies could help us define a "best practice" upon which to agree, as a community, to compute opacity tables for

retrievals in order to minimize these systematic effects. Furthermore, ongoing experimental work will be necessary to improve the completeness of the transition line databases and reduce discrepancies.

4.5. Differentiating the epochs

A quantitative approach to differentiate between the various scenarios is through the results of the retrieval analyses. We performed a first qualitative step in this direction in Sects. 3.2, 3.3, and 3.4, where we visually compared the retrieved P–T profiles and the posteriors for the planetary parameters and abundances. Through visual comparison, we find that differentiating the epochs via the retrieved P–T structure and planetary parameters is challenging. By studying the retrieved abundance posteriors, we find that the best candidates to perform such differentiation are O_3 and CH_4. This finding is especially interesting since the O_2–CH_4 pair is generally considered the strongest biosignature (see Lovelock 1965; Lederberg 1965) and O_2 can be constrained from O_3 through atmospheric chemistry models. Thus, the detection of one or both of these molecules will likely trigger follow-up observations and could allow us to separate between potentially alive and lifeless planets. However, a more in-depth characterization of the atmospheres is limited by the large variance in the posteriors of all these species, which typically exceeds one order of magnitude.

A more quantitative separation between the retrieved posterior distributions for the various epochs can be achieved by considering the difference between the cumulative posterior distribution functions of two epochs for a model parameter. This approach is similar to the Kolmogorov-Smirnov test (Kolmogorov 1933; Smirnov 1939), which is generally used to assess whether two samples are drawn from the same underlying distribution. Given a model parameter M with prior range $X = [X_{min}, X_{max}]$, we calculated the cumulative distribution $G^M(x)$ for $x \in X$ of the retrieved posterior $P(x)$ as follows:

$$G^M(x) = \int_{X_{min}}^{x} P(x')\, dx' \cdot \left(\int_{X_{min}}^{X_{max}} P(x')\, dx' \right)^{-1}. \quad (1)$$

We then compared the cumulative distribution functions $G_a^M(x)$ and $G_b^M(x)$ of two different epochs, a and b, by considering the maximum difference $\Delta := \Delta_{a-b}^M \in [0, 1]$ between them:

$$\Delta = \max |G_a^M(x) - G_b^M(x)|. \quad (2)$$

Thus, small values of Δ indicate that the compared posterior distributions only show small differences relative to each other. In this case, it is hard to differentiate between the retrieved posteriors. On the other hand, larger values of Δ indicate that the differences between the two posteriors are likely to correspond to different underlying true values of the considered parameter.

We can calculate Δ for all the combinations of the various scenarios and all parameters. We get particularly interesting results for CH_4 and O_3. Figure 10 shows the cumulative distribution functions for all the combinations of the clear sky scenarios (MOD-CF, NOE-CF, GOE-CF, and PRE-CF) calculated from the posteriors of CH_4 and O_3, for $R = 50$ and S/N = 10. In each subplot of the corner plot we annotate the values of Δ (percentage) corresponding to each combination. On the diagonals, the retrieved posteriors for every scenario are shown for reference. We keep the color scheme defined by Table 1. Regarding CH_4, we can fairly confidently distinguish between the clear prebiotic Earth (PRE-CF) and the Earth after the GOE (GOE-CF),

for which $\Delta = 95\%$, as well as between PRE-CF and the Earth after the NOE (NOE-CF), for which $\Delta = 90\%$. The distinction between the prebiotic Earth and the modern Earth (MOD-CF), as well as between the NOE and the GOE Earth is more difficult ($\Delta \leq 31\%$). For O_3, we observe a clear division into two subgroups: on the one side the modern and NOE Earth, where we have a clear detection of O_3, and on the other hand, the GOE and prebiotic Earth, where we only retrieve an upper limit on the abundance. The high value of $\Delta \sim 90\%$ between all combinations of MOD-CF or NOE-CF versus GOE-CF or PRE-CF clearly allows such a distinction. This is in agreement with what we concluded from Fig. 5 in Sect. 3.4. However, in contrast to the qualitative discussion based on the appearance of the posteriors, Δ provides a promising metric to quantify the magnitude of these differences.

In Fig. 11, we summarize the calculated Δ values for all combinations of cloud-free input spectra and the different R–S/N pairs considered in Sect. 4.3, for a total of four tables. Within the tables, each cell shows the Δ value (percentage) for a given parameter (columns) and a comparison of two specific scenarios (rows). The cells are also colored according to the value of Δ, with darker hues for larger values of Δ. As mentioned above, the biggest differences between the posteriors at $R = 50$, S/N = 10 can be observed for the molecules CH_4 and O_3. Furthermore, we observe some differences for CO_2 and H_2O, as well as for the P_0 posteriors. However, as discussed previously, these differences are rooted in degeneracies between the pressure and the abundances (see Sect. 4.1) and are not caused by large physical differences in the underlying atmospheres. These findings are generally still valid as we move to higher R and S/N. Similar conclusions can also be drawn for the cloudy inputs (see Appendix C).

Since we are able to confidently detect O_3 for a clear Earth after the NOE and for the modern Earth and we can distinguish these two epochs from earlier scenarios (prebiotic and GOE Earth), we can infer that LIFE would be able to detect traces of life as we know it in an Earth-like atmosphere when the abundance of O_2 has passed the 10% PAL threshold. This is consistent with other studies that focused on different wavelength ranges, such as the work by Kawashima & Rugheimer (2019) based on LUVOIR. The biosignature pair CH_4–O_3 might be even easier to detect when the abundance of O_2 is around 10% PAL (NOE Earth), rather than on modern Earth. The NOE Earth scenario is particularly favored since the atmosphere is filled with enough O_3 to be detectable, but a low enough abundance of O_2 to deplete the CH_4 in the atmosphere. These results are also consistent with the results shown in Kawashima & Rugheimer (2019). In other words, if LIFE were to observe the Earth at various stages of its evolution orbiting the Sun at a 10 pc distance, it would be able to detect strong indicators of life starting from around 0.8 Ga (NOE Earth). The detection of CH_4 with an upper limit on O_3 would also allow a tentative detection of potential biological activity up to 2.0 Ga (GOE Earth).

We must keep in mind that the epochs that we chose for our study are momentary "snapshots" in the continuous evolution of Earth, even though these four scenarios represent the major changes in our atmosphere. Still, other evolutionary paths are possible in the context of exoplanets, especially when considering other stellar classes. Realistically, all promising candidates would be followed-up with additional observations within the LIFE mission. It is beyond the scope of this work to conclusively infer the presence of a biosphere from the measured spectra of potentially habitable candidates, As discussed in other works such as Meadows et al. (2018) or Krissansen-Totton et al. (2022),
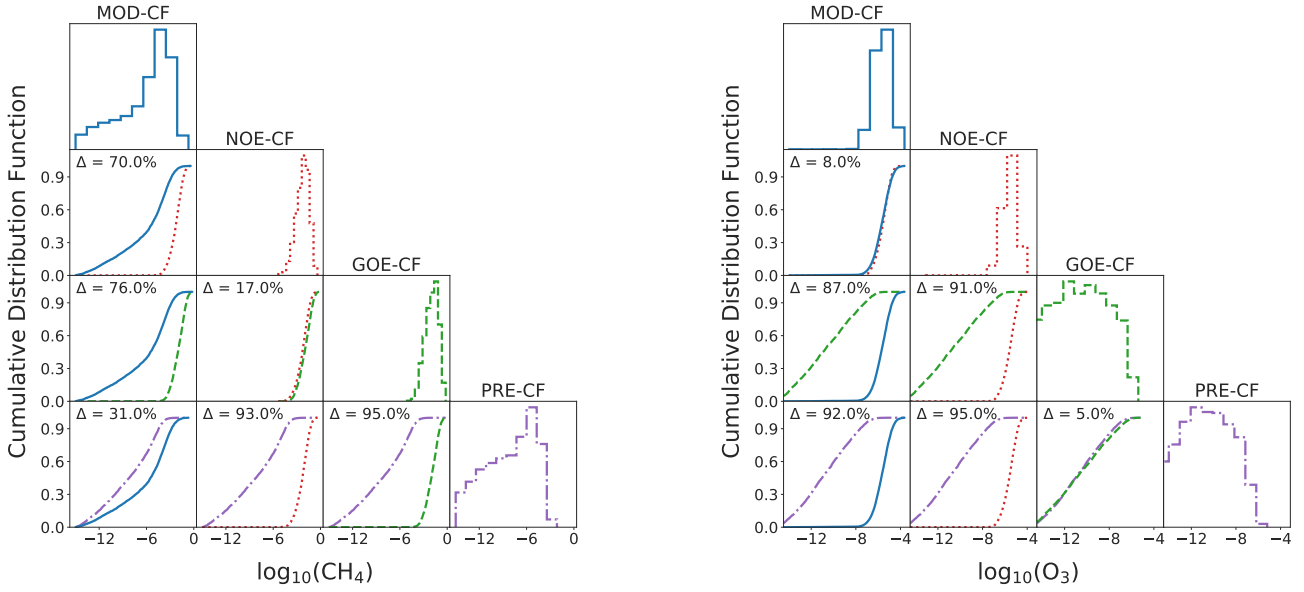
Fig. 10. Comparison of the cumulative distribution functions of the CH_4 posteriors (*left corner plot*) and the O_3 posteriors (*right corner plot*) for all the combinations of the clear sky scenarios (MOD-CF, NOE-CF, GOE-CF, and PRE-CF). The retrieved posteriors for each scenario are shown on the diagonal. Following the color scheme in Table 1, we show the posteriors and cumulative distribution functions as: solid blue lines (MOD-CF); dotted red lines (NOE-CF); dashed green lines (GOE-CF); and dash-dotted purple lines (PRE-CF).
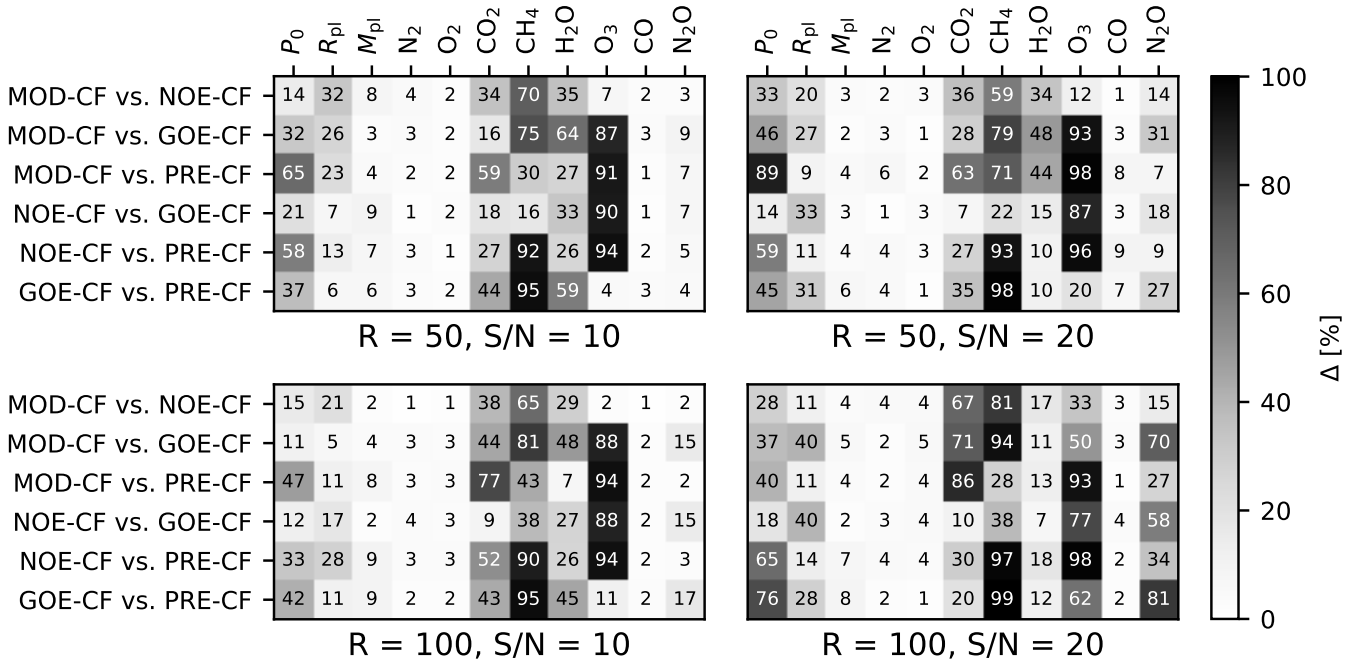
Fig. 11. Maximum difference, Δ, between the cumulative posteriors for the different model parameters for each combination of input spectra (cloud-free subset) and different R–S/N pairs. The background of each cell in the tables is related to the value of Δ (darker hues for larger Δ).

we would require a thorough discussion of the context information available for the observed planetary system before claiming a "life detection." However, the presented retrieval results are certainly an important piece of information for the development of frameworks for systematically assessing biosignature detections (e.g., Catling et al. 2018; Walker et al. 2018, see also the NFoLD Community Report from the Biosignatures Standards of Evidence Workshop[7]).

5. Conclusions

The Bayesian retrieval framework introduced in Paper III and extended here has delivered insightful answers to the questions introduced in Sect. 1. These can be summarized as follows:

1. LIFE can characterize prebiotic and biotic worlds. We can constrain the surface temperatures with an uncertainty of around 20 K. We can confirm, exclude, or give upper limits on the presence of several astrobiologically relevant molecules that show signatures in the MIR bands. In particular, LIFE can detect O_3 in the atmosphere if the

O_2 mass fraction is on the order of ∼10% PAL. In terrestrial atmospheres CH_4 can be constrained if its abundance is ∼0.1% in mass fraction. For lower abundances (around 10^{-6} in mass fraction), LIFE will detect an upper limit on CH_4 (SL-type posterior). Simultaneously constraining O_3 and CH_4 will be possible in atmospheres with an O_2 abundance of around 10% PAL. This is in agreement with other studies based on a different wavelength range (Kawashima & Rugheimer 2019). Such a result is relevant in terms of the detection of biosignatures in the atmospheres of habitable exoplanets.

2. Neglecting clouds in the retrievals could cause biases in the determination of the thermal structure of the atmosphere of cloudy exoplanets. However, cloud-free retrievals of cloudy spectra can still yield accurate results for what concerns the atmospheric composition.

3. We confirm that the minimum requirements in spectral resolution and S/N for a MIR mission such as LIFE found in Paper III are also sufficient for the scenarios considered here. Improving the S/N would allow for a clearer detection of O_3 and CH_4 even when these species are less abundant (until ∼10^{-7} in mass fraction). Therefore, a better characterization could be obtained by observing promising targets longer during the characterization phase of the mission (or by increasing the instrument throughput and/or aperture size).

4. We are able to demonstrate that inter-model comparison and retrieval is possible, with the caveats and limitations detailed in Sect. 4.4. The most important discrepancies in the retrievals are caused by the use of different opacity tables, in particular for what concerns the line wing cutoff treatment. Degeneracies and correlations in the posteriors appear as a result of the various relations among various parameters.

6. Next steps

Several new interesting questions and opportunities for more detailed studies arise from this work. First of all, we plan to work on a study that will take advantage of the model selection potential that Bayesian retrievals have to offer, for example by comparing retrievals including and excluding non-retrieved parameters (e.g., the CO abundance). We are also performing retrievals assuming various cloud models (Konrad et al., in prep.). Retrievals of hazy planets (see, e.g., Arney et al. 2016), as well as ocean worlds, might also help us further quantify the science potential of LIFE for a variety of different planet types.

Another interesting study would be to increase S/N and R to even higher values. This will not only evaluate the extreme limits of a concept such as LIFE but also help us better understand if retrievals are limited by R rather than S/N (e.g., due to unresolved narrow features at low R). It would also be useful to compare different R–S/N combinations, this time fixing the observing time. This would help us quantify the best R–S/N combination needed to optimize the characterization of a terrestrial atmosphere. Further work is needed to optimize the yield in the characterization phase of the LIFE mission concept. The estimates of the observation time needed to establish knowledge about the habitability and the presence of biologically relevant molecules in the atmosphere that we derived here are crucial pieces of information for these follow-up studies.

In this work, we only used simulated data obtained with the LIFE mission. However, in the future there will likely be more information available for each system and planet. Therefore, it will be important to put this study in context with other observations. For instance, joint retrievals of reflected light data obtained with LUVOIR or HabEx at optical-to-near-infrared wavelengths and thermal emission spectra as obtained by LIFE would provide useful insight into the synergies between the various missions.

One of the most important open questions regarding the ultimate goal of detecting extrasolar life will require us to put our results in the context of life detection frameworks (e.g., Green et al. 2021; Catling et al. 2018; Walker et al. 2018). Our ongoing retrieval efforts could be useful for the fine-tuning of such frameworks. These frameworks, in turn, would provide insight into the meaning and the likelihood of a potential biosignature detection, which would allow us to infer the presence of life-forms on another planet and justify such inferences.

References

Arney, G., Domagal-Goldman, S. D., Meadows, V. S., et al. 2016, Astrobiology, 16, 873
Astropy Collaboration (Robitaille, T. P., et al.) 2013, A&A, 558, A33
Astropy Collaboration (Price-Whelan, A. M., et al.) 2018, AJ, 156, 123
Barstow, J. K., Changeat, Q., Garland, R., et al. 2020, MNRAS, 493, 4884
Baudino, J.-L., Mollière, P., Venot, O., et al. 2017, ApJ, 850, 150
Bryson, S., Kunimoto, M., Kopparapu, R. K., et al. 2021, AJ, 161, 36
Buchner, J., Georgakakis, A., Nandra, K., et al. 2014, A&A, 564, A125
Burch, D. E., Gryvnak, D. A., Patty, R. R., & Bartky, C. E. 1969, J. Opt. Soc. Am., 59, 267
Campbell, I. H., & Squire, R. J. 2010, Geochim. Cosmochim. Acta, 74, 4187
Catling, D. C., Krissansen-Totton, J., Kiang, N. Y., et al. 2018, Astrobiology, 18, 709
Chen, J., & Kipping, D. 2016, ApJ, 834, 17
Chubb, K. L., Rocchetto, M., Yurchenko, S. N., et al. 2021, A&A, 646, A21
Cobb, A. D., Himes, M. D., Soboczenski, F., et al. 2019, AJ, 158, 33
Dannert, F., Ottiger, M., Quanz, S. P., et al. 2022, A&A, 664, A22 (Paper II)
Ertel, S., Defrère, D., Hinz, P., et al. 2020, AJ, 159, 177
Feautrier, P. 1964, Compt. Rend. Acad. Sci. (Serie non specifiee), 258, 3189
Feng, Y. K., Robinson, T. D., Fortney, J. J., et al. 2018, AJ, 155, 200
Feroz, F., Hobson, M. P., & Bridges, M. 2009, MNRAS, 398, 1601
Feulner, G. 2012, Rev. Geophys., 50, RG2006
Gaudi, B. S., Seager, S., Mennesson, B., et al. 2020, ArXiv e-prints, [arXiv:2001.06683]
Gharib-Nezhad, E., & Line, M. R. 2019, ApJ, 872, 27
Gordon, I. E., Rothman, L. S., Hill, C., et al. 2017, J. Quant. Spec. Radiat. Transf., 203, 3
Graham, R. J. 2021, Astrobiology, 21, 1406
Green, J., Hoehler, T., Neveu, M., et al. 2021, Nature, 598, 575
Gregory, B. S., Claire, M. W., & Rugheimer, S. 2021, Earth Planet. Sci. Lett., 561, 116818
Hartmann, J. M., Boulet, C., Brodbeck, C., et al. 2002, J. Quant. Spec. Radiat. Transf., 72, 117
Harvey, A. H., Gallagher, J. S., & Levelt Sengers, J. M. H. 1998, J. Phys. Chem. Ref. Data, 27, 761
Heller, R., Duda, J.-P., Winkler, M., Reitner, J., & Gizon, L. 2021, PalZ, 95, 563
Hunter, J. D. 2007, Comput. Sci. Eng., 9, 90
Jeffreys, H. 1998, The Theory of Probability, Oxford Classic Texts in the Physical Sciences (OUP Oxford), 432
Kaltenegger, L., & Traub, W. A. 2009, ApJ, 698, 519
Kaltenegger, L., Traub, W. A., & Jucks, K. W. 2007, ApJ, 658, 598
Karman, T., Gordon, I. E., van der Avoird, A., et al. 2019, Icarus, 328, 160

Kawashima, Y., & Rugheimer, S. 2019, AJ, 157, 213

Kolmogorov, A. 1933, Inst. Ital. Attuari, Giorn., 4, 83

Konrad, B. S., Alei, E., Angerhausen, D., et al. 2022, A&A, 664, A23 (Paper III)

Krissansen-Totton, J., Thompson, M., Galloway, M. L., & Fortney, J. J. 2022, Nat. Astron., 6, 189

Lederberg, J. 1965, Nature, 207, 9

Lee, E., Taylor, J., Grimm, S. L., et al. 2019, MNRAS, 487, 2082

Lovelock, J. E. 1965, Nature, 207, 568

Luo, G., Ono, S., Beukes, N. J., et al. 2016, Sci. Adv., 2, e1600134

Lyons, T. W., Reinhard, C. T., & Planavsky, N. J. 2014, Nature, 506, 307

Lyons, T. W., Diamond, C. W., Planavsky, N. J., Reinhard, C. T., & Li, C. 2021, Astrobiology, 21, 906

Madhusudhan, N. 2018, Handbook of Exoplanets, 2153

Marais, D. J. D. 2000, Science, 289, 1703

Márquez-Neila, P., Fisher, C., Sznitman, R., & Heng, K. 2018, Nat. Astron., 2, 719

Meadows, V. S., Reinhard, C. T., Arney, G. N., et al. 2018, Astrobiology, 18, 630

Mollière, P., Boekel, R. v., Dullemond, C., Henning, T., & Mordasini, C. 2015, ApJ, 813, 47

Mollière, P., van Boekel, R., Bouwman, J., et al. 2017, A&A, 600, A10

Mollière, P., Wardenier, J. P., van Boekel, R., et al. 2019, A&A, 627, A67

Mollière, P., Stolker, T., Lacour, S., et al. 2020, A&A, 640, A131

National Academies of Sciences, Engineering, and Medicine. 2021, Pathways to Discovery in Astronomy and Astrophysics for the 2020s (Washington, DC: The National Academies Press)

Olson, S. L., Schwieterman, E. W., Reinhard, C. T., & Lyons, T. W. 2018, Earth: Atmospheric Evolution of a Habitable Planet, eds. H. J. Deeg, & J. A. Belmonte (Cham: Springer International Publishing), 2817

pandas development team, T. 2020, https://doi.org/10.5281/zenodo.3509134

Peterson, B. M., Fischer, D., & LUVOIR Science and Technology Definition Team. 2017, in American Astronomical Society Meeting Abstracts, 229, 405.04

Quanz, S. P., Absil, O., Angerhausen, D., et al. 2021, Atmospheric characterization of terrestrial exoplanets in the mid-infrared: biosignatures, habitability & diversity, Exp. Astron.

Quanz, S. P., Ottiger, M., Fontanet, E., et al. 2022, A&A, 664, A21 (Paper I)

Rothman, L., Gordon, I., Barber, R., et al. 2010, J. Quant. Spectr. Rad. Transf., 111, 2139

Rothman, L. S., Gordon, I. E., Babikov, Y., et al. 2013, J. Quant. Spec. Radiat. Transf., 130, 4

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

Rugheimer, S., Kaltenegger, L., Zsom, A., Segura, A., & Sasselov, D. 2013, Astrobiology, 13, 251

Sharp, C. M., & Burrows, A. 2007, ApJS, 168, 140

Shields-Zhou, G., & Och, L. 2011, GSA Today, 21, 4

Skilling, J. 2006, Bayesian Anal., 1, 833

Smirnov, N. V. 1939, Bull. Math. Univ. Moscou, 2, 3

Sneep, M., & Ubachs, W. 2005, J. Quant. Spec. Radiat. Transf., 92, 293

Thalman, R., Zarzana, K. J., Tolbert, M. A., & Volkamer, R. 2014, J. Quant. Spec. Radiat. Transf., 147, 171

Thalman, R., Zarzana, K. J., Tolbert, M. A., & Volkamer, R. 2017, J. Quant. Spec. Radiat. Transf., 189, 281

Waldmann, I. P. 2016, ApJ, 820, 107

Walker, S. I., Bains, W., Cronin, L., et al. 2018, Astrobiology, 18, 779

Wolfe, J. M., & Fournier, G. P. 2018, Nat. Ecol. Evol., 2, 897

Yurchenko, S. N., & Tennyson, J. 2014, MNRAS, 440, 1649

Appendix A: Scattering of terrestrial exoplanets

As discussed in Mollière et al. (2020), `petitRADTRANS` was updated to treat scattering. This was done using the Feautrier method (Feautrier 1964). This is a third-order method that allows the treatment of radiative transfer equation in the diffusive regime.

The Feautrier method solves the angle- and frequency-dependent radiative transfer equation for both the planetary and the stellar radiation field. These can be treated separately, since the radiative transfer equation (Eq. A.1) depends only linearly from the intensity I:

$$\mu \frac{dI}{d\tau} = -I + S. \tag{A.1}$$

Here, $\mu = \cos\theta$ where θ is the angle between a light ray and the surface normal, τ is the optical depth, I is the intensity, and S is the source function.

Conceptually, for any direction μ of a ray, there also exists a ray in direction $-\mu$, where $\mu \in [-1, 1]$. It is possible to instead let μ run from 0 to 1 only, and define rays I_+ and I_- parallel and antiparallel to this direction. For one of these, the projection onto the atmospheric normal vector (defined by the scalar product) will be positive (going upward), while for the other one it will be negative (that is, going downward). Eq. A.1 can be therefore rewritten as

$$\frac{dI_+}{d\tau} = S - I_+ \tag{A.2}$$

$$\frac{dI_-}{d\tau} = -S + I_-. \tag{A.3}$$

To solve these, it is convenient to define other two variables,

$$I_J = \frac{1}{2}(I_+ + I_-) \tag{A.4}$$

$$I_H = \frac{1}{2}(I_+ - I_-), \tag{A.5}$$

such that Eqs. A.2 and A.3 become

$$\frac{dI_J}{d\tau} = -I_H \tag{A.6}$$

$$\frac{dI_H}{d\tau} = S - I_J. \tag{A.7}$$

Replacing I_H as defined by Eq. A.6 into Eq. A.7, we obtain Feautrier's equation:

$$\frac{d^2I_J}{d\tau^2} = I_J - S. \tag{A.8}$$

In this paper we only take thermal scattering (i.e., scattering of the planetary radiation) into account. We, therefore, neglect the scattering of the direct stellar contribution. However, since the radiative transfer equation depends only linearly on I_ν, the contribution of the stellar radiation can be treated as an additional term in the calculation (see Mollière et al. 2017). This term is also included in the latest version of `petitRADTRANS`

and we refer to the online documentation for a more detailed description[8].

Purely considering the planetary radiation, we define the boundary conditions at the top of the atmosphere:

$$I_+(P = 0, \mu) = 0 \quad \forall\mu, \tag{A.9}$$

meaning that there is no planetary radiation coming downward from the top of the atmosphere, and at the surface

$$I_-(P = P_{surf}, \mu) = e_{surf}\, B(T_{surf}) + a_{surf} J^{scat}(P_{surf}). \tag{A.10}$$

The constraint on I_- at the lower boundary is composed of the thermal emission of the surface itself (blackbody radiation scaled by the surface emissivity e_{surf}) and by a portion of the incoming planetary radiation that is reflected by the surface. The wavelength dependence of the effectiveness of the reflection depends on the "surface albedo" or "reflectance" a_{surf}. The average scattered intensity J^{scat} is the integral of I_+ over all the possible angles ($\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$, which corresponds to the light that comes from the top layers):

$$J^{scat}(P_{surf}) = \int_0^1 I_+(P_{surf})d\mu. \tag{A.11}$$

The boundary conditions translate, in terms of I_H and I_J, in

$$I_J(0) = -\frac{I_-(0)}{2} \tag{A.12}$$

and

$$I_J(P_{surf}) = \frac{1}{2}\left[I_+(P_{surf}) + e_{surf}\, B(T_{surf}) + a_{surf} J^{scat}(P_{surf})\right]. \tag{A.13}$$

It is possible to thus solve Eq. A.8 for $i \neq 1$, $i \neq N$ by discretization:

$$-\frac{\left(\frac{I_{J,i+1}-I_{J,i}}{\tau_{i+1}-\tau_i}\right) - \left(\frac{I_{J,i}-I_{J,i-1}}{\tau_i-\tau_{i-1}}\right)}{\left(\frac{\tau_{i+1}+\tau_i}{2} - \frac{\tau_i+\tau_{i-1}}{2}\right)} + I_{J,i} = S_i, \tag{A.14}$$

which can be expressed in matrix form by extracting the coefficients a_i, b_i, and c_i:

$$\begin{pmatrix} b_1 & c_1 & 0 & \cdots & 0 \\ a_2 & b_2 & c_2 & \ddots & \cdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & a_{N-1} & b_{N-1} & c_{N-1} \\ 0 & \cdots & 0 & a_N & b_N \end{pmatrix} \cdot \begin{pmatrix} I_{J,1} \\ I_{J,2} \\ \vdots \\ I_{J,N-1} \\ I_{J,N} \end{pmatrix} = \begin{pmatrix} S_1 \\ S_2 \\ \vdots \\ S_{N-1} \\ S_N \end{pmatrix}. \tag{A.15}$$

To take into account the boundary conditions, at $i = 1$ the value of a_1 is 0, while at $i = N$ both c_N and a_N will be 0, since there is no dependence from the $(N-1)$th layer in the boundary condition A.13; b_N, as a consequence, will be equal to 1.

---

[8] https://petitradtrans.readthedocs.io/en/latest/content/notebooks/emis_scat.html#Scattering-of-stellar-light

The tridiagonal matrix can be inverted to retrieve the corresponding values of I_J through multiple iterations. This iterative process is needed to correctly take into account the scattering contribution into the source function terms S_i. During the first iteration of the Feautrier's routine, the scattering contribution has yet to be properly calculated. The source function at this step corresponds to the thermal blackbody radiation produced by the atmospheric layer i at temperature T_i:

$$S_i = B_i = B(T_i). \tag{A.16}$$

For any other iteration, the model will consider the previous solution for I_J to calculate the new source function, which will then include the contribution of the photons that have been scattered in the previous steps. In the case of $i = N$, the source function must correspond to the right term in Eq. A.13, computed using the most recent estimate of $I_+(P_{surf})$ and $J^{scat}(P_{surf})$. This process can be accelerated through the accelerated lambda iteration and Ng methods (see Mollière et al. 2017, p. 75)).

From that value, it is possible to calculate I_H using Eq. A.6. The emergent flux at the top of the atmosphere can be then calculated as follows:

$$
\begin{aligned}
F &= \int_0^{2\pi} \int_0^{\pi/2} I(P = 0) \cos\theta \sin\theta d\phi d\theta \\
&= 2\pi \int_0^{\pi/2} I(P = 0)\mu d\mu = -4\pi \int_0^1 I_H(P = 0)\mu d\mu.
\end{aligned}
\tag{A.17}
$$

The iterations stop once the estimate of the flux has reached a convergence value.

Appendix B: Corner plots

Corner plots for the retrieval runs at the reference R and S/N are shown in this section. We grouped both the cloudy and the clear sky retrievals for each epoch in the same figure, in order to compare the results. Namely: Fig. B.1 shows the corner plots of the two modern Earth scenarios (MOD-CF and MOD-C); Fig. B.2 shows the NOE Earth scenarios (NOE-CF and NOE-C); the GOE Earth scenarios (GOE-CF and GOE-C) are in Fig. B.3; finally, the prebiotic scenarios (PRE-CF and PRE-C) are shown in Fig. B.4.

The models are color-coded according to Table 1. Also, the results for the clear sky retrievals are shown using dashed contour lines, while the cloudy models are represented using solid lines. The table on the top right of each figure shows the expected values for each parameter, together with the estimates and the 1σ uncertainty for the two scenarios.

	True Values	MOD-CF	MOD-C
a_4	0.232	$0.499 ^{+0.673}_{-0.331}$	$0.649 ^{+0.839}_{-0.416}$
a_3	18.644	$14.477 ^{+7.905}_{-6.307}$	$14.545 ^{+9.961}_{-7.056}$
a_2	110.331	$107.881 ^{+49.634}_{-44.764}$	$96.838 ^{+51.943}_{-44.645}$
a_1	177.997	$276.849 ^{+122.163}_{-131.739}$	$224.876 ^{+113.414}_{-109.663}$
a_0	293.601	$431.651 ^{+113.563}_{-119.198}$	$381.808 ^{+94.732}_{-87.722}$
$\log_{10} P_0$	0.007	$-0.732 ^{+0.509}_{-0.344}$	$-0.674 ^{+0.446}_{-0.328}$
R_{pl}	1.000	$1.025 ^{+0.085}_{-0.085}$	$0.987 ^{+0.076}_{-0.070}$
$\log_{10} M_{pl}$	0.000	$-0.000 ^{+0.313}_{-0.322}$	$0.022 ^{+0.316}_{-0.335}$
$\log_{10} N_2$	-0.113	$-7.239 ^{+4.323}_{-4.663}$	$-7.416 ^{+4.427}_{-4.362}$
$\log_{10} O_2$	-0.631	$-7.533 ^{+4.520}_{-4.534}$	$-7.934 ^{+4.501}_{-4.336}$
$\log_{10} CO_2$	-3.265	$-1.844 ^{+0.759}_{-0.930}$	$-1.944 ^{+0.807}_{-0.844}$
$\log_{10} CH_4$	-6.028	$-5.100 ^{+1.893}_{-5.039}$	$-5.429 ^{+2.172}_{-5.172}$
$\log_{10} H_2O$	-2.607	$-1.977 ^{+0.967}_{-1.127}$	$-2.339 ^{+1.047}_{-1.503}$
$\log_{10} O_3$	-6.436	$-5.621 ^{+0.714}_{-0.785}$	$-5.795 ^{+0.782}_{-0.832}$
$\log_{10} CO$	-7.215	$-7.748 ^{+4.636}_{-4.543}$	$-7.684 ^{+4.571}_{-4.475}$
$\log_{10} N_2O$	-6.343	$-8.126 ^{+4.213}_{-4.265}$	$-7.859 ^{+4.191}_{-4.322}$

Fig. B.1: Corner plot for the posterior distributions from the retrievals of the MOD-CF (dashed contour lines) and MOD-C (solid contour lines) scenarios. The black lines indicate the expected values for every parameter. The retrieved values (median and 1σ uncertainties) are shown in the table in the top-right corner, together with the expected values. The scenarios are color-coded according to Table 1.

	True Values	NOE-CF	NOE-C
a_4	1.559	$0.578\ ^{+0.733}_{-0.384}$	$0.468\ ^{+0.700}_{-0.309}$
a_3	17.562	$11.169\ ^{+8.774}_{-5.872}$	$10.208\ ^{+8.076}_{-5.457}$
a_2	72.414	$76.466\ ^{+52.025}_{-36.130}$	$68.343\ ^{+49.116}_{-33.276}$
a_1	140.294	$210.479\ ^{+135.169}_{-82.222}$	$179.649\ ^{+132.085}_{-73.795}$
a_0	307.747	$400.509\ ^{+114.303}_{-79.944}$	$373.847\ ^{+115.448}_{-72.460}$
$\log_{10} P_0$	0.017	$-0.614\ ^{+0.434}_{-0.312}$	$-0.675\ ^{+0.450}_{-0.355}$
R_{pl}	1.000	$0.952\ ^{+0.086}_{-0.086}$	$0.888\ ^{+0.060}_{-0.061}$
$\log_{10} M_{pl}$	0.000	$-0.061\ ^{+0.320}_{-0.307}$	$-0.012\ ^{+0.315}_{-0.329}$
$\log_{10} N_2$	-0.016	$-7.684\ ^{+4.733}_{-4.558}$	$-7.840\ ^{+4.640}_{-4.460}$
$\log_{10} O_2$	-1.623	$-7.763\ ^{+4.681}_{-4.259}$	$-7.593\ ^{+4.399}_{-4.383}$
$\log_{10} CO_2$	-1.807	$-1.127\ ^{+0.611}_{-0.855}$	$-1.338\ ^{+0.717}_{-0.774}$
$\log_{10} CH_4$	-3.635	$-2.189\ ^{+0.712}_{-0.934}$	$-2.345\ ^{+0.758}_{-0.909}$
$\log_{10} H_2O$	-2.143	$-3.093\ ^{+1.231}_{-3.451}$	$-5.858\ ^{+2.884}_{-5.188}$
$\log_{10} O_3$	-6.567	$-5.519\ ^{+0.564}_{-0.709}$	$-5.535\ ^{+0.649}_{-0.702}$
$\log_{10} CO$	-6.072	$-7.460\ ^{+4.454}_{-4.589}$	$-7.612\ ^{+4.459}_{-4.346}$
$\log_{10} N_2O$	-6.859	$-7.809\ ^{+4.206}_{-4.376}$	$-8.391\ ^{+4.059}_{-3.920}$

Fig. B.2: As for Fig. B.1 but for the NOE-CF and NOE-C scenarios.

	True Values	GOE-CF	GOE-C
a_4	0.996	$0.665\ ^{+0.874}_{-0.443}$	$0.522\ ^{+0.735}_{-0.346}$
a_3	12.886	$11.003\ ^{+9.241}_{-5.604}$	$10.180\ ^{+8.160}_{-5.299}$
a_2	59.984	$66.077\ ^{+41.068}_{-30.990}$	$64.387\ ^{+47.643}_{-30.642}$
a_1	123.764	$164.201\ ^{+108.345}_{-66.526}$	$169.269\ ^{+122.816}_{-69.804}$
a_0	290.306	$342.610\ ^{+94.435}_{-67.036}$	$358.076\ ^{+108.888}_{-66.047}$
$\log_{10} P_0$	0.008	$-0.371\ ^{+0.498}_{-0.389}$	$-0.593\ ^{+0.453}_{-0.346}$
R_{pl}	1.000	$0.967\ ^{+0.084}_{-0.083}$	$0.955\ ^{+0.076}_{-0.077}$
$\log_{10} M_{pl}$	0.000	$0.008\ ^{+0.320}_{-0.323}$	$-0.036\ ^{+0.333}_{-0.326}$
$\log_{10} N_2$	-0.005	$-7.566\ ^{+4.514}_{-4.611}$	$-7.639\ ^{+4.634}_{-4.621}$
$\log_{10} O_2$	-2.622	$-7.441\ ^{+4.245}_{-4.395}$	$-7.577\ ^{+4.667}_{-4.673}$
$\log_{10} CO_2$	-1.806	$-1.486\ ^{+0.702}_{-0.911}$	$-1.213\ ^{+0.625}_{-0.781}$
$\log_{10} CH_4$	-3.035	$-1.896\ ^{+0.774}_{-0.957}$	$-1.613\ ^{+0.680}_{-0.856}$
$\log_{10} H_2O$	-2.498	$-5.419\ ^{+2.604}_{-5.778}$	$-6.896\ ^{+3.765}_{-5.047}$
$\log_{10} O_3$	-7.412	$-10.583\ ^{+2.815}_{-2.769}$	$-10.581\ ^{+2.985}_{-2.758}$
$\log_{10} CO$	-7.206	$-7.327\ ^{+4.667}_{-4.558}$	$-7.390\ ^{+4.416}_{-4.445}$
$\log_{10} N_2O$	-7.963	$-7.196\ ^{+4.166}_{-4.672}$	$-7.021\ ^{+4.258}_{-4.947}$

Fig. B.3: As for Fig. B.1 but for the GOE-CF and GOE-C scenarios.

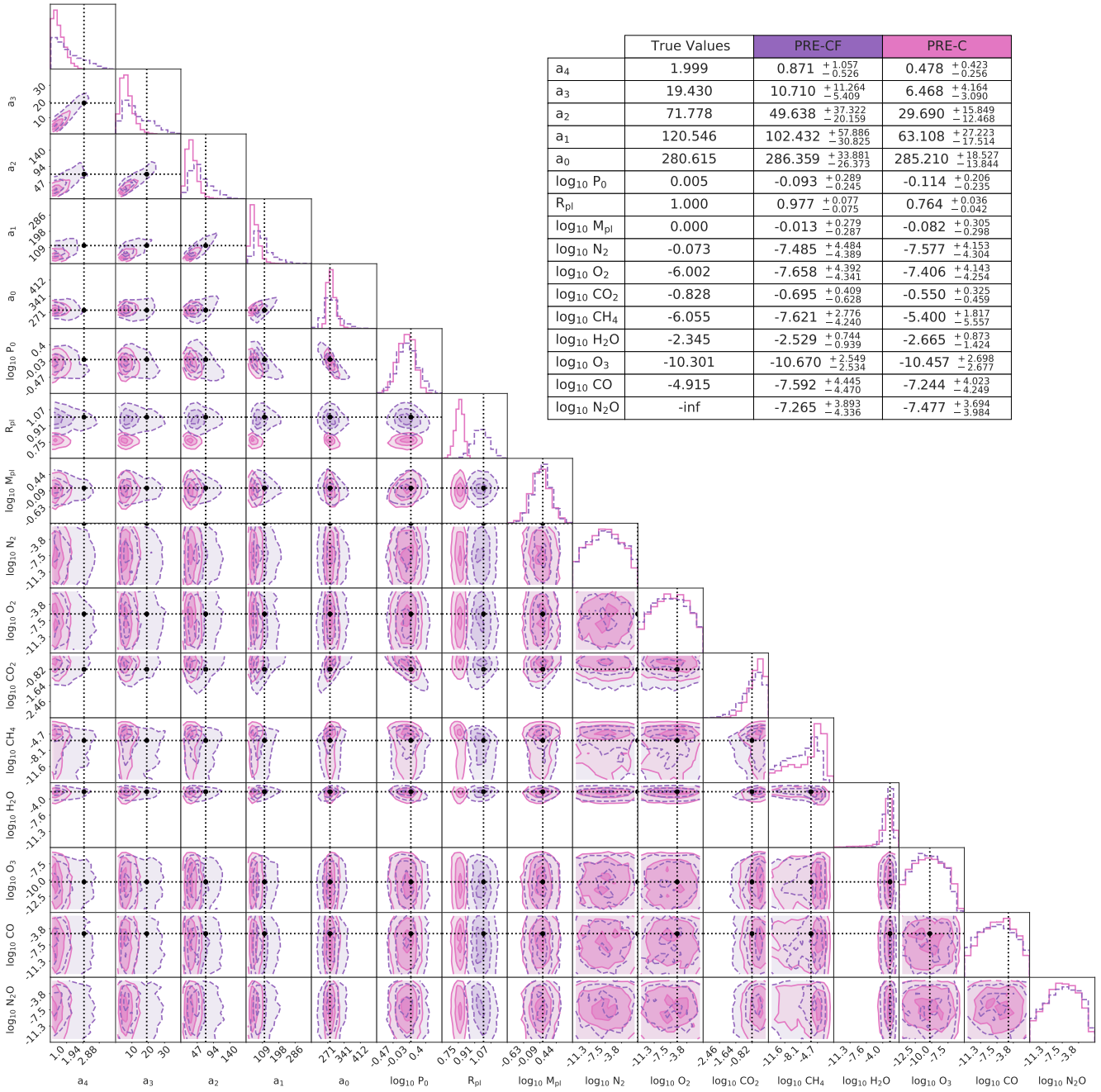|	True Values	PRE-CF	PRE-C
a_4	1.999	$0.871\ ^{+1.057}_{-0.526}$	$0.478\ ^{+0.423}_{-0.256}$
a_3	19.430	$10.710\ ^{+11.264}_{-5.409}$	$6.468\ ^{+4.164}_{-3.090}$
a_2	71.778	$49.638\ ^{+37.322}_{-20.159}$	$29.690\ ^{+15.849}_{-12.468}$
a_1	120.546	$102.432\ ^{+57.886}_{-30.825}$	$63.108\ ^{+27.223}_{-17.514}$
a_0	280.615	$286.359\ ^{+33.881}_{-26.373}$	$285.210\ ^{+18.527}_{-13.844}$
$\log_{10} P_0$	0.005	$-0.093\ ^{+0.289}_{-0.245}$	$-0.114\ ^{+0.206}_{-0.235}$
R_{pl}	1.000	$0.977\ ^{+0.077}_{-0.075}$	$0.764\ ^{+0.036}_{-0.042}$
$\log_{10} M_{pl}$	0.000	$-0.013\ ^{+0.279}_{-0.287}$	$-0.082\ ^{+0.305}_{-0.298}$
$\log_{10} N_2$	-0.073	$-7.485\ ^{+4.484}_{-4.389}$	$-7.577\ ^{+4.153}_{-4.304}$
$\log_{10} O_2$	-6.002	$-7.658\ ^{+4.392}_{-4.341}$	$-7.406\ ^{+4.143}_{-4.254}$
$\log_{10} CO_2$	-0.828	$-0.695\ ^{+0.409}_{-0.628}$	$-0.550\ ^{+0.325}_{-0.459}$
$\log_{10} CH_4$	-6.055	$-7.621\ ^{+2.776}_{-4.240}$	$-5.400\ ^{+1.817}_{-5.557}$
$\log_{10} H_2O$	-2.345	$-2.529\ ^{+0.744}_{-0.939}$	$-2.665\ ^{+0.873}_{-1.424}$
$\log_{10} O_3$	-10.301	$-10.670\ ^{+2.549}_{-2.534}$	$-10.457\ ^{+2.698}_{-2.677}$
$\log_{10} CO$	-4.915	$-7.592\ ^{+4.445}_{-4.470}$	$-7.244\ ^{+4.023}_{-4.249}$
$\log_{10} N_2O$	-inf	$-7.265\ ^{+3.893}_{-4.336}$	$-7.477\ ^{+3.694}_{-3.984}$

Fig. B.4: As for Fig. B.1 but for the PRE-CF and PRE-C scenarios.

Appendix C: Cloudy scenarios: Additional figures

In this section we provide additional plots for the cloudy scenarios. In Figs. C.1 and C.2 we show the retrieved exoplanet parameters and abundances for the different scenarios with varying R and S/N values. Finally, we plot in Fig. C.3 the maximum difference Δ between the cumulative posteriors for the different model parameters, for each combination of the cloudy scenarios and different R-S/N pairs.
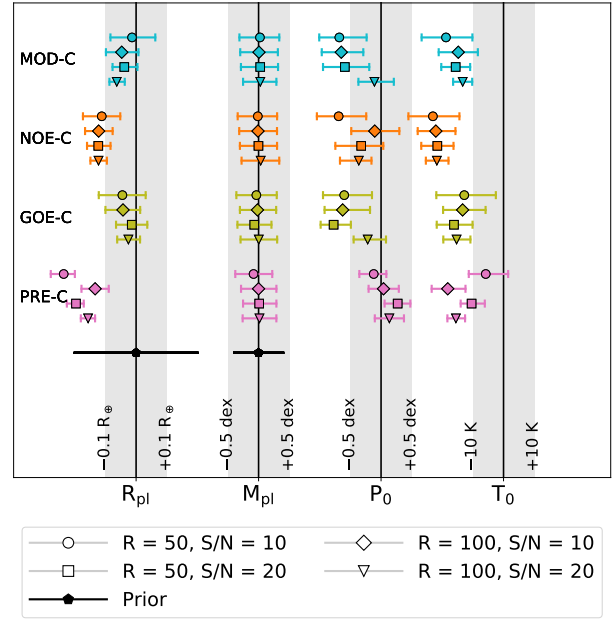

Fig. C.1: As for Fig. 6 but for the cloudy scenarios.
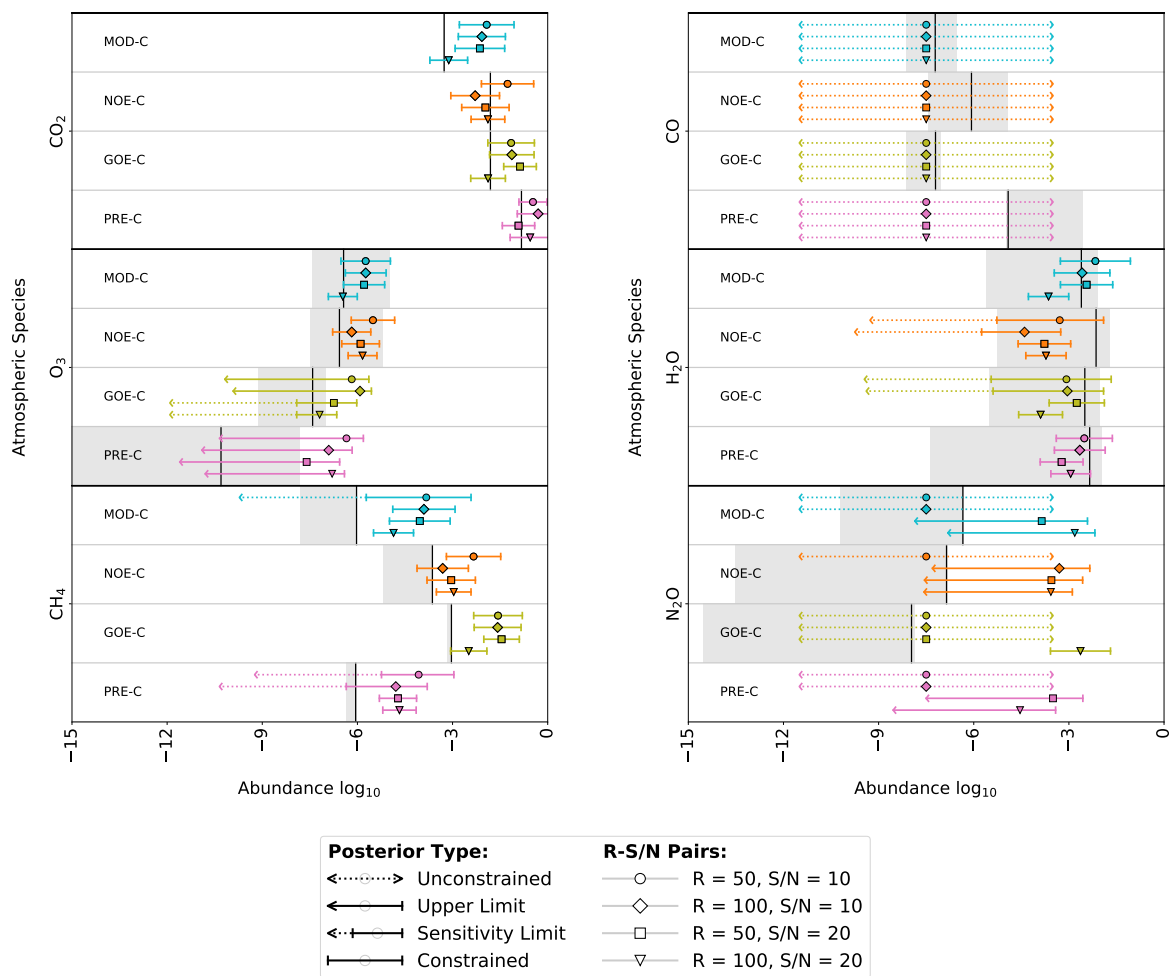
Fig. C.2: As for Fig. 7 but for the cloudy scenarios.
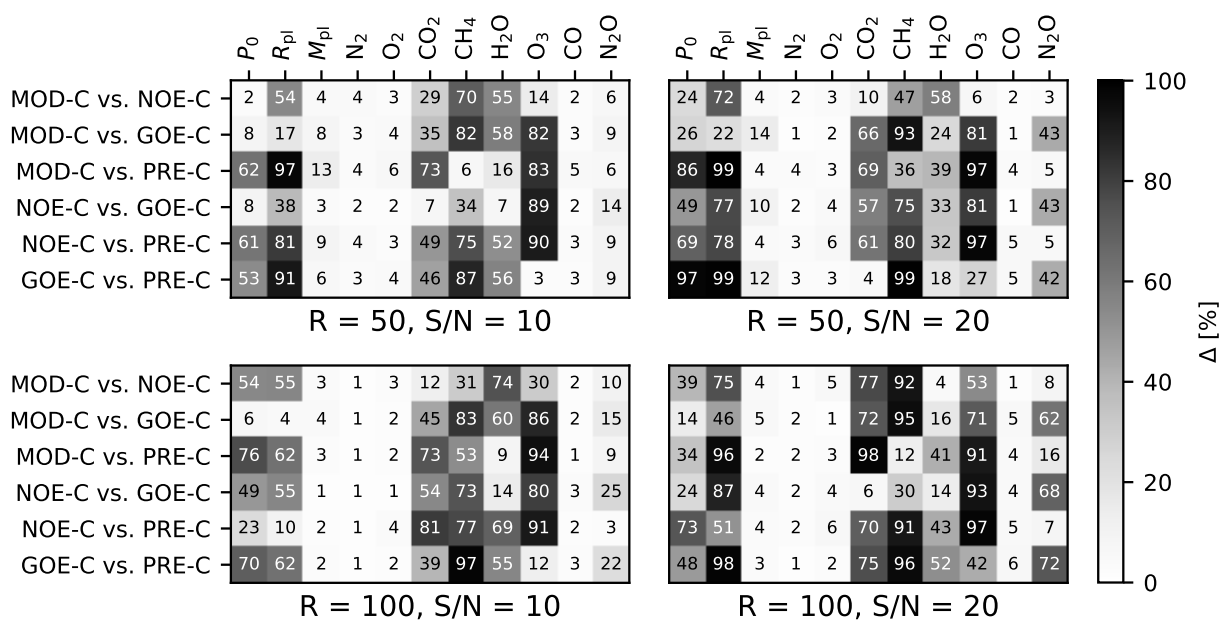
Fig. C.3: As for Fig. 11, but for the cloudy scenarios.

Appendix D: Bayes factor analysis: Other epochs

As described in Sect. 2, the theoretical spectral model was updated with respect to Paper III and it now takes into account additional physical processes. For the results presented in Sect. 3, we ran retrievals using the most updated version of the Bayesian framework.

The additional flexibility of `petitRADTRANS` now allows us to quantify the impact of CIA and scattering in retrievals. We ran additional retrievals on the clear sky scenarios for $R = 50$ and $S/N = 10$. In these retrievals, we altered the number of physical processes that were treated in the `petitRADTRANS` theoretical spectral model as follows: (1) including both CIA and scattering (setup used in Sect. 3); (2) excluding both CIA and scattering; (3) including scattering and excluding CIA; and (4) including CIA and excluding scattering.

In the runs where scattering is included, we consider self-scattering, surface scattering of the thermal radiation, and gaseous Rayleigh scattering (see Table 5 for references). We do not include aerosol and cloud scattering in the calculation. Since our theoretical spectral model neglects clouds, in this analysis we considered only the cloud-free scenarios. The effect of modeling cloudy spectra using a cloud-free retrieval model will be discussed in detail in Sect. 4.2.

To determine the theoretical spectral model configuration that best reproduces the input spectra we performed a Bayes factor analysis. The Bayes factor is defined as

$$K = \mathcal{Z}_{\mathcal{M}_1}(\mathcal{D})/\mathcal{Z}_{\mathcal{M}_2}(\mathcal{D}), \tag{D.1}$$

where $\mathcal{M}_1$ and $\mathcal{M}_2$ represent two different model configurations, each with their corresponding Bayesian evidence $\mathcal{Z}_{\mathcal{M}_i}(\mathcal{D})$ given the input data $\mathcal{D}$. In the case of Eq. D.1, the Bayes factor provides an indication whether $\mathcal{M}_1$ or $\mathcal{M}_2$ better describes the data. We can use the Jeffreys scale (see Table D.1) to interpret the values of the Bayes factor K. This approach was extensively described and used in Paper III, to which we refer for more details.

We calculate the Bayes factor corresponding to every possible combination of the four different setups previously outlined. We summarize the results obtained for the MOD-CF epoch in Fig. D.1.

Here, the diagonal shows the values of the Bayesian evidence $\mathcal{Z}_{\mathcal{M}_i}(\mathcal{D})$ of each of the four setups $\mathcal{M}_i$. The triangular matrix of six boxes below the diagonal is instead filled with the logarithm of the Bayes factors K (see Eq. D.1) for each combination of theoretical spectral model setups, as well as their interpretation according to the Jeffreys scale (Table D.1). The cells are color-coded according to the color bar in the lower panel, whose edges are determined by the Jeffreys scale.

Table D.1: Jeffreys scale (Jeffreys 1998).

$\log_{10}(K)$	Strength of evidence
$(-\infty, -2]$	Decisive support for $\mathcal{M}_2$
$(-2, -1]$	Strong support for $\mathcal{M}_2$
$(-1, -0.5]$	Substantial support for $\mathcal{M}_2$
$(-0.5, 0]$	Very weak support for $\mathcal{M}_2$
$(0, 0.5)$	Very weak support for $\mathcal{M}_1$
$[0.5, 1)$	Substantial support for $\mathcal{M}_1$
$[1, 2)$	Strong support for $\mathcal{M}_1$
$[2, \infty)$	Decisive support for $\mathcal{M}_1$

Notes. Scale for the interpretation of the Bayes factor $K = \mathcal{Z}_{\mathcal{M}_1}(\mathcal{D})/\mathcal{Z}_{\mathcal{M}_2}(\mathcal{D})$. Adapted from Paper III.

The blue and red color range adopted in the color bar was chosen deliberately in order to illustrate that, for any pair of models M_1 and M_2, a redder shade would mean that M_1 is more likely to reproduce the data compared to M_2, while a bluer shade would instead prefer M_2 over M_1. Our results show colors that lie somewhere in the middle of the range of the color bar, which corresponds to a value of $\log_{10}(K)$ generally very close to 0, as confirmed by the text within the cells. This means that there is no clear preference for any of the tested setups: we find no evidence that one of the considered setups outperforms the others in describing the input data. Including or excluding CIA and scattering (one or both) results in negligible differences in the retrieval results. This means that CIA in the spectra or/and spectral features induced by scattering are not detectable in retrieval studies at the considered R and S/N of the input. This analysis shows us that it is justifiable to neglect CIA and scattering in MIR retrievals of spectra with $R = 50$ and $S/N = 10$ (the minimum requirements for LIFE determined by Paper III), with negligible loss in the quality of the retrieval results.

The results for the remaining epochs exhibit similar behavior. In Fig. D.2 we show the results for the NOE Earth, Fig. D.3 shows the ones for the GOE Earth, and the results for the prebiotic Earth are shown in Fig. D.4.
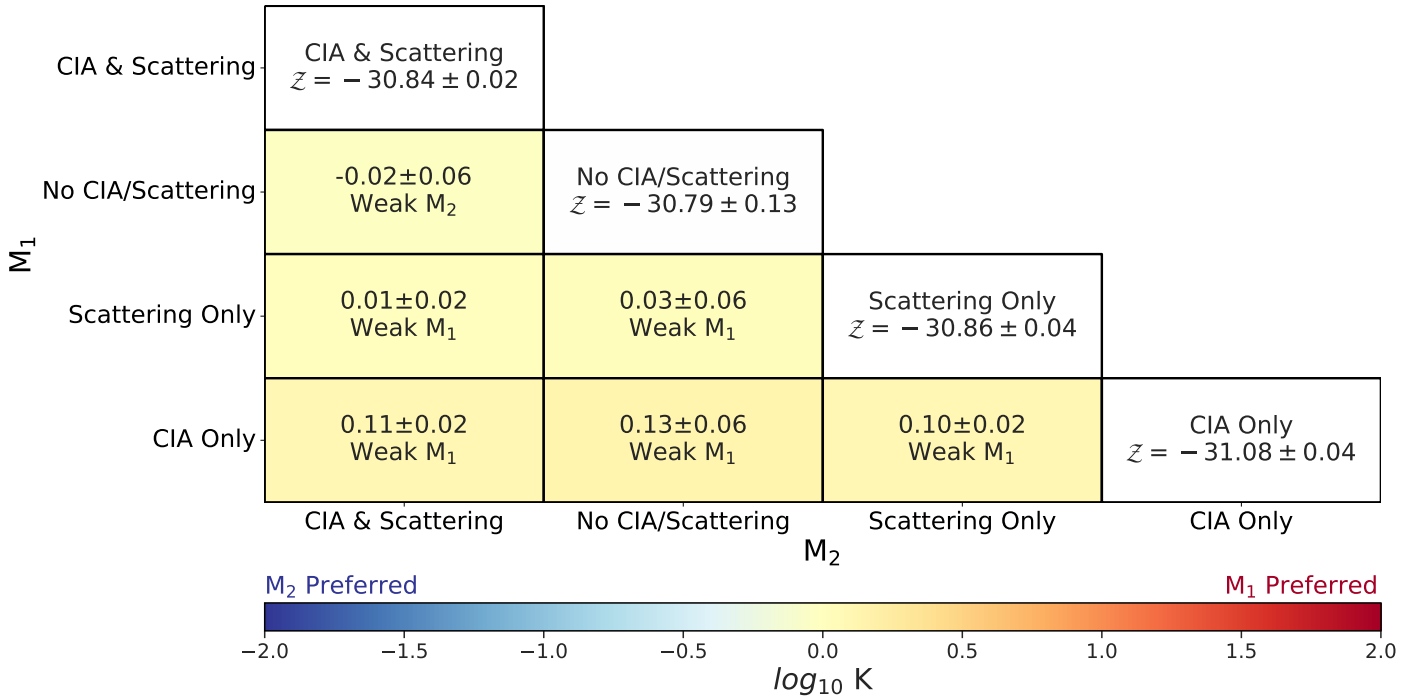
Fig. D.1: Bayesian evidence, $\mathcal{Z}_{\mathcal{M}_i}$, for each setup for the MOD-CF scenario (on the diagonal) and the Bayes factor for every pair of retrieval setups for the MOD-CF scenario (lower triangle). The cells in the lower triangle are color-coded according to the color bar, whose limits are determined by the Jeffreys scale (see Table D.1).
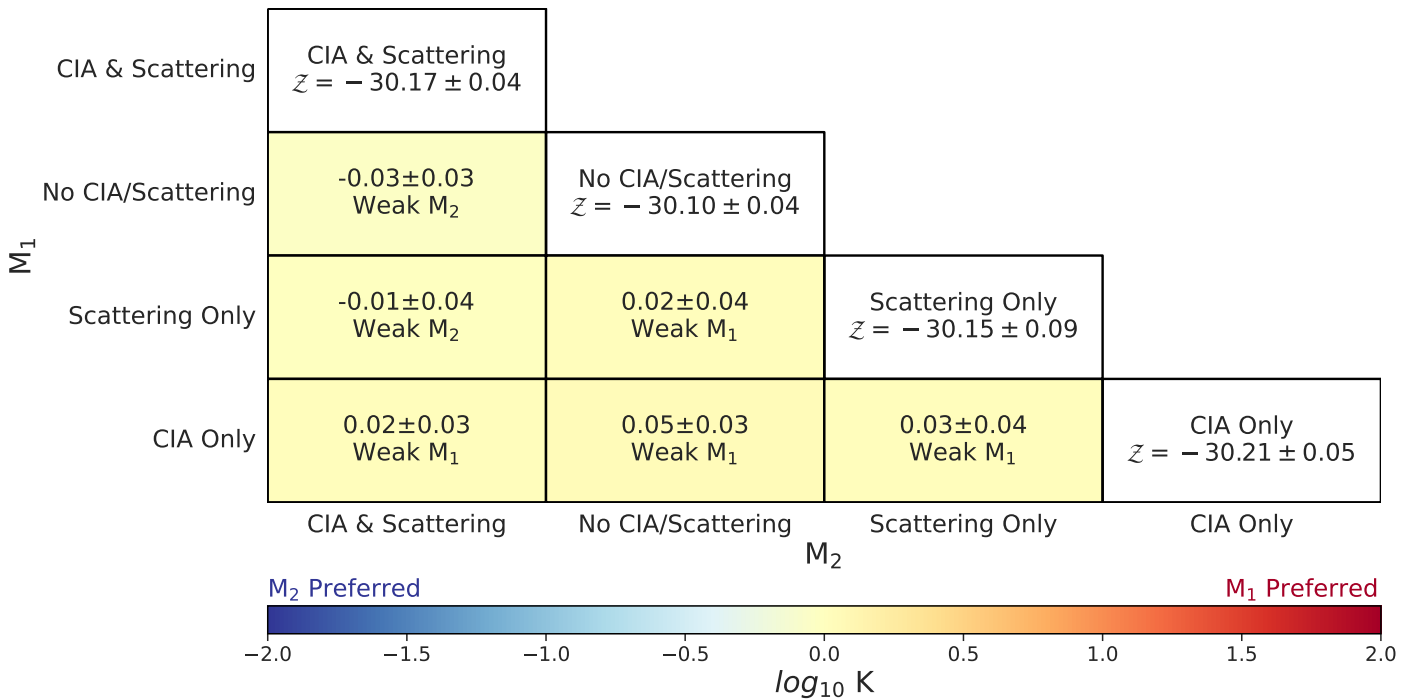

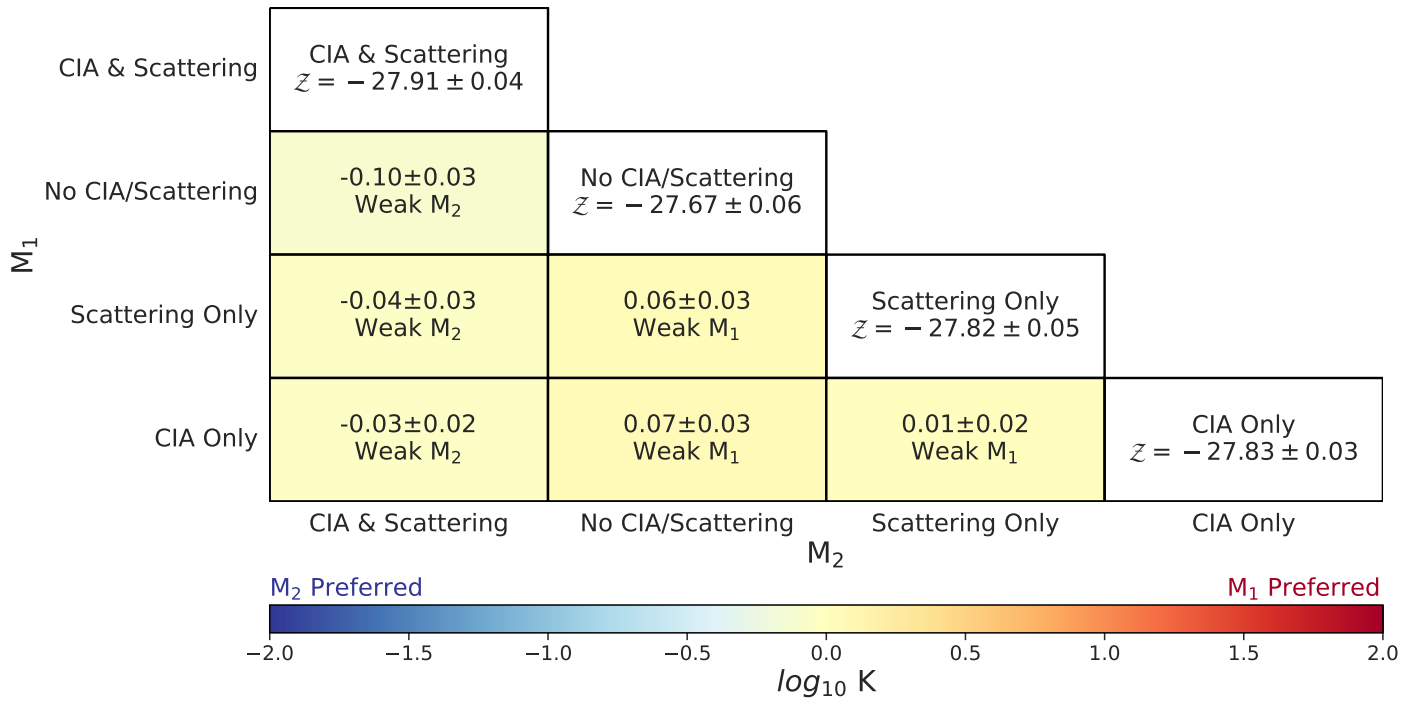Fig. D.2: As for Fig. D.1 but for the NOE-CF scenario.
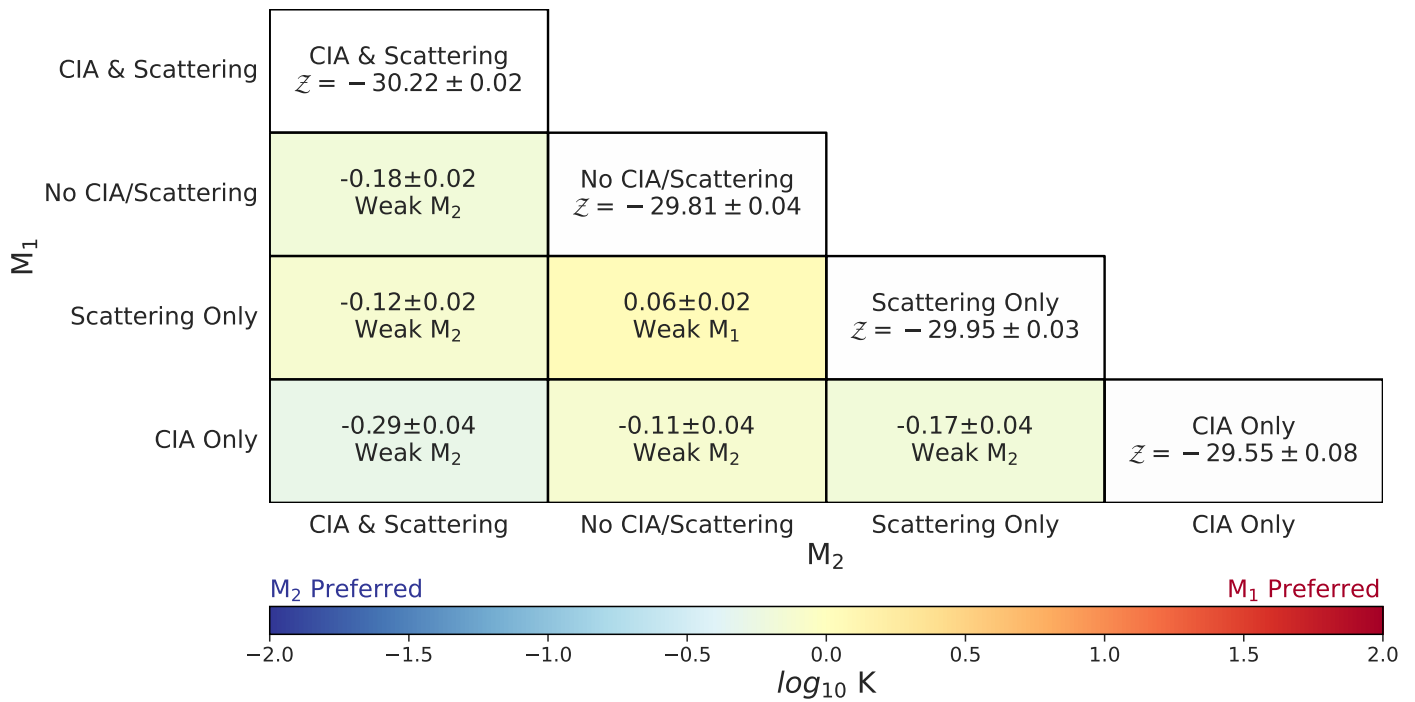
Fig. D.3: As for Fig. D.1 but for the GOE-CF scenario.

Fig. D.4: As for Fig. D.1 but for the PRE-CF scenario.