

Large Interferometer For Exoplanets (LIFE)

IX. Assessing the impact of clouds on atmospheric retrievals at mid-infrared wavelengths with a Venus-twin exoplanet[★]

B. S. Konrad^{1,2}, E. Alei^{1,2}, S. P. Quanz^{1,2,3}, P. Mollière⁴, D. Angerhausen^{1,2}, J. J. Fortney⁵, K. Hakim^{6,7}, S. Jordan⁸, D. Kitzmann^{2,9}, S. Rugheimer¹⁰, O. Shorttle^{8,11}, R. Wordsworth¹², and the LIFE Collaboration

¹ ETH Zurich, Institute for Particle Physics & Astrophysics, Wolfgang-Pauli-Str. 27, 8093 Zurich, Switzerland
e-mail: konradb@ethz.ch; sascha.quanz@phys.ethz.ch

² National Center of Competence in Research PlanetS, Gesellschaftsstrasse 6, 3012 Bern, Switzerland

³ ETH Centre for Origin and Prevalence of Life, Wolfgang-Pauli-Str. 27, 8093 Zurich, Switzerland

⁴ Max-Planck-Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany

⁵ Department of Astronomy and Astrophysics, University of California, Santa Cruz, CA 95064, USA

⁶ KU Leuven, Institute of Astronomy, Celestijnenlaan 200D, 3001 Leuven, Belgium

⁷ Royal Observatory of Belgium, Ringlaan 3, 1180 Brussels, Belgium

⁸ Institute of Astronomy, University of Cambridge CB3 0HA, UK

⁹ University of Bern, Center for Space and Habitability, Gesellschaftsstrasse 6, 3012 Bern, Switzerland

¹⁰ Department of Physics and Astronomy, York University, 4700 Keele Street, North York, Ontario 3MJ 1P3, Canada

¹¹ Department of Earth Sciences, University of Cambridge, Cambridge CB2 3EQ, UK

¹² School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA

Received 9 December 2022 / Accepted 13 March 2023

ABSTRACT

Context. Terrestrial exoplanets in the habitable zone are likely a common occurrence. The long-term goal is to characterize the atmospheres of dozens of such objects. The Large Interferometer For Exoplanets (LIFE) initiative aims to develop a space-based mid-infrared (MIR) nulling interferometer to measure the thermal emission spectra of such exoplanets.

Aims. We investigate how well LIFE could characterize a cloudy Venus-twin exoplanet. This allows us to: (1) test our atmospheric retrieval routine on a realistic non-Earth-like MIR emission spectrum of a known planet, (2) investigate how clouds impact retrievals, and (3) further refine the LIFE requirements derived in previous Earth-centered studies.

Methods. We ran Bayesian atmospheric retrievals for simulated LIFE observations of a Venus-twin exoplanet orbiting a Sun-like star located 10 pc from the observer. The LIFESIM noise model accounted for all major astrophysical noise sources. We ran retrievals using different models (cloudy and cloud-free) and analyzed the performance as a function of the quality of the LIFE observation. This allowed us to determine how well the atmosphere and clouds are characterizable depending on the quality of the spectrum.

Results. At the current minimal resolution ($R = 50$) and signal-to-noise ($S/N = 10$ at $11.2 \mu\text{m}$) requirements for LIFE, all tested models suggest a CO_2 -rich atmosphere ($\geq 30\%$ in mass fraction). Further, we successfully constrain the atmospheric pressure-temperature (P - T) structure above the cloud deck (P - T uncertainty ± 15 K). However, we struggle to infer the main cloud properties. Further, the retrieved planetary radius (R_{pl}), equilibrium temperature (T_{eq}), and Bond albedo (A_{B}) depend on the model. Generally, a cloud-free model performs best at the current minimal quality and accurately estimates R_{pl} , T_{eq} , and A_{B} . If we consider higher quality spectra (especially $S/N = 20$), we can infer the presence of clouds and pose first constraints on their structure.

Conclusions. Our study shows that the minimal R and S/N requirements for LIFE suffice to characterize the structure and composition of a Venus-like atmosphere above the cloud deck if an adequate model is chosen. Crucially, the cloud-free model is preferred by the retrieval for low spectral qualities. We thus find no direct evidence for clouds at the minimal R and S/N requirements and cannot infer the thickness of the atmosphere. Clouds are only constrainable in MIR retrievals of spectra with $S/N \geq 20$. The model dependence of our retrieval results emphasizes the importance of developing a community-wide best-practice for atmospheric retrieval studies.

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

1. Introduction

One major goal for the future of exoplanet science is to constrain the atmospheric structure and composition of a statistically significant number of terrestrial exoplanets. Special attention will be given to planets within or close to the habitable zone (HZ; [Kasting et al. 1993](#); [Kopparapu et al. 2013](#)) of the host star. Such exoplanets are expected to be common within our galaxy ([Petigura et al. 2013](#); [Foreman-Mackey et al. 2014](#); [Dressing &](#)

[Charbonneau 2015](#); [Bryson et al. 2020](#)), and have been detected within 20 pc of the Sun (e.g., [Anglada-Escudé et al. 2016](#); [Gillon et al. 2016, 2017](#); [Gilbert et al. 2020](#)). A powerful approach to characterize exoplanets is to analyze their spectra, which contain important information about relevant properties such as the atmospheric pressure-temperature (P - T) structure, the chemical composition, and the possible existence of clouds and their properties. If and how well an exoplanet property can be constrained depends on the wavelength regime covered by the spectrum and the accuracy with which the spectrum is measured.

[★] Webpage: <http://www.life-space-mission.com>

For terrestrial exoplanets orbiting their host star close to or within the HZ, detections are challenging, but possible, with current and approved future ground- and space-based observatories. However, these instruments will not be capable of obtaining detailed spectroscopic measurements for several dozens of such terrestrial exoplanets. Partially motivated by this goal, there is great interest in the community to develop a new generation of observatories. HabEx (Gaudi et al. 2020) and LUVOIR (The LUVOIR Team 2019), two flagship mission concepts that aim to directly detect and characterize HZ terrestrial exoplanets in reflected light (at ultraviolet, optical, and near-infrared or UV/O/NIR wavelengths), were evaluated in the Astro 2020 Decadal Survey in the United States (National Academies of Sciences, Engineering, and Medicine 2021). As a result, the space-based, UV/O/NIR flagship Habitable Worlds Observatory (HWO) was recommended. Additionally, the Voyage 2050 plan of the European Space Agency (ESA; Voyage 2050 Senior Committee 2021) recommended considering a large-scale, mid-infrared (MIR), space-based mission to characterize HZ terrestrial exoplanets via their thermal emission. The Large Interferometer For Exoplanets (LIFE) initiative aims to achieve this goal using a space-based MIR nulling interferometer (Kammerer & Quanz 2018; Quanz et al. 2021, 2022).

A first step in the LIFE design phase is to derive the requirements necessary to adequately characterize the atmospheres of nearby HZ terrestrial exoplanets. This includes constraining the wavelength coverage, spectral resolution, and instrument sensitivity. Previous studies in the LIFE series (Konrad et al. 2022; Alei et al. 2022a, hereafter Paper III and Paper V) derive first estimates for the required spectral quality. These studies use atmospheric retrievals (for recent reviews on retrievals, see e.g., Madhusudhan 2018; Deming et al. 2018; Barstow & Heng 2020) to derive quantitative estimates for important atmospheric and planetary parameters from a simulated or observed exoplanet spectrum. Both studies focus on characterizing Earth-like planets (Paper III – modern Earth, Paper V – Earth at various stages of its evolution; both assume the LIFESIM observation noise simulator from Dannert et al. 2022, hereafter Paper II). However, a future observatory should not only be able to characterize Earth-like exoplanets, but also discern Earth-like from non-Earth-like HZ exoplanets. In addition to being Earth-centric, our previous studies do not systematically investigate the effect of clouds on exoplanet characterization. Yet, since clouds influence an exoplanet’s spectrum (e.g., Kitzmann et al. 2011; Rugheimer et al. 2013; Vasquez et al. 2013; Komacek et al. 2020; Feinstein et al. 2023), a more detailed study of the impact of clouds is required to derive robust requirements for LIFE.

Earth-centered retrieval studies and retrieval studies on theoretical spectra of habitable worlds are often used to investigate the characterization performance for different quality spectra (e.g., von Paris et al. 2013; Brandt & Spiegel 2014; Feng et al. 2018; Léger et al. 2019; Carrión-González et al. 2020; Quanz et al. 2021; Robinson & Salvador 2023). Venus – to our knowledge – has not yet been considered in a comparable retrieval study. However, terrestrial exoplanets with a Venus-like insolation could maintain habitable conditions if a surface ocean is present (e.g., Yang et al. 2014; Way et al. 2016)¹. Further, exoplanets at the inner edge of the HZ (i.e., potentially Venus-like planets) are ideal targets for LIFE (Quanz et al. 2022). Finally, in

contrast to theoretical planet models, Venus and its atmosphere are a known outcome of planet formation and evolution and thus provide a realistic ground-truth for a retrieval study.

Despite Venus being Earth-like in size and mass, its atmospheric state and surface conditions are vastly different. In addition to a CO₂ dominated atmosphere, the mean surface pressure reaches 93 bar, significantly exceeding that on Earth. Further, at atmospheric pressures of 0.05 bar to 1 bar, a layer of opaque H₂SO₄ clouds covers the planet. This opaque cloud layer can lead to an ambiguity in characterization between thick and cloudy or tenuous and cloud-free atmospheres (Barstow et al. 2016; Lustig-Yaeger et al. 2019). Further, the detectability of H₂SO₄ clouds is of great interest, since it could provide constraints on the amount of liquid surface water (Loftus et al. 2019). Finally, the high atmospheric CO₂ content leads to strong atmospheric greenhouse heating, which raises the mean surface temperature to a hostile 730 K (for a recent review on Venus’ atmosphere, see, e.g., Taylor et al. 2018).

In this study, we reevaluate the LIFE requirements for wavelength range, spectral resolution (R), and signal-to-noise ratio (S/N) from Paper III. We focus on the key science application of distinguishing a cloudy Venus-like planet from an Earth-like planet. To this purpose, we studied a Venus-twin exoplanet using a modified version of our retrieval framework. The opaque cloud layer in Venus’ atmosphere required us to model clouds in our retrievals. This also allowed us to investigate how clouds affect exoplanet characterization with LIFE. Hence, this study improves the robustness of our instrument requirements and provides new insights into difficulties in exoplanet characterization via atmospheric retrievals.

In Sect. 2, we introduce the model used to simulate the Venus-twin MIR emission spectrum, our retrieval framework, and the LIFESIM observation noise model. The retrieval results for different quality spectra are presented in Sect. 3. In Sect. 4, we discuss implications of our results for the LIFE requirements. Important takeaway points are summarized in Sect. 5.

2. Methods

In Sect. 2.1, we introduce the atmosphere model used to simulate Venus’ MIR thermal emission spectrum and compare our spectrum to the literature. Next, we introduce our Bayesian retrieval routine and discuss updates with respect to previous versions (Sect. 2.2). In Sect. 2.3, we introduce the noise model used to generate the input spectra for the retrievals, the different atmospheric models fitted in the retrieval, and the assumed model parameter priors.

2.1. Cloudy Venus-twin model

As in Papers III and V, we used the 1D radiative transfer code `petitRADTRANS` (Mollière et al. 2019, 2020; Paper V) to model the MIR thermal emission spectrum of our Venus-twin exoplanet. `petitRADTRANS` passes a featureless black-body spectrum at the surface temperature through discrete atmospheric layers and models the interaction of each layer with the radiation. Further, it accounts for the scattering of photons by the atmosphere and the surface. This yields the MIR emission spectrum at the top of the atmosphere. Each layer is characterized by its temperature, pressure, and the opacity sources that are present. We provide a list of all model parameters along with the assumed true values in Table 1.

¹ Yet, the existence of large bodies of surface water on early Venus and Venus-like exoplanets is uncertain and heavily debated (e.g., Kasting & Harman 2021; Turbet et al. 2021). If Venus ever had liquid surface water, it was lost in a runaway greenhouse process (Kasting 1988).

Table 1. Parameters used in the different (retrieval) models, assumed true values, and prior distributions.

Parameter	Description	Truth	Prior	Model configuration			
				$\mathcal{M}_{\text{Op.}}^{\text{H}_2\text{SO}_4}$	$\mathcal{M}_{\text{Tr.}}^{\text{H}_2\text{SO}_4}$	$\mathcal{M}_{\text{Op.}}^{\text{H}_2\text{O}}$	\mathcal{M}^{CF}
a_3	P - T parameter (degree 3)	2.30 ⁽¹⁾	$\mathcal{U}(0, 10)$	✓	✓	✓	✓
a_2	P - T parameter (degree 2)	29.83 ⁽¹⁾	$\mathcal{U}(0, 500)$	✓	✓	✓	✓
a_1	P - T parameter (degree 1)	125.64 ⁽¹⁾	$\mathcal{U}(0, 1000)$	✓	✓	✓	✓
a_0	P - T parameter (degree 0)	344.94 ⁽¹⁾	$\mathcal{U}(0, 1000)$	✓	✓	✓	✓
$\log_{10}(P_0)$	Surface pressure [bar]	1.97 ⁽²⁾	$\mathcal{U}(-4, 3)$	×	✓	×	✓
R_{pl}	Planet radius [R_{\oplus}]	0.95 ⁽²⁾	$\mathcal{G}(0.95, 0.20)$	✓	✓	✓	✓
$\log_{10}(M_{\text{pl}})$	Planet mass [M_{\oplus}]	-0.09 ⁽²⁾	$\mathcal{G}(-0.09, 0.40)$	✓	✓	✓	✓
$\log_{10}(\text{CO}_2)$	CO ₂ mass fraction	-0.015 ⁽²⁾	$\log_{10}(1 - 10^{\mathcal{U}(-7,0)})$	✓	✓	✓	✓
$\log_{10}(\text{H}_2\text{O})$	H ₂ O mass fraction	-4.699 ⁽²⁾	$\mathcal{U}(-7, 0)$	✓	✓	✓	✓
$\log_{10}(\text{CO})$	CO mass fraction	-4.770 ⁽²⁾	$\mathcal{U}(-7, 0)$	✓	✓	✓	✓
$\log_{10}(\text{H}_2\text{SO}_4^{\text{cloud}})$	H ₂ SO ₄ cloud mass fraction	-5.097 ⁽³⁾	$\mathcal{U}(-7, -1)$	✓	✓	×	×
$\log_{10}(\text{H}_2\text{O}^{\text{cloud}})$	H ₂ O cloud mass fraction	-	$\mathcal{U}(-7, -1)$	×	×	✓	×
$\log_{10}(P_{\text{top}}^{\text{cloud}})$	Cloud top pressure [bar]	-1.30 ⁽⁴⁾	$\mathcal{U}(-6, 3)$	✓	✓	✓	×
$\log_{10}(P_{\text{span}}^{\text{cloud}})$	Cloud thickness [bar]	0.161 ⁽⁴⁾	$\mathcal{U}(-6, 3)$	✓	✓	✓	×
$\log_{10}(\bar{R}^{\text{cloud}})$	Mean cloud particle radius [cm]	-4.30 ⁽⁴⁾	$\mathcal{U}(-8, -3)$	✓	✓	✓	×
σ^{cloud}	Log-normal particle size spread	1.95 ⁽⁴⁾	$\mathcal{U}(1, 5)$	✓	✓	✓	×

Notes. In the third column we provide the values used to generate the Venus-twin spectrum introduced in Sect. 2.1. These values provide the ground truth for all retrievals. The fourth column lists the prior distributions used in the retrievals. With $\mathcal{U}(x, y)$, we denote a boxcar prior with lower threshold x and upper threshold y ; $\mathcal{G}(\mu, \sigma)$ indicates a Gaussian prior with mean μ and standard deviation σ . The last four columns summarize the model parameters used by the different forward models in the retrievals (✓ = used, × = unused; see Sect. 2.3.2) and the model color-coding used throughout the paper. $\mathcal{M}_{\text{Op.}}^{\text{H}_2\text{SO}_4}$ – opaque H₂SO₄ clouds (84% H₂SO₄, 16% H₂O by weight); $\mathcal{M}_{\text{Tr.}}^{\text{H}_2\text{SO}_4}$ – transparent H₂SO₄ clouds (84% H₂SO₄, 16% H₂O by weight); $\mathcal{M}_{\text{Op.}}^{\text{H}_2\text{O}}$ – opaque H₂O clouds (100% H₂O); \mathcal{M}^{CF} – cloud-free.

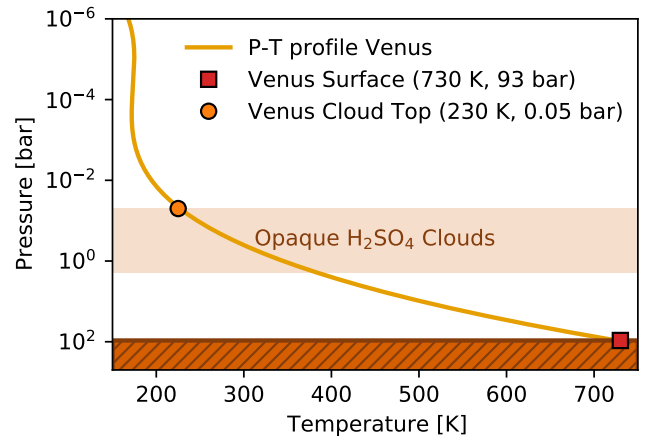
References. ⁽¹⁾Obtained by fitting Eq. (1) to the Venus P - T profile in Fig. 1 of Mueller-Wodarg et al. (2008); ⁽²⁾Near surface abundances from NASA’s planet factsheet: <https://nssdc.gsfc.nasa.gov/planetary/factsheet/venusfact.html>; ⁽³⁾Mean of values from Oschlisniok et al. (2012) and Krasnopolsky (2015); ⁽⁴⁾Mean values for the mode 1 cloud particles in Titov et al. (2018).

We parametrized Venus’ P - T structure of using a polynomial. In Papers III and V, this approach allowed us to minimize the number of model parameters and thus reduce the computational complexity of our retrieval. An extensive discussion, justifying this choice of P - T parametrization, is provided in the appendix of Paper III. Since, in contrast to Earth, Venus’ P - T profile does not exhibit a temperature inversion, a third order polynomial (four parameters) is sufficient for the present study:

$$T(P) = \sum_{i=0}^3 a_i P^i. \quad (1)$$

Here, P is the pressure and T the corresponding temperature of each atmospheric layer. The a_i terms are the parameters of the polynomial P - T model. In Fig. 1, we show our ground-truth Venus P - T profile, which we obtained by fitting the polynomial model to the Venus P - T profile from Mueller-Wodarg et al. (2008). The corresponding a_i values are given in Table 1.

To simulate the MIR emission of a Venus-twin, we accounted for different opacity sources. First, we modeled the MIR absorption and emission features of CO₂, H₂O, and CO (see Table 1 for the assumed mass fractions and Table 2 for the used line lists, broadening coefficients, and line cutoffs). For all three gases, we assumed constant vertical abundance profiles. Second, we modeled spectral features from collision-induced absorption (CIA) by CO₂, and Rayleigh scattering by all three molecules (see Table 3 for the used opacities). Third, we considered the opaque H₂SO₄ clouds (see Fig. 1), which is essential to accurately model the MIR thermal emission of Venus. We accounted for the H₂SO₄ clouds by adding a cloud slab to the atmosphere,


Fig. 1. Schematic structure of Venus’ atmospheric P - T profile and position of the opaque H₂SO₄ cloud layer.

which spanned multiple atmospheric layers. The cloud slab was characterized using five parameters: the pressure at the cloud-top $P_{\text{top}}^{\text{cloud}}$, the thickness of the cloud layer $P_{\text{span}}^{\text{cloud}}$ in bar, the mass fraction of the cloud forming substance, the mean cloud-particle radius \bar{R}^{cloud} , and the standard deviation σ^{cloud} of the log-normal cloud-particle size distribution. The parameter $P_{\text{top}}^{\text{cloud}}$ defined the uppermost atmospheric layer that contained clouds, while the difference $P_{\text{top}}^{\text{cloud}} - P_{\text{span}}^{\text{cloud}}$ the corresponding lowermost layer. Throughout the cloud slab defined by these two parameters, we assumed a constant mass fraction of the cloud forming

Table 2. Molecular line opacities used to calculate MIR spectra.

Molecule	Line list	Pressure broadening	Line cutoff
CO ₂	HITEMP	γ_{air}	BU69
H ₂ O	HITEMP	γ_{air}	HA02
CO	HITEMP	γ_{air}	HA02

References. HITEMP: Rothman et al. (2010); BU69: Burch et al. (1969); HA02: Hartmann et al. (2002).

Table 3. Continuum opacities used to calculate MIR spectra.

Opacity type	Material	Reference
CIA	CO ₂ –CO ₂	KA19
Cloud	H ₂ SO ₄ (liquid)	PW75
Cloud	H ₂ O (liquid)	SE81
Rayleigh	CO ₂	SU05
Rayleigh	H ₂ O	HA98
Rayleigh	CO	SU05

Notes. The cloud opacities were calculated from the indices of refraction of the cloud species via Mie scattering theory.

References. KA19: Karman et al. (2019); PW75: Palmer & Williams (1975); SE81: Segelstein (1981); SU05: Sneep & Ubachs (2005); HA98: Harvey et al. (1998).

H₂SO₄–H₂O solution (84% H₂SO₄, 16% H₂O by weight). All other atmospheric layers were modeled to be cloud-free (mass fraction of the cloud forming substance set to zero). Our cloud model further assumed homogeneous, spherical cloud particles of variable size. We assumed both \bar{R}^{cloud} and σ^{cloud} to be constant throughout the cloud deck. We calculated the pressure- and temperature-dependent opacities for the different cloud particle sizes from the wavelength-dependent index of refraction (see Table 3 for sources) using Mie scattering theory. For this calculation, we relied on the software presented in Min et al. (2005), which uses the codes from Toon & Ackerman (1981). Thereafter, we used the standard `petitRADTRANS` cloud modeling pathway to include clouds in the MIR Venus-twin spectrum.

In Fig. 2, we compare our MIR Venus-twin spectrum, to the simulated Venus spectra from NASA’s Planetary Spectrum Generator (PSG²; Villanueva et al. 2018) and Arney & Kane (2018). In contrast to our Venus-twin atmosphere, both models included additional atmospheric isotopes and trace gases and assumed altitude-dependent abundance profiles for all gases. Additionally, both models assumed a more complex atmospheric cloud structure (PSG: altitude-dependent volcanic clouds; Arney & Kane 2018: altitude-dependent multilayer H₂SO₄ solution clouds). We observe that the general shape of all three spectra is comparable and the differences are smaller or of similar magnitude as the assumed noise level. Furthermore, our MIR Venus-twin spectrum is not missing any significant spectral absorption features, despite not taking into account various atmospheric species and isotopes. There are two possible explanations for this finding: either these species have no significant spectral lines in the MIR (e.g., O₂ and N₂) or their atmospheric abundance above the opaque cloud layer is too low to cause a noticeable signature in Venus’ MIR spectrum (e.g., SO₂ and O₃). This finding justifies our approach of excluding these additional molecules from

² <https://psg.gsfc.nasa.gov>

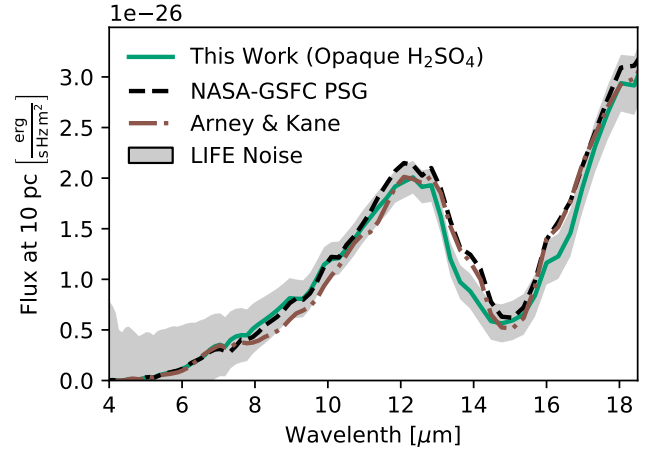


Fig. 2. Comparison of our Venus-twin MIR spectrum (opaque H₂SO₄ clouds) to other models. We show the flux received by an observer located 10 pc from the planet. The solid line is the MIR thermal emission spectrum of our model (calculated assuming the true values listed in Table 1). The gray-shaded region indicates the $S/N = 10$ LIFESIM uncertainty considered for our retrievals (see Sect. 2.3.1). The black-dashed line is the standard Venus emission spectrum from the NASA-GSFC Planetary Spectrum Generator (PSG, <https://psg.gsfc.nasa.gov>; Villanueva et al. 2018). The brown-dashed-dotted line is the MIR Venus spectrum from Arney & Kane (2018).

our Venus-twin model. Between 6 μm and 12 μm , we observe minor differences between the spectra. Since the spectrum in this wavelength range is predominantly determined by the clouds, the observed variance is likely rooted in the differences between the three cloud models. Additionally, we observe differences in the CO₂ absorption feature between 13 μm and 17 μm . While the PSG and the Arney & Kane (2018) models yield similar results, our model deviates. This deviance is most likely evoked by differences in the assumed P – T profiles, but might also be partially due to differences between the line lists, pressure broadening coefficients, or line cutoffs. However, since we are interested in assessing the impact of clouds on exoplanet characterization and whether the first requirements from Papers III and V are sufficient, the deviances are negligible for this study.

2.2. Bayesian retrieval framework

The study we present here relied on our Bayesian retrieval routine, which we first introduced in Paper III. Our routine relied on two subroutines. First, it applied the 1D radiative transfer code `petitRADTRANS` (Mollière et al. 2019, 2020; Paper V) to calculate the theoretical MIR spectrum corresponding to a given combination of values of the forward model parameters listed in Table 1. Second, the routine used `pyMultiNest` (Buchner et al. 2014), which efficiently samples the parameter space spanned by the prior probability distributions (or “priors”) of the forward model parameters to determine parameter combinations that fit the simulated Venus-twin observation well. This yielded posterior probability distributions (or “posteriors”) for the model parameters and estimates for the Bayesian evidence \mathcal{Z} of the model. The posteriors tell us how likely different combinations of model parameter values are. The evidence \mathcal{Z} measures how well the model fits the input spectrum and can be used for model selection (see Sect. 2.3.2). The `pyMultiNest` package is based on the `MultiNest` (Feroz et al. 2009) implementation of the Nested Sampling algorithm (Skilling 2006). In all retrievals we performed throughout this study, we ran `pyMultiNest` using

700 live points and a sampling efficiency of 0.3 (suggested for evidence evaluation by the MultiNest documentation³). An in-depth description of our atmospheric retrieval routine can be found in Paper III of the LIFE series. Here, we focus on the updates and improvements implemented since Paper III.

In Paper III, we did not consider the impact surface, atmospheric, and cloud scattering processes have on the MIR thermal emission. Including scattering in the forward model, would have significantly increased the required computation time per spectrum, making the study presented in Paper III unfeasible. However, in both Papers III and V, the surface and atmospheric scattering effects were shown to be negligible for the MIR thermal emission. Further, only cloud-free forward models were considered, which justified neglecting cloud scattering. However, for the cloudy Venus-twin retrievals we performed in this study, cloud scattering effects were no longer negligible. The limiting factor for the retrievals presented in Paper III was that `petitRADTRANS` calculated spectra at a predetermined resolution of $R = 1000$. Our retrieval routine then binned down the spectra to the resolution of the input spectrum. More recent improvements to `petitRADTRANS` enabled us to compute spectra directly at the resolution of the input spectrum (see Paper V for further information). With this update, the computation time required for one cloudy spectrum including scattering at $R = 50$ dropped from ≈ 20.0 s to just ≈ 1.5 s. This reduction in the computation time per spectrum enabled us to run retrievals accounting for scattering effects, making the present study feasible. We validate the updated retrieval routine in Appendix A.

2.3. Retrieval setup

In Sect. 2.3.1, we discuss the Venus-twin input spectrum and the noise model used in the retrievals. Thereafter, we introduce four atmospheric models we used as forward models during the retrievals (Sect. 2.3.2). Lastly, the prior distributions assumed in our retrievals are motivated in Sect. 2.3.3.

2.3.1. Input spectra and noise terms

We generated the Venus-twin input spectra for our retrieval study using `petitRADTRANS` and the cloudy Venus-twin model introduced in Sect. 2.1. We defined the resolution R of a spectrum as $\lambda/\Delta\lambda$, where λ was the wavelength at the center of a wavelength bin and $\Delta\lambda$ was the bin width. Further, we used noise models to estimate the wavelength dependent S/N. We defined the S/N of the input spectrum as the S/N at the 11.2 μm reference bin, since this bin did not coincide with strong absorption features from the considered atmospheric species.

For all LIFE retrievals in Sect. 3, we used the LIFESIM noise model introduced in Paper II. LIFESIM provides estimates for the wavelength-dependent S/N expected for observations with LIFE by accounting for noise contributions from the photon noise of the planet's emission, stellar leakage, and local- as well as exozodiacal dust emission. We hence implicitly assumed that a large future space mission like LIFE will not be dominated by instrumental noise terms (Dannert et al., in prep.). Possible consequences of this assumption are mentioned in Sect. 4.2. For our study, we assumed a Venus-twin exoplanet orbiting a G2V Star on an 0.72 AU orbit at a distance of 10 pc from the observer. We further set the exozodiacal dust emission to be three times the level of the local zodiacal light. This value corresponds to

the median exozodi level found for Sun-like stars in the HOSTS survey (Ertel et al. 2020).

In all retrievals, we interpreted the noise as uncertainty to the simulated spectral points and assumed that the noise does not impact the predicted flux values. As discussed in Feng et al. (2018) and the appendix of Paper III, randomizing the individual spectral data points according to the S/N would simulate more accurate observational instances. However, a retrieval study based on a single noise instance will result in biased estimates for the retrieval's characterization performance due to the random placement of the few spectral points. An ideal retrieval study should thus consider multiple (≥ 10) different noise realizations of each input spectrum and evaluate the instrument performance by considering the average retrieved parameter posterior. However, the vast number of different retrievals (12 retrievals for each of the four models introduced in Sect. 2.3.2 resulting in a total of 48 retrievals) we executed for this study and the average computation time per retrieval (~ 1 day on 20 CPUs) made such a study computationally unfeasible (≥ 5 months of total cluster time). In addition, in the Appendix of Paper III, we motivated that by retrieving the unrandomized input spectra we obtain reliable estimates for the average expected retrieval performance.

2.3.2. Atmospheric forward models in the retrievals

To test our retrieval framework's sensitivity for Venus' clouds, we analyzed how our routine performed for different atmospheric forward models. This approach enabled us to test if the LIFE design requirements from Papers III and V are sufficient to infer the presence of clouds and to accurately characterize the clouds in the atmosphere of a Venus twin. Additionally, this approach provided us with important new insights into the biases that arise when assuming an incorrect atmospheric model in a retrieval study. We considered four atmospheric forward models (see Table 1 for the parameter configuration of each model):

1. Opaque H_2SO_4 clouds – (14 parameters; $\mathcal{M}_{\text{Op}}^{\text{H}_2\text{SO}_4}$): As is true for Venus, we assumed that opaque H_2SO_4 clouds blocked the contributions from the lower atmospheric layers and surface to the outgoing MIR emission spectrum. By fixing the surface pressure P_0 to an arbitrary 10^4 bar, we forced the retrieval to add an opaque cloud layer to the atmosphere.

2. Transparent H_2SO_4 clouds – (15 parameters; $\mathcal{M}_{\text{Tr}}^{\text{H}_2\text{SO}_4}$): in contrast to the opaque model, we assumed that contributions from the lower atmosphere are not fully blocked. Therefore, we tried to retrieve for the surface pressure P_0 .

3. Opaque H_2O clouds – (14 parameters; $\mathcal{M}_{\text{Op}}^{\text{H}_2\text{O}}$): similar to the opaque H_2SO_4 model, we assumed an opaque cloud layer to be present and fixed the surface pressure P_0 to 10^4 bar. However, we assumed pure H_2O clouds to determine if we could identify the correct cloud species in retrievals.

4. Cloud-free – (10 parameters; \mathcal{M}^{CF}): we assumed no clouds to be present in the atmosphere. With this model, we investigated whether the presence of clouds in an atmosphere can be inferred at the considered input qualities.

We used Bayesian model selection to determine which model performed best as a function of the quality of the input spectrum. If we run two retrievals that assume different atmospheric models \mathcal{A} and \mathcal{B} , both retrieval results will be characterized by their evidences $\ln(\mathcal{Z}_{\mathcal{A}})$ and $\ln(\mathcal{Z}_{\mathcal{B}})$. We can use the evidences to identify the better fitting model via the Bayes factor K :

$$\log_{10}(K) = \frac{\ln(\mathcal{Z}_{\mathcal{A}}) - \ln(\mathcal{Z}_{\mathcal{B}})}{\ln(10)}. \quad (2)$$

³ <https://github.com/farhanferoz/MultiNest>

The Jeffreys scale (Jeffreys 1998, see Table 4) provides a possible interpretation for the value of the Bayes factor K .

2.3.3. Prior distributions

In Table 1, we provide a summary of the assumed prior distributions, which define the range of parameter space sampled by pyMultiNest. For the P – T parameters a_i and the surface pressure P_0 , we chose broad uniform priors such that the corresponding P – T profiles covered a wide range of atmospheric structures. For the abundances of the atmospheric species and the cloud parameters, we assumed broad and uniform priors that spanned large regions of parameter space. In contrast to Paper III, we assumed narrower abundance priors for H_2O and CO , since the lowest abundance detectable by our retrieval routine for the quality of input spectra considered is approximately 10^{-7} in mass fraction (cf. Paper III). For CO_2 , we used a prior that covers the full abundance range, but samples high abundances more densely. This prior allows us to better estimate the CO_2 abundance and better identify a potential upper limit.

As in Papers III and V, we chose a Gaussian prior for R_{pl} . For the mean, we assumed Venus’ true radius, for the standard deviation of 20% of the true value. This choice for R_{pl} was motivated by findings presented in Paper II, which demonstrated that the detection of a planet during LIFE’s search phase would yield such constraints for R_{pl} (for a terrestrial planet around the HZ, we expect a radius estimate R_{est} for the true radius R_{true} with $R_{\text{est}}/R_{\text{true}} = 0.97 \pm 0.18$). We then used the R_{pl} prior to derive a Gaussian prior for the planet mass $\log_{10}(M_{\text{pl}})$ using the statistical mass-radius relation Forecaster⁴ (Chen & Kipping 2016).

3. Retrieval results

Here, we present the retrieval results for the Venus-twin mock observations with LIFE for the different assumed forward models (see Sect. 2.3.2 and Table 1). In Sect. 3.1, we discuss the results obtained for a Venus-twin spectrum at the minimal LIFE requirements determined in Paper III (4 – 18.5 μm , $R = 50$, and $S/N = 10$). Thereafter, in Sect. 3.2, we discuss if and how the retrieval’s characterization performance is improved when considering higher quality spectra. To this purpose, we ran retrievals for various spectra of different wavelength coverage (4 – 18.5 μm , 3 – 20 μm), R (50, 100), and S/N (10, 15, 20).

3.1. Results for the current minimal LIFE requirements

We present the retrieval results for the different forward models at the minimal LIFE design requirements (4 – 18.5 μm , $R = 50$, $S/N = 10$) in Figs. 3 (spectrum residuals), 4 (P – T profile residuals), and 5 (posteriors). The full corner plots, the absolute retrieved P – T profiles and spectra, the wavelength- and pressure-dependent contribution to the emission spectrum, and tables with the retrieved values can be found in Appendix C.

3.1.1. Fit to the Venus-twin spectrum

The spectrum residuals in Fig. 3 show that the fit of all four forward models to the Venus-twin spectrum lies well within the LIFESIM noise level. This indicates that all models can reproduce the Venus-twin input with sufficient accuracy. Above 8 μm , the retrieved quantile envelopes of all forward models are similar, roughly centered on the truth, and smaller than the LIFESIM

Table 4. Jeffreys scale (Jeffreys 1998).

$\log_{10}(K)$	Probability	Strength of evidence
< 0	< 0.5	Support for \mathcal{B}
0–0.5	0.5–0.75	Very weak support for \mathcal{A}
0.5–1	0.75–0.91	Substantial support for \mathcal{A}
1–2	0.91–0.99	Strong support for \mathcal{A}
> 2	> 0.99	Decisive support for \mathcal{A}

Notes. Scale for interpreting the values of the Bayes’ factor $\log_{10}(K) = (\ln(\mathcal{Z}_{\mathcal{A}}) - \ln(\mathcal{Z}_{\mathcal{B}})) / \ln(10)$. The scale is symmetrical, i.e., negative values of $\log_{10}(K)$ correspond to very weak, substantial, strong, or decisive support for model \mathcal{B} .

noise. Below 8 μm , the quantile envelopes become larger and deviate more from the truth as the LIFESIM noise level increases. We discuss the origin of these deviations in Sect. 3.1.3. The spread of the quantiles is largest for the true model (opaque H_2SO_4 clouds). For the other models, the spread is smaller, but the residual deviates from the input. This indicates that the three models cannot reproduce the input accurately. However, due to the large LIFESIM errors in this wavelength range, these deviations will not affect the retrieval performance significantly.

3.1.2. Retrieved P – T structure

The P – T profile residuals in Fig. 4 show that the fit of all four models is best above Venus’ cloud layer (roughly between 10^{-1} bar and 10^{-4} bar). The retrieved means for the cloud-free and H_2SO_4 cloud forward models lie within ± 15 K of the truth. For the opaque H_2O cloud model, the deviations from the truth are larger (within ± 25 K). The uncertainty on the retrieved P – T structure in this pressure range is roughly ± 15 K for all four forward models. When considering higher or lower pressures, the deviations from the truth increase and the uncertainties grow. This behavior is due to a lack of significant spectral features from the high and low pressure atmospheric layers in the Venus-twin MIR spectrum (see emission contribution plots in Appendix C). The constraints on the percentile envelopes for these layers stem from extrapolation of the P – T model (nonphysical polynomial model). Thus, we cannot trust the P – T predictions for these atmospheric layers and cannot estimate Venus’ surface pressure P_0 and temperature T_0 accurately.

For the cloud-free and the transparent H_2SO_4 cloud model, the retrieved P_0 roughly corresponds to the position of the cloud-top in Venus’ atmosphere and the retrieved T_0 slightly overestimates the cloud-top temperature (by roughly 10 K for both models). Additionally, for the transparent H_2SO_4 cloud model, the large spread in the retrieved cloud-top pressure and temperature indicates that these two parameters are no longer well constrained. In contrast, $P_{\text{top}}^{\text{cloud}}$ is accurately retrieved and $T_{\text{top}}^{\text{cloud}}$ is slightly overestimated (by roughly 25 K for H_2O and 10 K for H_2SO_4 clouds) in the retrievals that assume opaque cloud forward models. Thus, we find accurate estimates for the position of the cloud-top in all retrievals.

3.1.3. Retrieved parameter posteriors

Lastly, we consider the retrieved posteriors displayed in Fig. 5. The figure further includes distributions for the planetary equilibrium temperature T_{eq} and Bond albedo A_{B} , which we derived from the posteriors using the method outlined in Appendix B.

⁴ <https://github.com/chenjj2/forecaster>

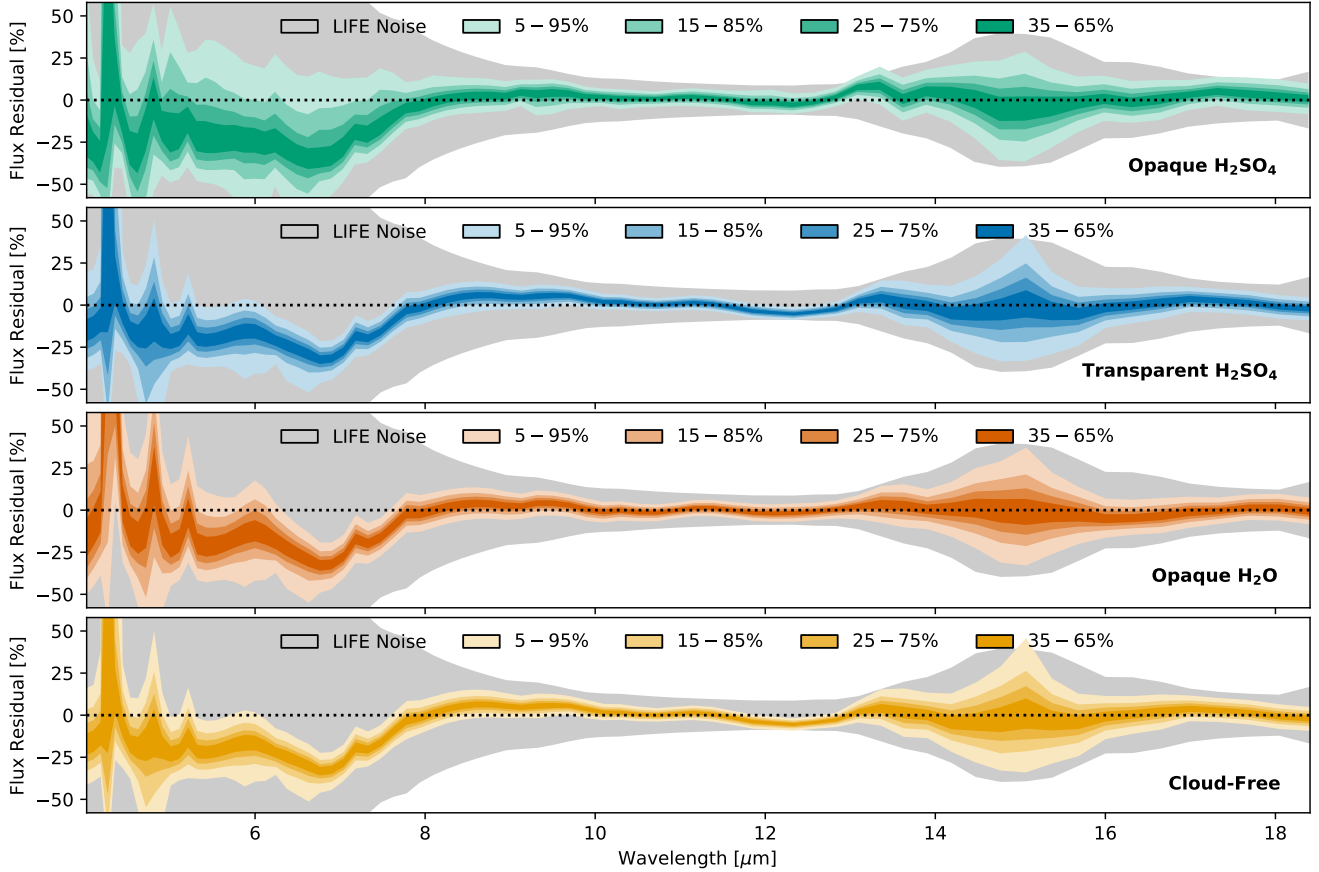


Fig. 3. Residuals of the spectra corresponding to the posteriors retrieved for the 4 – 18.5 μm , $R = 50$, LIFESIM $S/N = 10$ Venus-twin spectrum (opaque H_2SO_4 clouds). Color-shaded areas represent different residual quantiles. The gray area marks the 1σ LIFESIM noise level. Each panel contains the results for a different forward model (see Sect. 2.3.2). From top to bottom: Opaque H_2SO_4 clouds, transparent H_2SO_4 clouds, opaque H_2O clouds, and cloud-free.

First, we consider the results for the planetary surface pressure P_0 . We see, that if P_0 is a model parameter (transparent H_2SO_4 clouds or cloud-free models, see Table 1), the posterior is strongly constrained. However, the retrieved value does not correspond to Venus’ true surface pressure, but coincides with the cloud-top pressure ($\log_{10}(P_{\text{top}}^{\text{cloud}}[\text{bar}]) = -1.3$). This is in agreement with the findings for the P – T profiles we outlined in the previous section. The forward models that assumed an opaque cloud layer did not retrieve for P_0 and thus yielded no estimates.

The M_{pl} posterior is roughly Gaussian in log space ($\mu = -0.18$, $\sigma = 0.33$) for all forward models and is not strongly constrained over the assumed Gaussian prior ($\mu = -0.09$, $\sigma = 0.4$). This failure to further constrain M_{pl} was also observed in Papers III and V and is due to the well known degeneracy between the planet mass (surface gravity) and the abundances of the atmospheric trace gases (see also, e.g., Mollière et al. 2015; Feng et al. 2018; Madhusudhan 2018; Quanz et al. 2021).

For R_{pl} , the retrieved posterior strongly depends on the forward model. For the transparent H_2SO_4 and the cloud-free model, the R_{pl} posterior is roughly centered on the truth, approximately Gaussian ($\mu = 0.97 R_{\oplus}$, $\sigma = 0.05 R_{\oplus}$), and significantly constrained over the Gaussian prior ($\mu = 0.95 R_{\oplus}$, $\sigma = 0.20 R_{\oplus}$). In contrast, the R_{pl} posteriors for the forward models assuming opaque clouds are broader, non-Gaussian, and not centered on the truth. When assuming opaque H_2O clouds, the posterior is shifted relative to the true value and roughly Gaussian (slightly asymmetric, with a tail to larger radii). The retrieved median strongly underestimates the planet radius by $0.2 R_{\oplus}$. For

the opaque H_2SO_4 cloud forward model, the resulting R_{pl} posterior is significantly broader. We observe two separate peaks, one of which is centered on the truth. The other is shifted to the left and underestimates R_{pl} by approximately $0.18 R_{\oplus}$.

Since the T_{eq} and A_{B} distributions are derived from the R_{pl} posterior (see Appendix B), they inherit the forward model dependence of R_{pl} . Thus, retrievals using the transparent H_2SO_4 or the cloud-free forward model result in accurate, Gaussian-shaped estimates for T_{eq} ($\mu = 233 \text{ K}$, $\sigma = 5 \text{ K}$) and A_{B} ($\mu = 0.74$, $\sigma = 0.04$). For the forward models that assume an opaque cloud layer, the underestimation of R_{pl} results in an overestimation of T_{eq} . A higher T_{eq} can only occur if the planet retains more of the incident stellar radiation, which manifests itself in a lower Bond albedo A_{B} . As a result, we overestimate T_{eq} by roughly 20 K and underestimate A_{B} by approximately 0.1 for the opaque H_2O cloud model. For the opaque H_2SO_4 clouds, the T_{eq} posterior is similar to the R_{pl} posterior. It is non-Gaussian in shape, exhibits a peak at the true T_{eq} value, and extends significantly toward higher T_{eq} values. For the A_{B} distribution, the peak coincides with the truth, but the distribution shows a significant tail toward lower A_{B} values.

We retrieve high atmospheric CO_2 abundances ($\geq 30\%$ in mass fraction) for all forward models. Thus, we can easily differentiate between a Venus-like, CO_2 dominated atmosphere and an Earth-like atmosphere with lower CO_2 abundances. In contrast, H_2O and CO are not detected at the considered input quality, since the signatures in Venus’ spectrum lie below 8 μm and are therefore not significant compared to the high LIFESIM noise

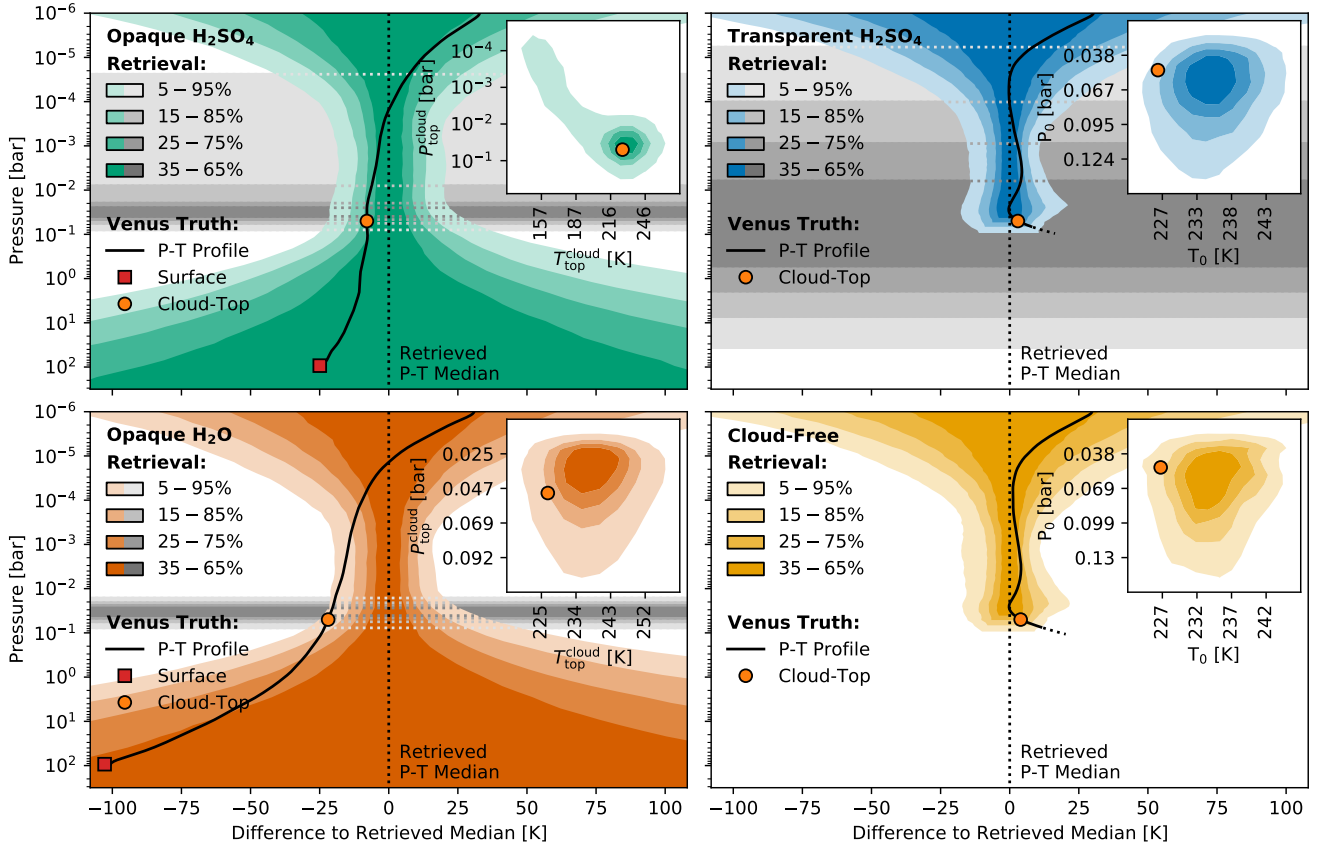


Fig. 4. Residuals of the P - T profiles retrieved for the 4 – 18.5 μm , $R = 50$, LIFESIM $S/N = 10$ Venus-twin spectrum (opaque H_2SO_4 clouds) relative to the retrieved median P - T profile. Color-shaded areas indicate percentiles of the P - T residuals. If the model includes clouds, the gray shaded regions indicate percentiles of the retrieved cloud-top pressure. The solid black line, the orange circular marker, and the red square marker represent the true profile relative to the retrieved median (for the transparent H_2SO_4 cloud and cloud-free models, we cannot plot the true surface and the P - T profile at pressures higher than retrieved surface pressure). In the top right, we plot the 2D P_0 - T_0 posterior (if retrieved; otherwise $P_{\text{top}}^{\text{cloud}}$ - $T_{\text{top}}^{\text{cloud}}$). Each panel summarizes the result for one of the four different forward models (see Sect. 2.3.2). From top-left to bottom-right: Opaque H_2SO_4 clouds, transparent H_2SO_4 clouds, opaque H_2O clouds, and cloud-free.

level. For H_2O , the drop in the posterior at high abundances rules out abundances $\geq 10^{-3}$. This limit on H_2O and the unconstrained CO abundance cause the drop in the spectrum residual below 8 μm observed in Sect. 3.1.1. Both posteriors extend to abundances significantly above the truth. Thus, on average, the spectra corresponding to the retrieved parameters have stronger H_2O and CO absorption features than the true Venus-twin spectrum, which leads to the observed drop in the residual below 8 μm .

Last, we consider the cloud parameter posteriors. The cloud-top pressure ($P_{\text{top}}^{\text{cloud}}$) posterior for both opaque cloud forward models provides a good approximation of the truth and is well described by a Gaussian ($\mu = -1.5$, $\sigma = 0.2$). Furthermore, we manage to retrieve a value for the minimal possible cloud thickness ($P_{\text{span}}^{\text{cloud}}$, $\log_{10}(P_{\text{span}}^{\text{cloud}} [\text{bar}]) \gtrsim -1.5 \pm 0.8$). Interestingly, even with the opaque H_2O forward model, which assumes a wrong cloud composition, we obtain accurate estimates for the position of the cloud deck in the atmosphere. In contrast, for the transparent H_2SO_4 cloud model, we do not manage to significantly constrain either the cloud-top position or the cloud thickness. The posteriors are flat and unconstrained with respect to the assumed priors. Similarly, also the cloud particle mass fraction (Species $^{\text{cloud}}$) is unconstrained for the transparent H_2SO_4 forward model.

The Species $^{\text{cloud}}$ posterior for the opaque H_2SO_4 forward model is strongly bimodal. The lower of the two peaks is

centered on the true H_2SO_4 mass fraction, while the other overestimates the abundance by roughly 3 dex. A more in-depth analysis of the posterior distributions (see corner plots in Appendix C) reveals a strong correlation between the retrieved cloud species abundance and R_{pl} . Interestingly, an overestimated cloud particle abundance is linked to an underestimated R_{pl} and thus also correlated with the T_{eq} and A_{B} posteriors. For the opaque H_2O forward model, the retrieved median abundance lies roughly 2.5 dex above the true H_2SO_4 cloud particle abundance.

Finally, the parameters describing the cloud particle size, \bar{R}^{cloud} and σ^{cloud} , are not well constrained for any of the forward models. In an in-depth analysis of the posteriors for the two opaque cloud models (see corner plots in Appendix C), we find a degeneracy between these two parameters. This indicates that a smaller \bar{R}^{cloud} can be compensated with a larger σ^{cloud} for the considered spectral quality. For the transparent H_2SO_4 model, we observe no degeneracy between the two parameters. The lack of constraints on all cloud parameters for the transparent H_2SO_4 model is caused by the addition of the surface pressure P_0 to the retrieval. The retrieval sets the surface at the cloud-top, which alleviates the need to model an opaque cloud layer.

3.2. Retrieval results for higher quality spectra

We now investigate how the retrieval results change if we consider higher quality Venus-twin spectra. For most model

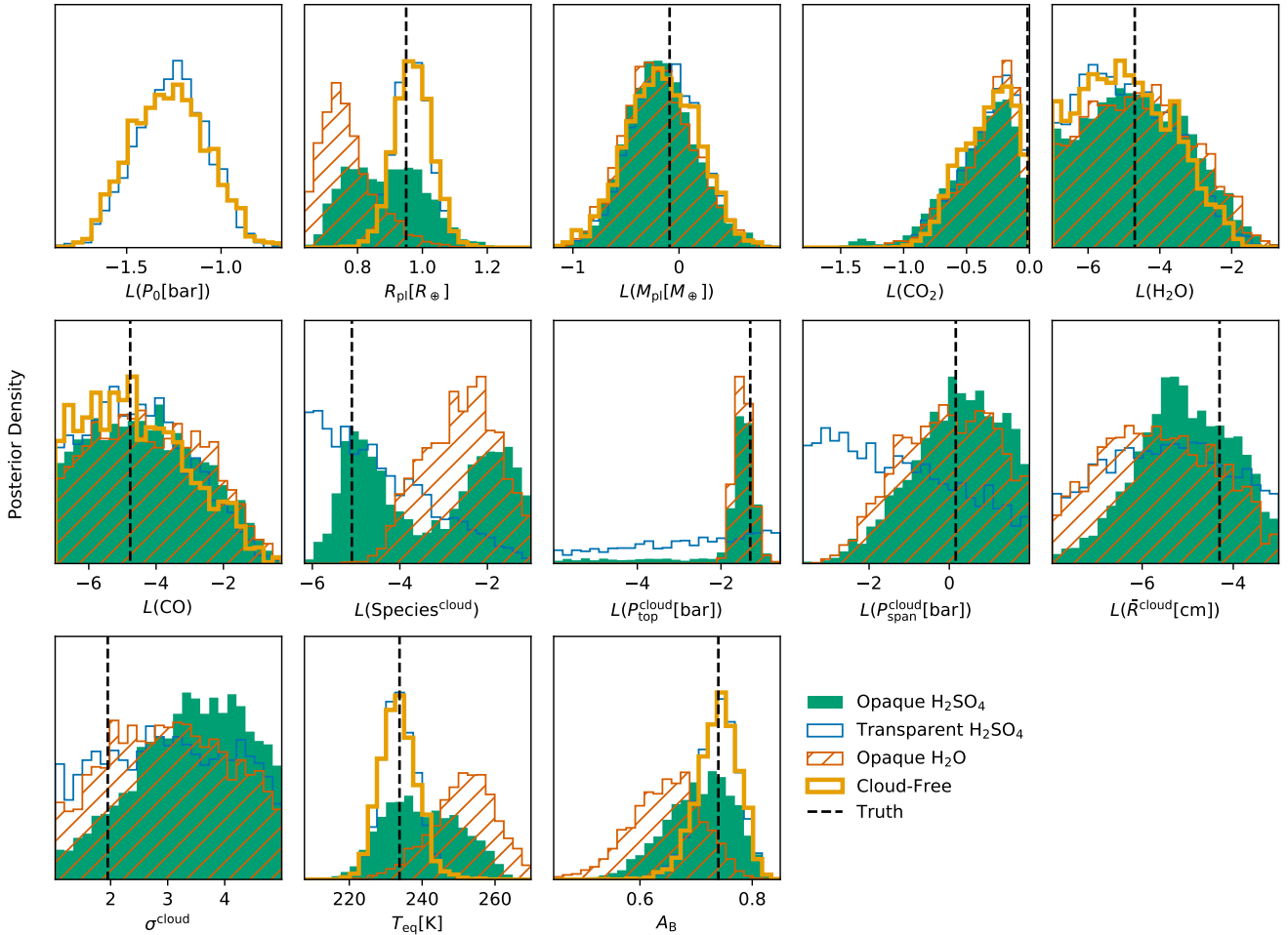


Fig. 5. Posteriors retrieved for the 4–18.5 μm , $R = 50$, LIFESIM $S/N = 10$ Venus-twin spectrum (opaque H_2SO_4 clouds) for different forward models (see Sect. 2.3.2). Here, $L(\cdot)$ abbreviates $\log_{10}(\cdot)$. We include the planetary equilibrium temperature T_{eq} and the Bond albedo A_{B} , which we derived from the posteriors (see Appendix B). Dashed black lines indicate the true values (see Table 1). For $\text{Species}^{\text{cloud}}$, the true value is the H_2SO_4 mass fraction in the Venus-twin model. Solid green area – opaque H_2SO_4 clouds; thin-blue outlined area – transparent H_2SO_4 clouds; orange-hatched area – opaque H_2O clouds; thick-yellow outlined area – cloud-free.

parameters, both increases in R and S/N do not significantly change the retrieval results. Generally, most parameters are better constrained as we move to higher quality spectra, but the general shape of the posterior distributions remains unchanged. Further, increasing the wavelength coverage from 4 – 18.5 μm to 3 – 20 μm does not significantly impact the results either. In Fig. 6, we focus on the parameter posteriors that significantly change when increasing the quality of the LIFESIM input spectrum (R_{pl} , abundance of cloud species, T_{eq} , and A_{B}). We plot the results for the 4 – 18.5 μm input spectra and exclude the intermediate $S/N = 15$ retrievals to increase readability. We provide the retrieved posteriors for all remaining model parameters and the retrieval results for the $S/N = 15$ and 3 – 20 μm input spectra in Appendix D.

For the opaque H_2SO_4 cloud model, the correlated bimodal nature of the R_{pl} and the H_2SO_4 abundance posteriors (see Sect. 3.1) diminishes strongly as we consider higher quality spectra. While neither of the two modes in these posteriors is preferred for the $R = 50$, $S/N = 10$ input spectrum (both peaks are equally high), the modes centered on the true values are clearly preferred for the higher quality spectra. For the $S/N = 20$ input spectra, the bimodalities disappear completely and both R_{pl} and the H_2SO_4 abundance are accurately determined.

For retrievals with the opaque H_2O cloud model, R_{pl} is underestimated by roughly $0.2 R_{\oplus}$ independent of the quality of the spectrum. Similarly, the retrieved median cloud particle abundance lies roughly 3 dex above the true H_2SO_4 abundance, irrespective of the R and S/N of the spectrum. However, for the $S/N = 20$ inputs, bimodalities in the R_{pl} and the H_2O cloud species posteriors emerge. These bimodalities are likely also present at lower S/N , but are not observable due to the larger uncertainties on the individual posterior models.

For the transparent H_2SO_4 and the cloud-free model, the constraint on R_{pl} is increased for higher R and S/N spectra. However, while the posterior is centered on the true value for the $R = 50$, $S/N = 10$ retrieval, R_{pl} is overestimated in retrievals of higher quality spectra. This bias is stronger for the cloud-free model ($\approx 0.05 R_{\oplus}$) and is likely due to differences between the model assumed to generate the Venus-twin input spectrum and the forward model. It is not observable at $R = 50$, $S/N = 10$, due to the larger uncertainties on the posterior. Finally, for the transparent H_2SO_4 cloud forward model, the retrieved cloud species abundance converges toward abundances below the true H_2SO_4 abundances as we consider higher quality LIFESIM spectra.

Lastly, we consider the T_{eq} and A_{B} distributions. When assuming the opaque H_2SO_4 model, the tail toward high T_{eq}

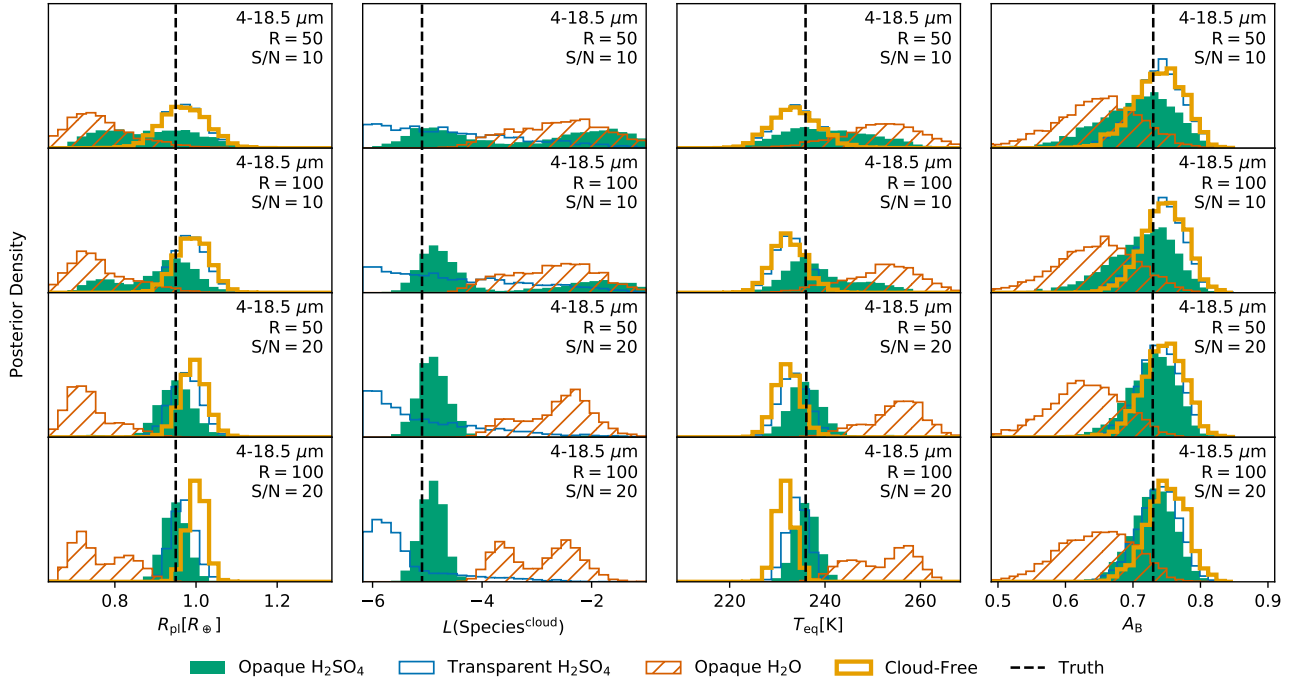


Fig. 6. Model parameter posteriors for increased quality 4 – 18.5 μm Venus-twin spectra (opaque H_2SO_4 clouds). Here, $L(\cdot)$ abbreviates $\log_{10}(\cdot)$. Dashed black lines indicate the true values (see Table 1). For $\text{Species}^{\text{cloud}}$, the true value is the H_2SO_4 mass fraction in the Venus-twin model. Solid green area – opaque H_2SO_4 clouds; thin-blue outlined area – transparent H_2SO_4 clouds; orange-hatched area – opaque H_2O clouds; thick-yellow outlined area – cloud-free. Columns (left to right) show the results for the planet radius R_{pl} , cloud species abundance $\text{Species}^{\text{cloud}}$, equilibrium temperature T_{eq} , Bond albedo A_{B} (T_{eq} and A_{B} were calculated following Appendix B; for the posteriors of all model parameters see Appendix D). Rows (top to bottom) represent different input qualities: $R = 50$, $S/N = 10$; $R = 100$, $S/N = 10$; $R = 50$, $S/N = 20$; $R = 100$, $S/N = 20$.

observed at $R = 50$, $S/N = 10$ diminishes analogously to the R_{pl} bimodality as we consider higher quality inputs. Consequentially, the tail toward low A_{B} decreases. Especially for $S/N = 20$ spectra, the A_{B} and T_{eq} distributions provide accurate Gaussian estimates that are centered on the truth. In contrast, systematic offsets from the truth emerge for T_{eq} for the transparent H_2SO_4 and the cloud-free forward models as we consider higher quality spectra. These offsets are linked to the offsets in the R_{pl} posteriors discussed above. Albeit less prominent, the systematic shifts also appear in the A_{B} distributions. They are less noticeable since the offset in T_{eq} is small compared to the uncertainties on the other parameters used to calculate A_{B} (see Appendix B). Finally, for the opaque H_2O cloud model, the T_{eq} and A_{B} estimates are not improved significantly as we move to higher R and S/N .

In summary, we observe significant changes in the posteriors of some parameters when considering higher quality spectra. While the transparent H_2SO_4 and the cloud-free forward models perform well at $R = 50$, $S/N = 20$, biases emerge for higher resolutions. This indicates that for high quality spectra, these models are likely not sufficient. In contrast, results for the opaque H_2SO_4 forward model are further refined with the input quality increase. While the estimates for many parameters are weak and biased at $R = 50$, $S/N = 10$, they improve significantly as we consider higher quality spectra (especially for $S/N = 20$). Finally, the results for the wrong opaque H_2O cloud model do not improve significantly when considering higher quality spectra.

4. Discussion

After summarizing the main results from our retrieval analysis in Sect. 3, we discuss how well one can characterize a Venus-twin exoplanet from simulated LIFE MIR observations of different

quality. In Sect. 4.1, we compare the performance of the different forward models to see whether we can find evidence for clouds by analyzing Venus’ MIR thermal emission spectrum. We further discuss potential alternative pathways for cloud inference. Thereafter, in Sect. 4.2, we address the limitations of our approach and motivate potential future studies.

4.1. Forward model selection and interpretation

In Sect. 3, we find that the retrieved P – T profile shape and the posterior distributions of the atmospheric gases (CO_2 , H_2O , and CO) exhibit only minor variations with the forward model and input quality. In contrast, the posteriors for R_{pl} and the cloud parameters, as well as the inferred distributions for T_{eq} and A_{B} depend significantly on the forward model and the input quality. This shows that incorrect model assumptions or an inadequate level of model complexity can result in incorrect exoplanetary characterization. These dependencies of the posteriors are problematic, because the true atmospheric structure and composition is unknown for an observed exoplanet. Consequentially, we will not be able to verify if the parameter values we retrieve assuming a forward model characterize the observed exoplanet correctly. Therefore, we require a method of determining an adequate forward model for a given exoplanet spectrum (Sect. 4.1.1). Furthermore, once an adequate forward model is determined, we need to understand how to link the obtained retrieval results to the conditions present on the exoplanet (Sect. 4.1.2).

4.1.1. Forward model selection via the Bayes factor

As outlined in Sect. 2.3.1, the Bayes’ factor K allows us to compare the retrieval performance of different forward models for a given exoplanet spectrum. Importantly, the Bayes’ factor K does

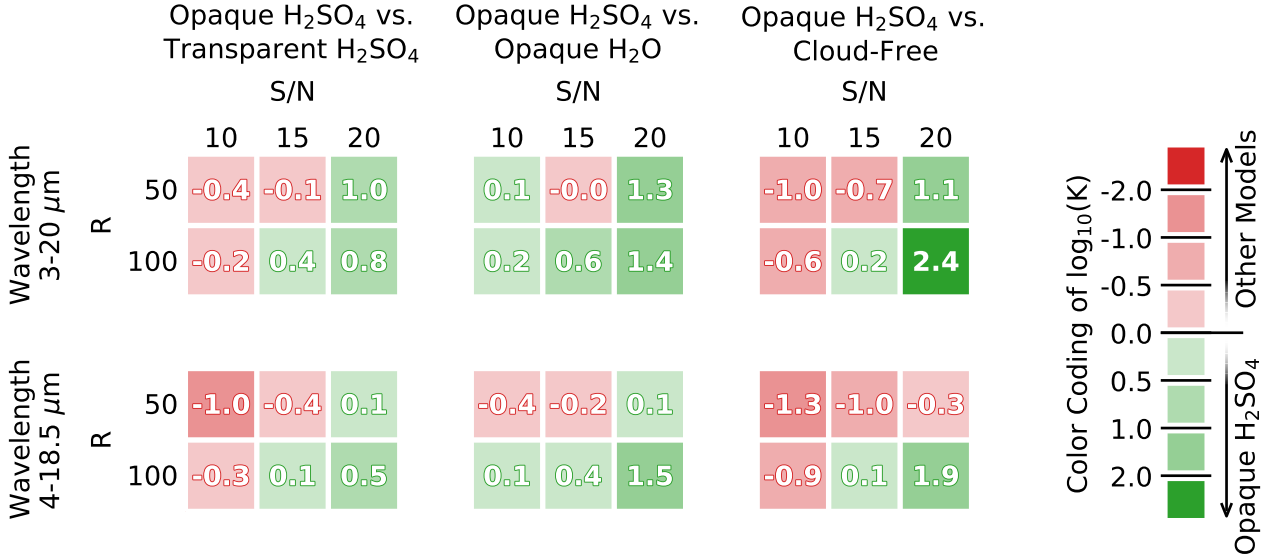


Fig. 7. Comparison of the retrieval performance of the true opaque H₂SO₄ cloud model with the other models via the Bayes factor K . We use the Jeffreys scale as metric for comparison (see Table 4, Jeffreys 1998) and consider different wavelength coverages, R , and S/N of the simulated Venus-twin LIFE observations. Green squares (positive $\log_{10}(K)$ values) indicate that the correct model (opaque H₂SO₄ clouds) is preferred, red squares (negative $\log_{10}(K)$ values) mark preference for the incorrect model. The intensity of coloring represents the preference strength.

Table 5. Model comparison via the Bayes factor K .

Compared models	$\log_{10}(K)$	Preferred model
$\mathcal{M}_{\text{Op.}}^{\text{H}_2\text{SO}_4}$ versus $\mathcal{M}_{\text{Tr.}}^{\text{H}_2\text{SO}_4}$	-0.4 ± 0.1	$\mathcal{M}_{\text{Tr.}}^{\text{H}_2\text{SO}_4}$
$\mathcal{M}_{\text{Op.}}^{\text{H}_2\text{SO}_4}$ versus $\mathcal{M}_{\text{Op.}}^{\text{H}_2\text{O}}$	0.1 ± 0.1	Either
$\mathcal{M}_{\text{Op.}}^{\text{H}_2\text{SO}_4}$ versus \mathcal{M}^{CF}	-1.0 ± 0.1	\mathcal{M}^{CF}
$\mathcal{M}_{\text{Tr.}}^{\text{H}_2\text{SO}_4}$ versus $\mathcal{M}_{\text{Op.}}^{\text{H}_2\text{O}}$	0.5 ± 0.1	$\mathcal{M}_{\text{Tr.}}^{\text{H}_2\text{SO}_4}$
$\mathcal{M}_{\text{Tr.}}^{\text{H}_2\text{SO}_4}$ versus \mathcal{M}^{CF}	-0.6 ± 0.1	\mathcal{M}^{CF}
$\mathcal{M}_{\text{Op.}}^{\text{H}_2\text{O}}$ versus \mathcal{M}^{CF}	-1.1 ± 0.1	\mathcal{M}^{CF}

Notes. Performance comparison of forward models for retrievals of the 4 – 18.5 μm, $R = 50$, $S/N = 10$ Venus-twin (opaque H₂SO₄ clouds) LIFESIM spectrum. We calculate $\log_{10}(K)$ for pairs of models with Eq. (2) and interpret its value via the Jeffreys scale (see Table 4, Jeffreys 1998).

not tell us if a model is correct or not. It provides a metric that measures which model out of a set of models is best suited to explain an observed exoplanet spectrum. Here, we investigate if clouds can be detected and characterized in retrievals for different quality input spectra. We do this by comparing the retrieval performance of the correct opaque H₂SO₄ cloud forward model to the other tested forward models using the Bayes’ factor K .

First, we compare the performance of different forward models at the minimal LIFE requirements from Paper III (4 – 18.5 μm, $R = 50$, and $S/N = 10$). We list the Bayes factor K for all combinations of forward models in Table 5. No forward model we tested is decisively ruled out or preferred ($|\log_{10}(K)| \leq 2$ for all comparisons). We only find slight performance differences between the forward models. Overall, the cloud-free model performs best (always preferred; $|\log_{10}(K)| \geq 0.6$), while the opaque cloud models perform worst (never preferred, despite the opaque H₂SO₄ cloud model being the true model). Importantly, the cloud-free model also uses fewer parameters than the cloudy scenarios (ten versus 14–15 parameters). The finding that the cloud-free forward model yields the

best retrieval performance indicates that the additional parameters required to model clouds are not justifiable.

For the Venus twin, this suggests that MIR retrievals at the current minimal LIFE requirements are not sufficient to find evidence for atmospheric clouds. Crucially, this does not rule out clouds. It merely indicates, that the considered spectrum is satisfactorily described by the cloud-free model. This agrees well with the findings for cloud-free retrievals on cloudy input spectra presented in Paper V.

Second, we investigate if our findings change for higher quality LIFESIM spectra. In Fig. 7, we compare the retrieval performance of the opaque H₂SO₄ cloud model (model used to produce the input spectra) to the other models via the Bayes factor K . Green fields mark positive $\log_{10}(K)$ values and thus preference for the true opaque H₂SO₄ cloud model. Red fields (negative $\log_{10}(K)$) indicate a preference for the incorrect model.

For spectra with $S/N \leq 20$, we observe that both the cloud-free and the transparent H₂SO₄ models perform better than (or comparable to) the opaque H₂SO₄ cloud model. Further, the cloud-free model outperforms the transparent H₂SO₄ model, as we see from the lower $\log_{10}(K)$ values. The two models assuming opaque clouds perform equally well. As before, these findings suggest that no evidence for clouds in Venus’ atmosphere can be found via retrieval studies of spectra with $S/N < 20$.

For $S/N = 20$ spectra, a preference for the opaque H₂SO₄ forward model emerges. This preference is generally stronger for spectra with larger wavelength coverage and higher R , since they contain more information. This suggests that an S/N of at least 20 is required to infer the presence of clouds in a retrieval study on the MIR thermal emission spectrum of a Venus-twin exoplanet.

Crucially, the S/N -dependent model preferences agree with our findings for the posteriors (see Sect. 3). For $S/N < 20$ spectra, the preferred cloud-free model yields good estimates for Venus’ atmospheric structure and composition above the cloud-top. The planetary parameters R_{pl} , T_{eq} , and A_{B} are correctly retrieved. Similarly, the transparent H₂SO₄ cloud model (second-best performance) approximates the aforementioned parameters well. The retrieved P_0 corresponds to the cloud-top, which

alleviates the need to model an opaque cloud layer. Thus, the cloud parameters are unconstrained. In contrast, the opaque cloud retrievals (lowest preference) yield weak and biased estimates for R_{pl} , T_{eq} , A_{B} , and the cloud parameters. This suggests, that for spectra with $S/N < 20$, the spectral information content is not sufficient to constrain these additional parameters. Hence, the cloud-free model with fewer parameters is preferred.

For $S/N = 20$ spectra, we notice significant changes in both the posteriors and the model preference. For the transparent H_2SO_4 and the cloud-free model, the estimates for R_{pl} , T_{eq} , and A_{B} are offset from the true value and thus yield biased estimates. For the opaque H_2O cloud model, we see no significant improvements in the posteriors over the $S/N < 20$ retrievals. For the opaque H_2SO_4 cloud model, which performs best on the $S/N = 20$ spectra, the biases we find for $S/N < 20$ spectra are no longer present. The posteriors for cloud and planet parameters are unbiased and provide good estimates. This suggests that at $S/N = 20$, the information content of the input spectrum is sufficient to justify the additional cloud parameters. These observations for the posteriors agree well with the shift in forward model preference from the cloud-free ($S/N < 20$) to the opaque H_2SO_4 cloud ($S/N = 20$) forward model.

In conclusion, we find that for low quality LIFESIM Venus-twin spectra ($S/N < 20$) cloud presence is not inferrable via MIR retrievals. For these inputs, the cloud-free model yields accurate estimates for fundamental planetary and atmospheric parameters. The accuracy of the constraints on these parameters are in accordance with the findings presented in Papers III and V. For the $S/N = 20$ spectra (especially if $R = 100$), we manage to find weak evidence for clouds in the atmosphere of the Venus twin and to constrain the cloud properties. We emphasize that our findings are based on the assumption of a Venus twin. However, the conclusion that clouds are hard to infer and constrain via retrievals of low quality MIR thermal emission spectra is likely generalizable to arbitrary terrestrial exoplanets. Further testing of this important result is foreseen for the future.

4.1.2. Interpretation of model selection results

Inferring cloud presence for terrestrial exoplanets via MIR thermal emission retrievals is challenging. At the minimal LIFE specifications, we find the cloud-free model to perform best, and thus no direct evidence for a cloud deck. There are two simple interpretations of the retrieval's preference for the cloud-free forward model. In the first interpretation, the retrieved surface pressure P_0 is incorrect. The true P_0 is larger and not retrieved correctly, since the atmospheric high pressure layers are optically thick and thus leave no signatures in the spectrum. For our study, this would suggest that the emission spectrum contains no information about the exoplanet's lower atmosphere (≥ 0.05 bar) and surface conditions. In the second interpretation, the retrieved P_0 corresponds to the truth and the surface contributes to the emission spectrum. In this case, the exoplanet would possess a thin atmosphere (≈ 0.05 bar). In both cases, the exoplanet is characterized by a high bond albedo ($A_{\text{B}} = 0.74 \pm 0.4$).

By considering the A_{B} of Solar System objects, we attempt to link the retrieved A_{B} to planet properties. In Table 6, we list the A_{B} of selected objects along with the main source of the MIR continuum emission. If an object has a low A_{B} (≤ 0.3), the continuum emission typically originates from the planet's surface (rocky objects with no or an optically thin atmosphere). For high A_{B} objects (≥ 0.5), the MIR continuum emission stems from either clouds (e.g., Venus) or predominantly ice and frost

Table 6. Bond albedos A_{B} of selected Solar System objects.

Object	A_{B}	MIR continuum emission	Reference
Mercury	0.08	Rocky surface	1
Venus	0.76	Clouds	2
Earth	0.30	Surface and clouds	3
Moon	0.14	Rocky surface	4
Mars	0.24	Rocky surface	5
Jupiter	0.53	Clouds	6
Europa	0.55	Icy surface	7
Saturn	0.34	Clouds	8
Tethys	0.67	Icy surface	9
Enceladus	0.81	Icy surface	9

Notes. The third column specifies the main source of the MIR continuum emission (for clear atmospheres the continuum originates from the surface, for opaque cloudy atmospheres it stems from the cloud deck).

References. (1) Mallama (2017); (2) Haus et al. (2016); (3) Pallé et al. (2003); (4) Matthews (2008); (5) Pleskot & Kieffer (1977); (6) Li et al. (2018); (7) Spencer et al. (1999); (8) Hanel et al. (1983); (9) Howett et al. (2010).

covered surfaces (e.g., Europa). Thus, our retrieval results suggest an exoplanet with either a cloudy atmosphere (first scenario) or an icy surface (second scenario). In the second scenario, the retrieved P_0 (≈ 0.05 bar) and the corresponding temperature ($T_0 = 235 \pm 4$ K) would allow for a water ice surface⁵. Such “snowball” states have occurred on Earth (Kirschvink 1992; Hoffman et al. 1998, 2017) and are also conceivable toward the inner edge of the HZ (Wordsworth 2021; Graham 2021). In a snowball state, the majority of the incident stellar radiation is reflected due to the high albedo of the planet's surface, which leads to low surface temperatures. Even for high incident radiation from the host star and large concentrations of atmospheric greenhouse gases, the planet surface can remain in a stable frozen state (Budyko 1969; Sellers 1969).

Yet, the icy scenario appears improbable given the exoplanet-star separation (0.72 AU) and the high retrieved levels of the strong greenhouse gas CO_2 ($\geq 30\%$ in mass fraction). In addition, the low retrieved surface pressure (≈ 0.05 bar) seems unlikely for an evolved Venus-sized exoplanet (e.g., Ortenzi et al. 2020). Lastly, the long-term stability of such a planet is uncertain and depends on various factors such as the rate of volcanic outgassing (e.g., Pierrehumbert 2010). However, the icy scenario cannot be ruled out solely via low quality MIR observations ($S/N < 20$). Increasing the spectrum's S/N to at least 20 allows us to infer cloud presence, yet still not robustly. Furthermore, an increase in S/N would require significantly more observation time (for $S/N = 20$ roughly four times longer than for $S/N = 10$). Thus, alternative cloud inference pathways are desirable to resolve this ambiguity in interpretation.

A potential remedy to the aforementioned ambiguity, is to leverage 1D or 3D (photo-)chemistry and climate models (e.g., Atmos⁶, ROCKE-3D⁷, or PlaSim⁸). Such a simulative approach can help us identify and rule out nonphysical planetary states.

⁵ This possibility of strongly underestimating the surface temperature of a cloudy exoplanet has already been discussed for habitable Earth-like exoplanets (e.g., Kitzmann et al. 2011).

⁶ <https://github.com/VirtualPlanetaryLaboratory/atmos>

⁷ <https://www.giss.nasa.gov/projects/astrobio/>

⁸ <https://www.mi.uni-hamburg.de/en/arbeitsgruppen/theoretische-meteorologie/modelle/plasim.html>

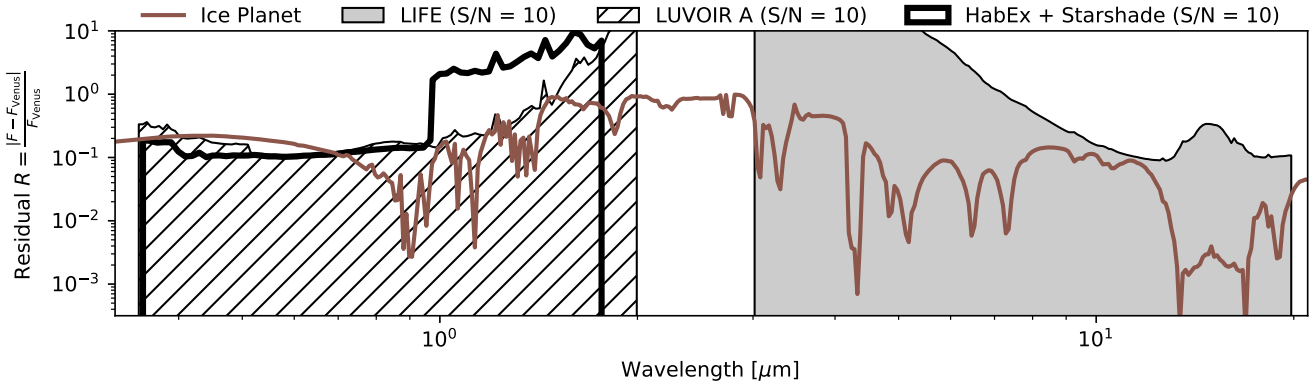


Fig. 8. Flux difference between a cloudy Venus-twin exoplanet (opaque H_2SO_4 clouds) and an icy world with a thin CO_2 -dominated atmosphere (see Appendix E for a comparison of the absolute fluxes). In brown, we plot the flux residual of the icy world with respect to the Venus twin as a function of wavelength. The gray shaded region indicates the 1σ LIFESIM noise level at $S/N = 10$. The thin, hatched, and black area represents the expected wavelength-dependent noise for the LUVOIR A mission concept (The LUVOIR Team 2019); the thick, solid black line represents the expected noise for the HabEx + Starshade mission concept (Gaudi et al. 2020).

For example, for the icy exoplanet scenario motivated above, there are two crucial questions that could be studied via climate simulations. First, one has to investigate if an icy surface together with the retrieved atmospheric structure and composition describes a physically possible and stable state. Since our retrieval framework does not model the atmospheric physics (e.g., convection, photochemistry), not all points in the posterior distribution result in stable atmospheres. In the case of the icy exoplanet, the atmosphere described by the retrieved posteriors might enter a rapid runaway greenhouse phase, which would lead to melting and subsequent evaporation of the surface H_2O . Under these circumstances, the icy scenario would be highly unlikely, and thus the cloudy scenario preferred. Studies similar to Boukrouche et al. (2021), Chaverot et al. (2022), or Graham et al. (2022) could help us better understand the stability of the icy scenario. If we find the icy state to be realistic, the second question to tackle is if it can be reached, given the exoplanets proximity to the host star and the high atmospheric CO_2 abundance. To answer this question, studies similar to Wordsworth (2021) or Graham (2021), that investigate a wide range of different atmospheres, could provide answers. Such studies are an example of potential future synergies between atmospheric retrievals and the theoretical modeling of atmospheres. However, recent intercomparison efforts show that both retrievals (e.g., Barstow et al. 2020) and atmospheric modeling (e.g., Sergeev et al. 2022) depend on the model choice. Community efforts, such as the CUISINES Working Group⁹, that benchmark, compare, and validate different models, are vital to the studies proposed above.

An alternative approach is to not only consider a terrestrial exoplanet’s MIR emission, but also the stellar light it reflects in the UV/O/NIR. While we do not expect the scenarios to differ noticeably in MIR emission, the different reflective properties of clouds and ice will significantly impact the UV/O/NIR reflected light spectrum. To test this approach, we simulated spectra between $0.3 \mu\text{m}$ and $21 \mu\text{m}$ using *petitRADTRANS*. For the first scenario, we used our cloudy Venus-twin exoplanet (opaque H_2SO_4 clouds; see Sect. 2.1 and Table 1). For the second scenario, we modeled an icy surface at a pressure of 0.05 bar (P_0 from cloud-free retrievals) and a cloud-free atmosphere with the same composition and $P-T$ profile as for the first scenario. We

used data from the ECOSTRESS Spectral Library¹⁰ (Baldridge et al. 2009; Meerdink et al. 2019) to model the wavelength-dependent reflectivity of the icy surface. An 80% frost- and 20% ice-covered surface yielded a total reflectance of 0.75 in the UV/O/NIR (\approx retrieved A_B). In Fig. 8, we plot the residual of the ice-planet spectrum relative to the cloudy Venus-twin spectrum and the LIFESIM noise ($S/N = 10$). We provide an analogous plot comparing the absolute fluxes in Appendix E. In the UV/O/NIR, we show the noise expected for observations of the same planets with LUVOIR A (The LUVOIR Team 2019) or HabEx + Starshade (Gaudi et al. 2020; $S/N = 10$ at $0.6 \mu\text{m}$, calculated with the NASA-GSFC Planetary Spectrum Generator¹¹; Villanueva et al. 2018). While the residual lies below the LIFESIM noise level in the MIR, it is significantly larger than both the expected LUVOIR and HabEx noise in the UV/O/NIR below $0.7 \mu\text{m}$. This indicates that the reflected light spectrum is more suitable to differentiate between the icy and cloudy scenario than the MIR thermal emission spectrum. It further exemplifies the complementarity of UV/O/NIR reflected light and MIR thermal emission observations and highlights the importance of following both strategies. Retrieval studies on combined UV/O/NIR and MIR spectra are foreseen for the future.

4.2. Limitations and future work

The study we present here provides us with first estimates for how well a Venus-like exoplanet could be characterized by LIFE. Further, we obtain insights into how atmospheric clouds can complicate retrieval studies and the interpretation of their output. As we are making several assumptions in our approach, our findings cannot readily be generalized to arbitrary science cases. Here, we discuss these limitations in detail.

First, we restricted ourselves to the study of a Venus twin. While the performance for individual model parameters will not generalize well, the more general findings provide insights into biases inherent to retrievals (e.g., the dependence of the parameter posteriors on the forward model). However, as suggested in Robinson & Salvador (2023), Solar System planets provide an excellent benchmark for retrievals, because these atmospheres

⁹ <https://nexss.info/cuisines/>

¹⁰ <https://speclib.jpl.nasa.gov>

¹¹ <https://psg.gsfc.nasa.gov>

are known to be physical and estimates for the ground-truth values of the model parameters are available. Performing retrievals on different Solar System planets will help us generalize our predictions for the characterization performance.

Second, there are limitations inherent to our theoretical Venus-twin input spectrum. We assumed a fully mixed atmosphere (vertically constant abundances). While the same assumption was made in Paper III, the input spectra in Paper V were based on variable abundance profiles. Nevertheless, the retrieval results from Papers III and V are comparable, indicating that this simplification does not heavily impact retrievals at the spectral qualities considered. We also treated the clouds in a simplified manner. More realistic cloud models, which consider P - T -dependent cloud particle sizes and abundances or spatial variations in the cloud deck, could affect the MIR emission spectrum measurably. Further, we neglected both temporal and spatial variances in the atmospheric structure and composition. Real (exo)planet emission spectra can vary with time and depend on the viewing geometry (e.g., Mettler et al. 2020, 2023). Retrieval studies on more realistic input spectra will provide more reliable estimates for LIFE's performance and are foreseen in the future.

Third, the limitations above are also valid for the forward models used in the retrievals. However, only limited increases in the forward model complexity are possible, as they lead to a substantial rise in the retrievals' computational complexity. For example, we only retrieved for molecules present in the Venus-twin atmosphere. Including additional molecules could lead to false positive detections of gases and a mischaracterization of the atmosphere. However, a first robustness study for false positive detections in the appendix of Paper III justifies our approach. Another simplification is that we use a 1D forward model. While this is not problematic here, since the Venus-twin input is also calculated with a 1D model, it will be wrong for retrievals of real spectra and spectra from 3D models. However, in a recent study, Robinson & Salvador (2023) compared the performance of their 1D retrieval suite (*rFast*) to results from a computationally expensive 3D retrieval (Feng et al. 2018). They concluded that for the R and S/N we considered here, 1D retrievals suffice to obtain a first order understanding of how the spectral quality affects the exoplanet characterization performance.

Fourth, we used *petitRADTRANS* to generate the Venus-twin input spectrum and as radiative transfer model in the retrieval. As discussed in Paper V and Barstow et al. (2020), systematic differences between the radiative transfer model used to generate the input spectrum and in the retrievals (e.g., differences in the used line lists, Alei et al. 2022b) can lead to biases in the posteriors. For retrievals on real exoplanet spectra, similar problems are unavoidable, since the radiative transfer model will never capture the full atmospheric physics and chemistry of the observed exoplanet. Thus, our results might be overly optimistic and it is indispensable to investigate the nature and magnitude of the resulting biases in future studies.

Last, important limitations are rooted in the LIFESIM noise model. Currently, LIFESIM models the dominant astrophysical noise terms but neglects systematic instrumental effects (Paper II). Ideally, instrumental noise contributions will not dominate LIFE's noise budget. Nevertheless, they will contribute to the observational noise by altering the relative distribution of noise across the wavelength range (Dannert et al., in prep.), which might affect the retrieval results. More accurate estimates will be possible once LIFE's optical, thermal, and detector designs have matured and are accounted for by LIFESIM. Also, we interpreted the LIFESIM noise as uncertainty on the

Venus-twin spectrum. Crucially, we did not randomize the values of individual spectral points. This decision might lead to overly optimistic results (see Sect. 2.3.1). We expect the low R and S/N cases to be more strongly affected by randomization. Further, the cloud inference and characterization capabilities could also be overly optimistic. For a detailed discussion on potential impacts of this simplification on the characterization performance, we refer to the appendix of Paper III.

5. Conclusions and outlook

In this study, we ran retrievals for a cloudy Venus twin orbiting a G2V star at a distance of 10 pc. The goal was to investigate how the minimal R and S/N requirements for LIFE defined in Paper III and verified in Paper V are affected by clouds.

We approximated Venus' MIR thermal spectrum using the 1D radiative transfer model *petitRADTRANS* (Mollière et al. 2019, 2020; Paper V) assuming a third order polynomial P - T structure, vertically constant CO_2 , H_2O , and CO abundances, and a uniform, Mie scattering H_2SO_4 - H_2O cloud slab. The LIFESIM tool (Paper II) simulated LIFE observations of the Venus twin. Using an updated version of the retrieval suite from Papers III and V, we ran retrievals for variable quality spectra (from the 4 – 18.5 μm , $R = 50$, $S/N = 10$ minimal LIFE requirements to 3 – 20 μm , $R = 100$, $S/N = 20$ spectra) assuming different (cloud-free and cloudy) forward models.

At the minimal LIFE requirements, we correctly retrieve the P - T structure above the cloud top (for pressures $\leq 0.03 \pm 0.02$ bar, T lies within $\leq \pm 25$ K of the truth) and find high CO_2 levels ($\geq 30\%$ in mass fraction). These findings allow us to discern Venus- from Earth-like exoplanets. Further, Venus' surface conditions are not constrainable via its MIR thermal emission, since the opaque atmospheric clouds block contributions from the lower atmospheric layers. The results for the planet radius R_{pl} , equilibrium temperature T_{eq} , Bond albedo A_{B} , and the cloud parameters depend strongly on the forward model choice. Overall, the cloud-free model yields the best estimates ($R_{\text{pl}} = 0.97 \pm 0.05 R_{\oplus}$, $T_{\text{eq}} = 233 \pm 5$ K, $A_{\text{B}} = 0.74 \pm 0.04$) and is favored by the Bayes factor analysis. This suggests that cloud presence cannot be inferred at the minimal LIFE requirements. For high quality spectra ($S/N \geq 20$), the parameter constraints increase and the model preference shifts toward the correct opaque H_2SO_4 cloud model. While this suggests that retrieval based cloud inference is possible with LIFE, other approaches, such as followup UV/O/NIR observations with a HWO-like telescope, offer an alternative and synergistic approach.

Crucially, we find that our retrieval results for important planetary parameters (R_{pl} , T_{eq} , A_{B}) strongly depend on the chosen forward model. An incorrect forward model or an inadequate level of forward model complexity (e.g., too complex given the quality of the input spectrum) can heavily bias retrieval results. This is a major concern since, for observations of real exoplanets, the atmospheric state will be unknown. Furthermore, the work we presented here suggests that model selection via the Bayes factor will likely be hard and thus the risk of over- or misinterpretation of the available data is high. While sufficient quality MIR spectra of Earth- or Venus-like exoplanets will not be available in the near future, the *James Webb* Space Telescope will measure transmission spectra as well as thermal emission of many exoplanets in the upcoming years. Atmospheric retrievals will be used to analyze these spectra (Cowan et al. 2015; Greene et al. 2016; Krissansen-Totton et al. 2018; Nixon & Madhusudan 2022). Therefore, working toward a community-wide common

approach for retrieval studies of exoplanet spectra is of great importance, as it would mitigate the risk of false characterization significantly and augment the comparability of different studies. Applied to empirical data from powerful future space missions, such as LIFE, the in-depth characterization of different types of terrestrial exoplanets seems within reach.

Acknowledgements. This work has been carried out within the framework of the NCCR PlanetS supported by the Swiss National Science Foundation under grants 51NF40_182901 and 51NF40_205606. S.P.Q. and E.A. acknowledge the financial support from the SNSF. S.R. acknowledges support from the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant [2022-04588]. Author contributions: B.S.K. carried out the analyses, created the figures, and wrote the manuscript. S.P.Q. initiated this project. S.P.Q. and E.A. provided regular guidance. All authors discussed the results and commented on the manuscript.

References

- Alei, E., Konrad, B. S., Angerhausen, D., et al. 2022a, *A&A*, **665**, A106
- Alei, E., Konrad, B. S., Mollière, P., et al. 2022b, *SPIE Conf. Ser.*, **12180**, 121803L
- Anglada-Escudé, G., Amado, P. J., Barnes, J., et al. 2016, *Nature*, **536**, 437
- Arney, G., & Kane, S. 2018, ArXiv e-prints [arXiv:1804.05889]
- Baldrige, A., Hook, S., Grove, C., & Rivera, G. 2009, *Rem. Sens. Environ.*, **113**, 711
- Barstow, J. K., & Heng, K. 2020, *Space Sci. Rev.*, **216**, 82
- Barstow, J. K., Aigrain, S., Irwin, P. G. J., Kendrew, S., & Fletcher, L. N. 2016, *MNRAS*, **458**, 2657
- Barstow, J. K., Changeat, Q., Garland, R., et al. 2020, *MNRAS*, **493**, 4884
- Boukrouche, R., Lichtenberg, T., & Pierrehumbert, R. T. 2021, *ApJ*, **919**, 130
- Brandt, T. D., & Spiegel, D. S. 2014, *PNAS*, **111**, 13278
- Bryson, S., Kunimoto, M., Kopparapu, R. K., et al. 2020, *AJ*, **161**, 36
- Buchner, J., Georgakakis, A., Nandra, K., et al. 2014, *A&A*, **564**, A125
- Budyko, M. I. 1969, *Tellus*, **21**, 611
- Burch, D. E., Gryvnak, D. A., Patty, R. R., & Bartky, C. E. 1969, *J. Opt. Soc. Am.*, **59**, 267
- Carrion-González, Ó., García Muñoz, A., Cabrera, J., et al. 2020, *A&A*, **640**, A136
- Chaverot, G., Turbet, M., Bolmont, E., & Lecote, J. 2022, *A&A*, **658**, A40
- Chen, J., & Kipping, D. 2016, *ApJ*, **834**, 17
- Cowan, N. B., Greene, T., Angerhausen, D., et al. 2015, *PASP*, **127**, 311
- Dannert, F. A., Ottiger, M., Quanz, S. P., et al. 2022, *A&A*, **664**, A22
- Deming, D., Louie, D., & Sheets, H. 2018, *PASP*, **131**, 013001
- Dressing, C. D., & Charbonneau, D. 2015, *ApJ*, **807**, 45
- Ertel, S., Defrère, D., Hinz, P., et al. 2020, *AJ*, **159**, 177
- Feinstein, A. D., Radica, M., Welbanks, L., et al. 2023, *Nature*, **614**, 670
- 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
- Foreman-Mackey, D., Hogg, D. W., & Morton, T. D. 2014, *ApJ*, **795**, 64
- Gaudi, B. S., Seager, S., Mennesson, B., et al. 2020, ArXiv e-prints [arXiv:2001.06683]
- Gilbert, E. A., Barclay, T., Schlieder, J. E., et al. 2020, *AJ*, **160**, 116
- Gillon, M., Jehin, E., Lederer, S. M., et al. 2016, *Nature*, **533**, 221
- Gillon, M., Triaud, A. H. M. J., Demory, B.-O., et al. 2017, *Nature*, **542**, 456
- Graham, R. J. 2021, *Astrobiology*, **21**, 1406
- Graham, R. J., Lichtenberg, T., & Pierrehumbert, R. T. 2022, *J. Geophys. Res.: Planets*, **127**, e2022JE007456
- Greene, T. P., Line, M. R., Montero, C., et al. 2016, *ApJ*, **817**, 17
- Hanel, R., Conrath, B., Kunde, V., Pearl, J., & Pirraglia, J. 1983, *Icarus*, **53**, 262
- Hartmann, J. M., Boulet, C., Brodbeck, C., et al. 2002, *J. Quant. Spec. Radiat. Transf.*, **72**, 117
- Harvey, A. H., Gallagher, J. S., & Sengers, J. M. H. L. 1998, *J. Phys. Chem. Ref. Data*, **27**, 761
- Haus, R., Kappel, D., Tellmann, S., et al. 2016, *Icarus*, **272**, 178
- Hoffman, P. F., Kaufman, A. J., Halverson, G. P., & Schrag, D. P. 1998, *Science*, **281**, 1342
- Hoffman, P. F., Abbot, D. S., Ashkenazy, Y., et al. 2017, *Sci. Adv.*, **3**, e1600983
- Howett, C., Spencer, J., Pearl, J., & Segura, M. 2010, *Icarus*, **206**, 573
- Jeffreys, H. 1998, *The Theory of Probability*, Oxford Classic Texts in the Physical Sciences (Oxford: OUP), 432
- Kammerer, J., & Quanz, S. P. 2018, *A&A*, **609**, A4
- Karman, T., Gordon, I. E., van der Avoird, A., et al. 2019, *Icarus*, **328**, 160
- Kasting, J. F. 1988, *Icarus*, **74**, 472
- Kasting, J. F., & Harman, C. E. 2021, *Nature*, **598**, 259
- Kasting, J. F., Whitmire, D. P., & Reynolds, R. T. 1993, *Icarus*, **101**, 108
- Kirschvink, J. L. 1992, in *The Proterozoic Biosphere: A Multidisciplinary Study*, eds. J. W. Schopf, C. Klein, & D. Des Maris (Cambridge University Press), 51
- Kitzmann, D., Patzer, A. B. C., von Paris, P., Godolt, M., & Rauer, H. 2011, *A&A*, **531**, A62
- Komacek, T. D., Fauchez, T. J., Wolf, E. T., & Abbot, D. S. 2020, *ApJ*, **888**, L20
- Konrad, B. S., Alei, E., Quanz, S. P., et al. 2022, *A&A*, **664**, A23
- Kopparapu, R. K., Ramirez, R., Kasting, J. F., et al. 2013, *ApJ*, **765**, 131
- Krasnopolsky, V. A. 2015, *Icarus*, **252**, 327
- Krissansen-Totton, J., Garland, R., Irwin, P., & Catling, D. C. 2018, *AJ*, **156**, 114
- Léger, A., Defrère, D., Muñoz, A. G., et al. 2019, *Astrobiology*, **19**, 797
- Li, L., Jiang, X., West, R. A., et al. 2018, *Nat. Commun.*, **9**, 3709
- Loftus, K., Wordsworth, R. D., & Morley, C. V. 2019, *ApJ*, **887**, 231
- Lustig-Yaeger, J., Meadows, V. S., & Lincowski, A. P. 2019, *ApJ*, **887**, L11
- Madhusudhan, N. 2018, *Handbook of Exoplanets*, eds. H. J. Deeg, & J. A. Belmonte (Springer), 2153
- Mallama, A. 2017, ArXiv e-prints [arXiv:1703.02670]
- Matthews, G. 2008, *Appl. Opt.*, **47**, 4981
- Meerdink, S. K., Hook, S. J., Roberts, D. A., & Abbott, E. A. 2019, *Rem. Sens. Environ.*, **230**, 111196
- Mettler, J.-N., Quanz, S. P., & Helled, R. 2020, *AJ*, **160**, 246
- Mettler, J.-N., Quanz, S. P., Helled, R., Olson, S. L., & Schwieterman, E. W. 2023, *AJ*, **946**, 82
- Min, M., Hovenier, J. W., & de Koter, A. 2005, *A&A*, **432**, 909
- Mollière, P., van Boekel, R., Dullemond, C., Henning, T., & Mordasini, C. 2015, *ApJ*, **813**, 47
- 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
- Mueller-Wodarg, I. C. F., Strobel, D. F., Moses, J. I., et al. 2008, *Neutral Atmospheres* (New York, NY: Springer New York), 191
- National Academies of Sciences, Engineering, and Medicine. 2021, *Pathways to Discovery in Astronomy and Astrophysics for the 2020s* (Washington, DC: The National Academies Press)
- Nixon, M. C., & Madhusudhan, N. 2022, *ApJ*, **935**, 73
- Ortzeni, G., Noack, L., Sohl, F., et al. 2020, *Sci. Rep.*, **10**, 10907
- Oschlinskiok, J., Häusler, B., Pätzold, M., et al. 2012, *Icarus*, **221**, 940
- Pallé, E., Goode, P. R., Yurchyshyn, V., et al. 2003, *J. Geophys. Res.: Atmos.*, **108**, 4710
- Palmer, K. F., & Williams, D. 1975, *Appl. Opt.*, **14**, 208
- Petigura, E. A., Howard, A. W., & Marcy, G. W. 2013, *PNAS*, **110**, 19273
- Pierrehumbert, R. T. 2010, *Principles of Planetary Climate* (Cambridge University Press)
- Pleskot, L. K., & Kieffer, H. H. 1977, *Icarus*, **30**, 341
- Quanz, S. P., Absil, O., Angerhausen, D., et al. 2021, *Exp. Astron.*, submitted
- Quanz, S. P., Ottiger, M., Fontanet, E., et al. 2022, *A&A*, **664**, A21
- Robinson, T. D., & Salvador, A. 2023, *Planet. Sci. J.*, **4**, 10
- Rothman, L., Gordon, I., Barber, R., et al. 2010, *J. Quant. Spec. Radiat. Transf.*, **111**, 2139
- Rugheimer, S., Kaltenecker, L., Zsom, A., Segura, A., & Sasselov, D. 2013, *Astrobiology*, **13**, 251
- Segelstein, D. J. 1981, Master Thesis, University of Missouri-Kansas City, USA
- Sellers, W. D. 1969, *J. Appl. Meteorol. Climatol.*, **8**, 392
- Sergeev, D. E., Lewis, N. T., Lambert, F. H., et al. 2022, *Planet. Sci. J.*, **3**, 214
- Skilling, J. 2006, *Bayesian Anal.*, **1**, 833
- Sneep, M., & Ubachs, W. 2005, *J. Quant. Spec. Radiat. Transf.*, **92**, 293
- Spencer, J. R., Tamppari, L. K., Martin, T. Z., & Travis, L. D. 1999, *Science*, **284**, 1514
- Taylor, F. W., Svedhem, H., & Head, J. W. 2018, *Space Sci. Rev.*, **214**, 35
- The LUVOIR Team 2019, ArXiv e-prints [arXiv:1912.06219]
- Titov, D. V., Ignatiev, N. I., McGouldrick, K., Wilquet, V., & Wilson, C. F. 2018, *Space Sci. Rev.*, **214**, 126
- Toon, O. B., & Ackerman, T. P. 1981, *Appl. Opt.*, **20**, 3657
- Turbet, M., Bolmont, E., Chaverot, G., et al. 2021, *Nature*, **598**, 276
- Vasquez, M., Schreier, F., Gimeno García, S., et al. 2013, *A&A*, **557**, A46
- Villanueva, G., Smith, M., Protopapa, S., Faggi, S., & Mandell, A. 2018, *J. Quant. Spec. Radiat. Transf.*, **217**, 86
- von Paris, P., Hedelt, P., Selsis, F., Schreier, F., & Trautmann, T. 2013, *A&A*, **551**, A120
- Voyage 2050 Senior Committee. 2021, *Voyage 2050 — Final Recommendations from the Voyage 2050 Senior Committee*
- Way, M. J., Del Genio, A. D., Kiang, N. Y., et al. 2016, *Geophys. Res. Lett.*, **43**, 8376
- Wordsworth, R. 2021, *ApJ*, **912**, L14
- Yang, J., Boué, G., Fabrycky, D. C., & Abbot, D. S. 2014, *ApJ*, **787**, L2

Appendix A: Cloud retrieval validation

Here, we validate the updated retrieval routine and test if the newly added cloud model is constrainable. For this retrieval test, we considered a high resolution ($R = 1000$), low noise (only photon noise, $S/N = 50$), and $3 - 20 \mu\text{m}$ cloudy Venus-twin MIR spectrum. For this validation retrieval, we used the opaque H_2SO_4 forward model and the priors specified in Table 1. We summarize the results in Fig. A.1.

The top left panel of Fig. A.1, shows the corner plot of the posterior distribution, excluding the $P-T$ parameters a_i for clarity (the full corner plot is in Appendix C). Additionally, we derived the equilibrium temperature T_{eq} and the Bond albedo A_B from the posteriors following the method outlined in Appendix B. We see that most posteriors are well constrained and centered on the true values. We especially emphasize the strong constraints on both R_{pl} and A_B . The ability to constrain R_{pl} and A_B simultaneously is unique to studies of the MIR thermal emission spectrum. Comparable constraints are not inferrable from reflected light spectra, due to a strong degeneracy between R_{pl} and A_B (Feng et al. 2018; Carrión-González et al. 2020).

Further, we find that the atmosphere is CO_2 dominated (abundance $\geq 85\%$) and manage to detect both H_2O and CO . Most cloud parameters are well retrieved, which indicates that there are spectral features from the clouds. Only $P_{\text{span}}^{\text{cloud}}$ is not well constrained. The posterior is best described by a step function, where any $P_{\text{span}}^{\text{cloud}} \geq 0.5$ bar is possible. This lower limit on $P_{\text{span}}^{\text{cloud}}$ corresponds to the thickness at which the cloud layer becomes opaque. Thicker cloud layers are not excluded since the spectrum contains no signatures from the lower atmospheric layers. For thinner layers, there are contributions from the lower atmospheric layers, which result in a bad fit to the input spectrum.

The model parameters are generally well retrieved. However, there are interesting features in some posteriors. First, the rectangular shape of the $P_{\text{top}}^{\text{cloud}}$ posterior is linked to the fixed atmospheric pressure grid of the opaque H_2SO_4 cloud forward model. All $P_{\text{top}}^{\text{cloud}}$ values in the rectangular peak fall within the same atmospheric layer and thus result in the same cloud-top. Their posterior probability is therefore equal. Thus, at the high R and S/N considered here, we are sensitive to the discrete nature of our atmosphere model. Cloud-tops in lower or higher layers are very unlikely and are linked to the secondary peaks seen in both the 2D and 1D posteriors of other parameters (e.g., M_{pl}).

Second, the strength of the M_{pl} constraint opposes the results from Papers III and V, where M_{pl} was less strongly constrained relative to the prior due to a known degeneracy between the mass (surface gravity) and the abundances of the atmospheric gases (M_{pl} and abundances simultaneously over- or underestimated; see also, e.g., Mollière et al. 2015; Feng et al. 2018; Madhusudan 2018; Quanz et al. 2021). This degeneracy is visible for CO_2 and H_2O , but not for CO (due to overall weaker abundance constraints). However, the CO_2 posterior converges toward the upper edge of the prior range and is more strongly constrained (tail to high abundances is cut off). This breaks the degeneracy and thus also M_{pl} is strongly constrained. The secondary peaks in the M_{pl} posterior correspond to the secondary peaks in the $P_{\text{top}}^{\text{cloud}}$ posterior (cloud top in different layers).

The top right panel of Fig. A.1 shows the constraints on the atmospheric $P-T$ structure relative to the retrieved median (we plot the relative retrieved $P-T$ profile, since deviances from the truth are more visible). The absolute $P-T$ plot is given in Appendix C. We visualize percentiles of the retrieved profiles as green-shaded regions and compare them to the true profile

(red square, orange circle, solid black line). Further, we plot the percentiles of the retrieved cloud-top as gray areas. The inset plot shows the 2D histogram of the retrieved $P_{\text{top}}^{\text{cloud}}$ and $T_{\text{top}}^{\text{cloud}}$ (cloud-top temperature; calculated from the retrieved $P-T$ profiles).

Both the $P-T$ profile above the cloud-top and the cloud-top are accurately retrieved. The retrieved $P-T$ structure above the cloud-top is centered on the true profile, and roughly 90% of the profiles lie within $\leq \pm 2$ K of the truth. Similarly, roughly 90% of the retrieved $P_{\text{top}}^{\text{cloud}}$ and $T_{\text{top}}^{\text{cloud}}$ values are distributed symmetrically around the true cloud-top and lie within $\leq \pm 0.01$ bar and $\leq \pm 3$ K of the truth. The 5 – 95% $P-T$ envelope shows an asymmetry toward positive relative temperatures. In the $P_{\text{top}}^{\text{cloud}}-T_{\text{top}}^{\text{cloud}}$ posterior, the 5 – 95% envelope exhibits a second isolated region (same $T_{\text{top}}^{\text{cloud}}$, lower $P_{\text{top}}^{\text{cloud}}$). These two outliers correspond to the cases discussed above where the cloud-top falls within a different atmospheric layer.

The uncertainty on the retrieved $P-T$ profile grows with decreasing pressure. This increase is due to a lack of spectral features from these optically thin, low-pressure layers in Venus' MIR spectrum at the R and S/N considered here. Similarly, the uncertainty on the retrieved $P-T$ structure increases below the cloud-top, because the optically thick cloud deck blocks spectral contributions from these atmospheric regions. This lack of spectral features can be seen in the emission contribution function provided in Appendix C. Constraints on the high- and low-pressure regions are obtained by extrapolating the $P-T$ model.

In the bottom panel of Fig. A.1, we show the residual of the emission spectra corresponding to the retrieved parameter posteriors relative to our Venus-twin input spectrum. We indicate percentiles of the residuals as green-shaded areas and the 1σ photon noise level as a gray area. Overall, the retrieval output is well centered on the input spectrum and the deviations from the truth mostly lie below 1%. The observed deviations grow larger for wavelengths where the flux is low and the photon noise is large (e.g., below $6 \mu\text{m}$, in the CO_2 band between $14 - 16 \mu\text{m}$). However, in general, the assumed forward model accurately reproduces the input spectrum, despite the points mentioned above. This demonstrates that our cloudy Venus-twin model is suited for our retrieval study.

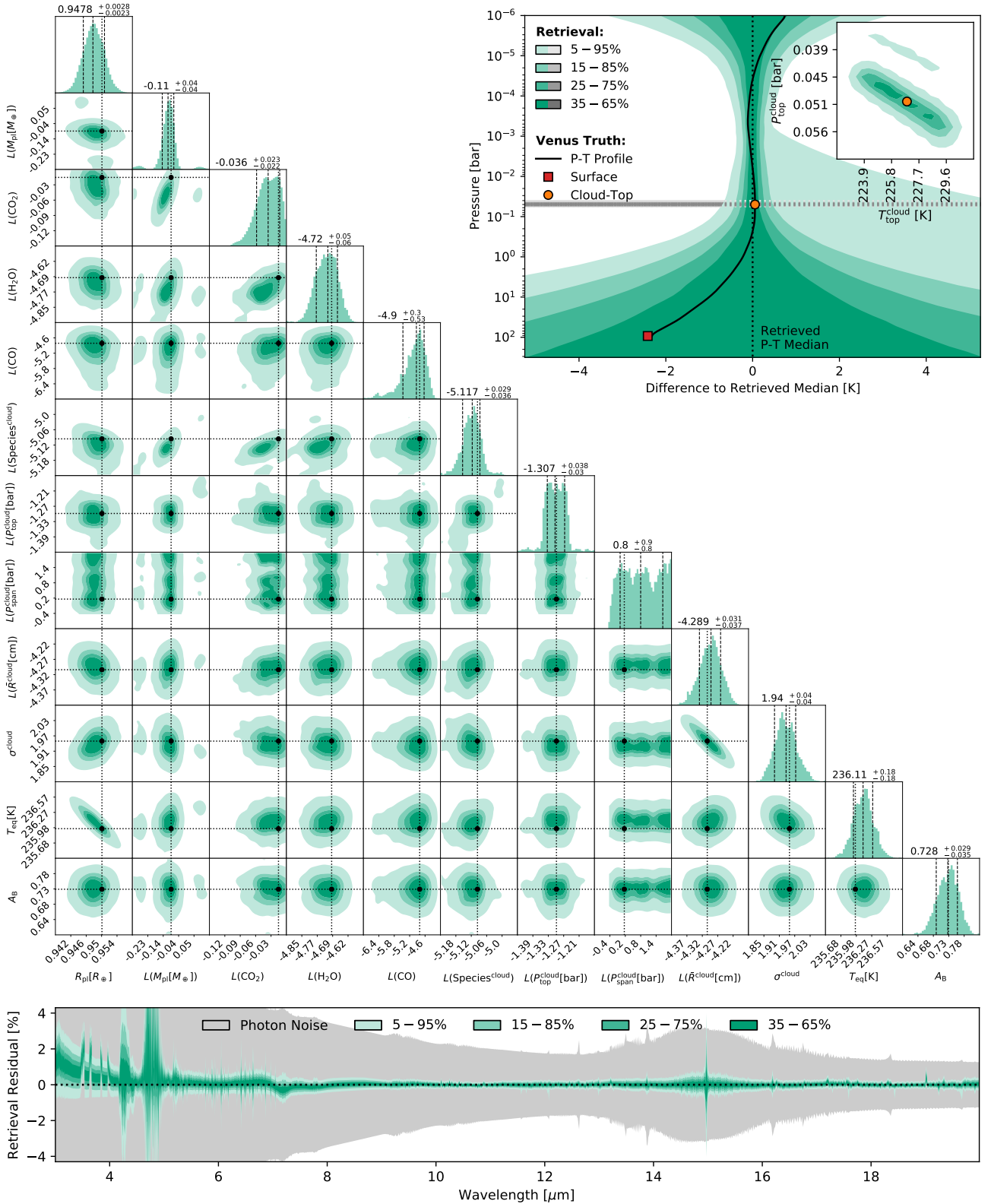


Fig. A.1. Results from the validation retrieval on the 3 – 20 μm , $R = 1000$, and photon noise $S/N = 50$ Venus-twin spectrum (opaque H_2SO_4 clouds) using the opaque H_2SO_4 cloud forward model. *Top left:* Corner plot of the posterior distribution of the forward model parameters. Here, $L(\cdot)$ stands for $\log_{10}(\cdot)$. We derived the equilibrium temperature T_{eq} and the Bond albedo A_{B} from the other posteriors (see Appendix B). The dotted black lines indicate the true values. In the 1D posteriors, we show the 16th, 50th, and 84th percentiles as black dashed lines. *Top right:* P – T profile residuals relative to the retrieved median P – T profile. Color-shaded regions indicate percentiles of the retrieved P – T profiles. The gray shaded regions indicate percentiles of the retrieved cloud-top pressure. The solid black line, the orange circular marker, and the red square marker represent the true Venus-twin P – T profile relative to the median retrieved P – T profile. In the inset figure, we plot the 2D P_0 – T_0 posterior. *Bottom:* Residuals of the spectra corresponding to the retrieved posteriors relative to the Venus-twin input. Color-shaded areas represent different quantiles of the residuals. The gray area marks the 1σ photon noise level.

Appendix B: Equilibrium temperature and Bond albedo

The planetary equilibrium temperature T_{eq} and the Bond albedo A_{B} are two parameters, which were of particular interest to us. These parameters provide important information about the energy budget of the planet. However, both parameters were not directly determined by our retrieval framework. Therefore, we required a method to derive their values from the retrieved posterior distributions of the model parameters.

In a first step, we generated a set of MIR spectra that was representative of the retrieved parameter posteriors. Next, we determined the planet's T_{eq} by fitting the black body emission, $F_{\text{BB}}(\lambda)$, expected for a spherical planet with radius R_{pl} to the continuum (between 8 μm and 11 μm) of these MIR spectra:

$$F_{\text{BB}}(\lambda) = \frac{2\pi hc^2}{\lambda^5} e^{-\frac{hc}{\lambda k_{\text{B}} T_{\text{eq}}}} \cdot \frac{R_{\text{pl}}^2}{d^2}. \quad (\text{B.1})$$

Here, d is the distance to the observer (10 pc in our case), λ the wavelength, h the Planck constant, c the speed of light, and k_{B} the Boltzmann constant. This provided us with a probability distribution for the planetary equilibrium temperature T_{eq} . From the T_{eq} distribution, we then calculated the probability distribution for A_{B} using the following relation:

$$A_{\text{B}} = 1 - 16\pi\sigma \frac{a_{\text{p}}^2 T_{\text{eq}}^4}{L_*}. \quad (\text{B.2})$$

Here, σ is the Stefan–Boltzmann constant, a_{p} is the semi-major axis of the planetary orbit around the host star, and L_* is the luminosity of the host star. For the A_{B} calculation, we assumed to know both a_{p} (Venus-like: $a_{\text{p}} = 0.72$ AU) and L_* (Sun-like: $L_* = 1 L_{\odot}$) with an uncertainty of $\pm 5\%$. For each sample from the T_{eq} distribution, we drew a random a_{p} and L_* value from two uncorrelated normal distributions and used these random values to calculate A_{B} . This yielded a probability distribution for the planetary Bond albedo A_{B} .

This approach for determining T_{eq} and A_{B} is advantageous, because correlations between the posteriors of the retrieved planetary parameters are taken into account when deriving the T_{eq} and A_{B} distributions. Furthermore, the resulting correlations between the distributions of T_{eq} , A_{B} , and the retrieved parameter posteriors are correct and can be analyzed.

Appendix C: Additional plots for retrieval results

Additional plots for the results from the validation retrieval (see Sect. A) and the retrievals of the 4 – 18.5 μm , $R = 50$, and $S/N = 10$ LIFESIM spectrum (see Sect. 3.1). We plot the retrieval results from the validation retrieval in Fig. C.1. For the 4 – 18.5 μm , $R = 50$, and $S/N = 10$ LIFESIM spectrum, we plot the results for the opaque H_2SO_4 forward model in Fig. C.2, for the transparent H_2SO_4 forward model in Fig. C.3, for the opaque H_2O forward model in Fig. C.4, and for the cloud-free forward model in Fig. C.5.

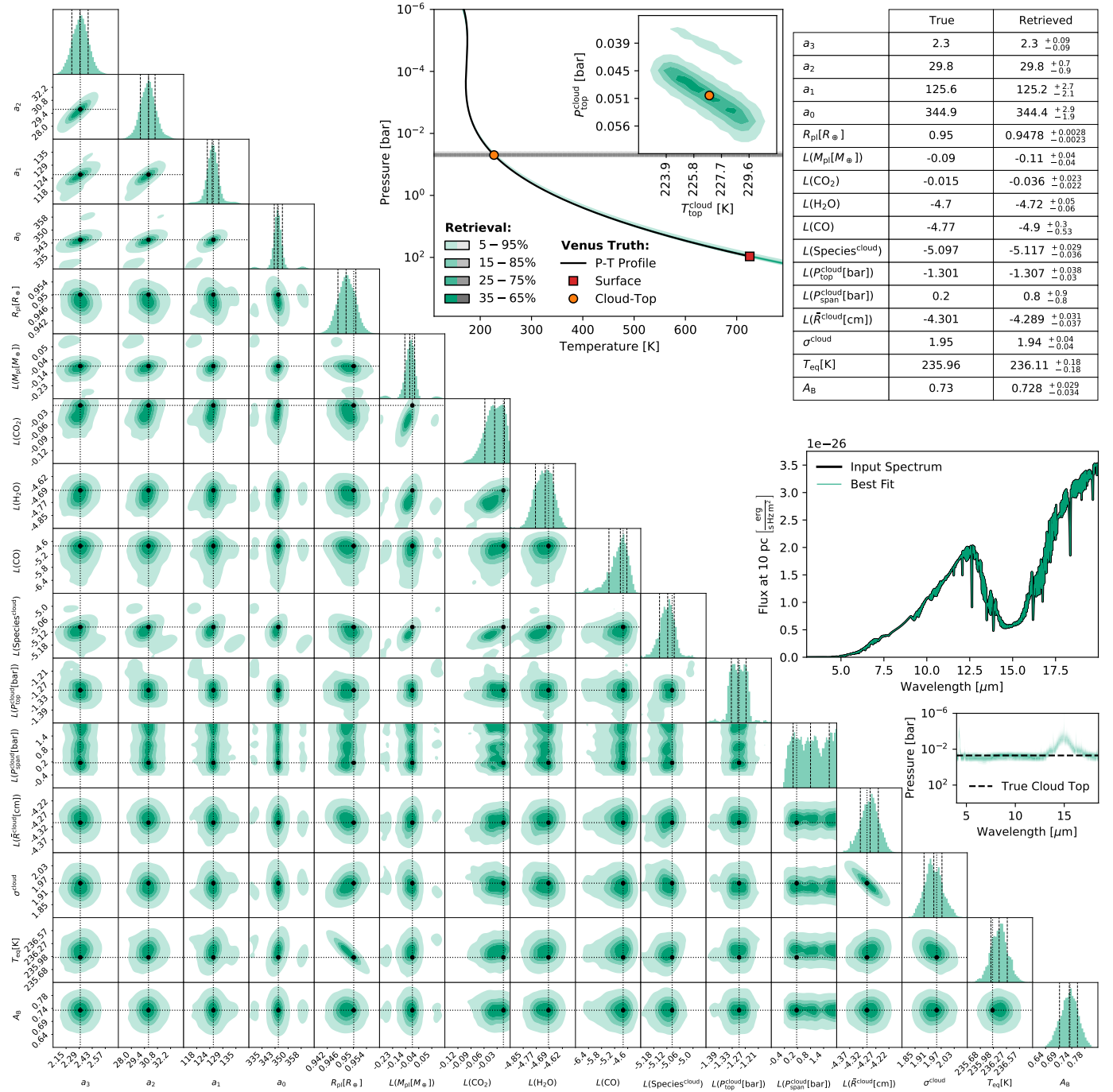


Fig. C.1. Results from the validation retrieval on the 3 – 20 μm , $R = 1000$, and photon noise $S/N = 50$ Venus-twin spectrum (opaque H_2SO_4 clouds) using the opaque H_2SO_4 cloud forward model (see Sect. A). *Lower left half*: Corner plot of the posterior distribution of the forward model parameters. Here, $L(\cdot)$ stands for $\log_{10}(\cdot)$. We derived the equilibrium temperature T_{eq} and the Bond albedo A_B from the other posteriors (see Appendix B). The dotted black lines indicate the true values. In the 1D posteriors, we show the 16th, 50th, and 84th percentiles as black dashed lines. *Top center*: $P-T$ profiles corresponding to the retrieved $P-T$ parameters. Color-shaded regions indicate percentiles of the retrieved $P-T$ profiles. The solid black line, the orange circular marker, and the red square marker represent the true Venus-twin $P-T$ profile. In the inset figure, we plot the 2D P_0-T_0 posterior. *Top right*: A table summarizing the true values of the forward model parameters and the 16th, 50th, and 84th percentiles of the parameter posteriors. *Center right*: Comparison of the retrievals best fit to the Venus-twin input spectrum. The thin colored line represents the best fit, the thick black line the Venus-twin input spectrum. The uncertainties on the best fit are too small to be visible. Below the spectrum plot, we show the mean wavelength-dependent contribution of the atmospheric layers to the emission spectrum corresponding to the retrieved parameter posteriors. Darker colors indicate higher contributions. The dashed black line indicates the position of the cloud-deck assumed to simulate the input spectrum (opaque H_2SO_4 clouds).

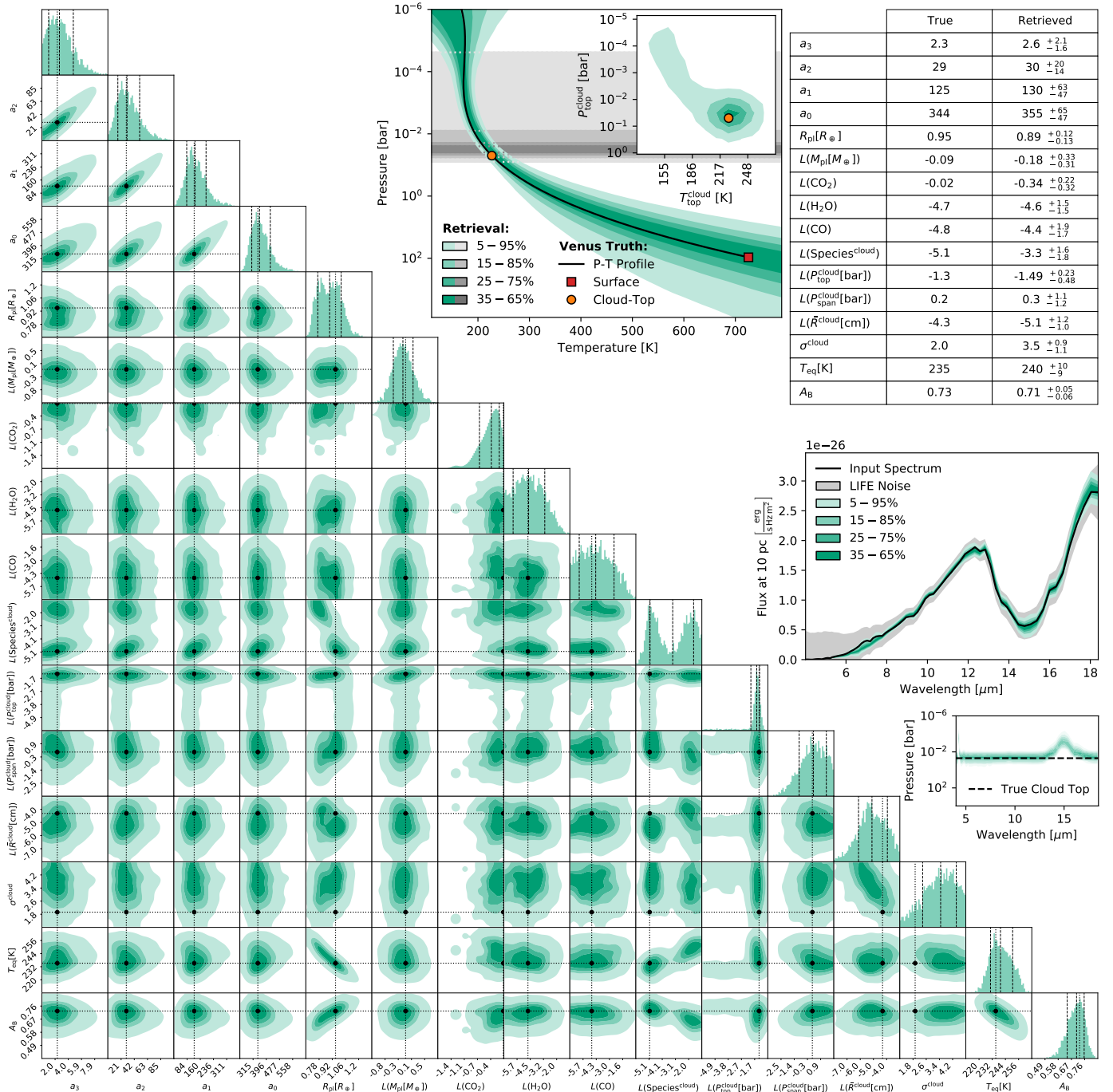


Fig. C.2. Results from the retrieval on the 4 – 18.5 μm , $R = 50$, and LIFESIM noise $S/N = 10$ Venus-twin spectrum (opaque H_2SO_4 clouds) using the opaque H_2SO_4 cloud forward model (see Sect. 3.1). *Lower left half:* Corner plot of the posterior distribution of the forward model parameters. Here, $L(\cdot)$ stands for $\log_{10}(\cdot)$. We derived the equilibrium temperature T_{eq} and the Bond albedo A_B from the other posteriors (see Appendix B). The dotted black lines indicate the true values. In the 1D posteriors, we show the 16th, 50th, and 84th percentiles as black dashed lines. *Top center:* P – T profiles corresponding to the retrieved P – T parameters. Color-shaded regions indicate percentiles of the retrieved P – T profiles. The gray shaded regions indicate percentiles of the retrieved cloud-top pressure. The solid black line, the orange circular marker, and the red square marker represent the true Venus-twin P – T profile. In the inset figure, we plot the 2D P_0 – T_0 posterior (if retrieved; otherwise $P_{\text{top}}^{\text{cloud}}$ – $T_{\text{top}}^{\text{cloud}}$). *Top right:* A table summarizing the true values of the forward model parameters and the 16th, 50th, and 84th percentiles of the parameter posteriors. *Center right:* Spectra corresponding to the retrieved posteriors in comparison to the Venus-twin input. Color-shaded areas represent different quantiles of the retrieved spectra. The solid black line represents the Venus-twin input spectrum. The gray area marks the 1σ LIFESIM noise level. Below the spectrum plot, we show the mean wavelength-dependent contribution of the atmospheric layers to the emission spectrum corresponding to the retrieved parameter posteriors. Darker colors indicate higher contributions. The dashed black line indicates the position of the cloud-deck assumed to simulate the input spectrum (opaque H_2SO_4 clouds).

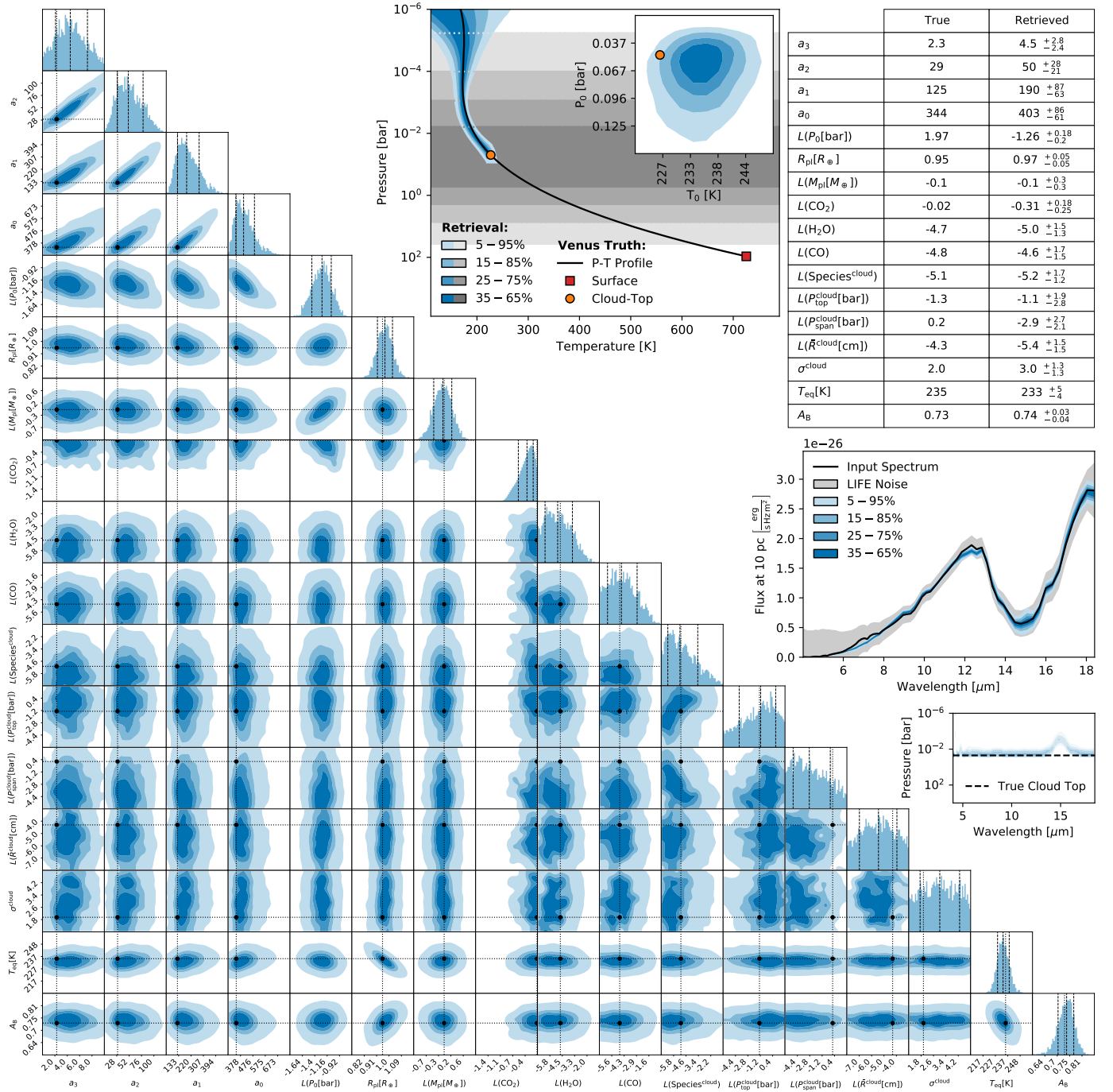


Fig. C.3. Same as Fig. C.2, but for the retrieval on the 4 – 18.5 μm , $R = 50$, and LIFESIM noise $S/N = 10$ Venus-twin spectrum (opaque H_2SO_4 clouds) using the transparent H_2SO_4 cloud forward model (see Sect. 3.1).

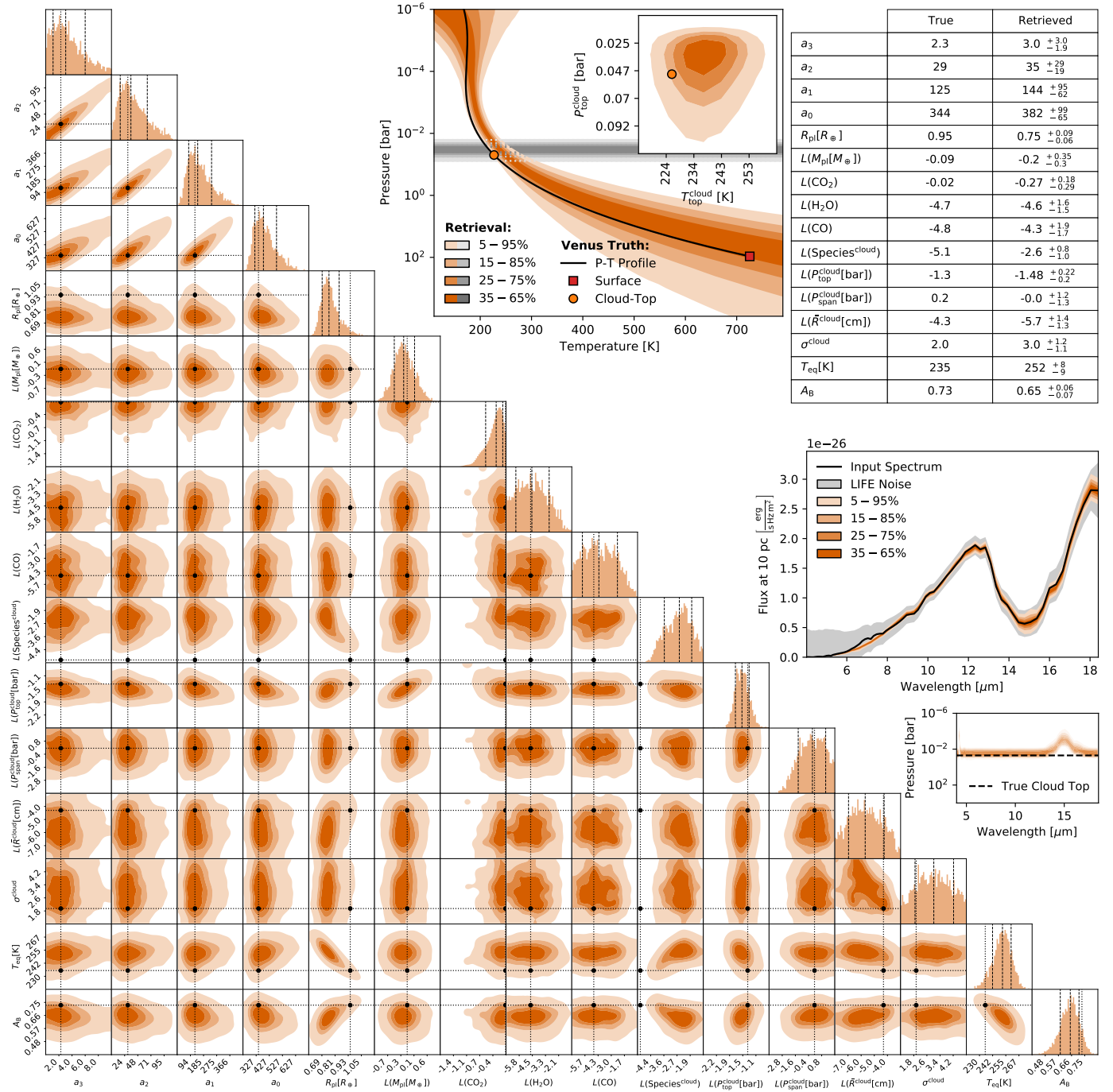


Fig. C.4. Same as Fig. C.2, but for the retrieval on the 4 – 18.5 μm , $R = 50$, and LIFESIM noise $S/N = 10$ Venus-twin spectrum (opaque H_2SO_4 clouds) using the opaque H_2O cloud forward model (see Sect. 3.1).

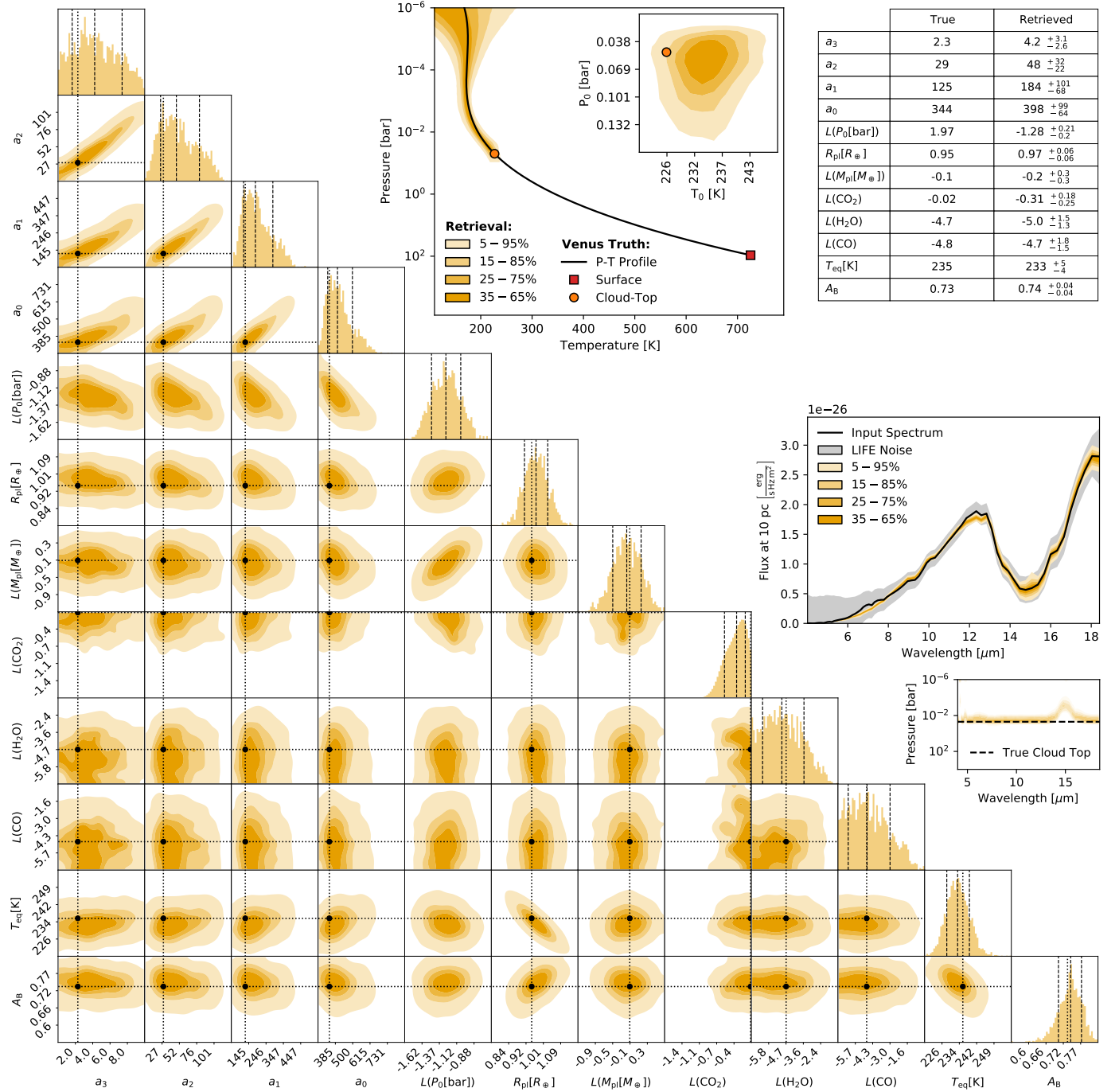


Fig. C.5. Same as Fig. C.2, but for the retrieval on the 4 – 18.5 μm , $R = 50$, and LIFESIM noise $S/N = 10$ Venus-twin spectrum (opaque H_2SO_4 clouds) using the cloud-free forward model (see Sect. 3.1).

Appendix D: Supplementary results from retrievals on higher quality input spectra

Here we plot the posteriors for all parameters found in the retrieval analysis for higher quality input spectra discussed in Sect. 3.2. We include the T_{eq} and A_{B} distributions, which we calculated from the posteriors following the method introduced in Appendix B. Further, we add the results for the intermediate $S/N = 15$ noise levels. In Figs. D.1 and D.2, we provide the retrieval results for the 3 – 20 μm input spectra. In Figs. D.3 and D.4, we display the results obtained for the 4 – 18.5 μm input spectra.

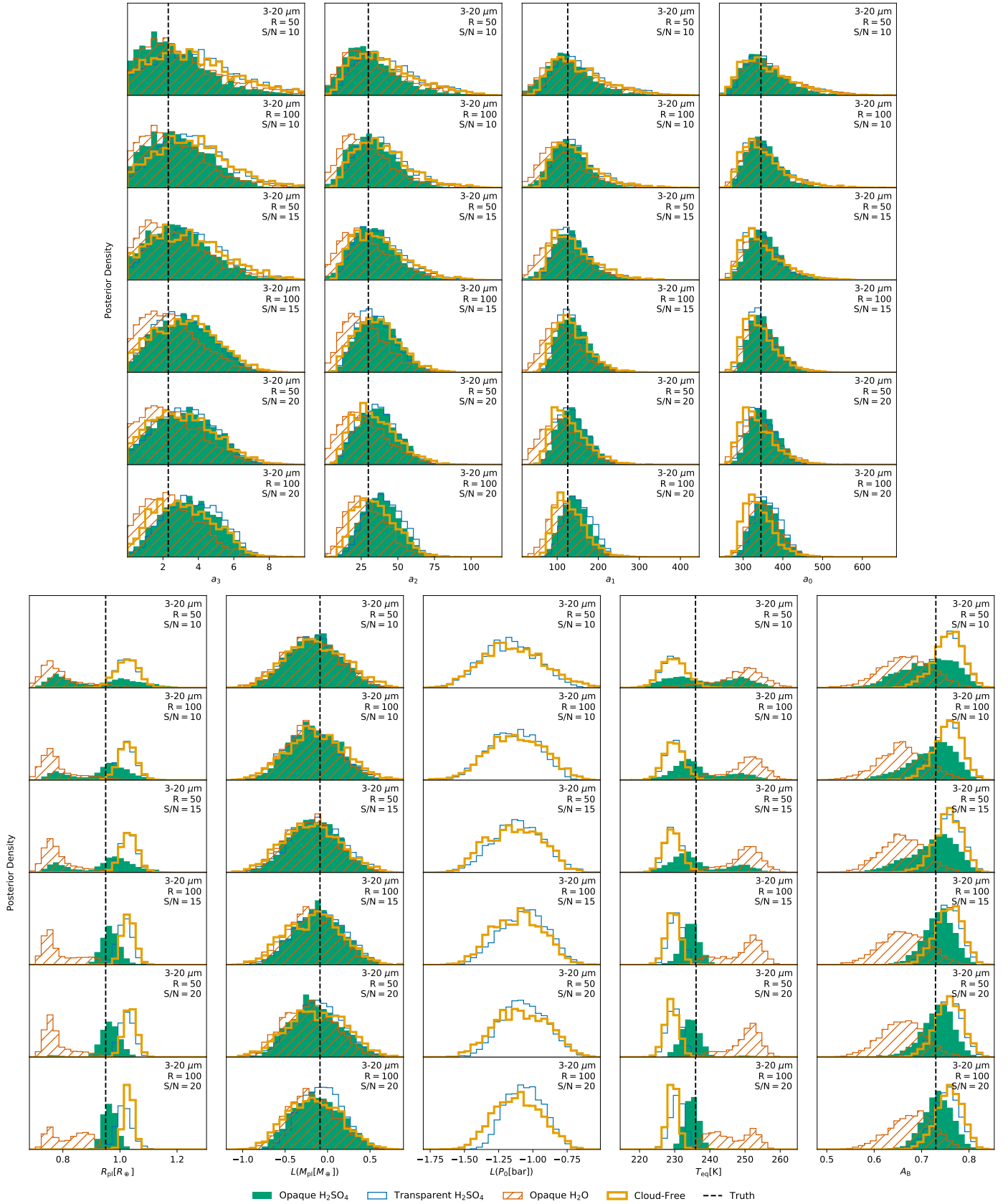


Fig. D.1. Model parameter posteriors for increased quality 3 – 20 μm Venus-twin spectra (opaque H_2SO_4 clouds). Here, $L(\cdot)$ abbreviates $\log_{10}(\cdot)$. Dashed black lines indicate the true values (see Table 1). Solid green area – opaque H_2SO_4 clouds; thin-blue outlined area – transparent H_2SO_4 clouds; orange-hatched area – opaque H_2O clouds; thick-yellow outlined area – cloud-free. Rows (top to bottom) represent different input qualities: $R = 50, S/N = 10$; $R = 100, S/N = 10$; $R = 50, S/N = 15$; $R = 100, S/N = 15$; $R = 50, S/N = 20$; $R = 100, S/N = 20$. Continuation in Fig. D.2.

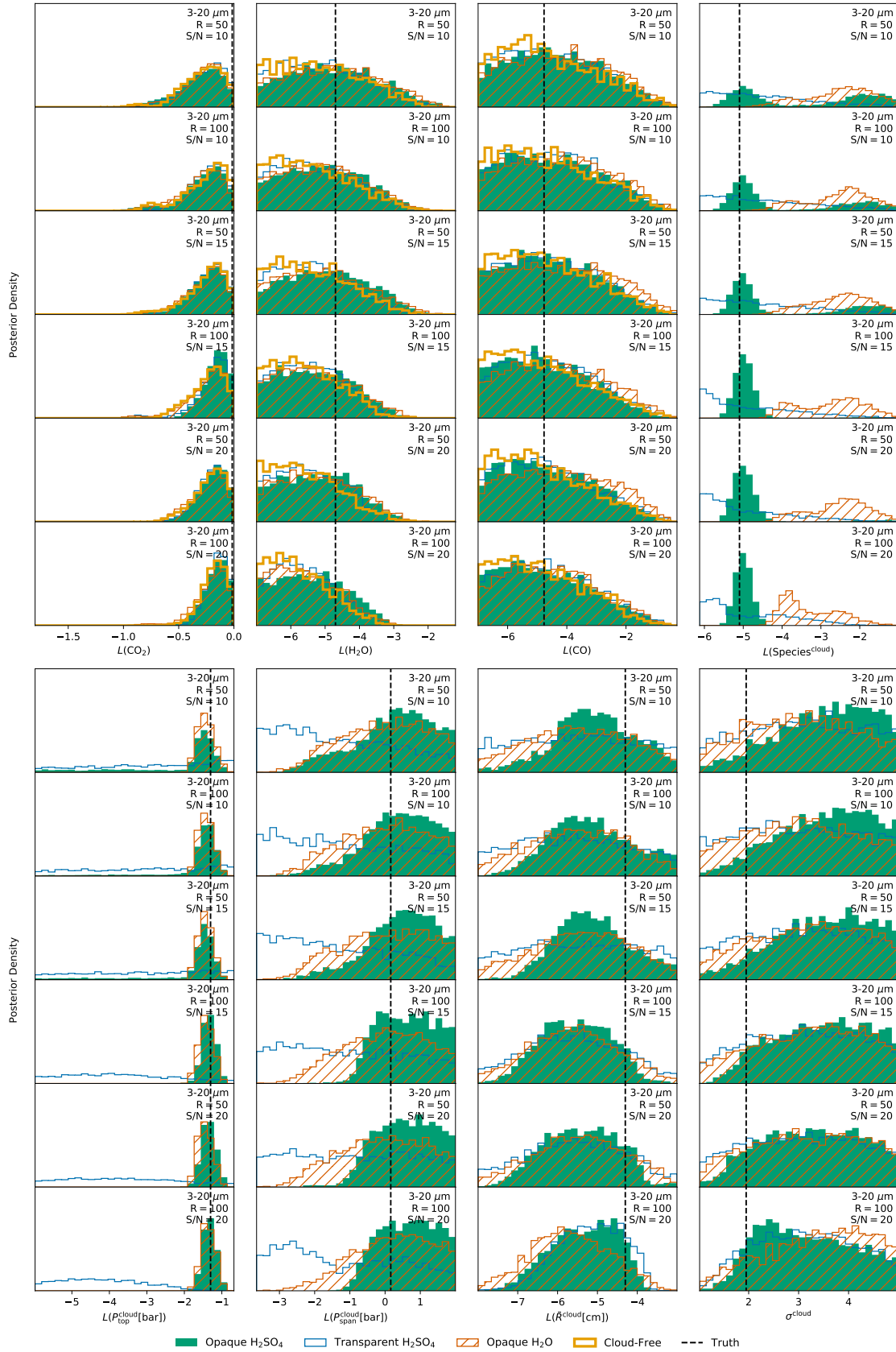


Fig. D.2. Continuation of Fig. D.1. For $\text{Species}^{\text{cloud}}$, the true value is the H_2SO_4 mass fraction in the Venus-twin model.

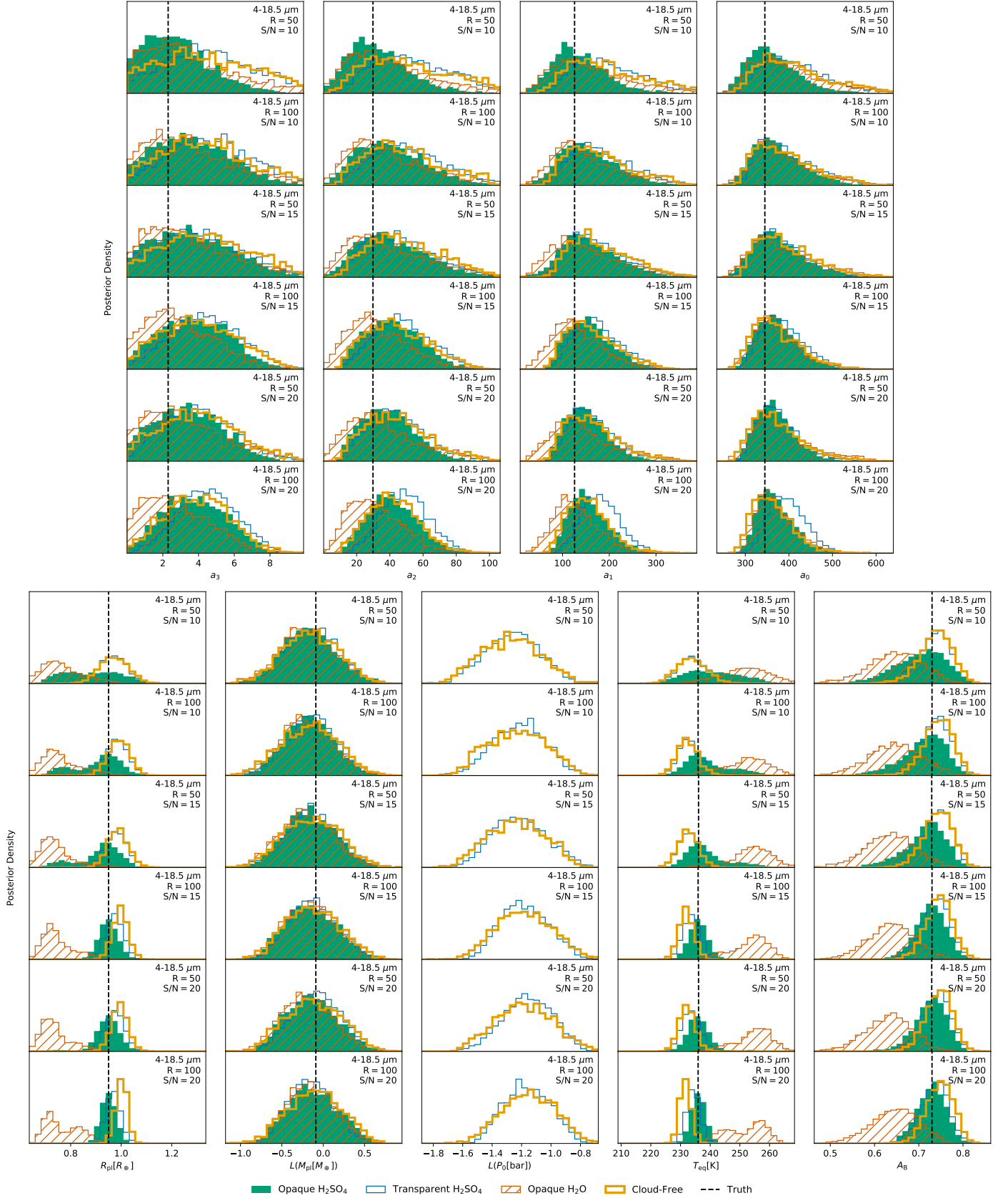


Fig. D.3. Model parameter posteriors for increased quality 4 – 18.5 μm Venus-twin spectra (opaque H_2SO_4 clouds). Here, $L(\cdot)$ abbreviates $\log_{10}(\cdot)$. Dashed black lines indicate the true values (see Table 1). Solid green area – opaque H_2SO_4 clouds; thin-blue outlined area – transparent H_2SO_4 clouds; orange-hatched area – opaque H_2O clouds; thick-yellow outlined area – cloud-free. Rows (top to bottom) represent different input qualities: $R = 50, S/N = 10$; $R = 100, S/N = 10$; $R = 50, S/N = 15$; $R = 100, S/N = 15$; $R = 50, S/N = 20$; $R = 100, S/N = 20$. Continuation in Fig. D.4.

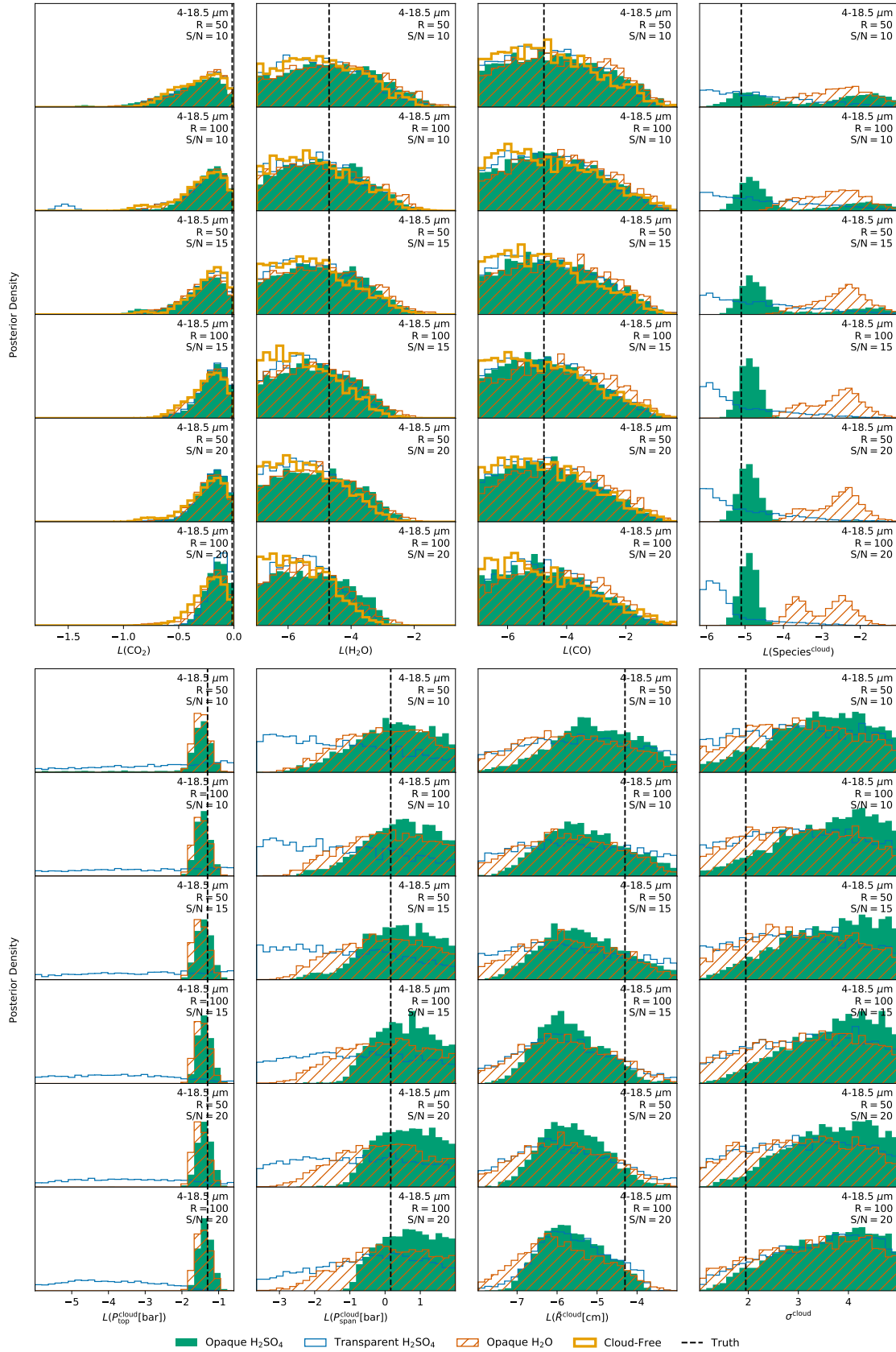


Fig. D.4. Continuation of Fig. D.3. For $\text{Species}^{\text{cloud}}$, the true value is the H₂SO₄ mass fraction in the Venus-twin model.

Appendix E: Cloudy versus icy scenario – Comparison of absolute fluxes

In Fig. E.1, we compare the absolute fluxes of the cloudy and icy exoplanet scenarios discussed in Sect. 4.1.2. In the UV/O/NIR, we show the noise expected for observations of the cloudy Venus-twin exoplanet with the HabEx + Starshade (Gaudi et al. 2020) or LUVOIR A (The LUVOIR Team 2019) instruments ($S/N = 10$ at $0.6 \mu\text{m}$, calculated with the NASA-GSFC Planetary Spectrum Generator; Villanueva et al. 2018). In the MIR, we show the wavelength-dependent expected LIFESIM noise ($S/N = 10$ at $11.2 \mu\text{m}$).

The increase in the HabEx noise at roughly $1 \mu\text{m}$ is caused by the HabEx + Starshade instrument design. Above $1 \mu\text{m}$, the throughput of the instrument is strongly decreased. Thus, the R of the spectrum above $1 \mu\text{m}$ is decreased to increase the S/N of the measured spectrum. However, the Venus-twin spectrum plotted in Fig. E.1 is constant in R over the full wavelength range considered. Further, the integration time is also constant over the full spectrum. Thus, the HabEx + Starshade noise above $1 \mu\text{m}$ is significantly larger than at shorter wavelengths.

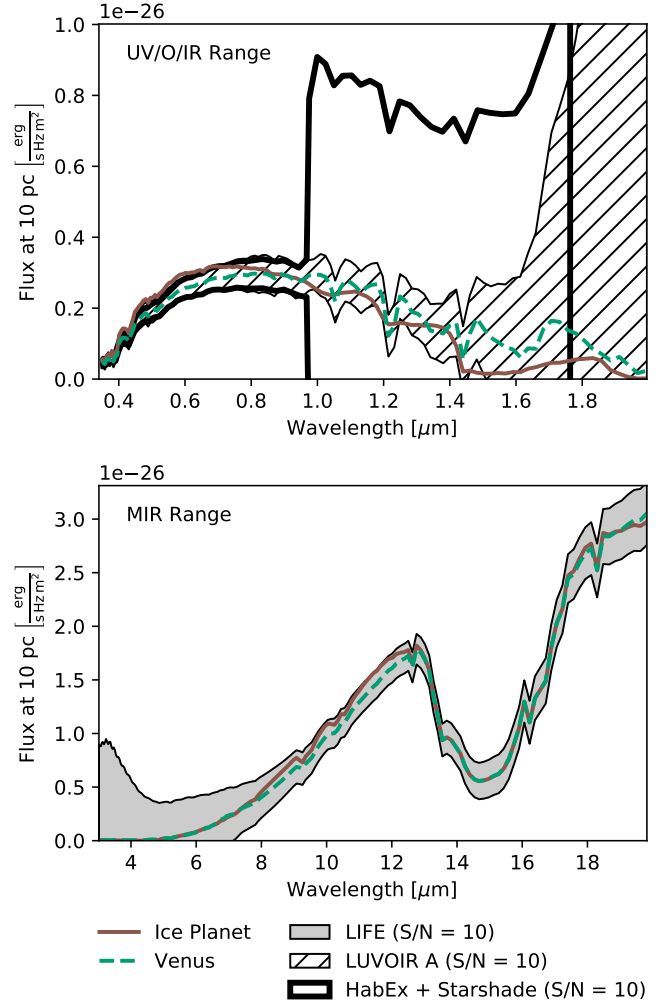


Fig. E.1. Flux of a cloudy Venus-twin exoplanet (opaque H_2SO_4 clouds; dashed green line) compared to an icy world with a thin CO_2 -dominated atmosphere (solid brown line). In the top panel, we plot the UV/O/NIR wavelength range. The hatched area represents the expected wavelength-dependent $S/N = 10$ noise for the LUVOIR A mission concept (The LUVOIR Team 2019). The thick, solid black line represents the expected $S/N = 10$ noise for the HabEx + Starshade mission concept (Gaudi et al. 2020). In the bottom panel, we plot the MIR wavelength range. The gray shaded region indicates the 1σ LIFESIM noise level at $S/N = 10$.