Astronomy & Astrophysics

Indications for very high metallicity and absence of methane in the eccentric exo-Saturn WASP-117b

Ludmila Carone[1], Paul Mollière[1], Yifan Zhou[2], Jeroen Bouwman[1], Fei Yan[3], Robin Baeyens[4], Dániel Apai[5,6], Nestor Espinoza[7], Benjamin V. Rackham[8,⋆], Andrés Jordán[9,10], Daniel Angerhausen[11,12,13,⋆⋆], Leen Decin[4], Monika Lendl[14], Olivia Venot[15], and Thomas Henning[1]

[1] Max-Planck-Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany
e-mail: `carone@mpia.de`
[2] Department of Astronomy/McDonald Observatory, The University of Texas, 2515 Speedway, Austin, TX, 78712, USA
[3] Institut für Astrophysik, Georg-August-Universität, Friedrich-Hund-Platz 1, 37077 Göttingen, Germany
[4] Instituut voor Sterrenkunde, KU Leuven, Celestijnenlaan 200D, 3001 Leuven, Belgium
[5] Steward Observatory, The University of Arizona, 933 N. Cherry Avenue, Tucson, AZ 85721, USA
[6] Lunar and Planetary Laboratory, The University of Arizona, 1629 E. Univ. Blvd., Tucson, AZ 85721, USA
[7] Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD, USA
[8] Department of Earth, Atmospheric and Planetary Sciences, and Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[9] Facultad de Ingeniería y Ciencias, Universidad Adolfo Ibáñez, Av. Diagonal las Torres 2640, Peñalolén, Santiago, Chile
[10] Millennium Institute for Astrophysics, Chile
[11] ETH Zürich, Inst. f. Teilchen- und Astrophysik, Zürich, Switzerland
[12] Center for Space and Habitability, University of Bern, Bern, Switzerland
[13] Blue Marble Space Institute of Science, Seattle, USA
[14] Observatoire de Genève, chemin des maillettes 51, 1290 Sauverny, Switzerland
[15] Laboratoire Interuniversitaire des Systèmes Atmosphériques (LISA), UMR CNRS 7583, Université Paris-Est-Créteil, Université de Paris, Institut Pierre Simon Laplace, Créteil, France

ABSTRACT

Aims. We investigate the atmospheric composition of the long-period (P_{orb} = 10 days) eccentric exo-Saturn WASP-117b. WASP-117b could be similar in atmospheric temperature and chemistry to WASP-107b. In mass and radius, WASP-117b is similar to WASP-39b, which allows a comparative study of these planets.
Methods. We analyzed a near-infrared transmission spectrum of WASP-117b taken with the *Hubble* Space Telescope (HST) WFC3 G141, which was reduced with two independent pipelines. High-resolution measurements were taken with VLT/ESPRESSO in the optical.
Results. We report the robust (3σ) detection of a water spectral feature. In a 1D atmosphere model with isothermal temperature, uniform cloud deck, and equilibrium chemistry, the Bayesian evidence of a retrieval analysis of the transmission spectrum indicates a preference for a high atmospheric metallicity [Fe/H] = $2.58^{+0.26}_{-0.37}$ and clear skies. The data are also consistent with a lower metallicity composition [Fe/H] < 1.75 and a cloud deck between $10^{-2.2}$ and $10^{-5.1}$ bar, but with weaker Bayesian preference. We retrieve a low CH_4 abundance of $<10^{-4}$ volume fraction within 1σ and $<2 \times 10^{-1}$ volume fraction within 3σ. We cannot constrain the equilibrium temperature between theoretically imposed limits of 700 and 1000 K. Further observations are needed to confirm quenching of CH_4 with $K_{zz} \geq 10^8$ cm^2 s^{-1}. We report indications of Na and K in the VLT/ESPRESSO high-resolution spectrum with substantial Bayesian evidence in combination with HST data.

Key words. methods: observational – planets and satellites: atmospheres

1. Introduction

The past years have revealed a large diversity in the atmospheres of transiting extrasolar gas planets colder than 1000 K and smaller than Jupiter, as well as a lack in methane in their atmospheres (e.g., Kreidberg 2015; Kreidberg et al. 2018; Wakeford et al. 2017, 2018; Benneke et al. 2019; Chachan et al. 2019). Whereas the atmospheric chemistry of hot Jupiters can be explained mainly with equilibrium chemistry, disequilibrium chemistry is expected to become more important for these cooler planets through vertical quenching, which results in an underabundance of CH_4 compared to predictions from equilibrium chemistry (Crossfield 2015). In principle, CH_4 is readily detectable alongside H_2O in the near- to mid-infrared.

While the quenching of CH_4 was first confirmed in brown dwarfs and later in directly imaged exoplanets (Barman et al. 2011a,b; Moses et al. 2016; Miles et al. 2018; Janson et al. 2013), observing disequilibrium chemistry in transiting tidally locked extrasolar gas planets has been more challenging. Quantifying CH_4 quenching reliably in transiting exoplanets to compare these disequilibrium chemistry processes to those occurring in brown dwarfs and directly imaged planets will be very illuminating to explore dynamical differences in different substellar atmospheres.

⋆ 51 Pegasi b Fellow.
⋆⋆ CSH Fellow.

Disequilibrium chemistry depends on vertical atmospheric mixing (K_{zz}), which provides a link between the observable atmosphere and deeper layers (see, e.g., Agúndez et al. 2014). Dynamical processes are further expected to be very different in tidally locked exoplanets compared to brown dwarfs just because the rotation and irradiation regime are different (Showman et al. 2014, 2019). For example, tidal locking slows down the rotation periods of transiting exoplanets to a few days, which is very slow in comparison to brown dwarfs with rotation periods shorter than one day, for example (see, e.g., Apai et al. 2017). Disequilibrium chemistry processes across a whole range of substellar atmospheres could thus shed light on deep atmospheric processes and on why the detection of methane has been proven to be very difficult so far in transiting exoplanets.

To this date, both methane and water have only been reliably detected for one mature extrasolar gas planet: at the dayside of the warm (600–850 K) Jupiter HD 102195b (Guilluy et al. 2019). For HAT-P-11b, the presence of methane was inferred with 1D atmosphere models for retrieval by Chachan et al. (2019), where the authors found that the observed steep rise in transit depth for wavelengths $\geq 1.5\,\mu m$ in their data is not present when they simulate the spectrum after removing CH_4 opacities from the model. Their data were obtained like the data we report here with the Wide Field Camera 3 (WFC3) on the *Hubble* Space Telescope (HST), using the G141 grism. The authors were unable to constrain CH_4 abundances further, however (Chachan et al. 2019). For the warm super-Neptune WASP-107b (Kreidberg et al. 2018), methane quenching is suggested because methane is absent (no discernible opacity source that would translate into an increase in transit depth for $\lambda \geq 1.5\,\mu m$ in their HST/WFC3 G141 data), while water was clearly detected. Benneke et al. (2019) also reported methane depletion for the mini-Neptune GJ 3470b compared to disequilibrium chemistry models.

Many of the transiting exoplanets for which disequilibrium chemistry may play a role were also found to be less massive than Jupiter, ranging from mini-Neptune to Saturn mass. These objects show a spread in metallicity that ranges from very low, $\leq 4\times$ solar, for the Neptune HAT-P-11b (Chachan et al. 2019) to very high values, $>100\times$ metallicity, for the exo-Saturn WASP-39b (Wakeford et al. 2018).

In this paper, the transiting exo-Saturn WASP-117b joins the ranks of super-Neptune-mass exoplanets with atmospheric composition constraints. WASP-117b is close in mass ($0.277 M_{Jup}$) and radius ($1.021 R_{Jup}$) to WASP-39b, which was found to be metal-rich (Wakeford et al. 2018). In temperature, it could be 900 K or colder and thus similar in atmospheric chemistry to WASP-107b (Kreidberg et al. 2018). Therefore this planet will aid a comparative analysis of methane content and metallicity from the Neptune- to Saturn-mass range.

Furthermore, WASP-117b orbits its quiet F-type main-sequence star on an eccentric orbit ($e = 0.3$) with a relatively long orbital period of ≈ 10 days (Lendl et al. 2014; Mallonn et al. 2019). Lendl et al. (2014) further report equilibrium temperatures of 900 and 1200 K during apastron and periastron, respectively, assuming an albedo of $\alpha = 0$ and efficient, uniform redistribution of absorbed stellar energy over the whole planet. When we use the same formalism to calculate atmospheric temperatures during one orbit, the planet would reside on its eccentric orbit for several days in the hot-temperature ($T > 1000$ K) regime, characterized by CO as the main carbon-bearing species at $p = 1$ bar for solar metallicity. The planet would also reside several days in the warm ($T < 1000$ K) regime, for which CH_4 becomes dominant in the observable atmosphere, and for which, depending on the strength of vertical mixing
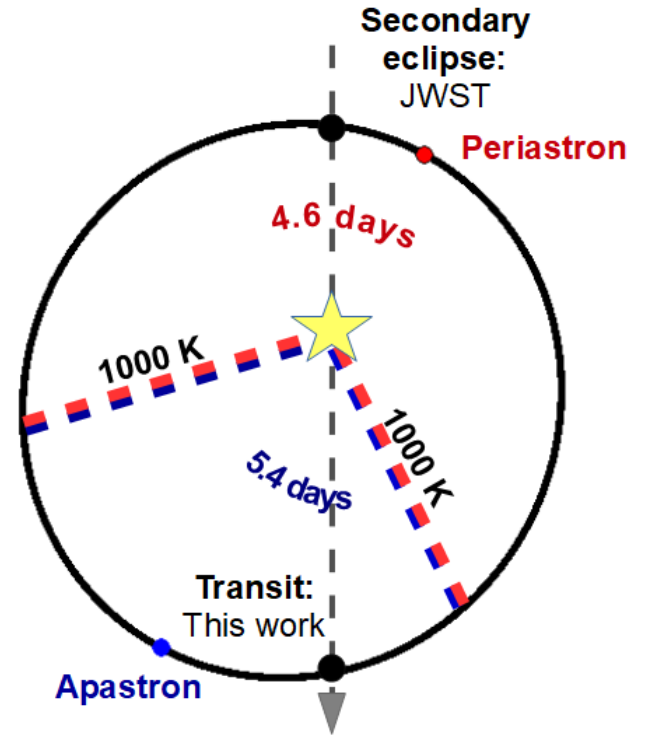

Fig. 1. Orbit orientation of WASP-117b and the expected equilibrium temperature, assuming albedo $\alpha = 0$ and to first-order instant adjustment of atmospheric temperature to changes of incoming stellar flux as used in Lendl et al. (2014). Temperatures about 1000 K for solar metallicity denote the transition in equilibrium chemistry between CO (for the hot regime $T_{eq} \gtrsim 1000$ K) and CH_4 (for the warm $T_{eq} \lesssim 1000$ K) (Crossfield 2015). In the latter case, however, formation of CH_4 is expected to be quenched because of disequilibrium chemistry. The planet is expected to remain 4.6 days in the hot regime and 5.4 days in the colder regime.

(K_{zz}) and atmospheric metallicity, disequilibrium chemistry of methane potentially becomes observable (Fig. 1). This is a different situation than for WASP-107b and WASP-39b, which reside on tighter circular orbits all the time in the same temperature and thus chemistry regime.

The atmospheric properties of WASP-117b will therefore shed further light on the diversity in basic atmospheric properties of transiting super-Neptunes such as atmospheric metallicity and CH_4 quenching at pressures deeper than 0.1 bar that can lead to depletion by several orders of magnitude compared to equilibrium chemistry for atmospheric temperatures colder than 1000 K. In addition, because of its the relatively long eccentric orbit, this exo-Saturn will also allow us to understand how exoplanets on wider noncircular orbits that are subjected to varying irradiation and atmospheric erosion differ from exoplanets on tighter, circular orbits.

Last but not least, we show that it is possible and worthwhile to characterize exo-Saturns on orbital periods of 10 days with single-epoch observations from space. Thus, our WASP-117b observations are prototype observations for other transiting exoplanets with orbital periods of 10 days and longer, such as K2-287 b (Jordán et al. 2019) and similar objects that were discovered after WASP-117b (Brahm et al. 2018; Jordán et al. 2020; Rodriguez et al. 2019).

We report here the first observations to characterize the atmosphere of the exo-Saturn WASP-117b (Sect. 2) in the near-infrared with HST/WFC3 (Sect. 2.1) and in the optical

with the Echelle SPectrograph for Rocky Exoplanets and Stable Spectroscopic Observations (ESPRESSO) on the Very Large Telescope (VLT) (Sect. 2.2). We investigate the significance of the water detection with HST/WFC3 and Na and K with VLT/ESPRESSO in Sect. 3.1. We perform atmospheric retrieval with an atmospheric model (Sect. 3.2) mainly on the HST/WFC3 data to constrain basic properties of the atmosphere of the exo-Saturn WASP-117b (Sects. 3.3 and 3.4). We then present improved stellar rotation and spin-orbit alignment parameters for WASP-117b based on the Rossiter-McLaughlin effect measured with VLT/ESPRESSO (Sect. 3.5).

We also compared our data with broadband transit observations obtained by the Transiting Exoplanet Survey Satellite (TESS) in the optical (Sect. 3.6) and discuss the effect of stellar activity to explain the discrepancy between the TESS transit depth and our WFC3/NIR transmission spectrum, which were obtained at different times (Sect. 3.7). To provide better context for the low CH_4 abundances that we find, we calculated a small set of representative chemical models for different temperatures and metallicities (Sect. 4). To guide future observations to further characterize the system, we also generated synthetic spectra based on our retrieved atmospheric models for wavelength ranges covered by the UVIS channel of HST/WFC3 and the *James Webb* Space Telescope (JWST; Sect. 5). We investigated the benefit of additional measurements to constrain the presence of a haze layer and to distinguish between high- ([Fe/H] > 2.2) and low- ([Fe/H] < 1.75) metallicity models. We discuss the atmospheric properties of WASP-117b and how they compare to other super-Neptunes such as WASP-39b and WASP-107b in Sect. 6. We provide a summary of our findings in Sect. 7 and outline steps for future observations of the eccentric exo-Saturn WASP-117b in Sect. 8.

2. Observations

We report here a near-infrared transmission spectrum measured with the HST (Program GO 15301, PI: L. Carone) and a high-resolution optical spectrum measured with VLT/ESPRESSO (PI: F. Yan). In the following, we describe the set of observations and the data reduction. For the HST/WFC3 raw data, we employed two completely independent data-reduction pipelines for a robust retrieval of the atmospheric signal. The relatively long transit duration of 6 h required a long and stable observation by both HST and VLT/ESPRESSO, which were successfully carried out with both instruments.

2.1. HST observation and data reduction

We observed one transit of WASP-117 b with the HST Wide Field Camera 3 (WFC3) on UT 20–21 September 2019. The transit observation consists of 11 consecutive HST orbits. This HST observation benefited from reobservations of the WASP-117b transit in July 2017 with broadband photometry using two small telescopes: one in Chile, the Chilean-Hungarian Automated Telescope 0.7 m (CHAT, PI: Jordán), and one in South Africa (1m, Los Cumbres Observatory). The combined observations allowed us to improve the uncertainties in mid-transit time from 2 h (based on Lendl et al. 2014) to 2 min (Mallonn et al. 2019).

For the HST measurement, an image of the target with the F126N filter, using NSAMP = 3, rapid readout mode was taken at the start of each orbit. This image was used to anchor the wavelength calibration for each orbit. Otherwise, we obtained
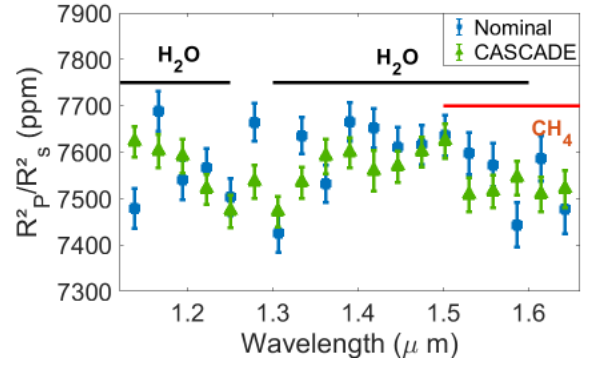

Fig. 2. Final HST/WFC3 transmission spectra from the nominal (blue) and CASCADE (green) pipeline, covering a wavelength range of 1.125–1.65 μm. The average transit depths are 7573 ± 43 ppm and 7553 ± 39 ppm, respectively. Black and red lines indicate the positions of the H_2O and CH_4 absorption bands, respectively, in the observed spectral range.

time-series spectra with the G141 grism, which covers the wavelength range 1.1–1.7 μm. We used the NSAMP = 15, SPARS 10 readout mode for these, in spatial scanning mode with a scan rate of 0.12 arcseconds sec^{-1} and scan direction "round trip".

For raw data reduction, we employed two separate and completely independent pipelines to ensure that the derived atmospheric spectra are robust. We call the two pipelines henceforth nominal pipeline and CASCADE pipeline; their respective data reduction procedures are described in Appendix A.

2.1.1. Final transmission spectra

After data reduction, as described in Appendices A.1 and A.2, we derived two transmission spectra (Fig. 2). We excluded points outside of the wavelength range 1.125–1.65 μm because the instrument transmission has steeper gradients in these regions than in the center of the spectrum. Jittering in the wavelength direction together with the strong variation in transmission profile change can introduce large systematic errors. Because a strong methane absorption feature covered by the WFC3/G141 grism is centered at 1.62 μm, where H_2O has no absorption, we can still judge whether methane is present or absent based on a spectrum covering 1.125–1.65 μm. This wavelength range was used for our further analysis.

The WASP-117b HST/WFC3 data indicate a muted water absorption spectrum with a slight drop in transit depth between 1.4 and 1.55 μm, which indicates H_2O absorption centered at 1.45 μm. There is potentially a second peak toward the shorter wavelengths at the edge of the spectral window. There is no strong CH_4 absorption feature toward the longest wavelengths in the observed spectral window. The feature is present in both spectra (nominal and CASCADE), using two independent reduction pipelines for the same dataset.

However, it also appears that the nominal spectrum is slightly shifted toward deeper transit depths than the CASCADE spectrum. Furthermore, some data points appear to deviate more than 3σ from each other. We attribute these deviations to differences at the raw light-curve extraction, as described in the next subsection.

2.1.2. Comparison of employed data pipelines

We find that the transmission spectra from the two pipelines demonstrate discrepancies (Fig. 2). First, there is disagreement

in the average transit depth. Calculating the standard error of the mean $\sigma/\sqrt{n}$ (SEM) gives values of $\text{SEM}_{\text{nominal}} = 18.2$ ppm for the nominal and $\text{SEM}_{\text{CASCADE}} = 11.0$ ppm for the CASCADE pipeline, respectively. The deviation between the mean transit depths is thus $1.1 \cdot \text{SEM}_{\text{nominal}}$ or $1.9 \cdot \text{SEM}_{\text{CASCADE}}$, which is a significant but not substantial ($>3\text{SEM}$) deviation. Second, there are more significant inconsistencies at several wavelength channels. At $1.28 \, \mu\text{m}$, the difference in transit depths is nearly 3σ.

We discuss the possible sources of these inconsistencies and the implications for our retrieval results further. The uncertainties in our transmission spectra include observational noise (photon noise, readout noise, dark current, etc.) and noise correlated with the detector systematics. The two data reduction pipelines, nominal, and CASCADE, are fully independent in the treatment of these uncertainties. In general, each pipeline consists of two modules: first, the light-curve extraction module, which reduces the time series of spectral images to spectral light curves; second, the light-curve modeling module, which fits instrument systematics and transit profile models to the light curve and measures transit depths. Both modules can cause discrepancies in the transmission spectrum.

To locate the source of the inconsistencies, we performed the following analysis: ee used the light-curve modeling module of the nominal pipeline (RECTE) to fit the uncalibrated light curves extracted in the CASCADE pipeline. We label the results "CASCADExRECTE".

In Fig. 3 (top panel), we compare the transmission spectra of nominal, CASCADE, and RECTExCASCADE pipelines without accounting for the different transit depths. In Fig. 3 (bottom panel), we scale the three spectra to the same average transit depth. The second comparison therefore shows deviations on the shape of the spectra more clearly. Furthermore, weighted residual sums of squares (RSS)

$$\text{RSS} = \sum_{i}^{n} \frac{(\text{spec(i)}_1 - \text{spec(i)}_2)^2}{(\text{uncertainty}^2(i)_1 + \text{uncertainty}^2(i)_2)}, \qquad (1)$$

where we sum over the total number n of wavelength bins i, were derived to evaluate the difference statistically. For nonnormalized spectra, the RSS between nominal and CASCADE is 35.2 (degrees of freedom, DOF = 19), and the RSS between CASCADExRECTE and CASCADE is 33.4 (DOF = 19). After normalization, the RSS between nominal and CASCADE is 32.4 (degrees of freedom, DOF = 18), and the RSS between CASCADExRECTE and CASCADE is 11.6 (DOF = 19).

This shows that with the same light-curve extraction module and scaling the spectra to the same average depth, the light-curve fitting modules in the two pipelines provide fully consistent results. This is also shown in Fig. 3 (bottom panel) by comparing the spectra CASCADExRECTE and CASCADE. All data points, except at $1.58 \, \mu\text{m}$, agree within 1σ with each other.

From these comparison results, we conclude that 1. the difference in light-curve extraction introduces random errors and causes the transmission spectra to differ at several wavelength channels; 2. the difference in systematic correction causes a slight offset ($<1\sigma$) in the average transit depths; and 3. after they are scaled to the same average depth, the spectra from the two pipelines agree with each other, that is, the shape of the spectra are consistent. The random errors and the offset yield apparent 3σ differences in single points in the final spectra (Fig. 2).

The largest inconsistency stems from the extracted light curves and is likely a consequence of the relatively strong telescope pointing drift during the WASP-117b observations. In
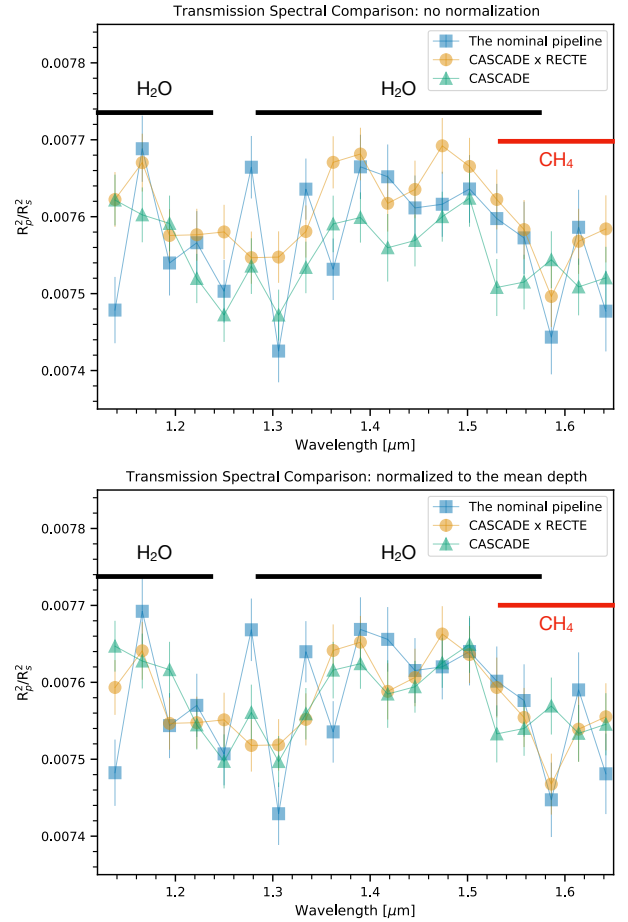
Fig. 3. Comparison of transmission spectra obtained with different pipelines. *Upper panel*: original spectra. *Lower panel*: spectra scaled to the same average depth. After scaling, the spectra using the same raw data extraction as CASCADE, but different systematic treatment (CASCADE and CASCADExRECTE), are consistent with each other.

the wavelength dispersion direction, the peak-to-peak drift distance is ~ 0.4 pixels, greater than the sub-0.1-pixel level normally seen in HST scanning mode observations (e.g., Kreidberg et al. 2014). Different treatments in aligning the spectra between the two pipelines can lead to the inconsistency of the final transmission spectra. To demonstrate this point, we performed another reduction with a slightly modified nominal pipeline, in which we applied a Gaussian filter to the images before aligning them in the dispersion direction. The kernel of the Gaussian filter was the same as the spectral resolution element of WFC3/G141. This treatment reduced the column-to-column pixel value variations and thus decreased the effect of pointing drift. We derived the transmission spectrum using this pipeline and compared it with that from CASCADE (Fig. 4). The two spectra are fully consistent with an RSS value of 15.6 (DOF = 19).

Regardless of the treatment of the pointing drift, the shape of the spectra derived from the nominal and CASCADE pipelines are consistent with each other. We therefore do not expect substantial discrepancies in the outcomes of the atmospheric retrieval for the two spectra.

Our goal is to robustly determine the planet properties, which are not dependent upon the choice of systematics models. We therefore used both data pipelines for a most robust interpretation of the single-epoch transit observation with HST/WFC3. In the following, we mainly use results derived from the nominal
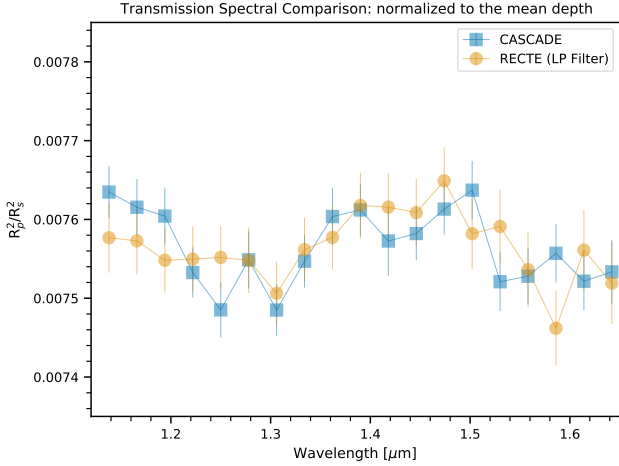
Fig. 4. Comparison of transmission spectra of a slightly modified nominal pipeline and the CASCADE pipeline. In the nominal pipeline, a Gaussian filter was applied to the images before wavelength registration to address the issue of telescope pointing drifts in the wavelength direction. The two spectra are fully consistent, which supports the assumption that discrepancies between two pipeline are due to their variations in wavelength registration in the raw light-curve extraction stage.

pipeline, which is well established for further analysis, for instance, with VLT/ESPRESSO (Sect. 2.2) and TESS data (Sect. 3.6).

2.2. VLT/ESPRESSO observations

To also probe gas-phase absorption of Na and K in the atmosphere of WASP-117b, we observed one transit of the planet on 24/25 October 2018 under the ESO program 0102.C-0347 with the ultrastable fibre-fed echelle high-resolution spectrograph ESPRESSO (Pepe et al. 2010) mounted on the VLT. The observation was performed with the 1-UT high-resolution and fast read-out mode. We set the exposure time as 300 s and pointed fiber B on sky. We observed the target continuously from 23:56 UT to 09:28 UT and obtained 98 spectra in total. These spectra have a resolution of $R \sim 140\,000$ and a wavelength coverage of 380–788 nm.

2.2.1. Data reduction

We reduced the raw spectral images with the ESPRESSO data-reduction pipeline (version 2.0.0). The pipeline produced spectra with sky background corrected by subtracting the sky spectra measured from fiber B from the target spectra measured from fiber A. The pipeline also calculated the radial velocity (RV) of the stellar spectrum with the cross-correlation technique. We discarded six spectra that have relatively low signal-to-noise ratios. Of the remaining 92 spectra, 59 spectra were observed during transit, and 33 spectra were observed during the out-of-transit phase.

2.2.2. Transmission spectra of sodium and potassium

To obtain the planetary transmission spectra of sodium (Na) and potassium (K), we implemented the procedures described below.

(1) Removal of telluric lines. There are telluric Na emission lines around the sodium doublet, which we corrected for using fiber B spectra. We further corrected the telluric H_2O

and O_2 absorption lines by employing the theoretical H_2O and O_2 transmission model described in Yan et al. (2015a). The corrections were performed in the Earth's rest frame.

(2) Removal of stellar lines. In order to remove the stellar lines, we first aligned all the spectra to the stellar rest frame by correcting the barycentric Earth radial velocity (BERV) and the stellar systemic velocity. We then generated a master spectrum by averaging all the out-of-transit spectra and divided each observed spectrum with this master spectrum. The residual spectra were then filtered with a Gaussian function ($\sigma \sim 3$ Å) to remove large-scale features.

(3) Correction for the CLV and RM effects. During the planet transit, the observed stellar line profile has variations originating from several effects. The Rossiter McLaughlin effect (Queloz et al. 2000) and the center-to-limb variation (CLV) effect (Yan et al. 2015b, 2017; Czesla et al. 2015) are two main effects. We followed the method described in Yan & Henning (2018) and Yan et al. (2019) to model the RM and CLV effects simultaneously. The stellar spectrum was modeled with the Spectroscopy Made Easy tool (Piskunov & Valenti 2017) and the Kurucz ATLAS12 model (Kurucz 1993). We used the stellar parameters from Lendl et al. (2014) except for the $v \sin i_\star$ and λ values, which were taken from the RM fit of the ESPRESSO RVs. The simulated line profile change due to the CLV and RM effects is weak and below the errors of the observed data. We subsequently corrected the CLV and RM effects for the obtained residual spectra.

(4) Obtaining the transmission spectrum. We shifted all the in-transit residual spectra to the planetary rest frame and added all these shifted spectra. Because of the high orbital eccentricity, the planetary orbital velocity changes from +10 km s^{-1} to +20 km s^{-1} during transit. In the planetary rest frame, the position of the stellar Na/K line therefore is approximately +15 km s^{-1} away from the expected planetary signal.

We investigated the Na D doublet lines (5891.584 Å and 5897.555 Å) and K D$_1$ line (7701.084 Å). The K D$_2$ line (7667.009 Å) is heavily affected by the dense O_2 lines and therefore was not used in the analysis. The listed wavelengths are in vacuum. The final transmission spectra are presented in Fig. 5. There are no strong absorption features (dip) at the expected wavelengths of the planetary Na/K lines within 3σ (indicated as dashed lines in the figure).

In the wavelength regions in which the stellar Na lines are located (i.e., ~0.3 Å away from the expected planetary signal) lie some spectral features. We attribute these features to the large errors of these points, however, because the flux inside the deep stellar Na line is significantly lower than the adjacent continuum.

The planetary Na and K lines in the VLT/ESPRESSO data are below 3σ significance. Thus, we decided not to perform atmospheric retrieval on the VLT/ESPRESSO data alone. Instead, we used the statistically significant atmosphere signal in the HST/WFC3 data in conjunction with VLT/ESPRESSO to strengthen the case that planetary Na and K lines are indeed present in the VLT/ESPRESSO data. This analysis is presented in Sect. 3.1.

3. WASP-117b atmospheric properties and improved planetary parameters

In this section, we first explore the significance of the water detection in the HST/WFC3 and the possible existence of
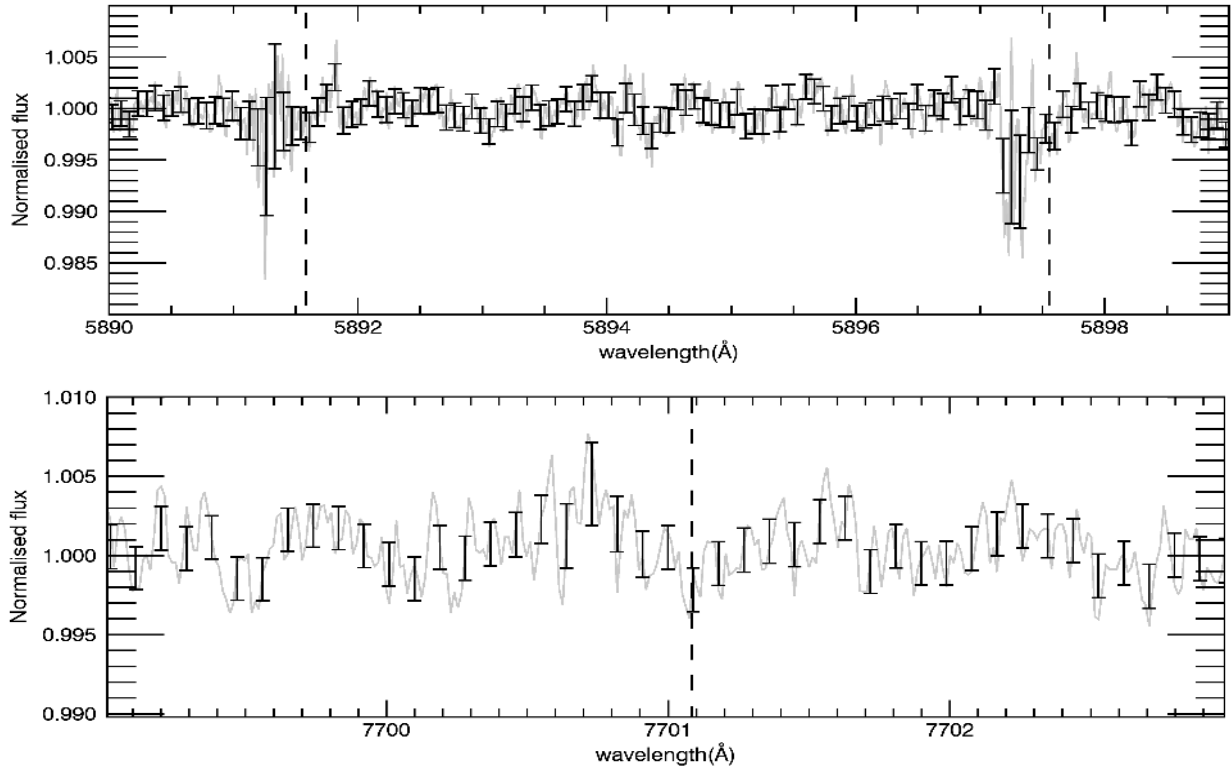
Fig. 5. Summary of the observation of WASP-117b with VLT/ESPRESSO. *Top*: normalised transmission spectrum around the Na D doublet lines. *Bottom*: normalised transmission spectrum around the K D$_1$ line. The center wavelengths of these absorption lines are indicated as dashed lines. The black data points are spectra binned every seven wavelength points, and the gray lines are original spectra. The stellar line profile change due to the CLV and RM effects is corrected. We display vacuum wavelengths. In normalized flux, planetary Na and K absorption appear as weak depressions around the dashed line.

weak Na and K in the VLT/ESPRESSO data. We then perform atmospheric retrieval with different models to interpret the atmospheric properties of the exo-Saturn WASP-117b.

3.1. Significance of water and Na/K detection

To identify the significance of the H$_2$O detection in the HST/WFC3 spectrum, we followed the approach of Benneke & Seager (2013) and ran a number of forward models of similar complexity against each other, using `petitRADTRANS` (Mollière et al. 2019). In this model, we assumed isothermal temperatures and a gray uniform cloud, which was modeled by setting the atmospheric opacity to infinity for $P > P_{cloud}$. Thus, P_{cloud} was treated as the pressure at the top of a fully opaque cloud. Furthermore, we included opacities of the following absorbers: H$_2$O, CH$_4$, N$_2$, CO, and CO$_2$. The mass fractions of absorbers are free parameters with priors in log space ranging from -10 to 0. For the remaining atmospheric mass, a mixture of H$_2$ and He was assumed with a ratio of 3:1. We retrieved the atmospheric reference pressure P_0 to reproduce the apparent size of the planet in the WFC3 wavelength range.

We quantified the significance of the observed molecular absorption features using the `MultiNest` sampling technique that enables quantifying and comparing model parameters and their significance. `MultiNest` is implemented within the python wrapper `PyMultiNest` (Buchner 2016).

Table 1 lists the (natural) log evidence (lnZ) and the Bayes factor B for different hypotheses. B is calculated through $B = \exp(\ln Z_{base} - \ln Z)$ from the `MultiNest` output for each model.

Figure 6 shows the full model, including H$_2$O, versus the model without H$_2$O for data derived with both pipelines.

Following Kass & Raftery (1995), we regard B values of 1–3, 3–20, 20–150, and >150 as 'weak' (or 'not worth more than a bare mention'), 'substantial', 'strong', and 'very strong' preference for a given hypothesis, respectively. Benneke & Seager (2013, Table 2), adopted from Trotta (2008), allows us further to translate B into lower limits on σ confidence levels. We note that Benneke & Seager (2013) adopted a different scale, where they labeled $B = 3$–12 weak, $B = 12$–150 moderate, and $B > 150$ strong.

Our statistical analysis yields a strong Bayesian preference ($B > 20$) for the detection of water in the WASP-117b observational data for both pipelines. This corresponds to a 3σ detection.

The same approach was also used for the combined HST/WFC3 (nominal) and VLT/ESPRESSO spectrum and the possible detection of Na and K. The full model was extended to also include K and Na opacities, again assuming their combined mass fraction to be a free parameter, but fixing the Na/K abundance ratio to the solar value (Asplund et al. 2009).

Figure 7 shows the full model, including Na and K, versus the model without Na and K for data derived with the nominal pipeline. The combined analysis of the HST/WFC3 and VLT/ESPRESSO yields still substantial ($B > 3$) evidence for Na and K in the data, which would correspond to $\approx 2.4\sigma$, using Benneke & Seager (2013, Table 3).

Corner plots of the full set of retrieved parameters for the full retrieval models and those without H$_2$O and without Na/K ($B > 3$) are presented in the appendix (Figs. B.1–B.8).

Table 1. Results of the Bayesian model comparison for H_2O in HST/WFC3 and Na/K in VLT/ESPRESSO.

Retrieval model	Parameters	Evidence $\ln Z$	Bayes factor B
HST/WFC3 for H_2O			
Full hypothesis	P_0, P_{clouds}, H_2O, CO_2, CH_4, CO, N_2	Nominal: -32.973 CASCADE: -19.394	Nominal: baseline CASCADE: baseline
H_2O removed	P_0, P_{clouds}, CO_2, CH_4, CO, N_2	Nominal: -36.042 CASCADE: -22.971	Nominal: 21.5 $(3\sigma_{H_2O})$ CASCADE: 35.8 $(3\sigma_{H_2O})$
CH_4 removed	P_0, P_{clouds}, H_2O, CO_2, CO, N_2	Nominal: -33.093 CASCADE: -19.300	Nominal: 1.1 CASCADE: 0.9
No clouds	P_0, H_2O, CO_2, CH_4, CO, N_2	Nominal: -33.828 CASCADE: -19.271	Nominal: 2.35 CASCADE: 0.88
HST/WFC3 and VLT/ESPRESSO for Na and K			
Full hypothesis	P_0, P_{clouds}, H_2O, CO_2, CH_4, CO, N_2, Na/K	Nominal: -174.929 CASCADE: -162.158	Nominal: baseline CASCADE: baseline
Na/K removed	P_0, P_{clouds}, H_2O, CO_2, CH_4, CO, N_2	Nominal: -176.918 CASCADE: -163.849	Nominal: 7.3 $(2.5\sigma_{Na,K})$ CASCADE: 5.4 $(2.4\sigma_{Na,K})$
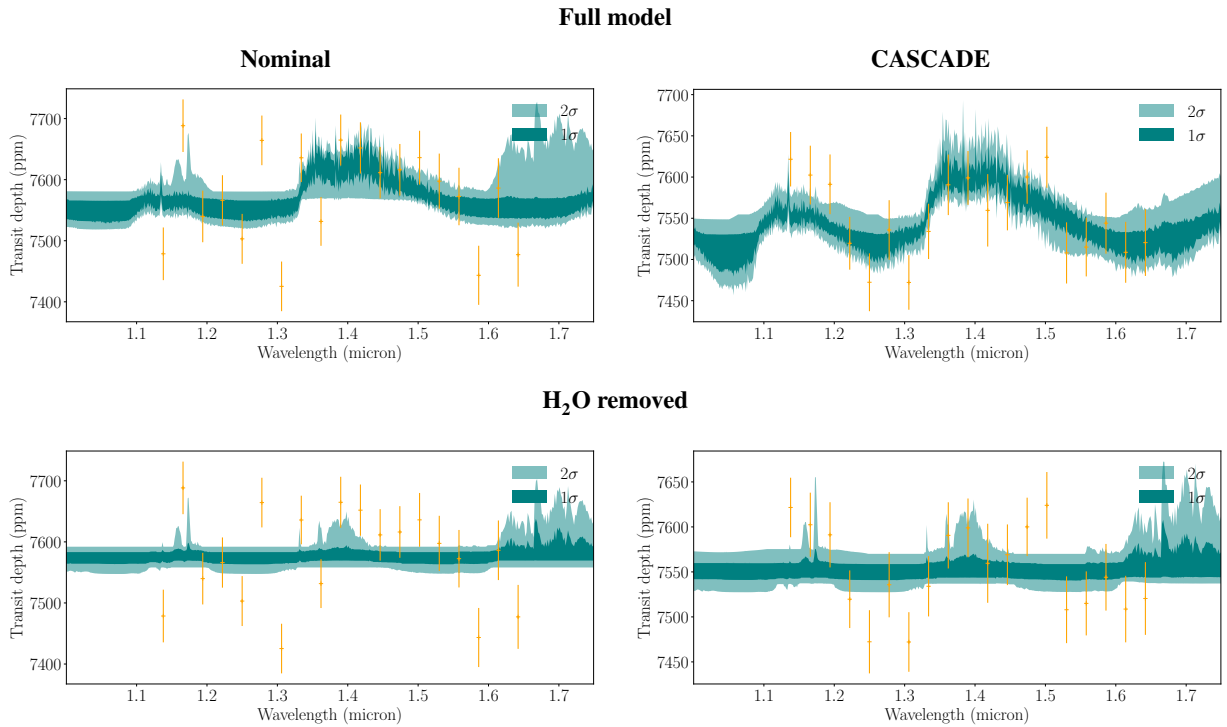
Full model

Fig. 6. Results of atmospheric retrieval for WASP-117b, using different opacities. *Top*: full retrieval model is fit to WASP-117b transmission spectrum reduced with the nominal pipeline (*left*) and CASCADE (*right*). *Bottom*: retrieval model with H_2O removed fit to WASP-117b transmission spectrum reduced with the nominal (*left*) and CASCADE pipeline (*right*). For all spectra, retrieved atmosphere models with confidence within 1σ (dark green) and 2σ (light green) are shown.

3.2. Atmospheric retrieval

To constrain the atmospheric parameters of WASP-117b further, we again ran forward atmospheric models using the open-source code petitRADTRANS (Mollière et al. 2019), this time retrieving the planetary metallicity and C/O value directly. Our retrieval setup was applied on the HST/WFC3 data first, and then on the combination of the HST/WFC3 and ESPRESSO data.

We constructed the retrieval forward model with the following rationale: in principle, retrievals of transmission spectra

allow many properties of the atmosphere to be constrained, such as the average terminator temperature and temperature gradient, and abundances of absorbers, and cloud properties such as the cloud base position, scale height, average particle sizes, and cloudiness fraction of the terminator (e.g., Barstow et al. 2013; Rocchetto et al. 2016; Line & Parmentier 2016; MacDonald & Madhusudhan 2017; Mollière et al. 2019; Barstow 2020). Differences between morning and evening terminators can also likely be constrained or lead to erroneous conclusions when they are ignored (MacDonald et al. 2020). The same holds for variations
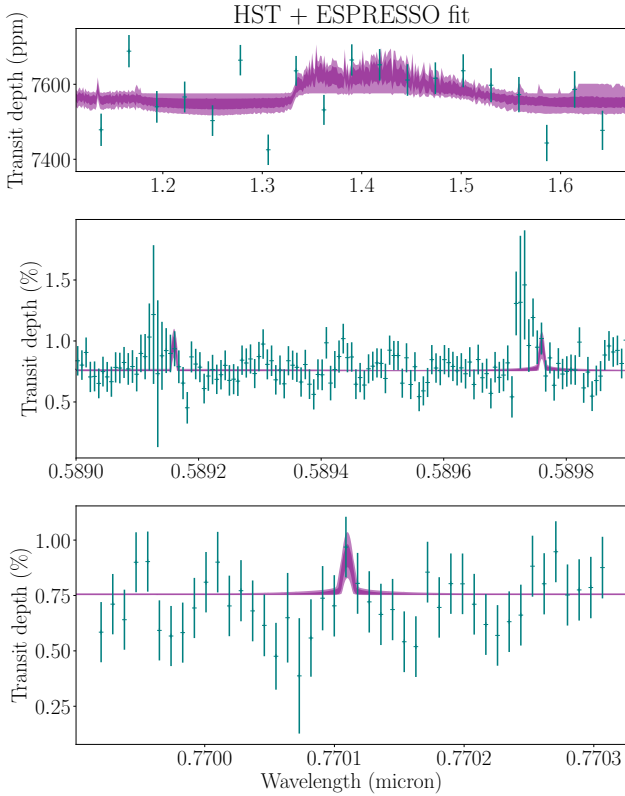
Fig. 7. Full retrieval model fit to the WASP-117b transmission spectrum in HST/WFC3 and VLT/ESPRESSO with the nominal pipeline with Na and K. The WFC3 data are shown in the *top panel*, the *middle panel* shows the Na doublet, and the *bottom panel* the K D_1 line. The retrieval model without Na and K (not shown) is a flat line in the optical and is indistinguishable from the full model without Na and K in the WFC3 near-infrared range. For all spectra, retrieved atmosphere models with confidence within 1σ (dark green) and 2σ (light green) are shown.

between the day- and nightside, probed across the terminator (Caldas et al. 2019; Pluriel et al. 2020). The number of free parameters, that is, the complexity of the retrieval model, needs to be justified by the quality of the data, however. Too complex models should be avoided for observations of low signal-to-noise ratio, and in general, there should be fewer free parameters than data points. This prevents overfitting the data. A useful criterion to judge whether one model is better than another, given the data, is the Bayes factor (Kass & Raftery 1995). The Bayes factor analysis penalizes those models that are too complex, given the data. However, a Bayes factor analysis should not be applied blindly. A simple (a few-parameter) model that gives a good fit to the data is always favored, even when the model assumptions are highly unphysical. The caveat in using a Bayesian factor analysis is that Bayes factors do not know physics.

Because the quality of our data is limited, we decided to keep the complexity of the retrieval 1D forward model low, with a limited number of free parameters (see Table 2). More specifically, we assumed an isothermal temperature structure, whereas the absorber abundances were not kept as free parameters, but were modeled with a chemical equilibrium model, using the chemistry code described in the appendix of Mollière et al. (2017). Our abundance treatment for WASP-117b is thus analogous to that of Kreidberg et al. (2018) for WASP-107b. WASP-107b could be similar in temperature and atmosphere chemistry to WASP-117b during transit.

Table 2. Models used for the retrieval of the WASP-117b atmosphere.

Model name	Prior in parameters	Short description
Model 1	$T \in [500, 1500]$ K $\log(P_0/1\mathrm{bar}) \in [-12, +2]$ $C/O \in [0.1, 1.6]$ $[Fe/H] \in [-1.5, +3]$ $\log(P_{cloud}) \in [-6, +2]$	High [Fe/H] & gray cloud
Model 2	As in Model 1 except $[Fe/H] \in [-1.5, +1.75]$	Low [Fe/H] & gray cloud
Model 3	As in Model 1 except $[Fe/H] \in [-1.5, +1.75]$ $\gamma \in [-6, 0]$ $\log(\kappa_0) \in [-10, +10]$	Low [Fe/H] & gray cloud & haze layer

While `petitRADTRANS` offers a wide range of different cloud parameterizations, we decided to only retrieve a gray cloud deck pressure, and optionally, a scattering-slope opacity to model small-particle hazes. The haze opacity is parameterized as

$$\kappa_{haze} = \kappa_0 \left(\frac{\lambda}{\lambda_0} \right)^{\gamma} \qquad (2)$$

in the retrieval model, where γ determines the steepness of the scattering slope, which we constrain to $\gamma > -6$ to avoid unphysically steep slopes (Pinhas & Madhusudhan 2017; Barstow 2020). The parameters κ_0 and λ_0 are reference opacities and wavelengths, where the first is freely retrieved and the latter is set to $\lambda_0 = 0.35 \, \mu$m. The unit of κ_0 is cm^2 g^{-1} and the haze opacity is assumed to be vertically constant.

Assuming a 1D temperature structure for transmission spectroscopy during transit could lead to underestimating the atmospheric temperature at the limbs, as indicated by MacDonald et al. (2020). In addition, patchy clouds could mimic situations of high enrichment (Line & Parmentier 2016), which is another possible limitation to keep in mind.

3.3. HST/WFC3 atmosphere retrieval

In the following, we discuss constraints on the atmospheric composition of WASP-117b based on atmospheric retrieval of the HST/WFC3 transmission spectrum using the nominal (Appendix A.1) and the CASCADE pipeline (Appendix A.2). We employed several models with physically motivated priors to constrain the atmospheric properties of WASP-117b. The models and their priors are listed in Table 2. Model 1 and Model 2 represent atmospheric models with isothermal temperature, equilibrium chemistry, and a gray cloud deck. Model 1 sets weak constraints on metallicity within $[-1.5, +3]$, Model 2 sets constraints to lower metallicity $[Fe/H] < 1.75$. Model 3 assumes the same parameters and priors as Model 2 and additionally adopts a haze layer.

3.3.1. Model 1: condensate clouds and weakly constrained [Fe/H]

Figure 8 displays the 2–98 and 16–84 percentile envelopes of the retrieved transmission spectra of Model 1, together with the WASP-117b HST data, reduced with the nominal and CASCADE pipeline, respectively. The corner plots with the
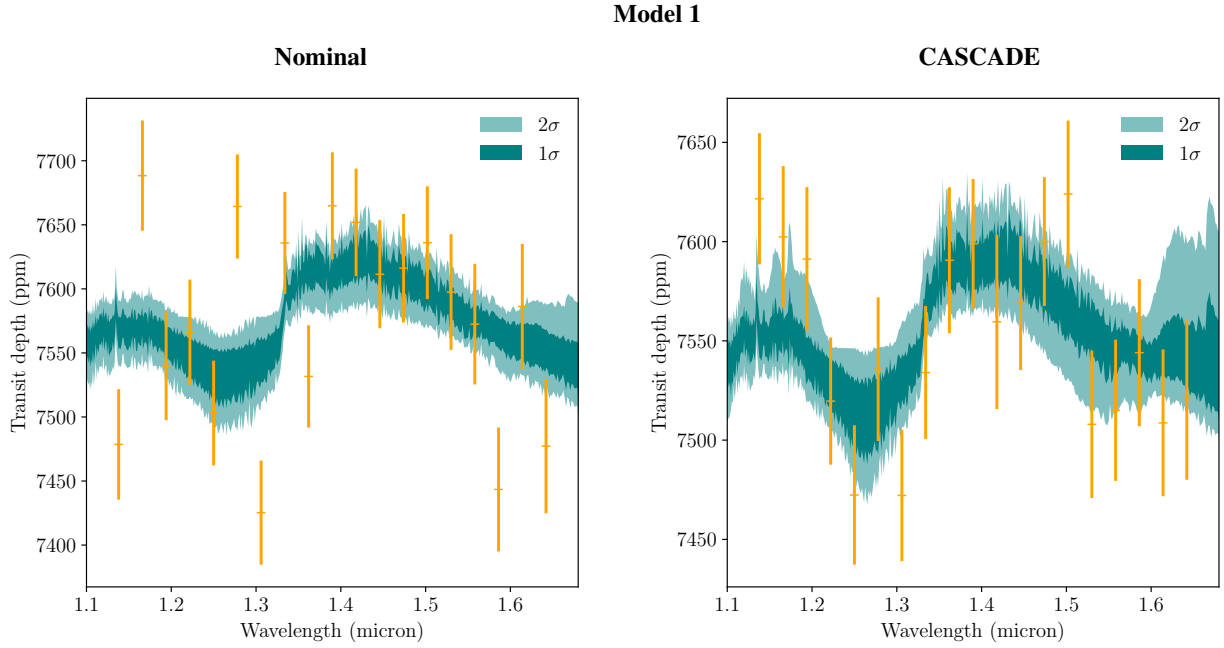
Model 1

Nominal **CASCADE**

Fig. 8. Results of atmospheric retrieval for WASP-117b with Model 1 (unconstrained [Fe/H] and gray cloud). *Left*: fit to the transmission spectrum reduced with the nominal data pipeline by Zhou et al. (2017). *Right*: fit to the transmission spectrum reduced with CASCADE. For both spectra, retrieved atmosphere models with confidence within 1σ (dark green) and 2σ (light green) are shown. These models were derived with `petitRADTRANS` (Mollière et al. 2019) and indicate a clear atmosphere of high metallicity, where the large mean molecular weight leads to a muted water spectrum.

Table 3. Median parameters and 68% confidence intervals for Models 1, 2, and 3 retrieved with HST/WFC3 spectra based on two different data-reduction pipelines.

Parameter	Model 1		Model 2		Model 3	
	Nominal	CASCADE	Nominal	CASCADE	Nominal	CASCADE
Temperature [K]	1129^{+228}_{-289}	821^{+264}_{-226}	1002^{+348}_{-350}	674^{+392}_{-135}	860^{+371}_{-257}	738^{+360}_{-157}
Pressure $\log(P_0/1\mathrm{bar})$	$-6.69^{+0.99}_{-1.21}$	$-7.20^{+1.11}_{-1.51}$	$-4.44^{+1.52}_{-0.97}$	$-4.45^{+1.47}_{-1.01}$	$-4.49^{+1.33}_{-1.13}$	$-4.89^{+1.23}_{-1.09}$
[Fe/H]	$2.58^{+0.26}_{-0.37}$	$2.55^{+0.28}_{-0.35}$	$0.66^{+0.85}_{-1.64}$	$0.54^{+0.97}_{-1.62}$	$0.41^{+0.97}_{-1.26}$	$0.67^{+0.81}_{-1.33}$
C/O	$0.25^{+0.17}_{-0.10}$	$0.27^{+0.24}_{-0.11}$	$0.37^{+0.30}_{-0.19}$	$0.36^{+0.41}_{-0.19}$	$0.40^{+0.48}_{-0.20}$	$0.38^{+0.40}_{-0.19}$
Cloud top $\log(P_{cloud}/1\mathrm{bar})$	$-1.32^{+2.24}_{-2.44}$	$-0.78^{+1.77}_{-1.77}$	$-3.69^{+1.42}_{-0.97}$	$-3.30^{+1.29}_{-1.03}$	$-3.69^{+1.27}_{-1.15}$	$-2.97^{+2.32}_{-1.49}$
$\log(\kappa_0)$	–	–	–	–	$-4.91^{+3.61}_{-3.10}$	$-1.99^{+2.84}_{-5.22}$
Scattering slope γ	–	–	–	–	$-3.55^{+1.89}_{-1.57}$	$-2.61^{+1.59}_{-2.09}$
Volume fractions within 1σ						
$\log(H_2O)$ at 10^{-4} bar	$-0.62^{+0.26}_{-0.51}$	$-0.75^{+0.34}_{-0.66}$	$-2.70^{+1.15}_{-1.63}$	$-2.68^{+1.17}_{-1.77}$	$-2.96^{+1.21}_{-1.50}$	$2.66^{+1.07}_{-1.66}$
$\log(CH_4)$ at 10^{-4} bar	$-13.19^{+3.87}_{-6.66}$	$-9.34^{+5.38}_{-5.92}$	$-9.89^{+5.39}_{-2.88}$	$-4.75^{+1.58}_{-5.43}$	$-7.55^{+3.63}_{-4.41}$	$-5.28^{+2.07}_{-5.05}$
3σ upper limits of CH_4 volume fractions						
CH_4 at 10^{-4} bar	$8.6 \cdot 10^{-2}$	$2.0 \cdot 10^{-1}$	$1.6 \cdot 10^{-2}$	$1.9 \cdot 10^{-2}$	$1.6 \cdot 10^{-2}$	$1.7 \cdot 10^{-2}$

retrieved properties based on Model 1 are displayed in Figs. B.9, B.10, and B.15, and the results are listed in Table 3. The retrieval based on the nominal spectrum yields a relatively high temperature of 1129^{+228}_{-289} K, whereas retrieval based on the CASCADE spectrum would indicate a rather low temperature of

821^{+264}_{-226}. The discrepancy in temperature can be explained by data-reduction differences that in the nominal spectrum cause an overall deeper transit depth, that is, a more inflated atmosphere, which can be achieved by attaining higher temperatures compared to the CASCADE spectrum. However, both temperatures
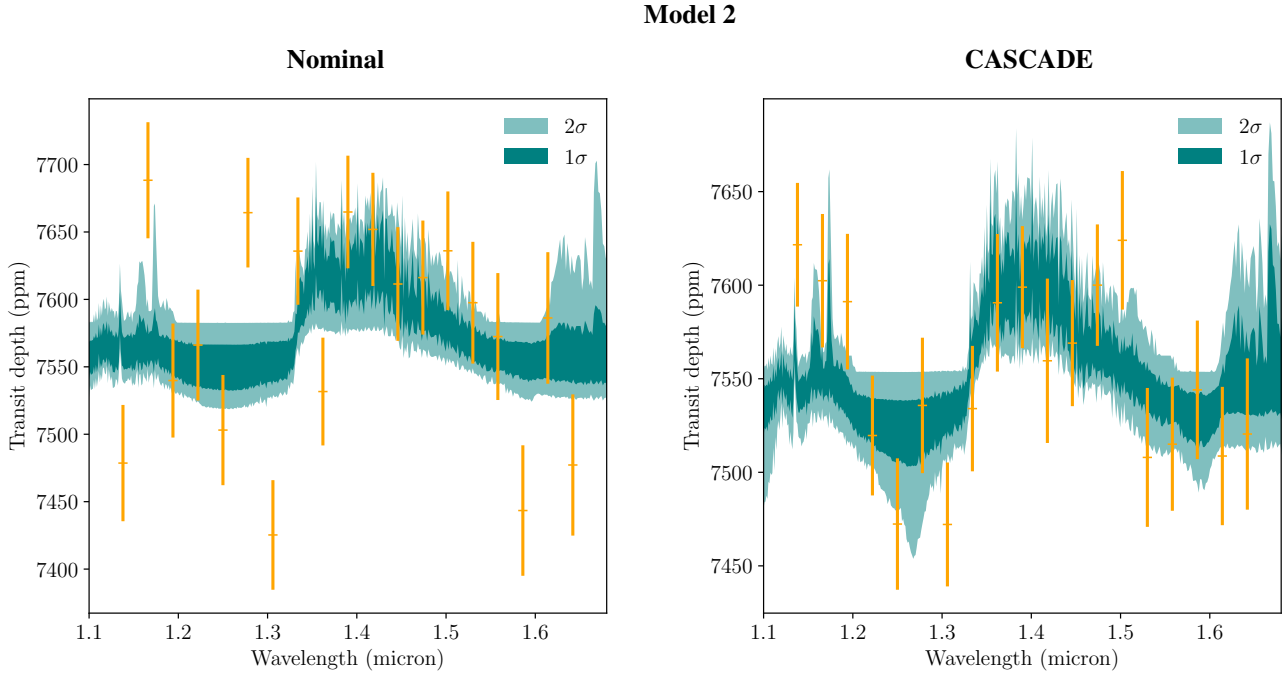
Model 2

Nominal **CASCADE**

Fig. 9. Results of atmospheric retrieval for WASP-117b with Model 2 (prior [Fe/H] < 1.75 and gray cloud). *Left*: fit to the transmission spectrum reduced with the nominal data pipeline by Zhou et al. (2017). *Right*: fit to the transmission spectrum reduced with the data pipeline CASCADE. For both spectra, retrieved atmosphere models with confidence within 1σ (dark green) and 2σ (light green) are shown. These models were derived with petitRADTRANS (Mollière et al. 2019) and indicate a cloudy atmosphere with a muted water feature.

are within one sigma of each other, which again indicates that this shift is not substantial. We therefore conclude that the atmospheric temperature is not well constrained with our data, at least with Model 1.

Retrieval with Model 1 further suggests very high metallicities [Fe/H] > 2.2 based on both spectra. At the same time, a cloud top located relatively deep in the atmosphere ($p > 10^{-4}$ bar) is inferred, and the reference pressure P_0 is set high in the atmosphere ($P_0 < 10^{-5.5}$) to reconcile the relatively high transit depth of this inflated exo-Saturn with the high metallicities. The high-metallicity solutions thus indicate relatively clear skies with very small atmospheric scale heights. These results indicate that the water feature is strongly muted and that condensate cloud modeling alone cannot properly account for the shape of the water signal. Instead, the model assumes a high mean molecular weight with more than 100× solar metallicity, regardless of which data reduction pipeline is used.

Furthermore, retrieval based on both the nominal and CASCADE spectra tend to yield subsolar C/O values with an upper limit of 0.42 and 0.51 respectively. Because of this, the CH_4 content is retrieved within 1σ to be below 1.1×10^{-4} volume fraction at $p = 10^{-4}$ bar when data retrieved with the CASCADE spectrum are used and even below 5.7×10^{-10} volume fraction when the nominal pipeline is used. However, we note that the 3σ range is very wide (see Fig. B.15), so that we cannot claim strong significance for these low CH_4 abundances.

Similar low C/O values were also found for WASP-107b by Kreidberg et al. (2018). These authors likewise attributed the low C/O values to the absence of CH_4 within the WFC3/G141 wavelength range, which is consistent with our retrieved low abundances of CH_4. If WASP-117b is colder than 800 K during transit and thus in a similar temperature range to WASP-107b, then the very low abundance of CH_4 could indicate disequilibrium chemistry. We explore this possibility in Sect. 4.

We furthermore note that we find a tail of solutions with lower metallicity and high cloudiness in the retrieved parameters based on a Model 1 fit to both spectra, the nominal and the CASCADE spectrum (see Figs. B.9 and B.10 in the [Fe/H] − P_{cloud} corners, indicated with a black ellipse). We therefore explore lower metallicity solutions more carefully in the next sections.

3.3.2. Model 2: condensate clouds and [Fe/H] < 1.75

We designed Model 2 with a constraint on allowed metallicities as agnostically as possible to select the tail of cloudy low-metallicity solutions of Model 1. We again adapted the simplest (uniform) cloud model for retrieval because the information content in the muted observed water spectrum observed with HST/WFC3 is too limited to warrant a more complex cloud model with patchy clouds, for example, as proposed by Line & Parmentier (2016).

We set an upper limit of 56× solar metallicity, that is, we imposed a prior for [Fe/H] < 1.75. Figure 9 displays the best fit of Model 2 to the WASP-117b HST data, reduced with the nominal and CASCADE pipelines. The detailed corner plots with the retrieved properties based on Model 2 are displayed in Figs. B.11, B.12, and B.16. Table 3 concisely lists the results of the retrieval.

We note that the retrieved atmospheric metallicity for WASP-117b is not well constrained for Model 2, with a tendency to still favor solutions at the higher end of the imposed prior. On the other hand, regardless of which data-reduction pipeline is used, nominal or CASCADE, in both cases we retrieve solutions with a better constrained cloud deck located higher in the atmospheric than in Model 1. With Model 2, the cloud deck is retrieved to lie between $p = 10^{-5.1}$ and $p = 10^{-2.2}$ bar. Because we excluded higher metallicities that reduce the scale height to very low values, the reference pressure P_0 is deeper in the atmosphere than
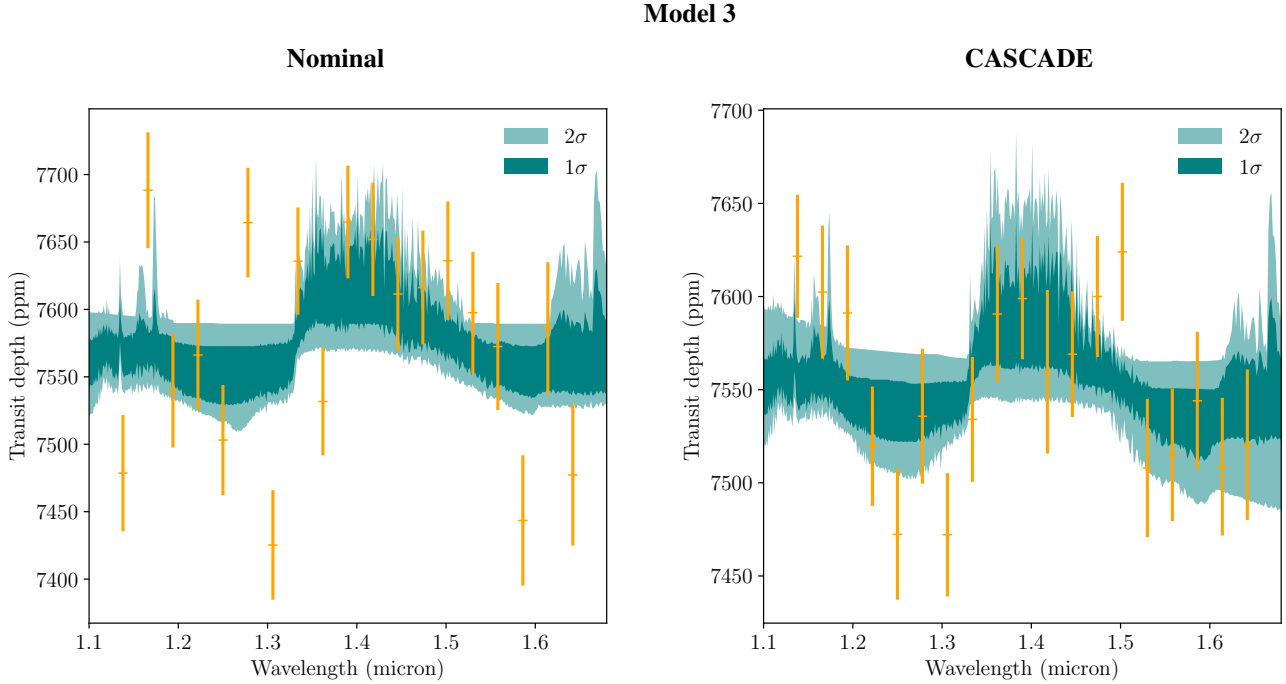
Model 3

Nominal   **CASCADE**

Fig. 10. Results of atmospheric retrieval for WASP-117b with Model 3 (gray cloud and haze layer with prior [Fe/H] < 1.75 and γ > −6). *Left*: fit to the transmission spectrum reduced with the nominal data pipeline by Zhou et al. (2017). *Right*: fit to the transmission spectrum reduced with CASCADE. For both spectra, retrieved atmosphere models with confidence within 1σ (dark green) and 2σ (light green) are shown. These models were derived with `petitRADTRANS` (Mollière et al. 2019) and indicate a cloudy atmosphere with a muted water feature.

in Model 1. In other words, Model 2 favors cloudy solutions to explain the muted water feature.

Subsolar C/O values that indicate low abundances of CH_4 are also favored by Model 2. However, we report a division between possible solutions within our constrained metallicity range. There is a tail of high C/O > 1 solutions (Figs. B.11 and B.12), which are correlated in the posterior of our retrieval with low metallicities ([Fe/H] < −0.5).

For the C/O > 1 solutions, we furthermore find that the CH_4 volume fraction at $p = 10^{-4}$ bar is greater than that of H_2O. For C/O < 1 solutions, the CH_4 abundances are always lower than the H_2O abundances. CH_4 is expected to dominate H_2O for high C/O ratios (Madhusudhan 2012; Mollière et al. 2015). However, we stress that these high C/O solutions only represent a small subset of solutions, associated with low metallicities [Fe/H] < −0.5, and they only occur when we impose a metallicity constraint of [Fe/H] < 1.75 on retrieval. The majority of the solutions are high-metallicity solutions ([Fe/H] > 1)) that still strongly favor solar to subsolar C/O ratios in Model 2 as well.

The low C/O ratios again appear to correspond to low CH_4 abundances (in any case, lower than the H_2O abundances), where we retrieved within 1σ CH_4 abundances below 6.7×10^{-4} volume fraction at 10^{-4} bar. For Model 2, the 3σ envelope of CH_4 abundances is also very large (Fig. B.16).

3.3.3. Model 3: condensate clouds, prior [Fe/H] < 1.75, and haze layer

Another possibility, which may explain the muted observed water spectrum, might be a haze layer on top of the cloudy atmosphere. This possibility was explored with Model 3. Figure 10 displays the best fit of Model 3 to the WASP-117b HST data, reduced with the nominal and CASCADE pipelines. The detailed corner plots with the retrieved properties based on

Model 3 are displayed in Figs. B.13, B.14, and B.17. A summary of the parameters retrieved with this model is given in Table 3.

For the CASCADE spectrum, Model 3 yields a bimodal posterior distribution for the haze opacity. The median value is $\log(\kappa_0) = -1.99^{+2.84}_{-5.22}$, with the high-$\kappa_0$ peak occurring at $\log(\kappa_0) = 0.15$ and the low peak occurring at $\log(\kappa_0) = -5.8$. The low peak is consistent with the posterior of Model 3 for the nominal spectrum, while the high peak indicates that some of the models are hazy with a moderate scattering slope of $\gamma = -2.61^{+1.59}_{-2.09}$ (Fig. 10, right). High values of κ_0 also correspond to a gray cloud deck that can lie deeper in the atmosphere than in Model 2. The water feature for some models inside the Model 3 framework is muted by the combined effect of condensed clouds and the haze layer.

For the nominal spectrum, however, Model 3 is clearly a poor fit (Fig. 10, left). Comparison of the corner plots Figs. B.13 and B.14 shows that the retrieved haze layer properties κ_0 and γ in the nominal spectrum are more uniformly spread across the allowed parameter range compared to the CASCADE spectrum. Generally, the retrieved parameters based on the nominal spectrum for Model 3 are well within 1σ of the parameters retrieved with the CASCADE spectrum as a basis, however. The subset of hazy solutions identified in the posterior of retrieval with Model 3 based on the CASCADE spectrum is therefore not substantial enough to indicate a strong disagreement of the data-reduction pipelines that are used in this work.

Like in Model 2, the metallicity is unconstrained in Model 3 within the imposed low-metallicity limits of [Fe/H] = [−1.5, 1.75], regardless of which pipeline is used. The temperature is again not well constrained in Model 3, as can be seen by comparing Figs. B.14 and B.12.

In Model 3, we again identify a tail of high C/O solutions as in Model 2. Most solutions, especially those with [Fe/H] > 1, however, favor subsolar C/O. As noted in previous subsections,

Table 4. Statistical analysis of WASP-117b HST data for different models.

Model	ln Z	Bayes factor B	χ^2
		Nominal spectrum	
Model 1	−31.59	Baseline	47.9
Model 2	−33.05	4.3	
Model 3	−34.5	18.4	
		CASCADE spectrum	
Model 1	−17.83	Baseline	19.1
Model 2	−20.46	13.9	
Model 3	−21.28	31.5	

Fig. 11. ESPRESSO RVs obtained on 25 October 2018 during a transit of WASP-117b. The barycentric Earth radial velocity and the stellar orbital velocity are corrected. The orange line shows the best-fit model of the Rossiter-McLaughlin effect.

low C/O values are indicative of low abundances of methane. This is again reflected by retrieved CH_4 abundances that are lower than 10^{-4} volume fraction at $p = 10^{-4}$ bar within 1σ for the nominal reduction pipeline. We note that fitting Model 3 to the CASCADE spectrum allows even a volume fraction up to $10^{-3.2}$ at $p = 10^{-4}$ bar within 1σ. The significance of the molecular absorption feature as well as that of a haze layer is investigated in the following.

3.4. Significance of differences in metallicity and haze layer

To quantify the different models that we used for retrieval, we now compare Models 1, 2, and 3 with each other, using the same method as in Sect. 3.1. In Table 4 we again list the (natural) log evidence (lnZ) and the Bayes factor B.

According to this statistical analysis, Model 1, a model with high atmospheric metallicity ([Fe/H] $\geq$ 2.2) and clear skies, fits the data best compared to the lower metallicity, cloudy models Model 2 and Model 3. Model 3 also assumes a haze layer on top. This is true for both pipelines.

The Bayesian preference, however, is higher for the CASCADE spectrum than for the nominal spectrum. Based on data reduced with the CASCADE pipeline, the preference is "substantial" to "strong". Based on data reduced with the nominal spectrum, the preference is still "substantial", that is, $1 - 2\sigma$.

Theoretical predictions of Thorngren & Fortney (2019) indicated that up to 200× solar metallicity is possible in principle for Saturn-mass exoplanets such as WASP-117b. We again point out the possible degeneracy of high metallicity with patchy cloud solutions (Line & Parmentier 2016).

We also added a χ^2 analysis for Model 1 to yet again quantify the agreement of the data derived from different pipelines in comparison with the best-fit physical model (Table 4). The χ^2 deviation as well as the scatter in data points below 1.35 μm in Fig. 1 indicates some problem with the nominal pipeline. We demonstrated in Sect. 2.1.2 that the discrepancy below 1.35 μm is likely caused by the difference in calibrating the telescope pointing movement in the dispersion direction. Therefore we performed retrieval analyses on spectra obtained using both pipelines. Because e we found fully consistent results, this confirms the high fidelity of our conclusions.

Our work illustrates that even with the well-established nominal pipeline, we have to be aware of instrumental effects that may yield a poorer performance than expected. It further highlights the importance of using independent data-reduction pipelines to confirm the analysis results. Still, better and
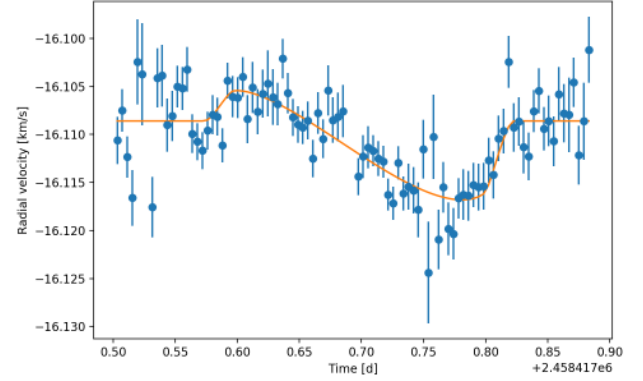
additional data in the optical range, for example, are needed for further constraints of cloud and haze coverage of WASP-117b and thus atmospheric metallicity. We have obtained additional observations of WASP-117b with VLT/ESPRESSO (Sect. 2.2) to support the HST/WFC3 observation. In addition, we can analyze WASP-117b transit data obtained by the TESS satellite (Sect. 3.6).

3.5. Rossiter-McLaughlin effect measured with VLT/ESPRESSO

Based on our high-resolution ESPRESSO data we were also able improve the constraints on the stellar rotation and spin-orbit alignment of WASP-117b through the Rossiter-McLaughlin (RM) effect. The RM of the system was initially measured by Lendl et al. (2014) using one transit dataset from the HARPS observation. We fit the RM effect using our ESPRESSO data to update the obliquity parameters. We applied a Markov chain Monte Carlo (MCMC) approach using the `emcee` tool (Foreman-Mackey et al. 2013). The RM effect was modeled with the python code `RmcLell` from the PyAstronomy library[1] (Czesla et al. 2019). The projected stellar rotation velocity ($v \sin i_\star$), projected spin-orbit angle (λ), and systemic velocity (V_{sys}) were set as free parameters, and we fixed other parameters to the values in Lendl et al. (2014). The RV curve together with the best-fit model are presented in Fig. 11. The retrieved parameters are listed in Table 5.

We did not use these updated parameters in the analyses for the rest of this work, for which the original data by Lendl et al. (2014) were already sufficient to reduce the HST/WFC3 and TESS data before the updated VLT/EPRESSO data were available to us. Still, we present the updated RM values here for completeness and for inclusion in future work.

3.6. TESS transit data

We furthermore found that TESS has acquired four observations of WASP-117 between 26 August 2018 and 13 October 2018 during Sectors 2 and 3 of the TESS primary mission, covering four consecutive transits of WASP-117b[2]. These transit data are more accurate than the transit depth measured from the ground,

---

[1] https://github.com/sczesla/PyAstronomy
[2] https://exofop.ipac.caltech.edu/tess/target.php?id=166739520

Table 5. Retrieved parameters from the Rossiter-McLaughlin effect.

This work	
$v \sin i_\star$	1.46 ± 0.14 km s^{-1}
Projected spin-orbit angle λ	$-46.9^{+5.5}_{-4.8}$ deg
Systemic velocity V_{sys}	-16.1086 ± 0.0004 km s^{-1}
Lendl et al. (2014)	
$v \sin i_\star$	$1.67^{+0.31}_{-0.24}$ km s^{-1}
Projected spin-orbit angle λ	-44 ± 11 deg

Table 6. WASP-117b transit depths.

Survey	Transit depth	Wavelength range
WASP South	8030^{+550}_{-480} ppm	0.4–0.7
TESS (ExoFOP)	7200 ± 150 ppm	0.6–1 μm
TESS (JULIET)	7267 ± 135 ppm	0.6–1 μm
WFC3 (nominal)	7573 ± 43 ppm	1.125–1.65 μm

as reported in the discovery paper (Lendl et al. 2014) from the WASP South survey, and they might place further constraints on the cloud properties of WASP-117b.

Interestingly, the reported TESS transit depth of 7200 ± 150 ppm was found to be very shallow (albeit within 3σ) compared to the value reported by Lendl et al. (2014) (8030^{+550}_{-480} ppm) and by our HST/WFC3 nominal transit depth (7573 ± 43 ppm). We therefore decided to perform an independent analysis of the published TESS photometry and the target-pixel files (TPFs) of this target, to verify this slight discrepancy. For the former, we used the PDC light curve published at MAST[3], which we fit using `juliet` (Espinoza et al. 2019). As priors for this fit, we used the eccentricity and argument of periastron as reported by Lendl et al. (2014) as priors ($e = 0.302 \pm 0.023$, argument of periastron $\omega = 242 \pm 2.7$ deg). To account for the systematic trends in the data, we used a Gaussian process (GP) with a Matèrn 3/2 kernel, whose parameters had wide priors; the timescale of the process had a prior between 10^{-5} and 1000 days, whereas the square-root of the variance of the process had a prior between 10^{-3} and 100 ppm. Independent GPs were used for each sector, which also had an added white-noise component modeled as a added Gaussian noise to the GP with a large prior on the square-root of its variance between 0.01 and 1000 ppm, and independent out-of-transit flux offsets with Gaussian priors centered on 0, but with a standard deviation of 10^5 ppm. For the transit depth and impact parameter we used the uninformative sampling scheme proposed in Espinoza (2018), which samples the entire range of physically plausible values for these parameters. A wide log-uniform prior was defined for the stellar density between 100 and 10 000 kg m^{-3}, and the efficient sampling scheme of Kipping (2013) was used to model the limb-darkening effect through a quadratic law, where the coefficients were left as free parameters of the fit. Finally, relatively wide priors were defined for the ephemerides of the orbit: a normal distribution centered on 10 with a standard deviation of 0.1 days for the period, and a normal distribution centered around 2 458 357.6 with a standard deviation of 0.1 days for the time-of-transit center. No dilution contamination was applied in this fit because this is corrected for in the PDC photometry.

Our `juliet` analysis of the PDC lightcurve reveals a transit depth within 1σ with the ExoFOP reported transit depth. The combined transit depths from WASP, TESS and the ExoFOP TESS depth are listed in Table 6. Next, we proceeded to test whether the PDC algorithm might dilute the transit signal through too strong detrending, for instance. To this end, we used the TPFs reported in MAST to extract our own photometry for this target. We used apertures consisting of 1, 2, and 3 pixels around the brightest pixel, which accounts for apertures of about

30″, 40″, and 50″ around the target. To detrend the data, we used pixel-level decorrelation (PLD, Deming et al. 2015), which was initially used for *Spitzer* but which has been successfully applied to photometric data from missions such as K2 (Luger et al. 2016). The detrending was performed simultaneously with the transit model defined above in order to simultaneously account for all the uncertainties in each aperture. The results from this fit were in excellent agreement with the transit depths obtained from the PDC algorithm; for example, for our smallest aperture, we obtained a transit depth of 7144 ± 177 ppm. Because no strong dilution is expected in this smaller aperture (judging from *Gaia* sources around WASP-117), this analysis gives us confidence that the relatively shallower TESS transit depths are in fact real and not related to dilution or detrending methods. This result implies that the TESS transit depth is shallow as a result of a physical mechanism: either spectral contamination due to stellar activity of the host star WASP-117 or due to clouds and hazes in the atmosphere of WASP-117b.

3.7. Stellar activity and clouds as a possible explanation for the TESS data discrepancy

Stellar activity can add systematic offsets to the measured transit depth of an exoplanet such as WASP-117b, which can make it difficult to compare observations at different wavelength ranges and taken at different epochs (Rackham et al. 2019). In the WASP-117b discovery paper, Lendl et al. (2014) did not detect any variability greater than 1.3 mmag with 95% confidence. We independently investigated the variability of WASP-117 using ASAS-SN Sky Patrol[4] (Shappee et al. 2014; Kochanek et al. 2017), which includes *V*-band photometric monitoring of WASP-117 for the 2014–2017 observing seasons taken with two cameras, 'bf' and 'bh'. We considered only the data for the 2014 and 2015 observing seasons using the bf camera. Data collected using the bh camera, including some of the 2015 season and all later seasons, appear to have a strong instrumental systematic. The RMS scatter of the data included in this analysis is 11.5 mmag, while the mean measurement uncertainty is 4.5 mmag, suggesting some variability. A Lomb-Scargle periodigram analysis (Lomb 1976; Scargle 1982), implemented in `astropy`[5], consistently reveals a periodicity with a full amplitude of 9.4 mmag, although the period with the largest power depends on the included observing season. Considering the ASAS-SN data, we conservatively adopted 9.4 mmag or 1% as the reference variability level for WASP-117 in this analysis.

We modeled the variability of WASP-117 using the method detailed by Rackham et al. (2018), including considerations for FGK dwarfs (Rackham et al. 2019). In brief, the approach involves adding spots and faculae to a large set of model

---

[3] https://archive.stsci.edu/

[4] https://asas-sn.osu.edu/
[5] http://www.astropy.org. Astropy is a community-developed core Python package for Astronomy (Astropy Collaboration 2013, 2018).

Table 7. Parameters used in the stellar heterogeneity analysis.

Parameter	Description	Value
T_{phot}	Photosphere temperature	5460 K
[Fe/H]	Metallicity	0.14
$\log g$	Surface gravity	4.37
T_{spot}	Spot temperature	4780 K
T_{fac}	Facula temperature	5600 K
R_{spot}	Spot radius	2°
α	Initial facula-to-spot areal ratio	10
A_{ref}	V-band variability amplitude	2%

Table 8. Stellar contamination factors in the TESS and WFC3/G141 bandpasses for the spots-only models.

	Band pass	Value
ϵ_{WFC3}	$1.1-1.7\,\mu m$	$1.014^{+0.031}_{-0.006}$
ϵ_{TESS}	$0.6-1.0\,\mu m$	$1.03^{+0.06}_{-0.01}$

Fig. 12. Modeled rotational variability amplitudes of WASP-117 in V band as a function of spot covering fraction. The median estimate (blue line) and 1σ confidence interval (shaded blue region) are shown.

photospheres to establish the probabilistic relation between active region coverage and rotational variability for a set of stellar parameters. The observed variability of a star is then used to estimate its active region coverage and the related stellar contamination of transmission spectra from the system. We set the photosphere temperature to the stellar effective temperature determined by Lendl et al. (2014). We furthermore adopted the surface gravity, stellar surface gravity, and metallicity from that work. We determined spot and facula temperatures from the scaling relations detailed by Rackham et al. (2019). We used the spot size and initial facula-to-spot areal ratio from the solar-like case outlined by Rackham et al. (2018). The method we used allows the facula-to-spot areal ratio to drift to lower values as a star becomes more spotted, as is observed on the Sun (Shapiro et al. 2014). See Rackham et al. (2018) for further details. Tables 7 and 8 summarize the adopted parameters.

Figure 12 illustrates the modeled variability of WASP-117 as a function of spot-covering fraction. We find that the adopted variability amplitude of WASP-117 corresponds to full-disk spot- and facula-covering fractions of $F_{spot} = 4^{+6}_{-2}\%$ and $F_{fac} = 29^{+19}_{-9}\%$, respectively, at 1σ confidence. If present outside of the transit chord, these active regions could alter the observed transit depths. Figure 13 illustrates the wavelength-dependent stellar contamination factor ϵ_λ, the ratio of the observed transit depth to the true transit depth (see Eq. (2) of Rackham et al. 2018) produced by these active region coverages over the HST/WFC3 G141 bandpass. We find the effect of unocculted spots dominates that of unocculted faculae, producing a net increase in transit depths ($\epsilon_\lambda > 1$). Over the complete G141 bandpass, we estimate that transit depths can be inflated by $1.4^{+3.1}_{-0.6}\%$ at 1σ confidence. This inflation is relative to a measurement in the same wavelength range that is completely unaffected by stellar activity. No strong spectral features are apparent in the contamination signal except for a slight increase toward shorter wavelengths: at 1.1 μm, transit depths may be increased by $1.5^{+3.8}_{-0.7}\%$, while at 1.7 μm the increase is limited to $1.1^{+2.3}_{-0.5}\%$. We note that while the scale of ϵ_λ is comparable to the precision of our HST/WFC3 G141 spectrum, the lack of strong spectral features in the contamination spectrum provides reassurance that the spectral features we see in the near-infrared spectrum do not result from stellar contamination.
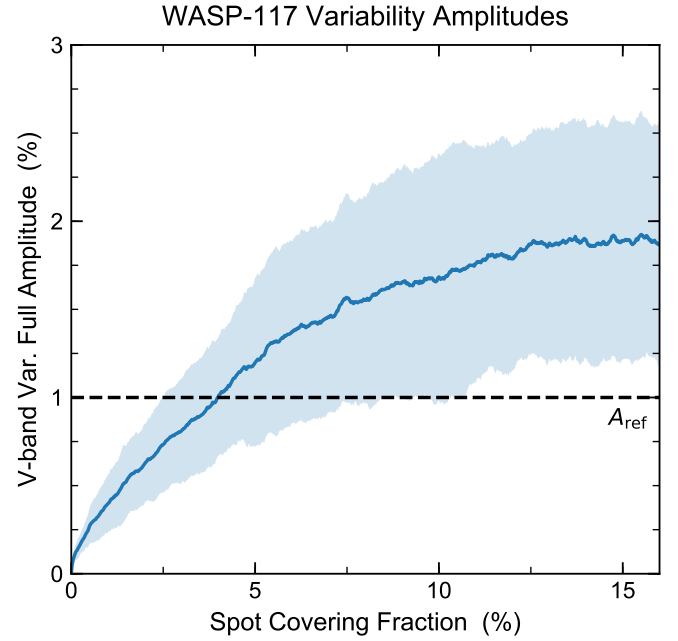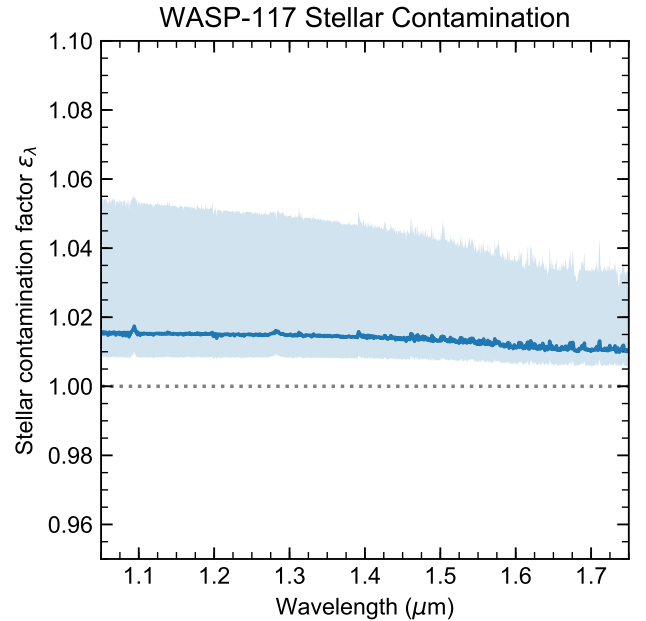

Fig. 13. Spectroscopic stellar contamination signal in the HST/WFC3 G141 bandpass produced by unocculted active regions on WASP-117. The median estimate (blue line) and 1σ confidence interval (shaded blue region) are shown.

Figure 14 shows the relative increase in planetary radius in the near-infrared (1.1μm) between 1 and 4%, the TESS transit depths, and HST/WFC3 spectrum. If the TESS data (observed in September 2018) had been taken at a time when the host star showed low activity, that is, very low coverage fraction of active regions compared to the time at which HST/WFC3 data were taken (observed in September 2019), then this difference in stellar activity between the two epochs might in principle explain the discrepancy between the optical and the near-infrared spectrum.
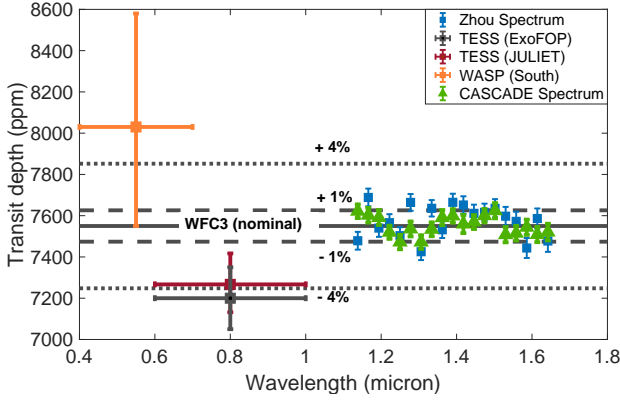
Fig. 14. HST transmission spectra from the nominal and CASCADE pipeline compared with TESS transit depth, reduced by ExoFOP, our independent reanalysis with JULIET, and the WASP South transit depth. The solid gray line denotes the nominal transit depth in the WFC3 wavelength range. The dashed and dotted lines show radius inflation or deflation by 1 and 4%, respectively.
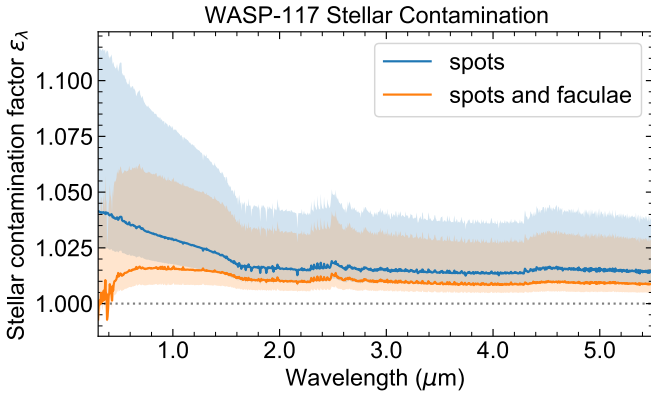

Fig. 15. Spectroscopic stellar contamination signal for 0.3–5.5 μm produced by unocculted active regions on WASP-117. The median estimates (blue line for spots and orange for spots and faculae,) as well as 1σ confidence intervals (shaded regions) are shown.

We note, however, that the raw HST/WFC3 data (Fig. A.2) do not display any obvious signs of stellar activity, for instance, in the form of spot-crossing events. Conversely, the VLT/ESPRESSO data (10/2019) taken at a few weeks after the TESS measurements do not show strong activity in the stellar Na and K lines. In other words, with the data at hand, there are no indications that the host star was more active in one epoch than in the other.

Alternatively, if the stellar activity between the two epochs did not differ drastically, unocculted bright regions (faculae) could be invoked to explain decreases in transit depths at visual wavelengths (i.e., $\epsilon < 1$). We consider this unlikely, however, because the TESS bandpass is redder than the wavelength where we would expect faculae to start dramatically decrease transit depths, which is typically around 0.5 μm (Rackham et al. 2019, Fig. 5). Furthermore, the scale of the decrease is too large between the TESS and WFC3 band. The transit depth is shallower by about 4% in the TESS band (Fig. 14). However, we expect a decrease smaller than 1% in transit depth due to the chromaticity of bright faculae between the TESS and WFC3 G141 bands (Fig. 15).

As we show in Sect. 5, no atmospheric model fit to the HST/WFC3 data yields transit depths shallower than 7400 ppm

in the optical wavelength range, including the very cloudy (Model 2) and hazy solutions (Model 3). Without additional measurements, it is therefore currently not possible to determine why the TESS transit depth is shallower (7200 ± 150) than the earlier WASP and the most recent HST/WFC3 measurements.

4. WASP-117b temperatures and chemistry composition

We were unable to use the TESS data to constrain atmospheric properties of WASP-117b. We furthermore were unable to constrain the atmospheric temperature of WASP-117b during transit based on the muted water feature observed with HST/WFC3 (Table 3).

However, it is still worthwhile to explore exemplary chemistry models for atmospheric compositions and temperature possible for WASP-117b based on the WFC3 data and theoretical reasoning. These physically based constraints (a) act as sanity checks for the simplified 1D retrieval, (b) still allow us to draw further conclusions, and (c) inform us about the benefit of future observations.

We calculated theoretical average planetary equilibrium temperatures $T_{\mathrm{Pl,inst}}$, assuming that the planet instantly adjusts to the stellar irradiation, that the planet is in equilibrium with the incoming stellar irradiation, and that the planet experiences global heat redistribution. This is the assumption that Lendl et al. (2014) used for their analysis, using albedo $\alpha = 0$. Here, we used Eq. (2) in Méndez & Rivera-Valentín (2017) with $\beta = 1$ (efficient heat redistribution over the whole planet) and broadband thermal emissivity $\epsilon = 1$, using stellar and planetary parameters from Lendl et al. (2014) and setting albedos $\alpha = 0, 0.3, 0.6$ within limits of cloud models for hot to warm Jupiters (Parmentier et al. 2016). $T_{\mathrm{Pl,inst}}$ gives the possible range of global average temperatures, assuming instant heat adjustment, global heat circulation, and different albedos for WASP-117b during one orbit, including at transit (Table 9, temperatures during transit set in bold). These temperatures were compared to the assumption that the planet temperature is constant and equal to the time-averaged planetary temperature $T_{\mathrm{Pl,ave}}$, where we used Eq. (16) of Méndez & Rivera-Valentín (2017) again with $\alpha = 0, 0.3, 0.6$. We stress that these temperatures are first-order assumptions to estimate the maximum possible temperature range during transit. A more comprehensive analysis of the temperature variation during one orbit requires the comparison of radiative and dynamical timescales, where the latter has to be estimated from a 3D circulation model (Lewis et al. 2014).

Table 9 shows that depending on albedo and heat adjustment assumptions, WASP-117b could assume globally averaged equilibrium temperatures T_{eq} between 700 and 1000 K during transit. In the following, we investigate the implications of equilibrium and disequilibrium chemistry within this temperature range for Model 1 (high metallicity [Fe/H] > 2.2) and Model 2 (low metallicity [Fe/H] < 1.75).

We computed a self-consistent pressure-temperature profile for the exo-Saturn WASP-117b using petitCODE (Mollière et al. 2015, 2017) and multiple 1D chemical kinetics models (Venot et al. 2012), incorporating vertical mixing and photochemistry. In order to systematically represent the best-fit retrieval models (see Table 3) as well as the parameter space of physically accepted solutions, we performed four base models with varying temperature and metallicity. We adopted values for atmospheric metallicities of 350 times and 5 times solar metallicity, consistent with Model 1 and Model 2, respectively. Furthermore, we picked

Table 9. Theoretical temperature constraints of WASP-117b during one orbit.

Time	Distance [a] a [AU]	$T_{Pl,inst}$			$T_{Pl,ave}$		
		$\alpha = 0$	$\alpha = 0.3$	$\alpha = 0.6$	$\alpha = 0$	$\alpha = 0.3$	$\alpha = 0.6$
Apastron	0.1230	899	822	715			
Transit	**0.1168**	**922**	**844**	**734**	**1019**	**932**	**810**
Periastron	0.0663	1224	1120	974			
Secondary eclipse	0.682	1207	1104	960			

Notes. [a]Based on Lendl et al. (2014).

equilibrium temperatures T_{eq} of 1000 and 700 K to represent the theoretically possible range of temperatures during transit. A C/O of 0.30 was adopted in accordance with the retrievals.

We acknowledge that a C/O of 0.3 may be too low because we assumed equilibrium chemistry during retrieval. We argue, however, that higher C/Os produce even higher CH_4 abundances for the same pressure and temperature range (see, e.g., Madhusudhan 2012). Thus, any constraints on CH_4 quenching by vertical mixing (K_{zz}) that we find for such a low C/O of 0.3 will also hold for higher C/Os and provide a lower limit for possible K_{zz}, for example. Therefore we decided given the lack of other constraints that adopting C/O = 0.3 is the best approach for the time being.

Another possible concern in basing disequilibrium chemistry on retrieved parameters is the difference in planetary atmospheric temperature structure used by both models. The former uses a self-consistent 1D atmospheric temperature profile (Fig. C.1, bottom right panel), based on petitCODE (Mollière et al. 2015, 2017), the retrieval code petitRADTRANS (Mollière et al. 2019) uses isothermal temperatures. However, we note that for $p \leq 10^{-3}$ bar the temperature is also approximately isothermal in the self-consistent 1D exoplanet atmosphere model. We therefore conclude that these differences in temperature do not play a strong role here.

Furthermore, for the pressure-temperature iteration we assumed a uniform dayside heat redistribution and a moderate intrinsic temperature of 400 K, in agreement with the planetary irradiation (Thorngren et al. 2019). Other stellar and planetary parameters were derived from Lendl et al. (2014). The pressure-temperature profiles computed with petitCODE, which were used in the chemical kinetics calculation, are shown in Fig. 16. For the chemical kinetics model we applied a vertically constant eddy diffusion coefficient $K_{zz} = 10^{10}$ cm^2 s^{-1}. We also applied photochemical reactions. To represent the spectral energy distribution of WASP-117, we used an ATLAS stellar model ($T = 6000$ K, $\log g = 4.5$) (Castelli & Kurucz 2003) between 168 nm and 80 μm, and a time-averaged solar UV spectrum (Thuillier et al. 2004) between 1 nm and 168 nm.

Figure 17 shows that the hot models with $T_{eq} = 1000$ K naturally have a low methane abundance in chemical equilibrium. This is exacerbated when disequilibrium chemistry is taken into account, as methane is quenched at even lower abundances. For the warm models with $T_{eq} = 700$ K, methane starts to dominate CO in chemical equilibrium in the higher atmospheric layers ($p < 1$ mbar for the high-metallicity case and $p < 40$ mbar in the low-metallicity case). In these models, vertical mixing also serves to reduce the methane abundance in the layers above the quenching level ($p < 100$ mbar for the high-metallicity case and $p < 1$ mbar in the low-metallicity case).

For low atmospheric temperatures ($T_{eq} = 700$ K), disequilibrium chemistry through vertical mixing is required to keep a

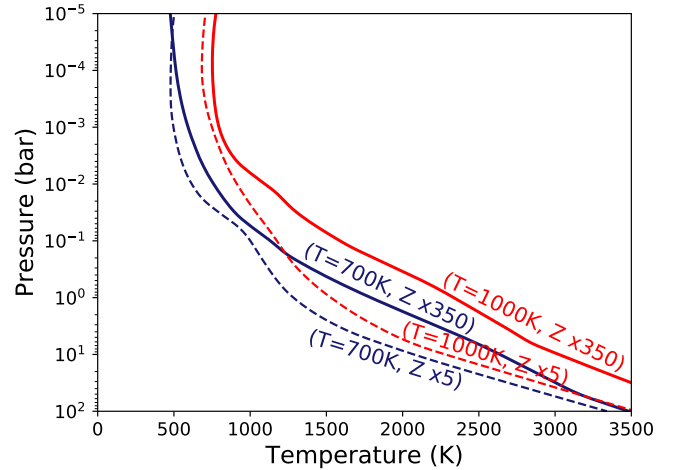**Fig. 16.** Pressure-temperature profiles computed with petitCODE for four different scenarios of WASP-117b: temperatures $T_{eff} = 1000$ K (red), $T_{eff} = 700$ K (blue), and metallicities $Z = Z_\odot \times 350$ (solid), $Z = Z_\odot \times 5$ (dashed). For all cases, a C/O of 0.3 and intrinsic temperature of 400 K were adopted. The profiles serve as input for the chemical kinetics calculations of Fig. 17.

low methane abundance ($<10^{-4}$ volume fraction) in the observable part of the atmosphere ($p < 0.1$ bar in transmission) for low and high atmospheric metallicity. Vertical mixing would be even stronger for lower atmospheric metallicity composition compared to high metallicity because in the former case, more CH_4 is produced at a given pressure level. The reason for this is twofold. First, for a given equilibrium temperature the lower metallicity case results in lower atmospheric temperatures (Fig. 16), leading to a higher methane abundance. Second, the lower metallicity increases the temperature of transition between CO / CH_4 (Lodders & Fegley 2002), again resulting in a higher methane abundance. Provided WASP-117b indeed has an equilibrium temperature of 700 K during transit, we might further constrain $K_{zz} \geq 10^8$ cm^2 s^{-1} (see Fig. C.1 and Appendix C, where we present an abundance study for this case as a function of K_{zz}). We note that this assessment is based on the median abundances that we retrieve for CH_4, with 1σ uncertainties on the upper limit of the CH_4 abundances as large as two orders of magnitude, see Figs. B.15–B.17.

A first disequilibrium versus equilibrium model comparison in the 1–5 μm range shows that JWST data can in principle distinguish between these two scenarios (Fig. 18): the absence of strong CH_4 bands, but also the presence of CO in the near-infrared, indicate disequilibrium chemistry. We investigate in the next section the prospect that JWST data might yield a tighter constraint on the CH_4 abundances, atmospheric temperatures, and disequilibrium chemistry in WASP-117b in more detail.
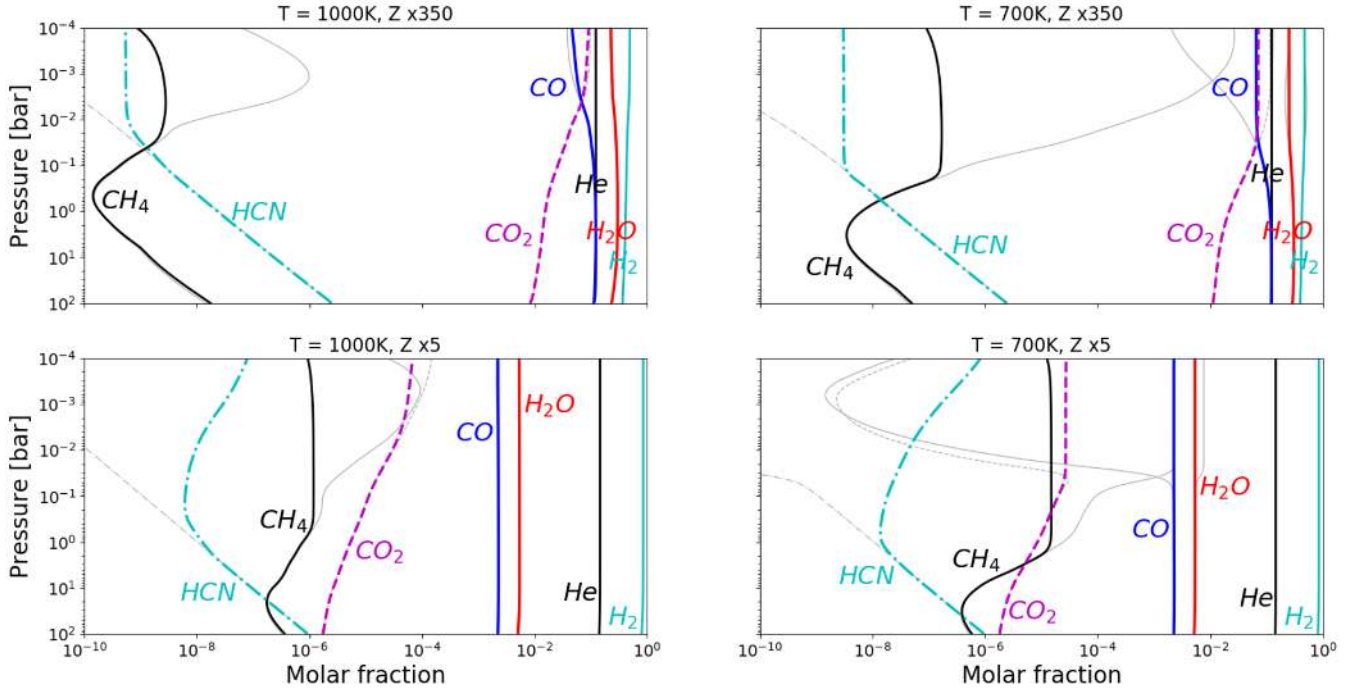
Fig. 17. Molecular abundances computed for hot (T_{eq} = 1000 K) or warm (T_{eq} = 700 K), and very ($Z_\odot \times 350$) or moderately ($Z_\odot \times 5$) enriched atmospheres, using the chemical kinetics code of Venot et al. (2012). The colored lines denote the disequilibrium abundances for the main molecules in the planetary atmosphere. The gray lines indicate the chemical equilibrium state. A constant $K_{zz} = 10^{10}$ cm^2 s^{-1} and C/O = 0.3 were adopted.
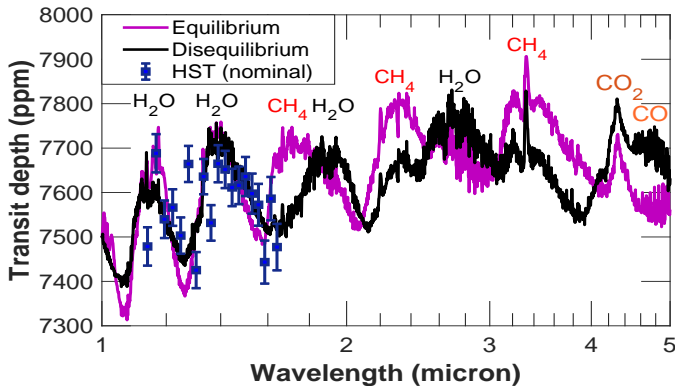

Fig. 18. Prediction of molecular features in the 1–5 μm wavelength range based on equilibrium (purple) and disequilibrium chemistry (black) cloud-free self-consistent forward models with methane quenching (black) for WASP-117b, assuming T_{eq} = 700 K and $K_{zz} = 10^8$ cm^{2}s and $Z = Z_0 \times 5$.

5. Prospects for future investigations with the HST and JWST

We investigated to what extent future observations with HST in the UV and optical, and JWST in the near- to mid-infrared, could constrain the properties of WASP-117b. Synthetic spectra based on the results of Model 1 (high metallicity), Model 2 (low metallicity and cloudy) and Model 3 (hazy) (see Sect. 3.3, Table 3) are shown in Fig. 19.

The possible atmospheric compositions retrieved for WASP-117b produce large differences in wavelength ranges between 0.3 and 0.5 μm (Fig. 19, top panels). An accuracy of 200 ppm could already be sufficient to distinguish between very hazy (extreme hazy solutions of Model 3, based on the CASCADE spectrum) and clear sky, heavy metal atmospheric composition (Model 1), based on the HST/WFC3 G141 grism spectrum reduced with the CASCADE pipeline (Fig. 19, top right). An accuracy of 100 ppm and better would be needed between 0.3 and 0.5 μm to distinguish between the medians of the clear sky, heavy metal, the gray cloudy solutions with lower metal content and the hazy solutions, regardless of the data pipeline (Fig. 19, top panels).

Wakeford et al. (2020) recently showed that it is in principle possible to accurately measure a transmission spectrum of a transiting exoplanet in this wavelength range using the WFC3/UVIS G280 grism. The authors reported an accuracy of 29–33 ppm for the broadband depth between 0.3 and 0.8 μm and 200 ppm in 10 nm spectroscopic bins. Wakeford et al. (2020) observed two transits of the hot Jupiter HAT-P-41b around a quiet F star with magnitude m_V = 11.08. Our target WASP-117b likewise orbits a quiet F star that is even brighter than the host star of HAT-P-42b. WASP-117 has a magnitude of m_V = 10.15.

We therefore estimate that two transit observations of WASP-117b might be sufficient to constrain a haze layer and potentially also the atmospheric metallicity using WFC3/UVIS. We did not consider here the optical transits in particular the TESS transit depth 7200 ± 150 ppm. All synthetic spectra have transit depths higher than 7400 ppm in the 0.8–1 μm range and do not fit the TESS data within 1σ (see Sect. 3.6).

In the JWST wavelength range with a single transit in each instrument, we expect clear H_2O features for the exo-Saturn WASP-117b in the JWST wavelength range (Fig. 19, bottom panels). Furthermore, some models with low-metallicity solutions (Model 2 and Model 3 framework) yield ~100 ppm strong CH_4 features at 3.4 μm. For example, the best-fit model for Model 2 based on the CASCADE spectrum predicts a clear CH_4 signal at 3.4 μm (Fig. 19, bottom left panel). This would correspond to a CH_4 abundance of 1.8×10^{-5} volume fraction (Table 3).
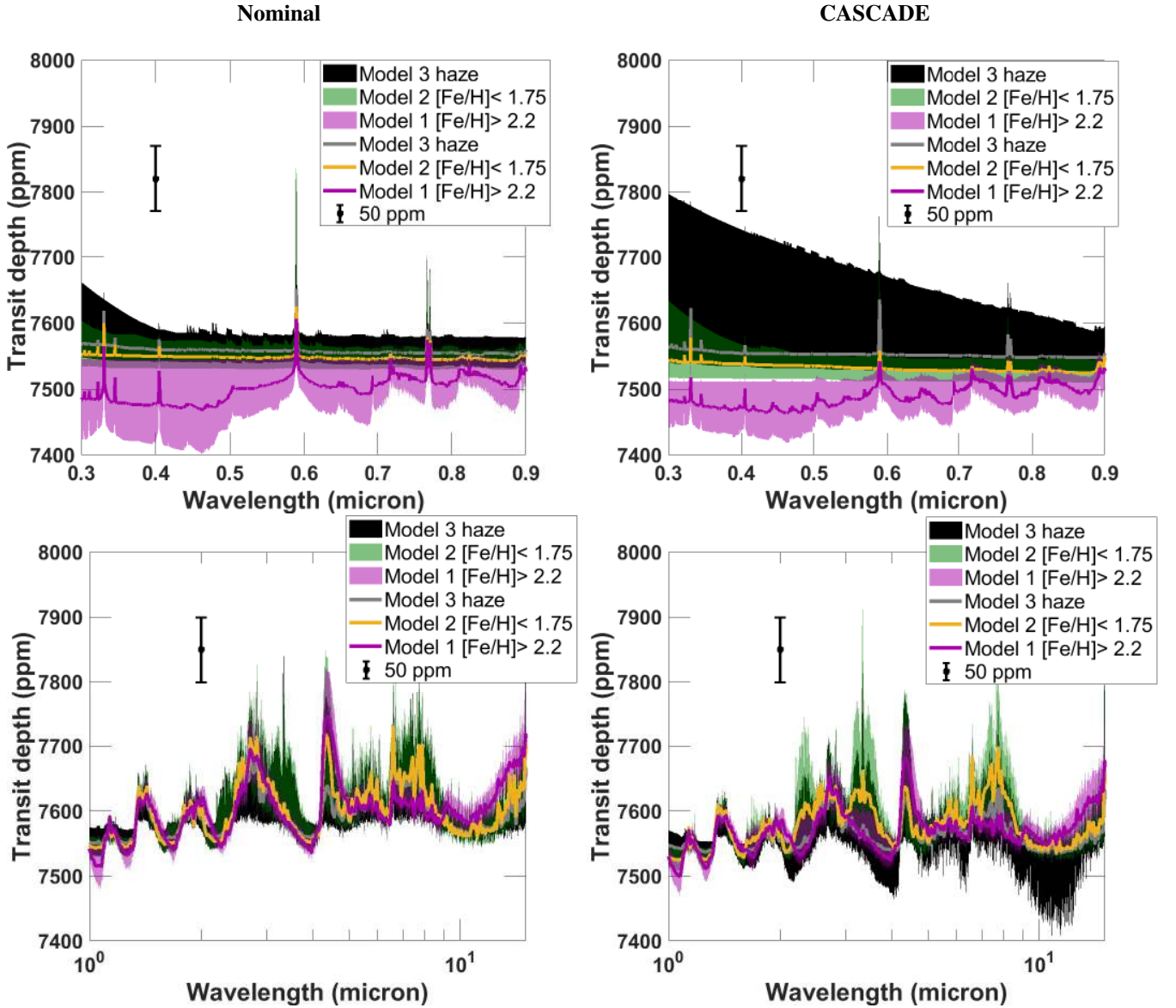
Nominal   **CASCADE**

Fig. 19. Synthetic spectra based on Model 1 (high metallicity, purple), Model 2 (low metallicity and cloudy, green) and Model 3 (haze, black) based on the WASP-117b spectrum reduced with the nominal pipeline (*left panels*), and the CASCADE pipeline (*right panels*). Colored areas show the 1σ envelope for the solutions for each model. Solid lines in gray, yellow, and light pink present the best fit for Model 1, Model 2, and Model 3, respectively. The spectra are shown for the wavelength range 0.3–0.9 μm (*upper panels*) and 1–15 μm (*lower panels*). In addition, a hypothetical data point with 50 ppm accuracy is shown for guidance.

Observing WASP-117b with JWST/NIRSpec should thus at the very least impose constraints on the CH_4 and H_2O content in the atmosphere. Observations of relatively high CH_4 abundances of 10^{-5} volume fraction could also disprove high metallicity through chemistry constraints, as shown in Sect. 4 (Fig. 17). Nonobservations of CH_4, on the other hand, might indicate either high-metallicity, high atmospheric equilibrium temperatures ($T_{eq} \gtrsim 1000$ K) during transit, or disequilibrium chemistry. In the latter case, we could also constrain vertical mixing (K_{zz}). It is also evident from Fig. 19 that CO_2 is a useful probe of metallicity, with the high-metallicity cases showing increased CO_2 absorption at 4.3 μm and at the red edge of the spectra, toward 15 μm.

Taking the expected instrument performance into account confirms our statement that NIRSpec should be capable, even with one single transit observation of WASP-117b, to resolve

differences in CO_2 at 4.3 μm (NIRSpec/G385M) that are due to different metallicity assumptions in atmospheric models (Fig. 20) as well as constrain CH_4 at 2.3 and 3.4 μm (NIRSpec/G235M). This is readily apparent in the right panel of Fig. 20 for CASCADE Model 2 (green) with enhanced CH_4 content compared to other retrieved models (see Table 3).

The comparison between spectra derived from the nominal and CASCADE pipelines (Fig. 20 left and right panel, respectively) also shows that the nominal H_2O features in the NIRSpec/G235M wavelength range are in the former case significantly enhanced compared to the latter, regardless of which underlying model is used. The differences in depth of the water absorption is mainly due to the differences in atmospheric temperatures assumed in the model medians. Models retrieved from the nominal HST/WFC3 spectrum are unconstrained in temperature within the prior range of 500–1500 K, so that
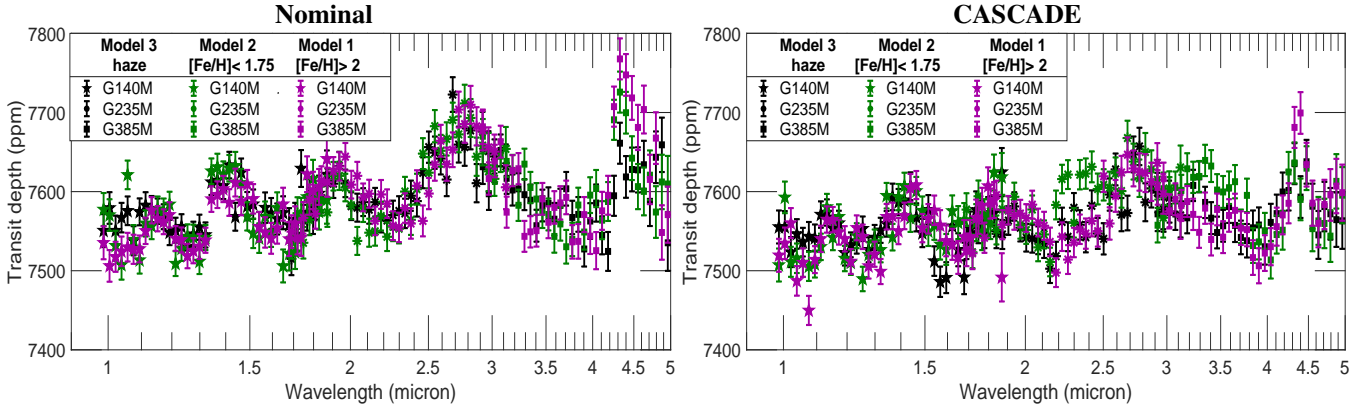
Fig. 20. Predicted JWST/NIRSpec observations in the 0.8–5 μm wavelength range based on the median of Model 1 (high metallicity, black), Model 2 (low metallicity and cloudy, green), and Model 3 (haze, purple) based on the WASP-117b spectrum reduced with the nominal pipeline (*left panel*), and the CASCADE pipeline (*right panel*). The model with estimated instrument performance was generated by *PandExo* (Batalha et al. 2017), assuming observation of a single transit with resolution $R \sim 30$. Hexagons, stars, and squares denote simulated NIRSpec G140M, G235M and G385M data, respectively.

the median model assumes 1000 K, whereas models retrieving the CASCADE HST/WFC3 spectrum retrieve lower median temperature of 800 K and cooler (Table 3).

The simultaneous measurement of H_2O, CH_4, CO, and CO_2 with the JWST thus promises to improve the constraint of the C/O of the planetary atmosphere and thus either confirm or refute the subsolar C/O that we tend to find for WASP-117b. The JWST will further be extremely important to constrain temperature, metallicity, CH_4 content, and thus methane chemistry in the exo-Saturn WASP-117b during transit. HST/UVIS observations will be more important to constrain hazes and cloudiness, which could also constrain atmospheric metallicity if an accuracy of 100 ppm or better is reached between 0.3 and 0.6 μm.

6. Discussion

This work shows that it is possible but also very challenging to investigate the atmospheric properties of a warm gas planet with an orbital period of 10 days and longer. The longer the orbital period, the longer the transit duration that needs to be captured in its entirety to perform accurate transmission spectroscopy. For WASP-117b, the total transit duration was 6 h, and a further 1–2 h of out-of-transit observations before and after transit were needed for calibration. Thus, a relatively large time investment was needed to obtain accurate transit spectroscopy with one single transit with HST (11 consecutive orbits) and VLT/ESPRESSO (10-h continuous observation). We also note that in 2018, there were only two occasions to capture the full transit with the VLT in its entirety from the ground. Otherwise, only partial transits were observable at night.

Atmosphere retrieval for WASP-117b favors, based on Bayesian analysis alone, a high-metallicity ([Fe/H] > 2.2), clear atmospheric composition (Model 1) "substantially" ($B > 4$) over the cloudy, low-metallicity solution (Model 2) as fit to the detected 3σ water spectrum. See Sect. 3.4 for an overview of the statistical evidence for the different models. We caution, however, against using Bayesian analysis of the HST/WFC3 data alone to rule out cloudy, low-metallicity solutions completely. We also caution that patchy clouds could mimic a high-metalliticy composition (Line & Parmentier 2016). We also note that Anisman et al. (2020) retrieved atmospheric properties of WASP-117b based on the same HST data with their own

pipeline. While these authors confirmed the H_2O detection, they were unable to place constraints on atmospheric metallicity or CH_4 abundances.

VLT/ESPRESSO observations of WASP-117b in the optical could not be used to constrain the atmospheric properties of WASP-117b further. The planetary Na and K signal is too small (2.4σ). We therefore conclude that exoplanets with long orbital periods should only be observed with ground-based telescopes after space-based observations have yielded first constraints on the cloud coverage and thus it is feasible to obtain a strong enough planetary signal in the optical.

TESS data could not be used for atmospheric constraints either because the TESS transit depth is abnormally shallow. We find that the TESS data are inconsistent with all atmospheric models that we explored, including cloudy and hazy atmosphere models (Sect. 5). Stellar activity is also unlikely to cause a 4% transit depth differences between the TESS and the WFC3/NIR wavelength range (Fig. 14). This example shows, however, how difficult it can be to combine measurements taken at two different epochs in two different wavelengths, even when the respective atmospheric measurements are precise and the host star is quiet (Rackham et al. 2019).

We maintain that a better constraint of the metallicity in the atmosphere of the eccentric WASP-117b is warranted as its properties would be complementary to other exoplanets on circular orbits in the same mass range: a clear sky, high-metallicity atmospheric composition would place WASP-117b significantly outside of the mass-metallicity relation inferred from the Solar System (Fig. 21, left panel). The exo-Saturn WASP-117b would thus join ranks with planets such as WASP-39b with comparatively high metallicity (Wakeford et al. 2018), and HAT-P-11b (Chachan et al. 2019), which do not follow the Solar System mass-metallicity relation either. A high atmospheric metallicity and its mass and radius would make WASP-117b very similar to WASP-39b. A cloudy, lower atmospheric metallicity composition (Model 2), on the other hand, would make WASP-117b more comparable to WASP-107b, WASP-43b and WASP-12b, which all follow the inferred Solar System mass-metallicity correlation (see also Thorngren & Fortney 2019). The values for the Solar System bodies, however, come with the caveat that even there the inferred metallicity depends on uncertainties such as core size and H_2O content (Thorngren & Fortney 2019).
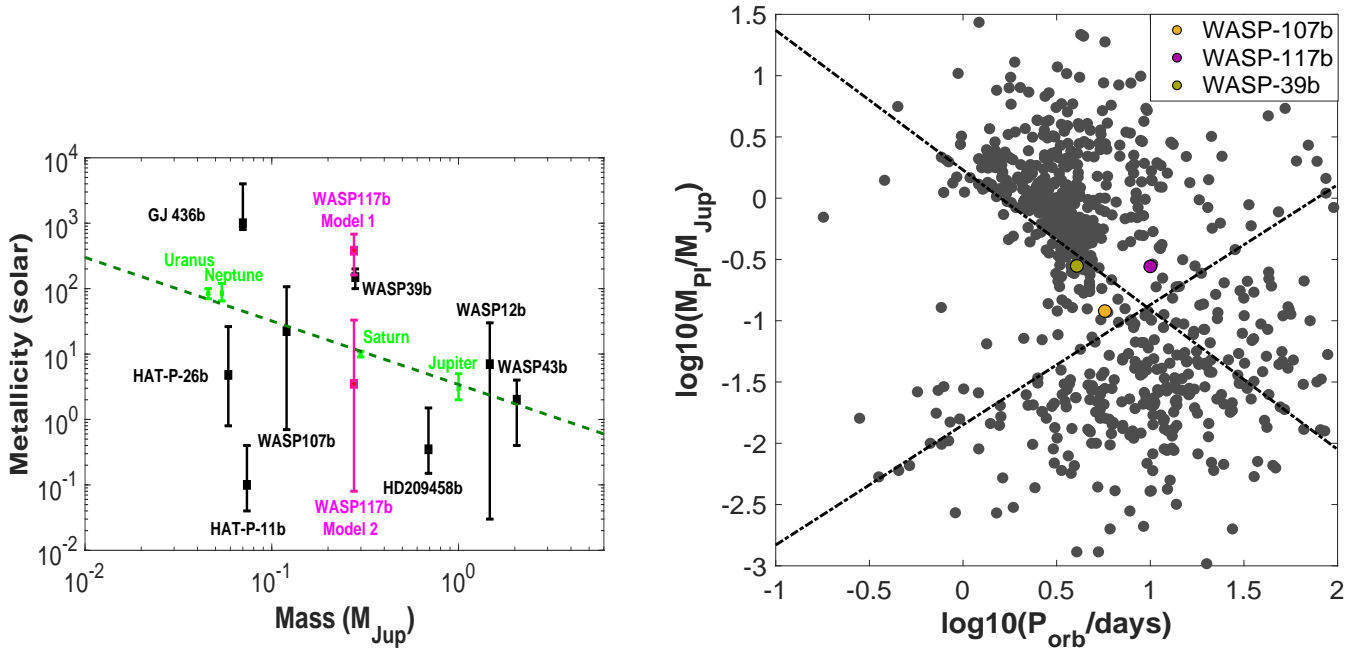
Fig. 21. Parameters of other known exoplanets in comparison with WASP-117b. *Left*: metallicity content derived for several exoplanets, adapted from Chachan et al. (2019). WASP-117b is mainly comparable in mass and metallicity to WASP-107b (Kreidberg et al. 2018), WASP-39b (Wakeford et al. 2018) and Solar System values (e.g. Wong et al. 2004; Helled & Guillot 2018; Chachan et al. 2019; Movshovitz et al. 2020) *Right*: all transiting planets with orbital periods shorter than 100 days and the Neptune-desert limits as proposed by Mazeh et al. (2016). WASP-39b and WASP-107b lie clearly inside the Neptune desert and are thus probably subject to erosion processes. WASP-117b lies outside the Neptune desert. (Derived from http://exoplanet.eu/ on 30.3.2020).

Furthermore, it is unclear in how far erosion processes have affected the atmospheric metallicity of extrasolar Neptune and Saturn-mass exoplanets (see, e.g., Owen & Lai 2018; dos Santos et al. 2019; Armstrong et al. 2020). The atmospheric properties of the eccentric WASP-117b could be illuminating here as well. The right panel of Fig. 21 shows that WASP-39b and WASP-107b lie in the Neptune desert (Mazeh et al. 2016), which is thought to reflect significant mass loss in this planetary mass and orbital period regime. The Neptune desert might also be called the sub-Jovian desert (Owen & Lai 2018) because the mass range of planets that are affected by photoevaporation is not strictly constrained to one Neptune mass. The super-Neptune WASP-107b was indeed found to experience loss of helium (Spake et al. 2018; Kirk et al. 2020) that could modify atmospheric metallicity over time. The exo-Saturn WASP-117b, on the other hand, should lie at 0.1168 AU at least during transit, which is well outside of the Neptune desert. The large distance during transit also explains the absence of deep Na and K features in the VLT/ESPRESSO data during transit. The atmospheric metallicity of WASP-117b could thus shed light on the question whether a high-metallicity atmospheric composition (if confirmed) is a sign of atmospheric erosion processes or instead of formation processes different from those of the Solar System.

The atmosphere of WASP-117b could also be illustrative for the study of disequilibrium processes. We found consistently low CH_4 abundances ($<10^{-4}$ volume fraction), expressed by our retrievals finding subsolar C/Os. Interpreting the atmospheric chemistry in more detail, however, requires the additional knowledge of the atmospheric temperature, which we were unable to constrain within the prescribed limits of 500–1500 K based on the WFC3 data alone.

We maintain, however, based on simple equilibrium temperatures calculations using Méndez & Rivera-Valentín (2017) that

temperatures between 700 and 1000 K are physical during the transit of the eccentric exo-Saturn WASP-117b (Sect. 4, Table 9). Chemical modeling for 700 and 1000 K for high ($350\times$ solar) and low metallicities ($5\times$ solar) further confirmed that very low CH_4 abundances in WASP-117b during transit are physically plausible. The 700 K models require disequilibrium chemistry to keep atmospheric CH_4 abundances below 10^{-4} volume fraction at $p = 10^{-4}$ bar, which we identified to first order as the main region that is probed during transmission. For the colder temperature models, we thus infer a vertical eddy diffusion coefficient of $K_{zz} \geq 10^8$ cm^2 s^{-1} to explain the absence of methane via CH_4 quenching.

The eddy diffusion coefficient K_{zz} is not trivially constrained for irradiated exoplanet atmospheres. Because these atmospheres tend to have deep, stably stratified radiative zones, the quenching point of most species typically does not lie within the (deeper) convective zone, and the traditional mixing-length theory breaks down. Instead, vertical mixing in irradiated exoplanets is described as being caused by medium- to large-scale atmospheric circulation, which when globally averaged may or may not act as a diffusive process (see, e.g., recent discussions in Zhang 2020; Showman et al. 2020). It is possible to study these atmospheric motions in 3D using general circulation models, and from these, derive order-of-magnitude parameterizations for the 1D eddy diffusion coefficient K_{zz}. For instance, for the hot Jupiter HD 209458b $K_{zz} = 5 \times 10^8$–10^9 cm^2 s^{-1} is derived at 1 bar (Parmentier et al. 2013; Menou 2019). We note that WASP-117b is cooler than HD 209458b, and K_{zz} should decrease with decreasing equilibrium temperature (Komacek et al. 2019). However, atmospheres with higher metallicities, such as we infer for WASP-117b, are also expected to have more vertical mixing than low-metallicity atmospheres (Charnay et al. 2015). Our tentative inference of $K_{zz} \geq 10^8$ cm^2 s^{-1} for the

(T_{eq} = 700 K)-model thus shows reasonable agreement with theoretically motivated expectations. However, a more sophisticated 3D model would be needed to further strengthen the agreement.

We argue that additional transit measurements of WASP-117b with HST/UVIS and the G280 grism could yield the necessary accuracy (200–100 ppm, Wakeford et al. 2020) to constrain a haze and cloud layer in WASP-117b. A low-resolution transmission spectrum in that range would reveal cloud properties due to the different scattering slopes between 0.3 and 0.5 μm for a clear atmosphere with high metallicity, a gray cloudy and hazy atmosphere with low metallicity (Fig. 5, upper panels). HST/UVIS measurements would thus already constrain atmospheric metallicity, which is degenerate with cloud properties in the HST/WFC3 range. In addition, HST/UVIS data could also potentially shed light on the origin of the shallow TESS transit depth. Such data would also warrant the use of a patchy cloud model (Line & Parmentier 2016).

Additional measurements with JWST/NIRSpec observations in the 0.8–5 micron region could further constrain CH$_4$ to at least $\leq 1.8 \times 10^{-5}$ volume fraction by observing in the 3.4 μm CH$_4$ band. A further constraint on CH$_4$ but also on CO at 4.7 μm would improve our understanding of atmosphere chemistry in WASP-117b, in particular, in conjunction with secondary eclipse observations to constrain temperature variations of WASP-117b during one orbit. The presence and absence of CH$_4$, or CO$_2$ at 4.3 micron would also constrain atmospheric metallicity and together with H$_2$O features confirm if WASP-117b indeed has a subsolar C/O. See also Guzmán-Mesa et al. (2020) for the capabilities of using JWST/NIRSpec to characterize the atmospheric chemistry of warm Neptunes.

7. Conclusions

We detected a 3σ water spectrum in the HST/WFC3 near-infrared transmission spectrum of WASP-117b. Comparison to one-dimensional atmospheric models assuming vertical isothermal temperatures, a uniform gray cloud coverage, and equilibrium chemistry (petitRADTRANS) shows that the water absorption is less pronounced than expected with a solar-type metallicity composition. Based on the preferences derived from the Bayes factor analysis alone, WASP-117b appears to have a high atmospheric metallicity ([Fe/H] = $2.58^{+0.26}_{-0.37}$ or $380^{+311}_{-217}\times$ solar, based on the nominal pipeline) with relatively clear skies. We cannot rule out lower atmospheric metallicity (Fe/H) < 1.75) with clouds and potentially an additional haze layer, however. We conclude that the data quality is too limited to warrant the application of a hazy atmosphere layer. The high metallicities that are strongly preferred by a Bayesian factor analysis are comparable to similarly high atmospheric metallicities ($151^{+46}_{-48} \times$ solar) found in WASP-39b, an exo-Saturn on a circular 4-day orbit around a cooler G type star (Wakeford et al. 2018).

While the basic atmospheric composition could be constrained, the temperature of WASP-117b during transit is unconstrained between 500–1500 K. Thus, we cannot identify whether the planet changes temperature during its eccentric 10-day orbit. Based on basic energy balance equations and realistic range of heat redistribution scenarios, we estimate that the planet could theoretically adopt equilibrium temperatures between 700 and 1000 K, depending on albedo and heat adjustment.

Because the temperature of WASP-117b during transit is not constrained, we cannot further test multiple possible hypotheses that may explain the low abundance of CH$_4$ in the atmosphere of WASP-117b during transit. If the atmospheric temperature is above 1000 K, then CH$_4$ cannot form and we do not expect to detect measurable levels of CH$_4$. If the temperature of WASP-117b is during transit 700 K, then higher CH$_4$ abundances than 10^{-4} volume fraction would form in the atmospheric levels that we probe ($p < 10^{-4}$ bar), assuming equilibrium chemistry. In the latter case, we would therefore need to invoke CH$_4$ quenching, and we could constrain the eddy diffusion coefficient to $K_{zz} \geq 10^8$ cm^2 s^{-1}. If WASP-117b is cool, it would thus show methane-quenching such as is indicated for the super-Neptune WASP-107b (Kreidberg et al. 2018).

Using VLT/ESPRESSO data alone, we do not find strong indications for Na and K within 3σ. In combination with HST/WFC3 data, however, we still find a "substantial" (2.4σ) preference for the presence of Na and K.

We also compared our WFC3 spectrum to TESS broadband observations of WASP-117b and found that the TESS transit depth is 4% shallower than the HST/WFC3 transit depth. We cannot explain this difference with systematic noise, stellar activity, or clouds.

Furthermore, with the new VLT/ESPRESSO data, we improved through the Rossiter-MacLaughlin effect measurements of the $v \sin i$ of the star and the projected spin-orbit angle to 1.46 ± 0.14 km s^{-1} and $-46.9^{+5.5}_{-4.8}$ deg.

8. Outlook

WASP-117b is one of the first warm eccentric exo-Saturns with a long orbital period (10 days), the atmosphere of which was characterized. While the observations proved to be challenging, they indicate that WASP-117b could have interesting properties such that further observations with the JWST and HST in the optical range would be warranted.

WASP-117b could be an ideal target to investigate how (disequilibrium) chemistry processes are controlled by precisely known, varying heating conditions on a 10-day orbit around a quiet F main sequence star. This studies would complement studies of eccentric exoplanets on tighter orbits (e.g., Lewis et al. 2013, 2014; Agúndez et al. 2014). WASP-117b is also a complementary observation target to the super-Neptunes WASP-107b and WASP-39b on circular, tighter orbits to aid our understanding of how erosion processes shape atmospheric properties such as metallicities in the Neptune-to-Saturn mass range.

Several promising extrasolar exo-Saturns on eccentric orbits with $P_{orb} \geq 10$ days around bright stars suitable for in-depth atmospheric characterization are currently known (Brahm et al. 2016, 2018, 2019; Jordán et al. 2019; Rodriguez et al. 2019). One such planet, K2-287 b, orbits a G-type star with a time-averaged equilibrium temperature of 800 K (Jordán et al. 2019), that is, it is at least 200 K cooler than WASP-117b would thus be even more ideal to study disequilibrium chemistry in warm eccentric gas planets on relatively wide orbits.

References

Agúndez, M., Venot, O., Selsis, F., & Iro, N. 2014, ApJ, 781, 68
Anisman, L. O., Edwards, B., Changeat, Q., et al. 2020, AJ, 160, 233
Apai, D., Karalidi, T., Marley, M. S., et al. 2017, Science, 357, 683
Armstrong, D. J., Lopez, T. A., Adibekyan, V., et al. 2020, Nature, 583, 39
Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, ARA&A, 47, 481
Astropy Collaboration (Robitaille, T. P., et al.) 2013, A&A, 558, A33
Astropy Collaboration (Price-Whelan, A. M., et al.) 2018, AJ, 156, 123
Barman, T. S., Macintosh, B., Konopacky, Q. M., & Marois, C. 2011a, ApJ, 735, L39
Barman, T. S., Macintosh, B., Konopacky, Q. M., & Marois, C. 2011b, ApJ, 733, 65
Barstow, J. K. 2020, MNRAS, 497, 4183
Barstow, J. K., Aigrain, S., Irwin, P. G. J., et al. 2013, MNRAS, 430, 1188
Batalha, N. E., Mandell, A., Pontoppidan, K., et al. 2017, PASP, 129, 064501
Benneke, B., & Seager, S. 2013, ApJ, 778, 153
Benneke, B., Knutson, H. A., Lothringer, J., et al. 2019, Nat. Astron., 3, 813
Brahm, R., Jordán, A., Bakos, G. Á., et al. 2016, AJ, 151, 89
Brahm, R., Espinoza, N., Jordán, A., et al. 2018, MNRAS, 477, 2572
Brahm, R., Espinoza, N., Jordán, A., et al. 2019, AJ, 158, 45
Brammer, G., Ryan, R., & Pirzkal, N. 2015, Source-dependent Master Sky Images for the WFC3/IR grisms, Space Telescope WFC Instrument Science Report (Washington DC: Nasa)
Buchner, J. 2016, Stat. Comput., 26, 383
Caldas, A., Leconte, J., Selsis, F., et al. 2019, A&A, 623, A161
Carnall, A. C. 2017, ArXiv e-prints [arXiv:1705.05165]
Castelli, F., & Kurucz, R. L. 2003, IAU Symp., 210, A20
Chachan, Y., Knutson, H. A., Gao, P., et al. 2019, AJ, 158, 244
Charnay, B., Meadows, V., & Leconte, J. 2015, ApJ, 813, 15
Crossfield, I. J. M. 2015, PASP, 127, 941
Czesla, S., Klocová, T., Khalafinejad, S., Wolter, U., & Schmitt, J. H. M. M. 2015, A&A, 582, A51
Czesla, S., Schröter, S., Schneider, C. P., et al. 2019, PyA: Python astronomy-related packages
Deming, D., Wilkins, A., McCullough, P., et al. 2013, ApJ, 774, 95
Deming, D., Knutson, H., Kammer, J., et al. 2015, ApJ, 805, 132
dos Santos, L. A., Ehrenreich, D., Bourrier, V., et al. 2019, A&A, 629, A47
Espinoza, N. 2018, Res. Notes Am. Astron. Soc., 2, 209
Espinoza, N., Kossakowski, D., & Brahm, R. 2019, MNRAS, 490, 2262
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Golub, G. H., Heath, M., & Wahba, G. 1979, Technometrics, 21, 215
Guilluy, G., Sozzetti, A., Brogi, M., et al. 2019, A&A, 625, A107
Guzmán-Mesa, A., Kitzmann, D., Fisher, C., et al. 2020, AJ, 160, 15
Helled, R., & Guillot, T. 2018, Internal Structure of Giant and Icy Planets: Importance of Heavy Elements and Mixing (Berlin: Springer), 44
Hoerl, A. E., & Kennard, R. W. 1970, Technometrics, 12, 55
Horne, K. 1986, PASP, 98, 609
Janson, M., Brandt, T. D., Kuzuhara, M., et al. 2013, ApJ, 778, L4
Jordán, A., Brahm, R., Espinoza, N., et al. 2019, AJ, 157, 100
Jordán, A., Brahm, R., Espinoza, N., et al. 2020, AJ, 159, 145
Kass, R. E., & Raftery, A. E. 1995, J. Am. Stat. Assoc., 90, 773
Kipping, D. M. 2013, MNRAS, 435, 2152
Kirk, J., Alam, M. K., López-Morales, M., & Zeng, L. 2020, AJ, 159, 115
Kochanek, C. S., Shappee, B. J., Stanek, K. Z., et al. 2017, PASP, 129, 104502
Komacek, T. D., Showman, A. P., & Parmentier, V. 2019, ApJ, 881, 152
Kreidberg, L. 2015, PASP, 127, 1161
Kreidberg, L., Bean, J. L., Désert, J.-M., et al. 2014, ApJ, 793, L27
Kreidberg, L., Line, M. R., Thorngren, D., Morley, C. V., & Stevenson, K. B. 2018, ApJ, 858, L6
Kurucz, R. L. 1993, ASP Conf. Ser., 44, 87
Lendl, M., Triaud, A. H. M. J., Anderson, D. R., et al. 2014, A&A, 568, A81
Lewis, N. K., Knutson, H. A., Showman, A. P., et al. 2013, ApJ, 766, 95
Lewis, N. K., Showman, A. P., Fortney, J. J., Knutson, H. A., & Marley, M. S. 2014, ApJ, 795, 150

Line, M. R., & Parmentier, V. 2016, ApJ, 820, 78
Lodders, K., & Fegley, B. 2002, Icarus, 155, 393
Lomb, N. R. 1976, Ap&SS, 39, 447
Luger, R., Agol, E., Kruse, E., et al. 2016, AJ, 152, 100
MacDonald, R. J., & Madhusudhan, N. 2017, MNRAS, 469, 1979
MacDonald, R. J., Goyal, J. M., & Lewis, N. K. 2020, ApJ, 893, L43
Madhusudhan, N. 2012, ApJ, 758, 36
Mallonn, M., von Essen, C., Herrero, E., et al. 2019, A&A, 622, A81
Mazeh, T., Holczer, T., & Faigler, S. 2016, A&A, 589, A75
Méndez, A., & Rivera-Valentín, E. G. 2017, ApJ, 837, L1
Menou, K. 2019, MNRAS, 485, L98
Miles, B. E., Skemer, A. J., Barman, T. S., Allers, K. N., & Stone, J. M. 2018, ApJ, 869, 18
Mollière, P., van Boekel, R., Dullemond, C., Henning, T., & Mordasini, C. 2015, ApJ, 813, 47
Mollière, P., van Boekel, R., Bouwman, J., et al. 2017, A&A, 600, A10
Mollière, P., Wardenier, J. P., van Boekel, R., et al. 2019, A&A, 627, A67
Moses, J. I., Marley, M. S., Zahnle, K., et al. 2016, ApJ, 829, 66
Movshovitz, N., Fortney, J. J., Mankovich, C., Thorngren, D., & Helled, R. 2020, ApJ, 891, 109
Nagao, M., & Matsuyama, T. 1979, Comput. Graph. Image Process., 9, 394
Owen, J. E., & Lai, D. 2018, MNRAS, 479, 5012
Parmentier, V., Showman, A. P., & Lian, Y. 2013, A&A, 558, A915
Parmentier, V., Fortney, J. J., Showman, A. P., Morley, C., & Marley, M. S. 2016, ApJ, 828, 22
Pepe, F. A., Cristiani, S., Rebolo Lopez, R., et al. 2010, SPIE Conf. Ser., 7735, 77350F
Pinhas, A., & Madhusudhan, N. 2017, MNRAS, 471, 4355
Pirzkal, N., Ryan, R., & Brammer, G. 2016, Trace and Wavelength Calibrations of the WFC3 G102 and G141 IR Grisms, Space Telescope WFC Instrument Science Report (Washington DC: Nasa)
Piskunov, N., & Valenti, J. A. 2017, A&A, 597, A16
Pluriel, W., Zingales, T., Leconte, J., & Parmentier, V. 2020, A&A, 636, A66
Queloz, D., Eggenberger, A., Mayor, M., et al. 2000, A&A, 359, L13
Rackham, B. V., Apai, D., & Giampapa, M. S. 2018, ApJ, 853, 122
Rackham, B. V., Apai, D., & Giampapa, M. S. 2019, AJ, 157, 96
Rocchetto, M., Waldmann, I. P., Venot, O., Lagage, P. O., & Tinetti, G. 2016, ApJ, 833, 120
Rodriguez, J. E., Quinn, S. N., Huang, C. X., et al. 2019, AJ, 157, 191
Samland, M., Bouwman, J., Hogg, D. W., Brandner, W., & Henning, T. 2021, A&A, 646, A24
Scargle, J. D. 1982, ApJ, 263, 835
Schölkopf, B., Hogg, D. W., Wang, D., et al. 2016, Proc. Natl. Acad.f Sci., 113, 7391
Shapiro, A. I., Solanki, S. K., Krivova, N. A., et al. 2014, A&A, 569, A38
Shappee, B. J., Prieto, J. L., Grupe, D., et al. 2014, ApJ, 788, 48
Showman, A. P., Lewis, N. K., & Fortney, J. J. 2014, ArXiv e-prints [arXiv:1411.4731]
Showman, A. P., Tan, X., & Zhang, X. 2019, ApJ, 883, 4
Showman, A. P., Tan, X., & Parmentier, V. 2020, Space Sci. Rev., 216, 139
Spake, J. J., Sing, D. K., Evans, T. M., et al. 2018, Nature, 557, 68
Thorngren, D., & Fortney, J. J. 2019, ApJ, 874, L31
Thorngren, D., Gao, P., & Fortney, J. J. 2019, ApJ, 884, L6
Thuillier, G., Floyd, L., Woods, T. N., et al. 2004, Adva. Space Res., 34, 256
Trotta, R. 2008, Contemp. Phys., 49, 71
van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., et al. 2014, PeerJ, 2, e453
Venot, O., Hébrard, E., Agúndez, M., et al. 2012, A&A, 546, A43
Wakeford, H. R., Sing, D. K., Kataria, T., et al. 2017, Science, 356, 628
Wakeford, H. R., Sing, D. K., Deming, D., et al. 2018, AJ, 155, 29
Wakeford, H. R., Sing, D. K., Stevenson, K. B., et al. 2020, AJ, 159, 204
Wang, D., Hogg, D. W., Foreman-Mackey, D., & Schölkopf, B. 2016, PASP, 128, 094503
Wong, M. H., Mahaffy, P. R., Atreya, S. K., Niemann, H. B., & Owen, T. C. 2004, Icarus, 171, 153
Yan, F., & Henning, T. 2018, Nat. Astron., 2, 714
Yan, F., Fosbury, R. A. E., Petr-Gotzens, M. G., et al. 2015a, Int. J. Astrobiol., 14, 255
Yan, F., Fosbury, R. A. E., Petr-Gotzens, M. G., Zhao, G., & Pallé, E. 2015b, A&A, 574, A94
Yan, F., Pallé, E., Fosbury, R. A. E., Petr-Gotzens, M. G., & Henning, T. 2017, A&A, 603, A73
Yan, F., Casasayas-Barris, N., Molaverdikhani, K., et al. 2019, A&A, 632, A69
Zhang, X. 2020, Res. Astron. Astrophys., 20, 099
Zhou, Y., Apai, D., Lew, B. W. P., & Schneider, G. 2017, AJ, 153, 243

Appendix A: Description of the HST data-reduction pipeline

We describe the two data reduction pipelines in more detail below. First, we describe the nominal pipeline, which we mainly use to infer atmospheric properties of WASP-117b. Then, we describe the CASCADE pipeline, which we use to understand the limitation of the data, impact of systematics and to provide an independent confirmation of derived atmospheric properties.

A.1. Nominal pipeline

We started our data reduction with the `ima` files, which were produced by the `CALWFC3` data reduction pipeline. The `ima` files contain all calibrated nondestructive reads. Before extracting the light curves, we first measured the subpixel-level telescope pointing drifts using the cross-correlation method and corrected the drift-induced wavelength calibration error by aligning images in the x-direction (dispersion direction) using linear interpolation shift. We then followed the standard procedures (Deming et al. 2013) to extract light curves from the `ima` files. First, differential images, which represent the spectrum collected in each nondestructive read, were obtained by subtracting each `ima` frame from its previous read. Second, we identified and corrected for bad pixels and cosmic rays. We first marked pixels with data quality flags of 4, 16, 32, and 256 as bad pixels. We then searched for spurious pixels caused by cosmic-ray hits by comparing the differential images with their median-filtered images and selected the 5σ outliers as cosmic-ray pixels. Pixels identified as either bad or cosmic rays were replaced by linear interpolations of their neighboring pixels. Third, we subtracted the sky background from each differential image. The sky background was estimated by taking the median values of pixels that were neither illuminated by any astrophysical sources nor by bad pixels. To select the background region, we manually constructed pixel masks to exclude all visible spectral traces. We also applied 5σ, 10-iteration sigma-clipping on the unmasked images to remove remaining bright pixels from background estimation. Finally, we measured the raw light curve of each column by taking a $R = 45$ pixel aperture photometry. We integrated spectroscopic light curves using a 5-pixel wide bin. The raw light curves are presented in Fig. A.1.

The raw light curves demonstrate systematic trends that are typical of WFC3/NIR observations (the "ramp effect"). Zhou et al. (2017) introduced a physically motivated model for this effect (RECTE) by assuming that the rise of the light curve at the beginning of each orbit is introduced by stimuli that gradually fill charge traps that are caused by detector defects. In the RECTE model, there are two populations of charge traps that differ in their trapping efficiencies and in the way they sustain lifetime. We corrected for the systematics and fit the transit profile simultaneously. We used RECTE to model the systematics. For the transit profile, we adopted the `batman` (Kreidberg 2015) code and assumed the linear limb-darkening formula. The parameters related to the properties of the charge trap, including the numbers, trapping efficiencies, and trapping lifetimes, were predetermined. There were four free parameters in the RECTE model. They were the initially trapped charges and the number of trapped charges added during Earth occultations for both trap populations. We also included linear trends (two free parameters) for light curves in both forward- and backward-scanning directions as part of the systematic model. We noted that the light curves demonstrated correlated changes with the telescope-pointing variations in the y-direction (cross-dispersion
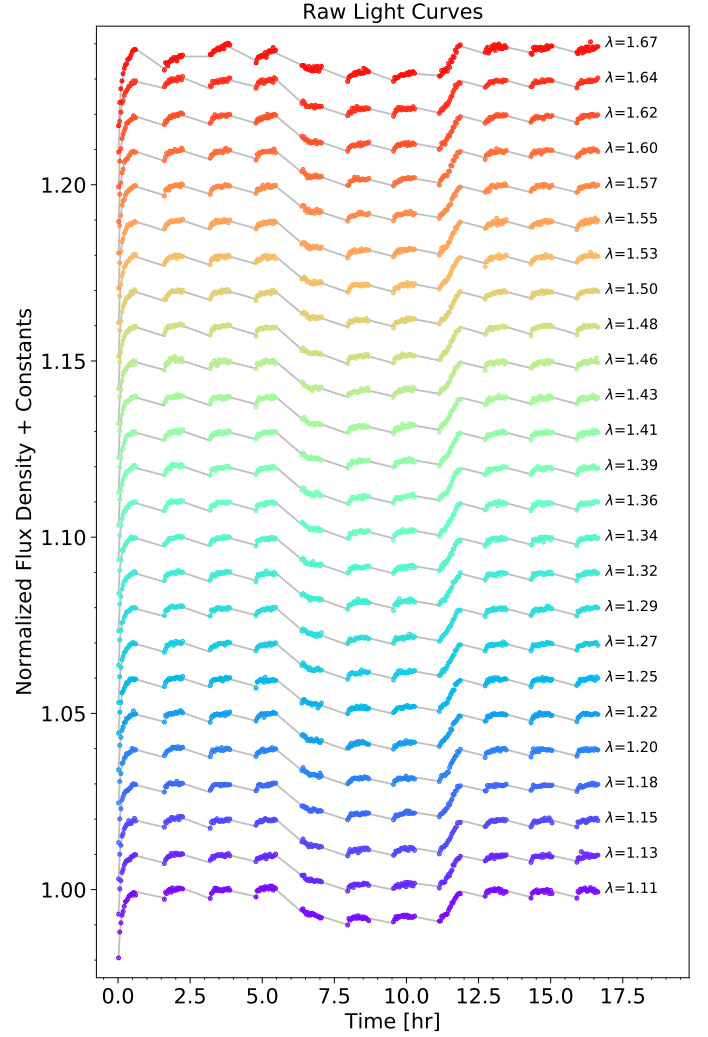

Fig. A.1. Raw light curves extracted from HST/WFC3 observations of WASP-117b. Each light curve is constructed by the integration of a 5-pixel-wide wavelength bin. Colors of the light curve represent their central wavelengths (using a rainbow color scheme, from red for the longest wavelength (1.67 μm) to blue for the shortest (1.11 μm)). Light-curve systematics are clearly visible, particularly at the first orbit of the observations. The gray curves show the best-fitting RECTE+transit profile models.

direction), which could be adequately modeled by a linear function of Δy. We added this Δy linear term as part of the systematics model as well.

We first conducted the joint transit+systematics fit on the 1.10–1.70 μm broadband light curve. In addition to the parameters for the systematics model described above, parameters of the transit mid-time, the transit depth, and the limb-darkening coefficient were optimized using the `emcee` code (Foreman-Mackey et al. 2013). The other planetary-system related parameters, including the orbital period, system semimajor axis, eccentricity, inclination, and the longitude of periastron, were fixed using values from the literature (Lendl et al. 2014; Mallonn et al. 2019), which were already good enough for the fit. We then conducted the fit to spectroscopic light curves using the same method except fixing the transit mid-time using the best-fitting value from the broadband fit. The corrected light curves and their best-fitting transit profile models are presented in Fig. A.2. The data for the derived spectrum is shown in Table D.1
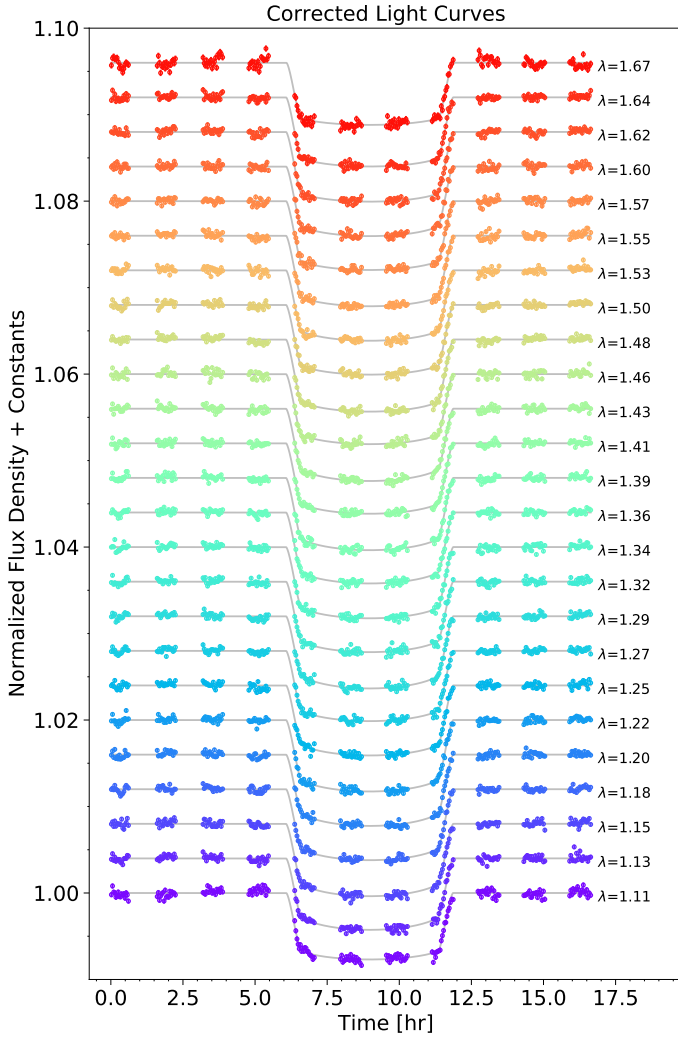
Fig. A.2. Systematics-corrected light curves of WASP-117b. Compared to those shown in Fig. A.1, the systematics are divided out of the light curves. The light curves are color-coded identically to Fig. A.1. The gray curves represent the best-fitting transit profile models.

A.2. CASCADE

In addition to the nominal pipeline, we also used the calibration of transit spectroscopy using causal data (CASCADe[6]) data-reduction package developed with the Exoplanet Atmosphere New Emission Transmission Spectra Analysis (ExoplANETS-A[7]) Horizon-2020 programme. The CASCADe pipeline also starts the data reduction with the `ima` intermediate data product, which were produced by the `CALWFC3` data reduction pipeline. As the data was observed using the `SPARS10` readout mode with `NSAMP = 15`, the first nondestructive readout of the 16 samples of the detector ramp stored in the `ima` data product was removed as it has a much shorter readout time than the other readouts. For the remaining readouts, we constructed pairwise differences between consecutive readouts resulting in 14 signal measurements per detector ramp.

To subtract the background signal from our target signal, we used the background-subtraction procedure outlined in Brammer et al. (2015). This entails fitting and subtracting "master sky"

images for the zodiacal light and a 1.083 μm emission line from the Earth's upper atmosphere appearing in exposures outside of the Earth's shadow. For details on the exact procedure, we refer to Brammer et al. (2015).

To identify and clean bad pixels, we first masked all pixels with a quality flag other than zero. We then employed an edge-preserving filtering technique similar to Nagao & Matsuyama (1979). We refer to this paper for details of the method. This procedure ensures that the spatial profile of the dispersed light is preserved. As directional filters we used a kernel defined by a two-dimensional Gausian kernel whose width is defined by the two standard deviations σ_x and σ_y of 0.1 and 3 pixels, which we rotated to obtain a series of different orientations. We flagged all pixels that deviated more than 4σ from the mean determined by the filter profile. After the filtering, we have three data products: one where all bad pixels are masked; a cleaned dataset where all bad pixels are replaced by a mean value determined by the optimal filter for those pixels, which we used to determine the relative telescope movements; and a smoothed data product on which we based our spectral extraction profile.

We used the image of the target with the F126N filter taken at the start of the first orbit to determine the initial wavelength solution and position of the time series of spectral images taken with the G141 grism. We implemented the method outlined in Pirzkal et al. (2016) to determine wavelength calibration and spectral trace from the acquisition image, using the *G141.F126N.V4.32.conf* configuration file for the correct polynomial solution. The initial wavelength and trace solution is expected to have an 0.1 pixel accuracy for the trace description and better than 0.5 pixel (~20 nanometers) accuracy for the overall wavelength solution. As our observations are performed using the spatial-scan mode and thus move back and forth around the initial detector position of the target system, we determined the center-of-light position of the first signal measurement to determine the absolute offset to the initial trace position. We then employed a cross-correlation technique in Fourier space, using the register_translation and warp_polar functions in the scikit-image package (van der Walt et al. 2014) version 16.2, to determine the relative positional shifts $(\Delta x, \Delta y)$, rotation $(\Delta\Omega)$ and scale changes of all subsequent exposures. The average change per scan is shown in Fig. A.3. We used these relative movements and rotations to update our wavelength solution for the spectral images at each time step.

We then optimally extracted (see Horne 1986) the 1D spectra from the spectral images using as an extraction profile for each time step the normalized filtered and smoothed spectral images. After the spectral extraction step, we rebinned the spectra to a uniform wavelength grid using the method outlined in Carnall (2017) to determine the corresponding flux values and error estimates, after which we created an average spectra per spectral scan. The resulting white-light curve of the uncalibrated time series data is shown in the top panel of Fig. A.4.

To calibrate the extracted spectral timeseries data and to extract the transmission spectrum of WASP 117b, we made use of the *half-sibling-regression* method developed by Schölkopf et al. (2016) using causal connections within the dataset to model both transit signal and any systematics, and which has been successfully applied to transit observations from the Kepler mission (Wang et al. 2016) and field-stabilized imaging data (Samland et al. 2021). For a detailed discussion and the mathematical proof of this method, we refer to Schölkopf et al. (2016). We briefly discuss the main concepts and our implementation of the method for the HST spectroscopic data.
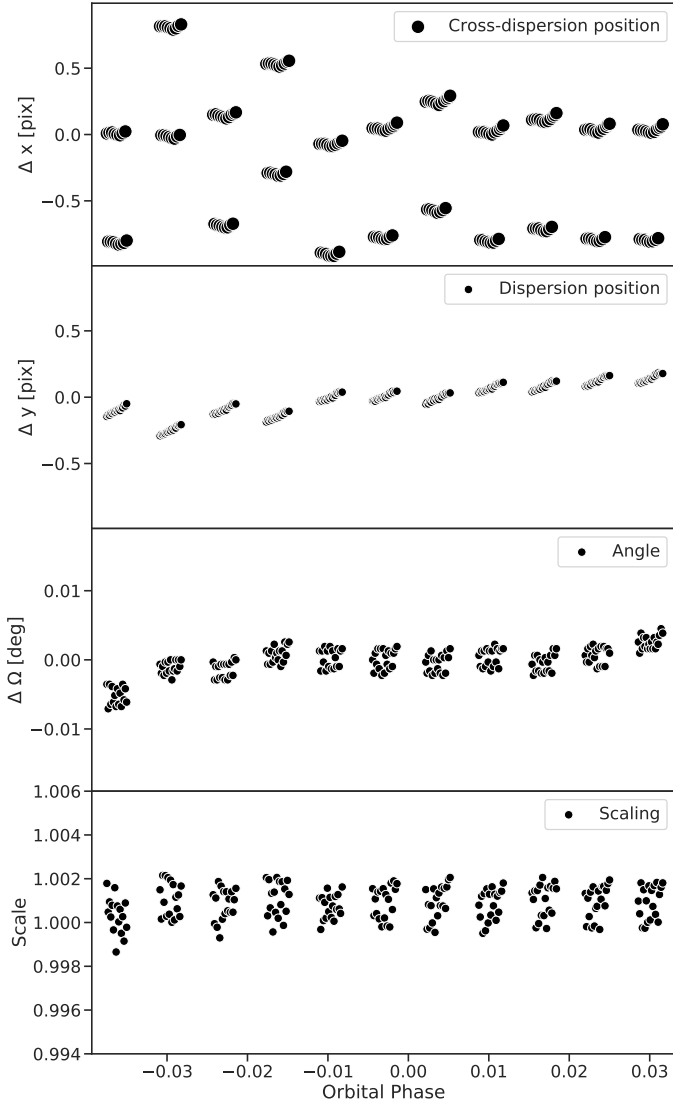
---

[6] https://jbouwman.gitlab.io/CASCADe/
[7] https://cordis.europa.eu/project/rcn/212911/factsheet/en

Fig. A.3. Derived pointing jitter and image distortion during the transit observation. *Top panel*: relative pointing movement in pixels for each spatial scan in the cross-dispersion direction. *Second panel*: relative pointing variations in the dispersion direction. *Third panel*: relative change in the rotation angle, and the bottom panel presents the change in size.

We assumed a time series of (spectral) image data $D_{i,k} \in \mathbb{R}^{M \times T}$, where M is the number of detector pixels and T is the number of discrete measurement times in the time series. The main idea is that all data $D_{i,k}$ as a function of detector position x_i at time t_k can be described by an additive noise model,

$$D(x_i, t_k \,|\, N) = S(x_i, t_k) + g(x_i, t_k \,|\, N) + \epsilon(x_i, t_k)$$
$$\forall i \in \{1, 2, \ldots, M\} \quad \forall k \in \{1, 2, \ldots, T\} \quad, \tag{A.1}$$

where S is the (transit) signal contribution we wish to measure, g denotes the functional form of the systematics affecting this measurement, and ϵ is the stochastic noise (e.g., photon noise). N here stands in for all hidden parameters that may cause the systematics. We can now split the dataset into two subsets,

$$\mathcal{P}_s = \{x_i \,|\, S(x_i, t_k) > 0\}, \quad \forall k \in \{1, 2, \ldots, T\}, \quad i \in \{1, 2, \ldots, N\}$$
$$\mathcal{P}_c = \{x_i \,|\, S(x_i, t_k) = 0\} \tag{A.2}$$
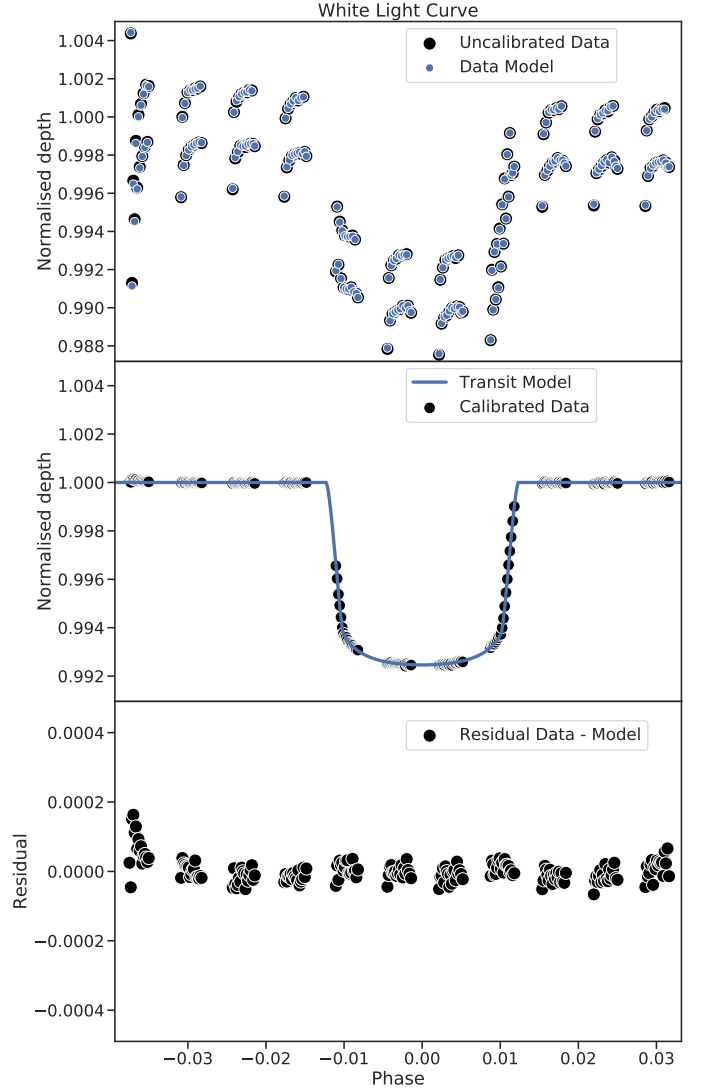
Fig. A.4. Calibration results of the spectral light curves with CASCADE. *Top panel*: wavelength-averaged uncalibrated light curves for each spatial scan and the wavelength-averaged fitted regression model. *Bottom panel*: fit residuals. *Middle panel*: derived calibrated light curve and transit model.

where the set $\mathcal{P}_s$ consists of all detector data that see the signal S and $\mathcal{P}_c$ all other detector data. In practice, these detector data are assigned to $\mathcal{P}_c$, which is far enough removed from the position of the source S on the detector such that the calibrator signal is independent of the signal of interest. This is typically a few times the full width at half maximum of the point spread function. Furthermore, as some of the systematics are unique to point sources (e.g., intrapixel sensitivity variations), $\mathcal{P}_c$ is limited to the pixels that see other point sources in case of multi-object observations (see also Wang et al. 2016). An estimate of the signal S can be made as follows:

$$\hat{S}(x_i, t_k) = D(x_i, t_k \,|\, N) - \mathbb{E}[D(x_i, t_k) \,|\, D(x_j, t_k), \Theta(t_k)],$$
$$\forall x_i \in \mathcal{P}_s, \quad \forall x_j \in \mathcal{P}_c, \quad \forall k \in \{1, 2, \ldots, T\} \quad, \tag{A.3}$$

where $\mathbb{E}$ is the expectation value, that is, the regression model for the dataset $D_{i,k}$ given the other data $D_{j,k}$ and any auxiliary data Θ_k, such as the derived telescope movement, as shown in Fig. A.3, that are correlated with the systematics present in the data.

In contrast to multi-object photometry such as the Kepler data, in the case of single-object spectra we use the dispersed light rather than the subset of data that contains the signals of other stars as regressors. To do this, we calibrate the observed signal at a particular wavelength with the data at other wavelengths. For the extracted HST spectra, we can rewrite Eq. (A.1) as

$$
D(\lambda_i, t_k \,|\, N) = S(\lambda_i, t_k) + g(\lambda_i, t_k \,|\, N) + \epsilon(\lambda_i, t_k)
$$
$$
\forall i \in \{1, 2, \ldots, M\} \quad \forall k \in \{1, 2, \ldots, T\} \tag{A.4}
$$

where the data become a function of the wavelength λ_i. The subsets of detector positions defining the data containing the signal of interest and the dataset used as calibrator described by Eq. (A.2) now becomes

$$
\mathcal{P}_s = \{\lambda_i\}, \quad i \in \{1, 2, \ldots, N\}
$$
$$
\mathcal{P}_c = \left\{ \left\{ \lambda_j \,|\, |\lambda_1 - \lambda_j| > \Delta\lambda \right\}, \ldots, \left\{ \lambda_j \,|\, |\lambda_N - \lambda_j| > \Delta\lambda \right\} \right\} \tag{A.5}
$$

where we used $\Delta\lambda = 0.07\mu$m, or about three resolution elements. For each λ_i in $\mathcal{P}_s$ we have a separate subset $\mathcal{P}_{c,i}$ of wavelengths that are used to calibrated the time-series data at λ_i. In practice, a linear regression model is sufficient to simultaneously model the transit signal and systematics. As the regressors can be colinear, we use the *ridge regression* estimator proposed by Hoerl & Kennard (1970). The strength of the regularization is determined by generalized cross-validation (Golub et al. 1979). The regression model for the transit signal and systematics thus becomes

$$
\operatorname*{argmin}_{\omega_i, \alpha_{i,j}} \left\| \underbrace{D(\lambda_i, t_k)}_{\text{data}} - \underbrace{\left(\omega_i \hat{S}(\lambda_i, t_k) + \sum_{\lambda_j \in \mathcal{P}_{c,i}} \alpha_{i,j} D(\lambda_j, t_k) \right)}_{\text{planet + systematics model}} \right\|^2
$$
$$
+ \underbrace{\kappa_i \left\| (\omega_i, \alpha_i) \right\|^2}_{\text{regularization}}, \forall \lambda_i \in \mathcal{P}_s, \tag{A.6}
$$

where $\hat{S}$ is the model for the transit signal S, and κ is the regularization parameter. The parameters ω and α are the fit parameters for the transit depth and the contributions of the time-series data from the calibration subset, respectively. We estimate the regularization parameter for each wavelength using a generalized cross-validation. In addition to the data-derived regressors, other linear terms can be added to Eq. (A.6), such as a constant value, time t, or $\Delta X(t_k)$, which is the relative cross-dispersion position, as shown in Fig. A.3. For our final solution we added a constant term and the relative cross-dispersion position because adding these terms gave the highest signal-to-noise ratio on the derived transit spectrum.

We do not need to know the absolute transit depth but only the relative transit depth, therefore we used the regression estimate as defined by Eq. (A.6) as follows:

$$
\frac{D(\lambda_i, t_k)}{\mathbb{E}[D(\lambda_i, t_k) \,|\, D(\lambda_j, t_k)] - \omega_i \hat{S}(\lambda_i, t_k)} \tag{A.7}
$$

to estimate the relative transit depth as a function of wavelength. The resulting combined regression model for the transit signal and systematics for the white-light curve is shown in the top panel of Fig. A.4. The residuals of the data minus the CASCADE regression model are shown in the lower panel of the same figure. The derived calibrated normalized light curve and transit model is shown in the middle panel of Fig. A.4. The data for the derived spectrum is shown in Table D.2 and in Fig. 2.

Appendix B: Corner plots for the retrieval of models we used for the atmosphere analysis of WASP-117b

In the following, the full corner plots as used in Sects. 3.1 and 3.2 are shown for the fit parameters and for each model we used to constrain the atmospheric properties of WASP-117b during transit based on HST/WFC3 data. We include the full corner plots here to allow a more thorough understanding of the interdependence of model parameters. Furthermore, the corner plots document the reliability of derived atmospheric properties.
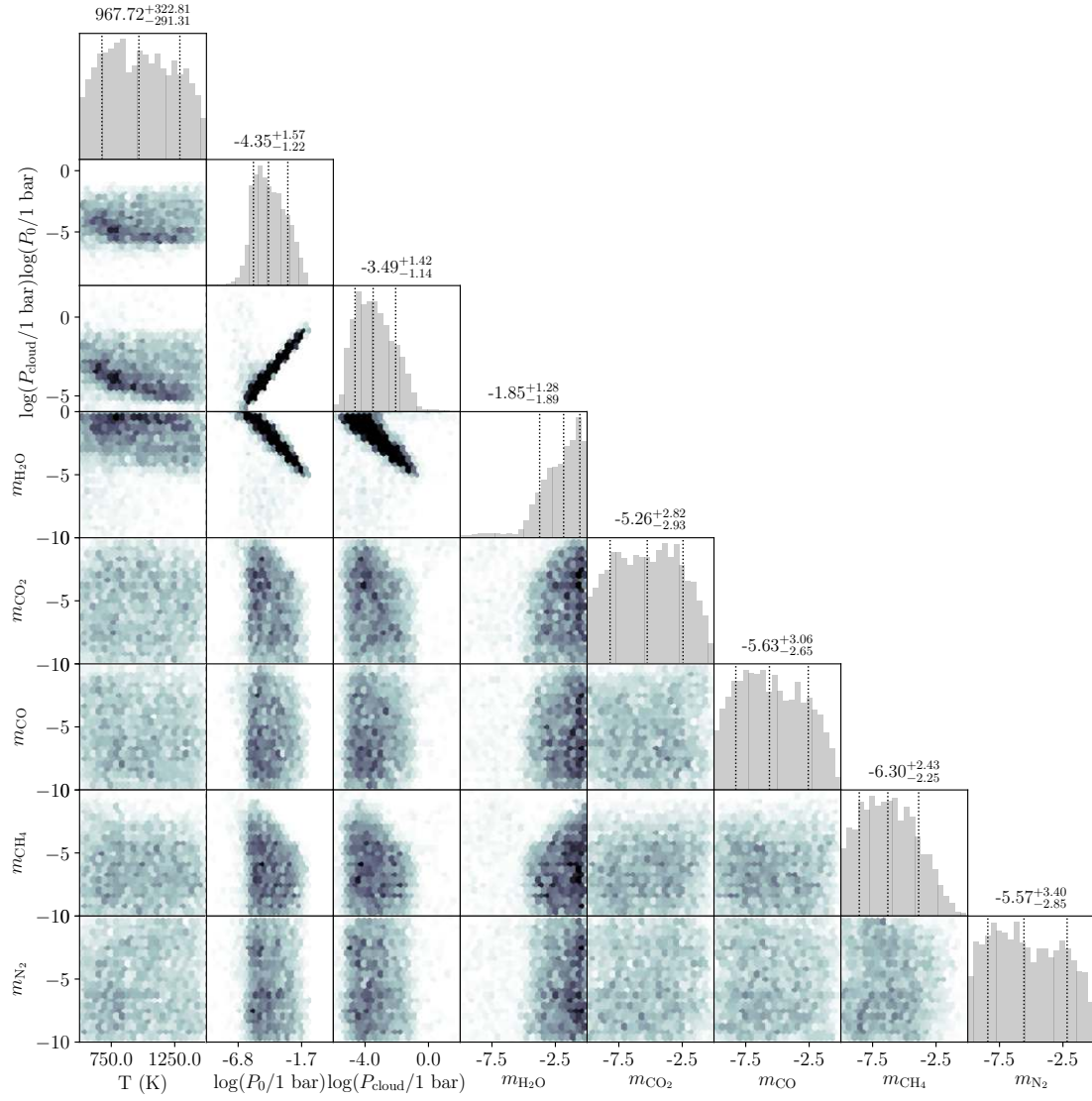
Full retrieval with H$_2$O - Nominal spectrum

Fig. B.1. Posterior probability distributions of the retrieval on HST/WFC3 dataset for WASP-117b, reduced with the nominal pipeline. The full retrieval including H$_2$O was applied (see Sect. 3.1). Median parameter values and 68% confidence intervals for the marginalized 1D posterior probability distributions are indicated with horizontal error bars.
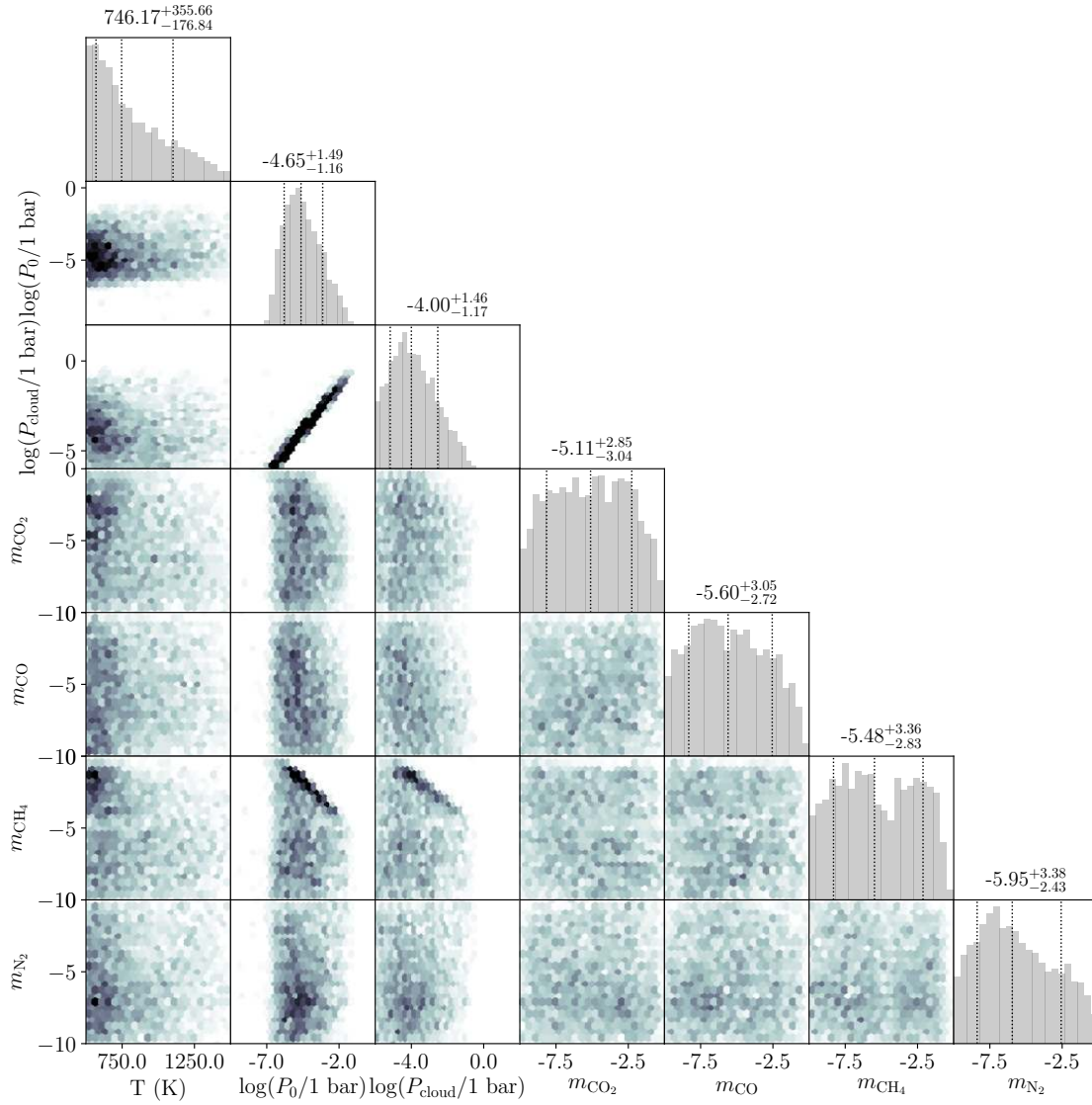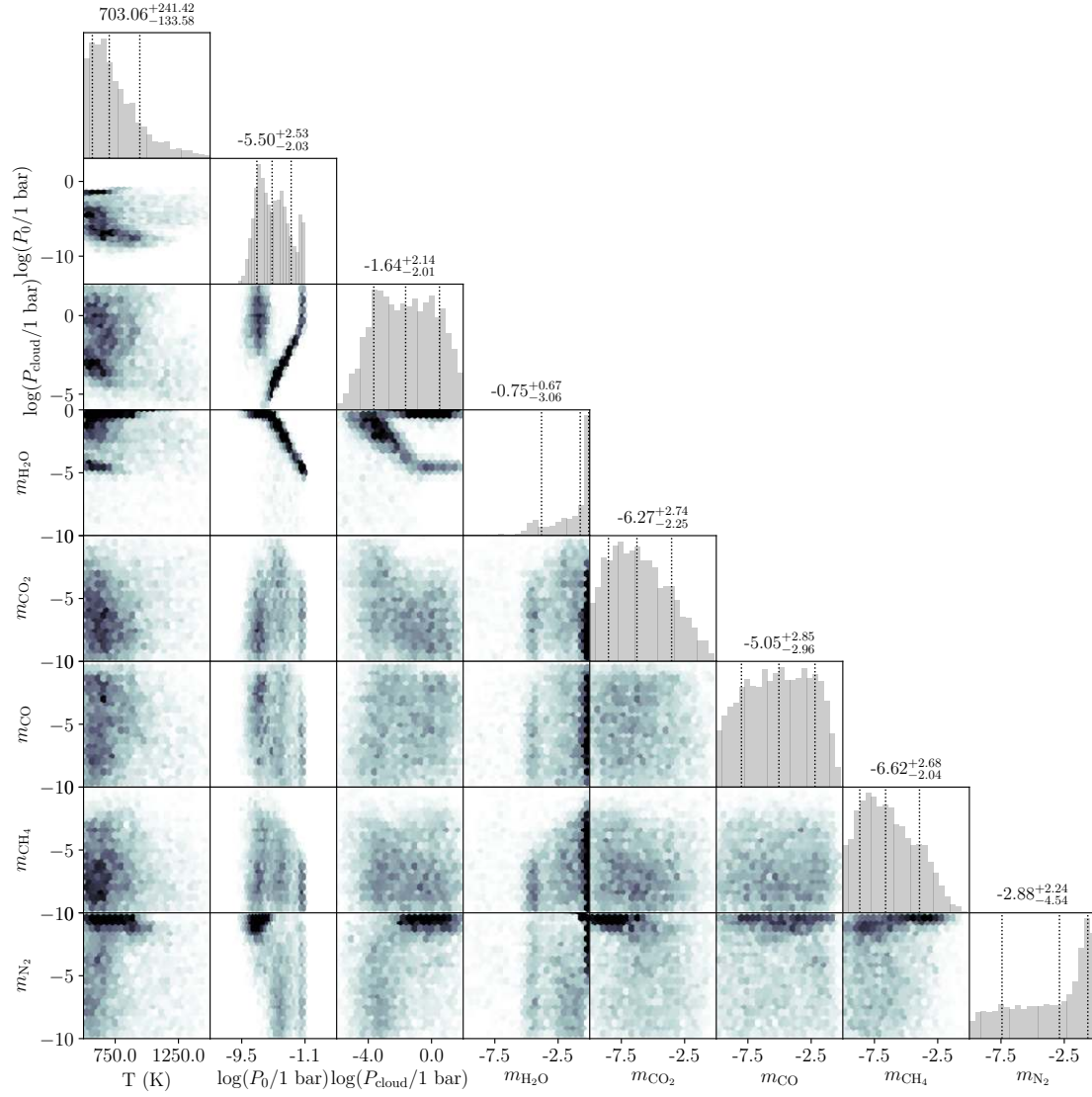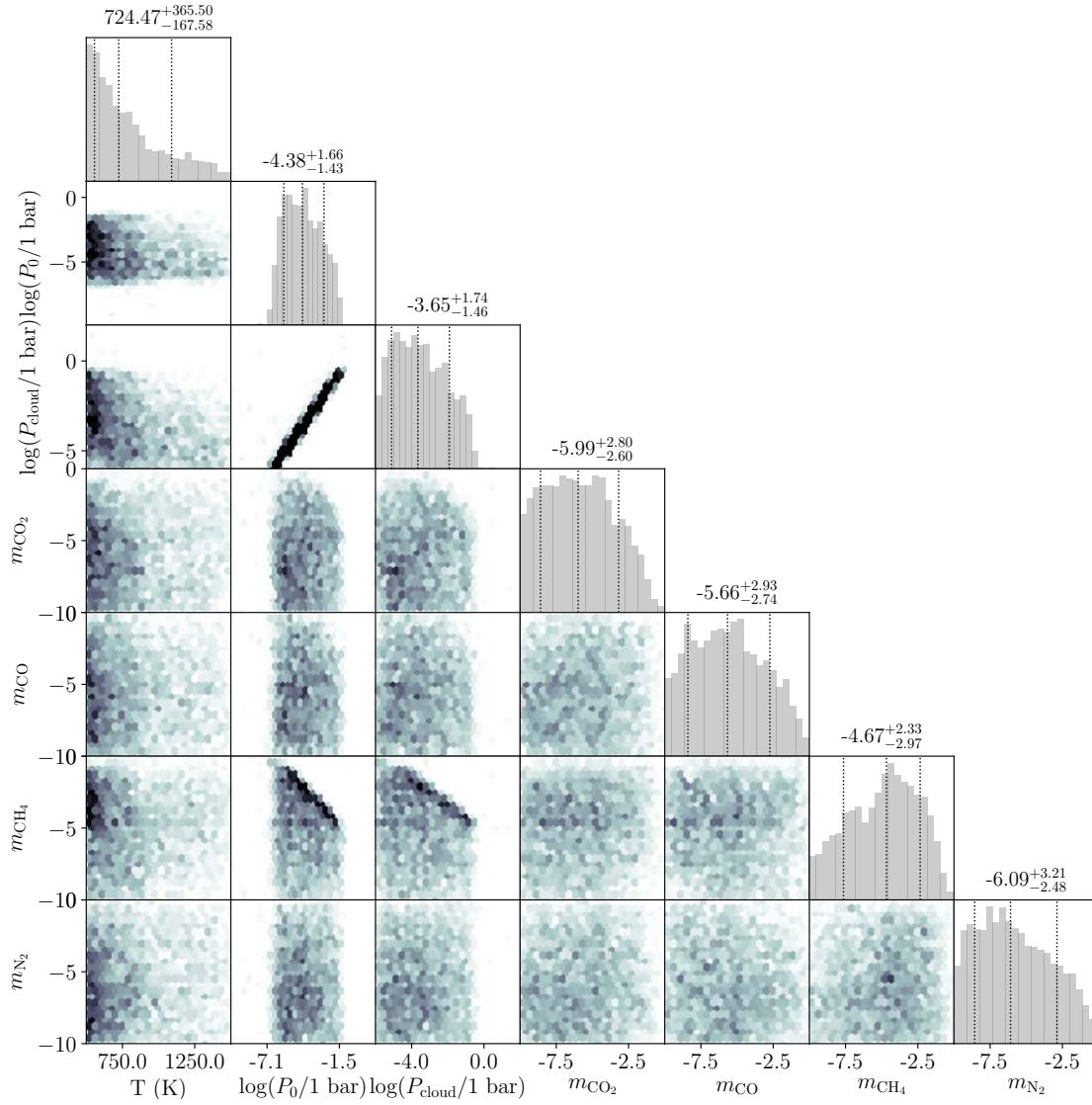
Retrieval without H$_2$O - Nominal spectrum

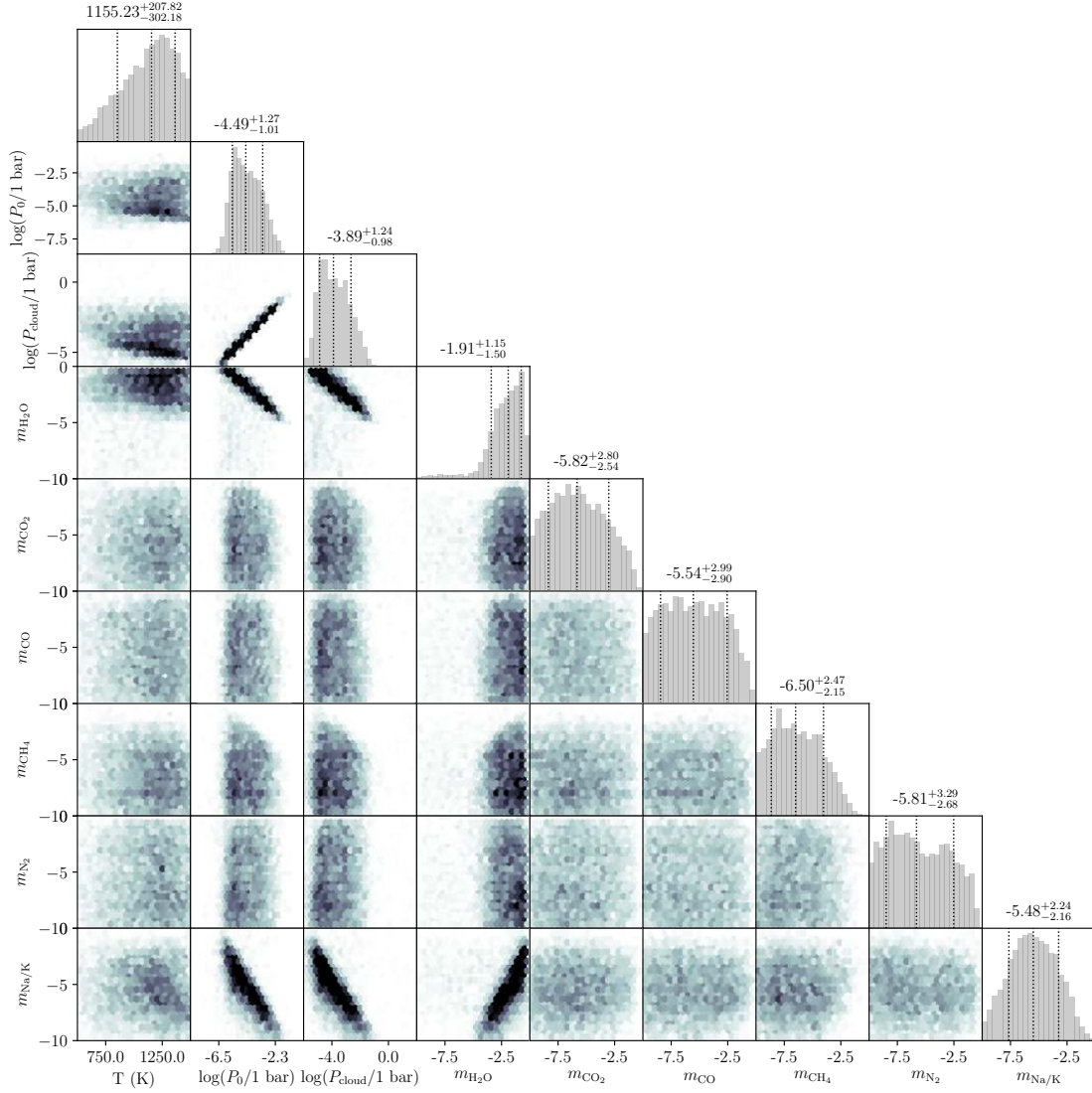**Fig. B.2.** Posterior probability distributions of the retrieval on HST/WFC3 dataset for WASP-117b, reduced with the nominal pipeline. The retrieval without H$_2$O was applied (see Sect. 3.1). Median parameter values and 68% confidence intervals for the marginalized 1D posterior probability distributions are indicated with horizontal error bars.

Full retrieval with H$_2$O - CASCADE spectrum

Fig. B.3. Posterior probability distributions of the retrieval on HST/WFC3 dataset for WASP-117b, reduced with the CASCADE pipeline. The full retrieval including H$_2$O was applied (see Sect. 3.1). Median parameter values and 68% confidence intervals for the marginalized 1D posterior probability distributions are indicated with horizontal error bars.

Retrieval without H$_2$O - CASCADE spectrum

Fig. B.4. Posterior probability distributions of the retrieval on HST/WFC3 dataset for WASP-117b, reduced with the nominal pipeline. The retrieval without H$_2$O was applied (see Sect. 3.1). Median parameter values and 68% confidence intervals for the marginalized 1D posterior probability distributions are indicated with horizontal error bars.
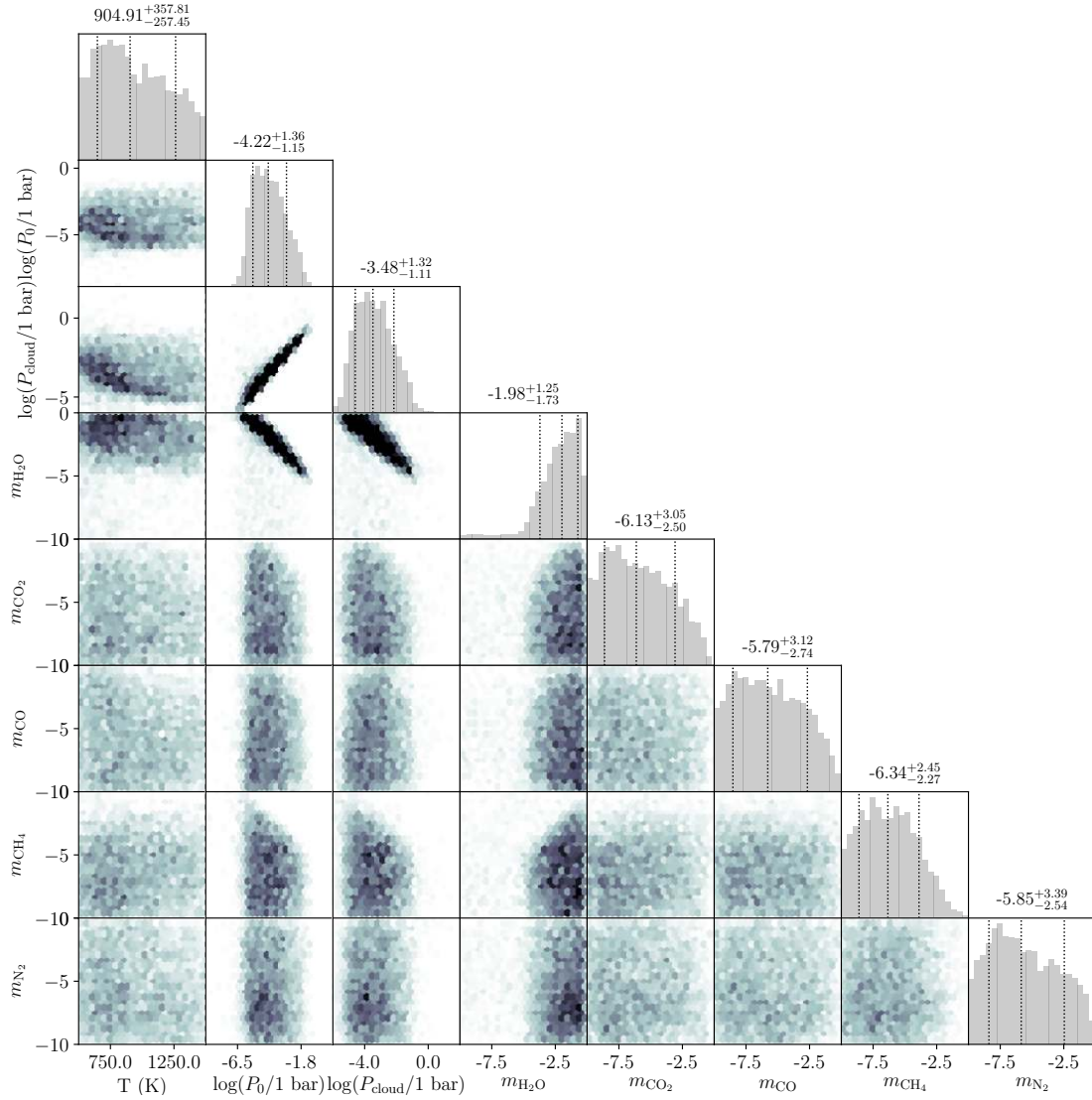
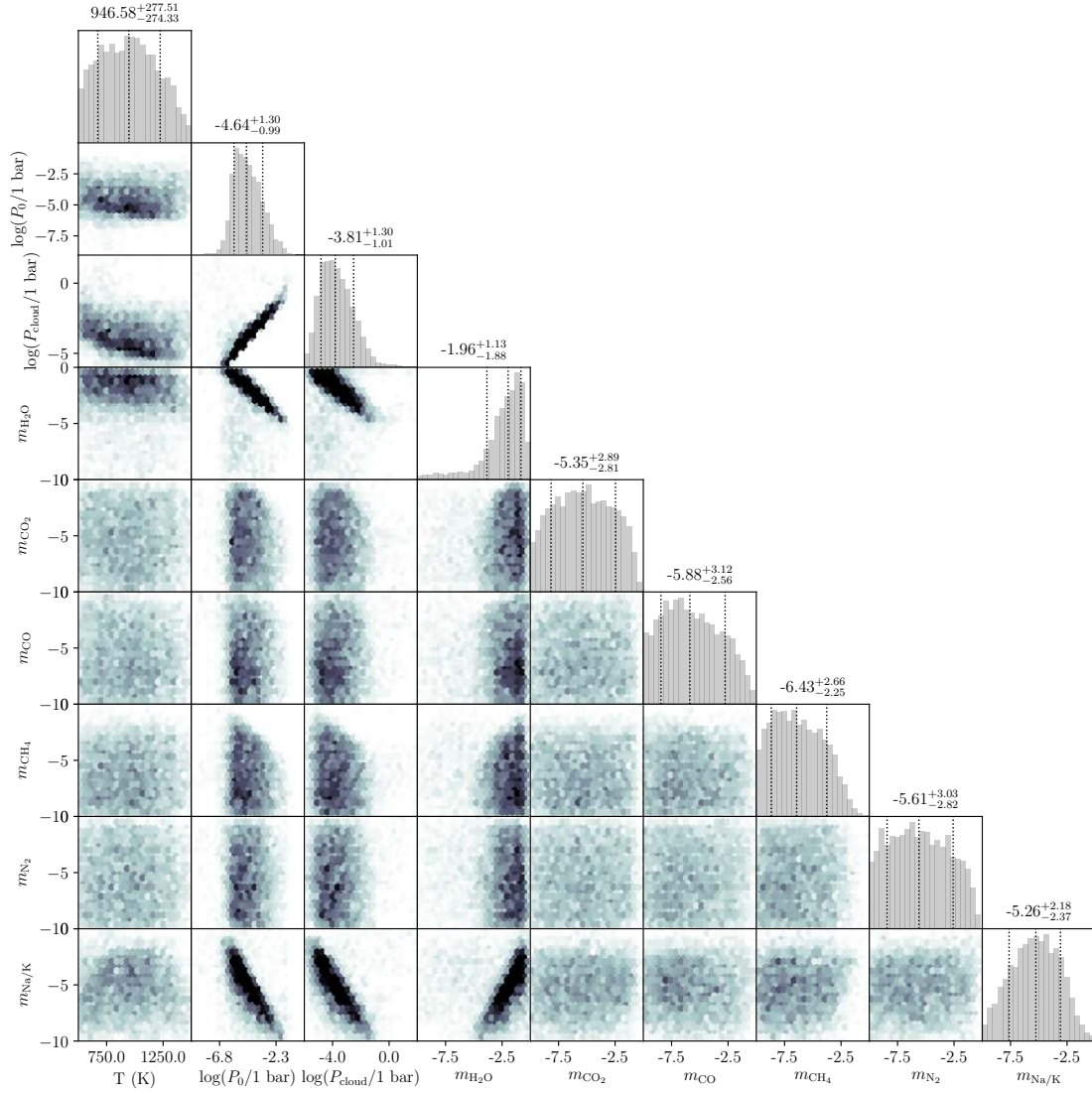**Full retrieval with Na/K - Nominal spectrum + VLT/ESPRESSO**

Fig. B.5. Posterior probability distributions of the joint analysis of the HST and VLT ESPRESSO dataset for WASP-117b, reduced with the nominal pipeline. The full retrieval with Na and K was applied (see Sect. 3.1). Median parameter values and 68% confidence intervals for the marginalized 1D posterior probability distributions are indicated with horizontal error bars.

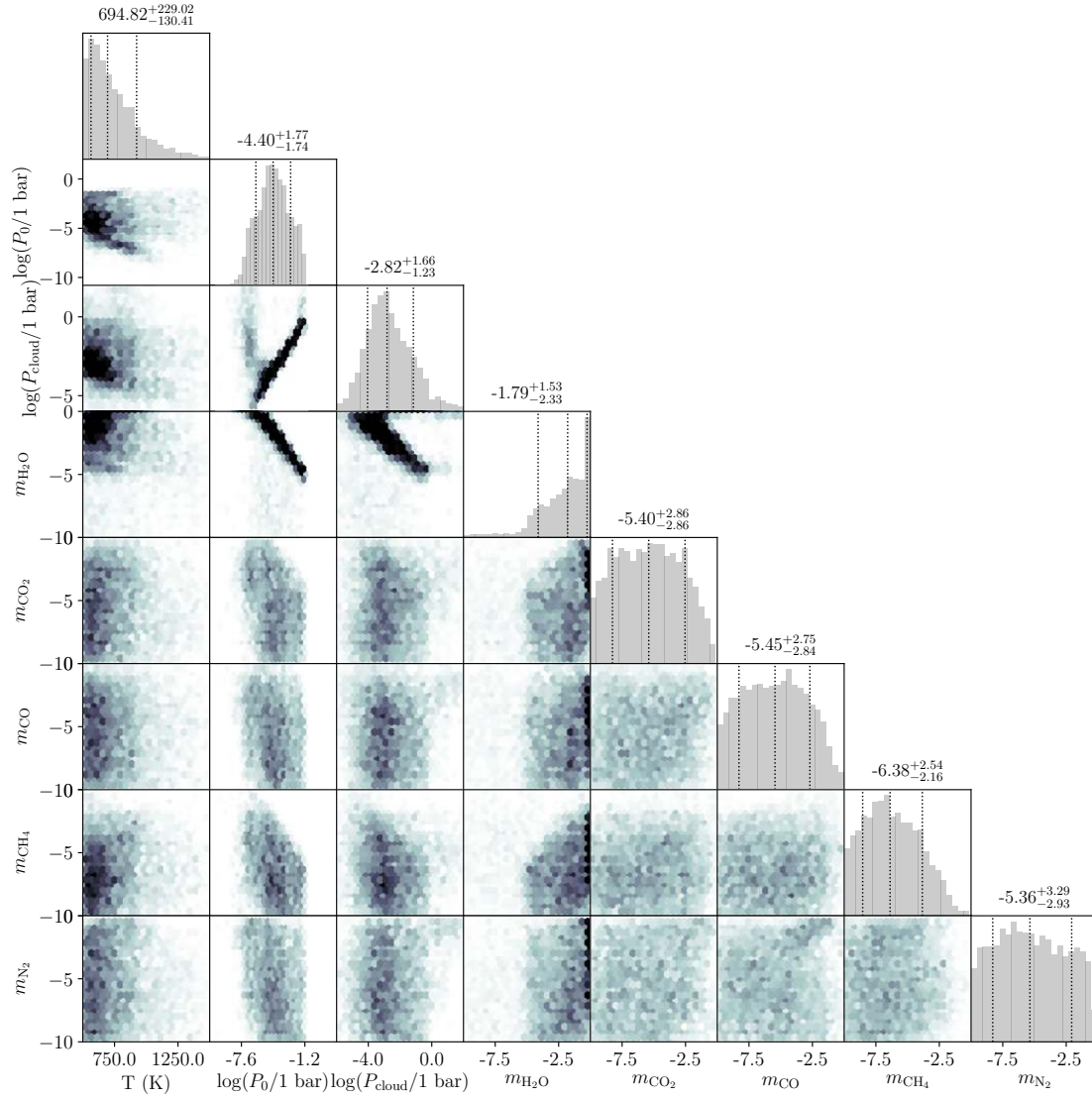**Retrieval without Na/K - Nominal spectrum + VLT/ESPRESSO**

Fig. B.6. Posterior probability distributions of the joint analysis of the HST and VLT ESPRESSO dataset for WASP-117b, reduced with the nominal pipeline. The retrieval without Na and K was applied (see Sect. 3.1). Median parameter values and 68% confidence intervals for the marginalized 1D posterior probability distributions are indicated with horizontal error bars.

Full retrieval with Na/K - CASCADE spectrum + VLT/ESPRESSO

Fig. B.7. Posterior probability distributions of the joint analysis of the HST and VLT ESPRESSO dataset for WASP-117b, reduced with the CASCADE pipeline. The full retrieval with Na and K was applied (see Sect. 3.1). Median parameter values and 68% confidence intervals for the marginalized 1D posterior probability distributions are indicated with horizontal error bars.

Retrieval without Na/K - CASCADE spectrum + VLT/ESPRESSO

Fig. B.8. Posterior probability distributions of the joint analysis of the HST and VLT ESPRESSO dataset for WASP-117b, reduced with the CASCADE pipeline. The retrieval without Na and K was applied (see Sect. 3.1). Median parameter values and 68% confidence intervals for the marginalized 1D posterior probability distributions are indicated with horizontal error bars.

Model 1 - Nominal spectrum
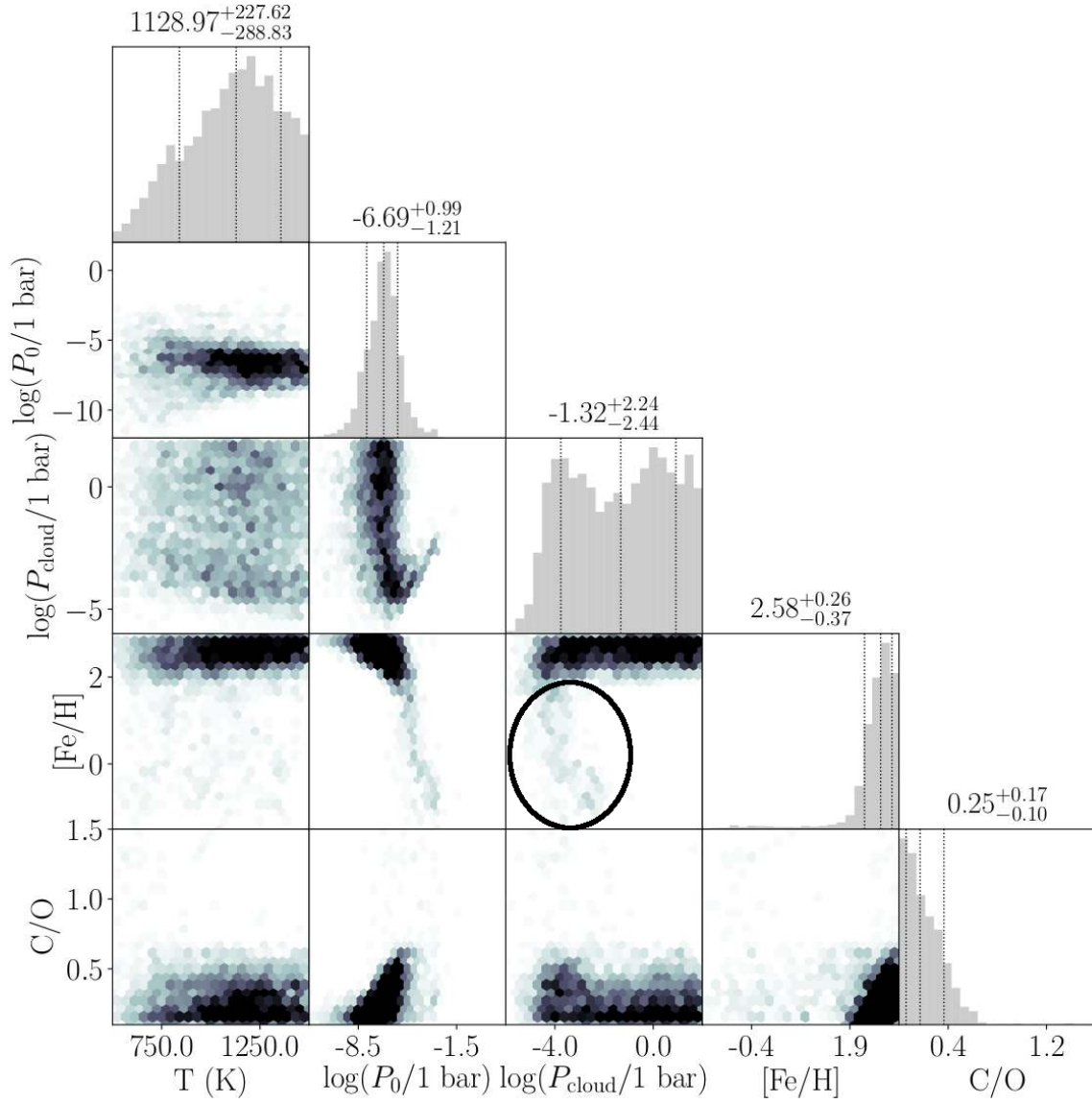
Fig. B.9. Posterior probability distributions for the HST dataset, reduced with the nominal pipeline. Model 1 was applied with unconstrained [Fe/H] and condensate clouds. Median parameter values and 68% confidence intervals for the marginalized 1D posterior probability distributions are indicated with horizontal error bars. The black ellipse highlights the subset of cloudy low-metallicity solutions, which are further explored with Model 2 and Model 3.
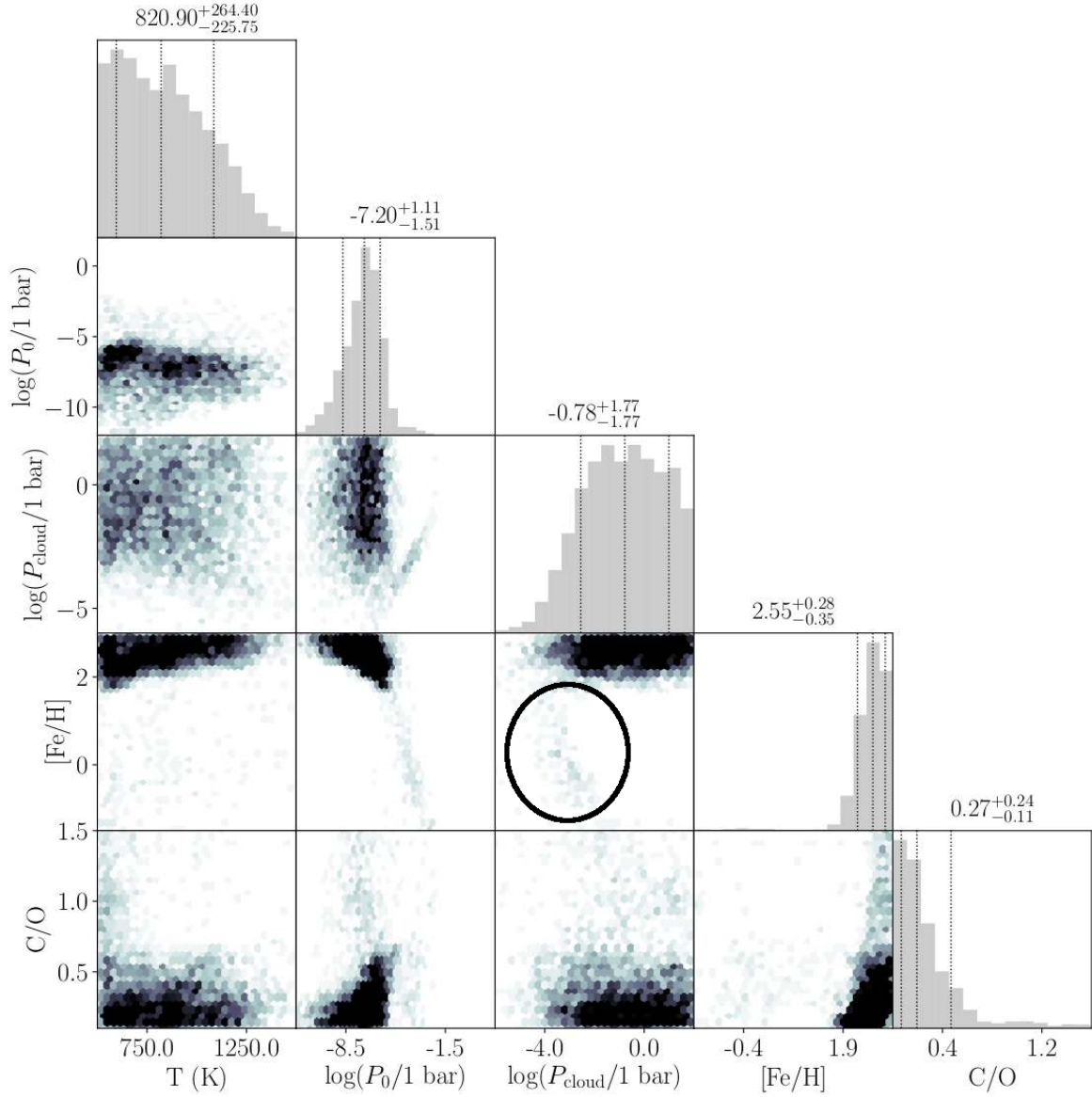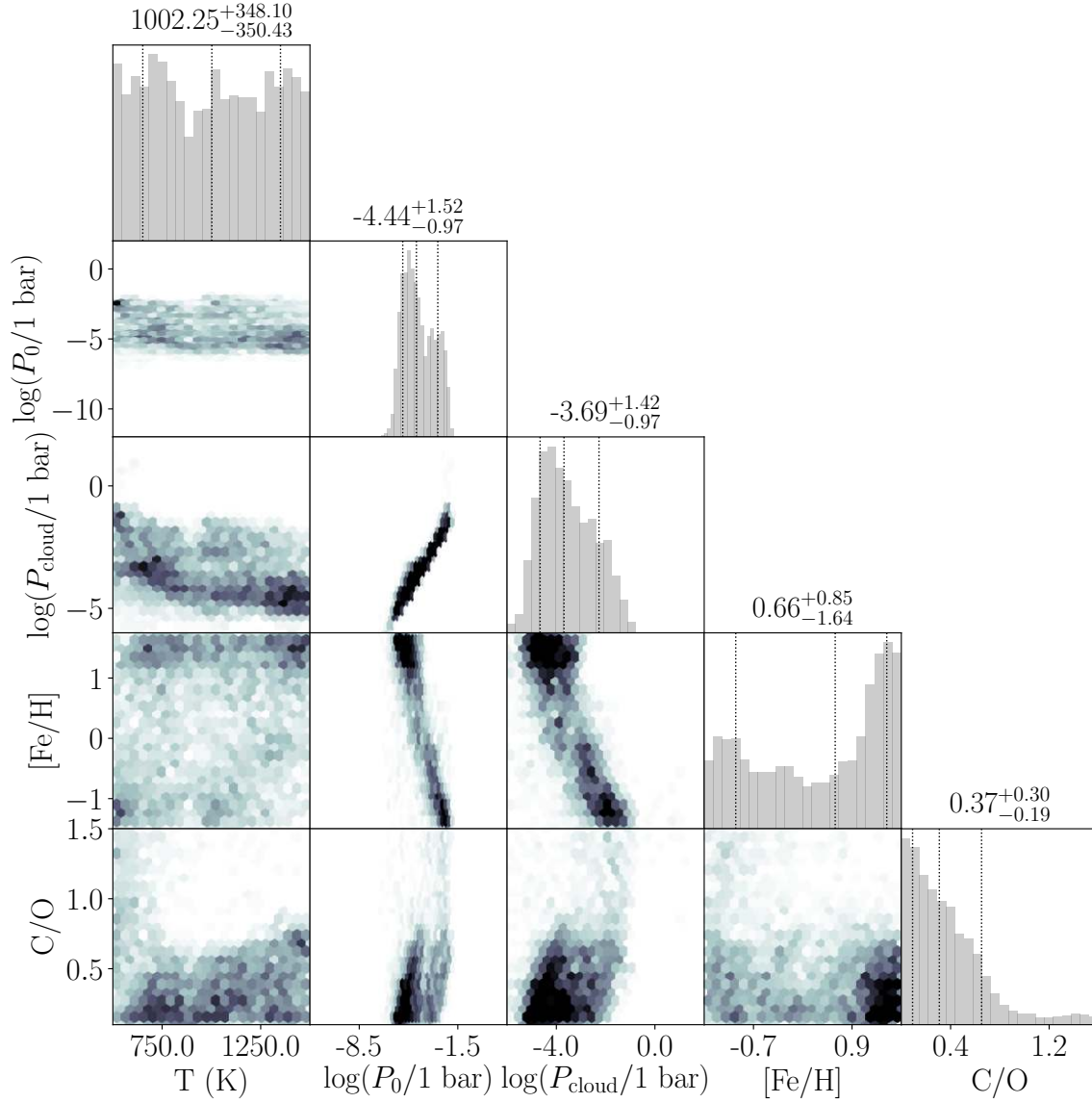
Model 1 - CASCADE spectrum

Fig. B.10. Posterior probability distributions for the HST dataset, reduced with the CASCADE pipeline. Model 1 was applied with unconstrained [Fe/H] and condensate clouds. Median parameter values and 68% confidence intervals for the marginalized 1D posterior probability distributions are indicated with horizontal error bars. The black ellipse highlights the subset of cloudy low metallicity solutions, which are further explored with Model 2 and Model 3.

Model 2 - Nominal spectrum

Fig. B.11. Posterior probability distributions for the HST dataset, reduced with the nominal pipeline. Model 2 was applied with prior [Fe/H] < 1.75 and condensate clouds. Median parameter values and 68% confidence intervals for the marginalized 1D posterior probability distributions are indicated with horizontal error bars.
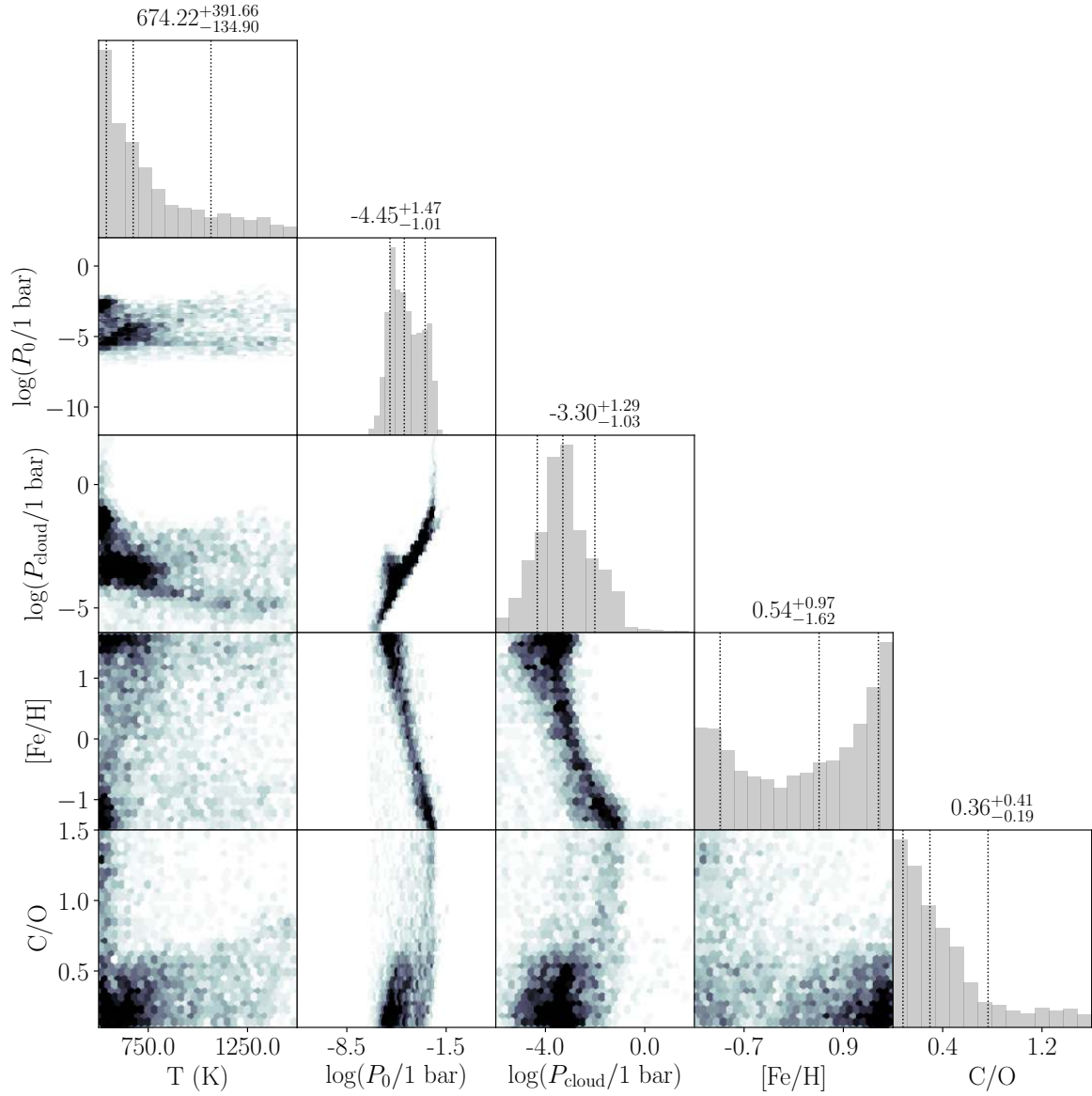
Model 2 - CASCADE spectrum

Fig. B.12. Posterior probability distributions for the HST dataset, reduced with the CASCADE pipeline. Model 2 was applied with prior [Fe/H] < 1.75 and condensate clouds. Median parameter values and 68% confidence intervals for the marginalized 1D posterior probability distributions are indicated with horizontal error bars.
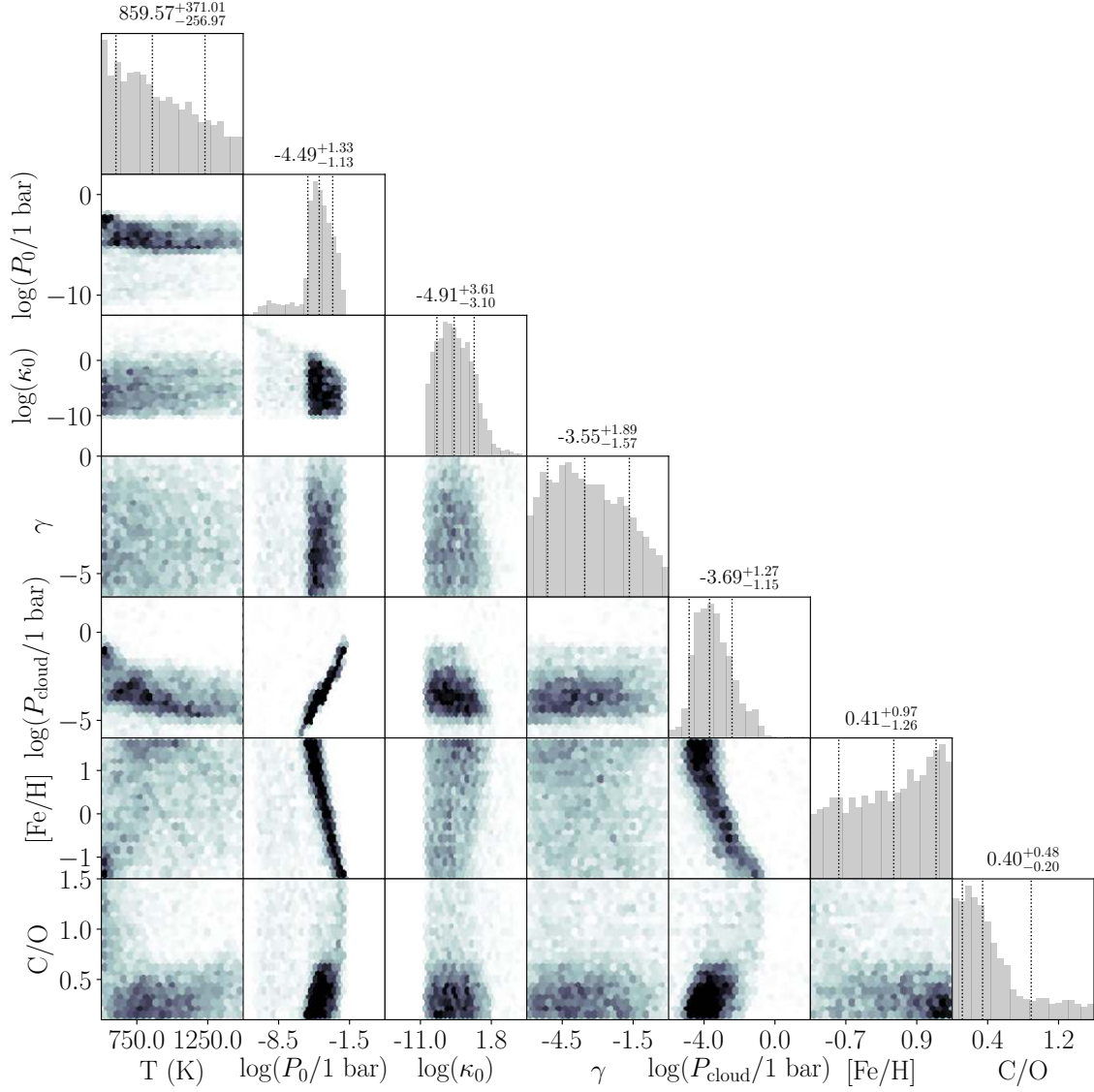
Model 3 - Nominal spectrum

Fig. B.13. Posterior probability distributions for the HST dataset, reduced with the nominal pipeline. Model 3 was applied with prior [Fe/H] < 1.75, condensate clouds, and a haze layer with a prior of $\gamma > -6$ on the scattering slope. Median parameter values and 68% confidence intervals for the marginalized 1D posterior probability distributions are indicated with horizontal error bars.
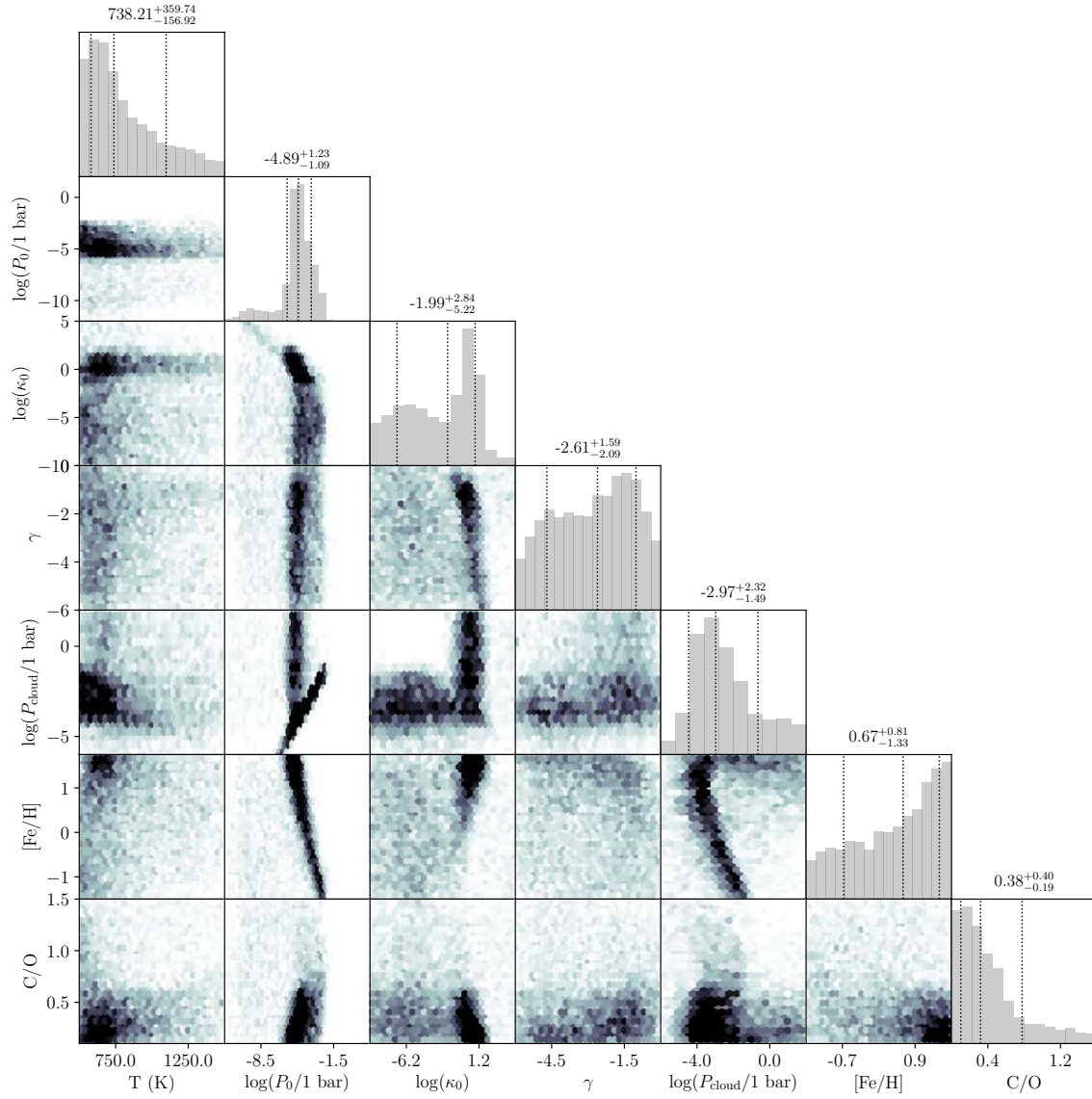
Model 3 - CASCADE spectrum

Fig. B.14. Posterior probability distributions of atmosphere properties of WASP-117b derived from WASP-117b HST observations. Raw data were reduced with the CASCADE pipeline. Model 3 was applied with prior [Fe/H] < 1.75, condensate clouds, and a haze layer with a prior of $\gamma > -6$ on the scattering slope. Median parameter values and 68% confidence intervals for the marginalized 1D posterior probability distributions are indicated with horizontal error bars.
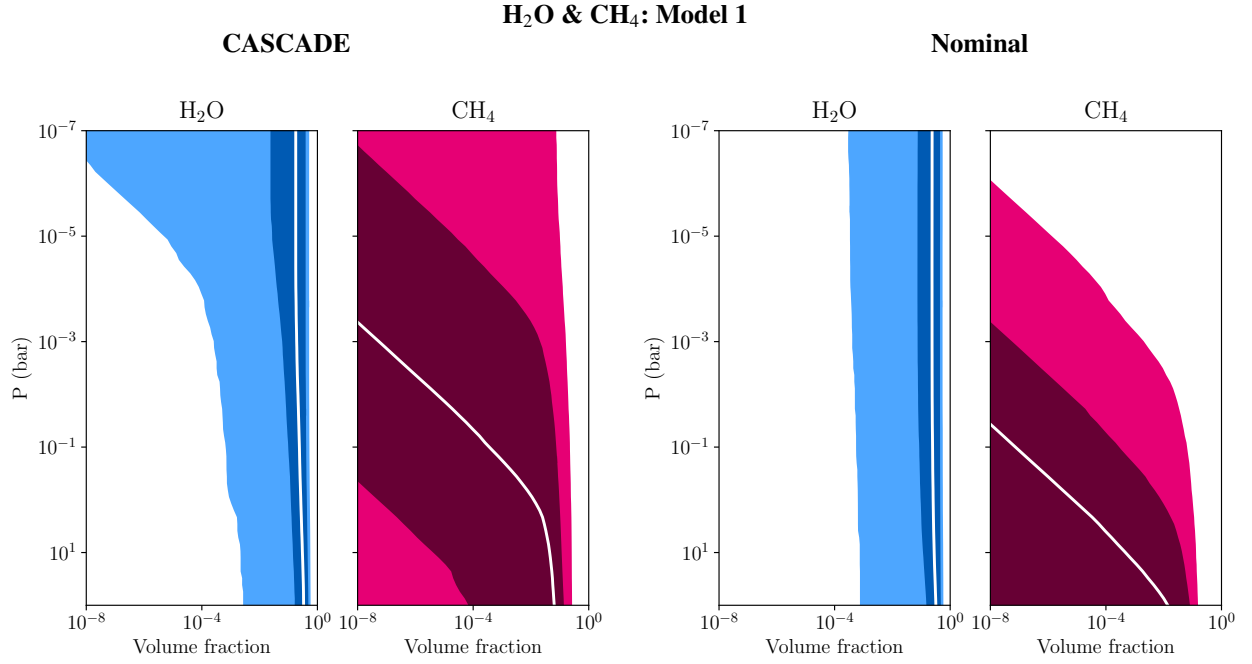
H$_2$O & CH$_4$: Model 1

CASCADE **Nominal**

Fig. B.15. Posterior probability distributions of H$_2$O and CH$_4$ content in the atmosphere of WASP-117b based on HST observations, reduced with the CASCADE (*left*) and nominal pipeline (*right*). Model 1 was applied, with condensate clouds. Median parameter values and 99.7% (light blue and red) and 68% (dark blue and red) confidence intervals are given.
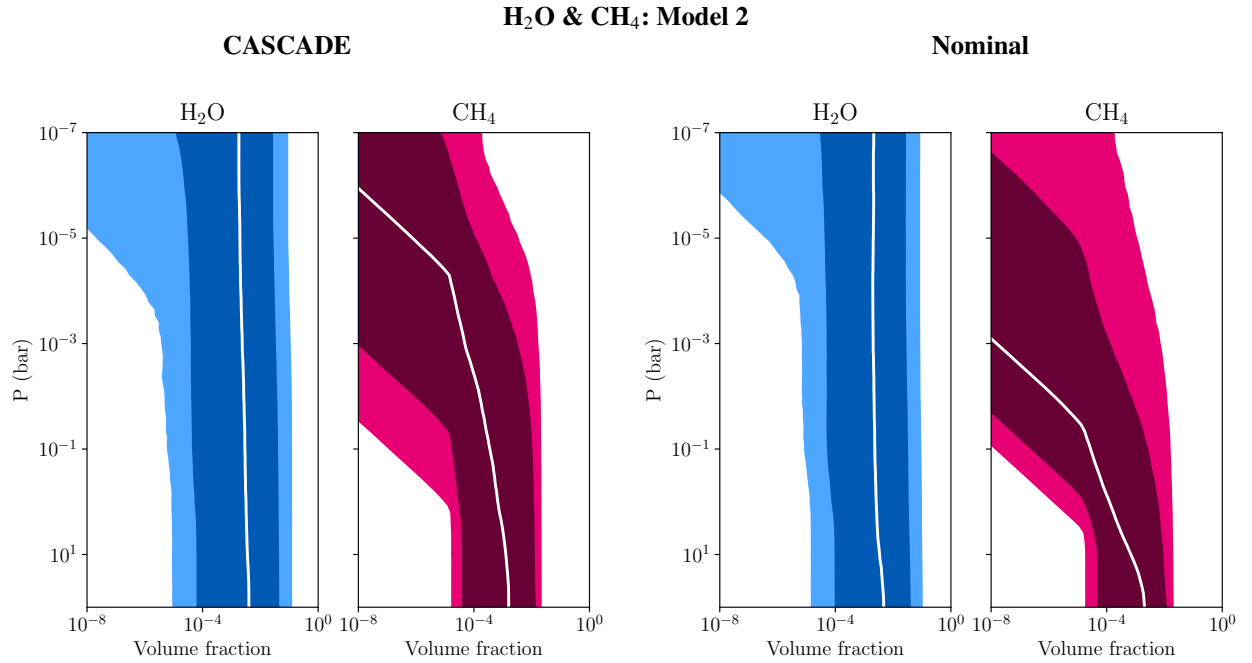
H$_2$O & CH$_4$: Model 2

CASCADE **Nominal**

Fig. B.16. Posterior probability distributions of H$_2$O and CH$_4$ content in the atmosphere of WASP-117b based on HST observations, reduced with the CASCADE (*left*) and nominal pipeline (*right*). Model 2 was applied, with condensate clouds and [Fe/H] < 1.75. Median parameter values and 99.7% (light blue and red) and 68% (dark blue and red) confidence intervals are given.

H$_2$O & CH$_4$: Model 3
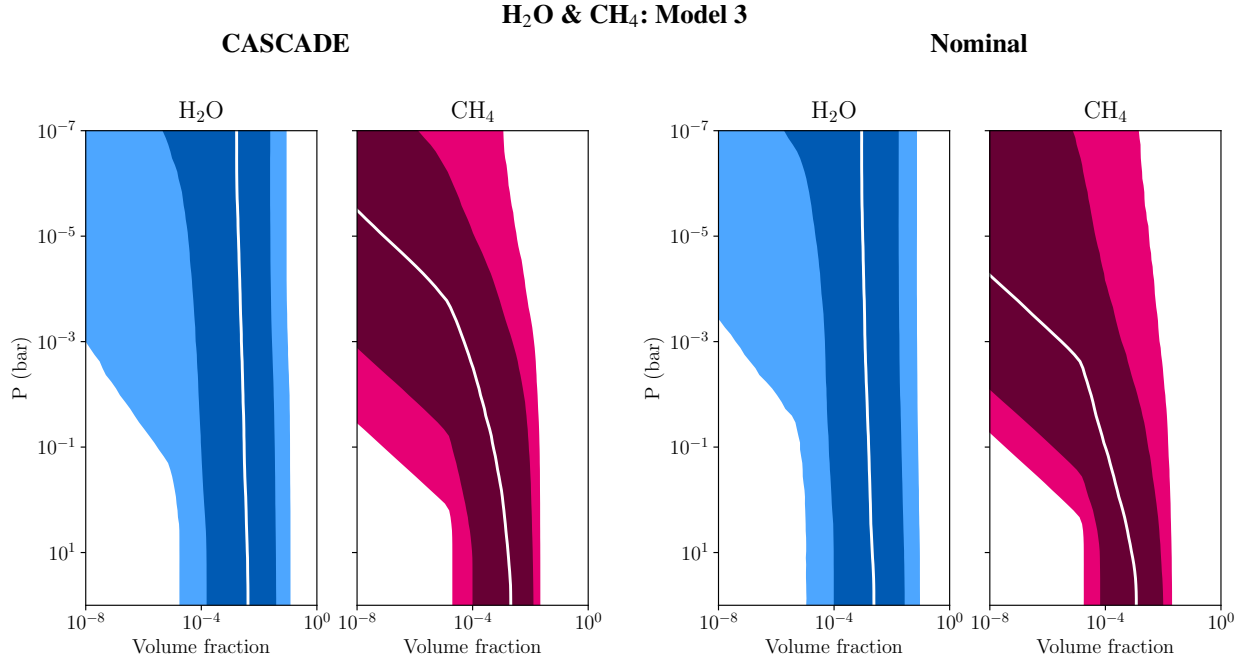

Fig. B.17. Posterior probability distributions of H$_2$O and CH$_4$ content in the atmosphere of WASP-117b based on HST observations, reduced with the CASCADE (*left*) and nominal pipeline (*right*). Model 3 was applied, with condensate clouds and [Fe/H] < 1.75 and a haze layer with scattering slope $\gamma < 3$. Median parameter values and 99.7% (light blue and red) and 68% (dark blue and red) confidence intervals are given.

Appendix C: Effect of vertical mixing on the methane abundance

To constrain the amount of mixing needed to reproduce the retrieved methane abundance, we performed additional simulations for the cool equilibrium temperature (T_{eq} = 700 K) and low-metallicity (5× solar) setup, with varying eddy diffusion coefficients: K_{zz} = 10^6, 10^8, 10^9, 10^{10}, and 10^{12} cm^2 s^{-1} (see

Fig. C.1). Here, chemical equilibrium considerations would indicate that the atmosphere of WASP-117b is dominated by H$_2$O and CH$_4$ above ~100 mbar. This is still the case for eddy diffusion coefficients up to ~10^6 cm^2 s^{-1}. For higher values of K_{zz}, 10^8 cm^2 s^{-1} and above, CH$_4$ becomes quenched at pressures of ~1 bar and deeper. For these strongly mixed cases, the dominant atmospheric molecules are H$_2$O and CO throughout the atmosphere.
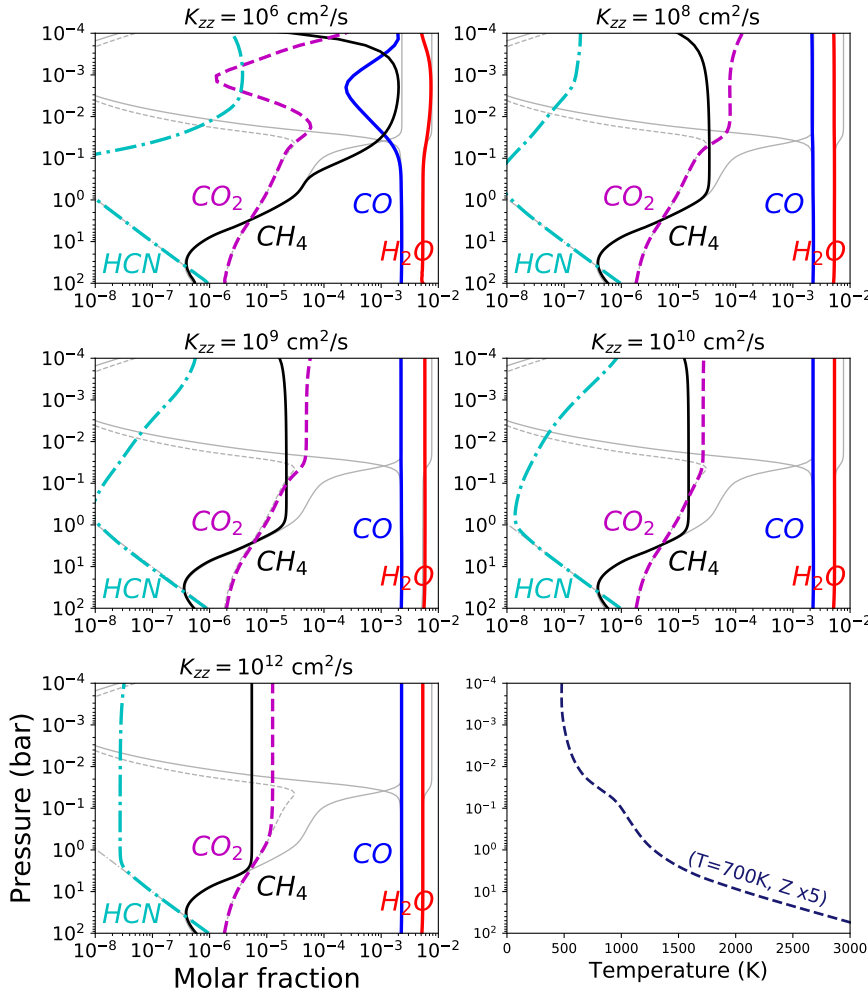
Fig. C.1. Disequilibrium abundances of the dominant molecules in the atmosphere of WASP-117b computed for different vertical eddy coefficients K_{zz} with the chemical kinetics code of Venot et al. (2012). The gray lines indicate the chemical equilibrium state. The *bottom right plot* shows the one-dimensional pressure-temperature profile of WASP-117b, calculated with petitCODE (Mollière et al. 2015, 2017). The model setup is based on the warm, slightly enriched scenario for WASP-117b (T_{eq} = 700 K, Z = 5× solar, C/O = 0.3).

Appendix D: HST transmission spectra

Table D.1. WASP-117b transmission spectra derived with the nominal pipeline.

Wavelength	Transit depth	Uncertainty
1.137955	0.00747861	4.312E-05
1.165961	0.00768836	4.304E-05
1.193967	0.00753981	4.226E-05
1.221973	0.00756615	4.110E-05
1.249979	0.00750309	4.087E-05
1.277985	0.00766432	4.064E-05
1.305991	0.00742532	4.063E-05
1.333997	0.00763581	3.978E-05
1.362003	0.00753166	3.996E-05
1.390009	0.00766481	4.177E-05
1.418015	0.00765191	4.198E-05
1.446021	0.00761142	4.215E-05
1.474027	0.00761617	4.229E-05
1.502033	0.00763603	4.401E-05
1.530039	0.00759753	4.515E-05
1.558045	0.00757245	4.706E-05
1.586051	0.00744342	4.838E-05
1.614057	0.00758623	4.878E-05
1.642063	0.00747719	5.2362E-05

Table D.2. WASP-117b transmission spectra derived with the CAS-CADE pipeline

Wavelength	Transit depth	Uncertainty
1.137955	0.00762161	3.305E-05
1.165961	0.00760239	3.566E-05
1.193967	0.00759120	3.623E-05
1.221973	0.00751966	3.202E-05
1.249979	0.00747235	3.511E-05
1.277985	0.00753574	3.619E-05
1.305991	0.00747213	3.311E-05
1.333997	0.00753405	3.352E-05
1.362003	0.00759060	3.677E-05
1.390009	0.00759895	3.262E-05
1.418015	0.00755960	4.392E-05
1.446021	0.00756905	3.370E-05
1.474027	0.00760003	3.245E-05
1.502033	0.00762398	3.702E-05
1.530039	0.00750796	3.711E-05
1.558045	0.00751503	3.558E-05
1.586051	0.00754406	3.703E-05
1.614057	0.00750870	3.696E-05
1.642063	0.00752046	4.042E-05