Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **521,** 1233–1252 (2023) https://doi.org/10.1093/mnras/stad480
Advance Access publication 2023 February 14

The hot Neptune WASP-166 b with ESPRESSO – III. A blue-shifted tentative water signal constrains the presence of clouds

M. Lafarga [1,2★†] M. Brogi,[1,3,4] S. Gandhi [1,2,5] H. M. Cegla,[1,2] J. V. Seidel,[6] L. Doyle,[1,2] R. Allart,[7] N. Buchschacher,[8] M. Lendl [8] C. Lovis[8] and D. Sosnowska[8]

[1]*Department of Physics, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK*
[2]*Centre for Exoplanets and Habitability, University of Warwick, Coventry CV4 7AL, UK*
[3]*INAF – Osservatorio Astrofisico di Torino, Via Osservatorio 20, I-10025 Pino Torinese, Italy*
[4]*Dipartimento di Fisica, Università degli Studi di Torino, via Pietro Giuria 1, I-10125 Torino, Italy*
[5]*Leiden Observatory, Leiden University, Postbus 9513, NL-2300 RA Leiden, the Netherlands*
[6]*European Southern Observatory, Alonso de Córdova 3107, Vitacura, Región Metropolitana, Chile*
[7]*Department of Physics, and Trottier Institute for Research on Exoplanets, Université de Montréal, Montréal H3T 1J4, Canada*
[8]*Observatoire Astronomique de l'Université de Genève, Chemin Pegasi 51b, CH-1290 Versoix, Switzerland*

Accepted 2023 February 8. Received 2023 January 31; in original form 2022 November 1

ABSTRACT

With high-resolution spectroscopy, we can study exoplanet atmospheres and learn about their chemical composition, temperature profiles, and presence of clouds and winds, mainly in hot, giant planets. State-of-the-art instrumentation is pushing these studies towards smaller exoplanets. Of special interest are the few planets in the 'Neptune desert', a lack of Neptune-sized planets in close orbits around their hosts. Here, we assess the presence of water in one such planet, the bloated super-Neptune WASP-166 b, which orbits an F9-type star in a short orbit of 5.4 d. Despite its close-in orbit, WASP-166 b preserved its atmosphere, making it a benchmark target for exoplanet atmosphere studies in the desert. We analyse two transits observed in the visible with ESPRESSO. We clean the spectra from the Earth's telluric absorption via principal component analysis, which is crucial to the search for water in exoplanets. We use a cross-correlation-to-likelihood mapping to simultaneously estimate limits on the abundance of water and the altitude of a cloud layer, which points towards a low water abundance and/or high clouds. We tentatively detect a water signal blue-shifted $\sim$5 km s^{-1} from the planetary rest frame. Injection and retrieval of model spectra show that a solar-composition, cloud-free atmosphere would be detected at high significance. This is only possible in the visible due to the capabilities of ESPRESSO and the collecting power of the VLT. This work provides further insight on the Neptune desert planet WASP-166 b, which will be observed with *JWST*.

Key words: instrumentation: spectrographs – methods: observational – techniques: spectroscopic – planets and satellites: atmospheres – planets and satellites: individual: WASP-166 b – exoplanets.

1 INTRODUCTION

High-resolution spectroscopy is used to detect and characterize the atmospheres of transiting planets, giving us information about their chemical composition, temperature profiles, and the presence of clouds and winds, mainly in hot, giant planets (see e.g. Birkby 2018, for a review). State-of-the-art instrumentation is pushing the precision of our measurements towards the detection and characterization of the atmospheres of cooler and smaller exoplanets (Neptune- and Earth-sized planets). One of the best-studied chemical species with high-resolution instruments is water. Water analyses have mainly been focused in the infrared wavelength range, because its spectrum presents several strong absorption bands, while in the optical range, there are only few weaker absorption bands in the red. Water vapour has been targeted by several ground-based, high-resolution infrared spectrographs such as CRIRES (Kaeufl et al. 2004), NIRSPEC (McLean et al. 1998), GIANO (Origlia et al. 2014), CARMENES NIR (Quirrenbach et al. 2016), and SPIRou (Donati et al. 2020). Observations with these instruments have led to the detection of water vapour in the atmospheres of several transiting and non-transiting exoplanets (e.g. Birkby et al. 2013, 2017; Brogi et al. 2014, 2016, 2018; Alonso-Floriano et al. 2019; Sánchez-López et al. 2019; Webb et al. 2020, 2022; Boucher et al. 2021). However, detections of water in the visible range remain challenging.

Esteves et al. (2017) and and Jindal et al. (2020) studied the presence of water in the super-Earth 55 Cancri e with several transits obtained with the optical, high-resolution spectrographs HDS (wavelength range 5240–7890 Å, Noguchi et al. 2002) on the 8.2-m Subaru telescope, ESPaDOnS (5060–7950 Å, Donati 2003) on the 3.6-m CFHT, and GRACES (3990–10 480 Å, Chene et al. 2014) on the 8.1-m Gemini North telescope. They did not detect water and ruled out the presence of water-rich atmospheres if cloud-free.

★ E-mail: marina.lafarga-magro@warwick.ac.uk
† UKRI Future Leaders Fellow.

Deibert et al. (2019) studied HDS and GRACES observations of HAT-P-12 b and WASP-69 b, two warm sub-Saturns with inflated radii. They also did not detect water, but injection tests suggest a cloudy atmosphere with a small amount of absorption, in agreement with other studies.

Allart et al. (2017) used HARPS (3780–6910 Å, Mayor et al. 2003) on ESO's La Silla 3.6-m telescope to look for water in the gas giant HD 189 733 b, focusing on the 6500 Å absorption band. The data used is too noisy to constrain the presence of water and the authors estimated that over 10 HARPS transits would be needed to have a constrain at a significant level. However, they also estimated that a significant detection would be feasible with only a single ESPRESSO transit, due to its increased collecting power and the fact that its wavelength range includes a stronger water band at ∼7400 Å.

With ESPRESSO (3782–7887 Å, Pepe et al. 2021) on ESO's 8.2-m VLT, Allart et al. (2020) studied WASP-127 b, a super-Neptune-mass planet with a radius larger than that of Jupiter, which makes it an extremely bloated planet. No water was found but, together with low-resolution data, the authors were able to constrain the pressure of a cloud-deck. Also with ESPRESSO, Sedaghati et al. (2021) observed the hot Jupiter WASP-19 b, which orbits a G8 V star in less than 1 d. Water was again not detected, but in this case, injection tests showed that it would only be detectable at high abundances and not feasible with ESPRESSO on an 8-m class telescope. The authors argued that this is the case due to the relatively faint host star ($V = 12.3$ mag) and the short transit duration (1.6 h, Cortés-Zuleta et al. 2020), which result in few in-transit observations with low signal-to-noise ratio (S/N).

Finally, Sánchez-López et al. (2020) reported a water detection in one out of three transits of HD 209 458 b using the 7000–9600 Å absorption bands present in the red part of the visible arm of CARMENES VIS (5200–9600 Å, Quirrenbach et al. 2016). Injection tests indicated that the lack of detection in the other two nights could be due to a lower S/N and a higher degree of telluric variability, which results in a worse telluric removal that hinders the detection of water.

As seen from the previous results, an important feature observed in the atmospheres of both hot and cool planets is the presence of clouds and/or hazes. Clouds and hazes reduce the strength of the features observed in an exoplanet spectrum, affecting the detectability of species such as water. Low-resolution observations of planets over a range of temperatures have shown muted water spectral features compared to what is expected for cloud-free atmospheres with solar metallicity (e.g. Sing et al. 2016; Stevenson et al. 2016; Barstow et al. 2017; Wakeford et al. 2017a,b, 2019; Pinhas et al. 2019; Benneke et al. 2019a,b; Kreidberg et al. 2020), and some low-mass, low-temperature planets even show completely featureless spectra (Kreidberg et al. 2014; Knutson et al. 2014a,b). These muted features can be attributed to either the presence of thick, high-altitude clouds, or to inherently low water abundances.

As opposed to low-resolution observations, which are sensitive to broad-band spectral features, high-resolution spectroscopy is able to resolve individual lines. The cores of absorption lines are formed higher up in the atmosphere than their wings. Therefore, high-resolution data is sensitive to high-altitude regions of the atmosphere and can probe above clouds.

The abundance of the species present in exoplanet atmospheres has been typically derived from low-resolution spectroscopy, which is sensitive to broad-band features over the continuum. Opposite to that, high-resolution observations do not preserve that continuum flux needed to measure abundances. However, it has recently been shown that by using a Bayesian framework, it is possible to recover abundances from the line-to-line and line-to-continuum contrast ratio alone (Brogi & Line 2019; Gibson et al. 2020; Line et al. 2021; Pelletier et al. 2021). Therefore, high-resolution spectroscopy is both sensitive to clouds and water abundance, and can break degeneracy between the two (Gandhi, Brogi & Webb 2020b; Hood et al. 2020).

Table 1. WASP-166 system properties used in this work.

Parameter	Value	Reference
$R_p/R_\star$	$0.05177^{+0.00063}_{-0.00035}$	Doyle et al. (2022)
R_p (R_J)	$0.6155^{+0.0306}_{-0.0307}$	Doyle et al. (2022)
$a/R_\star$	$11.83^{+0.29}_{-0.68}$	Doyle et al. (2022)
a (au)	$0.0668^{+0.0040}_{-0.0044}$	Doyle et al. (2022)
i_p (°)	$88.85^{+0.74}_{-0.94}$	Doyle et al. (2022)
t_0 (BJD)	$2458524.40869201^{+0.00030021}_{-0.00029559}$	Doyle et al. (2022)
T_{dur} (h)	$3.608^{+0.020}_{-0.015}$	Doyle et al. (2022)
P (d)	$5.44354215^{+0.00000307}_{-0.00000297}$	Doyle et al. (2022)
e	0.0	Hellier et al. (2019)
V_{sys} (km s^{-1})	23.532 ± 0.012	Doyle et al. (2022)
K_p (km s^{-1})	134.1 ± 8.1	This work
T_{eff} (K)	6050 ± 50	Hellier et al. (2019)
T_{eq} (K)	1270 ± 30	Hellier et al. (2019)

Note. Values from Doyle et al. (2022) have been derived using the same ESPRESSO observations as here, as well as *TESS* and NGTS photometry. In particular, V_{sys} has been measured from the out-of-transit cross-correlation functions of the ESPRESSO data and here we use the mean V_{sys} of the two nights (Doyle et al. 2022, see their table 1). K_p has been computed here based on the parameters from Doyle et al. (2022) (see Section 3.2 and Appendix B).

In this work, we use optical, high-resolution spectroscopy to study the presence of water and clouds on the transiting planet WASP-166 b, a bloated super-Neptune. WASP-166 b orbits a relatively bright ($V = 9.36$ mag), F9-type star in a close orbit of 5.4 d, at 0.06 au (Hellier et al. 2019; Bryant et al. 2020, see Table 1 for the system parameters adopted here). The planet has a mass of $0.101 \pm 0.005\ M_J$ (1.9 M_{Nep}) and a radius of $0.63 \pm 0.03\ R_J$ (1.8 R_{Nep}) (Hellier et al. 2019), and its orbit has been found to be aligned with the stellar spin (Hellier et al. 2019; Doyle et al. 2022; Kunovac Hodžić in preparation). It is located in the so-called 'Neptune desert', a dearth of Neptune-sized planets in close orbits around their host stars. The study of such planets can provide insight to their formation and evolution, and the existence of the desert.

Despite its close-in orbit, the planet has preserved its atmosphere, making it a benchmark target for exoplanet atmosphere studies in the Neptune desert. Seidel et al. (2020, 2022) recently confirmed the presence of sodium in the atmopshere of WASP-166 b with high-resolution ground-based transit observations obtained with the spectrographs HARPS (Mayor et al. 2003) and ESPRESSO (Pepe et al. 2021), respectively. In the optical, other than sodium, we also expect the presence of potassium (although its signature usually overlaps with strong absorption from telluric oxygen, challenging its detection) and water, which we study here (e.g. Fortney et al. 2008; Madhusudhan 2012; Moses et al. 2013; Woitke et al. 2018; Drummond et al. 2019). Other species with signatures in the optical such as CH_4 or NH_3 do not have reliable opacities below 0.5–1.0 micron, and hence have not been considered here. The planet is not hot enough to have other species with optical signatures such as Fe, TiO, or VO. WASP-166 is scheduled to be observed from space in the near-infrared with *JWST*, which should constrain the presence of molecules such as H_2O, CO, CH_4, CO_2, C_2H_2, HCN, and NH_3 in the planetary atmosphere (Mayo et al. 2021).

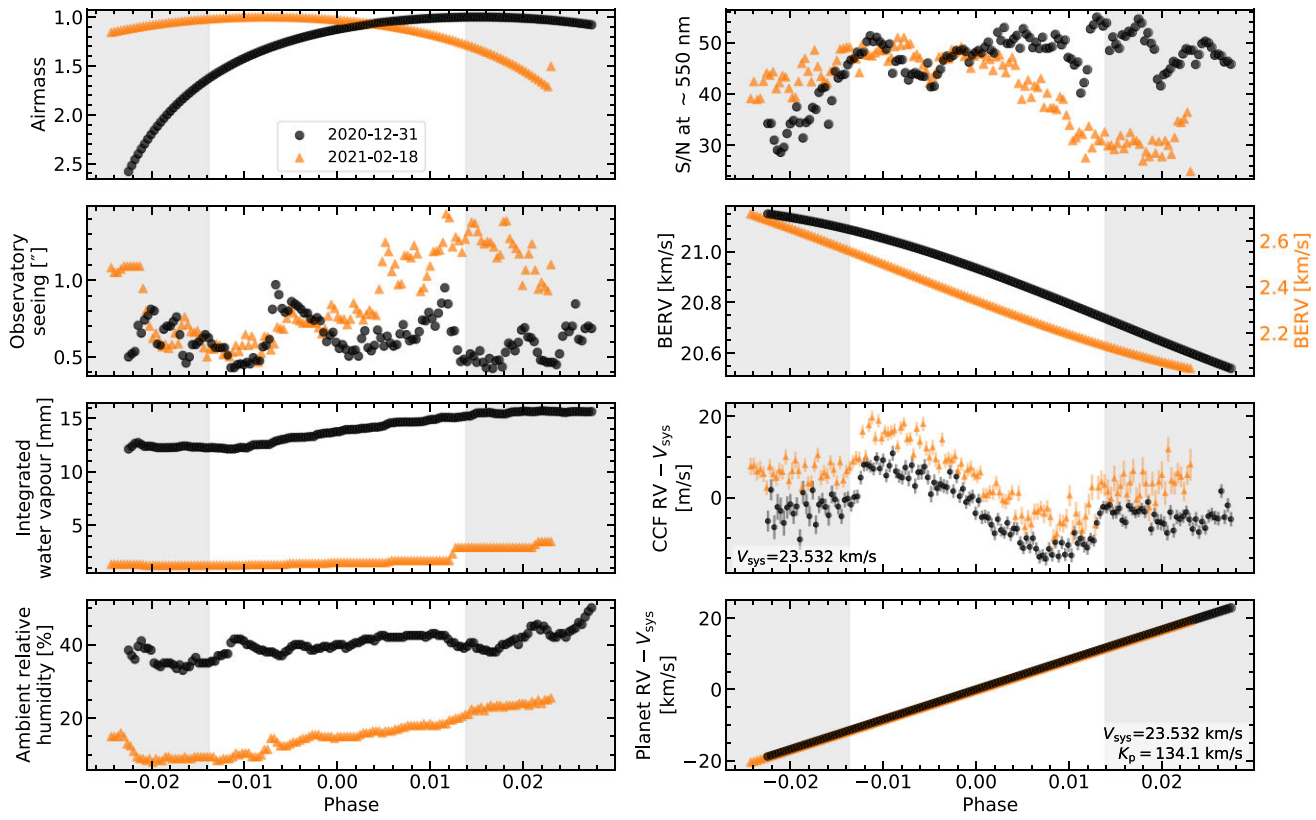

Figure 1. Observing conditions, S/N, and Earth, star, and planet RVs for the 2 transits observed, as a function of the planetary phase, where phase 0 corresponds to the mid-transit. Left, top to bottom panels: Airmass (mean between start and end of each observation), seeing (mean between start and end of each observation), integrated water vapour (mean between start and end of each observation), and ambient humidity. Right, top to bottom panels: S/N at ∼550 nm, barycentric Earth radial velocity (BERV, note the offset between nights), star RV from the DRS CCF, and planet RV. The V_{sys} of the system has been subtracted from the star and planet RVs (see Table 1). All parameters obtained from the observations FITS headers, except for the planet RV, which is computed from the orbital parameters of the system (see text). Grey areas indicate out-of-transit phases.

In Section 2, we describe the ESPRESSO observations used. Section 3 details the analysis performed, and in Section 4 we show and discuss the results obtained. We summarize our findings and conclude in Section 5.

2 OBSERVATIONS

We observed two full transits of WASP-166 b, on 2020 December 31 and 2021 February 18, with the high-resolution, optical (wavelength range 3782–7887 Å) spectrograph ESPRESSO (Pepe et al. 2021) installed on the VLT at the ESO Paranal Observatory, in Chile (ESO programme ID: 106.21EM, PI: H. M. Cegla). The observations were carried out in the 1-UT configuration (using UT1 on the first night and UT4 on the second) and high-resolution mode with 2 × 1 readout binning (HR21 mode, median resolving power of $R = 138\,000$). The target was observed with fibre A while fibre B was used to monitor the sky (i.e. simultaneous sky mode). The observations were reduced with the ESPRESSO Data Reduction Software[1] (DRS) version 2.3.1, which performs standard reduction steps for echelle spectra, including bias and dark subtraction, optimal order (2D spectra) extraction, bad pixel correction, flat-fielding and de-blazing, wavelength calibration, as well as extraction of sky spectrum from fibre B (see Pepe et al. 2021, for details). In our analysis, we used the blaze-corrected and sky-subtracted (i.e. corrected for telluric sky emission) 2D spectra.

The observations of each night cover the full planetary transit (transit duration 3.608 h, Doyle et al. 2022) and 2–3 h of out-of-transit baseline (in total, before and after the transit). The exposure time was set to 100 s to ensure an S/N sufficiently high to have photon-noise-dominated spectra (S/N ∼ 50 at 550 nm) and to obtain a good temporal cadence to sample the transit. These observations were initially obtained to perform a study of the Rossiter–McLaughlin effect, which requires a fine temporal cadence during the transit (see Doyle et al. 2022). In the first night, we obtained 80 in-transit observations and 26/40 observations before/after the transit, and in the second night, 81 in-transit observations and 30/27 out-of-transit observations before/after the transit.

For the two nights, most of the observations were taken at low airmass (<1.5, see Fig. 1 for an overview of the observing conditions). We discarded the first 8 observations of the first night (all out-of-transit observations) because they were taken at an airmass larger than 2.2, which is the maximum value for which the ESPRESSO ADC (Atmospheric Dispersion Corrector) is calibrated for. Additionally, in the second night, we discarded 3 observations taken during the post-transit baseline due to telescope vignetting.

We note that in the stellar RVs there is an offset of about 10 m s^{-1} in the systemic velocities of the two nights. These stellar RVs are obtained with the ESPRESSO DRS by computing the cross-correlation function with a suitable stellar mask. The reason for

[1] www.eso.org/sci/software/pipelines/espresso/espresso-pipe-recipes.html

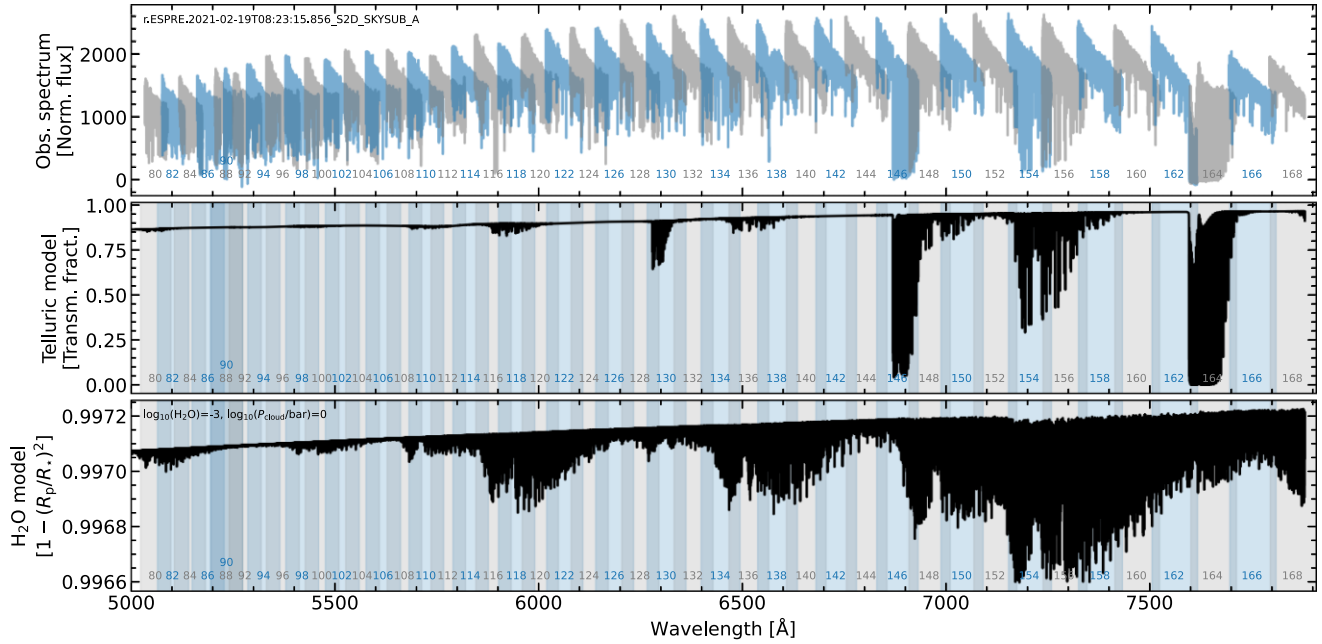

Figure 2. Example of an observed spectrum of WASP-166 (top), telluric template used to select telluric-affected regions (middle), and one of the H_2O models used to compute the CC functions (bottom). Grey and blue shaded regions and numbers indicate the different ESPRESSO orders (also colour coded in the WASP-166 spectrum in the top panel). The numbers correspond to the ESPRESSO slices (each order has two slices), starting at 0 for the first slice of the bluest order). We only show the wavelength range covering the spectral region used (where planetary model shows stronger absorption.

this offset is unknown but we attribute it to instrumental effects or differences in the observing conditions between the two nights. Regardless of the origin, the offset is too small to have any effect on our analysis (our precision is of about 1 km s^{-1}, 100 times larger than the offset). In the following, we consider as the systemic velocity of the system the average of the systemic velocities of each night.

The same observations have been used in Doyle et al. (2022) to study the Rossiter–McLaughlin effect, characterize centre-to-limb convection-induced variations, and refine the star–planet obliquity, and in Seidel et al. (2022) to detect the presence of sodium in the planetary atmosphere.

3 METHODS

3.1 Telluric correction: PCA

Spectroscopic observations taken from the ground are affected by spectral features produced by the Earth's atmosphere, known as telluric contamination. The ESPRESSO wavelength range is affected mainly by water (H_2O) and oxygen (O_2), which produce absorption lines at specific wavelength ranges with varying strength, from shallow lines called microtellurics to deep and strong lines with completely saturated cores. The strength of the lines can vary depending on the observing conditions, such as the airmass or the atmospheric water vapour content. The effect of tellurics is especially relevant when trying to study water in exoplanet atmospheres. This is because the planetary water absorption lines can overlap in wavelength space with the telluric water (see e.g. Fig. 2). Hence, we need to correct our observed spectrum from telluric lines.

To correct for telluric effects, we used a principal component analysis (PCA) on the observed spectral time series inspired by Giacobbe et al. (2021) (see also de Kok et al. 2013; Piskorz et al. 2016, 2017, Damiano et al. 2019, for other examples of works implementing PCA to study exoplanet atmospheres). We design our own automated algorithm to select the number of PCA components (described in Section 3.1.1) and to only feed into the PCA the spectral channels most affected by tellurics (Section 3.1.2).

The use of PCA to remove tellurics is based on the fact that, during the transit observations, the Earth and the target star remain stationary or quasi-stationary, while the target planet moves tens of km s^{-1} as it orbits around the star. Therefore, telluric and stellar spectral lines are always approximately located in the same pixels in the detector CCD, as they only experience a small shift in RV, while the planetary signal will shift noticeably in pixel space (see Fig. 1).

The PCA method consists in finding an orthogonal basis for the covariance matrix of the data in which the eigenvectors (also called principal components, PC) represent the direction of decreasing variance in the data. That is, the first vector or PC of the new basis has the direction of the maximum variance in the data, the second one has the direction of the second largest variance, and so on. Since the first PCs are the ones that describe most of the variance in the data, we can remove them to clean the data of the strongest telluric, stellar, and instrumental time-dependent variations.

In our case, the data matrix M is composed of the different obser- vations or frames as rows (nf) and the pixels or spectral channels as columns (nx). We work slice-by-slice, therefore, the steps described below are repeated for each slice, and for each night, separately. We note here that, conversely to other echelle spectrographs, ESPRESSO uses an APSU (anamorphic pupil slicer unit) that divides each or- der into two slices (i.e., the two slices corresponding to a specific order cover the same wavelength range). We treat each of the slices separately.

We detail our PCA implementation in Appendix A. To briefly summarise it here, we first cleaned the spectra from flux anomalies, standarised the data matrix M, and then performed the PCA. Instead of directly decomposing the covariance matrix of the data as in Giacobbe et al. (2021), we applied the PCA via a singular value decomposition (SVD).

In this work, we are only studying the presence of water in the planetary atmosphere of WASP-166 b. Therefore, we are mostly concerned in the removal of telluric lines from the observed data. The host star, WASP-166, is too warm to display any water in the stellar spectrum (spectral type F9 V and $T_{\rm eff} = 6050$, Hellier et al. 2019). The star is not especially active and we do not expect the presence of cool spots on the photosphere to be significant. Even if spots were present, their temperature contrast with the quiet photosphere is expected to be small, and hence, not sufficiently cool to display water either. Nevertheless, we want to note here that, in transmission spectroscopy, when studying planetary species that are also present in the stellar photosphere, one needs to account for the Rossiter–McLaughlin effect and centre-to-limb variations (CLV) across the stellar disc. This is because, during the transit, the planet occults different areas of the rotating stellar disc, which results in the in-transit stellar spectra being distorted (mainly depending on the projected stellar rotational velocity, the stellar obliquity, and the impact parameter). These distortions need to be accounted for to derive accurate and precise estimates of the planetary transmission spectra (see e.g. Brogi et al. 2016; Yan et al. 2017; Chiavassa & Brogi 2019; Hoeijmakers et al. 2020; Casasayas-Barris et al. 2021; Maguire et al. 2022; Seidel et al. 2022, for more details on such effects and strategies to account and correct for them).

3.1.1 Optimization of the number of PCA components per slice

Since different orders are differently affected by tellurics, we performed a per-slice optimization of the number of components *NC* to be removed when applying the PCA, which we describe in this section. To perform this optimization, we made use of the cross-correlation function (CC) of the observed spectra with a water model. We refer the reader to the following Section 3.2 for all the details on the CC computation.

For each slice affected by tellurics, we started by removing the first 2 components in the PCA. We then computed the CC of the resulting spectra with a water model and coadded the CCs of the in-transit observations in the barycentric rest frame. Coadding in the barycentric frame maximizes the presence of telluric residuals in the CC, which is what we are focusing on at this stage.

We then assessed the significance of the telluric signal by taking the value of minimum or maximum CC flux in the region ± 10 km s^{-1} (to cover the full telluric feature) around the mean BERV of the observations, and comparing it with the scatter (standard deviation) of the CC flux outside of this region. The atmospheric water vapour changes during the observations, increasing and decreasing from the overall trend dictated by the change in airmass. This causes negative and positive residuals in the processed spectra, which result in correlation and anticorrelation with the CC water template used. Therefore, when looking at the telluric signal in the CC, we considered both minima and maxima features (i.e. anticorrelation and correlation with the template). We considered a signal at the telluric position of the CC to be significant if the minimum (or maximum) flux is below (or above) 3.5 times the standard deviation of the flux of the rest of the CC. If the telluric signal is significant, we repeat the process but removing an additional PCA component. This goes on until the signal is not significant, or until the algorithm reaches the maximum number of components allowed. We set the maximum number of components to be removed to 15 (after removing over $\sim$15 components, injected planetary signals start to decrease in significance).

Although the aforementioned CC functions of our individual observations can show a minimum or a maximum at the expected telluric position, if we coadd the CC functions of the individual observations, the dominant feature at the telluric position in our case is a minimum, i.e. anticorrelation. This means that the observations with telluric residuals that anticorrelate with the models are more prominent than those observations with a positive correlation. The coadding of anticorrelated and correlated CCFs can result in a smearing of the overall signal. To check for that, we also computed the significance of the telluric peak in each individual observation. We observe that for all the cases where we have a significant signal in the 'observation-coadded' CC function, more than half of the individual observations also show a significant signal. Additionally, if more than half of the observations contain a significant telluric signal, so does the coadded CC function.

We note again that here, instead of coadding all the available observations, we coadded only the in-transit ones. This is because these observations are the only ones we use in the planet analysis, and therefore we are mostly concerned about the telluric effects in them. Aside from this, we noticed that the observations at high airmass (airmass higher than 2 at the beginning of the first night, and airmass close to 1.7 at the end of the second night) are the ones that show the strongest telluric signals in the CC function, being very distinct than those immediately after or before. If we included these high airmass observations in the coadded CC function, they heavily biased the significance of the telluric signal, so that the algorithm keeps removing components even though the in-transit telluric signal is not significant.

3.1.2 Selectively feeding telluric lines into the PCA

To try to further improve the telluric removal, instead of using the whole spectral range of each slice, we tested feeding into the PCA only the pixels affected by tellurics, i.e. pixels containing telluric lines. By doing this, the PCA should better trace the variability due to telluric changes. To determine the telluric-affected pixels, we used the ESO Sky Model Calculator[2] based on the Cerro Paranal Advanced Sky Model (Noll et al. 2012) to generate a telluric absorption model in the ESPRESSO wavelength range (see Fig. 2, middle panel). We interpolated the model to the observed wavelength grid of each slice and continuum-normalized it by fitting a cubic spline (we do this slice by slice). To fit the spline, we selected the pixel with maximum flux in windows of 25 pixels and avoided strong telluric bands that would bias the determination of the continuum. This results in a flat telluric spectrum normalized to one.

After this normalization, we flagged as telluric-affected all the pixels that overlap with a telluric line. We set the threshold to pixels where the telluric flux is below 0.998, which allows us to select most of the lines present in the ESPRESSO spectral range. The slices affected are 80–83, 96–103, 108–123, 128–141, and 146–169 (slice numbering starts at 0 for the first slice of the bluest order). When applying the PCA, only these pixels are used in the SVD. In orders with no telluric lines, we still used all the pixels to remove any systematics.

3.2 High-resolution cross-correlation spectroscopy

After correcting for tellurics with the PCA, we used the high-resolution cross-correlation spectroscopy (HRCCS) method to search for the presence of water in the atmosphere of WASP-166 b. Planetary water produces thousands of molecular absorption lines in the planetary transmission spectrum. This water signal, however,

[2]https://www.eso.org/observing/etc/skycalc

is below the noise level of the data. The HRCCS method coadds all the lines present in the transmission spectrum by cross-correlating the processed observations with an adequate spectral template of the planetary atmosphere. To compute the CC, the template is Doppler-shifted by a range of RV values, and, for each shift, we take the dot product with the observed data. This operation results in a cross-correlation function with much higher S/N than a single spectral absorption line, which enhances the planetary signal. This is because the S/N of the CC function scales with the square root of the number of lines coadded when computing the CC. In Section 3.2.1, we describe the different sets of CC models used, and in Sections 3.2.2 and 3.2.3, we explain the formalism used to compute the CC and assess the significance of the results within the cross-correlation-to-log likelihood (CC-to-log L) framework.

3.2.1 *Planetary atmosphere water models*

We generated primary eclipse spectra of WASP-166 b using GENESIS adapted for transmission spectroscopy (Gandhi & Madhusudhan 2017; Pinhas et al. 2018). GENESIS is a line-by-line numerical radiative transfer code that computes the transmission spectrum of the atmosphere given the atmospheric temperature and chemical abundance profile. The opacity of each species is computed on a grid of pressure–temperature (P–T) values for each wavelength to determine the overall optical depth of rays passing through the atmosphere and therefore the transit depth at each wavelength. We use a grid of fixed pressure values, between 100 and 10^{-7} bar and evenly spaced in log P. We assumed an isothermal temperature profile consistent with the equilibrium temperature of WASP-166 b, ∼1270 K. The chemical abundances are set as volume mixing ratios (VMR) assumed to be vertically constant throughout the atmosphere. We also included a wavelength-independent cloud deck at different pressures by setting all wavelengths to a very high opacity.

The models spanned a grid in H_2O abundance and cloud pressure, encompassing $\log_{10}(H_2O) = -1$ (highest abundance, in VMR) to -5 (lowest abundance), and cloud deck pressures of $\log_{10}(P_{cloud}/bar) = 0$ (lowest altitude) to -5 (highest altitude), both in steps of 0.5 dex (see Fig. 3 for examples). In total, we computed two grids of model spectra, one using an ExoMol POKAZATEL (Polyansky et al. 2018) line list and the other with a HITEMP (Rothman et al. 2010) line list (see Gandhi et al. 2020a, for further details on opacities). In addition, all models across both grids include collisionally induced absorption from H_2–H_2 and H_2–He interactions (Richard et al. 2012) and Rayleigh scattering due to H_2. Each model was generated at a spectral resolution of $R = 500\,000$ between 0.38 and 0.8 μm.

The models already include intrinsic pressure and temperature broadening. To better match the line shape of the expected observed planetary signal, we further broadened these model spectra by the instrument profile of the observations; for this, we used a Gaussian kernel with FWHM corresponding to the $R = 140\,000$ resolution of ESPRESSO (of ~ 2.14 km s^{-1}). We also computed the broadening due to planetary rotation (assuming it is tidally locked), which is of only 0.58 km s^{-1}. This is negligible compared to the instrument profile broadening, and hence, we do not consider it here (i.e. including it would only change the broadening from ~ 2.14 to ~ 2.22 km s^{-1}).

We used two different line lists because published water lines in the optical have not been extensively empirically verified. In the optical, water absorption bands are weaker than in the near-infrared. Due to this reduced strength, the accuracy and completeness of the model lines in the optical is expected to be worse than in the near-infrared, because their experimental verification is more challenging. Therefore, there could be differences between different line lists. To check for systematics due to these potential differences we decided to repeat our analysis using the two sets of line lists.

3.2.2 *CC-to-log* L *framework*

To assess the significance of any planetary signals, we followed the cross-correlation to log -likelihood framework (CC-to-log L, Zucker 2003; Brogi & Line 2019; Gibson et al. 2020). This is a Bayesian framework based on mapping the cross-correlation function to a log likelihood function. This allows us to accurately assess the significance of any detections by deriving confidence intervals, as well as to compare the performance of different models.

We used the CC-to-log L mapping proposed by Brogi & Line (2019)

$$\log(L) = -\frac{N}{2}\log[s_f^2 - 2R + s_g^2], \tag{1}$$

where s_f^2 is the variance of the observed spectrum, s_g^2 is the variance of the model used, R is the cross-covariance between the observed spectrum and model, and N is the number of points in the spectrum. The cross-correlation is contained in the above equation, since the correlation coefficient C is proportional to the cross-covariance R as

$$\mathrm{C} = \frac{R}{\sqrt{s_f^2 s_g^2}}. \tag{2}$$

In our case, all these refer to each individual spectral slice, because we are working slice-by-slice. The broadened models are sliced so that they are within the wavelength range of each order. We also spline-interpolated the models to the wavelength grid of each order, so that the number of data points N of the observed spectrum and model are the same. This interpolation is performed for every RV shift for which we compute the CC and log L functions.

We followed two different approaches to compute the CC and log L functions. Both methods lead to the same final result but have different advantages and drawbacks, as we describe in the following paragraphs. For more details on the implementation of each approach, we refer the reader to Appendix B.

In our initial or 'fast' approach, we compute the CC and log L functions of each slice for a fixed RV grid. Then, for each observation, the log L function of all the slices considered are coadded. Finally, the log L functions of the in-transit observations are coadded in time along the planet RV, as a function of K_p, from which we can then build the usual K_p–V_{sys} (or K_p–V_{rest} if V_{sys} has been subtract) maps. In the second or 'slow' approach, instead of computing the full CC and log L functions for a fixed grid of RV values, we only shift the model once to the expected planet RV (which is given by a pair of K_p and V_{sys} values), and compute a single CC and log L value. We repeat this for a range of K_p and V_{sys} pairs, which also results in the usual K_p–V_{sys} maps. With the slow approach, we are building the K_p–V_{sys} map pixel-by-pixel, while in the fast one, we directly get a full row of the map for each K_p considered.

The main advantage of the slow approach is that it allows us to process the model used in the cross-correlation through the same PCA as the data, which is not possible in the fast approach (see Appendix B for details about the model processing). This is important because the PCA might alter the planetary signal contained in the data by changing the line strength and shape. The models used in the fast approach do not contain any change due to the PCA, therefore, the match with a possible planetary signal will not be as good as if

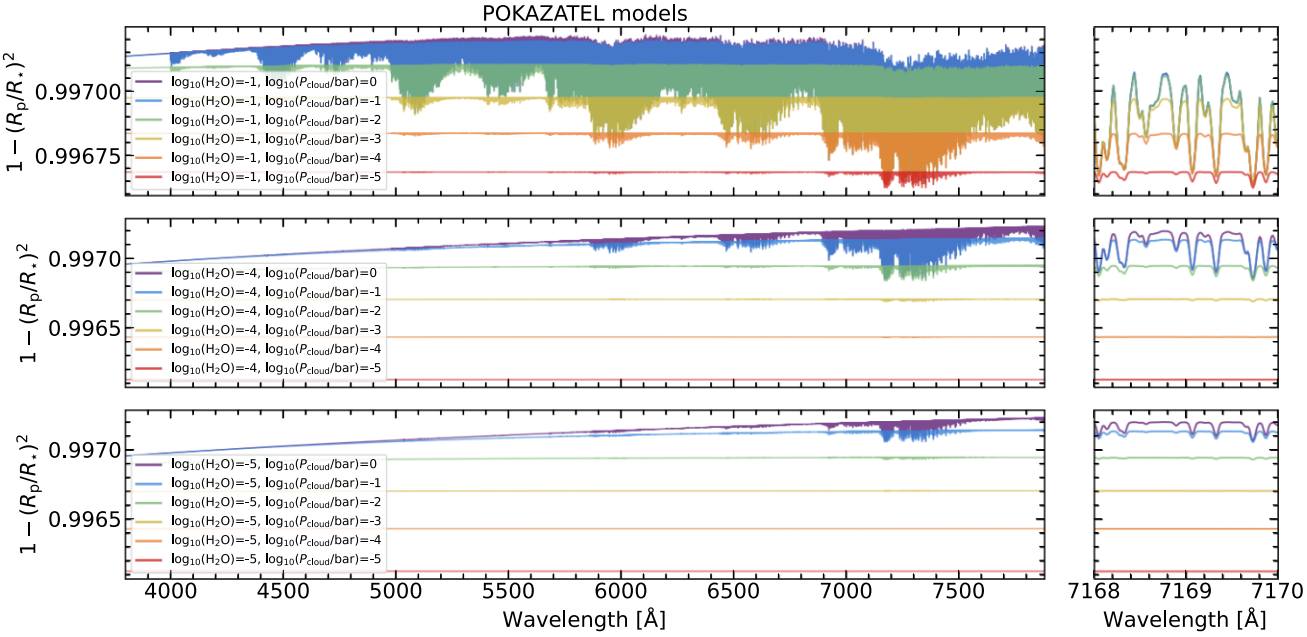

Figure 3. Left: POKAZATEL H_2O templates for WASP-166 b covering the ESPRESSO wavelength range for a range of water abundances ($\log_{10}(H_2O) = -1$ top, $\log_{10}(H_2O) = -4$ middle, and $\log_{10}(H_2O) = -5$ bottom), and a range of cloud deck pressures (depicted by various colours in all panels). Right: Zoom in on a region with strong absorption lines. This figure shows that the strength of the water absorption lines decreases as we decrease the abundance and/or decrease the cloud deck pressure.

the model has also been altered in the same way as the data. Due to this mismatch, a potential planetary detection might be weaker and biased in K_p and V_{sys}. Moreover, when performing the model comparison (see below Section 3.3), we could also misinterpret the water abundance and cloud deck pressure because the line depths do not match between model and data. Therefore, we expect more accurate results with the slow approach than with the fast one, because (1) we are computing the $\log L$ value at the exact K_p–V_{sys}, and not shifting and interpolating the whole function computed for a different K_p and V_{sys} pair, and (2) we are processing the model in the same way as the PCA modifies the data, which should result in a better match between model and data. However, the implementation of the fast approach is significantly faster than the slow one. Moreover, the fast approach allows us to study the behaviour of the telluric signal directly in the CC and $\log L$ functions, which is not possible with the slow approach. In the following, we refer to the two approaches as 'fast/unprocessed-model' and 'slow/processed-model'.

3.2.3 Confidence intervals

The CC-to-$\log L$ framework allows us to estimate confidence intervals for the K_p–V_{sys} maps (Brogi & Line 2019; Pino et al. 2020), to know which K_p–V_{sys} pair is more likely compared to all the pairs tested. According to Wilks' theorem (Wilks 1938), minus twice the difference between the $\log L$ values of two models ($\Delta \log L = -2(\log L_1 - \log L_2)$) follows a χ^2 distribution with number of degrees of freedom equal to the number of explored parameters. In our case, we are comparing the $\log L$ value of each K_p–V_{sys} pair (2 parameters) with the maximum $\log L$ of the map. I.e., we subtracted each $\log L$ value from the maximum $\log L$ of the map, which, if detected, should correspond to the planetary signal. We can then compute the p-value of this χ^2 distribution, from which we can finally derive the confidence interval value in units of standard deviation (σ) for a Normal distribution. Then, the model with the highest $\log L$ will have a σ of 0, with less likely models having increasing σ values.

We computed the confidence intervals for the data of each night separately and also on both nights combined. To combine the nights, we summed the $\log L$ values of each K_p–V_{sys} pair of both nights, and then computed the confidence intervals on this coadded $\log L$.

3.3 Model comparison

We also performed a likelihood ratio test for each of the 2 (POKAZATEL and HITEMP line lists) grids of 99 models (9 water abundances × 11 cloud top pressures) computed (see Section 3.2.1). This allows us to derive confidence intervals in both cloud top pressure and water abundance. We computed the K_p–V_{rest} map (we have subtracted the expected V_{sys}) for each of the 99 models. To compute the $\log L$ functions we followed the CC approach 2 as explained in Appendix B2, in which we modify the template in the same way as the PCA processes the data. We used a grid ranging from 90 to 150 km s^{-1} in K_p, in steps of 3 km s^{-1}, and from −20 to 5 km s^{-1} in V_{rest}, in steps of 1 km s^{-1}. This grid results in a reasonable computational time, is sufficiently fine to resolve any signals, and covers the expected planetary position as well as any tentative detections seen in our initial tests.

To identify the model with the highest significance, we compared their $\log L$ values, following the same idea as when computing confidence intervals for the different K_p–V_{sys} pairs explained above. Now, we have again two parameters: the water abundance $\log_{10}(H_2O)$ and the cloud deck pressure $\log_{10}(P_{cloud}/bar)$. In the K_p–V_{rest} map obtained for each model, we computed the maximum $\log L$ of an area around where the planet is expected. We used an area spanning ±10 km s^{-1} from the expected K_p, and −10 and +5 km s^{-1} from the expected V_{rest} (see Table 1). We used +5 km s^{-1} in V_{rest} instead of +10 km s^{-1} (i.e. which would be symmetric around the expected V_{rest}) because we are limited by the V_{rest} range covered. We tested

smaller and larger areas (from ± 5 km s^{-1} up to ± 20 km s^{-1} in both K_p and V_{rest}) and the results obtained do not change significantly. This gives us a $\log L_{max}$ for each model. We can then apply Wilks' theorem to obtain confidence intervals for the grid of models. As before, we computed twice the difference of the $\log L_{max}$ of each model from the absolute maximum of all models, derived the p-value from this distribution of $\Delta \log L$, and finally computed the confidence intervals in σ. This likelihood ratio test informs us about how likely the 99 models tested are compared to each other. The best model will then have a σ of 0, and the rest of the models will have larger values of σ. We performed this analysis on each night individually, as well as on both nights coadded (for which we used the K_p–V_{rest} $\log L$ map obtained by summing the $\log L$ values of each night).

3.4 Injection tests

We also tested the detectability of the water signal in our data by performing several injection tests using the H_2O models described in Section 3.2.1. We note that we do not use these injection tests to optimize our data analysis, but rather to assess the sensitivity of our data to a water signal. We tested different strengths of the model by scaling it to different values. To scale the model, we subtracted the mean of the model flux, which removes the average transit depth and leaves only the effect of the planet atmosphere. We then multiplied the residual spectrum (which is now only due to the planet atmosphere) by a scaling factor. We then brought back the original flux level by adding the original mean. We note that this scaling factor does not correspond to an increase or decrease in the H_2O abundance of the model. Increasing the abundance could lead strong lines to saturate, while this will not happen with a scaling factor. By using the scaling factor we only want to study its detectability.

Right before applying the PCA, we injected the scaled model to the in-transit observations. We performed this process slice-by-slice. To do this, we first Doppler-shifted the wavelength of the model by the corresponding RV of the planet at the time of each observation, so that the model shift reflect that of the actual planet, and interpolated the shifted model to the wavelength grid of each observation. We then multiplied the flux of each observation by the flux of the corresponding model. This way, each in-transit observation includes now a model of the planetary spectrum. We performed this step after the observed flux had been cleaned of bad pixels. After the injection, the standardization and the PCA are performed as explained above (see Section 3.1). We then computed the CC as explained above with the same model as injected.

4 RESULTS AND DISCUSSION

In this section, we first present the results from the tests performed to optimize the PCA algorithm. We then apply the optimum PCA algorithm to constrain the presence of water and clouds in the data via model comparison with a likelihood ratio test.

4.1 PCA optimization

As explained in the methodology section (3.1), we performed several tests with the goal of optimizing the performance of the PCA to minimize the presence of tellurics in the CC and $\log L$ functions. Here, we detail the results obtained. Unless explicitly stated, all figures in this section display CC functions and K_p–V_{rest} maps obtained using the fast/unprocessed-model CC approach (Appendix B1). This is because we wanted to directly study the shape of the CC functions, and in particular, the presence of telluric residuals, which is not possible with the slow/processed-model CC approach (Appendix B2). Moreover, to cover the same K_p–V_{rest} parameter space, the slow/processed-model method takes significantly longer computational time than the fast/unprocessed-model one, which in practice limits the K_p–V_{rest} values that we can sample, as well as the number of tests we can do. Therefore, to perform our initial tests, we decided to follow the first approach. This allowed us to test the optimal parameters for the PCA and identify any tentative planetary signals.

4.1.1 *Fixed number of PCA components*

Applying the PCA algorithm removing a fixed number of components for all the slices results in a strong telluric feature in the CC functions. We show this in the top panels of Figs 4 and 5 (black and orange lines), where we indicate the position of the telluric residuals in red. Fig. 4 shows the CC functions obtained for all the observations as a function of the orbital phase for the two nights (columns), for different tests performed (rows). Fig. 5 shows the in-transit CC functions coadded in planet rest frame for the two nights (columns) and different tests (different lines in both rows). We know that the observed feature is due to telluric contamination because it appears at the expected BERV and spans the entire sequence of observations (i.e. it is not phase dependent and is present in both in-transit and out-of-transit observations). In the CC functions, we see that the signal evolves in time from correlated (maxima) to anticorrelated (minima), as a result of the positive and negative residuals in the processed spectra. These residuals, in turn, come from the change in airmass and the changes in the atmospheric integrated water vapour column that changes during the observations, which increase and decrease the amount of water vapour above or below the overall trend. The telluric residuals do not perfectly correlate with changes in airmass and water vapour because we have applied the PCA and removed the first components prior to computing the CC functions. That is, the first PCA components removed contain part or most of the airmass and water vapour variations, and hence, the correlation is broken. We do not show them here, but if we calculate the same maps using the $\log L$ function (equation 1) rather than CC (equation 2), we see the same residuals. The telluric feature is also clearly seen in the form of maxima and minima in the K_p–V_{rest} maps produced after coadding the in-transit $\log L$ functions in planet-rest frame, see e.g. top panels in Fig. 6, where we plot the confidence intervals obtained for each night and for both nights coadded (columns).

As expected, the telluric signal decreases as we increase the number of components removed during the PCA, see top panels in Fig. 5 for examples removing 2 and 6 components (black and orange lines), but it is never completely removed. We notice that the removal seems to work better in the first night than in the second one, i.e. when the integrated water vapour is higher. We tested removing between 2 and 15 components on all the slices, but found no significant improvement, i.e. the telluric signal did not decrease further, after removing more than $\sim$6 components. We qualitatively explain the inability of PCA to de-trend telluric lines as follows. In optical observations where telluric lines are not prominent, their contribution to the total variance within one slice is also negligible. Since the SVD algorithm 'ranks' components based on their contribution to the variance, telluric residuals might potentially be absent in the first 15 components, which would instead be dominated by throughput and continuum variations. The residual level of correlation is similar to that expected from standard telluric removal algorithms e.g. direct modelling of the telluric spectrum, unless these residuals are heavily

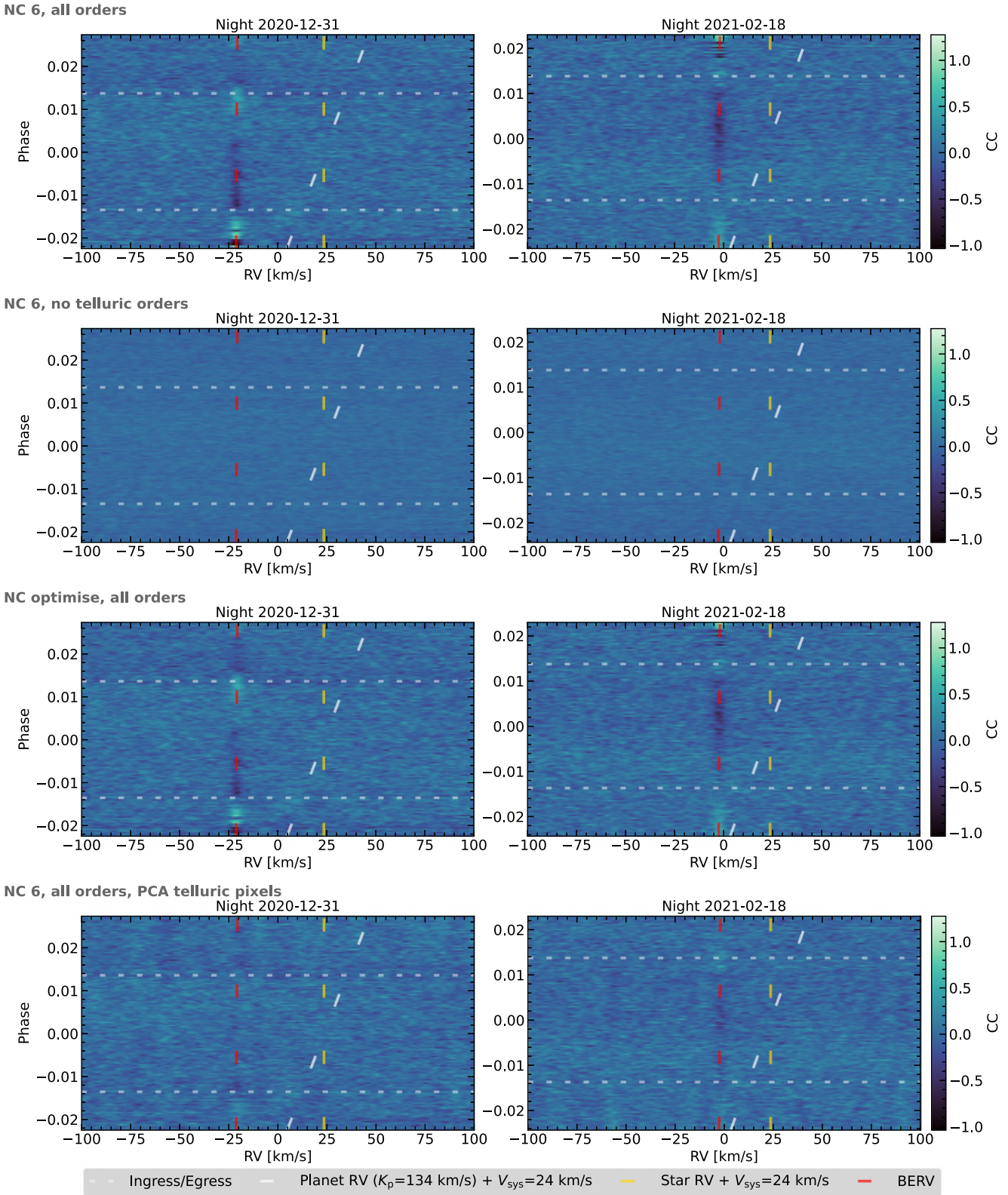

Figure 4. Comparison of telluric removal algorithms (Section 3.1 and subsections). Top row: CC of each observation as a function of the observation phase. Left-hand and right-hand panels correspond to the two transits observed. The white, yellow, and red dashed lines correspond to the planetary, stellar, and barycentric Earth RVs, respectively. The short-dash white lines indicate the transit ingress and egress. The CCs shown correspond to the coadding of the CCs of all the slices considered. The CCs have been computed with the $\log_{10}(H_2O) = -3$ and $\log_{10}(P_{cloud}/bar) = 0$ model using the fast/unprocessed-model approach, and in the PCA processing we removed 6 components for all slices. Second row: Same as top, but in this case only the CCs of slices 84–95, 104–107, 124–127, and 142–145 (slices with no or small telluric contamination) have been coadded. Third row: Same as top, but in this case the number of components removed per slice has been optimized. Bottom row: Same as top, but in this case only the pixels affected by tellurics have been used in the PCA.

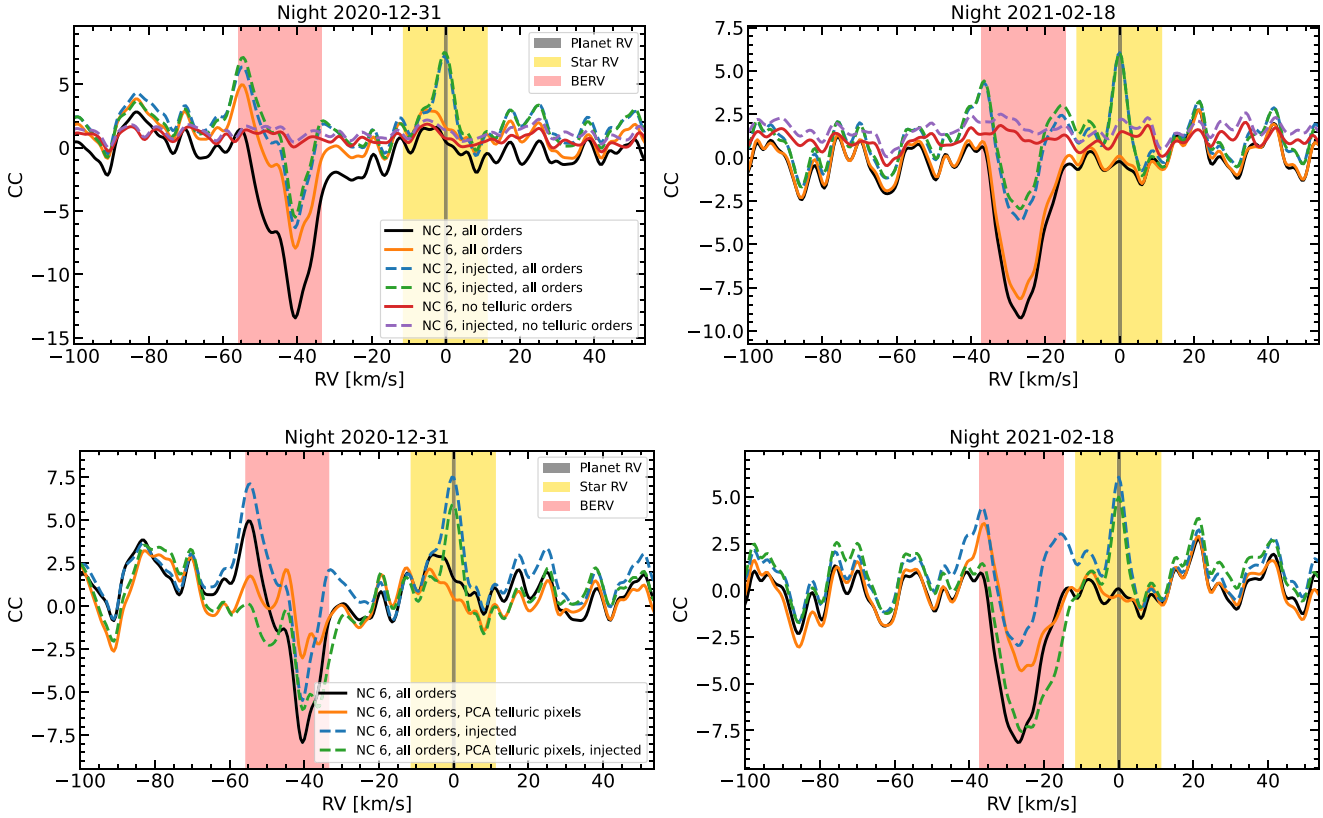

Figure 5. CC functions from a selected sub-set of tests performed to optimize the telluric removal via PCA. The CC shown are the result of coadding the CC of each in-transit observation in the planet rest frame, at the expected K_p and V_{sys}. Shaded grey, yellow, and red areas correspond to the expected planetary, stellar, and barycentric Earth RVs, respectively. Left and right correspond to the first and second night of observations, and top and bottom correspond to different sets of tests (see the legend).

masked prior to correlation. To improve the correction, we revised the SVD algorithm as explained in Section 4.1.3.

Injection tests: We also tested the behaviour of the PCA algorithm when injecting a planetary model (water abundance $\log_{10}(H_2O) = -3$, in VMR, and cloud deck pressure $\log_{10}(P_{cloud}/\text{bar}) = 0$ based on the POKAZATEL line list) in the data. The K_p–V_{sys} maps (Fig. 7, top) show that an injected planetary signal with a scaling of $\times 1$ (i.e. original strength) is recovered with high significance, despite the presence of tellurics in the data. The injected signal is clear in each night individually, with a higher confidence in the first night that increases when combining both nights. From the CC maps (Fig. 4), we see that in both nights, the expected planetary RV and the BERV do not overlap, which might help in obtaining a significant detection. Even when only removing 2 PCA components, the injected planetary signal is clearly detected in each night in the form of a peak in the CC and $\log L$ functions, see top panels in Fig. 5, where we compare removing 2 and 6 components (blue and green dashed lines, respectively). From these same tests we also see that the injected planetary signal is not affected by increasing the number of components removed. This indicates that the PCA components are not selecting the injected planetary signal, which is the behaviour we expect. We also note from Fig. 5 that the telluric residuals in the CC are slightly different if the model has been injected in the data (black, orange lines) or not (blue, green dashed lines), which indicates that part of the injected planetary signal could impact the PCA, despite the fact that its significance does not decrease.

Neglecting orders affected by tellurics: The orders where the telluric effect is stronger are those for which we see strong telluric absorption lines. These orders are also those where the planetary water shows the strongest absorption lines (see e.g. Fig. 2). Coadding the CC (or $\log L$) functions discarding these telluric-affected order slices (i.e. using only slices 84–95, 104–107, 124–127, and 142–145) results in a decrease in the telluric signal, see the CC functions in the second row of Fig. 4 and the top panels in Fig. 5, red line, which show no significant feature at the telluric position in RV space. The telluric residuals also disappear in the K_p–V_{rest} maps, see middle panels in Fig. 6. We note here that in these maps, all data points are within $\sim 4\sigma$ (or less) of one another. This means that none of the data points, i.e. none of the K_p–V_{rest} pairs, is more significant than any other. In other words, the points with the highest likelihood in maps without the telluric orders maps (i.e. confidence interval close to 0) are not significant.

If we now look at the cases where we have injected a water model, we note that the planetary signal that is clearly detected using all the orders also disappears. We see this in the top panels of Fig. 5, purple dashed line, where the clear signal at the injected RVs is no longer there, and the CC looks as flat as in the case where we have not injected a planet, as well as in the K_p–V_{rest} in the second row of Fig. 7. As happens in the case without any signal injected, now all data points in the K_p–V_{rest} maps are also within $\sim 4\sigma$ of one another, meaning that no data point is significant with respect to the others. The fact that the injected planetary signal is not seen here is not surprising. Despite the fact that the exoplanet temperature is significantly higher than the Earth's temperature, the main H_2O features are similar, and thus removing orders containing telluric H_2O also removes exoplanetary H_2O.

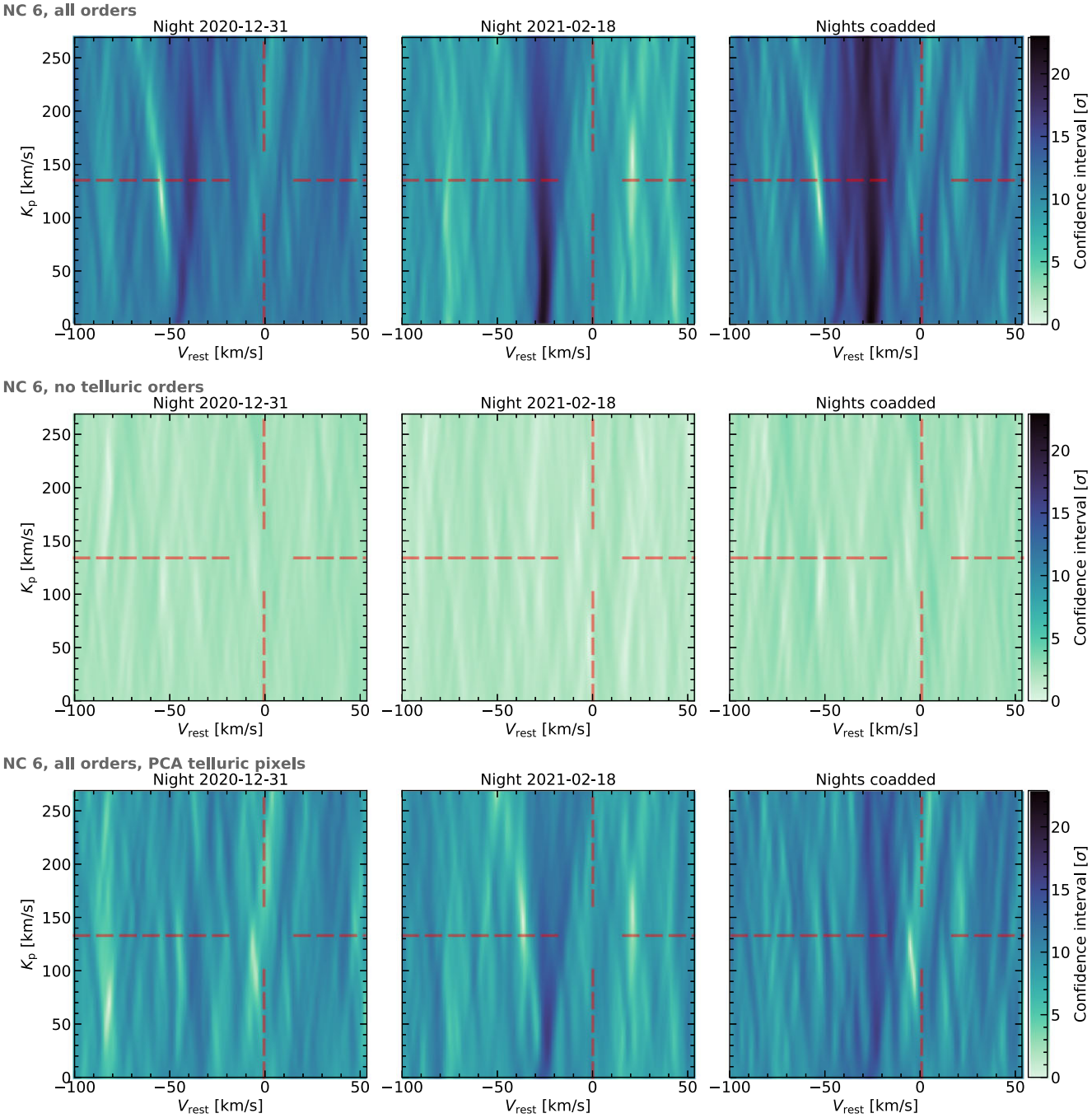

Figure 6. Top row: K_p–V_{rest} confidence interval maps obtained when removing 6 components from the PCA for all slices, using the $\log_{10}(H_2O) = -3$ and $\log_{10}(P_{cloud}/bar) = 0$ model based on the POKAZATEL line list to compute the CC, and coadding all the slices (this corresponds to coadding the in-transit CC shown in the top panel of Fig. 4). CCs computed using the fast/unprocessed-model approach. Red dashed lines indicate the expected K_p and V_{rest}. Left to right are the maps for the first, second, and coadded nights. Middle row: Same as top, but in this case, only the CCs of slices 84–95, 104–107, 124–127, 142–145 (slices with no or small telluric contamination) have been coadded (i.e. corresponds to coadding the in-transit CC shown in the above Fig. 4, second row). Bottom row: Same as top, but in this case only the pixels affected by tellurics have been used in the PCA (i.e. corresponds to coadding the in-transit CC shown in the above Fig. 4, bottom row).

4.1.2 Optimization of the number of PCA components per slice

We also optimized the number of components to be removed per slice using the method described in Section 3.1.1. The slices that are optimized, i.e. those that result in an increased number of components removed with respect to the initial, are in general those that contain strong telluric absorption lines: slices 80–83, 98, 99, 108–111, 116–121, 134–139, 147–155, 158–161, 168, and 169, see Fig. 2 above.

There is a relatively strong telluric band covering slices 128–131, and the strongest band of saturated O_2 lines in slices 162–165, for which the number of components are not optimized. In the case of the saturated band in slices 162–165, it is possible that the telluric lines are simply too strong for the PCA to be able to remove its effect; how-

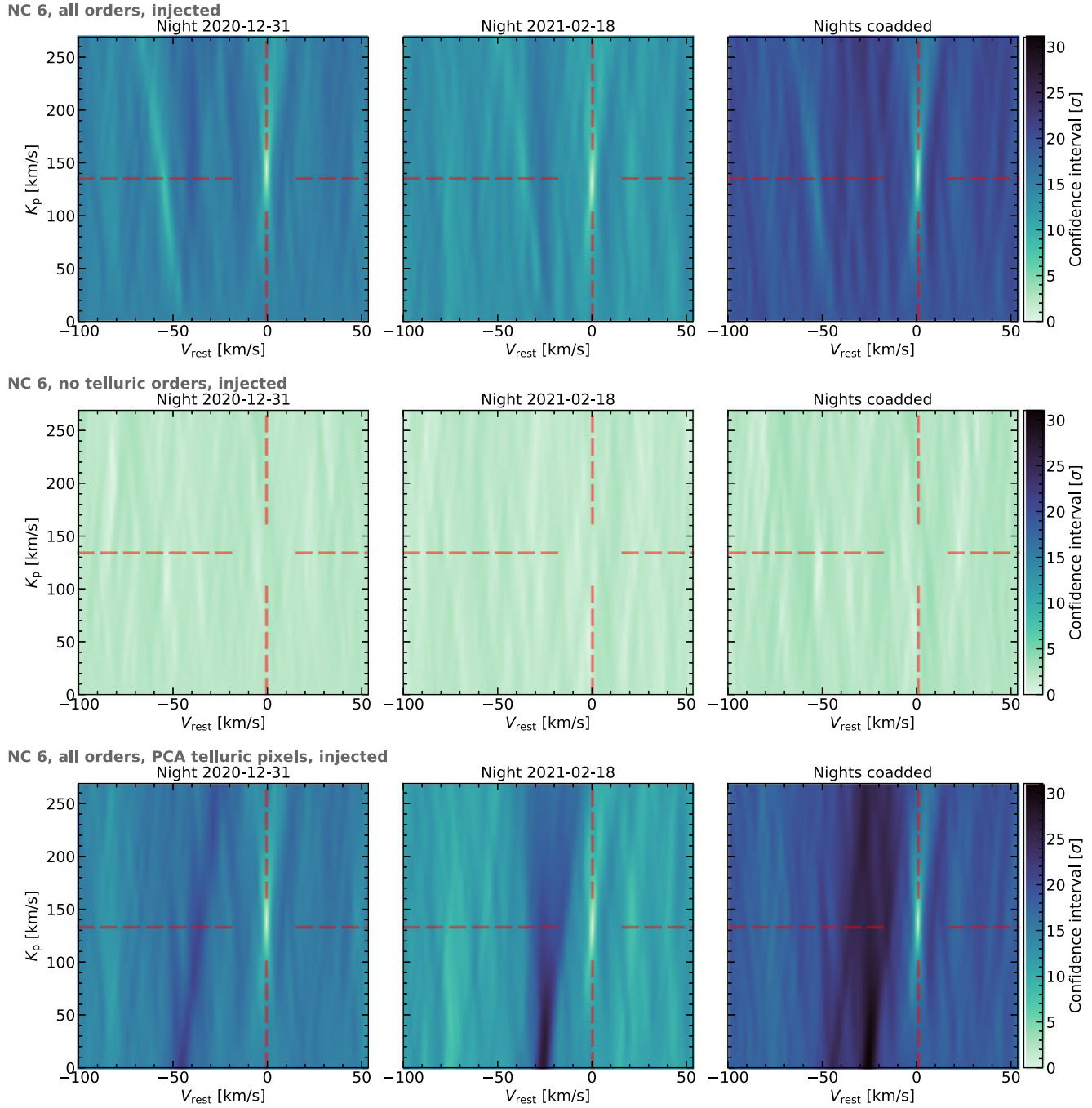

Figure 7. Same as Fig. 4, but in this case, the planetary model used in the CC has been initially injected in the observations with no scaling.

ever, this argument does not explain why the weaker band in slices 128–131 is not being properly removed. Further analysis is needed to understand these results and the behaviour of the PCA algorithm.

In most of the order slices that contain tellurics, both slices have an increased number of tellurics removed, but the final number of components is not always the same for both slices of the same order. This is not necessarily expected, and suggests that the PCA is selecting additional correlated noise different in both slices, rather than purely telluric signals, which should be the same in both slices. Again, further analysis of the PCA behaviour is needed to understand this difference.

For the two nights, most of the slices mentioned above are optimized. However, the final number of components also differs between nights for the same slices, which is expected since the tellurics behave differently in the two nights.

Fig. 4, third row, shows the CC functions obtained when applying this optimization. We notice that, despite removing a significantly higher number of components for the telluric-affected orders, this results is a CC map very similar to the one we obtain by removing a lower, fixed number of components. The K_p–V_{rest} map is also similar to this case, hence we have not included it here. This indicates that, despite still having strong telluric residuals in the CC, removing a

higher number of components does not result in a significant telluric removal.

4.1.3 Selectively feeding telluric lines into the PCA

As explained in Section 3.1.2, we modified the PCA algorithm to focus on the pixels affected by telluric lines, rather than using the full spectral order. We show the CC function and K_p–V_{rest} maps that we obtain in the bottom panels of Figs 4 and 6. In the CC function maps, we see some telluric residuals at the beginning of the first night, including part of the transit, as well as some faint residuals during the transit of the second night. This translates into a very faint signal in the K_p–V_{rest} map at the position where we expect the telluric residuals to be for the first night, and in a stronger residual for the second one. Compared to the results obtained using the full spectral order (top panels in Figs 4 and 6), in this case, the telluric contamination is significantly removed in both the CC functions and K_p–V_{rest} maps. This means that the PCA components removed are more effective in tracing the telluric variability if we only use the regions of the spectrum affected by tellurics rather than the whole wavelength range. This is again expected, because in the (sub-)matrix containing only telluric lines the latter will have a more noticeable contribution to the variance, and therefore will be ranked higher by the SVD algorithm.

In addition to using the telluric-affected pixels in PCA, we also performed the optimization of the number of components to be removed per slice, as done above. In this case, since the initial PCA components are already removing most of the telluric signal, the optimization algorithm did not detect a signal strong enough to continue removing components. Hence, for almost all the slices, the algorithm stops after the initial number of components considered has been removed. This means that the CC and K_p–V_{rest} maps look very similar if we apply the optimization and if we do not, and are not shown here.

Injection tests: With the new SVD algorithm, injected planetary signals are still recovered at high significance, see bottom panels of Fig. 7. As happened when using the full spectral range (Section 4.1.1), the telluric residuals look different if the planetary signal has been injected in the data or not; see bottom panels of Fig. 5, orange and green lines, and bottom panels in Figs 6 (non-injected) and 7 (injected case). The telluric residuals are stronger if the planetary signal has been injected. Similarly, if we now compare the case where only the telluric-affected regions are used in the PCA with the initial case where the full spectral range of the order is used (both with an injected planetary signal), the telluric residuals are different, see again Fig. 5, bottom, and top and bottom panels in Fig. 7. In general, for both nights, the tellurics are more significant if only the telluric-affected pixels are used in the PCA (bottom panels of Fig. 7) compared to the whole spectral order (top panels of Fig. 7), which is the opposite as what happens when no planetary signal is injected. As mentioned before, this indicates that the injection of a planetary model affects the telluric identification in the PCA algorithm, but this does not seem to affect the planetary signal itself, as it is clearly detectable in both cases.

4.1.4 A tentative H_2O signal from WASP-166 b?

The K_p–V_{rest} map of the first night obtained with the analysis in Section 4.1.3 above (i.e. with the modified PCA algorithm) shows a correlated signal close to the expected planetary position, about 5 km s^{-1} blue-shifted from the expected V_{rest} and extending about −30 km s^{-1} from the expected K_p (bottom left panel in Fig. 6). The signal is slightly significant with respect to its neighbouring points. While this is outside of the uncertainties of V_{rest} (or V_{sys}) and K_p (see Table 1), unaccounted atmospheric circulation at the km s^{-1} level has been shown to potentially alter V_{rest} and K_p measurements. The second night shows a similar structure but less prominent and not significant. This could be affected by the fact that the tellurics are less removed in the second night than in the first one, and hence, a possible planetary signal might be hidden in the telluric residuals. In addition, in RV space, the tellurics are closer to the expected V_{sys} in the second night compared to the first one. Despite this difference, the signal is still present when coadding both nights. It is also more significant with respect to its neighbouring points than in the first night alone. We will further discuss this candidate signal in Section 4.2.

4.2 Model comparison

In the previous section, we see that the modified PCA algorithm in which we use only the spectral regions affected by tellurics in the SVD results in less significant telluric residuals than any of the other tests performed. Therefore, to perform the model comparison between different theoretical models (as explained in Section 3.3), we used the data processed using the modified SVD algorithm, since it minimizes the telluric residuals. Moreover, to be able to compare the different models, it is important to correctly reproduce the line strength of the data. We can only guarantee this if the model used in the CC has been processed by the same PCA as the data, which we do here by using the slow/processed-model CC approach.

4.2.1 Grid search

Fig. 8 shows the confidence intervals obtained for the grid of 99 models tested. These results correspond to the models created with the POKAZATEL line list, but we obtain equivalent results for the HITEMP line lists. In each night separately, and also when coadding both nights, models with a high content of water and a cloud deck at high pressure ($\log_{10}(H_2O) \lesssim -3$ and $\log_{10}(P_{cloud}/bar) \gtrsim -2$, bottom-right quadrant of Fig. 8) are rejected with high confidence compared to the other models tested ($\gtrsim 6\sigma$, up to 15σ for some models). Models with the lowest cloud deck pressures and lowest water abundances (upper-left quadrant of the plot), are also excluded but only with $\sim 4\sigma$ confidence.

Overall, the preferred model is that with $\log_{10}(H_2O) = -4$ and $\log_{10}(P_{cloud}/bar) = 0$ (i.e. no cloud deck). There is a preference for intermediate models with low water content and high cloud deck pressure, or higher water content and lower cloud deck pressure (models coloured in light green in Fig. 8). These models are within a confidence interval of $\sim 2\sigma$ of the preferred model. This happens in all cases, i.e. for the two nights individually and both nights coadded.

The fact that the intermediate models are preferred over those with the lowest cloud deck pressure and lowest water content (upper-left quadrant of the plot) points towards a tentative detection of a water signal. If there was no planetary signal present, the preferred models will be those that have the shallowest absorption lines, i.e. those that are compatible with an almost flat spectrum (see models in Fig. 3). These are the models with the lowest water abundance and low-pressure clouds, i.e, models in the upper-left quadrant of Fig. 8), which are not preferred here. In other words, a non-detection would only exclude the bottom-right quadrant of the grid, but not the upper-left, as happens here. This is in qualitative agreement with the predictions of Gandhi et al. (2020b).

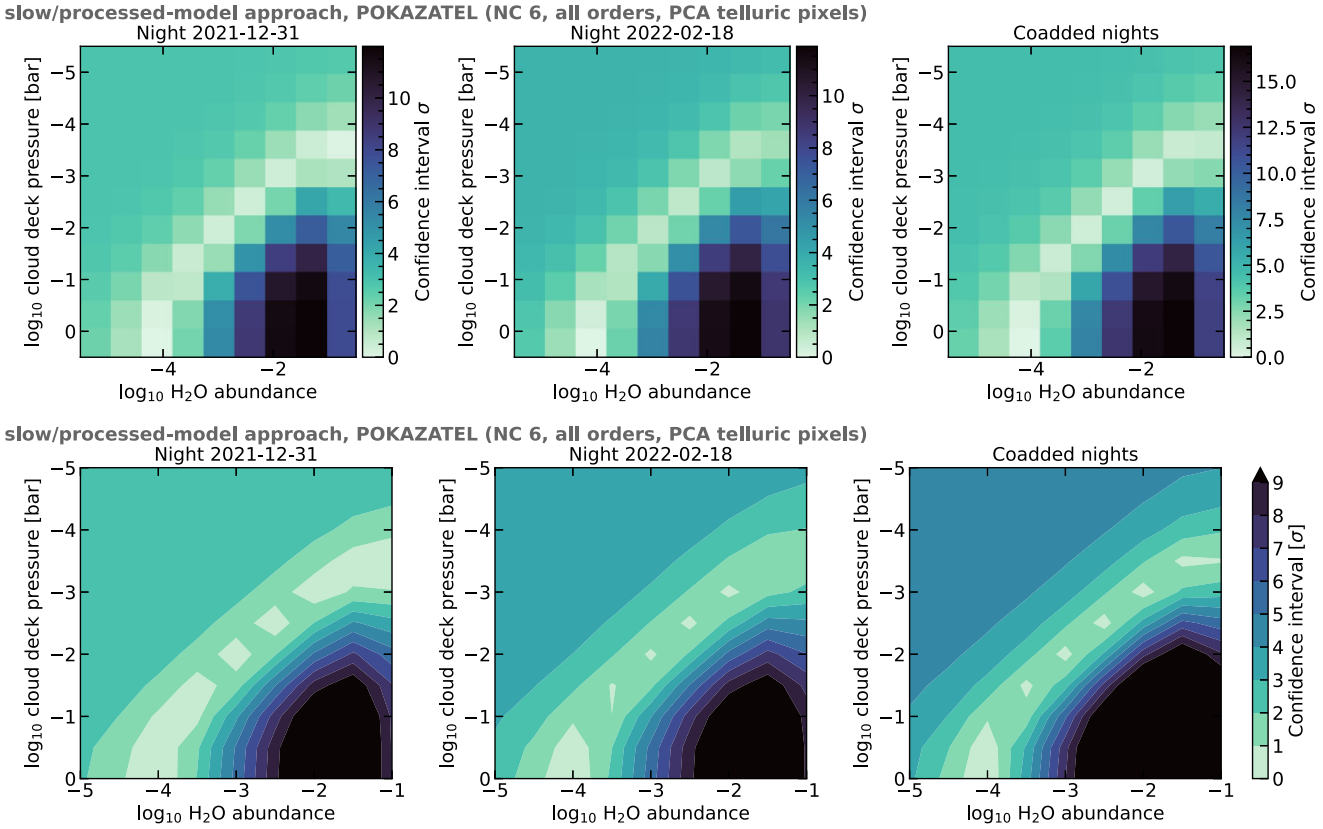

Figure 8. Top: Confidence intervals (colour) for the POKAZATEL model grids of different water abundances ($\log_{10}(H_2O) = -1$ to -5, in VMR, x-axis) and cloud deck pressures ($\log_{10}(P_{cloud}/bar) = 0$ to -5, y-axis) obtained with the slow/processed-model approach to compute the CCs. Left and middle panels are the results for each of the transits, and right, for both transits coadded. Bottom: Same as top, but here contour levels have been added to help visualize the confidence intervals.

We note that at low cloud pressures, models with water abundance $\log_{10}(H_2O) \simeq -2$ are more strongly rejected than those with higher water abundances ($\log_{10}(H_2O) = -1$). This is due to the higher mean molecular weight of the atmosphere with $\log_{10}(H_2O) = -1$ compared to the one with $\log_{10}(H_2O) = -2$. As we increase the abundance of water, for $\log_{10}(H_2O) \gtrsim -2$, the mean molecular weight of water-rich atmospheres becomes higher than at lower water abundances. This higher mean molecular weight reduces the scale height, which results in a decrease in the strength of the water absorption features (see Fig. C1 in Appendix C). Hence, due to the fact that for water abundances $\log_{10}(H_2O) \simeq -2$ the absorption features are stronger than for $\log_{10}(H_2O) = -1$, models with $\log_{10}(H_2O) \simeq -2$ are more strongly rejected (also see e.g. Gandhi et al. 2020b).

Our confidence interval analysis is relative to the 'best' model (the one with a highest likelihood). That is, the best model has by default a σ of 0 and the other models have then σ values relative to the best one. In general, it is not clear how to assess the absolute significance of the best model. Even commonly used signal-to-noise ratio approaches do not assess how good a model fits to the data in an absolute sense. To try to obtain a 'baseline' likelihood and assess the absolute significance of our models, we have performed an extra test with a 'flat' model, i.e. a model with flux equal to 1 at all wavelengths. We have used this flat model to compute the CC and $\log L$ functions with the data processed with the best PCA algorithm using the slow/processed-model method (including model processing), as done with the model grid above. We have then compared the $\log L$ obtained with this model with that of the 'best' model according to our grid analysis ($\log_{10}(H_2O) = -4$ and $\log_{10}(P_{cloud}/bar) = 0$) in the same way as we compared the different models in the grid above. That is, we performed a likelihood ratio and computed how many σs away the models are from each other. For the first night, the flat model is rejected with 3.9σ compared with the best one, for the second night, 4.2σ, and for both nights coadded, 4.9σ. This tells us that in the data, there is a signal (regardless of its origin, planetary or telluric) that is $\sim 5\sigma$ above a flat model. This test is similar to comparing the best model with models close to a flat line (i.e. those in the top left corner of our grid). Indeed, the difference in sigma obtained between the best model and those in the top left corner is similar to that obtained with respect to the flat model.

4.2.2 K_p–V_{sys} maps of the preferred model

In Fig. 9, we show a comparison of the K_p–V_{rest} maps obtained with the model favoured by our grid search in the previous section (model with $\log_{10}(H_2O) = -4$ and $\log_{10}(P_{cloud}/bar) = 0$), obtained with each of the two line lists considered, POKAZATEL (top) and HITEMP (bottom). As mentioned above, here we used the slow/processed-model CC approach, including model processing, as opposed to the results we show in the bottom panels of Fig. 6, in which we use the fast/unprocessed-model CC approach. Note also the smaller K_p and V_{rest} ranges explored here with the slow/processed-model approach, which are around a blue-shifted signal close to the expected planet position.

This blue-shifted signal close to the expected planet position, but with lower K_p value than expected, was already seen in the initial

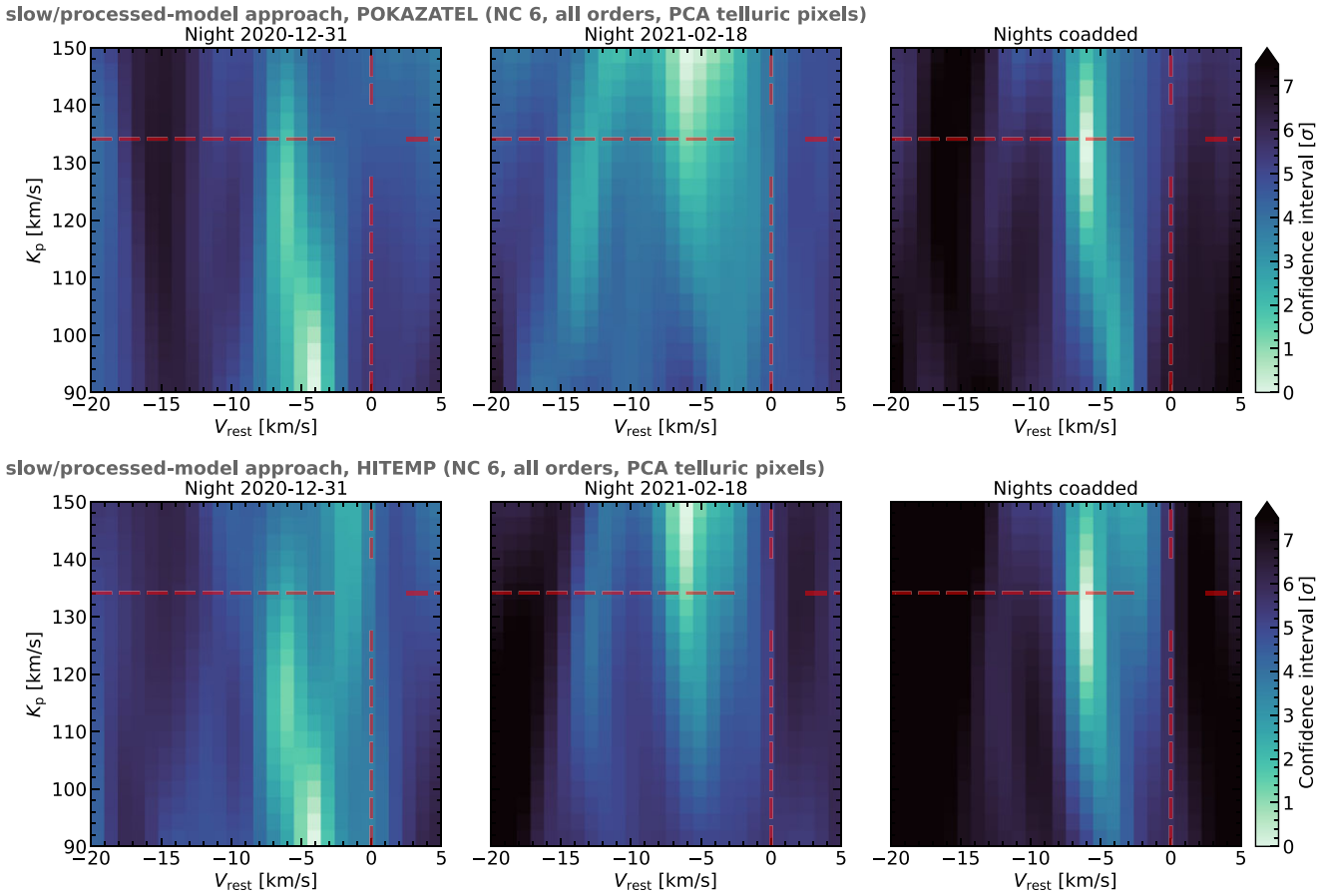

Figure 9. K_p–V_{sys} confidence interval maps obtained with the PCA algorithm that uses only telluric-affected regions, removing six PCA components, CCs computed with the slow/processed-model approach, and using the $\log_{10}(H_2O) = -4$ and $\log_{10}(P_{cloud}/bar) = 0$ model based on the POKAZATEL (top) and HITEMP (bottom) line list to compute the CC, and coadding all the order slices. Red dashed lines indicate the expected K_p and V_{sys}. Left to right are the maps for the first, second, and coadded nights.

tests with the fast/unprocessed-model CC approach (Fig. 6, bottom panels) and is still present in the first night. For the second night, this signal was less significant than in the first night in the initial tests. Now, with the slow/processed-model CC approach including the model processing, a signal also blue-shifted ~ 5 km s^{-1} appears, but it is shifted towards higher K_p values. This difference between the fast/unprocessed-model and slow/processed-model approach is something expected. With the fast/unprocessed-model method, we process the data through the PCA and then directly cross-correlate it with a model. However, in the slow/processed-model approach, we additionally process the model through the same PCA as the data before computing the cross-correlation. By doing this extra processing of the model, we modify the model in the same way as the data has been modified by the PCA. The fact that this signal is clearer in the second night using the slow/processed-model approach highlights the importance of model processing. By altering the model in the same way the PCA has altered the data, the CC should result in a better match between data and model, which is what we see here.

Coadding both nights results in an ~ 5 km s^{-1} blue-shifted signal at the expected K_p, which may be the result of the original signals in both nights being displaced in K_p in opposite directions. This blue-shifted signal is favoured with respect to the neighbouring points with $\sim 7\sigma$. For both line lists, the results obtained are equivalent, i.e. both nights show a blue-shifted signal displaced from the expected K_p, with very similar K_p–V_{rest} maps, and the signal is still present at high significance when both nights are coadded. Despite being at different K_p, the individual night signals are within 1 and 2σ (for the second and first night, respectively), of the coadded nights signal. Therefore, the signal observed when coadding both nights is not rejected by the results obtained for each night individually. We note that K_p is not strongly constrained in the transit observations that we are considering here, because they only span a small part of the total Keplerian orbit. Hence, it is hard to obtain good constraints in K_p, which could be the cause of the discrepant K_p values observed in these maps.

The -5 km s^{-1} shift in rest-frame velocity is outside of the uncertainties on the measured V_{rest} (or V_{sys}), which has been obtained from the same observations as we use here (see Table 1). The observed blue-shift could potentially be due to the presence of winds in the planetary atmosphere. The planetary Na lines detected by Seidel et al. (2022) in the same observations show a significant broadening with respect to the instrument profile, of 9.37 ± 0.95 km s^{-1}, which suggest that the Na is moving at high velocities, similarly to what we might be seeing here with H_2O.

The telluric residuals for the first and second nights are at ~ -50 and ~ -25 km s^{-1} of this signal, respectively. Therefore, we do not expect the pixels neighbouring the blue-shifted signal to be significantly affected by tellurics.

5 SUMMARY AND CONCLUDING REMARKS

In this work, we have analysed two transits of the inflated super-Neptune WASP-166 b observed with the optical, high-resolution spectrograph ESPRESSO. Using the high-resolution cross-correlation technique, we study the presence of water vapour and clouds in the atmosphere of the planet.

To correct for the presence of telluric signals which may interfere with a potential planetary signal, we start by applying a PCA algorithm on the observed spectra. We noticed that a standard PCA algorithm results in very strong telluric residuals in the CC and $\log L$ functions, as well as in the K_p–V_{rest} maps. Consequently, we performed several tests changing different parameters controlling the PCA to optimize the algorithm and obtain the best possible telluric removal. In particular, we explore the number of components removed, the spectral slices coadded, and the specific wavelengths (or pixels) used to compute the PCA components. We note here that our PCA optimization, differently from other studies in the infrared, is model independent, i.e. it is not performed by optimizing an injected signal but rather by minimizing the residual telluric noise. While this is arguably not the best choice to maximize S/N, it is a conservative choice to avoid any optimization biases as highlighted by e.g. Cabot et al. (2019) and Spring et al. (2022). A full comparison with alternative telluric removal methods such as telluric fitting with Molecfit (Kausch et al. 2015; Smette et al. 2015) or polynomial detrending (e.g. Snellen et al. 2010) is out of the scope of this work and will be the subject of future studies.

Increasing the number of components removed, whether if fixed or variable for each slice, slightly reduces the significance of the telluric residuals, but no improvement is found after removing more than ∼6 components. In all cases, relatively strong telluric residuals remain in the processed data. As expected, removing the spectral orders that are strongly affected by tellurics from the final coadded $\log L$ results in a reduction of these telluric residuals. However, injection tests show that the injected planetary signal, which is clearly detected using all the orders, even when strong telluric residuals are present, also disappears. This occurs because both telluric and planetary water show the strongest absorption lines (and hence, the strongest signal) in similar spectral ranges. Finally, we find that modifying the PCA algorithm so that it uses only the specific parts of the spectrum affected by telluric absorption (i.e. pixels that capture telluric lines), rather than using the whole spectral range, results in a significant decrease of the telluric residuals. This happens because, by feeding the algorithm only with telluric-affected regions, telluric-related variations are more noticeable, and hence, are ranked higher than other effects in the PCA components. Therefore, in our case, avoiding the ranges where tellurics are the strongest in order to mitigate telluric residuals and enhance a potential planetary detection is not a good solution because the planetary signal is also suppressed. Instead, a PCA algorithm fed with pre-defined wavelength ranges where tellurics are known to be present results in a significantly stronger telluric mitigation, whilst preserving any potential H_2O signals.

We then cross-correlated the spectra resulting from the optimized PCA algorithm with a grid of models covering a range of water abundances and cloud deck levels. We use the CC-to-$\log L$ Bayesian framework which allows us to robustly assess the significance of our results. We see that models with high water abundances and high cloud deck pressures, and models with low water abundances and low cloud deck pressures are significantly rejected. The preferred models are those with intermediate abundances and cloud deck pressures. These results are compatible with a potential detection of water in the atmosphere of WASP-166 b. If no water was detected, the preferred models would be those compatible with an almost flat spectrum, i.e. models with low water abundances and low cloud deck pressures, and only models with high water abundance and high cloud deck pressure would be excluded. We further tried to assess the significance of our best model by computing the CC function with a flat model. Compared to the best model, the flat one is rejected with 4σ to 5σ, meaning that, regardless of the origin, the data contain a signal $\sim 5\sigma$ above a flat line.

In the K_p–V_{rest} maps, we observe a correlated signal blue-shifted by about 5 km s^{-1} from the expected planetary RV. The signal observed in the two individual nights is shifted from the expected K_p by a few tens of km s^{-1} in opposite directions for each night. However, when both nights are coadded, the signal sits at the expected K_p and its significance is increased. The signals in the individual nights are within 1σ and 2σ from the coadded nights signal, meaning that the coadded nights signal is not strongly rejected by the individual night ones. Moreover, the transit observations analysed do not strongly constrain K_p because they only cover a small part of the total Keplerian orbit. The shift observed in the planetary V_{rest} could be due to winds in the planetary atmosphere. Global blue-shifts of the transmission spectrum of hot giant exoplanets have been predicted by several works (e.g. Miller-Ricci Kempton & Rauscher 2012; Showman et al. 2013; Rauscher & Kempton 2014). Such shifts have been observed in the optical through the Na doublet (e.g. Louden & Wheatley 2015; Wyttenbach et al. 2015) and tentatively reported in the infrared through CO (Snellen et al. 2010) and CO and H_2O (Brogi et al. 2016; Flowers et al. 2019). The study of the Na doublet at ∼589 nm with the same observations as those analysed here (Seidel et al. 2022) shows that the Na lines are significantly broadened. This suggests the presence of winds, which seems compatible with what we might be observing here with H_2O.

An important step in the likelihood-ratio analysis is that the models are processed through the same PCA algorithm as the data. This is necessary to avoid biases introduced by the PCA modifying any potential planetary signal during the telluric correction performed initially, since a PCA can alter both the strength and the shape of the planetary lines, resulting in spurious shifts in K_p and V_{sys}. By processing the models through the same PCA as the data, both data and models should have been modified in the same way, which should result in a better match when performing the cross-correlation. In our case, we see that if we use a model without processing it through the same PCA as the data, then the tentative blue-shifted signal is very weak in the second night compared to what we obtain if the model has been adequately processed. The slow/processed-model method is the only method attempting to reproduce the effects of the telluric removal on the model, and therefore it should be taken as the most reliable reference when quoting a detection. The fast/unprocessed-model method, despite being still common in the literature, is subject to biases with a large variety of telluric-removal algorithms, especially important when retrieving abundances, but also potentially affecting the measured value of K_p. Therefore, it is not surprising that the two methods give potentially different results, and such discrepancy does not imply that the tentative signal obtained with the slow method cannot be trusted. The biases of the fast/unprocessed-model method have been known for a few years now (e.g. the simulated tests in Brogi & Line 2019), and the slow/processed-model approach is standard among many research groups applying Bayesian analysis on high-resolution spectroscopy infrared data (e.g. recently Giacobbe et al. 2021; Line et al. 2021; Gibson et al. 2022; van Sluijs et al. 2022).

To create the grid of models covering several H_2O abundances and cloud deck pressures we used two different line lists, POKAZATEL

and HITEMP, resulting in two sets of 99 models each. Since we are working with ESPRESSO observations, our water models cover optical wavelengths, a range for which published line lists have not been extensively empirically verified, as mentioned above. Hence, we can expect worse accuracy in general and differences between the two different line lists. Water lines are known but the line positions are not necessarily accurate. This is key in high-resolution studies such as the one performed here. A lack of accuracy in the line positions could result in Doppler shifts of any expected signal. Incomplete line lists could make any existing planetary water signal weaker, but we do not expect any possible incompleteness to create shifts in the planetary signal. In the data analysed here, we see that both the individual K_p–V_{rest} maps obtained for each model and the final grid of models are very similar for both POKAZATEL and HITEMP models, which points towards a good agreement between both sets of lines. Despite that, lines could still be inaccurate or incomplete in similar ways, and this agreement does not add evidence to support a planetary origin for the tentative signal observed.

We note that when creating the models, we fixed their temperature and scaling factor, and only explored a range of water abundances and cloud deck pressures. We did not consider other sources of opacity. In other words, we did not perform a full atmospheric retrieval and have assumed that the parameters used to create our models are true, because our main goal was only to perform an initial assessment of the presence of water and clouds in the planetary atmosphere. Based on the tentative detection that we obtain, a full atmospheric retrieval is warranted to confirm the results reported here. Further observations of upcoming transits of WASP-166 b could also shed light on the differences obtained between the two nights studied here.

To summarize, we have analysed the presence of water and clouds in WASP-166 b using two transits observed with ESPRESSO. We use the cross-correlation technique with a grid of models covering a range of water abundances and cloud deck pressures. We find a tentative planetary signal blue-shifted 5 km s^{-1} from the expected planet velocity in the K_p–V_{rest} maps, which could be caused by winds in the atmosphere. A comparison of the different models used favours those with intermediate water abundances and cloud deck pressures. Models with a high water abundance and low cloud deck pressure are strongly rejected, and models with low water abundance and high cloud deck pressure are also not preferred. If no planetary signal was present, we would expect models compatible with a flat spectrum (i.e. low water abundance and high cloud deck pressure) to be favoured, which is not what we observe, hence reinforcing the tentative signal observed in the K_p–V_{rest} maps.

ACKNOWLEDGEMENTS

This work is based on observations made with European Southern Observatory (ESO) Telescopes at the La Silla Paranal Observatory under the programme ID 106.21EM. ML, HMC, and LD acknowledge funding from a UK Research and Innovation (UKRI) Future Leader Fellowship, grant number MR/S035214/1. SG is grateful to Leiden Observatory at Leiden University for the award of the Oort Fellowship. RA is a Trottier Postdoctoral Fellow and acknowledges support from the Trottier Family Foundation. This work was supported in part through a grant from Fonds de Recherche du Québec Nature et technologies (FRQNT). MLe acknowledges support of the Swiss National Science Foundation under grant number PCEFP2_194576. The contribution of MLe has been carried out within the framework of the NCCR PlanetS supported by the Swiss National Science Foundation under grants 51NF40_182901 and 51NF40_205606. We thank G. Frame for useful discussions of this work. We thank the anonymous referee for the very helpful and thorough review of the article. This work made use of `NUMPY` (Harris et al. 2020), `SCIPY` (Virtanen et al. 2020), `ASTROPY` (Astropy Collaboration et al. 2013, 2018), and `MATPLOTLIB` (Hunter 2007).

DATA AVAILABILITY

The ESPRESSO data used in this work is publicly available from the ESO archive under programme ID 106.21EM.

REFERENCES

Allart R., Lovis C., Pino L., Wyttenbach A., Ehrenreich D., Pepe F., 2017, A&A, 606, A144
Allart R. et al., 2020, A&A, 644, A155
Alonso-Floriano F. J. et al., 2019, A&A, 621, A74
Astropy Collaboration, 2013, A&A, 558, A33
Astropy Collaboration, 2018, AJ, 156, 123
Barstow J. K., Aigrain S., Irwin P. G. J., Sing D. K., 2017, ApJ, 834, 50
Benneke B. et al., 2019a, Nat. Astron., 3, 813
Benneke B. et al., 2019b, ApJ, 887, L14
Birkby J. L., 2018, preprint (arXiv:1806.04617)
Birkby J. L., de Kok R. J., Brogi M., de Mooij E. J. W., Schwarz H., Albrecht S., Snellen I. A. G., 2013, MNRAS, 436, L35
Birkby J. L., de Kok R. J., Brogi M., Schwarz H., Snellen I. A. G., 2017, AJ, 153, 138
Boucher A. et al., 2021, AJ, 162, 233
Brogi M., Line M. R., 2019, AJ, 157, 114
Brogi M., de Kok R. J., Birkby J. L., Schwarz H., Snellen I. a. G., 2014, A&A, 565, A124
Brogi M., de Kok R. J., Albrecht S., Snellen I. A. G., Birkby J. L., Schwarz H., 2016, ApJ, 817, 106
Brogi M., Giacobbe P., Guilluy G., de Kok R. J., Sozzetti A., Mancini L., Bonomo A. S., 2018, A&A, 615, A16
Bryant E. M. et al., 2020, MNRAS, 494, 5872
Cabot S. H. C., Madhusudhan N., Hawker G. A., Gandhi S., 2019, MNRAS, 482, 4422
Casasayas-Barris N. et al., 2021, A&A, 647, A26
Chene A.-N. et al., 2014, 9151, 915147
Chiavassa A., Brogi M., 2019, A&A, 631, A100
Cortés-Zuleta P., Rojo P., Wang S., Hinse T. C., Hoyer S., Sanhueza B., Correa-Amaro P., Albornoz J., 2020, A&A, 636, A98
Damiano M., Micela G., Tinetti G., 2019, ApJ, 878, 153
Deibert E. K., de Mooij E. J. W., Jayawardhana R., Fortney J. J., Brogi M., Rustamkulov Z., Tamura M., 2019, AJ, 157, 58
de Kok R. J., Brogi M., Snellen I. a. G., Birkby J., Albrecht S., de Mooij E. J. W., 2013, A&A, 554, A82
Donati J. F., 2003, in Javier T.-B., Jorge S. A., eds, ASP Conf. Ser. Vol. 307, Solar Polarization. Astron. Soc. Pac., San Francisco, p. 41
Donati J.-F. et al., 2020, MNRAS, 498, 5684
Doyle L. et al., 2022, MNRAS, 516, 298
Drummond B., Carter A. L., Hébrard E., Mayne N. J., Sing D. K., Evans T. M., Goyal J., 2019, MNRAS, 486, 1123
Esteves L. J., de Mooij E. J. W., Jayawardhana R., Watson C., de Kok R., 2017, AJ, 153, 268
Flowers E., Brogi M., Rauscher E., Kempton E. M. R., Chiavassa A., 2019, AJ, 157, 209
Fortney J. J., Lodders K., Marley M. S., Freedman R. S., 2008, ApJ, 678, 1419
Gandhi S., Madhusudhan N., 2017, MNRAS, 472, 2334
Gandhi S. et al., 2020a, MNRAS, 495, 224
Gandhi S., Brogi M., Webb R. K., 2020b, MNRAS, 498, 194
Giacobbe P. et al., 2021, Nature, 592, 205
Gibson N. P. et al., 2020, MNRAS, 493, 2215
Gibson N. P., Nugroho S. K., Lothringer J., Maguire C., Sing D. K., 2022, MNRAS, 512, 4618
Harris C. R. et al., 2020, Nature, 585, 357

Hellier C. et al., 2019, MNRAS, 488, 3067
Hoeijmakers H. J. et al., 2020, A&A, 641, A123
Hood C. E. et al., 2020, AJ, 160, 198
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Jindal A. et al., 2020, AJ, 160, 101
Kaeufl H.-U. et al., 2004, 5492, 1218
Kausch W. et al., 2015, A&A, 576, A78
Knutson H. A., Benneke B., Deming D., Homeier D., 2014a, Nature, 505, 66
Knutson H. A. et al., 2014b, ApJ, 794, 155
Kreidberg L. et al., 2014, Nature, 505, 69
Kreidberg L. et al., 2020, preprint (arXiv:2006.07444)
Line M. R. et al., 2021, Nature, 598, 580
Louden T., Wheatley P. J., 2015, ApJ, 814, L24
Madhusudhan N., 2012, ApJ, 758, 36
Maguire C., Gibson N. P., Nugroho S. K., Ramkumar S., Fortune M., Merritt S. R., de Mooij E., 2022, MNRAS, stac3388
Mayo A., Dressing C., Fortenbach C., Giacalone S., Harada C., Turtelboom E., 2021, JWST Proposal. Cycle 1, 2062
Mayor M. et al., 2003, The Messenger, 114, 20
McLean I. S. et al., 1998, 3354, 566
Miller-Ricci Kempton E., Rauscher E., 2012, ApJ, 751, 117
Moses J. I. et al., 2013, ApJ, 777, 34
Noguchi K. et al., 2002, PASJ, 54, 855
Noll S., Kausch W., Barden M., Jones A. M., Szyszka C., Kimeswenger S., Vinther J., 2012, A&A, 543, A92
Origlia L. et al., 2014, 9147, 91471E
Pelletier S. et al., 2021, AJ, 162, 73
Pepe F. et al., 2021, A&A, 645, A96
Pinhas A., Rackham B. V., Madhusudhan N., Apai D., 2018, MNRAS, 480, 5314
Pinhas A., Madhusudhan N., Gandhi S., MacDonald R., 2019, MNRAS, 482, 1485
Pino L. et al., 2020, ApJ, 894, L27
Piskorz D. et al., 2016, ApJ, 832, 131
Piskorz D. et al., 2017, AJ, 154, 78
Polyansky O. L., Kyuberis A. A., Zobov N. F., Tennyson J., Yurchenko S. N., Lodi L., 2018, MNRAS, 480, 2597
Quirrenbach A. et al., 2016, in, Takami H., Simard L.,Evans C. J., eds, Proc. SPIE Conf. Ser. Vol. 9908, Ground-Based and Airborne Instrumentation for Astronomy VI. SPIE, Bellingham, p. 990812
Rauscher E., Kempton E. M. R., 2014, ApJ, 790, 79
Richard C. et al., 2012, J. Quantum Spectrosc. Radiat. Transfer, 113, 1276
Rothman L. S. et al., 2010, J. Quantum Spectrosc. Radiat. Transfer, 111, 2139
Sánchez-López A. et al., 2019, A&A, 630, A53
Sánchez-López A. et al., 2020, A&A, 643, A24
Sedaghati E. et al., 2021, MNRAS, 505, 435
Seidel J. V. et al., 2020, A&A, 641, L7
Seidel J. V. et al., 2022, MNRAS, 513, L15
Showman A. P., Fortney J. J., Lewis N. K., Shabram M., 2013, ApJ, 762, 24
Sing D. K. et al., 2016, Nature, 529, 59
Smette A. et al., 2015, A&A, 576, A77
Snellen I. A. G., de Kok R. J., de Mooij E. J. W., Albrecht S., 2010, Nature, 465, 1049
Spring E. F. et al., 2022, A&A, 659, A121
Stevenson K. B., Bean J. L., Seifahrt A., Gilbert G. J., Line M. R., Désert J.-M., Fortney J. J., 2016, ApJ, 817, 141
van Sluijs L. et al., 2022, preprint (arXiv:2203.13234)
Virtanen P. et al., 2020, Nat. Methods, 17, 261
Wakeford H. R. et al., 2017a, Science, 356, 628
Wakeford H. R. et al., 2017b, ApJ, 835, L12
Wakeford H. R., Wilson T. J., Stevenson K. B., Lewis N. K., 2019, Res. Notes Am. Astron. Soc., 3, 7
Webb R. K., Brogi M., Gandhi S., Line M. R., Birkby J. L., Chubb K. L., Snellen I. A. G., Yurchenko S. N., 2020, MNRAS, 494, 108
Webb R. K., Gandhi S., Brogi M., Birkby J. L., de Mooij E., Snellen I., Zhang Y., 2022, MNRAS, 514, 4160
Wilks S. S., 1938, Ann. Math. Stat., 9, 60
Woitke P., Helling C., Hunter G. H., Millard J. D., Turner G. E., Worters M., Blecic J., Stock J. W., 2018, A&A, 614, A1
Wyttenbach A., Ehrenreich D., Lovis C., Udry S., Pepe F., 2015, A&A, 577, A62
Yan F., Pallé E., Fosbury R. a. E., Petr-Gotzens M. G., Henning T., 2017, A&A, 603, A73
Zucker S., 2003, MNRAS, 342, 1291

APPENDIX A: PCA IMPLEMENTATION DETAILS

Before applying the PCA, the observations of each slice are cleaned from bad pixels. We corrected for flux anomalies caused by cosmic rays. To do this, we first identified outliers by performing a sigma-clip on values deviating more than $+3$ and -6 times the standard deviation of the slice flux (values tailored for these data), and then corrected the identified spikes and the adjacent pixels on each side by linear interpolation of the neighbouring points. After this, we fitted the continuum of the slice with a linear polynomial and divided by the best fit to remove instrumental slopes. We also flag channels with low S/N (channels with median flux lower than 2 per cent of the overall median flux of all channels), which will not be used when computing the PCA eigenvectors.

We then normalized (divided) each observation (each row) by its median flux value. This is done to account for variations in the light throughput in the different observations, so that all of them now have the same baseline flux. Each pixel or channel (each column) has its mean subtracted, so that the data matrix is centred. Then each channel is divided by its standard deviation, so that the matrix is now standardized. Hence, each channel has mean equal 0 and a standard deviation of 1.

We note that the previous step of fitting and dividing by a linear polynomial to remove instrumental offsets is not strictly necessary, because any instrumental offset is removed afterwards when standardizing each channel. However, removing these instrumental slopes is needed to correctly flag channels with low S/N (if not normalized, the flagged channels would be biased due to the instrumental slope).

We then performed a PCA, but instead of directly decomposing the covariance matrix of the data as in Giacobbe et al. (2021), we performed it via an SVD of the standardized data matrix (e.g. de Kok et al. 2013). If M is the matrix with our standardized data, with dimensions $nf \times nx$ (i.e. rows $\times$ columns), the covariance matrix of the data is then given by $MM^{\mathrm{T}}/(nf-1)$. This can be diagonalized into $MM^{\mathrm{T}}/(nf-1) = WDW^{\mathrm{T}}/(nf-1)$, where W contains the eigenvectors or principal components, and D is a diagonal matrix containing the eigenvalues of a new basis. We can decompose the data matrix M via an SVD into $M = U\Sigma V^{\mathrm{T}}$, where Σ is an $nf \times nx$ diagonal matrix containing the singular values of M, U is an $nf \times nf$ matrix whose columns contain the left singular vectors of M, and V is an $nx \times nx$ matrix whose columns contain the right singular vectors of M. The singular vectors are a set of orthogonal unit vectors, hence making a new orthonormal basis. If we now consider M in terms of its SVD, it can be shown that its covariance matrix is $MM^{\mathrm{T}}/(nf-1) = (U\Sigma V^{\mathrm{T}})(U\Sigma V^{\mathrm{T}})^{\mathrm{T}}/(nf-1) = (U\Sigma V^{\mathrm{T}})(V\Sigma U^{\mathrm{T}})/(nf-1) = U\Sigma^2 U^{\mathrm{T}}/(nf-1)$. That is, the singular vectors of U are equivalent to the principal components of the covariance matrix.

We then created a new matrix containing the first NC columns of U (i.e. the first NC eigenvectors of the SVD of M or, what is the same, the first NC eigenvectors or principal components of the covariance matrix), where NC stands for number of components. We used this new matrix to perform a multilinear regression with the

initial matrix of data in order to determine the best-fitting coefficients (i.e. the eigenvalues) for the linear combination of the chosen *NC* components. This resulted in a fit to the data that should contain most of the telluric, stellar, and instrumental variations, as captured by the first *NC* components of the PCA. We then divided the initial matrix of data by this fit and subtract 1. By doing this, we obtained the residuals of the observed data where the telluric, stellar, and instrumental variations captured by the *NC* components considered have been removed.

Additionally, we applied a high-pass filter in the spectral direction to remove residual instrumental effects from the final processed data. Specifically, we first filter out pixels whose scatter (standard deviation) is larger than 2 times the median scatter of that specific pixel in all observations. We then fit 2-degree polynomials to each filtered observation to capture any residual instrumental effects, and subtract them out of each observation.

APPENDIX B: CROSS-CORRELATION IMPLEMENTATION DETAILS

B1 Fast/unprocessed-model CC approach

In this approach, we compute the full CC and $\log L$ functions over a grid of -100 to 100 km s^{-1}, in steps of 0.5 km s^{-1} (which corresponds to the ESPRESSO pixel width). That is, we shifted the model to each RV step, interpolated it to the wavelength grid of the observed spectra, and computed the CC and $\log L$ functions following equations (1) and (2). This is performed slice-by-slice for all the observations of each night, using the different models described in Section 3.2.1. For a specific model, this results in a CC and $\log L$ function per slice, per observation, and per night. For each observation, we then combined the $\log L$ functions of each slice by simply coadding them. There is no need to weight the different slices because the $\log L$ already contains information about the different S/N of each slice. This results in a single $\log L$ function per observation, per night.

To enhance the planet signal, the $\log L$ functions of the in-transit observations (where we expect the planetary signal to be) need to be coadded in the planet rest frame. To do this, we shift them by the corresponding planetary orbital velocity V_p, which we computed with the following equation (for which we assume that the planet has no eccentricity, Hellier et al. 2019)

$$V_p(t) = V_{sys} + K_p \sin\left[2\pi\varphi(t)\right], \tag{B1}$$

where V_{sys} is the systemic velocity of the system, K_p is the planet orbital radial velocity semi-amplitude, and $\varphi(t)$ is the planet orbital phase. The phase is defined as

$$\varphi(t) = \frac{t - t_0}{P}, \tag{B2}$$

where t_0 is the mid-transit time, and P the orbital period of the planet, so that $\varphi = 0$ corresponds to mid-transit. After shifting all the $\log L$ functions by the corresponding V_p, we only need to coadd them. When performing the shift to planet rest frame, we spline-interpolated the $\log L$ functions of each observation to a common RV grid. This way, each point of the $\log L$ functions of each observation can be directly summed. This results in a single $\log L$ function per night.

We perform this coadding for different V_p, computed using the expected V_{sys} and a range of K_p from 0 km s^{-1} to twice the expected value in steps of 1 km s^{-1} (following equation B1). We obtain the expected K_p using the following equation with the most up-to-date literature values (see Table 1)

$$K_p = \frac{2\pi}{P} a \sin(i_p), \tag{B3}$$

where a is the planet semimajor axis, and i_p, the orbital inclination. By doing this, we can then produce the usual K_p–V_{sys} maps (or K_p–V_{rest} if we subtract V_{sys}), since the RV grid of the $\log L$ function is equivalent to sampling different V_{sys} (see equation B1). In our case, since we included V_{sys} in the computation of V_p, the maps are in the planet rest frame, rather than in the systemic frame.

B2 Slow/processed-model CC approach: precise implementation and model processing

In the second approach, we compute a single CC and $\log L$ value for each pair of K_p and V_{sys} considered. So rather than performing the cross-correlation with the same model shifted by a range of RV steps, we only shift the model once for each pair of K_p and V_{sys} values, and use this shifted model to compute a single point of the CC and $\log L$ functions. As mentioned before in Section 3.2.2, this approach allows us to process the model through the same PCA as the data prior to performing the cross-correlation. This is key to avoid biases and should result in a better match between model and data. In the following, we explain this model processing and the computation of the $\log L$ with this slow/processed-mode l approach.

To process the planetary water model through the same PCA as the data, we first created a data matrix with the same dimensions as the original spectra ($nf \times nx$) containing the model that will be used to compute the CC (instead of the observed data). For the rows corresponding to in-transit observations, the model matrix contains the model shifted to the expected planet RV, interpolated to the same wavelength grid as each observation. For the rows corresponding to out-of-transit observations, the data matrix contains only ones.

As explained in Section 3.1, the linear regression of the data with the selected PCA components results in a fit matrix that should only contain (if the PCA works as expected) the fitted tellurics and stellar lines, and changes in flux due to varying airmass and throughput. It also contains the overall drop in flux due to the planet transit, i.e. the broad-band transmission planet spectrum. We want to inject the planetary water model to this fit matrix so that the model contains the same variability as the data. However, our model matrix already contains the drop in flux due to the planet transit, because the models are expressed in units of $1 - (R_p/R_\star)^2$. Therefore, before injecting the models, we need to normalize them to remove this effect. We do this by dividing the in-transit observations in the model matrix by their mean (we do not need to apply any change to the out-of-transit observations, which are simple a flat spectrum at flux one). After this, we injected the normalized model to the fit from the data by multiplying the two matrices.

We then apply the full PCA processing to this last matrix (including the out-of-transit observations), as done originally with the data. That is, we use the same number of components and bad-pixel masks, and perform the centering and standardization, singular value decomposition, and linear regression. This results in a matrix with the processed model per observation, which should have been altered by the PCA in the same way as the real data.

We then computed the CC and $\log L$ of the in-transit observations using the same method as in the first approach (i.e. equations 1 and 2). In this case, however, we have already shifted the template to the expected planet RV (for a specific pair of K_p and V_{sys} values). Therefore, we only compute the CC and $\log L$ once for

each observation. To get a single $\log L$ value per observation, we then directly sum the $\log L$ values that we obtain for each slice. Since the $\log L$ has been computed with the model already shifted to the expected planet RV, we can directly sum the $\log L$ of all the observations, as we are already in planet rest frame, and there is no need for interpolation as in the first approach. This directly gives us a data point on the K_p–V_{sys} maps. The processing of the model depends on the chosen K_p and V_{sys} values, therefore, we repeated this whole process (model processing and computation of a single CC and $\log L$ value) for each pair of K_p and V_{sys} values considered, resulting in the full $K_p - V_{sys}$ map (or again, $K_p - V_{rest}$ if we subtract V_{sys}).

APPENDIX C: POKAZATEL MODELS FOR HIGH WATER ABUNDANCES

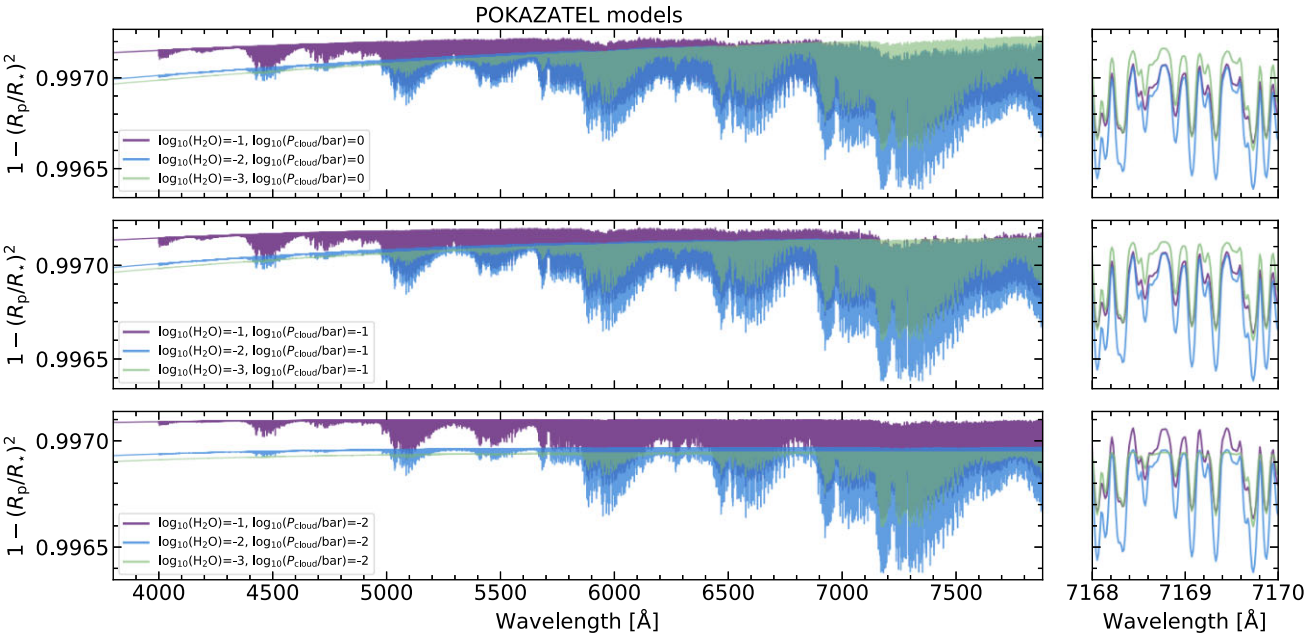

Figure C1. Left: POKAZATEL H_2O templates for WASP-166 b covering the ESPRESSO wavelength range for a range of cloud deck pressures ($\log_{10}(P_{cloud}/bar) = 0$ top, $\log_{10}(P_{cloud}/bar) = -1$ middle, $\log_{10}(P_{cloud}/bar) = -2$ bottom), and a range of water abundances (depicted by the various colours in all panels). Right: Zoom in on a region with strong absorption lines. This figure shows the decrease in absorption strength for water rich atmospheres ($\log_{10}(H_2O) = -1$, purple) compared to lower abundances ($\log_{10}(H_2O) = -2$, blue) due to the increase in mean molecular weight for water-rich atmospheres (see Section 4.2.1). For lower water abundances ($\log_{10}(H_2O) = -3$, green), the strength of the absorption features decreases as expected due to the decrease in water content.

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