Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **543,** 3456–3473 (2025) https://doi.org/10.1093/mnras/staf1659
Advance Access publication 2025 September 30

Testing the performance of cross-correlation techniques to search for molecular features in *JWST* NIRSpec G395H observations of transiting exoplanets

Emma Esparza-Borges,[1,2⋆] Mercedes López-Morales,[3] Enric Pallé,[1,2] Vladimir Makhnev,[4] Iouli Gordon,[4] Robert Hargreaves,[4] James Kirk,[5] Claudio Cáceres,[6] Ian J. M. Crossfield,[7] Nicolas Crouzet,[8] Leen Decin,[9] Jean-Michel Désert,[10] Laura Flagg,[11] Antonio García Muñoz,[12] Joseph Harrington,[13] Karan Molaverdikhani,[14] Giuseppe Morello,[15,16] Nikolay Nikolov,[3] Arif Solmaz,[17] Benjamin V. Rackham[18,19] and Seth Redfield[20]

Affiliations are listed at the end of the paper

Accepted 2025 September 24. Received 2025 September 21; in original form 2025 May 27

ABSTRACT

Cross-correlations techniques offer an alternative method to search for molecular species in *James Webb Space Telescope* (*JWST*) observations of exoplanet atmospheres. In a previous article, we applied cross-correlation functions for the first time to *JWST* NIRSpec/G395H observations of exoplanet atmospheres, resulting in a detection of CO in the transmission spectrum of WASP-39b and a tentative detection of CO isotopologues. Here, we present an improved version of our cross-correlation technique and an investigation into how efficient the technique is when searching for other molecules in *JWST* NIRSpec/G395H data. Our search results in the detection of more molecules via cross-correlations in the atmosphere of WASP-39b, including H_2O and CO_2, and confirms the CO detection. This result proves that cross-correlations are a robust and computationally cheap alternative method to search for molecular species in transmission spectra observed with *JWST*. We also searched for other molecules (CH_4, NH_3, SO_2, N_2O, H_2S, PH_3, O_3, and C_2H_2) that were not detected, for which we provide the definition of their cross-correlation baselines for future searches of those molecules in other targets. We find that that the cross-correlation search of each molecule is more efficient over limited wavelength regions of the spectrum, where the signal for that molecule dominates over other molecules, than over broad wavelength ranges. In general, we also find that Gaussian normalization is the most efficient normalization mode for the generation of the molecular templates.

Key words: methods: observational – exoplanets – planets and satellites: atmospheres – methods: data analysis – techniques: spectroscopic – planets and satellites: composition.

1 INTRODUCTION

Before the launch of the *James Webb Space Telescope* (*JWST*), bonafide detections of molecular species in exoplanet atmospheres from space were mostly limited to H_2O (e.g. D. K. Sing et al. 2016). The limited wavelength coverage of available space-borne instrumentation at the time prevented the detection of other molecules. *JWST* has changed that by providing for the first time access to a wider range of wavelengths, which are sensitive to a larger number of molecular species. Within 3 yr, the catalogue of molecules detected in exoplanet atmospheres using *JWST* has expanded significantly, with high-confidence detections of H_2O, CO_2, CO, SO_2 (e.g. L. Alderson et al. 2023; E. Esparza-Borges et al. 2023; A. D. Feinstein et al. 2023; D. Grant et al. 2023; JWST Transiting Exoplanet Community Early Release Science Team 2023; Z. Rustamkulov et al. 2023; J. Kirk et al. 2025), and CH_4 (e.g. T. J. Bell et al. 2023). Detections of more exotic molecules, such as CS_2 (B. Benneke et al. 2024; M. Holmberg & N. Madhusudhan 2024), have also been reported.

A standard approach to identifying molecular species in exoplanet atmospheres is via retrieval models that, although computationally expensive, are able to estimate molecular abundances, atmospheric temperatures, and elemental ratios (e.g. A. Banerjee et al. 2024; M. Basilicata et al. 2025; S. K. T. Kanumalla et al. 2025; J. Kirk et al. 2025; S. P. Schmidt et al. 2025). Cross-correlation functions (CCFs; e.g. I. A. G. Snellen et al. 2010; M. Brogi et al. 2012; F. Rodler, M. Lopez-Morales & I. Ribas 2012) offer another approach, which is arguably less efficient than retrievals at constraining atmospheric parameters, but more time efficient at detecting molecular features and less sensitive to systematics such as bumps and offsets that appear in transmission spectra and that are often attributed to instrumental, limb-darkening effects, etc, and vary between pipeline reductions and observing epochs (e.g. E. M. May et al. 2023). Additionally, high-resolution observations increase the sensitivity to molecular signals, and can separate the planet's signal from the stellar signal via Doppler shifts (I. Snellen et al. 2015). CCFs are regularly used

---

⋆ E-mail: emma.esparza.borges@gmail.com

in very high-resolution observations from the ground (e.g. M. Brogi et al. 2016; H. J. Hoeijmakers et al. 2019, 2020; F. Yan et al. 2019; M. Stangret et al. 2020; H. M. Tabernero et al. 2021; I. Carleo et al. 2022; L. Nortmann et al. 2025), but have now also been demonstrated to work on the highest resolution modes of *JWST*, specifically in transmission and direct imaging observations of exoplanet atmospheres with NIRSpec G395H (E. Esparza-Borges et al. 2023; S. Gandhi et al. 2023).

The CCFs detection of CO in the atmosphere of WASP-39b (E. Esparza-Borges et al. 2023, hereafter EB23), was successful as a pilot programme with a fairly simplified approach for a targeted molecule, a narrowly limited wavelength range, and normalizing the data using a simple polynomial function. In this paper, we present the results of additional work undertaken (1) to examine whether detections of CO in *JWST* NIRSpec G395H observations can be improved by expanding the approach adopted in EB23, and (2) to test the performance of the technique when searching for a number of other molecules expected to be stable in the atmospheres of hot gas giants like WASP-39b.

Section 2 briefly summarizes the methodology used in EB23, and describes how that methodology has been expanded and generalized to be applied to other molecules. Section 3 shows the results when applying the new methods to the observed *JWST* NIRSpec/G395H transmission spectrum of WASP-39b. Our findings are summarized and discussed in Section 4.

2 METHODOLOGY

In EB23, we presented a CCF method to search for CO in the observed transmission spectrum of WASP-39b with *JWST* NIRSpec/G395H (ERS-1366; PI: Batalha). Briefly, this method consisted in extracting the NRS2 detector transmission spectrum of WASP-39b at the detector pixel level and cross-correlating that transmission spectrum with a pure CO template generated using the public radiative transfer code PETITRADTRANS (P. Mollière & I. A. G. Snellen 2019), using the known parameters of the planet (F. Faedi et al. 2011; L. Mancini et al. 2018).The cross-correlation was only applied to a portion of the *JWST* NIRSpec/G395H NRS2 spectrum between 4.6 and 5.0 μm, where the CO signal was expected to dominate over noise and over additional potential contributions from other molecules (see fig. 1 in EB23). Prior to performing the cross-correlation search, both the observed and template spectra were normalized between those wavelengths using a fourth-order polynomial function. The peak of the resultant CCF – expected to appear shifted in velocity to a value consistent with the relative velocity of the system with respect to *JWST* at the time of the observations – was then normalized using the average of the CCF values over a range of Doppler velocities with low CCF values (see fig. 3 in EB23). We then reported a CCF signal-to-noise ratio (CCF–SNR), defined as the ratio between each value of the CCF and the average value over that region (see Figs 3, C1, D1, D2, and section 4 in EB23).

Here, we describe our refined approach to search for CO, and extend the method to other molecules expected to (1) be stable at temperature and pressure conditions typical of hot, gas giant exoplanets, and (2) have absorption features in the wavelength range covered by *JWST* NIRSpec G395H (i.e. 2.71–3.72 μm for NRS1 detector and 3.82–5.18 μm for NRS2 detector), based on the publicly available line lists HITEMP (The high-temperature molecular spectroscopic database, L. S. Rothman et al. 2010) and HITRAN (The high-resolution transmission database of molecular spectral parameters, I. E. Gordon et al. 2022).The molecules included in our study are summarized in Table 1, together with references to the source of their line lists. The individual line-intensities of each molecule over the wavelength range covered by *JWST* NIRSpec/G395H are shown in Appendix Fig. B1. In the following sections, we describe our refined analysis process, highlighting the parts that are similar to, or differ from, the process followed in EB23.

Table 1. Molecules searched in this study and details of their linelists.

Molecule	Line list name	Line list origin
H_2O	‘H2O_main_iso’	HITEMP[a]
CO_2	‘CO2_main_iso’	HITEMP[a]
CO	‘CO_all_iso’	HITEMP[b]
CH_4	‘CH4_main_iso’	HITEMP[c]
NH_3	‘NH3_main_iso’	HITRAN[d]
SO_2	‘SO2_HITRAN20_pRT’	HITRAN[d]
N_2O	‘N2O_HAPI2_P’	HITEMP[e]
H_2S	‘H2S_main_iso’	HITRAN[d]
PH_3	‘PH3_main_iso’	HITRAN[d]
O_3	‘O3_main_iso’	HITRAN[d]
C_2H_2	‘12C2 − 1H2’	HITRAN[d]

Notes. [a] HITEMP (L. S. Rothman et al. 2010).
[b] HITEMP update for CO (G. Li et al. 2015).
[c] HITEMP update for CH_4 (R. J. Hargreaves et al. 2020).
[d] HITRAN2020 (I. E. Gordon et al. 2022).
[e] HITEMP update for N_2O (R. J. Hargreaves et al. 2019).

2.1 Transmission spectrum

To test our technique, we use the same *JWST* ERS-1366 NIRSpec/G395H NRS2 spectrum of WASP-39b presented in EB23. Expanding from EB23, we also use the *JWST* ERS-1366 NIRSpec/G395H NRS1 spectrum of WASP-39b extracted using the TIBERIUS pipeline (J. Kirk et al. 2017, 2021) in the same manner as described in EB23. The extracted pixel level transmission spectra, with mean resolving powers of $R \sim 2380$ for NRS1 and $R \sim 3454$ for NRS2 are shown in Fig. 1. The figure also shows the cross-sections for the molecules detected in that spectrum by previous analyses (L. Alderson et al. 2023; E. Esparza-Borges et al. 2023; A. D. Feinstein et al. 2023; D. Grant et al. 2023; Z. Rustamkulov et al. 2023).

2.2 Generation of templates

We generated template transmission spectra for each molecule across the wavelengths covered by *JWST* NIRSpec/G395H, using the radiative transfer code PETITRADTRANS (P. Mollière & I. A. G. Snellen 2019). For the simulated planet, we adopted the same stellar and planetary parameters used in the EB23 (F. Faedi et al. 2011). We used line lists of all the molecules listed in Table 1 from HITEMP (L. S. Rothman et al. 2010; G. Li et al. 2015; R. J. Hargreaves et al. 2019, 2020) when available, using HITRAN2020 (I. E. Gordon et al. 2022) line lists otherwise, and assumed atmospheres composed purely of each molecule of interest. As in EB23, we first calculated each template transmission spectrum using the line-by-line radiative transfer mode of PETITRADTRANS with resolving power $\lambda/\Delta\lambda = 10^6$, and then resampled each spectrum with SPECTRES (A. C. Carnall 2017) to match the resolution of the observations.

An updated HITEMP line list of CO_2 has since become available (R. J. Hargreaves et al. 2025), this line list does not affect the conclusions of this work, but is recommended for use in future investigations due to the increased accuracy and completeness.

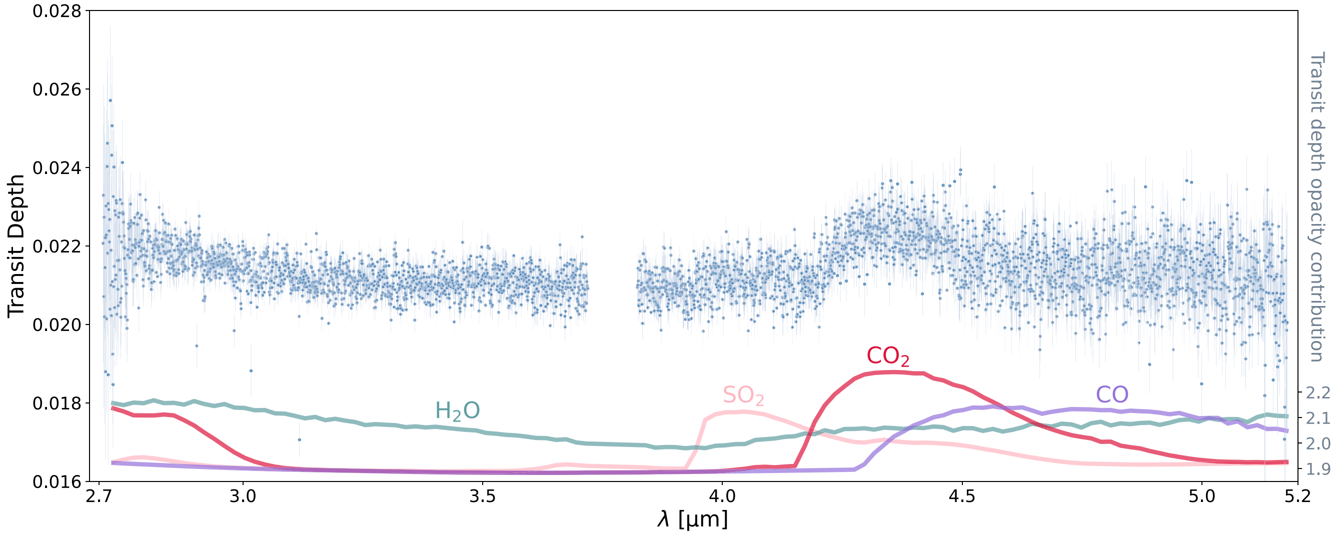

Figure 1. Transmission spectrum of WASP-39b obtained from the JTEC ERS Program NIRSpec G395H observations across NRS1 and NRS2 detectors extracted at pixel-level by the TIBERIUS pipeline. The right axis shows the relative opacity contributions of H_2O, SO_2, CO_2, and CO, which are expected to be the dominant spectral features in this spectral range in WASP-39b.

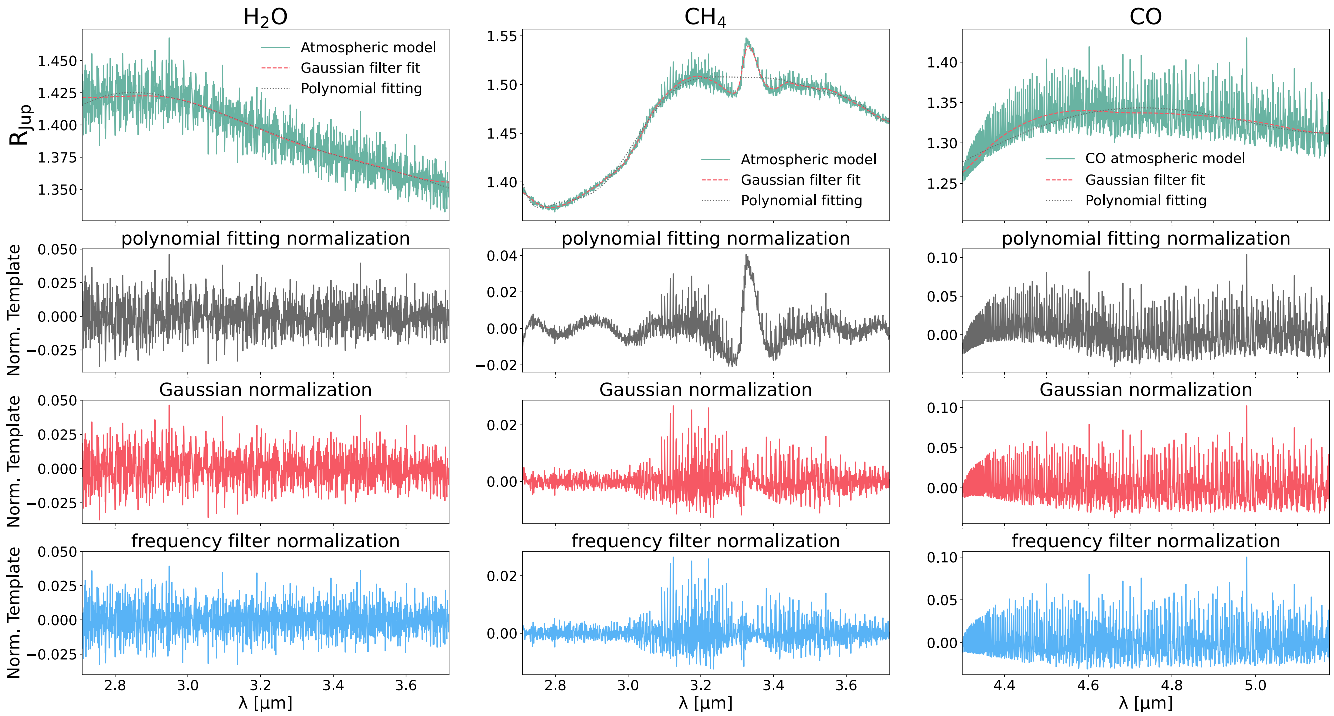

Figure 2. Examples of the template extraction from the atmospheric models at *JWST*'s spectral resolution (top panels) for three molecules, H_2O (left column panels), CH_4 (middle column panels), and CO (right column panels), using three different normalization approaches. The second, third, and final rows show the resulting normalized templates using the polynomial fitting, Gaussian, and high-pass frequency filtering approaches, respectively.

2.3 Generalization of the spectra normalization step

As an improvement from the EB23 analysis, we generalize the normalization step by performing three different normalization approaches and selecting a posteriori the most suitable approach for each molecule. The three normalization approaches implemented are: polynomial fitting (as in EB23), one-dimensional Gaussian filter fitting (frequently used in high-resolution cross-correlation studies, e.g. Y. Yang et al. 2024), and frequency filtering with a Butterworth high-pass filter (S. Butterworth 1930). To illustrate the effectiveness of each approach works, we show in Fig. 2 examples for H_2O, CH_4, and CO.

For the one-dimensional Gaussian filter normalization approach we used the `gaussian_filter1d` function in `SciPy.ndimage` (P. Virtanen et al. 2020) to fit the spectrum, setting the standard deviation for the Gaussian kernel at $\sigma = 100$. For the frequency filtering approach, we used the function `butter` in `SciPy.signal` (P.

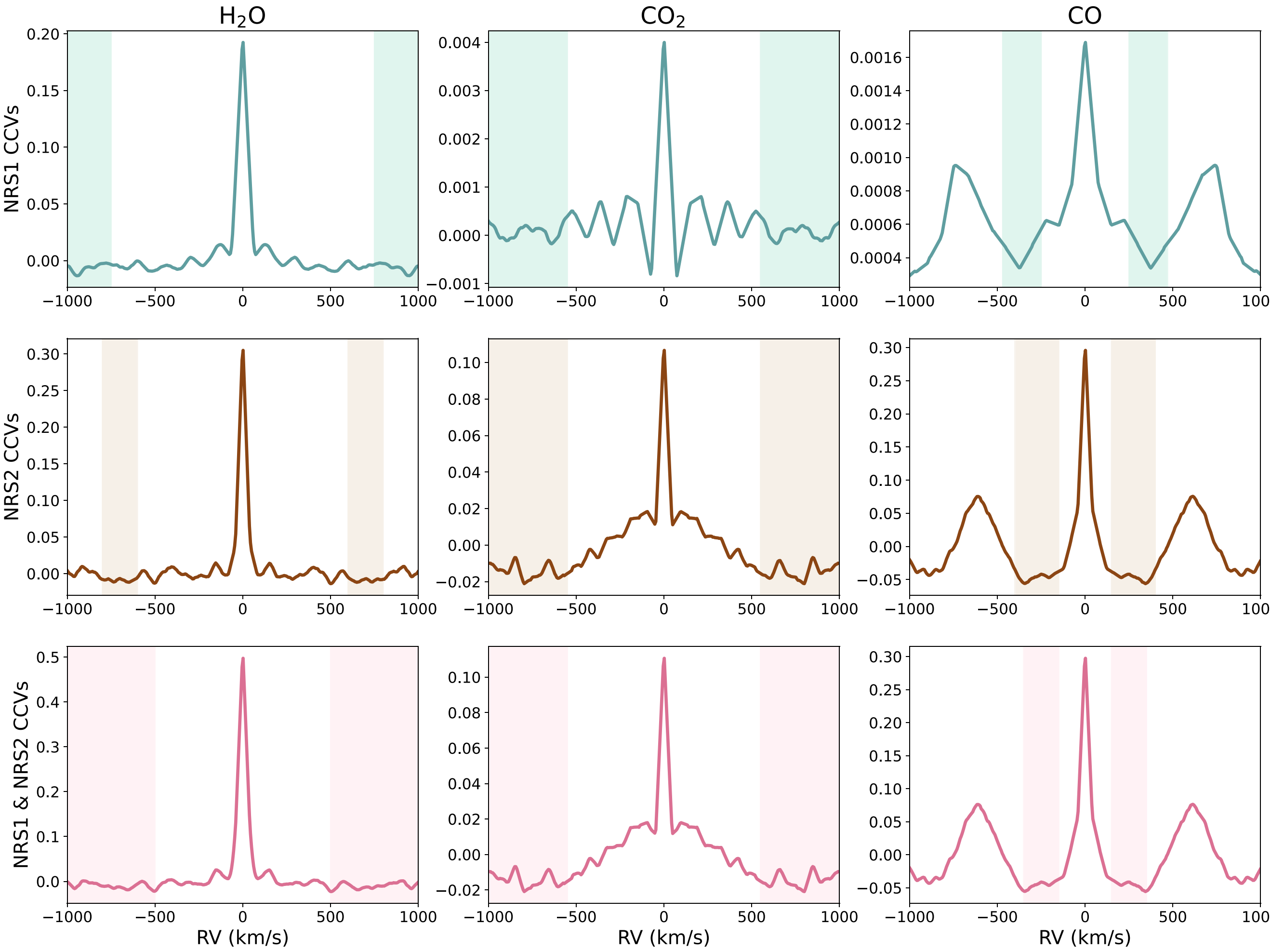

Figure 3. Definition of the baseline regions (shaded regions, see Table D1) with respect to the template–template CCF within the optimal wavelength ranges using the Gaussian normalization on NRS1 (top panels), NRS2 (middle panels), and both ranges combined (bottom panels) for each detected molecule, i.e, H_2O (left column), CO_2 (central column), and CO (right column). In Appendix C, we show an analogous plot considering the frequency filter normalization instead.

Table 2. Definition of the optimal wavelength ranges for each detected molecule.

Molecule	NRS1 wavelength Range (μm)	NRS2 wavelength Range (μm)
H_2O	[2.71, 3.72]	[3.82, 4.20] ∪ [4.60, 5.18]
CO_2	[2.71,3.00]	[4.00,4.80]
CO	[2.71,3.20]	[4.60,5.00]

Virtanen et al. 2020) to compute the coefficients of a fifth-order Butterworth high-pass filter, setting the critical cut-off frequency at Wn = 500 and using a sampling frequency of fs = 10000. We used the function `sosfiltfilt` in `SciPy.signal` to apply the filter to the spectrum. For a description of the approach based on polynomial fitting, see EB23.

The performance of the normalization approaches depends on the shape and spectral features of the molecular absorption, i.e. slopes/trends or Gaussian-like/broad spectral features. To illustrate the performance of each normalization approach, we show in Fig. 2 the results of normalizing the H_2O and CH_4 bands in the NRS1 spectrum and the CO band in NRS2 with each normalization approach. This example shows how the H_2O and CH_4 bands require different normalization approaches to extract a normalized molecular template suitable for cross-correlations. Thus, in the context of a systematic search of different molecules, it is necessary to use different normalization approaches and select the one that works best for each molecule. In general, we find that 1D Gaussian filter fitting works best for the molecules detected in this study, as described in more detail in Section 3. In addition, we find that the polynomial fitting normalization is the least efficient normalization mode, as it might introduce artefacts in the generated templates for being less suitable to fit molecules presenting broad spectral features or curved trends. For this reason, we discarded the polynomial fitting normalization in this study.

2.4 Cross-correlation analysis

The next step of our analysis consists of cross-correlating the normalized pixel-level spectrum of WASP-39b with the normalized template of each molecule. We followed the procedure described in EB23, which we briefly summarize here for the reader's convenience.

We used the `crosscorrRV` algorithm in PYASTRONOMY (S. Czesla et al. 2019) to perform cross-correlations. This function

Doppler-shifts the template over a range of radial velocity (RV) offsets, evaluates the shifted template using linear interpolation to match the wavelength points of the observed spectrum and calculates the cross-correlation function at each RV shift as

$$\mathrm{CC(RV)} = \sum_{\lambda} \mathrm{S}(\lambda) \times \mathrm{T}(\lambda - \Delta_{\lambda,\mathrm{RV}}), \tag{1}$$

where $\Delta_{\lambda,\mathrm{RV}}$ is the Doppler shift implemented as $\Delta_{\lambda,\mathrm{RV}} = \lambda(\mathrm{RV}/c)$, $\mathrm{T}(\lambda - \Delta_{\lambda,\mathrm{RV}})$ is the shifted template, and $\mathrm{S}(\lambda)$ is the observed spectrum. Additionally, before computing the cross-correlation, we zero-padded the template to avoid edge artefacts in the cross-correlation.

Cross-correlations were performed over a range of RVs between ± 1000 km s^{-1}, in 6 km s^{-1} steps. As explained in section 4 of EB23, 6 km s^{-1} corresponds to a fraction of the RV pixel-to-pixel sampling. To test how these results depend on the resolution in RV space, we repeated this cross-correlation analysis for wider RV steps (20, 43, and 64 km s^{-1}). We present the results of these tests in Appendix A. These results show consistent molecular detections, being the CCF maxima at the expected velocity (Fig. A1) and without substantial variations in the detection significances when varying the RV step in the cross-correlation (Fig. A2).

We also performed the cross-correlation of each molecule's template with itself, to identify secondary CCF peaks and to identify regions in velocity space where the cross-correlation values (CCVs) of the CCF are low and can therefore be used as cross-correlation baseline values (see Fig. 3).

The last step of the method is to define the baseline regions of the CCF for each molecule, with respect to which the significance of the detected peak in the CCF is reported. This is a similar step to the one presented in section 4 of EB23 for the CO band in the NRS2 spectrum, where we defined the baseline of the CCF in the low-CCV/valley regions between the central peak and the secondary peaks (if any) in the template–template CCF, which indicate the RV regions (with respect to the velocity of the central peak) where the CCVs of the data–template CCF are expected to be dominated by noise. In Fig. 3, we show the definition of the baseline regions for the detected molecules (H_2O, CO_2, and CO) compared with their template–template CCFs.

Additionally, to facilitate the interpretation of the resultant CCFs we converted the CCFs from cross-correlation values (CCVs) to signal-to-noise ratio (SNR) values following the conversion procedure described in section 4 of EB23 and computing them as

$$\mathrm{SNR(RV)} = \frac{\mid \mathrm{CCV(RV)} - \mu_{\mathrm{baseline}} \mid}{\sigma_{\mathrm{baseline}}}, \tag{2}$$

where μ_{baseline} and $\sigma_{\mathrm{baseline}}$ are the mean and standard deviation of the CCVs in the baseline regions.

3 CROSS-CORRELATION RESULTS

We carried out cross-correlations between the normalized observed spectrum and the normalized template for each molecule listed in Table 1. For each molecule, we systematically performed cross-correlations across the whole wavelength ranges covered by the NRS1 and NRS2 detectors, first separately (Figs E1 and E2) and then jointly (Fig. E3). We repeated this process for two different normalization modes: subtracting a Gaussian filter (Figs E1 and E3) and filtering by a high-pass frequency filter (Figs E2 and E3). For each molecule, we compared the resulting observations–template CCF with the template–template CCF (shown in dashed lines in Figs 4, 5, 6, E1, E2, and E3), with the aim of identifying the presence of secondary peaks in the CCFs. To perform this comparison we put the template–template CCF in the planet's rest frame by introducing an offset in RVs that accounts for the systemic velocity of WASP-39b system (-58.4421 km s^{-1}) and the *JWST* barycentric velocity at the time of the observations (-28.888 km s^{-1}).

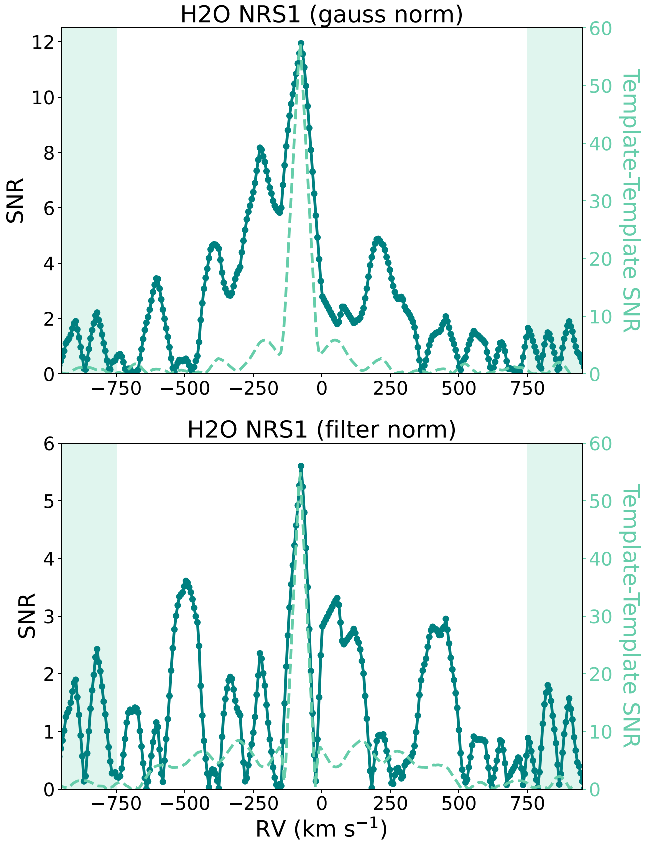

Figure 4. Result of the cross-correlation search of H_2O along the NRS1 detector (2.71–3.72 μm) using the Gaussian normalization mode (top) and the frequency–filter normalization mode (bottom). The solid-dotted line shows the resulting data–template CCF and the dashed line shows the template–template CCF shifted towards the expected RV offset produced by the barycentric velocities of WASP-39 system and *JWST* at the time of the observations.

For each molecule, we performed an additional cross-correlation search over a portion of the observed transmission spectrum that we defined as the optimal wavelength range considering the distribution of lines for a given molecule, and avoiding the wavelength ranges where other molecules might dominate (Fig. B1). The optimal wavelength ranges considered for each molecule are listed in Table 2. For instance, we masked the portion of the spectrum between 4.2 and 4.6 μm in the optimal wavelength range definition for all molecules except CO_2, as in this range the opacity contribution of CO_2 clearly dominates over any other molecular feature.

Using our generalized methodology, we confirm the detection of H_2O, CO_2, and CO via cross-correlation. Below we discuss the results that achieve the highest detection significance for these molecules in Sections 3.1, 3.2, and 3.3, which are generally obtained from the cross-correlation search over the optimal wavelength range for each molecule. We analyse the significance of these detections in Section 3.4. In addition, we show the results of the systematic cross-correlation search throughout the NRS1 and NRS2 detectors in Appendix E. We report non-detections for the rest of molecular

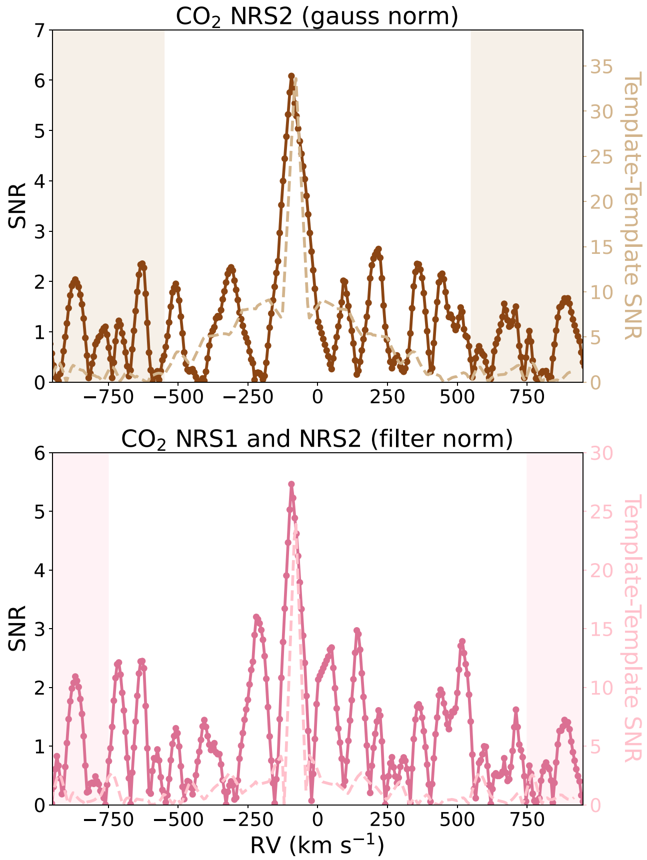

Figure 5. Results of the cross-correlation search of CO_2 along the optimal wavelength range of NRS2 detector using the Gaussian normalization mode (top) and along the wavelength ranges of NRS1 and NRS2 detectors using frequency-filter normalization (bottom). The solid-dotted line shows the data–template CCF and the dashed line shows the template–template CCF.

species investigated, which are CH_4, NH_3, SO_2, N_2O, H_2S, PH_3, O_3, and C_2H_2. These non-detections are in agreement with the expected atmospheric composition of WASP-39b (L. Alderson et al. 2023; E. Esparza-Borges et al. 2023; A. D. Feinstein et al. 2023; D. Grant et al. 2023; JWST Transiting Exoplanet Community Early Release Science Team 2023; Z. Rustamkulov et al. 2023; S.-M. Tsai et al. 2023), except for the non-detection of SO_2, which is further discussed in Section 4.

3.1 H_2O detection

Among all studied molecules, water produces the most significant cross-correlation detection. H_2O is detected both in the NRS1 (2.71–3.72 μm) and NRS2 (3.82–5.18 μm) detectors independently, as well as in their combined analysis. These cross-correlation functions converted to SNR are shown in the top panels of Figs E1, E2, and E3, respectively, where the SNRs are defined as the ratios between the CCVs and the average CCVs of the baseline regions as defined in Section 2.4 and Fig. 3. These CCFs show central peaks that appear shifted towards the expected RV offset corresponding to the barycentric velocities of WASP-39b system and *JWST* at the time of the observations.

In Fig. 4, we show the CCF resulting from the search throughout the NRS1 detector's wavelength range compared to the shifted template–template CCF for the same wavelength range, where the H_2O detection is the most significant for both normalization modes.

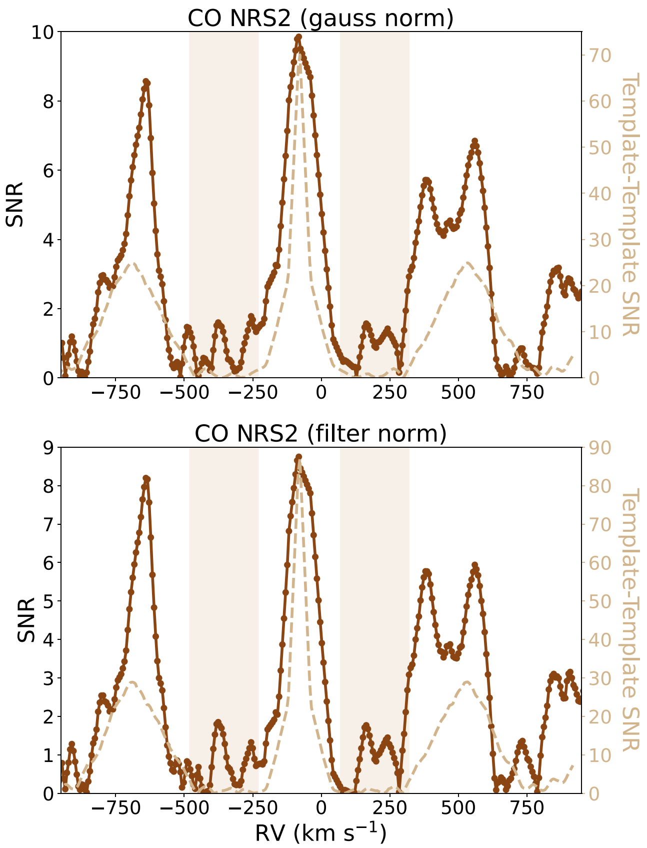

Figure 6. Results of the cross-correlation search of CO along the optimal wavelength range for NRS2 detector (4.6–5.0 μm) using the Gaussian normalization mode (top) and the frequency–filter normalization mode (bottom). The solid-dotted line shows the data–template CCF and the dashed line shows the template–template CCF.

We obtained that the central peak of the CCF is more prominent when using the Gaussian fitting subtraction normalization ($SNR = 11.9$) than with the frequency filtering normalization ($SNR = 5.6$). Besides this, we find that H_2O is confidently detected in the NRS1 detector regardless of the normalization mode while in the NRS2 the detection is strongly favoured by the use of the Gaussian normalization mode.

3.2 CO_2 detection

The presence of CO_2 is detected in the NRS2 detector and in the combination of the NRS1 and NRS2 detectors with both normalization modes. However, the signal of CO_2 is not retrieved by cross-correlations in the NRS1 detector's wavelength range alone through any of the normalization modes. This is expected because of the predominance of H_2O lines in the NRS1 spectral range, which are probably diluting the cross-correlation signal of CO_2, and also because the blue edge of the NRS1 detector is a very noisy region where the observed spectrum has a low SNR.

In Fig. 5, we show the cross-correlation result on NRS2 detector using the Gaussian normalization mode (top panel) and on the combination of the NRS1 and NRS2 detectors with the frequency filter normalization (bottom panel), which correspond to the most significant CO_2 detections for each normalization mode with $SNR = 6.1$ and $SNR = 5.5$, respectively. Both results are obtained using the optimal wavelength range definition (Table 2), where the CO_2 opacity contribution is expected to be significant or dominant considering the rest of the expected molecules.

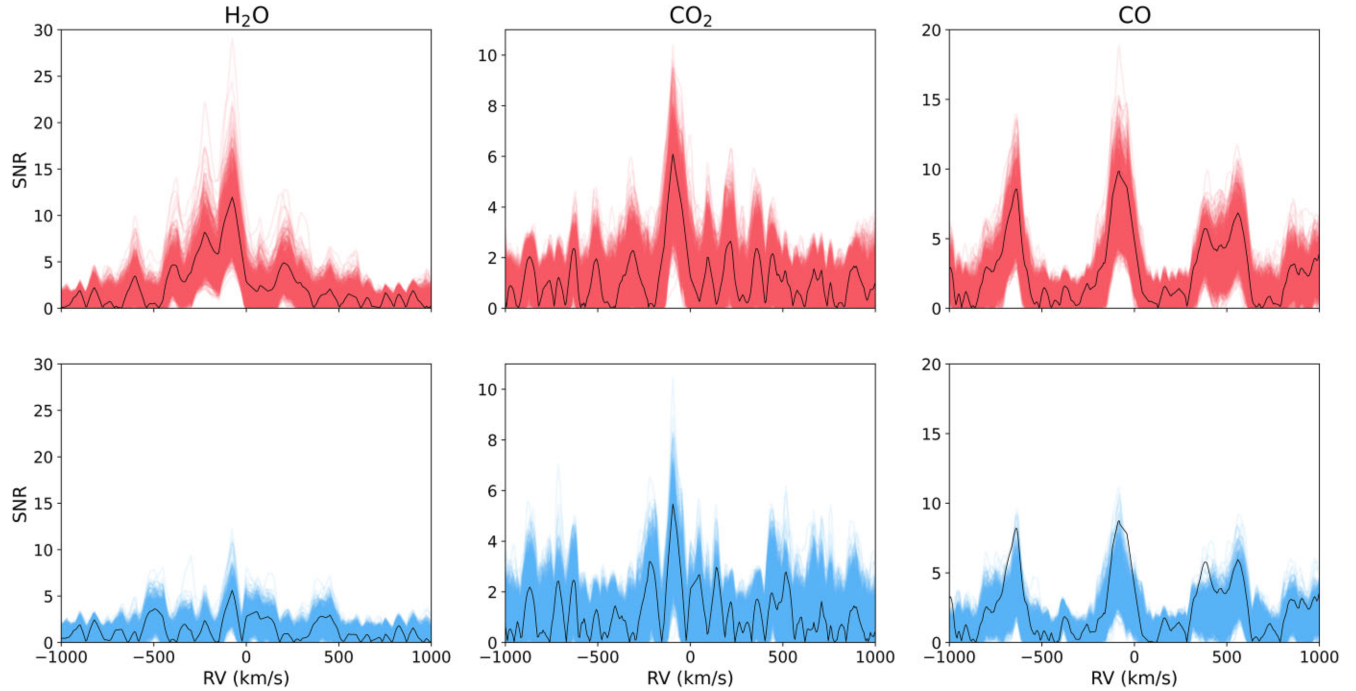

Figure 7. Results of the cross-correlation search of H_2O (left panels), CO_2 (middle panels), and CO (right panels) along their optimal wavelength ranges over the 2000 samples of the transmission spectrum (those shown in Figs 4, 5, and 6). Black-solid lines show the CCFs resulting from the cross-correlation search over the original transmission spectrum. Top panels show the results using the Gaussian normalization mode and bottom panels show the results using the frequency filter normalization mode.

3.3 CO detection

We confirm the previous detections of CO (EB23; D. Grant et al. 2023) through this improved cross-correlation methodology, achieving a similar detection confidence than that obtained in EB23. CO is detected in NRS2 detector and in the combination of NRS1 and NRS2 detectors with both normalization modes. Fig. 6 shows the cross-correlation result on the optimal wavelength range of NRS2 detector, (4.6–5.0 μm), using the Gaussian normalization (top panel) and the frequency filter normalization (bottom panel). In this work, we generated only one CO template using a line list containing all CO isotopologues, as here we focus on testing the performance of the technique among different molecules. The selection of different CO isotopologues for the generation of different CO templates and their impact on the cross-correlation results were preliminarily analysed in EB23 and will be further studied in future dedicated articles.

3.4 Monte Carlo Markov Chain sampling

The relative SNR of the CCF peaks shown in Figs 4, 5, and 6 depends on the ‘instantaneous’ noise structure of the transmission spectrum as observed. Therefore, to evaluate the significance of the reported SNRs for the detected molecules we performed a Monte Carlo Markov Chain (MCMC) procedure, where each point on the transmission spectrum is resampled randomly within the transit depth uncertainties, generating 2000 samples of the transmission spectrum. Then, we performed the cross-correlation of each sampled transmission spectrum with each molecular template and converted the resulting CCF to SNR following the same methodology described in Section 2.

In Fig. 7, we show the results of the cross-correlations search of H_2O, CO_2, and CO (along their optimal wavelength ranges) for the 2000 transmission spectrum samples, using the Gaussian fitting normalization (top panels) and the frequency filtering normalization (bottom panels).

For each sampled CCF we identified the maximum SNR value of the central peak (SNR_{peak}) and its RV position. Fig. 8 shows the sampled distribution of SNR_{peak} values and the distribution of RV positions at which the maxima of the central peak appear, for each detected molecule and for both normalization modes. To determine the statistical significance of each detection, we measured the median value and 1σ percentiles on each SNR_{peak} distribution. In Table 3, we report the significance of each molecular detection by indicating the median SNR_{peak} and 1σ percentiles for each molecule. These results show that the Gaussian fitting achieves the highest detection significance for all molecules, being the most efficient normalization mode.

4 CONCLUSIONS AND DISCUSSION

We present an improved, generalized cross-correlation methodology for the search of molecular species in *JWST* transmission spectra. We applied this methodology to the WASP-39 b JTEC ERS Program NIRSpec G395H data set.

We demonstrate that cross-correlation effectively detects multiple molecular species in *JWST* transmission spectra, proving particularly effective for molecules with mid-to-low SNR and narrow spectral features (e.g. H_2O, CO). This technique is more sensitive to individual spectral lines than to broad molecular features. The performance of this technique is limited for the detection of broad spectral features because the normalization step removes low-frequency signals characteristic of a specific molecule, e.g. Gaussian-like features in CO_2 around 4.4 μm or SO_2 around 4.0 μm, and slopes in H_2O. However, this fact is also an advantage, since the remaining high-frequency

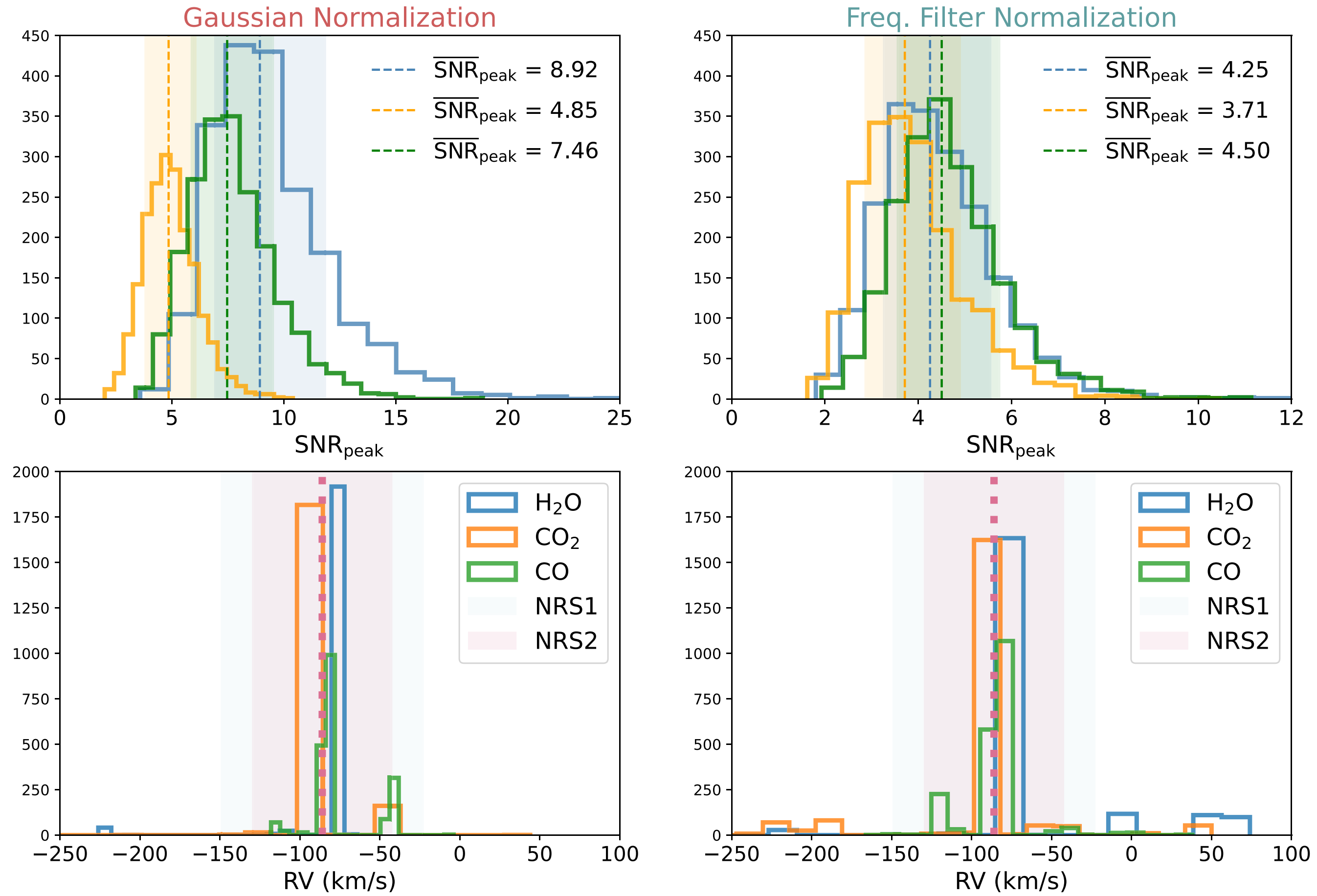

Figure 8. Distribution of the CCF central peak maximum SNR values (SNR_{peak}) and their RV positions from the cross-correlation results over the transmission spectrum samples for each detected molecule and for both normalization modes. The top panels show the SNR_{peak} distribution for each detected molecule, indicating their median values (dashed line) and their 1σ percentiles (shaded regions), and the bottom panels show the distribution of the RV position of the central peak's maximum value for each molecule compared to the expected systemic velocity of WASP-39b relative to *JWST* at the time of the observations (-87 km s^{-1}, pink dashed line) and the mean RV pixel size of each detector (shaded regions). The left panels show the results for the Gaussian normalization mode and the right panels show the results for the frequency filter normalization mode.

Table 3. Median of SNR_{peak} for each detected molecule and their 1σ uncertainties. The columns show this result for both normalization modes (Gaussian and frequency filter).

Molecule	SNR_{peak} (Gaussian)	SNR_{peak} (Freq. filter)
H_2O	$8.9^{+2.9}_{-2.0}$	$4.3^{+1.3}_{-1.0}$
CO_2	$4.9^{+1.2}_{-1.0}$	$3.7^{+1.2}_{-0.8}$
CO	$7.5^{+2.1}_{-1.6}$	$4.5^{+1.2}_{-1.0}$

pattern signals of those absorptions are less prone to systematics (e.g. bumps in the spectrum).

We confirm the detections of H_2O, CO_2, and CO, achieving detection significances of $\text{SNR}_{\text{peak}} = 8.9^{+2.9}_{-2.0}$, $\text{SNR}_{\text{peak}} = 4.9^{+1.2}_{-1.0}$, and $\text{SNR}_{\text{peak}} = 7.5^{+2.1}_{-1.6}$, respectively. Our analysis shows that – for all detected molecules – the use of the Gaussian normalization mode for the generation of templates leads to higher detection significances than the frequency filter normalization mode.

We find that the detection significance does not have any substantial variation with the selection of the cross-correlation RV step. However, we note a slight increase in the uncertainties in Table A1, especially in the case of H_2O for the 64 km s^{-1} step, which might be related to an aliasing effect that depends on the density of the spectral lines. This effect needs to be explored in more detail in future studies, in particular studying how it might affect some molecules more than others and how it impacts the SNR values estimated using the CCF peak divided by the standard deviation of a predefined baseline.

Our results show that the molecular detections are more efficient when doing the cross-correlation search over the optimal wavelength ranges, which are defined as the spectral regions where the contribution to the opacity of each molecular feature is expected to be dominant or considerable considering the relative contribution of the rest of molecules present in the spectrum.

Exoplanet atmospheres are composed by multiple molecular species, and the signals of multiple molecules can overlap, specially at the lower spectral resolution levels produced by *JWST* observations. This introduces additional noise in the CCFs that can lower the significance of the detections. However, this is not an issue in terms of identifying the individual absorptions pattern of an specific molecule. While this study focuses on identifying individual molecules in a transmission spectrum through cross-correlations, the developed methodology can be expanded to the search for combinations of molecules, which would allow measuring (or placing limits) in abundance ratios.

ACKNOWLEDGEMENTS

This work is based on observations made with the NASA/ESA/CSA *JWST*. The data were obtained from the Mikulski Archive for Space Telescopes at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-03127 for *JWST*. These observations are associated with programme *JWST*-ERS-01366. Support for program *JWST*-ERS-01366 was provided by NASA through a grant from the Space Telescope Science Institute. This work was supported by grant JWST-ERS-01366.033-A. EE-B and EP acknowledge funding from the Spanish Ministry of Economics and Competitiveness through project PGC2018-098153-B-C31. EE-B acknowledges financial support from the European Union and the State Agency of Investigation of the Spanish Ministry of Science and Innovation (MICINN) under the grant PRE2020-093107 of the Pre-Doc Program for the Training of Doctors (FPI-SO) through FSE funds. JK acknowledges financial support from Imperial College London through an Imperial College Research Fellowship grant. CC acknowledges support by ANID BASAL project FB210003. LD acknowledges support from the KU Leuven IDN grant IDN/19/028 and the MC-ITN CHAMELEON grant. This research was supported by the Excellence Cluster ORIGINS which is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC-2094 – 390783311. GM acknowledges financial support from the Severo Ochoa grant CEX2021-001131-S and from the Ramón y Cajal grant RYC2022-037854-I funded by MCIN/AEI/1144 10.13039/501100011033 and FSE+. This material is based upon work supported by the National Aeronautics and Space Administration under Agreement No. 80NSSC21K0593 for the programme 'Alien Earths'. The results reported herein benefited from collaborations and/or information exchange within NASA's Nexus for Exoplanet System Science (NExSS) research coordination network sponsored by NASA's Science Mission Directorate. J-MD acknowledges the research programe VIDI New Frontiers in Exoplanetary Climatology with project number 614.001.601, which is (partly) financed by the Dutch Research Council (NWO).

DATA AVAILABILITY

The data were obtained from the Mikulski Archive for Space Telescopes at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-03127 for *JWST*. The specific observations can be accessed via http://dx.doi.org/10.17909/gdm2-0q65. These observations are associated with programme JWST-ERS-01366.

REFERENCES

Alderson L. et al., 2023, Nature, 614, 664
Banerjee A. et al., 2024, ApJ, 975, L11
Basilicata M. et al., 2025, A&A, 695, A107
Bell T. J. et al., 2023, Nature, 623, 709
Benneke B. et al., 2024, preprint (arXiv:2403.03325)
Brogi M., Snellen I. A. G., de Kok R. J., Albrecht S., Birkby J., de Mooij E. J. W., 2012, Nature, 486, 502
Brogi M., de Kok R. J., Albrecht S., Snellen I. A. G., Birkby J. L., Schwarz H., 2016, ApJ, 817, 106
Butterworth S., 1930, Experimental Wireless and the Wireless Engineer, 7, 536
Carleo I. et al., 2022, AJ, 164, 101
Carnall A. C., 2017, preprint (arXiv:1705.05165)
Czesla S., Schröter S., Schneider C. P., Huber K. F., Pfeifer F., Andreasen D. T., Zechmeister M., 2019, Astrophysics Source Code Library, record ascl:1906.010
Esparza-Borges E. et al., 2023, ApJ, 955, L19 (EB23)
Faedi F. et al., 2011, A&A, 531, A40
Feinstein A. D. et al., 2023, Nature, 614, 670
Gandhi S., de Regt S., Snellen I., Zhang Y., Rugers B., van Leur N., Bosschaart Q., 2023, ApJ, 957, L36
Gordon I. E. et al., 2022, J. Quant. Spec. Radiat. Transf., 277, 107949
Grant D. et al., 2023, ApJL, 949, L15
Hargreaves R. J. et al., 2019, J. Quant. Spec. Radiat. Transf., 232, 35
Hargreaves R. J., Gordon I. E., Rey M., Nikitin A. V., Tyuterev V. G., Kochanov R. V., Rothman L. S., 2020, ApJS, 247, 55
Hargreaves R. J., Gordon I. E., Huang X., Toon G. C., Rothman L. S., 2025, J. Quant. Spec. Radiat. Transf., 333, 109324
Hoeijmakers H. J. et al., 2019, A&A, 627, A165
Hoeijmakers H. J. et al., 2020, A&A, 641, A123
Holmberg M., Madhusudhan N., 2024, A&A, 683, L2
JWST Transiting Exoplanet Community Early Release Science Team, 2023, Nature, 614, 649
Kanumalla S. K. T., Line M., Weiner Mansfield M., Welbanks L., 2025, in AAS Meeting Abstracts. p. 316.06
Kirk J., Wheatley P. J., Louden T., Doyle A. P., Skillen I., McCormac J., Irwin P. G. J., Karjalainen R., 2017, MNRAS, 468, 3907
Kirk J. et al., 2021, AJ, 162, 34
Kirk J. et al., 2025, MNRAS 537 3027
Kirk J. et al., 2025, MNRAS, 537, 3027
Li G., Gordon I. E., Rothman L. S., Tan Y., Hu S.-M., Kassi S., Campargue A., Medvedev E. S., 2015, ApJS, 216, 15
Mancini L. et al., 2018, A&A, 613, A41
May E. M. et al., 2023, ApJ, 959, L9
Mollière P., Snellen I. A. G., 2019, A&A, 622, A139
Nortmann L. et al., 2025, A&A, 693, A213
Rodler F., Lopez-Morales M., Ribas I., 2012, ApJ, 753, L25
Rothman L. S. et al., 2010, J. Quant. Spec. Radiat. Transf., 111, 2139
Rustamkulov Z. et al., 2023, Nature, 614, 659
Schmidt S. P. et al., 2025, preprint (arXiv:2501.18477)
Sing D. K. et al., 2016, Nature, 529, 59
Snellen I. et al., 2015, A&A, 576, A59
Snellen I. A. G., de Kok R. J., de Mooij E. J. W., Albrecht S., 2010, Nature, 465, 1049
Stangret M., Casasayas-Barris N., Pallé E., Yan F., Sánchez-López A., López-Puertas M., 2020, A&A, 638, A26
Tabernero H. M. et al., 2021, A&A, 646, A158
Tsai S.-M. et al., 2023, Nature, 617, 483
Virtanen P. et al., 2020, Nat. Methods, 17, 261
Yan F. et al., 2019, A&A, 632, A69
Yang Y., Chen G., Wang S., Yan F., 2024, AJ, 167, 36

APPENDIX A: IMPACT OF RADIAL VELOCITY SAMPLING IN THE SNR OF THE DETECTIONS

We repeated the analysis of the detected molecules varying the RV step in the cross-correlations, testing 6, 20, 43, and 64 $\mathrm{km\,s^{-1}}$. We found that the detection significance does not have any substantial variation with the selection of the cross-correlation RV step. We note a slight increase in the uncertainties in Table A1, especially in the case of H_2O for the 64 $\mathrm{km\,s^{-1}}$ step, which might be related to an aliasing effect that depends on the density of the spectral lines. This effect needs to be explored in more detail in future studies, in particular studying how it might affect some molecules more than others and how it impacts the SNR values estimated using the CCF peak divided by the standard deviation of a predefined baseline.

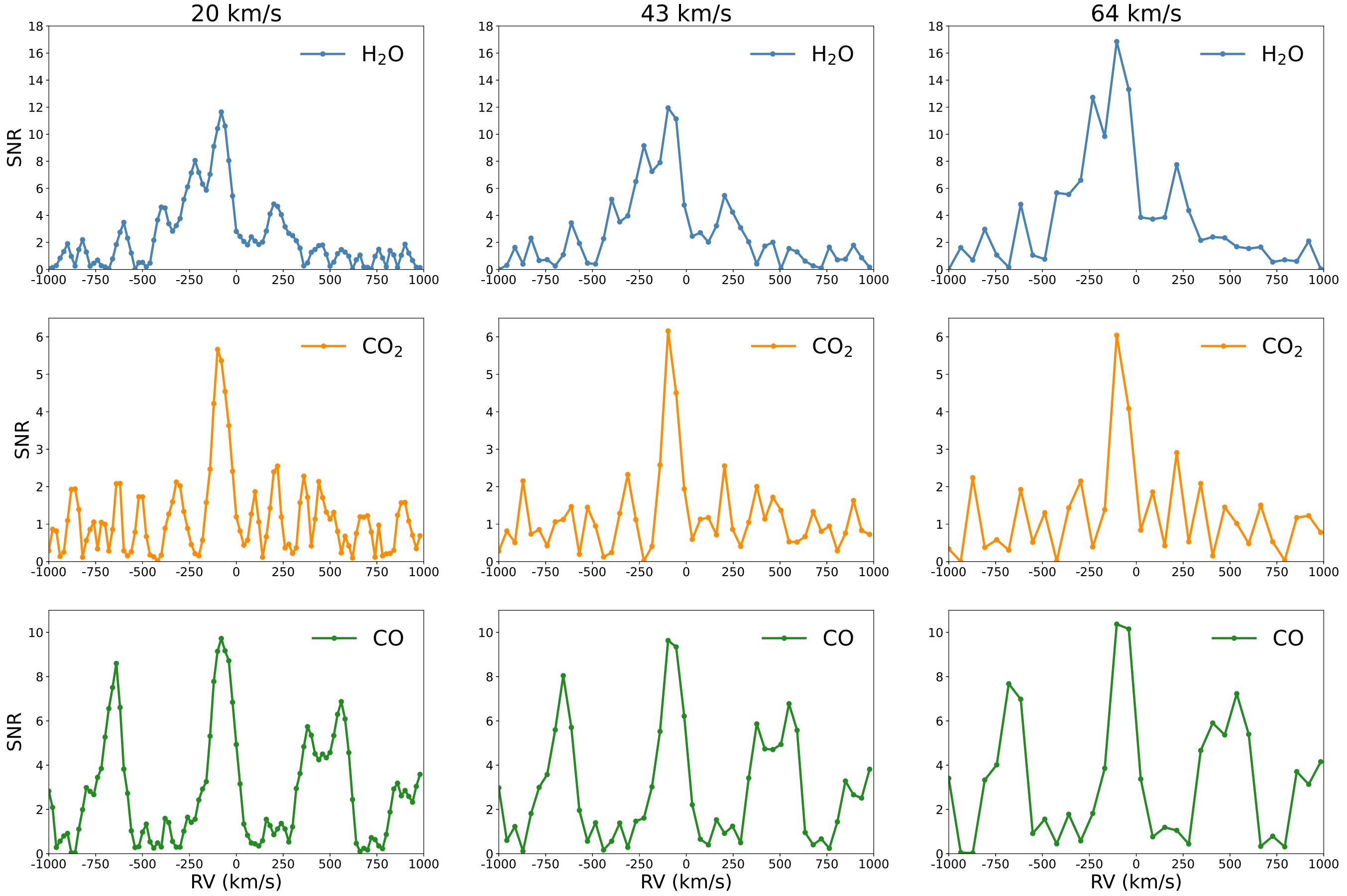

Figure A1. Signal-to-noise ratio (SNR) of the cross-correlation function (CCF) as a function of radial velocity (RV) for H_2O (top row), CO_2 (middle row), and CO (bottom row) across three different CCF sampling intervals: 20 km s^{-1} (left), 43 km s^{-1} (middle), and 64 km s^{-1} (right). All panels use the Gaussian normalization mode and apply the cross-correlation over the optimal wavelength ranges defined for each molecule. The consistent detection peaks at the expected velocities across different RV steps demonstrate the robustness of the molecular detections and show that SNR estimates are not significantly affected by the choice of velocity sampling.

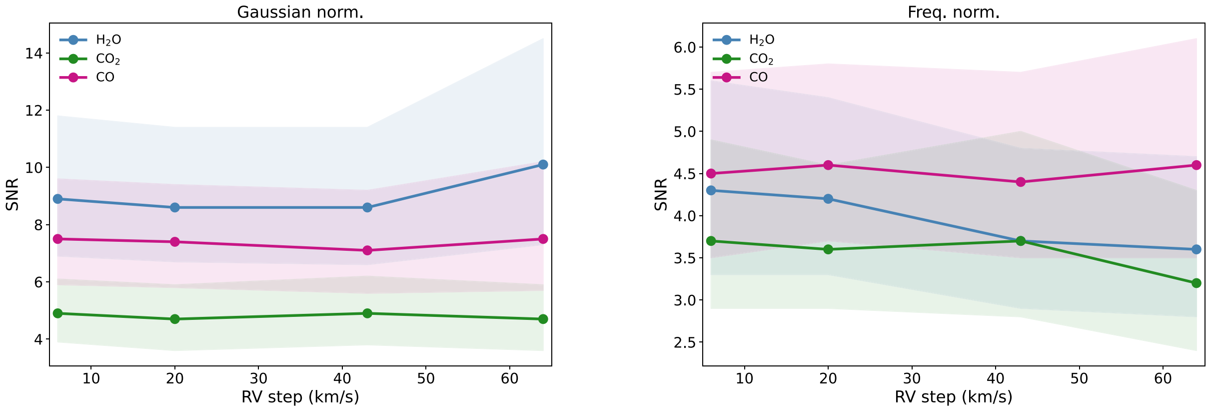

Figure A2. Median SNR values of the CCF central peaks for H_2O (blue), CO_2 (green), and CO (magenta) as a function of RV sampling step size (6–64 km s^{-1}), using Gaussian normalization (left) and frequency-filter normalization (right). Shaded regions indicate the 1σ intervals from 2000 Monte Carlo samples of the transmission spectrum. The consistency of the SNR values across different RV steps confirms that the detection significance is not strongly affected by the choice of velocity resolution, validating the robustness of the molecular detections.

Table A1. Peak signal-to-noise ($\mathrm{SNR_{peak}}$) values for the cross-correlation detection of H_2O, CO_2, and CO across different RV step sizes (6–64 $\mathrm{km\,s^{-1}}$), using Gaussian (Gauss) and frequency-filter (Freq) normalization modes.

	$\mathrm{SNR_{peak}}$							
Molecule	64 $\mathrm{km\,s^{-1}}$ (Gauss)	64 $\mathrm{km\,s^{-1}}$ (Freq)	43 $\mathrm{km\,s^{-1}}$ (Gauss)	43 $\mathrm{km\,s^{-1}}$ (Freq)	20 $\mathrm{km\,s^{-1}}$ (Gauss)	20 $\mathrm{km\,s^{-1}}$ (Freq)	6 $\mathrm{km\,s^{-1}}$ (Gauss)	6 $\mathrm{km\,s^{-1}}$ (Freq)
H_2O	$10.1^{+4.4}_{-2.8}$	$3.6^{+1.1}_{-0.8}$	$8.6^{+2.8}_{-2.0}$	$3.7^{+1.1}_{-0.8}$	$8.6^{+2.8}_{-1.9}$	$4.2^{+1.2}_{-0.9}$	$8.9^{+2.9}_{-2.0}$	$4.3^{+1.3}_{-1.0}$
CO_2	$4.7^{+1.2}_{-1.1}$	$3.2^{+1.1}_{-0.8}$	$4.9^{+1.3}_{-1.1}$	$3.7^{+1.3}_{-0.9}$	$4.7^{+1.2}_{-1.1}$	$3.6^{+1.0}_{-0.7}$	$4.9^{+1.2}_{-1.0}$	$3.7^{+1.2}_{-0.8}$
CO	$7.5^{+2.7}_{-1.8}$	$4.6^{+1.5}_{-1.1}$	$7.1^{+2.1}_{-1.5}$	$4.4^{+1.3}_{-0.9}$	$7.4^{+2.0}_{-1.6}$	$4.6^{+1.2}_{-0.9}$	$7.5^{+2.1}_{-1.6}$	$4.5^{+1.2}_{-1.0}$

APPENDIX B: CROSS-SECTION MOLECULAR CONTRIBUTIONS

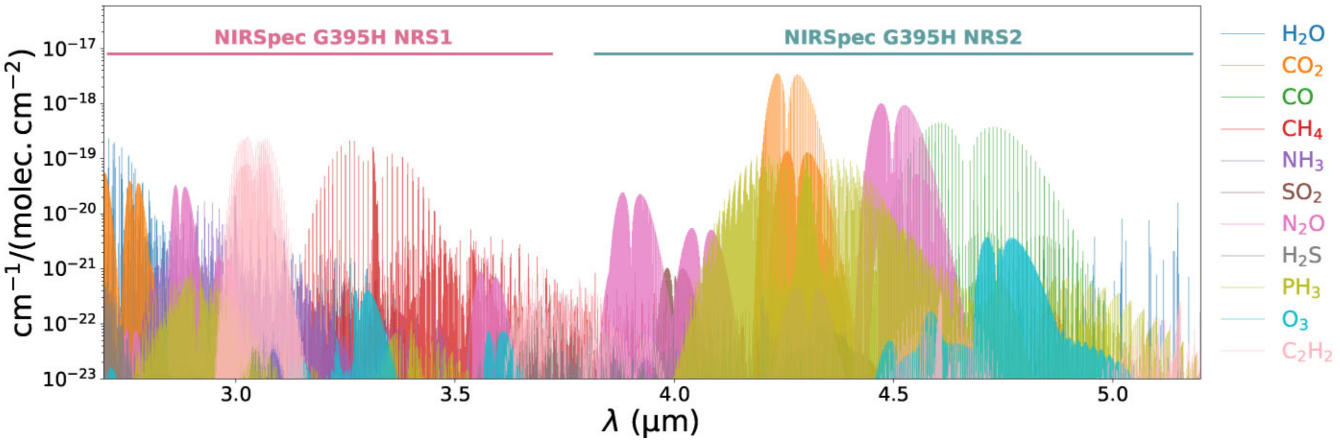

Figure B1. Line-intensities of all the molecules considered in this study. The cross-sections are obtained from the HITEMP and HITRAN data bases as listed in Table 1.

APPENDIX C: TEMPLATE–TEMPLATE CCFS AND BASELINE DEFINITIONS FOR THE FREQUENCY FILTER NORMALIZATION

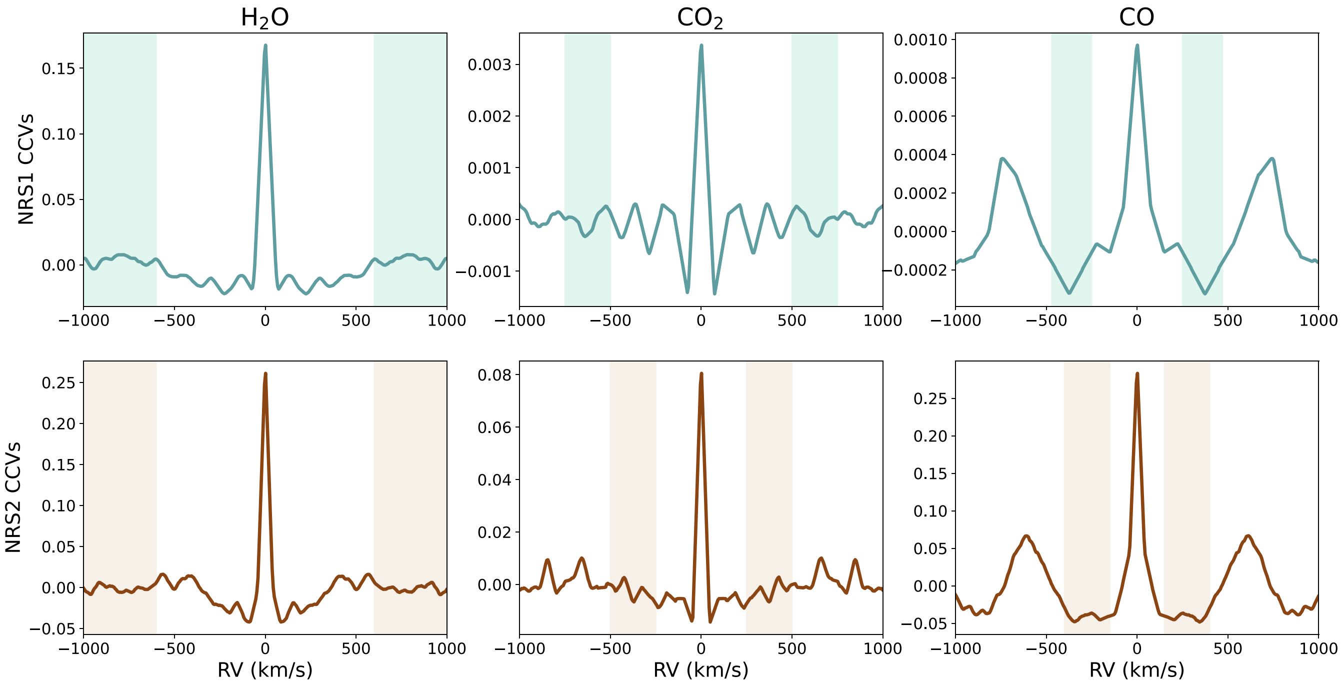

Figure C1. Definition of the baseline regions (shaded regions) with respect to the template–template CCF within the optimal wavelength ranges using the frequency filter normalization on the NRS1 (top panels) and the NRS2 (bottom panels) for each detected molecule.

APPENDIX D: TEMPLATE–TEMPLATE CCFS AND BASELINE DEFINITIONS FOR THE NON-DETECTED MOLECULES

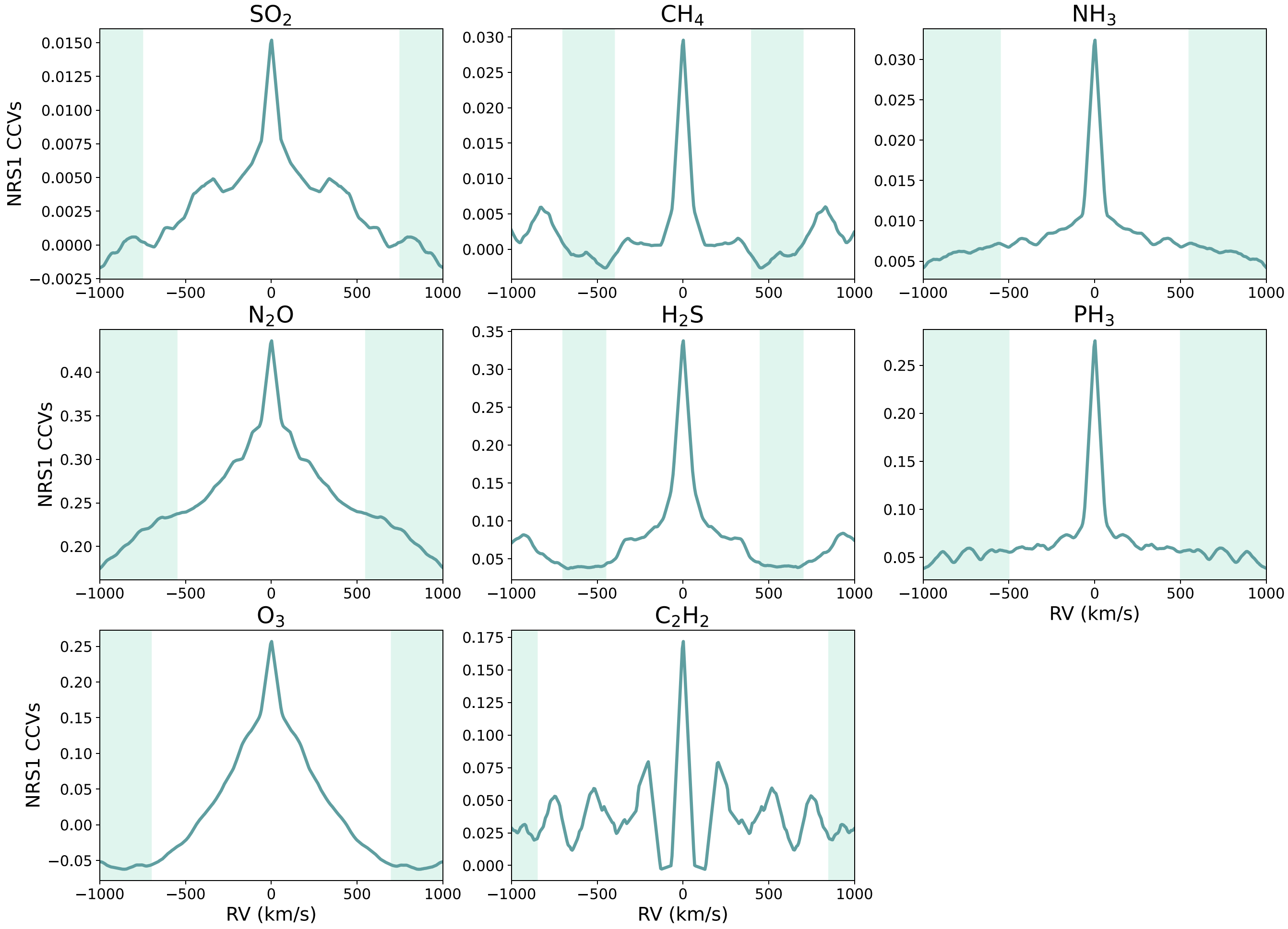

Figure D1. Definition of the baseline regions (shaded regions) with respect to the template–template CCF within the optimal wavelength ranges using the Gaussian normalization on NRS1 for the non-detected molecules.

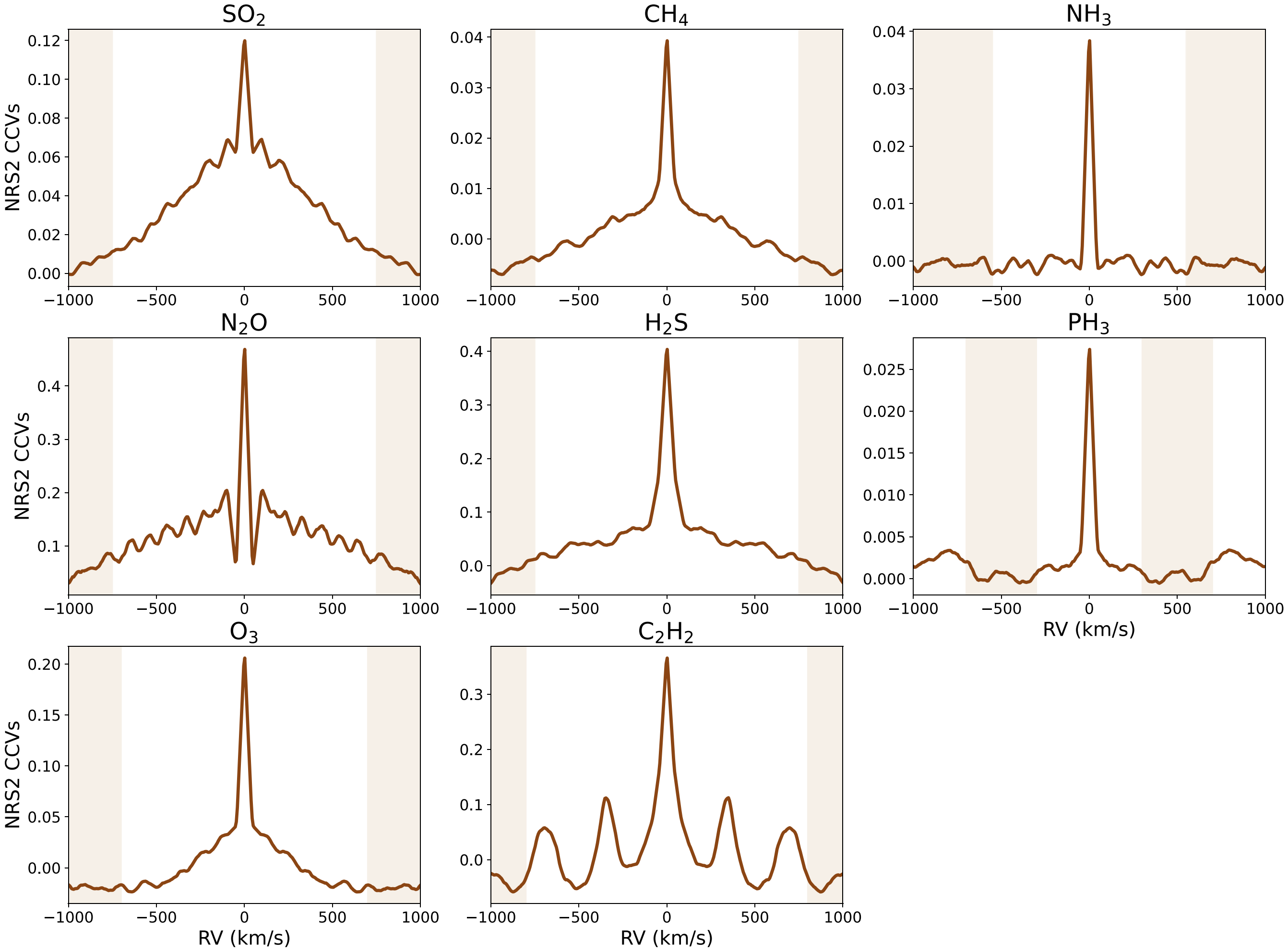

Figure D2. Definition of the baseline regions (shaded regions) with respect to the template–template CCF within the optimal wavelength ranges using the Gaussian normalization on NRS2 for the non-detected molecules.

Table D1. Definition of the RV baselines for non-detected molecules.

Molecule	NRS1 RV (km s^{-1})	NRS2 RV (km s^{-1})
H_2O	[−1000, −750] ∪ [750, 1000]	[−800, −600] ∪ [600, 800]
CO_2	[−1000, −550] ∪ [550, 1000]	[−1000, −550] ∪ [550, 1000]
CO	[−470, −250] ∪ [250, 470]	[−400, −150] ∪ [150, 400]
SO_2	[−1000, −750] ∪ [750, 1000]	[−1000, −750] ∪ [750, 1000]
CH_4	[−700, −400] ∪ [400, 700]	[−1000, −750] ∪ [750, 1000]
NH_3	[−1000, −550] ∪ [550, 1000]	[−1000, −550] ∪ [550, 1000]
N_2O	[−1000, −550] ∪ [550, 1000]	[−1000, −750] ∪ [750, 1000]
H_2S	[−700, −450] ∪ [450, 700]	[−1000, −750] ∪ [750, 1000]
PH_3	[−1000, −500] ∪ [500, 1000]	[−700, −300] ∪ [300, 700]
O_3	[−1000, −700] ∪ [700, 1000]	[−1000, −700] ∪ [700, 1000]
C_2H_2	[−1000, −850] ∪ [850, 1000]	[−1000, −800] ∪ [800, 1000]

APPENDIX E: RESULTS OF THE SYSTEMATIC CROSS-CORRELATION SEARCH THROUGHOUT NRS1 AND NRS2 DETECTORS.

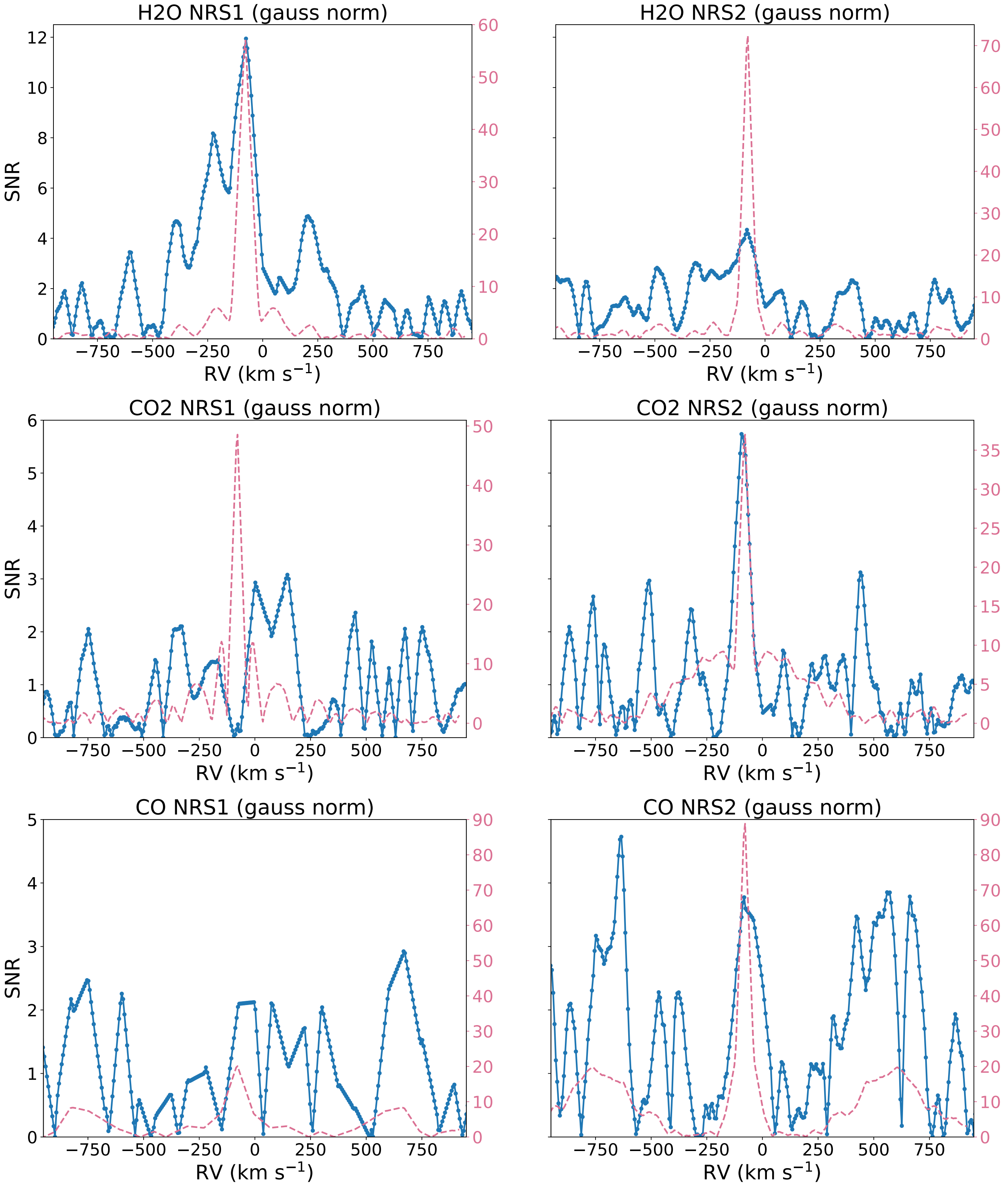

Figure E1. Results of the systematic cross-correlation search of H_2O, CO_2, and CO along the whole wavelength range of NRS1 and NRS2 detectors using Gaussian normalization. The blue dash–dotted line shows the data–template CCF and the pink dashed line shows the template–template CCF.

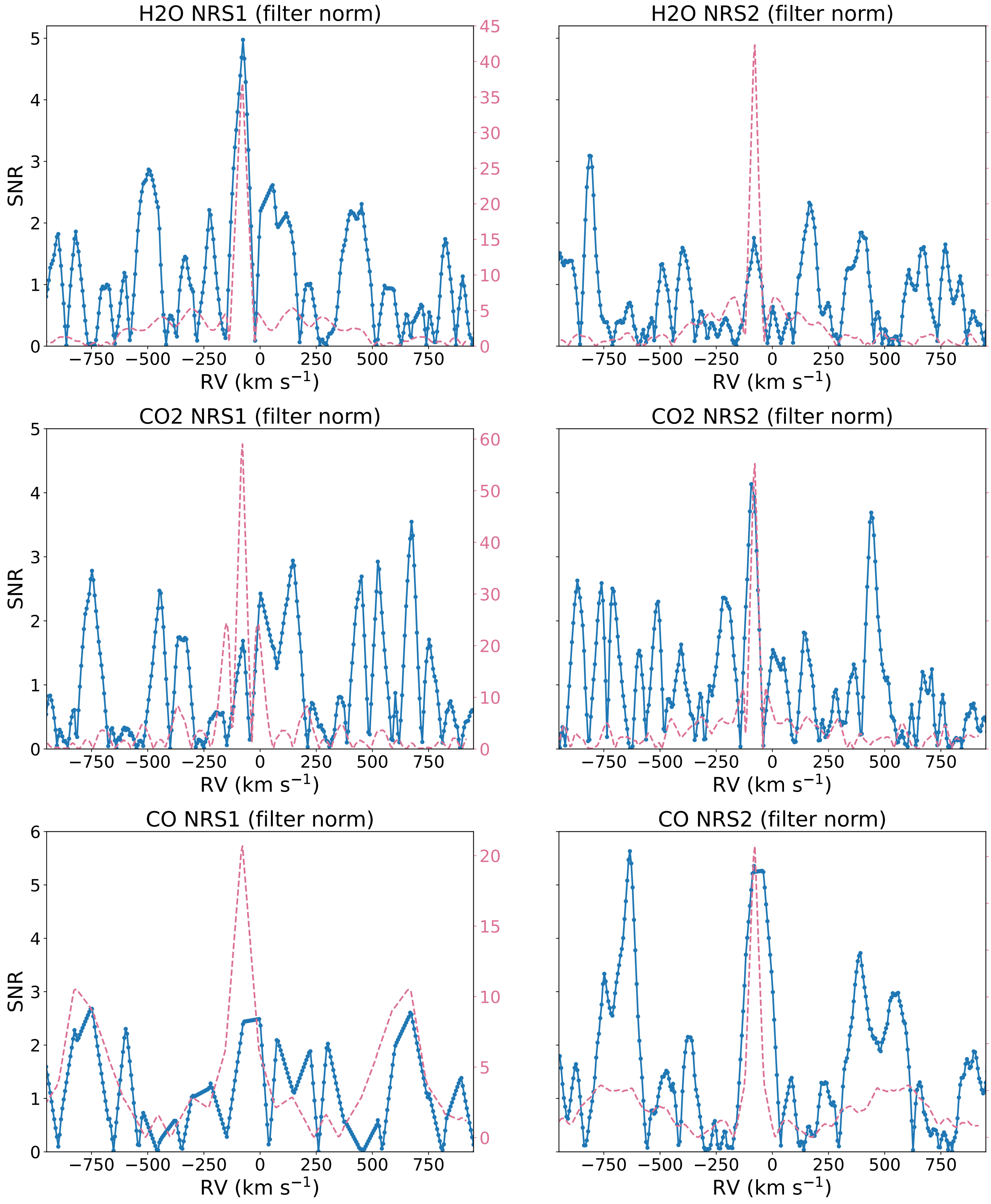

Figure E2. Results of the systematic cross-correlation search of H_2O, CO_2, and CO along the whole wavelength range of NRS1 and NRS2 detectors using frequency-filter normalization. The blue dash–dotted line shows the data–template CCF and the pink dashed line shows the template–template CCF.

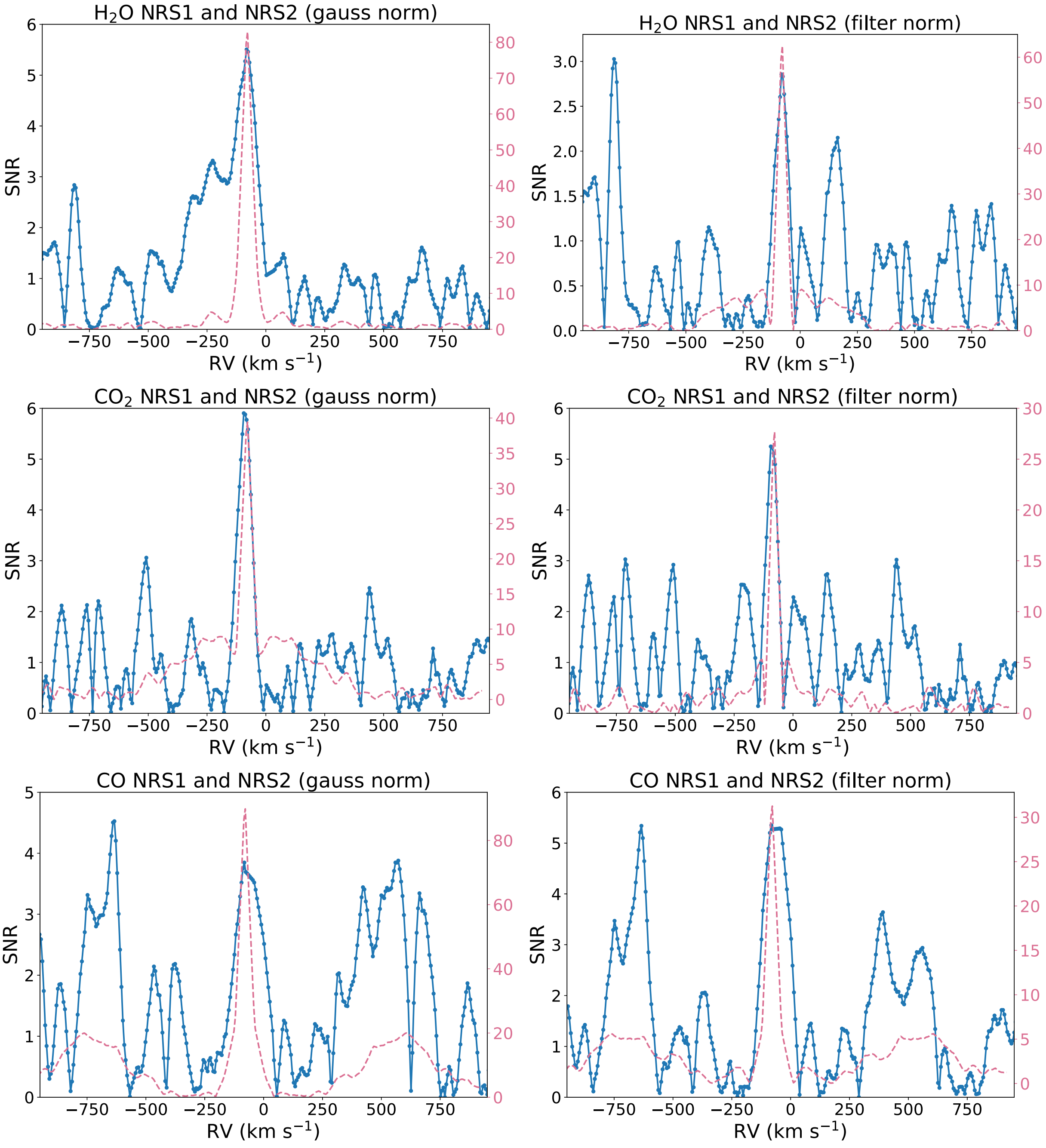

Figure E3. Results of the systematic cross-correlation search of H_2O, CO_2, and CO along the whole wavelength range of NRS1 and NRS2 detectors using both normalizations. The blue dash–dotted line shows the data–template CCF and the pink dashed line shows the template-template CCF.

[1] Instituto de Astrofísica de Canarias, E-38200 San Cristóbal de La Laguna, Tenerife, Spain
[2] Departamento de Astrofísica, Universidad de La Laguna, E-38206 San Cristóbal de La Laguna, Tenerife, Spain
[3] Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA
[4] Center for Astrophysics | Harvard & Smithsonian, Atomic and Molecular Physics Division, Cambridge, MA 02138, USA
[5] Department of Physics, Imperial College London, Prince Consort Road, London SW7 2AZ, UK
[6] Instituto de Astrofisica, Departamento de Fisica y Astronomia, Facultad de Ciencias Exactas, Universidad Andres Bello, Santiago, 8370992, Chile
[7] Department of Physics and Astronomy, University of Kansas, Lawrence, KS 66045, USA
[8] Kapteyn Astronomical Institute, Rijksuniversiteit Groningen, Postbus 800, NL-9700 AV Groningen, the Netherlands
[9] Institute of Astronomy, KU Leuven, Celestijnenlaan 200D, 3001 Leuven, Belgium
[10] Anton Pannekoek Institute for Astronomy, University of Amsterdam, 1098 XH Amsterdam, the Netherlands
[11] Johns Hopkins University, Department of Physics & Astronomy, Baltimore, MD 21218, USA
[12] Paris-Saclay, Université Paris Cité, CEA, CNRS, AIM, F-91191 Gif-sur-Yvette, France
[13] Planetary Sciences Group, Department of Physics and Florida Space Institute, University of Central Florida, Orlando, FL 32826, USA

[14]*Universitäts-Sternwarte, Ludwig-Maximilians-Universität München, Scheinerstrasse 1, D-81679 München, Germany*
[15]*Instituto de Astrofísica de Andalucía (IAA–CSIC), Glorieta de la Astronomía s/n, E-18008 Granada, Spain*
[16]*INAF – Osservatorio Astronomico di Palermo, Piazza del Parlamento, 1, I-90134 Palermo, Italy*
[17]*İstanbul Health and Technology University, Department of Mechatronic Engineering, 34445 İstanbul, Türkiye*
[18]*Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[19]*Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[20]*Astronomy Department and Van Vleck Observatory, Wesleyan University, Middletown, CT 06459, USA*

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