# A formally motivated retrieval framework applied to the high-resolution transmission spectrum of HD 189733 b

ATMOSPHERIC PHYSICS OF EXOPLANETS

Find the full article here!


**Doriann Blain[1,†], Alejandro Sánchez-López[2,3], Paul Mollière[1]**

[1]Max-Planck-Institut für Astronomie

[2]Leiden Observatory, [3]Instituto de Astrofísica de Andalucía


petitRADTRANS 3

## Introduction

**HD 189733 b** is a hot Jupiter ($T_{eq}$ ≈ 1200 K) and is one of the most studied exoplanet to date. It is roughly 13% more massive and larger than Jupiter. It orbits a close K2 V star and has one of the highest TSM (≈ 770). However, in the literature, the values reported for the **$H_2O$ abundance** (-4 to -1.5 ±0.5 $\log_{10}$ MMR) and **$V_{rest}$** (-8 to -2 ±0.5 km/s) **are inconsistent**.

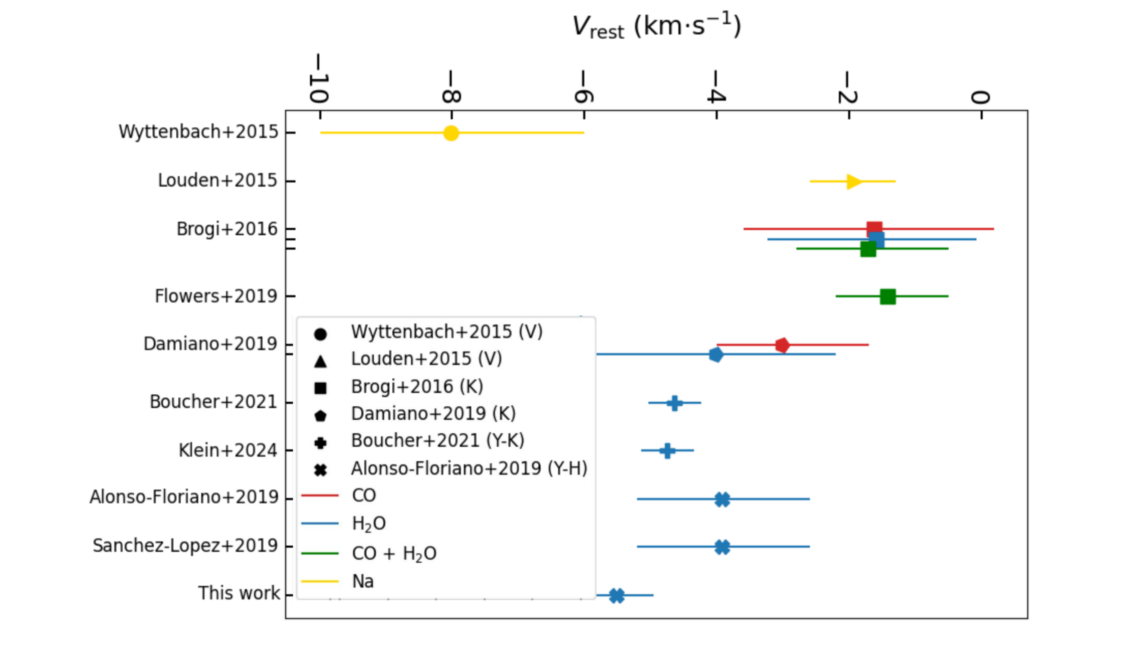


Figure 1: Inferred rest velocities ($V_{rest}$) for HD 189733 b from the literature. Colors indicate from which species the value was inferred; shapes indicate the source of the data.

**Groud-based observations** can benefit from higher resolving powers ($R \approx 100{,}000$) compared to space-based observations (e.g. $R \lesssim 3000$ for the JWST). This allows for unambiguous species detections, and grants access to atmospheric kinematics ($K_P$, $V_{rest}$) from the Doppler effect.

**Bayesian inferences** ("retrievals") are commonly performed to retrieve information such as species abundances, using the log-likelihood function:

$$\ln(L) = -\frac{1}{2}\sum\left(\frac{F - M_\theta}{U}\right)^2,$$

where $F$ is the data, $M_\theta$ is the forward model with parameters $\theta$, and $U$ are the data uncertainties. For space-based observations, $F \approx M_\Theta + N$, where $\Theta$ are the "true" parameters and $N$ is the noise. In that case, calculating $\ln(L)$ is straightforward. For ground-based observations, $F$ contains deformations and telluric lines ($D$) not modelled in $M_\theta$. These are removed by a **preparing pipeline**, which deforms the spectrum (see Fig. 2).

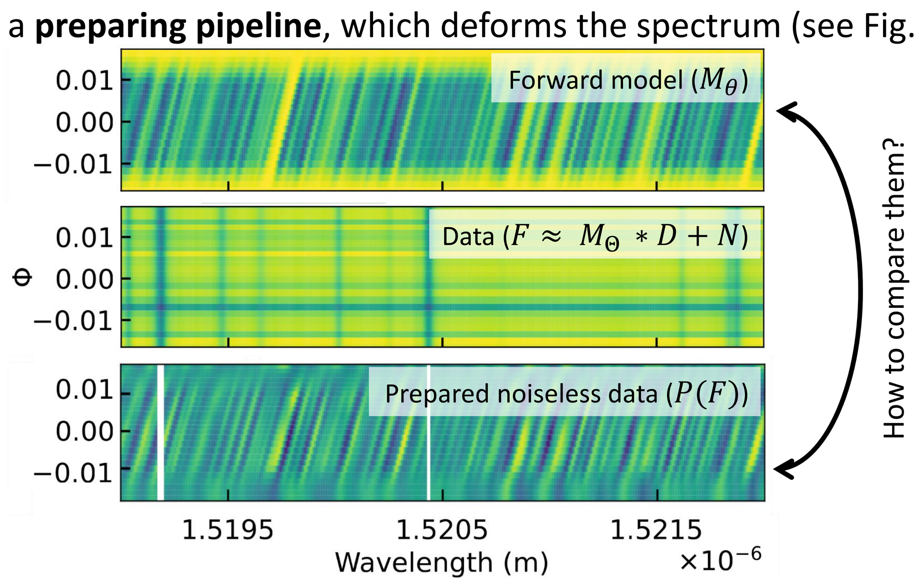


Figure 2: Example HR ground-based (normalized) model, data and prepared data.

**How does this step must be taken into account to perform a retrieval?**
Methods have been developed to tackle this issue (e.g. Brogi & Line 2019, Gibson et al., 2022), but they are not developed in a formal way.

## Objectives

- Develop a formally motivated framework to retrieve high-resolution ground-based data.
- Use the framework on high-resolution data to characterize HD 189733 b's atmosphere.

## Methodology

- **CARMENES** "benchmark" data (0.96—1.71 µm, $R \approx 80{,}000$, Fig. 3) from previous HD 189733 b observations (Alonso-Floriano et al. 2019).
- Upgraded version of the spectral modelling package **petitRADTRANS 3** (Mollière et al., 2019; Nasedkin et al., 2024; Blain et al. 2024 b), including a high-resolution retrieval framework (see Fig. 4).
- Preparing pipeline ("**Polyfit**") using 2nd-order polynomials to fit the data and to remove instrumental deformations as well as telluric and stellar lines.
- Nested sampling retrieval algorithm **MultiNest** (Feroz et al. 2009; Buchner et al., 2014).

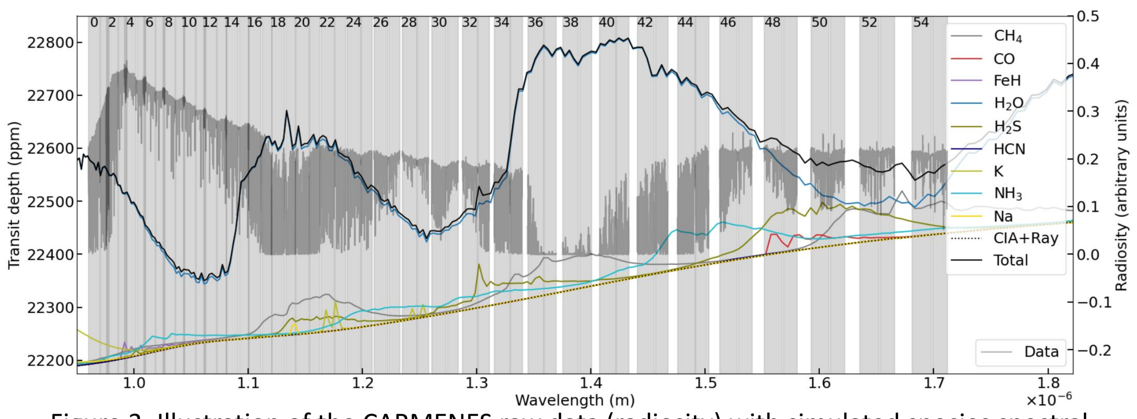


Figure 3: Illustration of the CARMENES raw data (radiosity) with simulated species spectral contributions (transit depth). The CARMENES orders are numbered and indicated as grey areas.

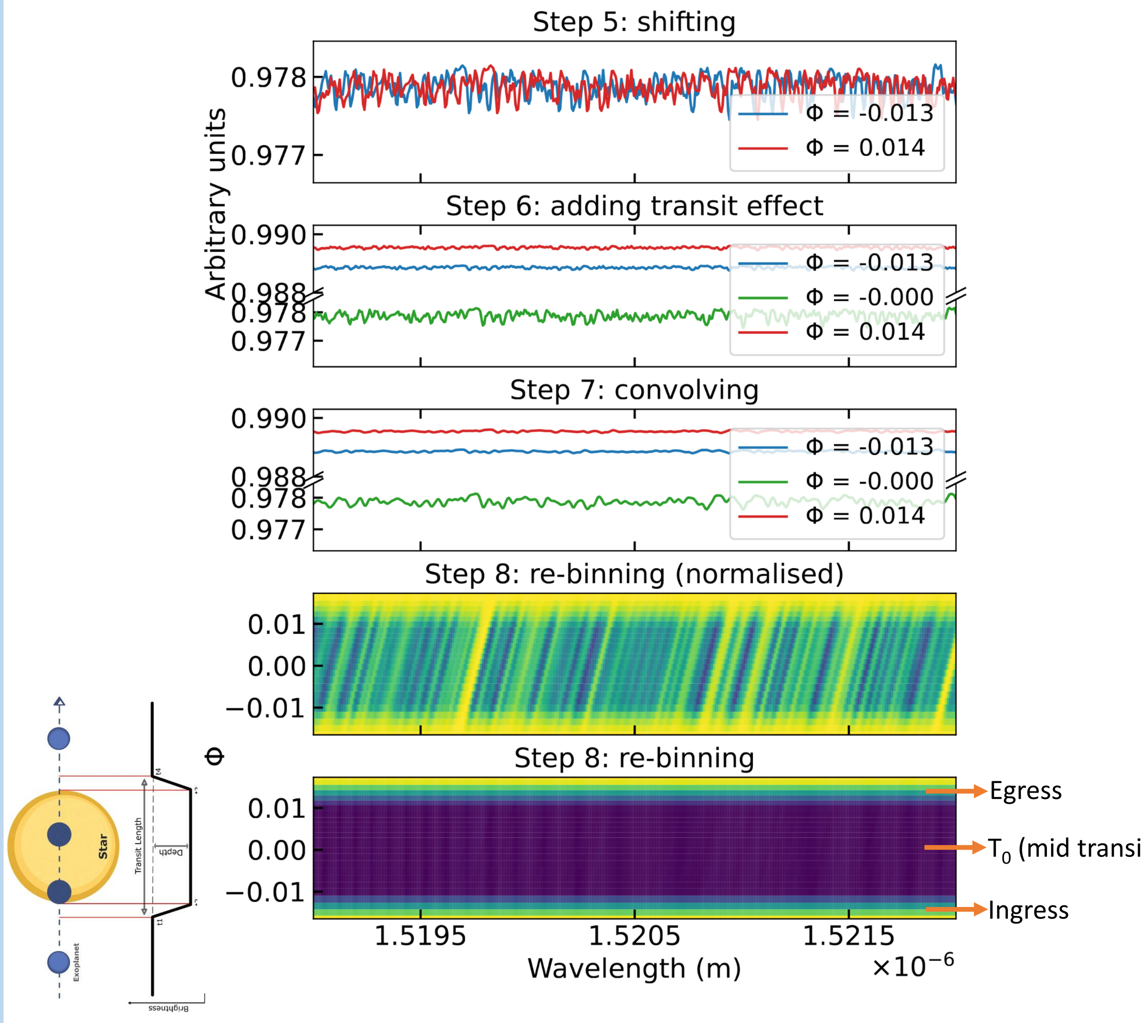


Figure 4: Illustration of the model construction process. Steps 1—4 don't involve HR operations and are not represented. In the penultimate row, the spectra have been normalized over wavelengths for illustrative purposes. The bottom panel represents a forward model without modification.

## Log-likelihood function for HR data

### Summarized demonstration

**Assumption 1**: the HR ground-based data can be represented as:

$$F = M_\Theta * D + N,$$

where $D$ represents the instrumental deformations, as well as the telluric and stellar lines. All the terms above are 3D matrices (order, exp., λ).

**Polyfit** divides the data by 2 polynomials fits (over λ, then over exposures). Hence, its effect on the data is:

$$P_R(F) \equiv F * R_F,$$

where $R_F$ ("preparing matrix") is the inverse of the product of the 2 fits.

**Assumption 2**: $R_F \approx A / D + B$, where $A$ and $B$ represent the pipeline's imperfections. $A$ depends only on $M_\Theta$, $B$ depends on $M_\Theta$, $D$ and $N$.

**Assumption 3**: $A/D \gg B$, i.e., the pipeline has negligible residuals.

**Assumption 4**: at the end of the retrieval, $M_\theta \approx M_\Theta$ ($\theta \to \Theta$).

With all the assumptions above, it can be shown that the log-likelihood function **when using Polyfit-like pipelines** is:

$$\boxed{\ln(L) \approx -\frac{1}{2}\sum\left(\frac{P_R(F) - P_R(M_\theta)}{U_R}\right)^2,}$$

where $U_R = U * |R_F| * \sqrt{n}$ are the prepared data uncertainties and $n$ is the product of the variance correction factors of the 2 fits. The above function **is not valid when using e.g. SysRem** (Tamuz et al., 2005), where the systematics are subtracted, not divided.

### Tests

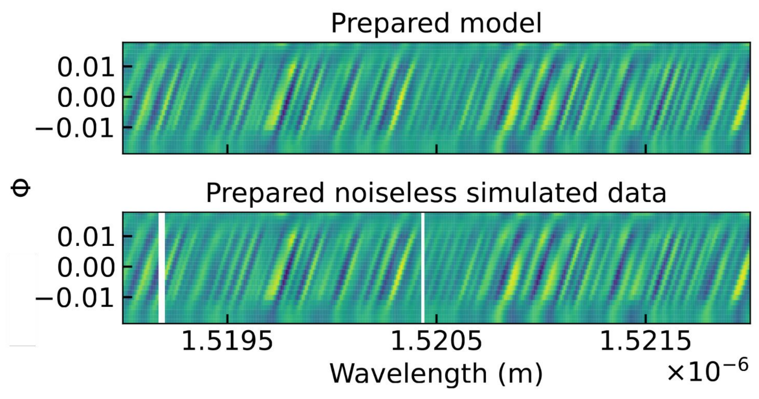


Figure 5: Comparison between a forward model and noiseless simulated data, prepared by Polyfit. The prepared noiseless simulated data are the same as in the middle row of Fig. 2 (noise removed).

Pipeline biases can be inferred with a retrieval on *noiseless* data.

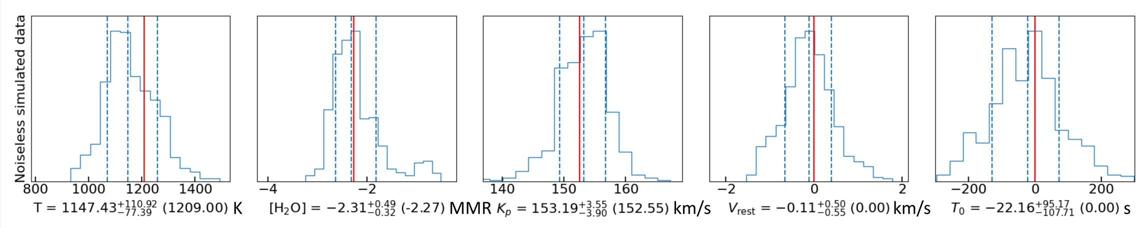

Figure 6: Posterior probability distributions for noiseless simulated data. The solid red lines correspond to the model input values (also indicated between parenthesis).

## Retrieving the mid-transit time ($T_0$)

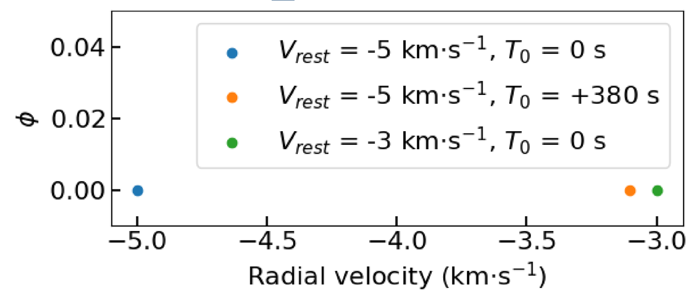


Figure 7: rest velocities at mid transit assuming HD 189733 b's $K_P$. An offset in the mid transit time (orange), i.e. from **inaccurate data timestamps**, can **bias** the retrieved $V_{rest}$. If the transit effect is modelled, $V_{rest}$ and $T_0$ are **not degenerated**.

## Results

### CCF analysis on CARMENES data with Polyfit

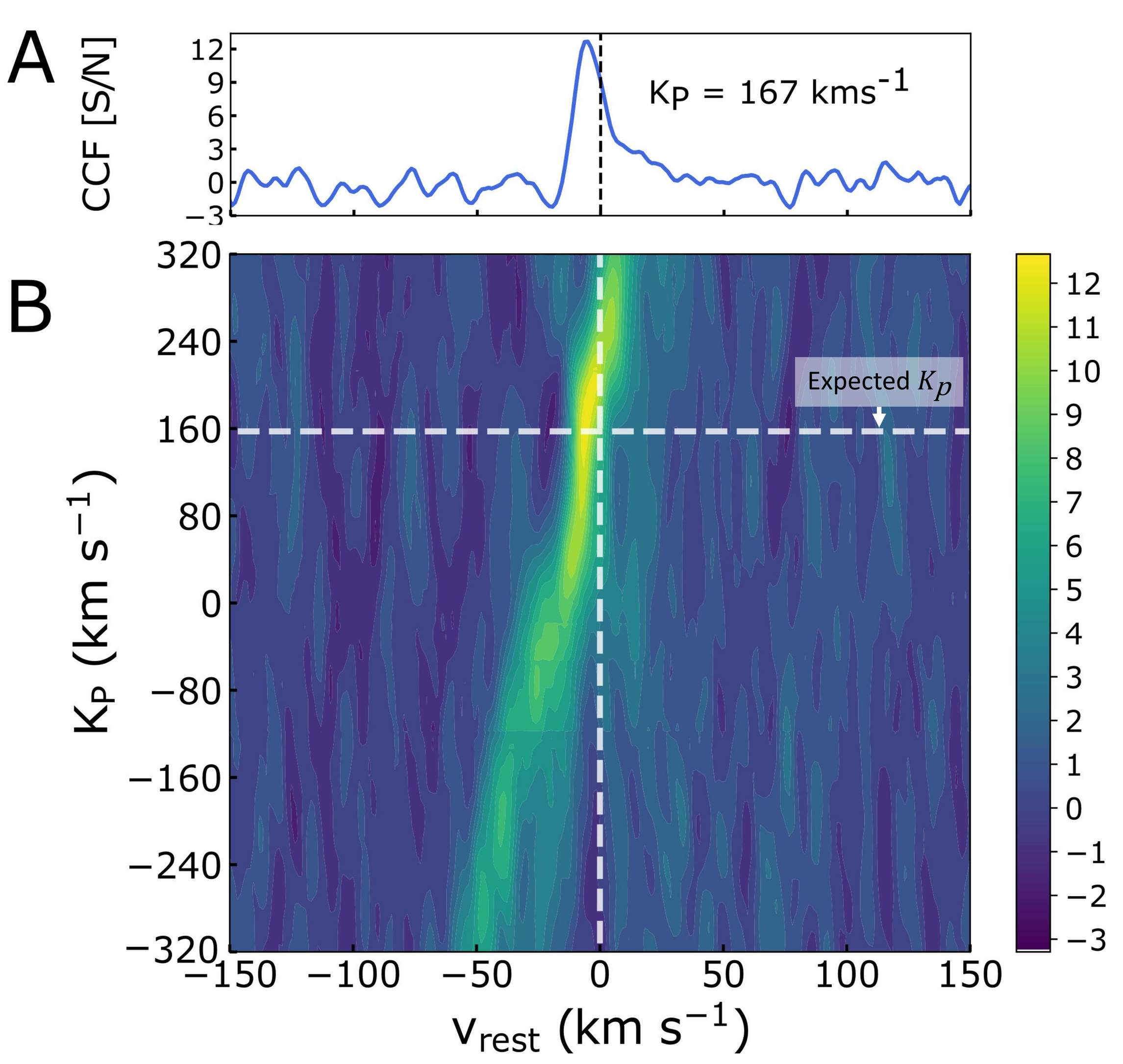


Figure 8: A: slice through the maximum significance CCF (167 km/s) showing the detected CCF peak. B: Cross-correlation map (expressed as SNR) of potential signals for H2O with respect to the exoplanet rest-frame velocity (horizontal axis) and Kp (vertical axis).

## Results

### Retrieval on CARMENES data with Polyfit

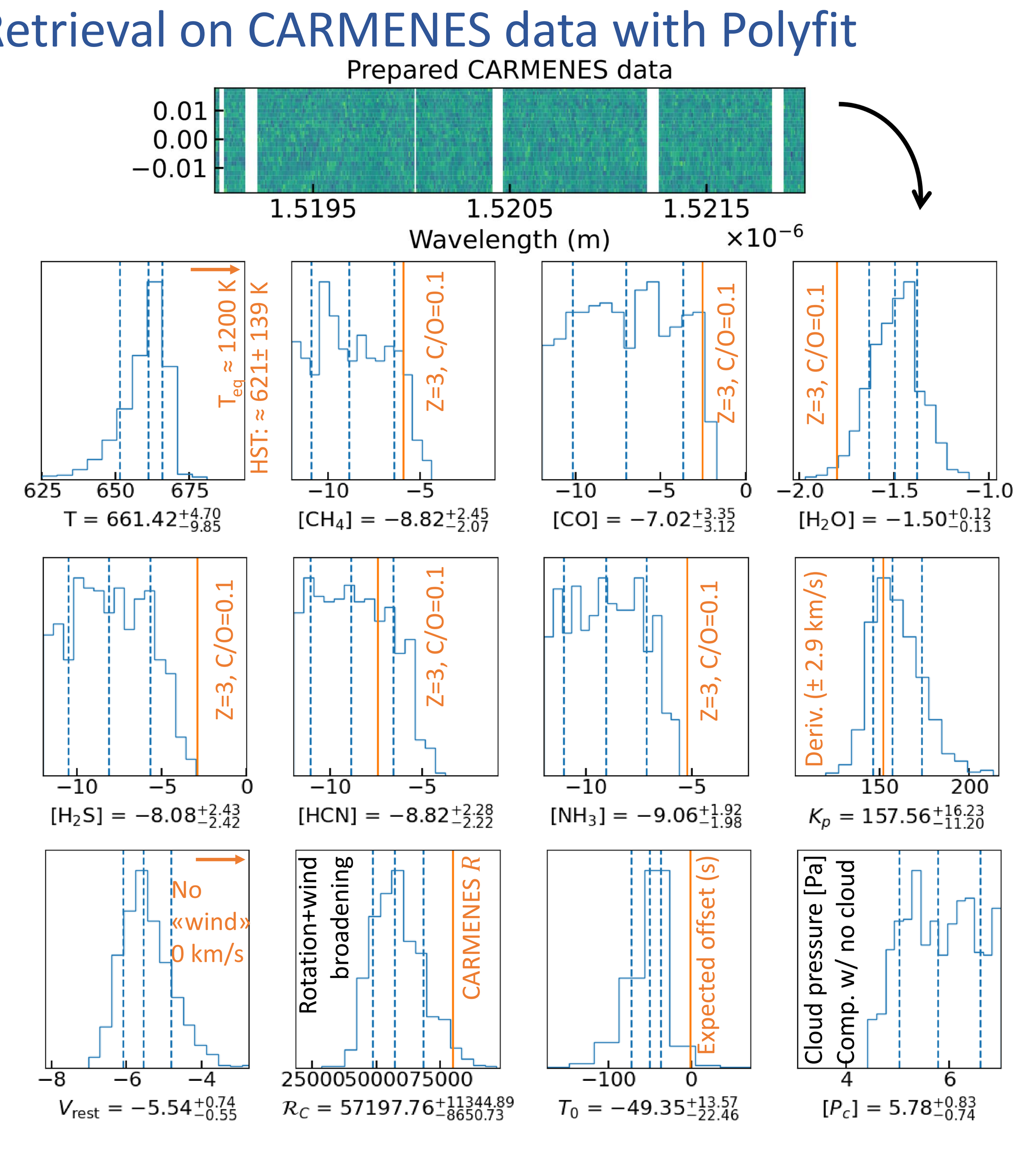


Figure 9: top row: sample of order #46 of the CARMENES prepared data. Bottom rows: posterior probability distributions for the CARMENES data, on all selected orders.

## Results

### Test retrieval on CARMENES data with SysREM

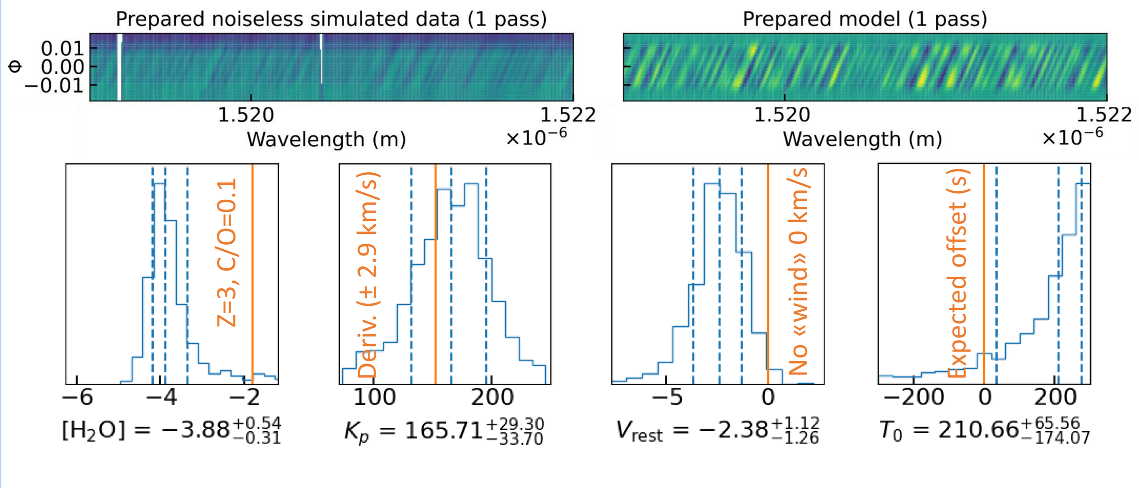


Figure 10: top row: (left) simulated data prepared with 1 SysREM pass and (right) a prepared forward model with 1 SysREM pass. Bottom: posterior probability distributions for the CARMENES data with 1 SysREM pass (on both the data and the forward models), on 46 selected orders

## Conclusion

- The retrieval and CCF results are consistent with each other, although the retrieval has smaller error bars and provides more details.
- Retrieval results for HD 189733 b are consistent with a super-solar metallicity, a sub-solar C/O ratio, and a significant spectral blueshift. Retrieved T is consistent with HST results (e.g., Tsiaras et al., 2018).
- Retrieving the mid-transit time is an important sanity check.
- The above framework and its demonstration are only valid for a subset of preparing pipelines. Invalid frameworks may lead to biased results.


References:
Alonso-Floriano, F., Sánchez-López, A., Snellen, I., et al. 2019, A&A, 621, A74
Brogi, M., de Kok, R. J., Albrecht, S., et al. 2016, ApJ, 817, 106
Brogi, M., & Line, M. R. 2019, AJ, 157, 114
Boucher, A., Darveau-Bernier, A., Pelletier, S., et al. 2021, AJ, 162, 233
Buchner, J., Georgakakis, A., Nandra, K., et al. 2014, A&A, 564, A125
Damiano, M., Micela, G., & Tinetti, G. 2019, ApJ, 878, 153
Klein, B., Debras, F., Donati, J.-F., et al. 2023, MNRAS, 527, 544
Feroz, F., Hobson, M. P., & Bridges, M. 2009, MNRAS, 398, 160
Flowers, E., Brogi, M., Rauscher, et al. 2019, AJ, 157, 209
Gibson, N. P., Nugroho, S. K., et al., 2022, MNRAS, 512, 4618
Louden, T., & Wheatley, P. J. 2015, ApJL, 814, L24
Mollière, P., Wardenier, J. P., van Boekel, R., et al. 2019, A&A, 627, A67
Nasedkin et al., 2024. Journal of Open Source Software, 9(96), 5875
Sánchez-López, A., Alonso-Floriano, F., López-Puertas, M., et al. 2019, A&A, 630, A53
Tamuz, O., Mazeh, T., & Zucker, S. 2005, MNRAS, 356, 1466
Tsiaras, A., Waldmann, I. P., Zingales, T., et al. 2018, AJ, 155, 156
Wyttenbach, A., Ehrenreich, D., Lovis, C., Udry, S., & Pepe, F. 2015, A&A, 577, A62



†**Contact:** Doriann BLAIN (doriann.blain@gmail.com)