Machine Learning for automatic identification of new minor species

Frédéric Schmidt, Guillaume Cruz Mermy, Justin Erwin, Séverine Robert, Lori Neary, Ian Thomas, Frank Daerden, Bojan Ristic, Manish Patel, Giancarlo Bellucci, et al.

Machine Learning for automatic identification of new minor species

Frédéric Schmidt [1], Guillaume Cruz Mermy[1], Justin Erwin[2], Séverine Robert[2], Lori Neary[2], Ian R. Thomas[2], Frank Daerden[2], Bojan Ristic[2], Manish R. Patel[3], Giancarlo Bellucci[4], Jose-Juan Lopez-Moreno[5], Ann-Carine Vandaele[2]

[1] *Université Paris-Saclay, CNRS, GEOPS, 91405, Orsay, France,* [2] *Belgian Institute for Space Aeronomy (BIRA-IASB), Avenue Circulaire, 3 B-1180 Brussels Belgium,*[3] *School of Physical Sciences, The Open University, Milton Keynes, MK7 6AA, U.K.,*[4] *INAF-Istituto di Astrofsica e Planetologia Spaziali, Rome, ITALY,*[5] *Instituto de Astrofísica de Andalucía CSIC*

Abstract

One of the main difficulties to analyze modern spectroscopic datasets is due to the extremely large amount of data. For example, in atmospheric transmittance spectroscopy, the solar occultation channel (SO) of the NOMAD instrument onboard the ESA ExoMars2016 satellite called Trace Gas Orbiter (TGO) had produced ∼10 millions of spectra in ∼20000 acquisition sequences since the beginning of the mission in April 2018 until 15 January 2020. Other datasets are even larger with ∼billions of spectra for OMEGA onboard Mars Express or CRISM onboard Mars Reconnaissance Orbiter. Usually, new lines are discovered after a long iterative process of model fitting and manual residual analysis. Here we propose a new method based on unsupervised machine learning, to automatically detect new minor species. Although precise quantification is out of scope, this tool can also be used to quickly summarize the dataset, by giving few endmembers ("source") and their abundances.

The methodology is the following: we proposed a way to approximate the dataset non-linearity by a linear mixture of abundance and source spectra (endmembers). We used unsupervised source separation in form of non-negative matrix factorization to estimate those quantities. Several methods are tested on synthetic and simulation data. Our approach is dedicated to detect minor species spectra rather than precisely quantifying them. On synthetic example, this approach is able to detect chemical compounds present in form of 100 hidden spectra out of 10^4, at 1.5 times the noise level. Results on simulated spectra of NOMAD-SO targeting CH_4 show that detection limits goes in the range of 100-500 ppt in favorable conditions. Results on real martian data from NOMAD-SO show that CO_2 and H_2O are present, as expected, but CH_4 is absent. Nevertheless, we confirm a set of new unexpected lines in the database, attributed by ACS instrument Team to the CO_2 magnetic dipole.

Keywords: spectroscopy, atmosphere, data mining, machine learning, unsupervised, source separation, non-negative matrix factorization

1. Introduction

In modern exploration science, one has to face a major challenge : how to learn something new from analyzing a large dataset collection while taking into account what we already know. If the current knowledge outweight the analysis, the discovery of new elements may be difficult. Usually, in the field of spectroscopy, one can compare laboratory spectra, model and observation spectra. Going back and forth leads to discovery of new lines by identifying unexpected residuals in the observation data (not expected by the model). Sometimes, initial identification of lines can be wrong. As an example, spectroscopic evidence of atmospheric CO_2 ice cloud was reported after the discovery of an emission spike at a wavelength of 4.3 μm from Mariner 6 and 7 infrared probings of the bright martian limb (Herr and Pimentel, 1970), but this spectral feature was mistaken for a resonant scattering band of CO_2 fluorescence (López-Valverde et al., 2005).

For one single spectrum, one can use simulation algorithm (see for instance Faisal et al., 2020). For large datasets, simplest ideas would be to scrutinize average spectra, or potential band depth distribution. Unfortunately, in the case of low signal-to-noise ratio (SNR, defined as signal / standard deviation of noise), such methods fail (as will be illustrated in the toy example). Analyzing residuals after modeling is a good method but it requires a lot of work.

Several statistical tools with various approaches have been proposed, such as the Principal Component Analysis (PCA) (Penttilä et al., 2018; Geminale et al., 2015), or Independent Component Analysis (ICA) (Shashilov et al., 2006; Erard et al., 2009), but most of them require a human operator to pick endmembers and trends since those methods are nothing more than a change of representation. Furthermore, none of these methods guarantees positivity of the *component* (which are sometimes also called *source*), which can be problematic during the interpretation. Recently, advanced machine learning methods based on non-negative matrix factorization have been proposed (Lee and Seung, 1999; Moussaoui et al., 2008; Dobigeon et al., 2009; Schmidt et al., 2010; Gillis and Glineur, 2012; Hinrich and Mørup, 2018). This approach is completely different from PCA/ICA: each source is positive and represents an endmember / a trend. A source is not one spectrum extracted from the dataset but a statistical reconstruction. By using this approach, the human operator doesn't have to identify endmembers/trends anymore, since they are automatically picked by the algorithm in form of source. Furthermore when there are statistical / spectral correlations between sources PCA/ICA fails because it assumes orthogonality / independence, which is not the case for non-negative matrix factorization.

Based on this new approach, we propose a tool:

- to give an overview and quickly summarize a large and complex spectroscopic dataset with simple variables

- to detect potential new spectroscopic features (unexpected minor species, new absorption lines,...)
- to be performed in a fully blind way (without prior information on neither the spectra, nor the abundances).

The target observation type of this study is solar occultation. This measurement principle has been proposed as early as 1900, an interesting review was published by Smith and Hunten (1990). Several recent instruments used this technique to investigate the composition of the Earth's (SCIAMACHY/ENVISAT Bovensmann et al. 1999), Mars' (SPICAM Bertaux et al., 2000) or Venus' atmospheres (SPICAV Bertaux et al., 2007). Here we will focus on the recent NOMAD instrument (Vandaele et al., 2015), and especially the SO channel, designed to study the Martian atmosphere and its trace gases, such as methane. Indeed the presence of CH_4 on Mars is a very hot topic for the planetary science community (Giuranna et al., 2019; Korablev et al., 2019; Moores et al., 2019). In the present article, we propose to apply the tool for potential CH_4 detection. Nevertheless, the approach can be extended to other types of spectroscopic measurements.

2. Dataset

We propose here to focus on the Nadir and Occultation for MArs Discovery (NOMAD) instrument onboard ESA's ExoMars Trace Gas Orbiter and especially the Solar Occultation (SO) channel (Vandaele et al., 2015). NOMAD is a compact, high-resolution, dual channel IR spectrometer (SO and LNO) coupled with a highly miniaturized UV-visible spectrometer (UVIS), capable of operating in different observation modes: solar occultation, nadir and limb.

The SO channel operates at wavenumbers from 2320 cm^{-1} to 4550 cm^{-1} (wavelength 2.2 to 4.3 μm), using an echelle grating with a groove density of 4 lines/mm in a Littrow configuration in combination with an Acousto-Optic Tunable Filter (AOTF) for spectral order selection. The width of the selected spectral ranges varies from 20 to 35 cm^{-1} depending on the selected diffraction order. The detector is an actively cooled HgCdTe Focal Plane Array. SO achieves an instrument line profile resolution of 0.15 cm^{-1}, corresponding to a resolving power $\lambda/\Delta\lambda$ of approximately 25000. All details of the instrument are available in Neefs et al. 2015 and Vandaele et al. 2018. The orders with the maximum sensitivity to CH_4 are: 119, 134 and 136. We will use the data from the beginning of the mission in April 2018 until 15 January 2020, in calibration version 1p0a. Due to temperature change, the spectral registration varies, producing a shift up to $\sim$10 spectels. We corrected it by aligning the full dataset to a reference spectra (arbitrarily choosen with the maximum band depth of water) by cross-correlation. No sub-spectel resampling has been performed but a simple shift. When the calibration will be improved, this step will most probably be replaced by a routine correction. The data are available on the ESA/Planetary Science Archive after a 6 months embargo period.

3. Method

In this section, we first describe the data pretreatment required for non-negative matrix factorization purpose followed by the data mining method.

3.1. Data pretreatment

After calibration, the NOMAD SO spectra are in transmittance $T = I/I_0$, depending on wavenumber ν, with I the observed light intensity trough the atmosphere and I_0 the solar spectra measured outside the atmosphere.

Assuming that the atmosphere is homogeneous, and that multiple scattering and refraction are negligible (Smith and Hunten, 1990; Bovensmann et al., 1999), the optical depth τ is a linear combination of $E(\nu)$ the total extinction, and ϵ the slant column density, for each chemical species i:

$$\tau(\nu) = -\log T(\nu) \approx \sum_{i=1}^{N_S} E_i(\nu).\epsilon_i + MC(\nu) \tag{1}$$

with N_S, the total number of species and $MC(\nu)$ a modeled continuum described below.

The slant column density ϵ is directly related to the total number of particles $N(s)$ along the line of sight s:

$$\epsilon = \int N(s)ds \tag{2}$$

While the extinction by gas is usually highly structured, absorption by particles, scattering by molecules and particles, and also reflection at the surface are broadband features. Such large features are modeled by a continuum $MC(\nu)$, often taken as a polynomial, that is filtered out.

The problem with this continuum removal rationale is that when the optical depth is large, the SNR is decreased and the noise effect on continuum removal amplified (see Sup. Mat.).

Instead of using this rationale, we propose to first correct for the continuum $C(\nu)$ in the transmittance space:

$$T^*(\nu) = T(\nu) - C(\nu) \tag{3}$$

Then convert the spectra into absorbance:

$$X(\nu) = 1 - T^*(\nu) \tag{4}$$

The final step is the linear mixture :

$$X(\nu) \approx \sum_{i=1}^{N_S} S_i(\nu).A_i \tag{5}$$

with $S_i(\nu)$ the source spectra and A_i the spectral abundance. In this description, the physical meaning of $S_i(\nu)$ and A_i is lost but the apparent SNR

is dramatically increased, which is much more important for our analysis. Nevertheless, assumptions required in eq. 1 are usually not relevant. Radiative transfer model used for precise quantification is highly non-linear.

One has to consider that this unsupervised linear unmixing problem is already very difficult for machine learning. Solving non-linear model in a unsupervised way is a research area that is clearly not solved yet. In addition, we would like to focus on spectral detection, rather than quantification. Thus, we will focus on $S(\nu)$ much more than A. We will show that for linear, but also non-linear simulation and real data, meaningful $S(\nu)$ can be retrieved. Due to non-linearity, A may differ significantly from truth, but the big tendencies should be respected. After the quick-look analysis, estimating S_i and A, one must go back to the real data. The most trivial strategy is to pick the spectra X out of the collection, with the highest abundance of a selected source S_i.

In the following, we will use the continuum estimation $C(\nu)$ using asymmetric least square (Eilers and Boelens, 2005), with parameters : $\nu_{smooth} = 10^3$ and $p = 1 - 10^{-2}$, 10 number of iterations.

3.2. Non negative matrix factorization

For a collection of spectra, eq. 5 can be written in matrix form $\mathbf{X}_{kj} \approx \mathbf{S}_{ki}.\mathbf{A}_{ij}$, with i the source index (from 1 to N_S) , j the observation index (from 1 to N_O) and k the wavenumber index (from 1 to N_ν). Thus, one have to estimate $\mathbf{S}$ and $\mathbf{A}$, by minimizing the objective function:

$$F = \|\mathbf{X} - \mathbf{S}.\mathbf{A}\|^2 \tag{6}$$

with $\|.\|$, the Frobenius norm (usual L_2 norm).

Several algorithms have been proposed to solve this problem, subject to positivity (both $\mathbf{S}$ and $\mathbf{A}$ are non-negative). Such problem is called Non negative Matrix Factorization (NMF). This constraint is important to keep the physical meaning, but also to promote sparsity of $\mathbf{S}$ (a signal is sparse when most of the values are close to zero except several non-zero values). Let $\dot{\mathbf{S}}$ and $\dot{\mathbf{A}}$ be the estimation of those quantities.

MU. We propose to use the Multiplicative Updates (MU) of Lee and Seung (1999) accelerated by Gillis and Glineur (2012). We used the convergence parameter $\alpha_{MU} = 1$. Other alternative algorithms are possible but give equivalent results since they minimize the same cost function. This algorithm has the advantage of very fast computation time but the result may depend on initialization.

BPSS2. We propose to test another kind of algorithm: the Bayesian Prior Source Separation (Moussaoui et al., 2006; Dobigeon et al., 2009), that has been optimized (Schmidt et al., 2010), hereafter called BPSS2. This algorithm has the main advantage to account for extra constraint : the sum-to-one or sum-lower-than-one on the abundances ($\sum_i A_{ij} = 1$) that also promotes sparsity of $\mathbf{S}$. This algorithm, based on Monte Carlo approach is much more time consuming. One

approach to reduce the computation time is to select only relevant spectra out of the dataset (Moussaoui et al., 2008), but then the statistics may be biased (Schmidt et al., 2010). Thanks to the advances of computer capabilities, we propose to treat the full dataset. This kind of algorithm is very slow but since the formulation is Bayesian, it converge toward an unique solution.

psNMF. In order to regularize the problem of eq. 6, one can add an extra penalization term to enforce sparsity on $\mathbf{A}$ (only few non zeros elements in $\mathbf{A}$) (Kim and Park, 2007) :

$$F = \|\mathbf{X} - \mathbf{S}.\mathbf{A}\|^2 + \lambda \|\mathbf{A}\|_1 \tag{7}$$

With $\|.\|_1$, the L_1 norm. The first term is called data attachment term (the usual squared difference). The second is called regularization term. The problem with this approach, is that hyperparameter λ is not known and has to be tuned manually. A recent approach has been proposed to solve this problem in the Bayesian framework (Hinrich and Mørup, 2018). The main idea is to encompass all variables and hyperparameters in a unique problem that is estimated with variational update principle. We will refer this algorithm to probability sparse NMF (psNMF). This algorithm has the advantage to have a reduced computation time and no hyperparameter tuning. It also has a regularization term to avoid strong dependence of the initialization on the final solution.

In order to estimate the precision of the reconstruction, we used the Root Mean Square Difference $RMSD$:

$$RMSD = \frac{\sqrt{\left\langle \left(\mathbf{X} - \dot{\mathbf{S}}.\dot{\mathbf{A}}\right)^2 \right\rangle}}{\langle \mathbf{X} \rangle} \tag{8}$$

With $\langle . \rangle$, the mean.

Once the sources are estimated, we quantify their relevance for the global dataset. From the total reconstruction $\dot{\mathbf{X}}_{kj}$=$\dot{\mathbf{S}}_{ki}.\dot{\mathbf{A}}_{ij}$, for all i, we can estimate the contribution of source i', that is to say: $\dot{\mathbf{X}}^i_{kj}$=$\dot{\mathbf{S}}_{ki'}.\dot{\mathbf{A}}_{i'j}$. Thus, the relevance of source i is defined as:

$$R^i = \frac{\left\langle \left| \dot{\mathbf{X}}^i - \dot{\mathbf{X}} \right| \right\rangle}{\left\langle \dot{\mathbf{X}} \right\rangle} \tag{9}$$

This definition is convenient since the sum of all R^i is one (this property is only present when sources and abundances are positive) and we can easily estimate the % contribution of each source in the final reconstruction. One has to note that relevance is not a measure of presence or not of a minor specie (for instance CH_4) but a measure of how important is the source over the dataset. Major species, should always have a larger relevance than minor species. In the following, we plot all sources results by decreasing order of relevance.

3.3. Band depth (BD)

We used the following band depth definition, difference of the geometric mean of two reference wavenumbers in the continuum, compared to the band:

$$BD = X(\nu_l)^{\frac{\nu_c-\nu_l}{\nu_r-\nu_l}}.X(\nu_r)^{\frac{\nu_r-\nu_c}{\nu_r-\nu_l}} - X(\nu_c) \tag{10}$$

with X the observed spectra in transmittance, ν_c the wavenumber of the center of band, ν_l the wavenumber of the reference level on the left (smaller wavenumber), ν_r the wavenumber of the reference level on the right (larger wavenumber).

4. Synthetic tests

We simulated several synthetic observations in different conditions, to mimic the case of NOMAD-SO. The first section describes a simple toy model example and the second one presents extensive tests of this toy model with various cases. By *hidden spectra*, *hidden compounds* and *hidden CH_4*, we always refer to a spectral dataset with a dominant major component (here water) and a minor specie (here CH_4). The goal of the proposed approach is to pick up a source, containing CH_4 only.

4.1. Toy example

4.1.1. Synthetic dataset

In order to demonstrate the usefulness of our method, we propose here a toy example in a very difficult case. We will see that usual method fails detection but our method is able to detect the hidden compounds.

For this toy example, we simulate a linear mixture of $N_O = 10^4$ observations spanned over $N_\nu = 320$ spectels (see fig. 1) similar to order 136 of NOMAD-SO. Each spectrum is a mixture of a spectra of water vapor S_{H_2O} (coming from one actual source estimated from real data using psNMF) and theoretical methane S_{CH_4} from Villanueva et al. (2018), with corresponding abundances A_{H_2O}, A_{CH_4}:

$$X = S_{H_2O}.A_{H_2O} + S_{CH_4}.A_{CH_4} + n \tag{11}$$

The noise n is assumed to be a Gaussian process with a standard deviation of σ=0.001 and no bias: $n = \mathscr{G}(0, \sigma)$. All spectra contain pure water vapor with a coefficient following $A_{H_2O} = 5/6.\beta(1, 10) + 1/6.\mathcal{U}(0, 1)$, a mixture of beta ($\beta$) distribution for 5/6 of the sample and an uniform ($\mathcal{U}$) distribution for 1/6 of the sample. This process mimics well the water vapor band depth distribution (BD, see definition in section 3.3) of the real dataset (see Fig.2). As the baseline of S_{H_2O} is not zero, we also mimic baseline correction errors. In addition 100 spectra out of 10000 contain methane with $A_{CH_4} = 1$, such that the band depth of S_{CH_4} is at 3-σ level. Please note that the model to generate the data is not fulfilling the sum-to-one constraint, but fully fulfilling the positivity constraint. Given the defined noise and signal level, the noise $RMSD$ is 0.16.

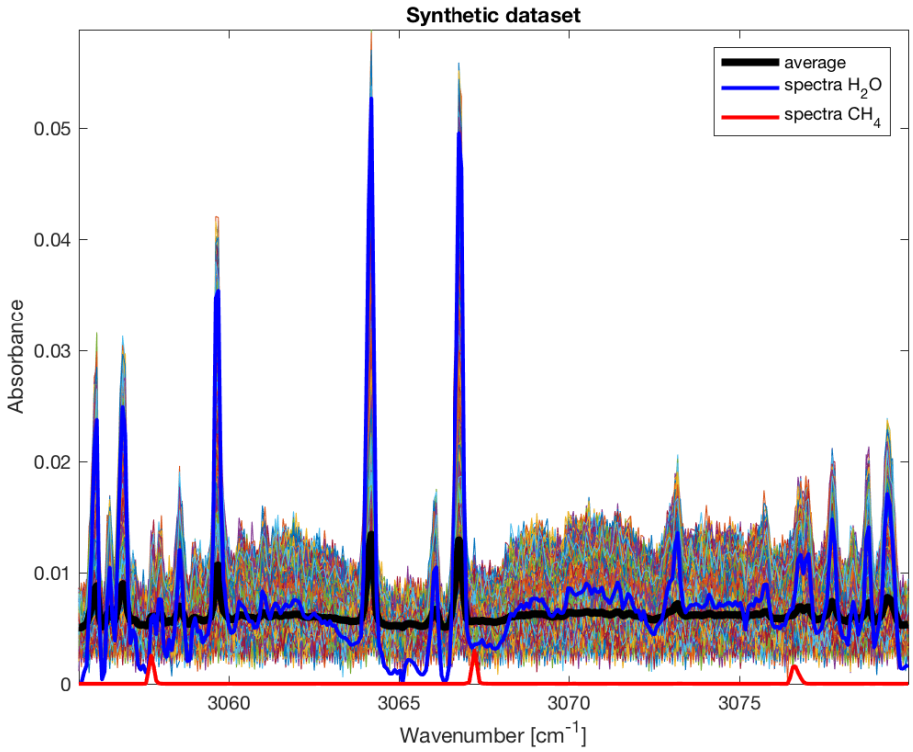

Figure 1: Synthetic dataset containing 10^4 spectra with various abundances of H_2O and 100 containing CH_4 at 3-σ level of the noise. In blue the reference spectra S_{H_2O} of H_2O (coming from actual data analysis). In red the reference spectra S_{CH_4} of CH_4 (from theoretical data).

The final synthetic dataset is represented in Fig. 1.

In order to check the quality of the estimation, we simply compute the correlation coefficient between S_{CH_4} and the estimated N_S sources $\dot{\mathbf{S}}$, using:

$$Q = corr\left\{S_{CH_4}, \dot{\mathbf{S}}_{:\mathbf{i}}\right\} \tag{12}$$

The ith source with the maximum correlation is identified to CH_4 contribution. The value to the maximum correlation is used as metric to assess the quality of the retrieval.

4.1.2. Results

By plotting the 10000 samples of the dataset, one is able to identify easily the H_2O bands. Nevertheless, we cannot observe the target CH_4 in the average spectrum, even at 3-σ level, because it is lost in the baseline changes.

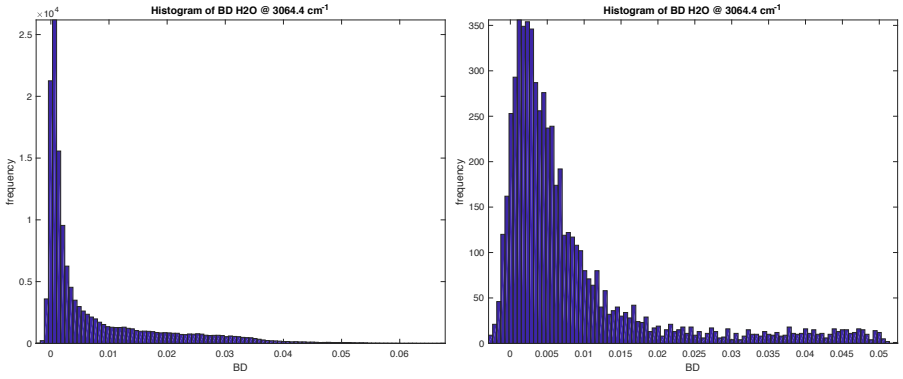

Figure 2: Water vapor Band Depth distribution (left) in the real observation (right) modeled by the toy example.

The second simple tool for detection would be the analysis of the band depth. Figure 3 (left) shows the histogram of the main CH_4 band that exhibits no sign of the presence of CH_4 (no asymmetry in the positive part). Figure 3 (right) represents the 100 spectra with the maximum CH_4 BD at 3067.2 cm^{-1}. Again, no particular elements can be used to argue for detection.

Figure 4 represents the results from the non-negative matrix factorization using psNMF algorithm. One can clearly identify both H_2O and CH_4 sources. Since those 2 chemical compounds are not correlated in abundance, (A_{H_2O} and A_{CH_4} are independent), two different source spectra are identified. Please note that the relevance of source 4 is very low (0.4%), meaning that only 0.4% of the variability in the dataset is due to CH_4, a very low value, as expected for minor species.

In this case, the correlation coefficient between estimated abundances $\dot{\mathbf{A}}_{\mathbf{4}:}$ and true ones A_{CH_4} is 0.73. Since the quantification of abundance is a more difficult problem, we will not pay excessive attention on this parameter.

4.1.3. Convergence and computation time

We set the MU algorithm convergence to relative difference of the cost function $< 10^{-8}$ and a maximum running time of 1000 seconds. For psNMF, we set the relative difference of the cost function to $< 10^{-7}$ and a maximum iteration to 2000. For BPSS2, we compute a minimum burn in of 1000 iterations and after that when the long term statistics (1000 last iterations) of the Markov Chain is close to the short term statistics (100 last iterations), convergence is considered to be reached. Then another 1000 iterations are computed to estimate the final solution statistics.

We run the 3 identified tools 10 times on the same dataset with different noise realization, and compute mean and standard deviation from these 10 experiments. Results are presented in Table 1. One can clearly see that the even if the convergence is set, there is a high variability in MU results, due to the

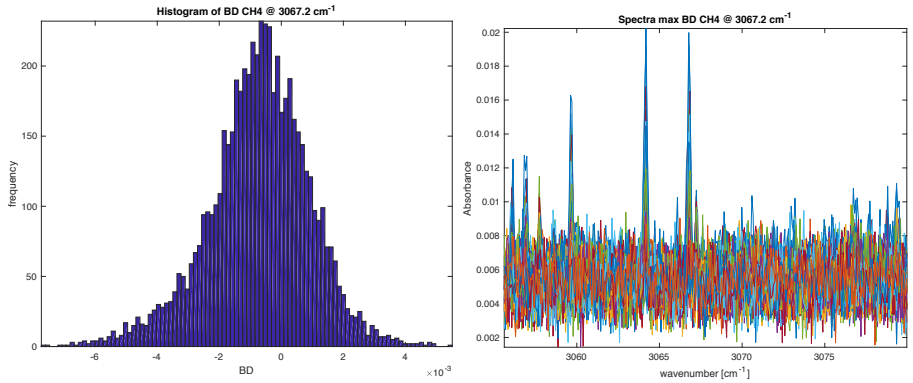

Figure 3: (left) Histogram of Band Depth at 3067.2 cm^{-1} from the dataset containing 100 CH_4 at 3-σ level out of 10^4 spectra. (right) 100 spectra with the maximum Band Depth at 3067.2 cm^{-1} specific of CH_4. Signal is dominated by water and by noise. No specific signature of CH_4 is visible.

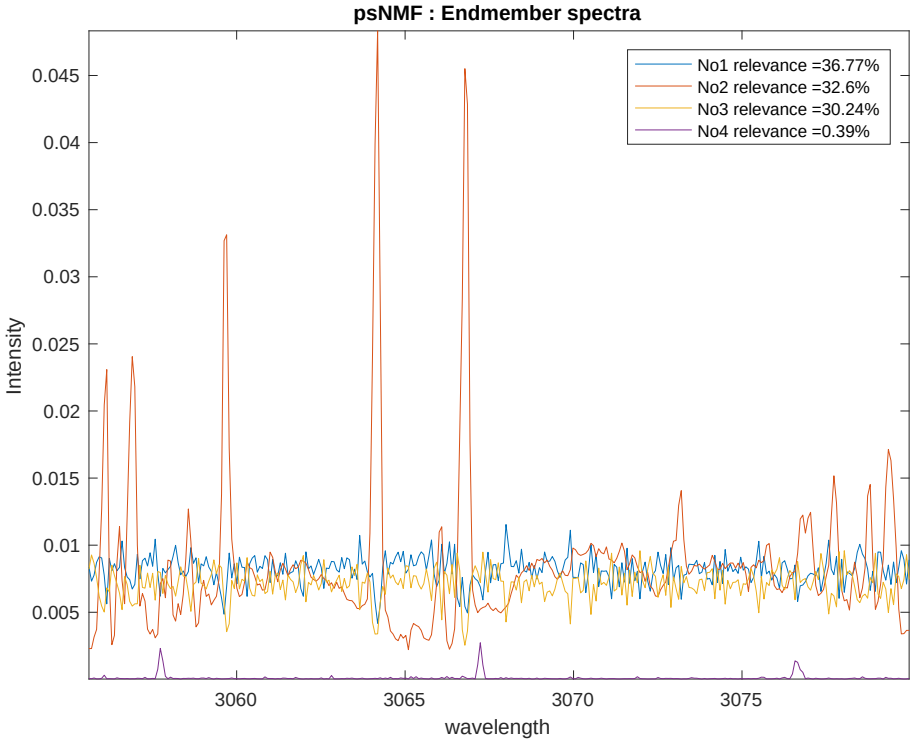

Figure 4: Results of the psNMF algorithm for $N_S = 4$. Sources 1 and 3 are identified to the level with significant noise contribution, source 2 is identified to H_2O (correlation coef. with groundtruth 0.99), and source 4 is CH_4 (correlation coef. with groundtruth 0.98). Relevance is computed from Eq. 9.

	MU	psNMF	BPSS2
Quality Q	0.35±0.12	0.822 ±0.005	0.41±0.06
$RMSD$ relative error	0.1455 ± 2.10^{-6}	0.1461 ± 5.10^{-6}	0.1468 ± 3.10^{-4}
Computation time (s)	13±8	46±9	413±21

Table 1: Results (mean and standard deviation) from 10 realizations of a toy synthetic example with $N_S = 5$ (in agreement with next section on synthetic tests), $N_O = 10000$, $N_\nu = 320$ and 300 CH_4 spectra hidden at a level of 1 std of the noise. Quality is computed as a correlation coefficient (see Eq. 12). $RMSD$ is computed from Eq. 8. Computation time is expressed in second.

lack of regularization. On this particular example, the best is clearly psNMF algorithm.

The $RMSD$ is computed for all cases and shown in Table 1. We can observe that the value is almost equivalent, around 0.146, for all method but MU is slightly better, due to the fact that the cost function has no other term. MU algorithm is just minimizing the reconstruction. As a comparison, the $RMSD$ expected for a perfect reconstruction of the signal (and not the noise) of this toy example is 0.16. With 5 sources (significantly more than the 3 sources we define in this toy example), the noise is also fitted, as expected.

The quality Q is the only parameter to assess the quality of the algorithm to detect minor specie (here CH_4). In this particular toy example, psNMF seems to be the best algorithm, providing a source correlated with groundtruth CH_4 with a correlation coefficient up to 0.8. We will extensively test this performance in the next section.

We also estimate the computation time on a 2.9 GHz Intel Core i7 with 16 Go DDR3 RAM as an example. All algorithms are implemented in ©Matlab using parallelized matrix computation. Results, presented in Table 1, demonstrate that MU is faster than psNMF but both are clearly less resources consuming than BPSS2. From the computation time and efficiency, we excluded BPSS2 from the next tests.

4.2. Extended synthetic tests

For the first set of tests, we used the same toy model described in section 4.1, except with 100 CH_4 spectra hidden at a level of 2 and 3 standard deviation of the noise (this number is called "factor above noise level"). In order to have robust results, we made 10 realizations and averaged the results.

Figure 5 represents the results as a function of the number of sources N_S. It presents two quality indicators of the results: the average correlation coefficient Q (see Eq. 12) and the fraction of realization with acceptable results (with $Q > 0.5$). We can observe that the psNMF is always better than MU on average at cost of an higher variability (higher standard deviation). Adding sources seems to always increase the detection until reaching a plateau around $N_S = 5$. Adding more sources will not drastically increase/decrease the source estimation. Nevertheless, it requires more computation time for a larger number

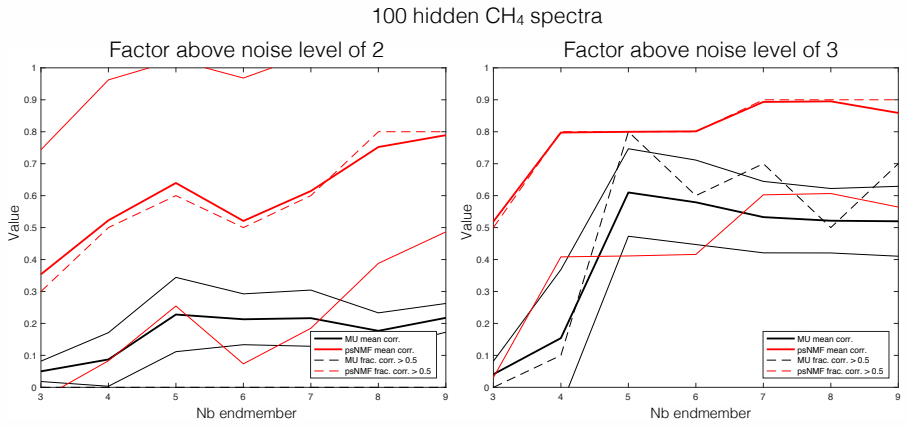

Figure 5: Results of the MU and psNMF algorithm for $N_S = 3$ to 9, $N_O = 10000$, $N_\nu = 320$, as a function of the number of source. The average Q of 10 realizations of the best estimated source (thick lines and standard deviation in thin lines) and the fraction of acceptable results (with $Q > 0.5$). (left) with a factor above noise level of 2 (right) with factor above noise level of 3.

of source (approximately x2 between 3 and 9 sources but the computation time always stays below 200 seconds).

For the second set of tests, we used the same toy model, except with 50 and 100 CH_4 spectra hidden at a level of 0.7, 1, 1.2, 1.5, 2.0, 2.5 and 3 standard deviation of the noise (this number is called "factor above noise level"). In order to have robust results, we made 10 realizations and averaged the results. Results are always with $RMSD < 0.18$ with an average ~ 0.16. $RMSD$ from the noise level is 0.16 whatever the experiment (the CH_4 is low enough so that it's contribution to $RMSD$ is negligible), so the reconstruction is in average as expected.

Figure 6 presents two quality indicators of the results: the average correlation coefficient Q (see Eq. 12) and the fraction of realization with acceptable results (with $Q > 0.5$). Both indicators indicate that the method psNMF clearly outperforms MU at high factor above noise level. From our visual inspection of the results, we define the detection limit when at least 50% of the results are with $Q > 0.5$ (correlation coefficient > 0.5). This definition is debatable but there is no absolute way of defining it. Figure 6 shows that the detection limit is at 1.5 factor above noise level for 100 hidden spectra case, around 2 for 50 hidden spectra. Below this limit, none of the method is able to detect the CH_4 spectra from the noise. For 20 hidden spectra, even at a factor above noise level of 3, none of the methods is able to detect the CH_4 spectra. One can also note that the psNMF is less stable since the standard deviation is much larger.

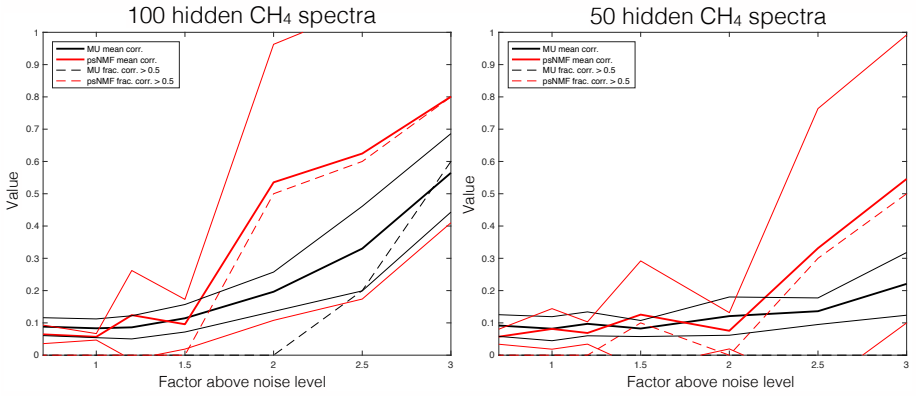

Figure 6: Results of the MU and psNMF algorithm for $N_S = 5$, $N_O = 10000$, $N_\nu = 320$, as a function of the factor above noise level. The average Q of 10 realizations of the best estimated source (thick lines and standard deviation in thin lines) and the fraction of acceptable results (with $Q > 0.5$). (left) with 100 hidden CH_4 spectra (right) with 50 hidden CH_4 spectra

5. Simulation of NOMAD-SO

5.1. Simulation dataset

This second dataset has been generated with the most precise direct model, taking into account the full non-linear radiative transfer and instrumental effects to produce synthetic transmittance, highly comparable with actual observations. Synthetic transmittances were made for real NOMAD-SO observation files using the relevant geometry and instrument parameters to attempt to include the variability inherent in the true measurements.

Model atmospheres for each occultation were developed from the GEM-Mars general circulation model (Neary and Daerden, 2018; Daerden et al., 2019). The output of the model were provided for 1 Martian day every 10 solar longitude, and 48 timesteps per Martian day. Atmospheric profiles were developed for each occultation by interpolating the model temperature and pressure to the solar longitude, local solar time, latitude, longitude, and tangent altitude relative to the areoid.

To construct the simulated transmittance spectra, the high resolution irradiances were computed for each occultation assuming a spherically symmetry and the tangent atmosphere developed from GEM-Mars for several different abundance of methane and water, which were simulated as constant volume mixing ratios. The spectroscopic data for methane and water were taking from HITRAN 2016 using CO_2 broadening (Gordon et al., 2017; Gamache et al., 2016; Fissiaux et al., 2014). The instrument forward model was then applied to each simulation by considering the AOTF bandpass, instrument Instrument Line Shape (ILS), blaze function, spectel to wavenumber calibration, and the contribution of light coming from the main order and nearby orders. The final synthetic transmittance spectra is the ratio of this low-resolution irradiance to the top-of-atmosphere low resolution irradiance.

	CH_4 [ppt]	H_2O [ppm]	fraction of CH_4[%]	noise level
Value	0; 100; 500; 1000	0; 10; 100	1; 5; 10; 50; 100	0.001; 0.0001

Table 2: Simulation parameters. Fraction of CH_4 is fraction of spectra containing methane hidden in the simulation dataset.

The AOTF/echelle instrument was modeled using the latest available calibration (Liuzzi et al., 2019; Aoki et al., 2019), considering order addition from $+/-2$ nearby orders (5 total). The spectral calibration of NOMAD-SO varies because it is affected by the instrument temperature, and is provided for each individual NOMAD spectra. The 320 spectels cover the range 3056.1 cm^{-1} to 3080.4 cm^{-1} with a wavenumber step of 0.0763 cm^{-1}.

No simulation of dust has been performed. Due to the limited spectral range on a single order, about 25 cm^{-1}, the major effect of dust and other aerosols is relatively flat baseline, which we remove at the pre-treatment of the spectra. When dust is optically thick, then non-linearity may appear that are out of the scope of this simulation.

The simulation dataset consist of 12486 spectra, simulating observations of order 136 in the same configuration as the 106 solar occultations actually observed from May to December 2018.

We add to the dataset a random noise with standard deviation of 0.001 and 0.0001 in order to simulate the instrumental noise (corresponding to SNR of 100 and 1000 approximately).

We hide spectra containing CH_4 in a fraction of the total number of spectra from 1% to 100% in a random manner. In real observation, CH_4 may be spatially / temporally coherent but the number of scenarios is infinite. We feel that the random case is interesting enough to be tested. One has to note that contrarily to the previous toy model of section 4, here abundance are quantitative abundance in the atmosphere.

The simulation parameters are summed up in table 2.

5.2. Detection limits

We applied the psNMF method with $N_S = 5$, which is the most promising one from the previous analysis. We compute the analysis 10 times for 10 different random noise realizations and average the results in order to present robust conclusion. We select a pure CH_4 and a pure H_2O spectra (noted P_{CH_4} and P_{H_2O}) from the simulation as reference spectra.

5.2.1. Methods to analyze the results

The main difference with the toy model section in 4 is that H_2O and CH_4 may be highly mixed in the sources. Simple correlation coefficient to pick the best source is thus not efficient enough. We propose here another approach to estimate the best source.

For each estimated source $\dot{S}_{:i}$, we analyze it as a linear mixture of P_{H_2O} and P_{CH_4}:

$$\dot{S}_{:i} = P_{H_2O}.\alpha_{H_2O,i} + P_{CH_4}.\alpha_{CH_4,i} \quad (13)$$

This problem is called supervised detection algorithm since P_{H_2O} and P_{CH_4} are known, contrary to the general one, presented in Eq. 5, where source spectra are not known. The source i^* with the maximum α_{CH_4,i^*} is selected as the best target CH_4 source, called *best source* hereafter.

We then propose to use three indicators of good detection :

- *Fraction of the 4 main CH_4 peaks detected* (at 3057.7, 3063.4, 3067.2 and 3076.6 cm^{-1}). This is computed using the peak detection algorithm from ©Matlab on both simulation and best source with a tolerance of 2 spectels, i.e. detected peaks can be 2 spectels off the expected one. The peak must be with a maximum amplitude larger than 1/1000 the maximum of $\dot{S}_{:i^*}$ to be considered significant. Please note that even there are only 5 possible fraction (0, 0.25, 0.5, 0.75 and 1), since we average on 10 realizations, any number can appear.

- *Mean distance to the expected center.* Mean distance in spectel between the CH_4 peaks detected in the best source and the reference one.

- *Abundance of CH_4 in the source.* α_{CH_4} (from Eq. 13), which describes the amplitude of the CH_4 peaks in the best source.

5.2.2. Analysis of the results

Figure 7 summarizes all the results. Fraction of the 4 main CH_4 peaks detected in the most relevant source has always a standard deviation < 0.43 and a mean value of 0.06 over the 10 realizations. The Mean distance to the expected center has always a standard deviation < 0.40 and a mean value of 0.07 over the 10 realizations. The abundance of CH_4 in the source has always a standard deviation < 0.05 and a mean value of 0.005 over the 10 realizations.

This figure shows that the detection limits clearly depend on CH_4 density, but also on the fraction of hidden CH_4 and noise level, as expected. Abundance of CH_4 in the source α_{CH_4} maximum is 25%, meaning that in any cases H_2O is dominating the best source and so both CH_4 and H_2O are present in each best source. This is because CH_4 is a minor specie (as expected from the conditions of our simulation), its absorption band generally follows the air-mass, as H_2O does. So there is no particular source for CH_4 only.

When more than two lines are detected, we can consider it as a detection. This limit is reached for $CH_4 \geq 500$ ppt for 10 and 100 ppm of H_2O. Nevertheless, the detection limits lies between 100 and 500 ppt in the case of 10 ppm of H_2O vapor since the detection is perfect (100% of the 4 main CH_4 peaks detected) occurs for a fraction of CH_4 5 to 50%. Interestingly, the optimum detection is not when 100% of the spectra contains CH_4, but more between 5-50 %. This behavior is due to the statistics that is richer when also CH_4 is lacking in certain spectra. When 100% of spectra contain CH_4, the statistical variability of the dataset is mainly due to airmass (atmosphere is assumed to

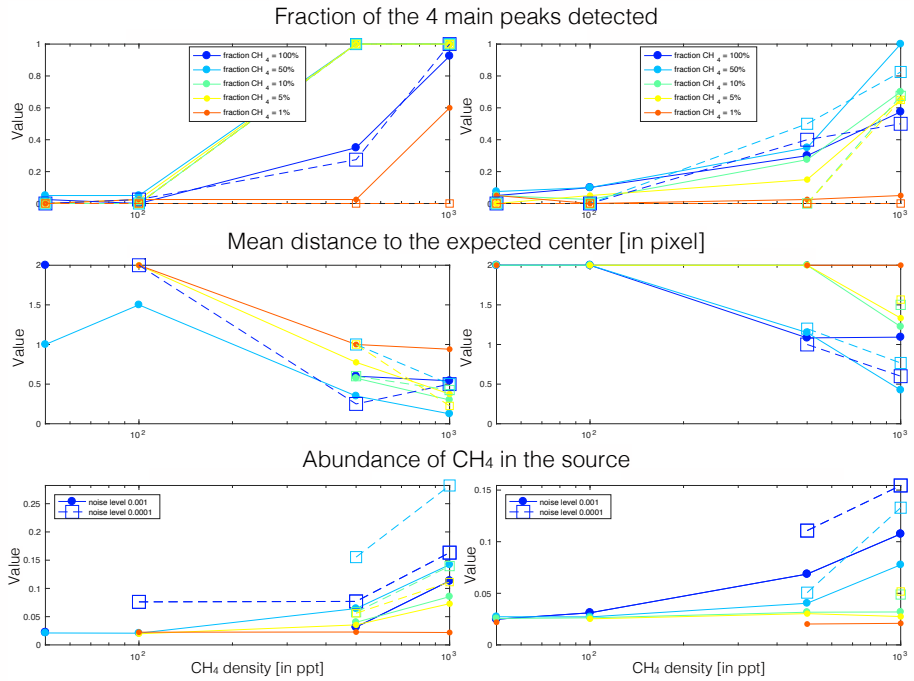

Figure 7: Results of the psNMF algorithm for $N_S = 5$ on simulation dataset, averaged over 10 noise realizations, for different noise levels (0.001 and 0.0001) and different fractions of hidden CH_4 (1%, 5%, 10%, %, 100%). Hidden CH_4 are taken within the same orbital sequences. The left panels represent results for 10 ppm of water vapor and the right ones for 100 ppm of H_2O. From top to bottom, we show: a) Fraction of the 4 main CH_4 peaks detected in the *best source* ; b) Mean distance to the expected center in spectel and c) Abundance of CH_4 in the source α_{CH_4}. Please note that the absence of plotted data means that no source was successfully detected.

be well mixed). So both CH_4 and H_2O are varying together and there is less statistics to base the detection on.

Noise level does not affect first the fraction of the 4 main CH_4 peaks but increases the spectral shift of the band center. In addition, it clearly affects the abundance and thus the band depth.

In conclusion, from this simulation analysis, one could expect detection limits of CH_4 in the range 100-500 ppt when operating in favorable conditions.

6. Real data analysis

In this section, we report the results of actual NOMAD data, focusing on diffraction orders with potential CH_4 lines: 119, 134 and 136, are shown respectively on Fig. 8, 9 and 10. We used the 821 ingress and egress transit orbits for order 119, 2358 orbits for order 134 and 703 for order 136. We filter spectra with SNR > 100. Results are compared with NOMAD simulations (Villanueva

	119	134	136
N_O	134045	365985	140064
$N_S = 4$	0.476	0.575	0.634
$N_S = 5$	0.456	0.553	0.609
$N_S = 6$	0.442	0.553	0.585
$N_S = 10$	0.410	0.484	0.544

Table 3: Number of spectra N_O and $RMSD$ relative errors for 4 to 10 number of sources N_S resulting from the analysis of all observations of NOMAD data up to 15 January 2020, using the psNMF algorithm. $RMSD$ is computed from Eq. 8.

et al., 2018) using the calibration pipeline. This process adds ghost lines from adjacent orders, as in real data. Table 3 summarizes the relative error and the number of spectra. The approach here is to compute the analysis with psNMF using $N_S = 5$ in agreement with the previous section.

Please remind that our approach is fully blind: no spectral information has been included in the analysis (nothing about H_2O, CO_2 or CH_4).

For all orders, sources of H_2O are estimated, as expected. Also a source presenting a residual of the continuum is always present. Due to non-linearities of the radiative transfer, the acquisition process (temperature dependence) and the wavenumber shift, the molecular species appears sometimes in different sources.

Order 136 gives the 1 source related to the background and 4 sources related to H_2O. All 4 sources of water have the peaks but with different relative intensities and wavenumber shift.

For order 119, both CO_2 and H_2O lines are identified (see Fig. 8). Since those two components are uncorrelated, separated sources are found by the algorithm.

Interestingly, order 134 presents a source with unexpected lines. The main lines are at positions : 3016.70, 3017.07, 3018.12, 3019.54, 3020.90, 3022.25, 3023.60, 3024.96, and 3027.29 cm^{-1}. These lines has been also detected in the ACS instrument data and attributed to CO_2 magnetic dipole transition (Trokhimovskiy et al., 2020). Further analysis shall be done to compare both NOMAD AND ACS data.

Solar lines are never appearing in the sources. They are self-corrected by the calibration since we don't use a reference solar spectra but the solar observation during the transit when the tangent altitude is so high that there is no martian atmosphere (typically > 200 km).

None of the analyzed orders presents sources related to CH_4.

7. Discussions and Conclusion

We implemented a new strategy to analyze spectroscopic datasets. This strategy is fully unsupervised, so that any kind of absorption bands can be discovered. The amount of prior information required is thus very low. The computation can be done on a regular hardware for the most common database and within reasonable amount of time ($\sim$100000 spectra).

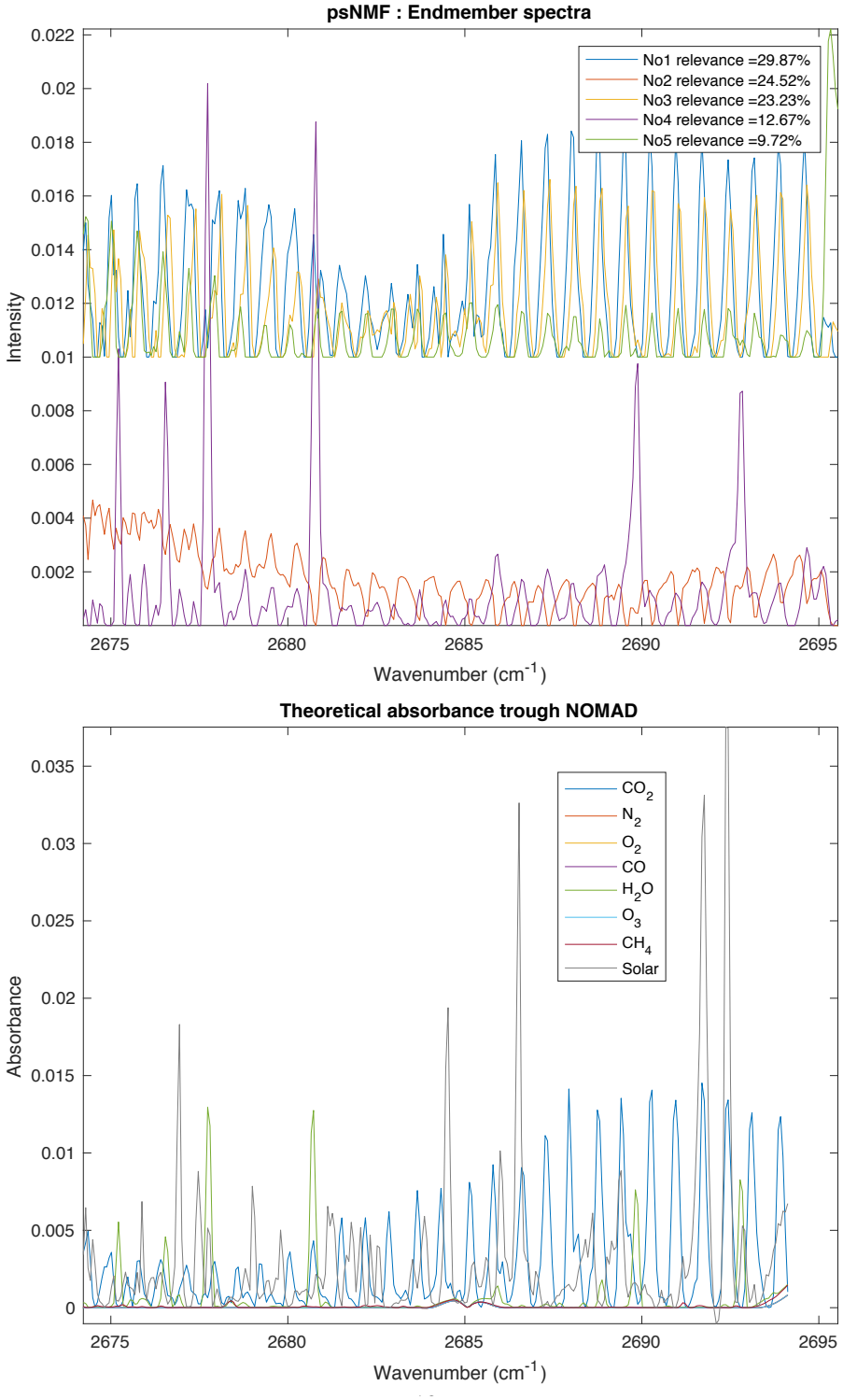

Figure 8: Results of the psNMF algorithm for the diffraction order 119 for $N_S = 5$. The sources 1, 3 and 5 are identified to CO_2 (shift of 0.01 for clarity). The source 2 is identified to the background level (continuum misestimation). The source 4 is identified to H_2O. No source seems to be related to CH_4.

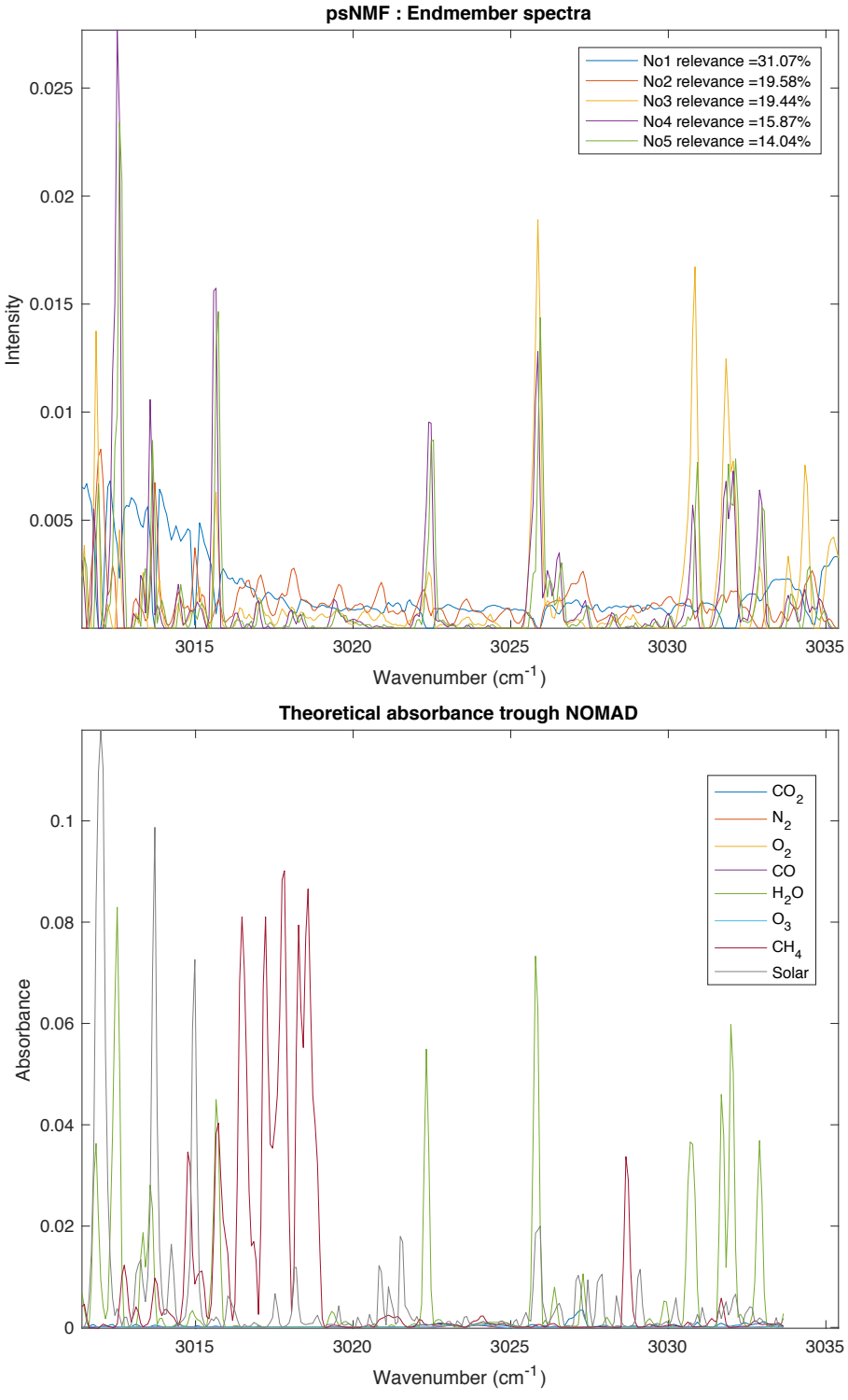

Figure 9: Results of the psNMF algorithm for the diffraction order 134 for $N_S = 5$. The source 1 is identified to the background level (continuum misestimation), the sources 3, 4 and 5 are identified to H_2O. The sources 2 present unmodeled lines that are not present in the spectroscopic database. These lines has been first detected in the ACS instrument data and attributed to CO_2 magnetic dipole transition (Trokhimovskiy et al., 2020). No source seems to be related to CH_4.

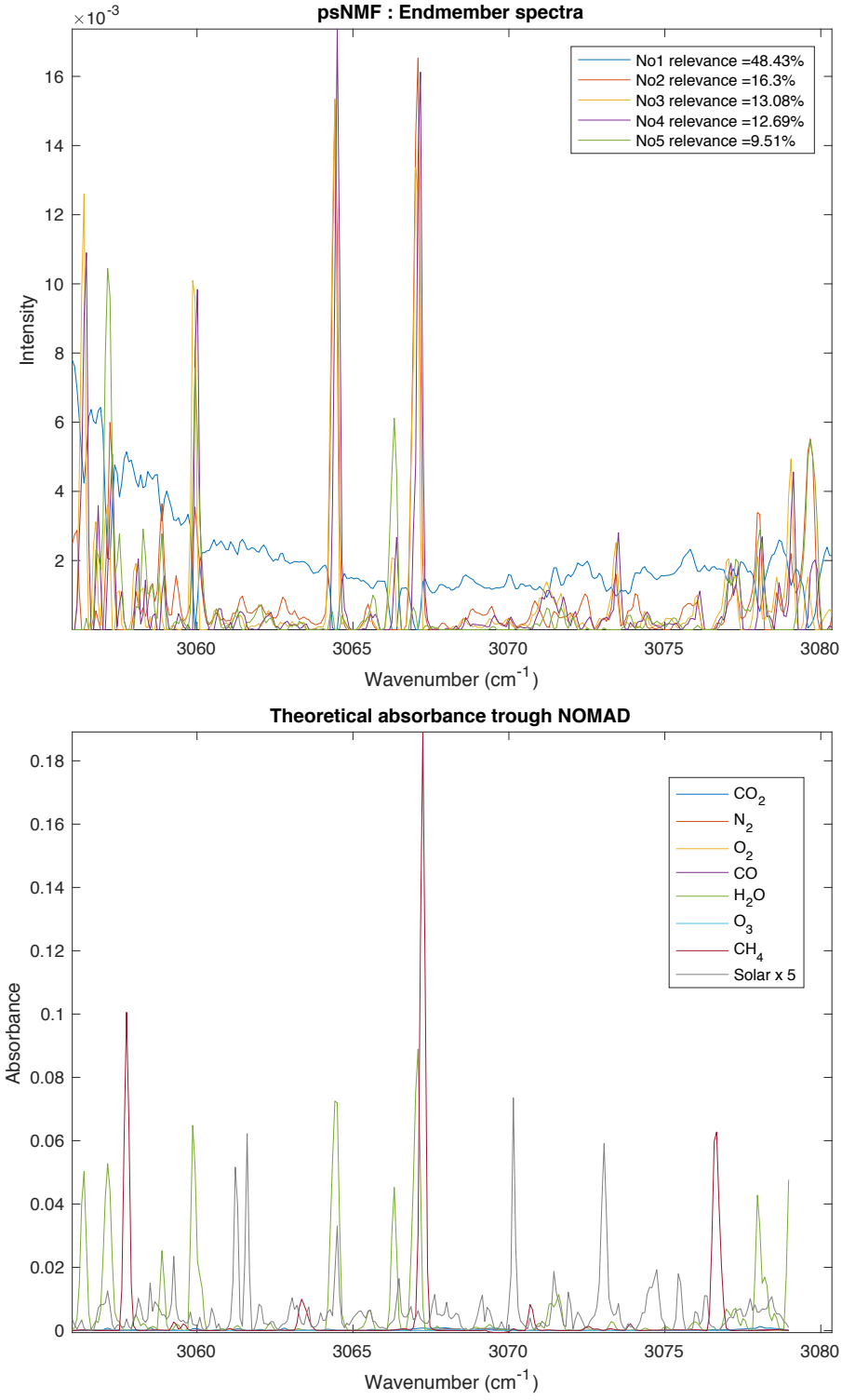

Figure 10: Results of the psNMF algorithm for the order 136 for $N_S = 5$. The source 1 is identified to the level background (continuum misestimation), the sources 2, 3, 4 and 5 are identified to H_2O, either directly either from the adjacent orders. No source seems to be related to CH_4.

We illustrate the approach for typical atmospheric spectroscopy. We first put forward a synthetic test, based on simple linear mixing to give a toy example and to identify the best promising algorithm. The psNMF clearly outperformed MU and BPSS2.

Then we proposed a simulation, based on realistic radiative transfer and instrumental effects, applied on NOMAD-SO spectra. The detection limits goes below 500 ppt in favorable conditions, with reduced H_2O and low noise level. The same range of detection limits is reach with usual approach of model fitting at a much higher computation cost and analysis effort. Given the simplicity of use, this tool may be relevant to handle large and complex datasets at first glance. As a perspective, analysis of residuals after the non-linear retrieval of the data may lower the detection limits. One can then test if the residuals are simply Gaussian noise, or if they may contain interesting features.

Interestingly, a molecular specie not well mixed in the atmosphere can be most easily detected with our approach.

The last section presented the results of the application on real NOMAD-SO data, using orders 119, 134 and 136, selected as they are representative of the baseline strategy of measurements in NOMAD, allowing characterization of H_2O and potential detection of CH_4. The outcome is that no CH_4 has been identified, but H_2O and CO_2 are detected. Interestingly a new set of spectral lines has been discovered in the NOMAD data. These lines has been first detected in the ACS instrument data and attributed to CO_2 magnetic dipole transition (Trokhimovskiy et al., 2020). We thus confirm their presence with our current analysis.

One way to go back to the data is to pick the real data with the highest source contribution $\dot{\mathbf{A}}$. Our quicklook analysis is thus only a starting point of a more complete scientific analysis. This second step will require much more prior information (chemical compounds, fundamental spectroscopic constants, radiative transfer model, ...).

Future work should apply the proposed approach to other datasets, such as other NOMAD-SO orders, or other spectroscopic datasets (including hyperspectral images) from laboratory measurements, ground based telescopes or space-born spectrometers. The approach is generic enough to treat datasets that can be at first order approximated to a linear mixture.

Acknowledgements

We acknowledge support from the "Institut National des Sciences de l'Univers" (INSU), the "Centre National de la Recherche Scientifique" (CNRS) and "Centre National d'Etudes Spatiales" (CNES) through the "Programme National de Planétologie" and the ExoMars TGO programs. The NOMAD experiment is led by the Royal Belgian Institute for Space Aeronomy (BIRA-IASB), assisted by Co-PI teams from Spain (IAA-CSIC), Italy (INAF-IAPS), and the United Kingdom (Open University). This project acknowledges funding by the Belgian Science Policy Office (BELSPO), with the financial and contractual coordination by the ESA Prodex Office (PEA 4000103401, 4000121493), by Spanish Ministry of Science and Innovation (MCIU) and by European funds under

grants PGC2018-101836-B-I00 and ESP2017-87143-R (MINECO/FEDER), as well as by UK Space Agency through grants ST/R005761/1, ST/P001262/1, ST/R001405/1 and ST/R001405/1 and Italian Space Agency through grant 2018-2-HH.0. This work was supported by the Belgian Fonds de la Recherche Scientifique - FNRS under grant number 30442502 (ET-HOME). The IAA/CSIC team acknowledges financial support from the State Agency for Research of the Spanish MCIU through the Center of Excellence Severo Ochoa award for the Instituto de Astrofísica de Andalucía (SEV-2017-0709). US investigators were supported by the National Aeronautics and Space Administration. Canadian investigators were supported by the Canadian Space Agency.

Aoki, S., Vandaele, A. C., Daerden, F., Villanueva, G. L., Liuzzi, G., Thomas, I. R., Erwin, J. T., Trompet, L., Robert, S., Neary, L., Viscardy, S., Clancy, R. T., Smith, M. D., Lopez-Valverde, M. A., Hill, B., Ristic, B., Patel, M. R., Bellucci, G., Lopez-Moreno, J.-J., the NOMAD team, 2019. Water vapor vertical profiles on mars in dust storms observed by tgo/nomad. Journal of Geophysical Research: Planets 124 (12), 3482–3497.

Bertaux, J.-L., Fonteyn, D., Korablev, O., Chassefière, E., Dimarellis, E., Dubois, J., Hauchecorne, A., Cabane, M., Rannou, P., Levasseur-Regourd, A., Cernogora, G., Quemerais, E., Hermans, C., Kockarts, G., Lippens, C., Maziere, M., Moreau, D., Muller, C., Neefs, B., Simon, P., Forget, F., Hourdin, F., Talagrand, O., Moroz, V., Rodin, A., Sandel, B., Stern, A., oct 2000. The study of the martian atmosphere from top to bottom with SPICAM light on mars express. Planetary and Space Science 48 (12-14), 1303–1320.

Bertaux, J.-L., Nevejans, D., Korablev, O., Villard, E., Quémerais, E., Neefs, E., Montmessin, F., Leblanc, F., Dubois, J., Dimarellis, E., Hauchecorne, A., Lefèvre, F., Rannou, P., Chaufray, J., Cabane, M., Cernogora, G., Souchon, G., Semelin, F., Reberac, A., Ransbeek, E. V., Berkenbosch, S., Clairquin, R., Muller, C., Forget, F., Hourdin, F., Talagrand, O., Rodin, A., Fedorova, A., Stepanov, A., Vinogradov, I., Kiselev, A., Kalinnikov, Y., Durry, G., Sandel, B., Stern, A., Gérard, J., oct 2007. SPICAV on venus express: Three spectrometers to study the global structure and composition of the venus atmosphere. Planetary and Space Science 55 (12), 1673–1700.

Bovensmann, H., Burrows, J. P., Buchwitz, M., Frerick, J., Noël, S., Rozanov, V. V., Chance, K. V., Goede, A. P. H., jan 1999. SCIAMACHY: Mission objectives and measurement modes. Journal of the Atmospheric Sciences 56 (2), 127–150.

Daerden, F., Neary, L., Viscardy, S., Muñoz, A. G., Clancy, R., Smith, M., Encrenaz, T., Fedorova, A., 2019. Mars atmospheric chemistry simulations with the gem-mars general circulation model. Icarus 326, 197–224.

Dobigeon, N., Moussaoui, S., Tourneret, J.-Y., Carteret, C., Dec. 2009. Bayesian separation of spectral sources under non-negativity and full additivity constraints. Signal Processing 89 (12), 2657–2669.

URL http://www.sciencedirect.com/science/article/B6V18-4W9XDSW-2/2/f3d4b6f457b91e5ccfcce8ffcf41bb18

Eilers, P. H., Boelens, H. F., 2005. Baseline correction with asymmetric least squares smoothing.

Erard, S., Drossart, P., Piccioni, G., Jan. 2009. Multivariate analysis of visible and infrared thermal imaging spectrometer (virtis) venus express nightside and limb observations. J. Geophys. Res. 114, –.
URL http://dx.doi.org/10.1029/2008JE003116

Faisal, M., Windholz, L., Kröger, S., apr 2020. Systematic investigations of the hyperfine structure constants of niobium i levels. part i: Constants of upper odd parity energy levels between 16,672 and 31,025 cm-1 and discovery of a new level. Journal of Quantitative Spectroscopy and Radiative Transfer 245, 106873.

Fissiaux, L., Delière, Q., Blanquet, G., Robert, S., Vandaele, A. C., Lepère, M., mar 2014. CO2-broadening coefficients in the $\nu 4$ fundamental band of methane at room temperature and application to CO2-rich planetary atmospheres. Journal of Molecular Spectroscopy 297, 35–40.

Gamache, R. R., Farese, M., Renaud, C. L., aug 2016. A spectral line list for water isotopologues in the 1100–4100 cm-1 region for application to CO2-rich planetary atmospheres. Journal of Molecular Spectroscopy 326, 144–150.

Geminale, A., Grassi, D., Altieri, F., Serventi, G., Carli, C., Carrozzo, F., Sgavetti, M., Orosei, R., D'Aversa, E., Bellucci, G., Frigeri, A., 2015. Removal of atmospheric features in near infrared spectra by means of principal component analysis and target transformation on mars: I. method. Icarus 253 (0), 51 – 65.
URL http://www.sciencedirect.com/science/article/pii/S0019103515000640

Gillis, N., Glineur, F., apr 2012. Accelerated multiplicative updates and hierarchical ALS algorithms for nonnegative matrix factorization. Neural Computation 24 (4), 1085–1105.

Giuranna, M., Viscardy, S., Daerden, F., Neary, L., Etiope, G., Oehler, D., Formisano, V., Aronica, A., Wolkenberg, P., Aoki, S., Cardesín-Moinelo, A., de la Parra, J. M.-Y., Merritt, D., Amoroso, M., apr 2019. Independent confirmation of a methane spike on mars and a source region east of gale crater. Nature Geoscience 12 (5), 326–332.

Gordon, I., Rothman, L., Hill, C., Kochanov, R., Tan, Y., Bernath, P., Birk, M., Boudon, V., Campargue, A., Chance, K., Drouin, B., Flaud, J.-M., Gamache, R., Hodges, J., Jacquemart, D., Perevalov, V., Perrin, A., Shine, K., Smith, M.-A., Tennyson, J., Toon, G., Tran, H., Tyuterev, V., Barbe, A., Császár, A., Devi, V., Furtenbacher, T., Harrison, J., Hartmann, J.-M., Jolly, A., Johnson,

T., Karman, T., Kleiner, I., Kyuberis, A., Loos, J., Lyulin, O., Massie, S., Mikhailenko, S., Moazzen-Ahmadi, N., Müller, H., Naumenko, O., Nikitin, A., Polyansky, O., Rey, M., Rotger, M., Sharpe, S., Sung, K., Starikova, E., Tashkun, S., Auwera, J. V., Wagner, G., Wilzewski, J., Wcislo, P., Yu, S., Zak, E., 2017. The hitran2016 molecular spectroscopic database. Journal of Quantitative Spectroscopy and Radiative Transfer 203, 3 – 69.

Herr, K. C., Pimentel, G. C., Jan. 1970. Evidence for Solid Carbon Dioxide in the Upper Atmosphere of Mars. Science 167, 47–49.

Hinrich, J. L., Mørup, M., 2018. Probabilistic sparse non-negative matrix factorization. In: Latent Variable Analysis and Signal Separation. Springer International Publishing, pp. 488–498.

Kim, H., Park, H., may 2007. Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis. Bioinformatics 23 (12), 1495–1502.

Korablev, O., , Vandaele, A. C., Montmessin, F., Fedorova, A. A., Trokhimovskiy, A., Forget, F., Lefèvre, F., Daerden, F., Thomas, I. R., Trompet, L., Erwin, J. T., Aoki, S., Robert, S., Neary, L., Viscardy, S., Grigoriev, A. V., Ignatiev, N. I., Shakun, A., Patrakeev, A., Belyaev, D. A., Bertaux, J.-L., Olsen, K. S., Baggio, L., Alday, J., Ivanov, Y. S., Ristic, B., Mason, J., Willame, Y., Depiesse, C., Hetey, L., Berkenbosch, S., Clairquin, R., Queirolo, C., Beeckman, B., Neefs, E., Patel, M. R., Bellucci, G., López-Moreno, J.-J., Wilson, C. F., Etiope, G., Zelenyi, L., Svedhem, H., Vago, J. L., The ACS and NOMAD Team, apr 2019. No detection of methane on mars from early ExoMars trace gas orbiter observations. Nature 568 (7753), 517–520.

Lee, D. D., Seung, H. S., Oct. 1999. Learning the parts of objects by non-negative matrix factorization. Nature 401 (6755), 788–791.
URL http://dx.doi.org/10.1038/44565

Liuzzi, G., Villanueva, G., Mumma, M., Smith, M., Daerden, F., Ristic, B., Thomas, I., Vandaele, A., Patel, M., López-Moreno, J., Bellucci, G., 2019. Methane on mars: New insights into the sensitivity of ch4 with the NOMAD/ExoMars spectrometer through its first in-flight calibration. Icarus 321, 671–690.

López-Valverde, M., López-Puertas, M., López-Moreno, J., Formisano, V., Grassi, D., Maturilli, A., Lellouch, E., Drossart, P., aug 2005. Analysis of non-LTE emissions at in the martian atmosphere as observed by PFS/mars express and SWS/ISO. Planetary and Space Science 53 (10), 1079–1087.

Moores, J. E., Gough, R. V., Martinez, G. M., Meslin, P.-Y., Smith, C. L., Atreya, S. K., Mahaffy, P. R., Newman, C. E., Webster, C. R., mar 2019. Methane seasonal cycle at gale crater on mars consistent with regolith adsorption and diffusion. Nature Geoscience 12 (5), 321–325.

Moussaoui, S., Brie, D., Mohammad-Djafari, A., Carteret, C., 2006. Separation of non-negative mixture of non-negative sources using a bayesian approach and mcmc sampling. Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Signal Processing, IEEE Transactions on] 54 (11), 4133–4145.

Moussaoui, S., Hauksdóttir, H., Schmidt, F., Jutten, C., Chanussot, J., Brie, D., Douté, S., Benediktsson, J., Jun. 2008. On the decomposition of mars hyperspectral data by ica and bayesian positive source separation. Neurocomputing 71 (10-12), 2194–2208.
URL http://www.sciencedirect.com/science/article/B6V10-4RV17HX-4/1/739950d227add850ec0720718c1c2362

Neary, L., Daerden, F., 2018. The gem-mars general circulation model for mars: Description and evaluation. Icarus 300, 458–476.

Neefs, E., Vandaele, A. C., Drummond, R., Thomas, I. R., Berkenbosch, S., Clairquin, R., Delanoye, S., Ristic, B., Maes, J., Bonnewijn, S., Pieck, G., Equeter, E., Depiesse, C., Daerden, F., Ransbeeck, E. V., Nevejans, D., Rodriguez-Gómez, J., López-Moreno, J.-J., Sanz, R., Morales, R., Candini, G. P., Pastor-Morales, M. C., del Moral, B. A., Jeronimo-Zafra, J.-M., Gómez-López, J. M., Alonso-Rodrigo, G., Pérez-Grande, I., Cubas, J., Gomez-Sanjuan, A. M., Navarro-Medina, F., Thibert, T., Patel, M. R., Bellucci, G., Vos, L. D., Lesschaeve, S., Vooren, N. V., Moelans, W., Aballea, L., Glorieux, S., Baeke, A., Kendall, D., Neef, J. D., Soenen, A., Puech, P.-Y., Ward, J., Jamoye, J.-F., Diez, D., Vicario-Arroyo, A., Jankowski, M., sep 2015. NOMAD spectrometer on the ExoMars trace gas orbiter mission: part 1—design, manufacturing and testing of the infrared channels. Applied Optics 54 (28), 8494.

Penttilä, A., Martikainen, J., Gritsevich, M., Muinonen, K., feb 2018. Laboratory spectroscopy of meteorite samples at UV-vis-NIR wavelengths: Analysis and discrimination by principal components analysis. Journal of Quantitative Spectroscopy and Radiative Transfer 206, 189–197.

Schmidt, F., Schmidt, A., Treguier, E., Guiheneuf, M., Moussaoui, S., Dobigeon, N., 2010. Implementation strategies for hyperspectral unmixing using bayesian source separation. Geoscience and Remote Sensing, IEEE Transactions 48 (11), 4003–4013.

Shashilov, V. A., Xu, M., Ermolenkov, V. V., Lednev, I. K., nov 2006. Latent variable analysis of raman spectra for structural characterization of proteins. Journal of Quantitative Spectroscopy and Radiative Transfer 102 (1), 46–61.

Smith, G. R., Hunten, D. M., 1990. Study of planetary atmospheres by absorptive occultations. Reviews of Geophysics 28 (2), 117.

Trokhimovskiy, A., Perevalov, V., Korablev, O., Fedorova, A. F., Olsen, K. S., Bertaux, J.-L., Patrakeev, A., Shakun, A., Montmessin, F., Lefèvre, F., Lukashevskaya, A., jul 2020. First observation of the magnetic dipole CO_2 absorption band at 3.3 μm in the atmosphere of mars by the ExoMars trace gas orbiter ACS instrument. Astronomy & Astrophysics 639, A142.

Vandaele, A., Neefs, E., Drummond, R., Thomas, I., Daerden, F., Lopez-Moreno, J.-J., Rodriguez, J., Patel, M., Bellucci, G., Allen, M., Altieri, F., Bolsée, D., Clancy, T., Delanoye, S., Depiesse, C., Cloutis, E., Fedorova, A., Formisano, V., Funke, B., Fussen, D., Geminale, A., Gérard, J.-C., Giuranna, M., Ignatiev, N., Kaminski, J., Karatekin, O., Lefèvre, F., López-Puertas, M., López-Valverde, M., Mahieux, A., McConnell, J., Mumma, M., Neary, L., Renotte, E., Ristic, B., Robert, S., Smith, M., Trokhimovsky, S., Auwera, J. V., Villanueva, G., Whiteway, J., Wilquet, V., Wolff, M., dec 2015. Science objectives and performances of NOMAD, a spectrometer suite for the ExoMars TGO mission. Planetary and Space Science 119, 233–249.

Vandaele, A. C., , Lopez-Moreno, J.-J., Patel, M. R., Bellucci, G., Daerden, F., Ristic, B., Robert, S., Thomas, I. R., Wilquet, V., Allen, M., Alonso-Rodrigo, G., Altieri, F., Aoki, S., Bolsée, D., Clancy, T., Cloutis, E., Depiesse, C., Drummond, R., Fedorova, A., Formisano, V., Funke, B., González-Galindo, F., Geminale, A., Gérard, J.-C., Giuranna, M., Hetey, L., Ignatiev, N., Kaminski, J., Karatekin, O., Kasaba, Y., Leese, M., Lefèvre, F., Lewis, S. R., López-Puertas, M., López-Valverde, M., Mahieux, A., Mason, J., McConnell, J., Mumma, M., Neary, L., Neefs, E., Renotte, E., Rodriguez-Gomez, J., Sindoni, G., Smith, M., Stiepen, A., Trokhimovsky, A., Auwera, J. V., Villanueva, G., Viscardy, S., Whiteway, J., Willame, Y., Wolff, M., jun 2018. NOMAD, an integrated suite of three spectrometers for the ExoMars trace gas mission: Technical description, science objectives and expected performance. Space Science Reviews 214 (5).

Villanueva, G., Smith, M., Protopapa, S., Faggi, S., Mandell, A., sep 2018. Planetary spectrum generator: An accurate online radiative transfer suite for atmospheres, comets, small bodies and exoplanets. Journal of Quantitative Spectroscopy and Radiative Transfer 217, 86–104.