Advanced data pre–processing for Comprehensive two–dimensional Gas Chromatography with Vacuum Ultraviolet Spectroscopy detection

Aleksandra Lelevic, Vincent Souchon, Christophe Geantet, Chantal Lorentz, Maxime Moreaud

Journal of Separation Science, 2021, doi: 10.1002/jssc.202100528

Advanced data preprocessing for Comprehensive two-dimensional Gas Chromatography with Vacuum Ultraviolet Spectroscopy detection

Aleksandra Lelevic[1,2,*], Vincent Souchon[1],Christophe Geantet[2], Chantal Lorentz[2], Maxime Moreaud[1,*]

1. IFP Energies nouvelles, Rond-point de l'échangeur de Solaize BP 3 69360 Solaize France
2. Univ Lyon, Université Claude Bernard Lyon 1, CNRS, IRCELYON, F-69626, Villeurbanne, France

* Authors for correspondence: aleksandra.lelevic@ifpen.fr, maxime.moreaud@ifpen.fr

Running title: Advanced data preprocessing for GC×GC-VUV

Mrs. Aleksandra Lelevic, 74 Route de Genève, 69140, Rillieux-la-Pape, France, contact@aleksandralelevic.com

Abbreviations

1D GC – One-dimensional gas chromatography

2D GC – Two-dimensional gas chromatography

GC×GC – Comprehensive two-dimensional gas chromatography

S/N – Signal-to-noise

VUV – Vacuum ultraviolet spectroscopy

Key words: Vacuum ultraviolet spectroscopy, two-dimensional chromatography, data preprocessing, baseline correction, noise reduction

Abstract

Comprehensive two-dimensional gas chromatography with vacuum ultraviolet detection results in sizable data for which noise and baseline drift ought to be corrected. As the data is acquired from multiple channels, preprocessing steps have to be applied to the data from all channels while being robust and rather fast with respect to the significant size of the data. In this study, we have described advanced data preprocessing techniques for such data which were not available in the existing commercial software solutions and which were dedicated primarily to noise and baseline correction. Noise reduction was performed on both the spectral and the time dimension. For the baseline correction, a morphological approach based on iterated convolutions and rectifier operations was proposed. On the spectral dimension, much less noisy and reliable spectra were obtained. From a quantitative point of view, mentioned preprocessing steps significantly improved signal to noise ratio for the analyte detection (circa 6 times in this study). These preprocessing methods were integrated into the plug im! platform (https://www.plugim.fr/).

1. Introduction

Analysis of the multidimensional chromatography data can be extremely complex, hence strategies and approaches for the data analysis must be carefully devised in order to obtain the most from the available information. Before applying intricate data analysis techniques, comprehensive two-dimensional gas chromatography (GC×GC) data must be subjected to preprocessing. Preprocessing techniques are applied with the goal of eliminating irrelevant chemical variations and often include noise reduction procedures, baseline correction, normalization, and retention time alignment [1,2]. After data preprocessing, the challenge is usually to identify compounds' peaks and perform quantification [2].

Vacuum ultraviolet detector (VUV) has been recently successfully hyphenated with GC×GC analysis [3–7]. This hyphenation has a lot of potential for the exploration of complex samples, particularly in the field of fuel analysis. GC×GC-VUV data processing is specific to the chosen application; however, it generally begins with integrating an interval of a spectral band. This operation is called 'spectral filtering'. Further processing procedure can be reduced to the rather classical sequence of steps: baseline correction, detection of peaks corresponding to compounds of interest and peak integration [1,8]. Concerning softwares for the processing and analysis of GC×GC-VUV data, previous studies mainly rely on homemade Matlab software [3,4] and GC Image [9] which is as far as our knowledge the only commercial two-dimensional chromatography software supporting direct import of the GC×GC-VUV data.

In this work, we have proposed a strategy for the advanced preprocessing of the GC×GC-VUV data. Adaptive noise reduction and robust and fast baseline correction methods were described. For the baseline correction, a morphological approach based on iterated convolutions and rectifier operations was presented. VUV detector blank signal subtraction was also performed. This data preprocessing approach was evaluated regarding the enhancement of the detection

limits in the GC×GC-VUV analysis as well as the improvement of the quality of the data obtained for real world complex samples.

2. Data preprocessing methods

2.1. Noise reduction

GC×GC-VUV data can be quite noisy, owing to the high acquisition frequencies which are necessary for the GC×GC analysis. As the VUV detector possesses moderate sensitivity, in the case of the GC×GC-VUV analysis it is important to apply the noise reduction as this can help to significantly improve the obtained signal to noise (S/N) ratio [10].

Our approach for lowering the noise employs a wide range of noise reduction filters whose parameterization, linked to the size of the calculation neighborhood, adapts automatically and locally to the signal. Classical linear filtering techniques express the smoothed value at one point as a linear combination of the values of the samples located in the interval around the considered point. This filtering can be interpreted in terms of a scale corresponding intuitively to the length and the weighting applied. The choice of the right filtering scale is crucial in order to preserve as much as possible the local properties of the signal we wish to analyze. However, this good scale (unknown a priori) can vary from one point to another, and most of the time leads the user to choose filtering parameters causing a compromise on the whole signal.

2.1.1. Noise reduction on the spectral dimension

Our most important strategy to reduce the noise for the GC×GC-VUV data was based on the observation that any VUV spectrum inherently possesses monotonous nature, without high frequency variations. Thus, VUV spectrum at every point of the GC×GC chromatogram can be smoothed efficiently by choosing a number of points in the averaging window, calculating the mean value in the window and repeating this operation until the entire observed spectral range is covered (e.g., 125-430 nm).

2.1.2. Noise reduction on the time dimension

Above mentioned approach ensured most of the GC×GC-VUV data noise reduction, however an additional smoothing was also applied on the chromatographic dimension (for all wavelength channels). This approach was already published in [11] and can be described as follows.

Let *s(x)* be a signal with noise to correct and composed by a set of peaks. Let us agree on the terms, "smoothing the noise" in our context, means eliminating peaks of the signal *s* corresponding to noise while modifying as little as possible the other peaks of *s* corresponding to useful information. It is assumed that noise peaks correspond to peaks of lower intensity and width than peaks containing useful information. The elimination of a peak is equivalent to transforming *s* locally for this peak into a slowly varying function passing through this peak at best. On the other hand, *s* must be modified as little as possible (or not at all) for the peaks containing information (see Figure 1). Let f_α be a smoothing function parameterized by the variable α. The higher the α is, the more important the smoothing is. Such a function can be a Gaussian kernel. Let *M* be the maximum value that parameter α can reach. All peaks of *s* with width greater than *M* are considered relevant information. Finally, let *m* be an incremental step for the parameter α. For each point *x*, a parameter α is calculated to adapt the smoothing function locally. The optimal parameters $\alpha_*(x)$ are obtained for the minimum residue between two consecutive smoothed functions:

$$\alpha_*(x) = \alpha \left| \min_{\alpha \in [m,M]} |s * f_\alpha(x) - s * f_{\alpha - m}(x)| \right. \quad (1)$$

The corrected noise signal is then obtained by convolution with $f_{\alpha_*(x)}$:

$$s * f_{\alpha_*(x)} \quad (2)$$

It should be noted that eq. 1 can be seen in a continuous framework as the search for each x of the parameter α allowing to obtain at x the minimum of the derivative with respect to α.

The advantages of the proposed approach for the noise correction on the chromatographic dimensions include: only one parameter to tune, conservation of peaks intensity while maximally reducing the noise (adaptive), fast computation time with real time preview. Disadvantages include: necessity of trial and error adjustment.

2.2. Baseline correction

Observed detector signal in general can be defined as a sum of the useful information, high frequency perturbation i.e. noise (no information) and low frequency trend i.e. baseline (no information). Baseline drift is a common problem in chromatographic studies. Baseline drift in the case of the GC×GC-VUV analysis can occur due to the column stationary phase bleed and/or low-frequency variations in the detector and/or instrument parameters (e.g., flow). Baseline correction is performed with the goal to improve analytes' detection and quantification which otherwise may be affected by baseline perturbations. Baseline correction methods ought to be carefully administered in order to preserve the relevant information (even broadened coeluting peaks which might be incorrectly attributed to low frequency baseline variations) but also overfitting ought to be avoided. In the case of multichannel detectors (such as VUV), baseline correction ought to be applied for all channels, following by baseline noise being centered on zero for the entire length of the chromatographic separation.

Baseline correction can be based on physical or mathematical approaches. Physical approaches usually require changes in the instrumental set-up. These are for example wavelength-shifted excitation [12–14] and time gating techniques [15–17]. Mathematical approaches mainly include polynomial fitting based methods, penalized least square based methods, first derivative

based methods, peak detection and interpolation methods, peak detection and interpolation, wavelet transform based methods and morphology based methods [18,19].

Morphological approach to baseline correction consists in defining a baseline as a rigid / deformable element which can be placed under a curve (kind of elastic more or less stretched under a curve). Basic morphological operators include: erosion, dilation, opening and closing. Mathematical morphology is mostly applied in the image processing field. For the baseline correction, the application of mathematical morphology is considered mostly for the Raman spectroscopy signal [18,20–23]. The advantages of the mathematical morphology for the baseline correction include possibility of fast, robust and automated methods which can be favorable for many applications, but also that they are well adapted to the analysis of 1D GC and 2D GC data. Early implementation of the approach based on morphology operations was provided in the work by Perez-Pueyo et al. [20] who proposed an automated approach based on changing the size of the structuring elements and calculating the opening of the spectrum until three consecutive equal openings are obtained. Further top-hat transformation (subtraction of an opening operation to the signal) was applied to obtain a baseline free spectrum. Disadvantages of this approach involved possible distortion of Raman peaks and estimation of the baseline which does not correspond to the conventional smooth nature of the background signal. Chen and Dai [18] subsequently modified this approach by proposing iteration of the morphological operations leading to the gradual estimation of the baseline. Koch et al. [23] developed an approach involving iterations of erosion operations which provides the estimation of the baseline together with mollification which smooths the baseline. Chen et al. [19] proposed automated baseline correction based on alternating sequential filters (iterations of closing and opening operations) and convolutions. Chen and Hsu [24,25] developed an approach involving iterations of convolutions and erosion operations.

Morphological closing operation, with horizontal segment or disc for example, leads to interesting results. Disadvantages of this approach however include: discrete implementation, high time computation and standard image processing algorithms are not directly transposable to the signal (for point (x,y), x can be discrete but y has real value). Segment with local fitting length can be also employed [2,26] producing similar result to closing operation.

The approach proposed in this work for the GC×GC-VUV data baseline estimation was taken from a previous patent [27] and belongs to approaches based on iterations of convolutions and thresholding operations [19,23,24]. At each iteration, a baseline passing through the fluctuations of signal 's' was obtained from a convolution operation. A rectifier was used to extract the lower part of this curve. The smoothing operation was repeated only on this lower part, thus allowing to gradually constrain the curve to position itself close to the local minima of s. One iteration of this operator was formulated as follows:

$$F_\alpha(s) = s * f_\alpha - \mathrm{ReLU}(s * f_\alpha - s) \quad (3)$$

$$\mathrm{ReLU}(x) = x^+ = \max(0,x) \quad (4)$$

with f_α a Gaussian kernel with parameter α.

Baseline was defined as n compositions of F_α:

$$b = (F_\alpha \circ F_\alpha \circ \ldots \circ F_\alpha)(s) * f_\alpha \quad (5)$$

$$b=(F_{\alpha})^{n}(s) * f_{\alpha} \quad (6)$$

Illustration of the proposed approach is provided in Figure 2. Figure 3 shows the influence of increasing the number of iterations.

Advantages of the proposed approach for the baseline correction include: only two parameters, very fast computation with real time preview accessible for GC×GC-VUV data (> 1Gb), handling signal with mixtures of low and high peak intensities, no model dependency for baseline. Disadvantages include: non-intuitive settings, i.e. need to make a "trial and error" adjustment.

2.3. Detector spectral blank subtraction

Inherent to the GC×GC-VUV data is also a detector blank signal which is varying depending on the observation wavelength. It can be seen in Figure S5 in Supporting Information, which shows measured blank detector signal, that noise is more dominant at certain wavelengths, i.e. in the beginning and the end of chromatogram. If detector blank is not subtracted, then VUV spectrum extracted from a chromatogram zone may contain a summed contribution of detector blank signal which can distort the spectrum.

Our approach for subtracting the detector blank consisted in estimating a contribution of detector in a GC×GC chromatogram zone without the analyte signal (see Figure S6) and then calculating its average for a single pixel (Figure S7). This detector blank spectrum was then subtracted from spectra at each pixel of the GC×GC chromatogram.

3. Materials and methods

For the investigation of the improvement of detection limits for the GC×GC-VUV data through application of described preprocessing methods, a standard test mixture was first analyzed. It

consisted of various hydrocarbons diluted in toluene, all of purity 95% or greater (detailed composition was provided in the Supporting Information). Also analyzed is was a real world sample, a gas oil diluted in toluene, provided by IFP Energies Nouvelles (Solaize, France).

For GC×GC-VUV experiments, an Agilent 7890A gas chromatograph under hydrogen as carrier gas equipped with a G3486A CFT forward fill/flush differential flow modulator was employed (Agilent Technologies, Inc., Santa Clara, CA, United States). A normal configuration column set was chosen: DB-1 column (100% dimethyl polysiloxane; 20 m, 0.1 mm ID, 0.4 μm; Agilent Technologies, Inc.) was used in the first dimension whereas BPX-50 (50% Phenyl Polysilphenylene-siloxane, 3.2 m, 0.25 mm ID, 0.25 μm; SGE Analytical Science, Ringwood, Australia) was used in the second dimension. Normal configuration for the column set was preferred as it permits separation of hydrocarbons according to increasing carbon number in the 1^{st} dimension and their polarity in the 2^{nd} dimension and generates highly ordered 2D chromatograms with good orthogonality for a number of samples.

For gas oil analysis: Split injections were performed with a temperature programmed MMI Agilent inlet operated in split mode (1 μl injected, 80:1 split ratio). Injection port was heated to 300°C, then ramped to 330°C at 500°C/min, where it remained isothermal during 5 min. Flow rates in first and second dimensions were set to 0.15 ml/min (inlet pressure 23.2 psig, average velocity 16.5 cm/s) and 13 ml/min (modulator pressure 8.33 psig, average velocity 368 cm/s), respectively. Oven temperature program was 50 °C (3 min) to 325 °C at 2.5 °C/min. Modulation period was set to 4.5 s while modulation injection time was set to 0.19 s. Modulation parameters were optimized according to our previous work [28].

For standard mixture analysis: 1 μl injections with 50:1 split ratio, same temperature programming as for the gas oil analysis. Flow rates in the first and the second dimension were

220 the same as in gas oil analysis, so were the modulation parameters. Oven temperature program
221 was 50 °C (3 min) to 245 °C at 2.5 °C/min.

222 VGA-101 (VUV Analytics, Inc., Austin, TX, United States) detector was employed. VUV
223 conditions were as follows: wavelength range, 125–430 nm; acquisition frequency, 33.33 Hz;
224 flow cell and transfer line temperature 325 °C, make-up gas pressure 0.15 psig.

225 Agilent ChemStation B.04.03-SP1 and VUVisionTM 3.0.1 were used for the GC×GC-VUV
226 data acquisition. VUV Data Exporter v6.10 was used for VUV data export in 'txt' format. plug
227 im! software was employed for the data preprocessing and extraction of spectra [29]. VUV
228 Spectra LLB v5.05.522 was used for the estimation of spectral similarity with VUVison spectral
229 library spectra. GC Image 2.7 software was employed for the data integration.

230

231 ## 4. Results and discussion

232 ### 4.1. Improving spectral data quality with preprocessing

233 To illustrate the importance of the data preprocessing for the GC×GC-VUV data, a gas oil was
234 first analyzed and a summed VUV spectrum was extracted for the zone of gas oil chromatogram
235 where alkenes and cycloalkanes typically coelute (Figure 4). As absorbance is additive, having
236 a reliable spectrum from this zone may allow to discern individual quantities of these two
237 groups through spectral decomposition. Extracted spectrum from the raw data (see Figure 5A
238 trace in grey) illustrates the necessity of the preprocessing as the spectrum was very noisy and
239 it exhibited a shape that is not characteristic for the mixture of the two mentioned groups of
240 compounds (see the VUV library spectra for the two representatives of alkenes and
241 cycloalkanes in Figure 5B).

242 For the noise correction, first VUV spectral filtering (noise reduction on spectral dimension)
243 was applied which ensured the majority of the noise smoothing effect. Figure S1 in the

Supporting Information illustrates the influence of the choice of the averaging window size on the recalculated spectrum. A window size (M) of 15 points ensured the maximum noise reduction without the loss of the significant spectral features. M value corresponds to half the size of the convolution window (i.e. 31 points were used for the averaging which corresponds to a spectral range of 6.2 nm – VUV data wavelength step for the used instrument was fixed at 0.2 nm). All points on the spectral dimension were recalculated and the dimensionality of the data was preserved. Satisfactory performance was obtained over the entire GC×GC chromatogram as VUV spectra features are in general not very elaborate and hydrocarbons exhibit broad absorption bands. Subsequently, additional chromatogram smoothing on the time dimension was applied. Figure S2 shows the influence of the choice of the noise filter value. It can be seen how analyte signal was preserved while noise was significantly reduced when choosing filter value M = 5, i.e. for each point the averaging was adjusted automatically with a maximum number of 11 points (corresponding to a time window of 0.33 s). After applying noise correction, summed spectrum was again extracted from alkenes and cycloalkanes zone (Figure 5A in blue).

Further preprocessing was applied in the form of the baseline correction. Figures S3 and S4 illustrate the estimation of the baseline based on the choice of the number of iterations and convolution kernel size. As it can be seen a very good estimation was obtained with these parameters set to 15 and 100, respectively. Gaussian kernel size ought to correspond approximately to the width of the widest peak of interest in the chromatogram, as if the length is too short loss of signal for wider peaks will be incurred. Very important is to add that GC×GC-VUV data can be very large in size (often >1Gb) which is why very fast computation with real time preview are advantages of our approach.

After applying baseline correction, extracted spectrum (Figure 5A in red) was finally more in line with what can be expected for a mixture of alkenes and cycloalkanes. However, absorbance

never reached zero at wavelengths above 200 nm, which is expected for species in question (conjugated diolefins can however demonstrate absorbance up to 240 nm) and was even demonstrating an increasing trend at around 240 nm which on the other hand could have been explained only by the detector blank signal. To correct for this artificial contribution, detector blank (Figure S7) was subtracted from spectra at each pixel of the GC×GC chromatogram. Finally, improved alkenes/cycloalkanes mixture spectrum was obtained (Figure 5A in green). In the final spectrum in green, the expected pi-bond absorption bond of alkenes centered at 180-190 nm was observed. In this way, extracted reliable spectra for zones containing coeluting species can be further employed to estimate proportions of each of the species in the mixture spectrum provided their reference spectra per unit mass are known.

4.2. Improving detection limits with data preprocessing

In order to investigate the effect of the GC×GC-VUV data preprocessing on the quantitative performance and estimated detection limits, dilutions of the test mixture containing various hydrocarbons were prepared: 2x, 10x, 20x, 50x, 100x, 150x, 200x, 300x and 500x. These solutions contained respectively ca. 5000 ppm, 1000 ppm, 500 ppm, 200 ppm, 100 ppm, 70 ppm, 50 ppm, 30 and 20 ppm of the present analytes, including linear alkanes, alkenes, cycloalkanes and monoaromatics. Test mixture along with all the diluted solutions was analyzed by GC×GC-VUV. Obtained chromatogram for the test mixture solution is shown in Figure 6.

First, GC×GC-VUV data integration was performed using the GC Image software. Average absorbance in the 125-240 nm spectral range was calculated as no signal was obtained for longer wavelengths, i.e. in the 240-430 nm range. Integration was performed for the data for which baseline correction, the only available preprocessing operation available in GC Image, was applied.

Subsequently, all acquired data was subjected to the developed preprocessing steps, which involved noise reduction, baseline correction and detector blank subtraction according to the approaches described previously. Parameters chosen for the preprocessing algorithms were kept the same as for the gas oil data.

Figure 7A illustrates the extracted VUV spectra obtained for n-decane, 1-decene and m-xylene from the GC×GC-VUV chromatogram of the 150x dilution of the test mixture before any preprocessing. As observed previously for the gas oil sample, significantly improved spectra were obtained after preprocessing (Figure 7B). m-xylene spectrum was clearly representative of monoaromatic species, 1-decene demonstrated profile characteristic of alkenes, while n-decane exhibited monotonous spectrum as expected for n-paraffins. Spectral similarity match with corresponding VUV library spectrum was determined for all three analytes before and after data processing. Improved values for both Chi^2 and R^2 were obtained also demonstrating the importance of the data preprocessing (Table S4 in Supporting Information).

After preprocessing, data was imported in the GC Image for the peak integration. For each solution, peak volumes of present species were calculated as well as their signal to noise ratio. Detailed results for the raw data and preprocessed data are provided in the Supporting Information in Table S2 and S3 and were summarized for 3 selected analytes in Figure 8.

For all compounds, ca. 6× improvement of the S/N ratio was obtained when preprocessing was applied. S/N ratio was the highest for the aromatic species and it decreased with the decrease of the degree of unsaturation. Improving S/N ratio naturally helped in improving the limit of detection. We have observed that below ca. 50 ppm of analyte concentration, signal to noise ratio was very low and at this stage peaks were difficult to distinguish from the noise. This can be seen on the example of m-xylene peak in Figure 9A (concentration ca. 20 ppm). After baseline and noise correction this peak was more easily detected, useful information was

preserved while noise was much reduced (Figure 9B). Estimation of spectral similarity with m-xylene VUV library spectrum is provided in Table S5 in Supporting Information demonstrating much improved Chi^2 and R^2 values.

5. Conclusions

In this work, we have proposed advanced preprocessing techniques for the GC×GC-VUV data. For the baseline correction, a morphological approach based on iterated convolutions and rectifier operations was presented. Approaches for the noise reduction and detector blank signal subtraction were also proposed. Improvement of spectral information extracted from real world samples or model mixture containing hydrocarbon species after applying developed preprocessing steps was demonstrated. Moreover, signal to noise ratio improvement of circa 6 times was observed for all model mixture species. All preprocessing methods have been fully integrated in the plug im!, a free open-access signal and image processing software (https://www.plugim.fr.). This software and dedicated modules for the GC×GC-VUV data processing are available to any user and provide open access capabilities which are complementary to currently existing software solutions.

Data availability statement

Data available on request from the authors.

References

[1] Pierce, K. M., Kehimkar, B., Marney, L. C., Hoggard, J. C., Synovec, R. E., Review of chemometric analysis techniques for comprehensive two dimensional separations data. *J. Chromatogr. A* 2012, 1255, 3–11.

[2] Celse, B., Moreaud, M., Duval, L., Cavagnino, D., Gas Chromatography and 2D-Gas Chromatography for Petroleum Industry. The Race for Selectivity. Editions TECHNIP 2013, pp. 99–151.

[3] Gröger, T., Gruber, B., Harrison, D., Saraji-Bozorgzad, M., Mthembu, M., Sutherland, A., Zimmermann, R., A Vacuum Ultraviolet Absorption Array Spectrometer as a Selective Detector for Comprehensive Two-Dimensional Gas Chromatography: Concept and First Results. *Anal. Chem.* 2016, 88, 3031–3039.

[4] Gruber, B., Groeger, T., Harrison, D., Zimmermann, R., Vacuum ultraviolet absorption spectroscopy in combination with comprehensive two-dimensional gas chromatography for the monitoring of volatile organic compounds in breath gas: A feasibility study. *J. Chromatogr. A* 2016, 1464, 141–146.

[5] Zoccali, M., Schug, K. A., Walsh, P., Smuts, J., Mondello, L., Flow-modulated comprehensive two-dimensional gas chromatography combined with a vacuum ultraviolet detector for the analysis of complex mixtures. *J. Chromatogr. A* 2017, 1497, 135–143.

[6] Jennerwein, K. M., Eschner, M., Wilharm, T., Application of GCxGC-VUV and GCxGC-FID for the analysis of common gasoline samples, middle distillates and crude oil distillation cuts. *PEFTEC 2019 Poster* 2019.

[7] Wang, F. C. Y., Comprehensive Two-Dimensional Gas Chromatography Hyphenated with a Vacuum Ultraviolet Spectrometer to Analyze Diesel-A Three-Dimensional Separation (GC × GC × VUV) Approach. *Energy and Fuels* 2020, 34, 8012–8017.

[8] Bertoncini, F., Courtiade-Tholance, M., Thiébaut, D., Gas Chromatography and 2D-Gas Chromatography for Petroleum Industry : The Race for Selectivity. Editions

TECHNIP 2013.

[9] GC Image: Software for Multidimensional Chromatography, https://www.gcimage.com/index.html.

[10] Lelevic, A., Souchon, V., Moreaud, M., Lorentz, C., Geantet, C., Gas chromatography vacuum ultraviolet spectroscopy: A review. *J. Sep. Sci.* 2020, 43, 150–173.

[11] Moreaud, M., Duval, L., Methode d'analyse Chimique Comportant Un Lissage de Diagramme Par Filtre Localement Auto Adaptatif, Patent No. 2 984 490, publ. date 2013.

[12] Gebrekidan, M. T., Knipfer, C., Stelzle, F., Popp, J., Will, S., Braeuer, A., A shifted-excitation Raman difference spectroscopy (SERDS) evaluation strategy for the efficient isolation of Raman spectra from extreme fluorescence interference. *J. Raman Spectrosc.* 2016, 47, 198–209.

[13] McCain, S. T., Willett, R. M., Brady, D. J., Multi-excitation Raman spectroscopy technique for fluorescence rejection. *Opt. Express* 2008, 16, 10975.

[14] De Luca, A. C., Dholakia, K., Mazilu, M., Modulated raman spectroscopy for enhanced cancer diagnosis at the cellular level. *Sensors (Switzerland)* 2015, 15, 13680–13704.

[15] Morris, M. D., Matousek, P., Towrie, M., Parker, A. W., Goodship, A. E., Draper, E. R. C., Kerr-gated time-resolved Raman spectroscopy of equine cortical bone tissue. *J. Biomed. Opt.* 2005, 10, 014014.

[16] Matousek, P., Towrie, M., Stanley, A., Parker, A. W., Efficient rejection of fluorescence from Raman spectra using picosecond Kerr gating. *Appl. Spectrosc.* 1999,

53, 1485–1489.

[17] Draper, E. R. C., Morris, M. D., Camacho, N. P., Matousek, P., Towrie, M., Parker, A. W., Goodship, A. E., Novel assessment of bone using time-resolved transcutaneous Raman spectroscopy. *J. Bone Miner. Res.* 2005, 20, 1968–1972.

[18] Chen, Y., Dai, L., An Automated Baseline Correction Method Based on Iterative Morphological Operations. *Appl. Spectrosc.* 2018, 72, 731–739.

[19] Chen, H., Xu, W., Broderick, N. G. R., An Adaptive and Fully Automated Baseline Correction Method for Raman Spectroscopy Based on Morphological Operations and Mollification. *Appl. Spectrosc.* 2019, 73, 284–293.

[20] Perez-Pueyo, R., Soneira, M. J., Ruiz-Moreno, S., Morphology-based automated baseline removal for raman spectra of artistic pigments. *Appl. Spectrosc.* 2010, 64, 595–600.

[21] Li, Z., Zhan, D. J., Wang, J. J., Huang, J., Xu, Q. S., Zhang, Z. M., Zheng, Y. B., Liang, Y. Z., Wang, H., Morphological weighted penalized least squares for background correction. *Analyst* 2013, 138, 4483–4492.

[22] Liu, H., Zhang, Z., Liu, S., Yan, L., Liu, T., Zhang, T., Joint baseline-correction and denoising for raman spectra. *Appl. Spectrosc.* 2015, 69, 1013–1022.

[23] Koch, M., Suhr, C., Roth, B., Meinhardt-Wollweber, M., Iterative morphological and mollifier-based baseline correction for Raman spectra. *J. Raman Spectrosc.* 2017, 48, 336–342.

[24] Chen, Y. S., Hsu, Y. C., Effective and efficient baseline correction algorithm for Raman spectra. *Lect. Notes Eng. Comput. Sci.* 2019, 2239, 295–298.

[25] Chen, Y.-S., Hsu, Y.-C., IAENG Transactions on Engineering Sciences. WORLD SCIENTIFIC 2020, pp. 21–32.

[26] Moreaud, M., Duval, L., Procédé d'analyse de Signaux Issus de Chromatographie Ou de Diffraction Par Estimation de La Ligne de Base, Patent No. 2 984 509, publ. date 2013.

[27] Itthirad, F., Moreaud, M., Bouabdellah, M., Reconstruction 3d Surfacique Micrométrique, Patent No. EP3074759, publ. date 2015.

[28] Lelevic, A., Souchon, V., Geantet, C., Lorentz, C., Moreaud, M., Quantitative performance of forward fill/flush differential flow modulation for comprehensive two-dimensional gas chromatography. *J. Chromatogr. A* 2020, 1626, 461342.

[29] "plug im!" an open access and customizable software for signal and image processing, https://www.plugim.fr.

List of Figures

and in green-with noise and baseline correction and detector blank subtraction; B) example of the VUV spectrum of a cycloalkane (ethyl cyclohexane) and an alkene (1-tetradecene) from the VUVision® spectral library.

Figure 6 GC×GC-VUV chromatogram of a test mixture-2x dilution (125 nm).

Figure 7 Extracted summed VUV spectra for n-decane, 1-decene and m-xylene (150x dilution, ca. 70 ppm): A) without preprocessing, B) after noise, baseline correction and detector blank subtraction.

Figure 8 log S/N ratio vs. log Concentration for n-decane, 1-decene and m-xylene for data with baseline correction in GC Image (in red) and with full preprocessing in plug im! software (in blue). S/N ratio was determined with GC Image (ratio of peak value to the estimated peak-to-peak noise).

Figure 9 Column II 1D view of the chromatogram (Avg. Abs. 125-240 nm) for the m-xylene peak: A) without preprocessing; B) with baseline, noise correction and detector blank subtraction.

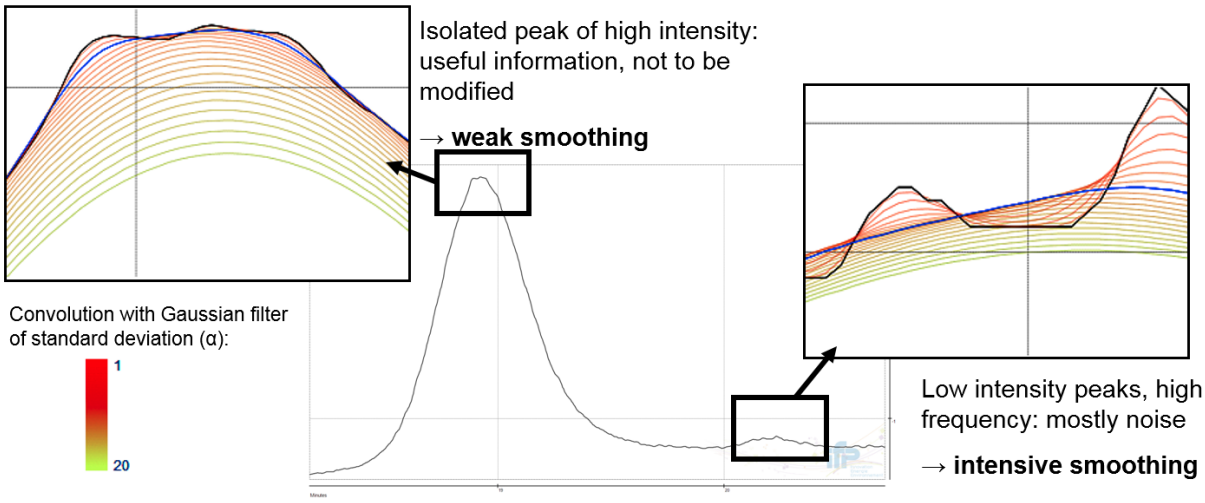

Isolated peak of high intensity: useful information, not to be modified
→ weak smoothing
Convolution with Gaussian filter of standard deviation (α):
1
20
Low intensity peaks, high frequency: mostly noise
→ intensive smoothing

Figure 1

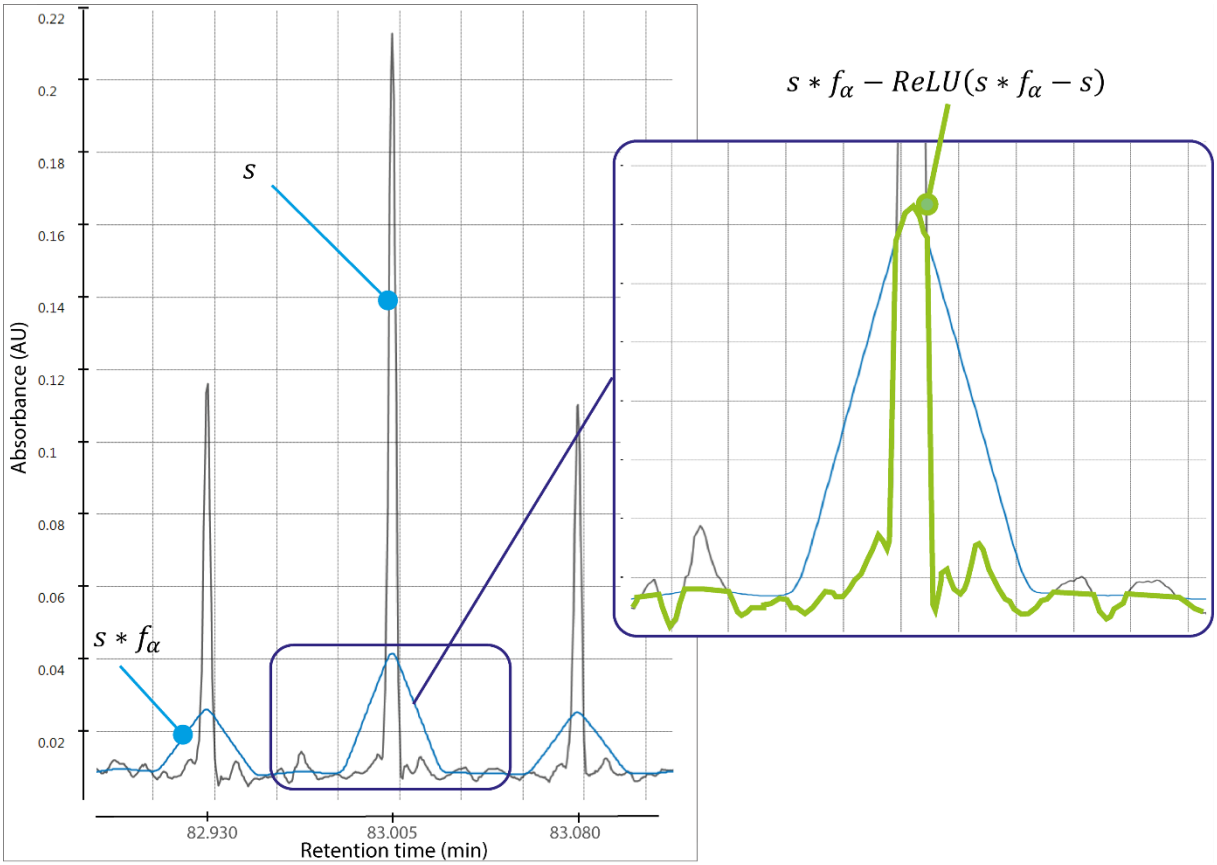
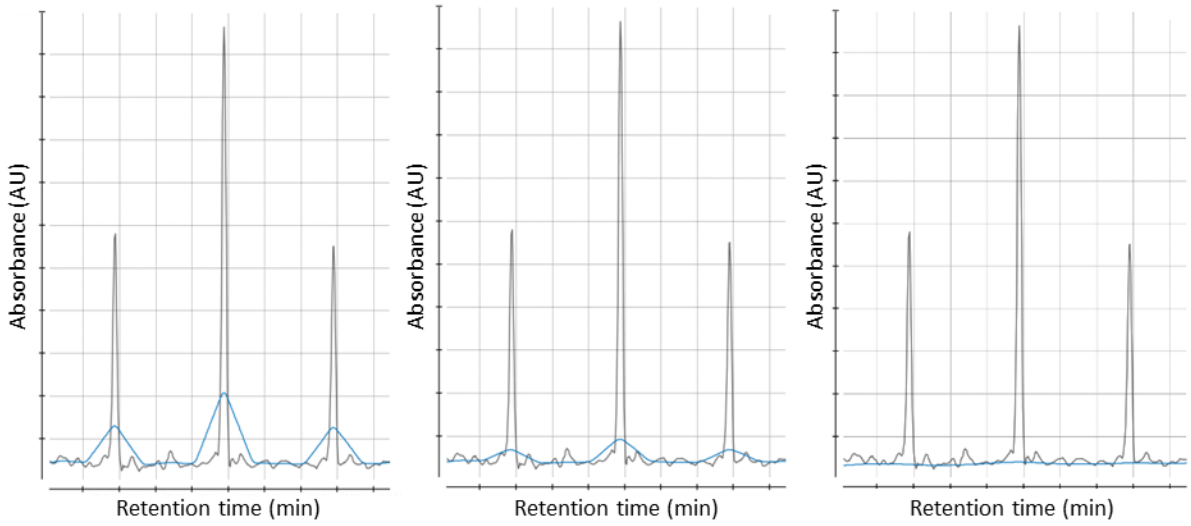

0.22
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
Absorbance (AU)
s
$s * f_\alpha$
$s * f_\alpha - ReLU(s * f_\alpha - s)$
82.930
83.005
83.080
Retention time (min)

Figure 2



$s * f_\alpha$ $(F_\alpha)^2(s) * f_\alpha$ $(F_\alpha)^{10}(s) * f_\alpha$

Figure 3

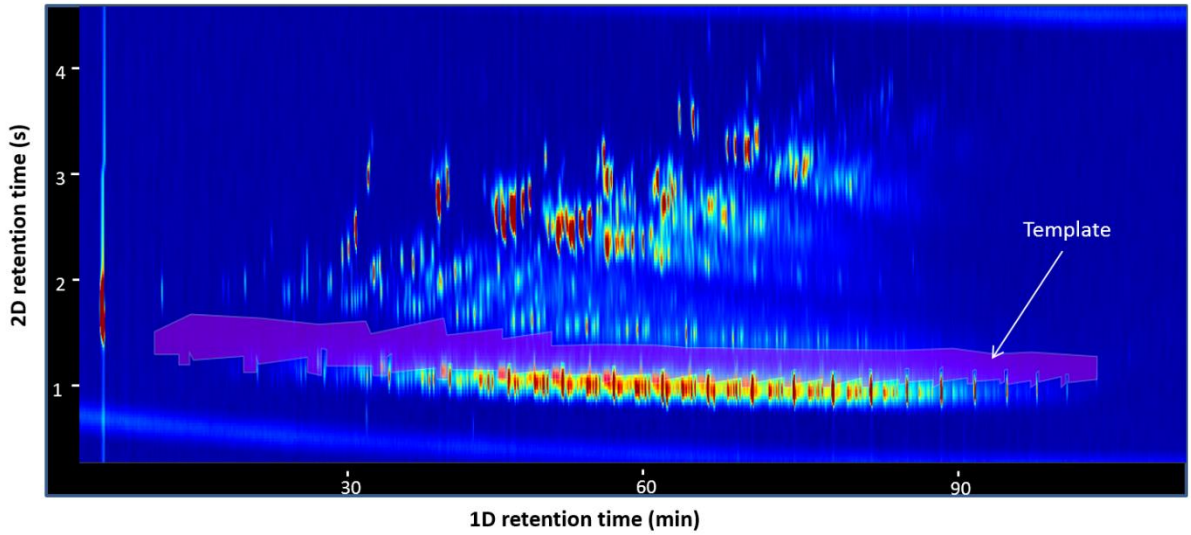

2D retention time (s)
4
3
2
1
30
60
90
1D retention time (min)
Template

Figure 4

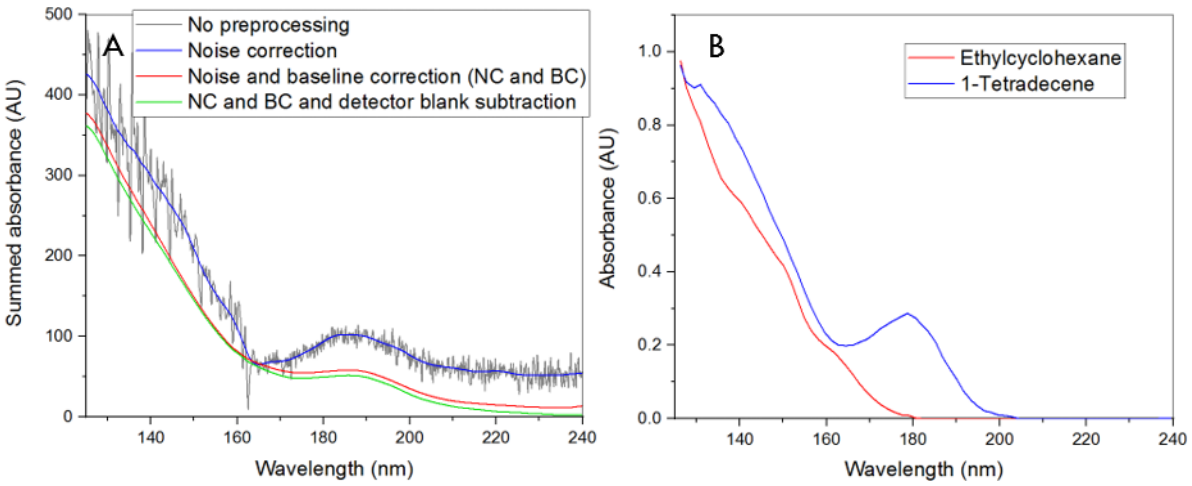

A
No preprocessing
Noise correction
Noise and baseline correction (NC and BC)
NC and BC and detector blank subtraction
Summed absorbance (AU)
Wavelength (nm)
B
Ethylcyclohexane
1-Tetradecene
Absorbance (AU)
Wavelength (nm)

Figure 5

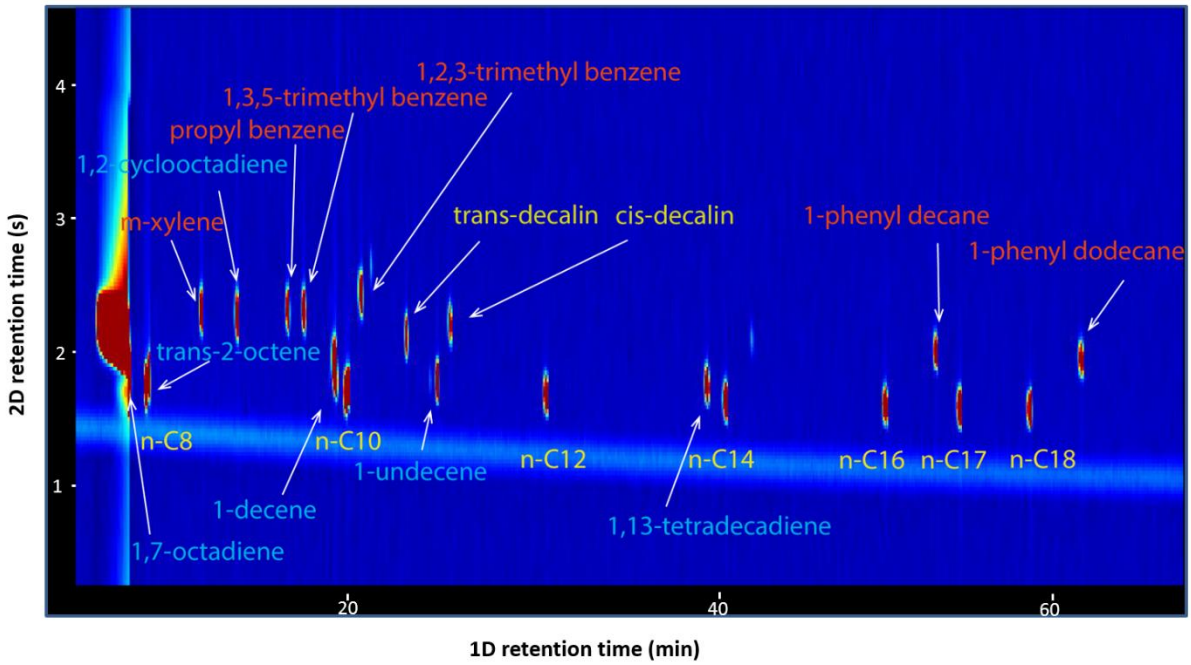

1,2,3-trimethyl benzene
1,3,5-trimethyl benzene
propyl benzene
1,2 cyclooctadiene
m-xylene
trans-decalin
cis-decalin
1-phenyl decane
1-phenyl dodecane
trans-2-octene
n-C8
n-C10
1-undecene
n-C12
n-C14
n-C16
n-C17
n-C18
1-decene
1,7-octadiene
1,13-tetradecadiene
2D retention time (s)
1D retention time (min)
4
3
2
1
20
40
60

Figure 6

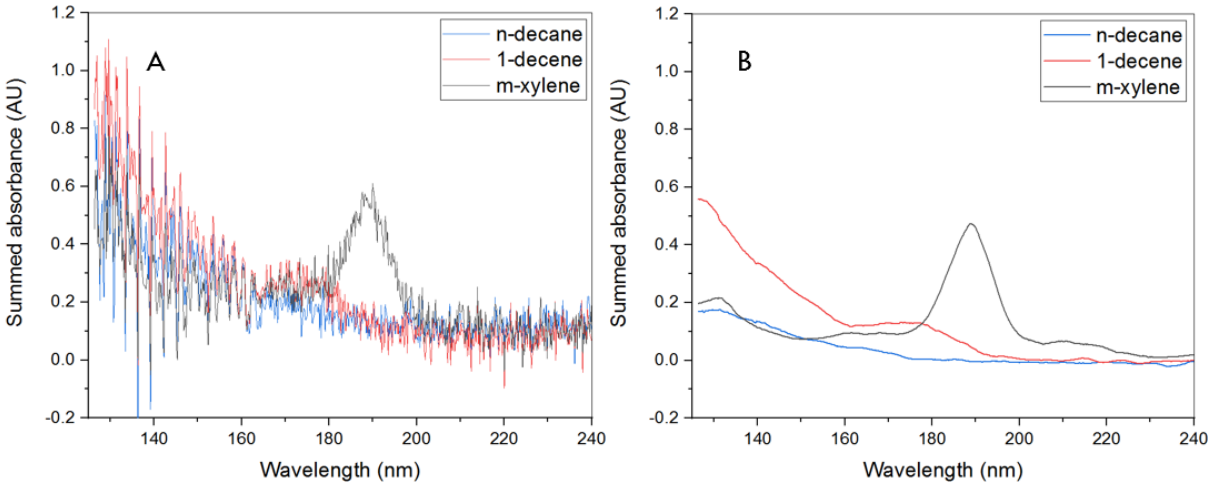

A
B
n-decane
1-decene
m-xylene
Summed absorbance (AU)
Wavelength (nm)
1.2
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
140
160
180
200
220
240

Figure 7

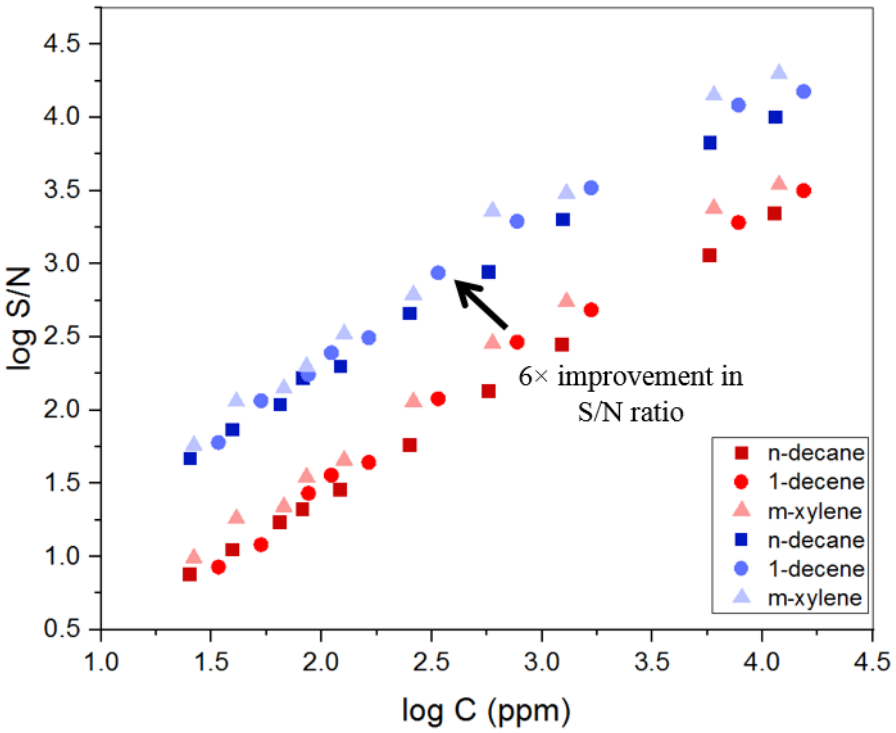

log S/N
log C (ppm)
6× improvement in
S/N ratio
n-decane
1-decene
m-xylene
n-decane
1-decene
m-xylene
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5

Figure 8

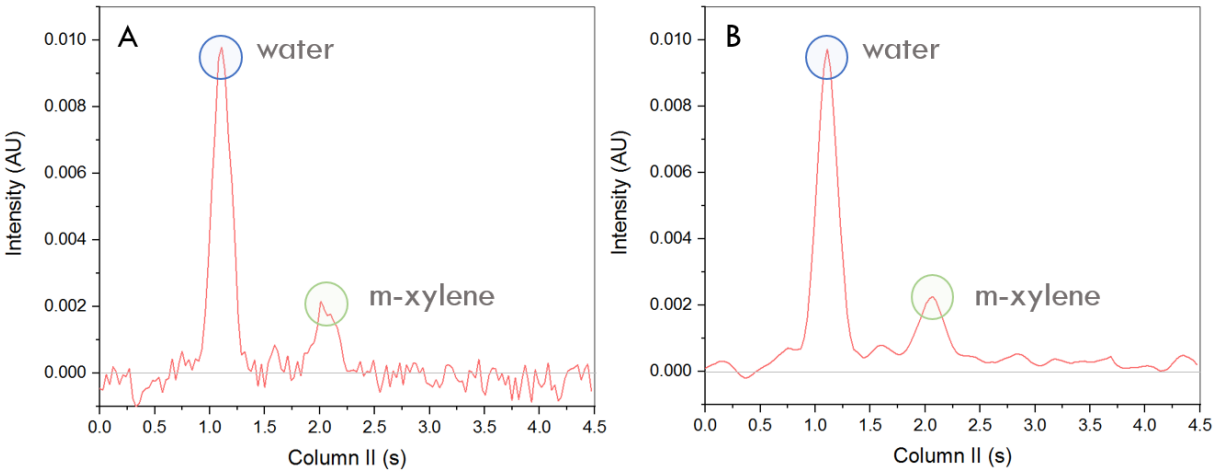

A
water
m-xylene
Intensity (AU)
0.010
0.008
0.006
0.004
0.002
0.000
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
Column II (s)
B
water
m-xylene
Intensity (AU)
0.010
0.008
0.006
0.004
0.002
0.000
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
Column II (s)

Figure 9

SUPPORTING INFORMATION

Advanced data preprocessing for Comprehensive two-dimensional Gas Chromatography with Vacuum Ultraviolet Spectroscopy detection

Aleksandra Lelevic[1,2,*], Vincent Souchon[1],Christophe Geantet[2], Chantal Lorentz[2], Maxime Moreaud[1,*]

1. IFP Energies nouvelles, Rond-point de l'échangeur de Solaize BP 3 69360 Solaize France
2. Univ Lyon, Université Claude Bernard Lyon 1, CNRS, IRCELYON, F-69626, Villeurbanne, France

* Authors for correspondence: aleksandra.lelevic@ifpen.fr, maxime.moreaud@ifpen.fr

Table S1 Composition of standard mixture in toluene for investigation of the limits of detection of the GC×GC/VUV method.

Compound	m/m% in solution
n-C8	1.07
n-C10	1.14
n-C12	0.96
n-C14	1.07
n-C16	1.07
n-C17	1.26
n-C18	1.03
1,13-tetradecadiene	1.07
1,2-cyclooctadiene	1.05
trans-2-octene	1.06
1,7 octadiene	1.00
cis-decalin	0.59
trans-decalin	0.70
1-decene	1.54
1-undecene	0.97
1,3,5-trimethylbenzene	1.06
1,2,3-trimethylbenzene	1.36
1-phenyldecane	0.84
1-phenyldodecane	0.89
m-xylene	1.19
Propylbenzene	1.09

Table S2 Peak volume and signal to noise ratio for selected peaks, for different dilutions of the test mixture with baseline correction in GC Image. GC Image was used for integration of Avg. 125-240 nm chromatogram.

Compound	Conc (ppm)	Blob volume	S/N	Conc (ppm)	Blob volume	S/N	Conc (ppm)	Blob volume	S/N	Conc (ppm)	Blob volume	S/N	Conc (ppm)	Blob volume	S/N
n-C10	11400	11.54	2214.63	5770	6.20	1147.38	1238	1.41	280.33	571	0.65	135.85	251	0.33	57.46
n-C12	9600	4.59	1122.90	4859	2.46	680.08	1042	0.56	235.26	481	0.28	112.96	211	0.13	46.00
n-C14	10700	5.01	1484.44	5416	2.65	917.50	1162	0.63	280.78	536	0.29	123.66	235	0.13	50.59
n-C16	10700	4.90	1050.22	5416	2.64	847.81	1162	0.59	294.33	536	0.30	111.33	235	0.15	56.85
n-C17	12600	5.85	1658.04	6378	3.16	969.53	1368	0.69	350.43	631	0.34	162.32	277	0.17	87.47
n-C18	10300	4.85	1122.15	5213	2.65	968.00	1118	0.59	254.27	516	0.29	162.41	226	0.13	52.61
1-decene	15400	16.77	3167.40	7795	9.17	1914.10	1672	1.98	484.52	772	0.94	292.33	338	0.46	119.76
1-undecene	9700	6.07	1783.45	4910	3.33	1058.84	1053	0.72	229.03	486	0.34	137.06	213	0.17	59.26
m-xylene	11900	15.63	3478.45	6023	8.62	2395.94	1292	1.87	551.68	596	0.85	287.17	261	0.39	114.49
Propylbenzene	10900	13.18	2572.56	5517	7.20	1887.33	1183	1.59	558.55	546	0.73	282.18	240	0.33	121.60
1,3,5-trimethylbenzene	10600	13.90	3075.52	5365	7.58	2023.89	1151	1.64	607.21	531	0.78	260.15	233	0.37	110.20
1,2,3-trimethylbenzene	13600	14.92	3066.36	6884	8.25	2206.04	1477	1.77	591.59	682	0.85	259.77	299	0.38	112.28
trans-decalin	7000	2.89	773.79	3543	1.58	496.78	760	0.35	100.73	351	0.16	53.80	154	0.09	26.68
cis-decalin	5900	2.46	642.82	2986	1.35	397.74	641	0.32	78.37	296	0.14	47.94	130	0.08	21.73
1-phenyldecane	8400	6.84	2040.22	4252	3.77	1245.21	912	0.81	357.23	421	0.39	200.63	185	0.19	73.10
1-phenyldodecane	8900	6.99	1821.50	4505	3.77	1415.59	966	0.83	365.46	446	0.39	155.85	196	0.19	69.32
1,2-cyclooctadiene	10500	9.38	2171.68	5315	5.00	1304.07	1140	1.08	256.46	526	0.51	152.18	231	0.28	65.33
1,13-tetradecadiene	10700	7.56	2311.23	5416	4.10	1687.16	1162	0.89	399.65	536	0.42	176.76	235	0.22	79.22
Compound	Conc (ppm)	Blob volume	S/N	Conc (ppm)	Blob volume	S/N	Conc (ppm)	Blob volume	S/N	Conc (ppm)	Blob volume	S/N	Conc (ppm)	Blob volume	S/N
n-C10	121.4	0.17	28.78	81.98	0.13	21.17	64.64	0.11	17.16	39.37	0.07	11.16	25.25	0.06	7.27
n-C12	102.2	0.05	19.93	69.04	0.07	14.62	54.43	0.03	6.92	33.16	0.02	6.26	21.26	0.02	5.87
n-C14	113.9	0.08	24.08	76.95	0.04	14.49	60.67	0.04	9.94	36.95	0.04	8.29	23.70	0.02	5.68
n-C16	113.9	0.07	28.02	76.95	0.05	19.18	60.67	0.05	14.95	36.95	0.02	8.14	23.70	0.02	6.28
n-C17	134.2	0.10	40.27	90.61	0.06	28.39	71.44	0.06	17.77	43.52	0.04	13.66	27.91	0.03	6.59
n-C18	109.7	0.07	23.50	74.07	0.06	20.37	58.40	0.04	12.07	35.57	0.03	10.96	22.81	0.04	6.55
1-decene	164.0	0.22	44.03	110.75	0.15	36.04	87.32	0.12	27.06	53.19	0.07	12.08	34.11	0.06	8.50
1-undecene	103.3	0.06	20.10	69.76	0.05	12.08	55.00	0.03	9.23	33.50	0.03	5.66	21.49	0.00	
m-xylene	126.7	0.21	45.71	85.58	0.17	34.87	67.47	0.10	22.00	41.10	0.08	18.19	26.36	0.04	9.80
Propylbenzene	116.1	0.16	50.17	78.39	0.11	31.71	61.80	0.09	23.80	37.65	0.06	12.06	24.14	0.02	7.71
1,3,5-trimethylbenzene	112.9	0.19	43.29	76.23	0.12	30.32	60.10	0.10	24.51	36.61	0.06	15.55	23.48	0.03	7.19
1,2,3-trimethylbenzene	144.8	0.17	52.74	97.80	0.12	35.56	77.11	0.08	24.44	46.97	0.07	16.66	30.12	0.03	7.33
trans-decalin	74.5	0.04	10.15	50.34	0.02	8.57	39.69	0.03	7.80	24.18	0.02	4.61	15.51	0.00	6.04
cis-decalin	62.8	0.03	11.94	42.43	0.02	8.83	33.45	0.02	6.46	20.38	0.01	4.39	13.07	0.00	5.29
1-phenyldecane	89.4	0.09	34.62	60.41	0.07	23.32	47.63	0.04	16.89	29.01	0.03	9.85	18.61	0.02	7.73
1-phenyldodecane	94.8	0.09	33.19	64.00	0.06	19.77	50.46	0.04	12.27	30.74	0.03	7.81	19.71	0.03	6.23
1,2-cyclooctadiene	111.8	0.11	30.37	75.51	0.08	20.03	59.54	0.07	12.19	36.26	0.03	7.87	23.26	0.03	7.31
1,13-tetradecadiene	113.9	0.10	37.28	76.95	0.08	27.16	60.67	0.05	20.10	36.95	0.04	11.07	23.70	0.03	7.79

Table S3 Peak volume and signal to noise ratio for selected peaks, for different dilutions of the test mixture with preprocessing in plug im! software (noise, baseline correction and detector blank subtraction). GC Image was used for integration of Avg. 125-240 nm chromatogram.

Compound	Conc (ppm)*	Blob volume	S/N	Conc (ppm)	Blob volume	S/N	Conc (ppm)	Blob volume	S/N	Conc (ppm)	Blob volume	S/N	Conc (ppm)	Blob volume	S/N
n-C10	11400	12.32	10146.80	5770	6.64	6750.49	1238	1.52	2032.03	571	0.72	884.71	251	0.44	462.64
n-C12	9600	4.88	6937.21	4859	2.70	4574.22	1042	0.63	1419.38	481	0.33	702.99	211	0.17	332.55
n-C14	10700	5.48	7550.75	5416	2.82	5527.42	1162	0.68	1341.73	536	0.34	761.13	235	0.16	297.42
n-C16	10700	5.07	6177.10	5416	2.79	4789.25	1162	0.64	1899.53	536	0.35	689.09	235	0.18	370.86
n-C17	12600	6.06	7815.90	6378	3.25	4361.52	1368	0.74	1698.04	631	0.36	613.32	277	0.21	465.15
n-C18	10300	4.96	7377.31	5213	2.73	5190.62	1118	0.66	1416.61	516	0.34	624.20	226	0.16	316.81
1-decene	15400	7.25	11995.34	7795	9.44	12148.47	1672	2.11	3302.40	772	1.06	1951.60	338	0.55	867.24
1-undecene	9700	7.22	11506.68	4910	3.55	6845.71	1053	0.81	1231.01	486	0.38	879.22	213	0.22	346.71
m-xylene	11900	7.99	12667.19	6023	9.10	14171.27	1292	1.97	3041.43	596	0.94	2299.64	261	0.45	616.58
Propylbenzene	10900	6.44	11269.94	5517	7.71	12197.06	1183	1.70	3101.11	546	0.82	1741.96	240	0.39	722.42
1,3,5-trimethylbenzene	10600	17.44	15028.42	5365	8.84	12930.31	1151	1.90	4016.70	531	0.97	1661.07	233	0.44	933.01
1,2,3-trimethylbenzene	13600	3.10	6467.49	6884	7.94	12848.24	1477	1.81	3441.14	682	0.89	1655.29	299	0.43	682.90
trans-decalin	7000	2.62	3789.06	3543	1.70	3138.50	760	0.40	550.18	351	0.21	366.57	154	0.14	166.04
cis-decalin	5900	15.82	20950.58	2986	1.43	2819.71	641	0.36	418.80	296	0.19	281.52	130	0.10	131.11
1-phenyldecane	8400	14.38	15829.45	4252	4.03	6629.14	912	0.91	1935.31	421	0.44	1018.36	185	0.24	469.11
1-phenyldodecane	8900	14.10	14534.28	4505	3.94	7912.15	966	0.92	2479.57	446	0.43	801.70	196	0.23	425.39
1,2-cyclooctadiene	10500	10.05	12620.19	5315	5.25	8248.81	1140	1.14	1429.86	526	0.59	1080.24	231	0.34	405.50
1,13-tetradecadiene	10700	16.47	19964.52	5416	4.49	10182.49	1162	0.97	2176.97	536	0.48	1234.71	235	0.27	486.49
Compound	Conc (ppm)	Blob volume	S/N	Conc (ppm)	Blob volume	S/N	Conc (ppm)	Blob volume	S/N	Conc (ppm)	Blob volume	S/N	Conc (ppm)	Blob volume	S/N
n-C10	121.4	0.21	200.64	81.98	0.16	167.22	64.64	0.20	110.00	39.37	0.07	73.87	25.25	0.00	
n-C12	102.2	0.07	115.54	69.04	0.08	74.94	54.43	0.06	46.51	33.16	0.03	37.47	21.26	0.05	27.19
n-C14	113.9	0.13	164.23	76.95	0.06	74.78	60.67	0.07	54.58	36.95	0.06	46.37	23.70	0.08	29.71
n-C16	113.9	0.11	174.38	76.95	0.09	113.75	60.67	0.10	80.15	36.95	0.05	49.65	23.70	0.05	25.59
n-C17	134.2	0.13	196.66	90.61	0.10	150.95	71.44	0.08	93.77	43.52	0.06	70.81	27.91	0.02	20.49
n-C18	109.7	0.09	131.06	74.07	0.08	147.08	58.40	0.06	74.34	35.57	0.07	57.09	22.81	0.02	26.58
1-decene	164.0	0.26	313.12	110.75	0.19	246.48	87.32	0.17	175.41	53.19	0.11	116.02	34.11	0.10	60.21
1-undecene	103.3	0.10	129.24	69.76	0.07	71.78	55.00	0.05	52.48	33.50	0.06	44.92	21.49	0.04	28.24
m-xylene	126.7	0.27	333.19	85.58	0.22	199.27	67.47	0.14	140.90	41.10	0.11	114.86	26.36	0.07	69.40
Propylbenzene	116.1	0.22	341.12	78.39	0.16	215.14	61.80	0.12	142.46	37.65	0.08	73.42	24.14	0.05	42.85
1,3,5-trimethylbenzene	112.9	0.23	298.26	76.23	0.14	215.10	60.10	0.12	161.49	36.61	0.13	95.32	23.48	0.02	16.39
1,2,3-trimethylbenzene	144.8	0.27	279.86	97.80	0.19	210.57	77.11	0.14	148.90	46.97	0.12	93.17	30.12	0.07	33.97
trans-decalin	74.5	0.08	82.85	50.34	0.05	52.65	39.69	0.05	47.40	24.18	0.03	30.74	15.51	0.02	20.77
cis-decalin	62.8	0.07	80.38	42.43	0.05	58.05	33.45	0.07	32.99	20.38	0.03	25.17	13.07	0.04	19.82
1-phenyldecane	89.4	0.14	207.39	60.41	0.10	133.76	47.63	0.10	84.03	29.01	0.05	51.30	18.61	0.06	24.29
1-phenyldodecane	94.8	0.13	164.04	64.00	0.08	121.18	50.46	0.09	70.28	30.74	0.06	46.84	19.71	0.06	39.40
1,2-cyclooctadiene	111.8	0.16	207.64	75.51	0.14	119.11	59.54	0.10	85.26	36.26	0.10	47.78	23.26	0.05	35.51
1,13-tetradecadiene	113.9	0.15	223.72	76.95	0.11	153.79	60.67	0.08	112.60	36.95	0.09	70.54	23.70	0.08	33.75

* At this level detector saturation could be present.

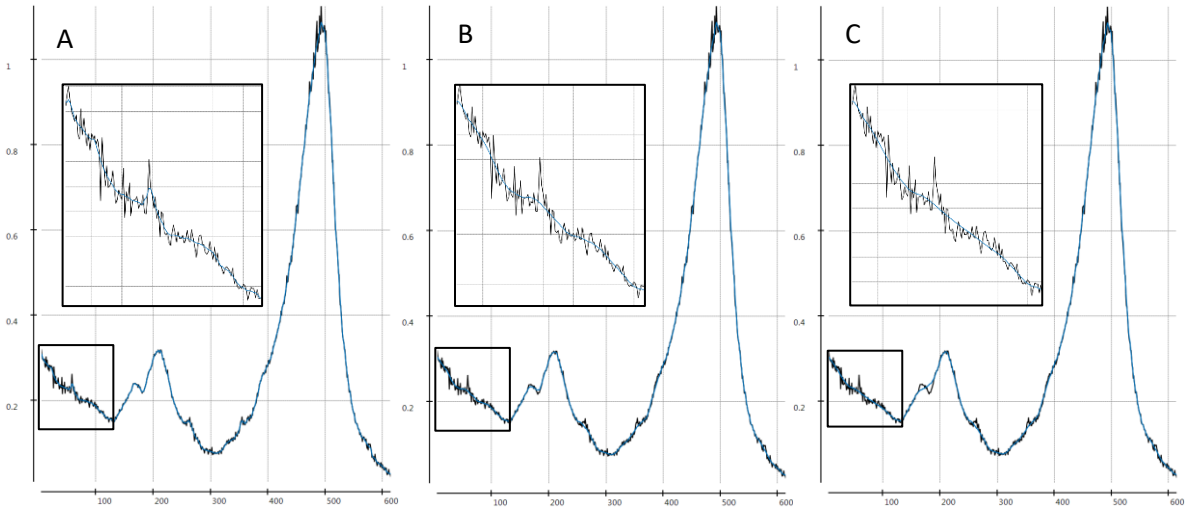


Figure S1 Spectrum at a chosen chromatogram point (in grey) smoothed by choosing following filter M values (in blue) (A) 5, (B) 15, (C) 20.

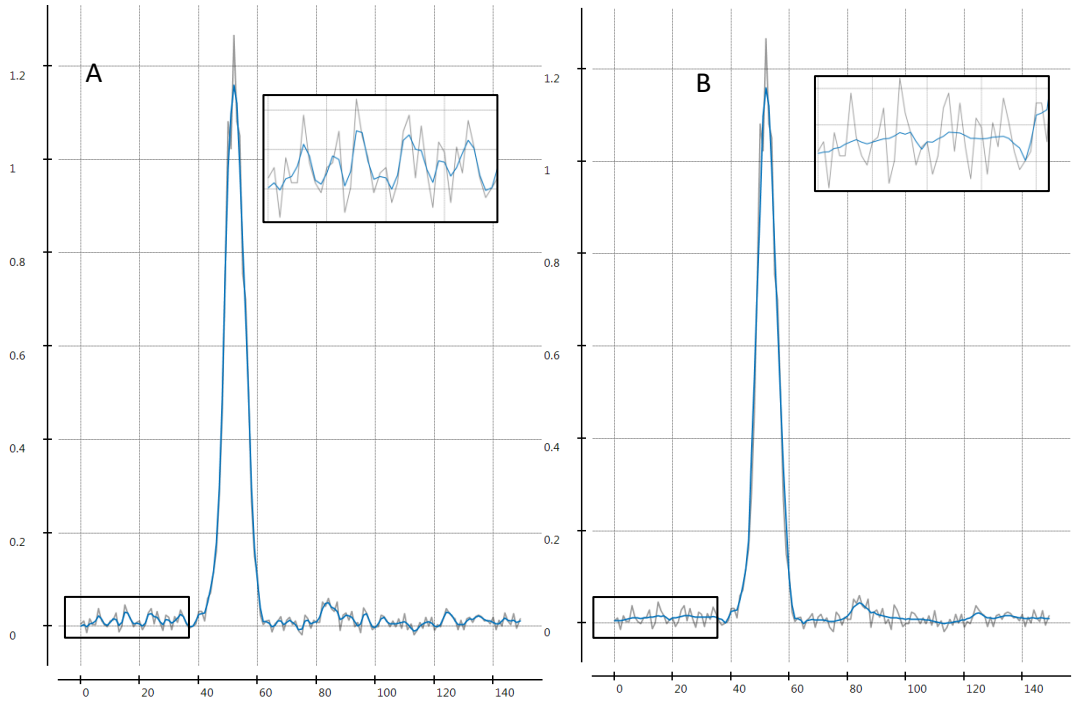


Figure S2 Illustration of the noise correction: input signal in grey and resultant signal in blue. Noise filter M value (A) 2, (B) 5. One modulation period 125 nm chromatogram is shown.

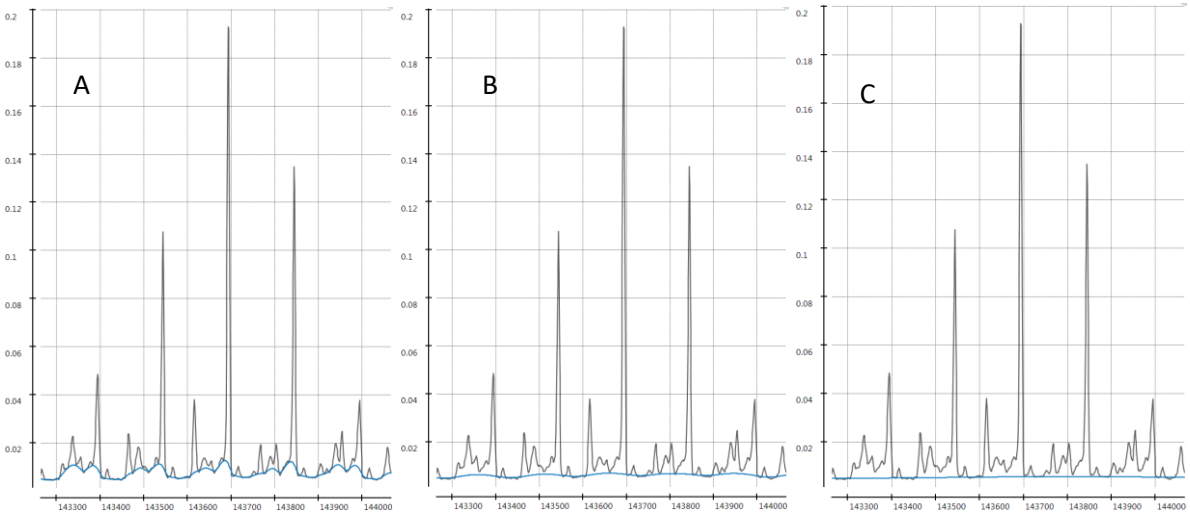


Figure S3 Estimated baseline (in blue) for a VUV signal (in grey) depending on the chosen convolution kernel sizes; (A) 5, (B) 20, (C) 100, number of iterations is 15 (125 nm).

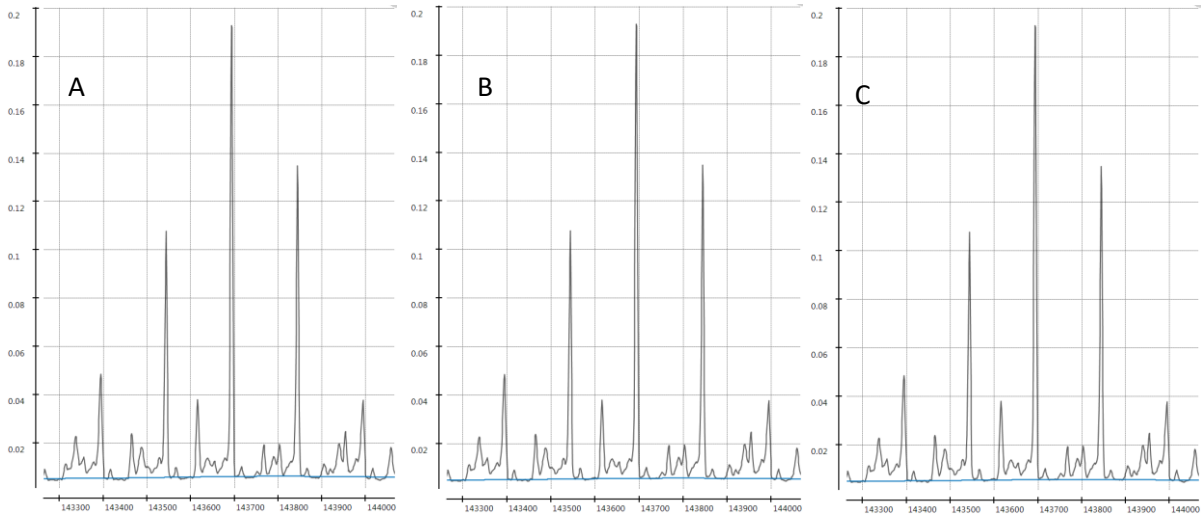


Figure S4 Estimated baseline (in blue) for a VUV signal (in grey) depending on the chosen number iterations; (A) 5, (B) 10, (C) 15, convolution kernel size is 100 (125 nm).

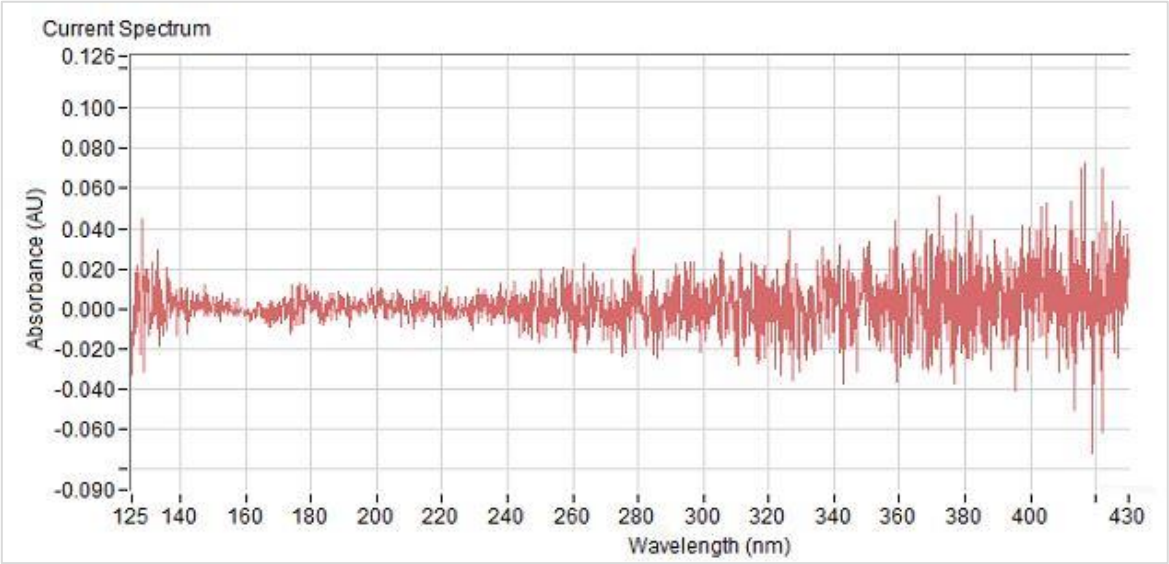


Figure S5 Measured detector blank spectrum.

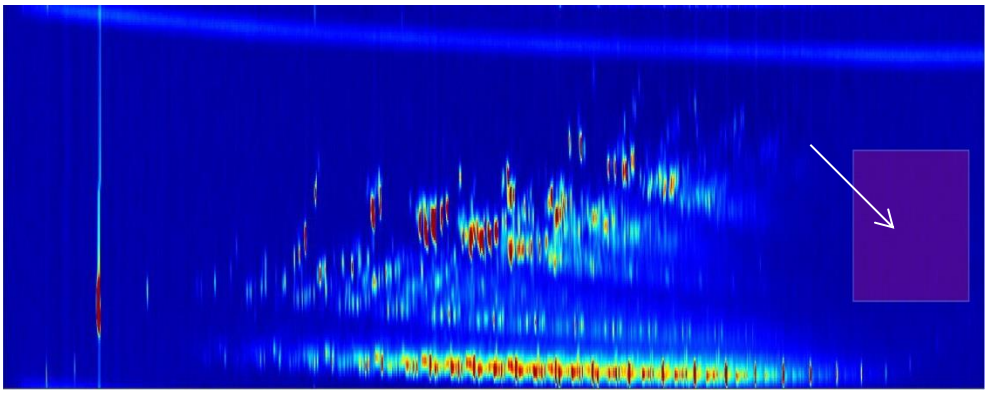


Figure S6 Selection of a zone for extracting detector blank signal.

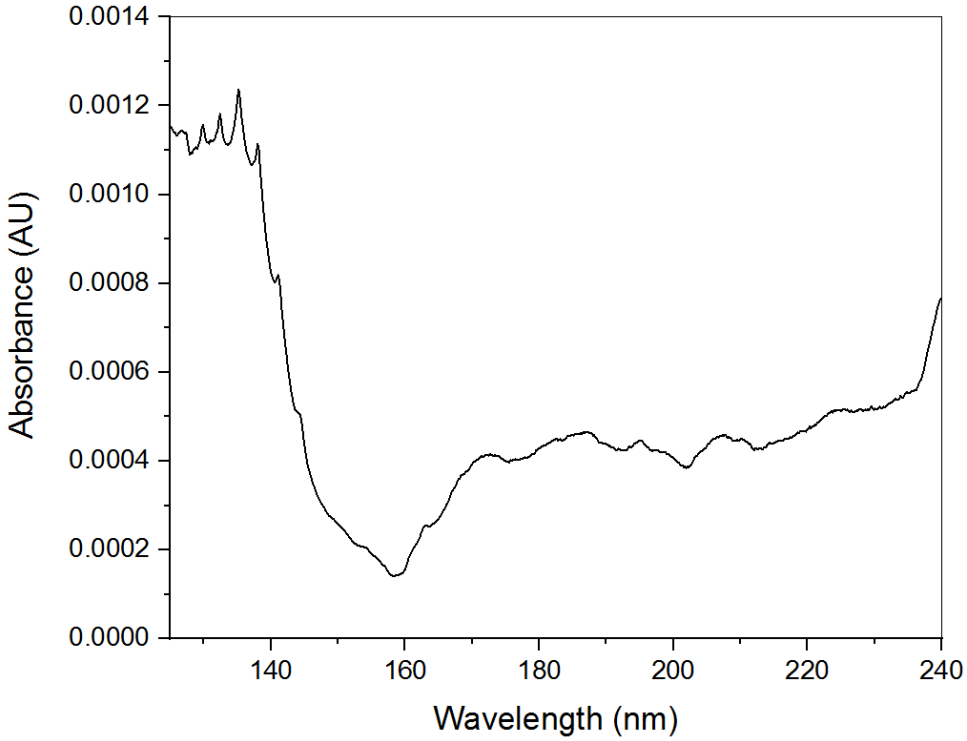


Figure S7 Detector blank signal.

Table S4 Estimation of spectral similarity with corresponding VUV library spectrum - test mixture 150 x dilution.

Compound	No preprocessing		With preprocessing	
	Chi²	**R²**	**Chi²**	**R²**
n-decane	0.008170	0.48	0.001530	0.96
1-decene	0.002730	0.87	0.000836	0.97
m-xylene	0.010200	0.61	0.000474	0.98

Table S5 Estimation of spectral similarity with corresponding VUV library spectrum - test mixture 500 x dilution.

Compound	No preprocessing		With preprocessing	
	Chi²	**R²**	**Chi²**	**R²**
m-xylene	0.007580	0.25	0.001640	0.94