



SERS mapping combined with chemometrics, for accurate quantification of methotrexate from patient samples

He, Peihuan; Dumont, Elodie; Göksel, Yaman; Slipets, Roman; Schmiegelow, Kjeld; Chen, Quansheng; Zor, Kinga; Boisen, Anja

Published in:
Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy

Link to article, DOI:
[10.1016/j.saa.2023.123536](https://doi.org/10.1016/j.saa.2023.123536)

Publication date:
2024

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
He, P., Dumont, E., Göksel, Y., Slipets, R., Schmiegelow, K., Chen, Q., Zor, K., & Boisen, A. (2024). SERS mapping combined with chemometrics, for accurate quantification of methotrexate from patient samples. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, 305, Article 123536. <https://doi.org/10.1016/j.saa.2023.123536>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



Contents lists available at ScienceDirect

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy

journal homepage: www.journals.elsevier.com/spectrochimica-acta-part-a-molecular-and-biomolecular-spectroscopy

SERS mapping combined with chemometrics, for accurate quantification of methotrexate from patient samples

Peihuan He^{a,d,f,g,1}, Elodie Dumont^{a,b,1,*}, Yaman Göksel^{a,b}, Roman Slipets^{a,b}, Kjeld Schmiegelow^c, Quansheng Chen^{e,f,*}, Kinga Zor^{a,b}, Anja Boisen^{a,b}

^a Center for Intelligent Drug Delivery and Sensing Using Microcontainers and Nanomechanics (IDUN), Department of Health Technology, Technical University of Denmark, Kgs. Lyngby, 2800, Denmark

^b BioInnovation Institute Foundation, Copenhagen N, 2200, Denmark

^c Department of Paediatrics and Adolescent Medicine, Rigshospitalet University Hospital, Copenhagen 2100, Denmark

^d School of Grain Science and Technology, Jiangsu University of Science and Technology, Zhenjiang 212100, PR China

^e College of Ocean Food and Biological Engineering, Jimei University, Xiamen 361021, PR China

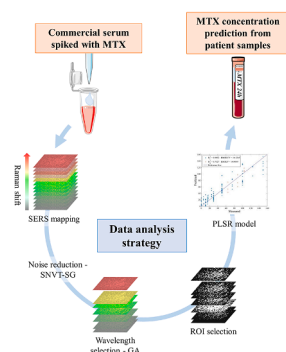
^f School of Food and Biological Engineering, Jiangsu University, Zhenjiang 212013, PR China

^g School of Materials Science and Engineering, Jiangsu University of Science and Technology, Zhenjiang 212100, PR China

HIGHLIGHTS

- Multivariate data analysis of SERS maps enabled MTX prediction in patient samples.
- More accurate quantification with multivariate vs. univariate data analysis.
- The model can be built directly in commercial serum and applied to patient samples.
- The genetic algorithm identified specific wavenumbers from complex SERS spectra.
- Image threshold segmentation was applied to select relevant map pixels.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:

SERS mapping
Genetic algorithm (GA)
Image threshold segmentation
Methotrexate (MTX)
Patient samples
Multivariate quantitative data analysis

ABSTRACT

Despite the technological development in Raman instrumentation that has democratized access to 2D sample scanning capabilities, most quantitative surface-enhanced Raman scattering (SERS) analyses are still performed by only acquiring a single or a few spectra per sample and performing univariate data analysis on those. This strategy can however reach its limit when analytes need to be detected and quantified in complex matrices. In that case, surface fouling and competition between the target analyte and interfering compounds can impair

* Corresponding authors at: Technical University of Denmark, Kgs. Lyngby, 2800, Denmark (E. Dumont).

E-mail addresses: elodu@dtu.dk (E. Dumont), qschen@ujs.edu.cn (Q. Chen).

¹ These authors contributed equally to this work.

<https://doi.org/10.1016/j.saa.2023.123536>

Received 17 May 2023; Received in revised form 28 August 2023; Accepted 12 October 2023

Available online 14 October 2023

1386-1425/© 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

univariate SERS data analysis, underlining the need for a more in-depth data analysis strategy based on exploiting of full-spectrum information.

In this paper, a multivariate data analysis strategy was developed, for analyzing SERS maps of methotrexate (MTX) from patient samples, including all steps from baseline correction, selection of wavelength, and the relevant pixels in the map (image threshold segmentation), as well as quantitative model construction based on partial-least squares regression.

Among the different baseline correction methods evaluated, standard normal variable transformation and Savitzky-Golay smoothing proved to be more suitable, while the genetic algorithm wavelength screening method was able to screen out MTX-related SERS spectral regions more efficiently. Importantly, with the here-developed process, it was sufficient to use MTX-spiked commercial serum when building quantitative models, removing the need to work with MTX-spiked patient samples, and consequently enabling time- and resource-saving quantitative analyses. Besides, the developed multivariate data analysis approach showed superior performances compared with univariate analysis, with 30 % improved sensitivity (detection limit of 5.7 μM), 25 % higher reproducibility (average relative standard variation of 15.6 %), and 110 % better accuracy (average prediction error of -10.5 %).

1. Introduction

Surface-enhanced Raman scattering (SERS) has been reported to be a powerful analysis tool with applications notably in medicine [1,2], food safety [3,4], and environment monitoring [5]. The main advantage of SERS compared to Raman is that the Raman signal of the target analyte, either bonded or adsorbed onto the surface of SERS-active substrates, is significantly enhanced, therefore enabling sensitive detection and quantification even in complex samples [6].

Both qualitative and quantitative information can be gathered from SERS spectra. A widespread practice in quantitative SERS is to combine colloids with the fast single-point measurement acquisition setting [7–10]. Thanks to the relative homogeneity of nanoparticle suspensions, collecting a single or a few spectra can give an overview of the analyzed sample within seconds and can be conducted even with handheld devices [7,8]. With the view of achieving more reproducible results and improving batch-to-batch variability, the development of solid SERS substrates has expanded, as they could provide more defined hotspot geometries [11].

Compared with single-point analysis, SERS mappings of larger substrate areas can provide more representative, objective, and accurate information about the analyzed samples by accounting for spot-to-spot intensity fluctuations [12–14]. These fluctuations typically occur, especially with solid SERS substrates and at low analyte concentrations, as only a few molecules located in the hotspots contribute to the recorded SERS spectra [15]. The data collected during SERS mappings generally consists of a three-dimensional data cube, with two spatial (length*width, $x*y$) and one spectral (Raman shifts, λ) dimension, consequently rich in information. SERS mappings have been applied in various areas, including the detection of food contaminants [16,17], quantification of bacterial secondary metabolite [18], or detection of drugs in human biological matrices [1,19]. However, in most of these cases, exhaustive exploitation of the hyperspectral data was not performed; on the contrary, solely a mean spectrum was extracted as output from the maps for univariate quantitative analysis, primarily based on peak height or area under the curve analysis of one or few defined spectral features [1,16,18].

While discriminative SERS data are generally handled with chemometrics tools, it is still not a standard rule to apply the latter to quantitative SERS data [20–25]. Moreover, surprisingly, chemometrics algorithms are almost exclusively used on quantitative data collected from single-point analyses [7–10,26–28], with hardly any applications to maps in the literature [18,29,30], to the best of our knowledge.

The most common algorithm applied to SERS data is noise reduction pre-processing. Different kinds of interfering signal and noise can appear in SERS spectra, notably from Raman spectrophotometer, SERS substrate, or interfering compounds in the sample matrix [31]. Therefore, interfering signals in SERS mapping data may increase while expanding the scanned area. Spectral pre-processing such as Savitzky-Golay

smoothing (SG) [7,8,18,26], multiple scattering corrections (MSC) [7–9], standard normal variable transformation (SNVT) [7–9,28], wavelet transform [32], first derivative (FD) [7–9,18,28,29], and second derivative (SD) [7–9,28] are some examples of algorithms, previously used on SERS data, that can be applied to reduce the spectral noise.

Furthermore, applying algorithms to the analysis of spectral data is also essential to improve the quantification performances of SERS, especially considering complex sample matrices [6,15]. SERS mapping data suffer from dimensionality issues where a high number of uninformative, undesired and/or interfering wavelength variables need to be removed. Therefore, selecting the most informative wavelengths is critical in constructing robust quantification models. Additionally, the number of samples is often much smaller than the number of wavelength variables collected, which can lead to overfitting, i.e., great prediction performances in training samples but poor prediction performances on newly introduced samples [15]. Both of the above aspects in data analysis of SERS spectra lead to a significant challenge in building a cost-effective and robust calibration model [33]. Wavelength selection algorithms used for this purpose enable retaining informative variables and removing uninformative or interfering ones, thereby improving the prediction performance of calibration models. Yet, applying these algorithms to quantitative SERS mapping data is still not a common practice, which could be due to the complexity of the required data analysis, which is not automated. However, algorithms such as genetic algorithm (GA) [8,9,28], competitive adaptive reweighted sampling (CARS) [7,8,10], uninformative variable elimination (UVE) [34], successive projection algorithm (SPA) [7], and synergy interval partial least squares (SI) [28] have been used to improve SERS quantitative models. The selection of variables is based on the model's performance, generally using partial least square regression (PLSR) through iterative selection, which yields the subset of variables with the best model performance [7–10,28]. However, variable selection strategies have specific constraints, and the algorithms perform differently according to the data set. Comparison and screening of algorithms performances is hence required for each new type of data set.

Combining SERS mapping with multivariate spectral data analysis would notably benefit the field of SERS-based bioanalysis [35]. Biological matrices, like blood or serum, are complex, with variable composition and patient-to-patient differences. Even in the case of the same patient, the sample matrix composition could be influenced by physiological and pathological factors, such as sampling time, food intake, or the presence of an underlying condition [35]. This can entail significant spectral variations in terms of both peak intensity and peak (dis)appearance, making the reliability of a univariate quantitative model low [35].

Previously, Göksel *et al.* developed a nanopillar-assisted separation (NPAS) SERS-based detection method to quantify methotrexate (MTX) in the serum of leukemic pediatric patients receiving high-dose MTX therapy [1]. Due to the NPAS process, MTX was well localized in a

specific area on the SERS substrates and univariate data analysis, based on a single spectral feature peak intensity, was successfully implemented on the averaged SERS spectra when building calibration curves even in commercial serum samples. However, when working with patient samples, the high intra- and inter-patient serum variability greatly influenced MTX peak intensity and led to spectral feature variations. To mitigate this high variability and enable quantification of MTX in patient samples, the serum of each patient, collected before MTX infusion, was used to build a calibration curve for each patient by spiking it with MTX. Still, despite this fact, the vast information included in the SERS maps was not fully utilized, leading to poor prediction ability, with an average prediction error of 21 %. Although a calibration curve was built for each patient, MTX peak intensity could be impacted by interfering serum species competing with MTX for hotspots. Moreover, matrix components could even produce interfering SERS peaks, decreasing the accuracy of MTX quantification. Additionally, constructing calibration curves for every patient is not a viable strategy in clinical practice. A single quantification model, ideally built using commercial serum, would instead be preferred to predict MTX concentrations from all patient samples.

Accordingly, the current work aimed to develop an accurate, rapid, reproducible, and easy-to-implement data analysis method for quantifying MTX in patient samples by fully exploiting the previously published SERS mapping dataset [1]. To that end, chemometrics algorithms were applied to reduce spectral noise (SNVT, MSC, FD, detrend, and SG), to screen and select meaningful wavelength variables (CARS, GA and SI), and to select the regions of the SERS substrate carrying significant MTX information (image threshold segmentation). As indicated earlier, this in-depth data analysis is one of the few where spectral data collected in maps are used for building an analytical model for quantification.

2. Materials and methods

2.1. Description of the used dataset

All the SERS data used in this work was collected and presented by Göksel *et al.* [1]. More information regarding experimental parameters can be found in Section 1 of the Supporting Information. The data set consisted of 2 quantitative models (i and ii) and 1 type of data used to test the models (iii):

- (i) SERS maps of commercial serum samples spiked with MTX at concentrations of 0, 5, 10, 15, 25, 30, 50, 75, 100, 125 or 150 μM . Each concentration level consisted of between 3 and 16 maps ($n = 3 - 16$).
- (ii) SERS maps of patient's serum samples drawn before administration of MTX, and spiked with MTX at concentrations of 0, 25, 50, 75, 100 or 125 μM . Each concentration level consisted of between 3 and 12 maps ($n = 3 - 12$).
- (iii) SERS maps of patient's serum samples collected 24 h after MTX administration ($n = 3$ maps for each patient), with a known MTX concentration (immunoassay measurement performed in Rigshospitalet, Copenhagen, Denmark). Three maps were acquired for each sample ($n = 3$).

The SERS maps comprised around 43×43 spectra and 1686 wavelengths per spectrum. Each map was collected onto a new piece of SERS substrate from a different aliquot of the same sample (technical replicates, n).

2.2. Baseline correction methods

To reduce the noise in the SERS spectra and correct their baseline, the FD (window size of 17), MSC, detrend, SNVT and SG (window size of 15, derivative order of 2) algorithms were applied as pre-processing.

2.3. Variable selection methods

As the SERS spectra contain numerous wavenumbers that do not include useful information for MTX quantification, variable selection methods were employed to select the most meaningful variables to include in the PLSR model. Both discontinuous wavelength screening (CARS and GA) and continuous wavelength screening (SI) methods were tested.

In the CARS algorithm, the importance of wavelength variables was determined by the absolute value of PLSR model coefficients [36]. CARS algorithm mainly included the following steps: (i) Monte Carlo method for sampling, and random selection of the calibration set samples; (ii) deletion of the wavelength variables with lower contribution rates for the PLSR model (with max 15 principal components) based on an exponentially decreasing function; (iii) adaptive weighted sampling adopting the "survival theory of the fittest" to eliminate variables competitively and retain variables whose weight was greater than the threshold (20); and (iv) comparison of cross-validation root mean square error (RMSECV) values acquired by 5-fold cross validation in subsets after 50 sampling times, and selection of the subset with the smallest RMSECV value.

In GA algorithm, the initial population (30 chromosome samples) of the subset was randomly generated by binary coding, where "1" represents the selection of the corresponding wavelength and "0" means the non-selection [37,38]. The population evolved through the probability of crossover (0.5) and the probability of mutation (0.01) operations. The variables were screened based on the principle of "survival of the fittest", i.e., variables with high predictive ability for the PLSR model were included in the next population. Finally, after 100 iterations, the population converged to the wavelength variables containing necessary information. Genetic algebra was used as the termination condition. When it reached the optimized algebra number, the search process in GA was terminated [38,39].

The SI algorithm divided the spectra evenly into $N = 20$ sub-intervals, removed one sub-interval, and established a PLSR model with the remaining $N-1$ sub-intervals [40]. The first excluded subinterval replaced each subinterval in the PLSR model established by $N-1$ subintervals, and the stability and prediction accuracy of these sub-models were assessed through RMSECV. Among these, the most stable sub-model was used as the first SI-PLS model, and then the above steps were repeated for the remaining $N-1$ sub-interval. Finally, the model with the smallest RMSECV was found from the established sub-models as the SI-PLS model, and this sub-interval was also the optimal wavelength variable subset.

All the above-described variable selection methods (CARS, GA and SI) were combined with PLSR. The samples were divided into training and prediction sets (sample ratio of 2/1). The RMSECV, root mean square error of prediction (RMSEP), and determination coefficient in the training set (R_t^2) and the prediction set (R_p^2) were calculated to assess the performance of the PLSR model. To further test the accuracy of the prediction results of the model, the predicted concentrations of patient samples were compared with their nominal values, and relative standard deviation (RSD) and prediction error (PE) were calculated.

Generally, reliable models have higher R^2 value and lower RMSE value, and the difference between R_t^2 and R_p^2 and between RMSECV and RMSEP should be as small as possible.

2.4. Optimization of the image threshold segmentation

As mentioned earlier, only certain parts of the SERS maps contain MTX-related information. Appropriate threshold segmentation can be applied to the maps to extract spectral information only from the regions where MTX molecules are located.

When MTX is present at a specific location of the SERS substrate, the SERS intensity at the characteristic MTX wavelengths should be relatively higher than the background signal. Therefore, threshold

segmentation can isolate these areas with high SERS intensity and then calculate the average SERS intensity in these areas (Supporting Information). The PLSR models can subsequently be built with these averaged intensity data.

2.5. Data analysis software

All the data pre-processing and chemometrics analysis were performed using MATLAB Version 2014b (Eigenvector Research Inc., Wenatchee, WA, USA).

3. Results and discussion

In our previous paper [1], when using univariate, single spectral feature-based data analysis on the same data set used in this work, a calibration curve had to be built for each patient serum to quantify the MTX concentration in the respective patient sample. This practice is very time- and sample-consuming, thus not suitable for quantitative analysis in real life. On the contrary, the optimal solution would be to construct a single quantitative model, preferably with commercial serum, to accurately predict MTX concentration in patient samples. The model would ideally be able to extract relevant MTX features without being influenced or sensitive to matrix variations and interferences.

3.1. Comparison of different baseline correction methods

When acquiring SERS maps, it is common that the baseline moves up and down or drifts (Figure S1A) [41]. These perturbations will reduce the stability of the model, so it is necessary to pre-process the SERS spectrum before a quantitative model is built. This work compared different baseline correction methods, such as SNVT, MSC, FD, detrend, and SG, to select the most suitable SERS maps.

In the spectra obtained using the MSC pretreatment method, several abnormal spectra were obtained, which did not occur in the raw spectra (Figure S1B). Many spectral features (peaks) obtained after FD pretreatment were negatives (Figure S1C), which made FD pretreatment unsuitable for the subsequent image analysis.

Compared with the raw spectra, the detrend and SNVT methods produced pre-processed spectra with similar baselines (Figure S1D and E, respectively). These methods were thus better suited to SERS maps.

As the next step, PLSR models were built on the data pre-processed with these 4 methods (detrend, SNVT, MSC, FD) to compare them. The R_t^2 , R_p^2 , RMSECV and RMSEP are displayed in Table 1, and it can be seen that, among the models, the one based on SNVT-corrected spectra showed better predictions.

Although SNVT could correct the baseline effectively, noise can clearly be observed in the pre-processed spectra (Figure S1E). An additional SG smoothing processing was therefore employed to reduce the noise in the SERS maps (Figure S1F), which resulted in improved model performances (Table 1, group 6); therefore this SNVT-SG processing was used for further data analysis.

Table 1

PLSR model results after spectra pre-processing with different baseline correction methods. The PLSR models were constructed using 128 spiked commercial serum samples, based on the full spectrum and the whole map pixels, and used to predict the same samples.

Group	Baseline correction method	R_t^2	R_p^2	RMSECV	RMSEP	PC
				(μM)	(μM)	
1	None	0.69	0.70	23.61	24.32	12
2	MSC	0.80	0.50	18.52	43.63	10
3	FD	0.69	0.77	23.48	21.40	11
4	Detrend	0.68	0.73	23.82	22.90	11
5	SNVT	0.78	0.84	19.03	17.54	12
6	SNVT-SG	0.79	0.84	18.67	17.29	12

To further verify the effectiveness of the SNVT-SG pretreatment, it was applied before univariate data analysis using MTX characteristic peak (679 cm^{-1}) and compared to non-baseline corrected data [1] (Table S1). As can be observed in Table S1, univariate models based on 2 points baseline-corrected spectra used in our previous work also produced better prediction results than non-baseline corrected data [1].

3.2. Comparison of variable selection methods

While the SERS spectrum of MTX has several characteristic peaks (as indicated by pink arrows in Fig. 1, Figure S2 and locations specified in Table S2), in our previous work, we only focused on a single peak to perform univariate quantitative analysis [1]. In this work, we aimed to improve the prediction accuracy by considering several MTX characteristic peaks in the quantification model. As can be observed in Fig. 1, only a few wavenumbers contain relevant information assigned to MTX. Furthermore, despite implementing sample pretreatment (protein precipitation and NPAS) prior detection, interferences originating from the complex serum matrix could remain in the pretreated sample and compete with MTX for the substrate surface. This matrix effect is further accentuated in the case of patient samples, where, due to the disease and used therapy, the serum composition can broadly vary, and the presence in serum of co-administered drugs is expected together with the target analyte.

In summary, it would be highly beneficial not to rely only on a single spectral feature peak from the SERS spectra, for data analysis. Selecting the characteristic analyte peaks manually is time-consuming and it can be subjective. An automatic wavelength screening method for finding these distinct peaks would be much more efficient.

As described in the Material and Methods section, 3 different variable screening algorithms, namely GA, CARS, and SI, were investigated in this work to select the most relevant wavelengths. PLSR models were then constructed using 128 spiked commercial serum samples based on these chosen variables. The variable screening algorithms were compared and their suitability was assessed by looking at the prediction accuracy of MTX from patient samples (Table 2).

GA is a method based on the theory of natural selection and biological evolution [42]. Each spectrum with wavelength variables in GA

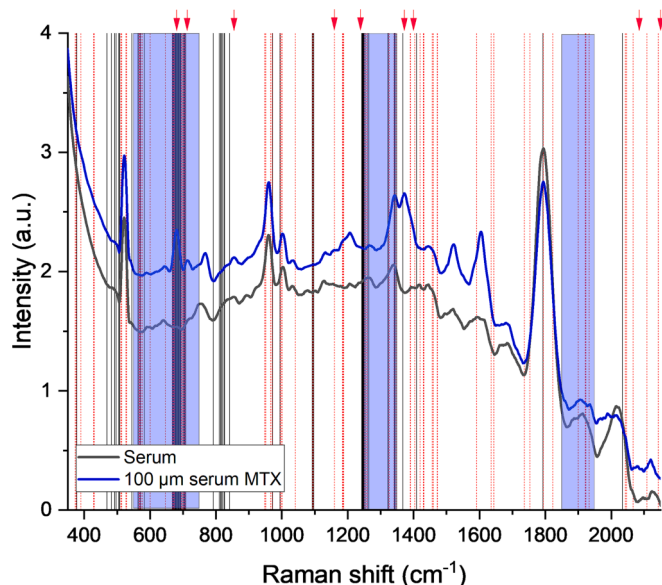


Fig. 1. SERS spectra of MTX in spiked commercial serum after SNVT-SG pre-processing. The characteristic SERS peaks of MTX are highlighted by pink arrows. The plain black and dashed red vertical lines display the wavelengths selected by the GA and CARS algorithms, respectively. The highlighted purple areas are the spectral regions selected by the SI algorithm.

Table 2

Comparison of different variable selection algorithms in terms of number of selected wavenumbers.

Variable selection methods	Number of selected variables	Informative variables	Uninformative variables (%)	Informative variable included (%)
GA	63	27	36 (57 %)	17
CARS	76	22	54 (71 %)	14
SI	208	60	148 (71 %)	38
None	935	157	778 (83 %)	100

it is regarded as a biological individual. The spectral group with the highest explained variance will be selected as the final selected wavelength variable [39]. CARS can also play a powerful and influential role in determining wavelength variables. CARS does not only reduce the influence of collinear variables on the model, but also further avoids over-fitting in the modeling process. It hence improves the prediction of the model [36]. GA and CARS algorithms screen out discontinuous variables. Nørgaard *et al.* introduced SI as a method for selecting continuous variables [40]. The accurate model is obtained by dividing the spectrum into smaller equidistant subintervals, then comparing the preferred combined spectral regions.

As shown in Table 2 and Fig. 1, 63, 76 and 208 wavenumbers were selected from the whole SERS spectrum by GA, CARS and SI, respectively. Among these algorithms, GA selected the lowest proportion of uninformative variables (Table S2). The proportion of informative variables included in the model by GA were between CARS and SI, while more MTX characteristic peaks were selected by GA than by CARS and SI (Table S2). Therefore, among these 3 methods, GA was the most efficient for spectrum screening and for selecting MTX characteristic peaks.

Table 3 and Figure S3 show the MTX prediction results from 3 patient samples ($n = 3$) based on PLSR models constructed with 128 spiked commercial serum samples, comparing the 3 different variable selection methods. When predicting MTX concentration from patient samples, the RMSECV of the 3 PLSR models were quite close. However, the RMSEP of the PLSR model based on GA-selected wavenumbers was lower than the ones based on CARS and SI-selected wavenumbers. Based on these results, we concluded that the PLSR model constructed using variables screened by GA predicts better MTX concentration than models built using CARS and SI-selected variables. The difference between Rt^2 and Rp^2 , and RMSECV and RMSEP in the PLSR models based on GA-selected wavenumbers were also lower than that for the 2 other models (CARS and SI), which proves the improved performance of the PLSR model constructed by GA-selected variables.

In addition, it can be noticed in Table 3 that the GA method yielded the lowest SD, RSD, and PE. Therefore, the model established on wavelengths selected by the GA algorithm provided the best prediction

Table 3

Comparison of PLSR model results for different variable selection methods, using the top 20 % pixels of the maps, obtained by image threshold segmentation. Spectra were pre-processed using SNVT-SG.

Group	Variable screening method	Reference conc.	Predicted conc.	SD	RSD	PE	Rt^2	Rp^2	RMSECV	RMSEP	PC
		(μM)	(μM)	(μM)	(%)	(%)			(μM)	(μM)	
1	GA	104	111.80	4.27	3.82	7.55	0.85	0.73	16.13	14.81	4
		64.3	56.90	13.14	23.11	-11.58					
		88	77.10	15.20	19.72	-12.38					
2	CARS	104	96.27	7.06	7.34	-7.44	0.84	0.73	16.70	38.54	16
		64.3	21.80	19.43	89.13	-66.09					
		88	43.29	12.76	29.48	-50.81					
3	SI	104	81.79	7.24	8.85	-21.36	0.85	0.54	16.29	29.82	10
		64.3	34.91	17.18	49.20	-45.71					
		88	68.88	24.45	35.50	-21.73					
4	None	104	130.03	3.71	2.85	25.03	0.83	0.56	16.96	24.59	8
		64.3	68.03	24.81	36.46	5.80					
		88	85.00	22.02	25.91	-3.41					

results, which was consistent with the fact that the proportion of informative variables included by GA was the highest.

When comparing the results of the PLSR model based on SNVT-SG pre-processing and GA-selected variables (Group 1 in Table 3) with the univariate model based on 679 cm^{-1} MTX peak intensity (Group 3 in Table S1), it could be noticed that the SD, RSD, and PE obtained with the multivariate model were lower than for the univariate model. Based on this finding, we concluded that the multivariate model was more accurate than the univariate one and therefore, the SNVT-SG-GA analysis method was selected for further data analysis.

3.3. Influence of the sample matrix on the model performance

The model's performance is affected by the data analysis method and the nature of the samples included in the model. As a general rule, the prediction results of PLSR models are the best when the samples used to build the model are similar to the ones that need to be predicted. In this work, serum from two different sources was used to build the models: commercial serum and serum collected from patients before the initiation of their MTX therapy. It is expected that serum originating from patients would have more variability (patient-to-patient and intra-patient) and be a more complex matrix than the somewhat standardized commercial serum due to the altered physiological condition of patients. Therefore, the effect of the sample matrix was also evaluated in this study.

When spiked patient samples were used to construct the PLSR model to predict the MTX concentration in patient samples undergoing MTX therapy (Group 1 in Table 4), the overall values of RSD, PE, RMSECV and RMSEP were the lowest. Furthermore, it can be seen from Figure S4A that the patient samples were predicted with higher accuracy in that case. On the contrary, when a mix of spiked commercial and patient serum samples were used to build the PLSR model (Group 2 in Table 4) to predict the MTX concentration, the overall values of RSD, PE, RMSEP and RMSECV were much higher. The model could only accurately predict the concentration of 1 of the 3 patient samples (Figure S4B).

The MTX concentration prediction results were the worst when only spiked commercial serum samples were used to build PLSR models (Group 3 in Table 4). The RMSEP, the difference between RMSEP and RMSECV, and Rt^2 and Rp^2 were quite large, and only 1 patient sample could be accurately predicted (Figure S4C).

By comparing these 3 groups of results, it was thus concluded that the PLSR model built with only spiked commercial serum samples resulted in the poorest prediction performances for patient samples. In contrast, as expected, the model built with only patient samples yielded the best prediction performances. When the samples used to create the models and for prediction are comparable, it is easier to eliminate similar spectral interferences, and the resulting model shows better

Table 4

Influence of the sample type on PLSR results for the prediction of MTX concentration in 3 patient samples ($n = 3$). Spectra were pre-processed using SNVT-SG-GA. The top 20 % pixels of the maps, selected by image threshold segmentation, were used to build the models.

Group	Samples for training	Reference conc.	Predicted conc.	SD	RSD	Prediction error	Rt ²	Rp ²	RMSECV	RMSEP	PC
		(μM)	(μM)	(μM)	(%)	(%)			(μM)	(μM)	
1	48 spiked patient samples	104	119.17	7.51	6.30	14.59	0.85	0.79	18.45	16.24	6
		64.3	55.80	4.20	7.53	-13.22					
		88	86.17	20.28	23.53	-2.08					
2	48 mixed samples (24 spiked patient and 24 spiked commercial samples)	104	105.01	5.10	4.85	0.97	0.87	0.81	18.35	20.23	6
		64.3	42.20	16.89	40.03	-34.37					
		88	68.97	8.01	11.62	-21.63					
3	48 spiked commercial samples	104	108.89	6.40	5.88	4.70	0.89	0.51	16.67	22.30	5
		64.3	55.20	24.38	44.17	-14.15					
		88	74.92	24.04	32.09	-14.87					

prediction results. It is worth noting that models constructed and verified with the same batch of samples are prone to overfitting, meaning that the models can only be used to predict samples from this batch. The prediction results are usually poor when testing new samples from different natures. This is why a large number of samples from a wide range of sources are generally collected in multivariate modeling [43,44]. Nevertheless, as samples from only a few different patients (3) were included in the model, the prediction accuracy of the latter could be further improved by including samples from a larger number of patients. In this way, the physiological and pathological variability of serum in-between patients and within the same individuals would be taken into account in the model.

3.4. Influence of the number of samples on the model performance

We also investigated the model's performance in correlation with the number of samples included. Especially for highly variable and complex matrices, such as serum and other biological matrices, it is essential to have enough samples to build the model, to consider this high variability and create a robust predictive model.

The performance of PLSR models constructed with 64, 82, 100 or 112 spiked commercial serum samples were thus compared to investigate the effect of sample number (Table 5 and Figure S5). While the Rt² and Rp² of models built with 64 and 82 samples were similar, increasing the number of samples reduced the RMSEP (Table 5 group 1 and 2). When the number of samples increased to 100, difference between Rt² and Rp², and RMSEP were lower, (Table 5 group 3). Further increasing the number of samples to 128, increased the Rp² and decreased the RMSEP, and the differences between Rt² and Rp², and between RMSECV and RMSEP were also the smallest (group 4 in Table 5). We also found that this model could accurately predict the concentration of almost all patient samples (group 4 in Table 5 and Figure S5D).

Table 5

Influence of the amount of samples used to build the model on PLSR results for the prediction of MTX concentration in 3 patient samples ($n = 3$). Spectra were pre-processed using SNVT-SG-GA. The top 20 % pixels of the maps, selected by image threshold segmentation, were used to build the models.

Group	Samples for training (commercial serum)	Reference conc.	Predicted conc.	SD	RSD	Prediction error	Rt ²	Rp ²	RMSECV	RMSEP	PC
		(μM)	(μM)	(μM)	(%)	(%)			(μM)	(μM)	
1	64	104	139.29	5.23	3.76	33.94	0.92	0.56	14.25	41.82	11
		64.3	39.73	36.82	92.67	-38.21					
		88	58.70	33.99	57.91	-33.30					
2	82	104	125.91	1.72	1.37	21.07	0.91	0.53	13.60	26.05	7
		64.3	61.19	25.93	42.38	-4.84					
		88	80.93	28.64	35.38	-8.03					
3	100	104	123.86	4.44	3.59	19.10	0.86	0.68	16.77	20.20	4
		64.3	62.00	17.61	28.40	-3.58					
		88	90.27	22.12	24.50	2.58					
4	128	104	111.80	4.27	3.82	7.55	0.85	0.73	16.13	14.81	4
		64.3	56.90	13.14	23.11	-11.58					
		88	77.10	15.20	19.72	-12.38					

It was thus concluded that increasing the number of samples included in the models could help mitigate sample-to-sample spectral variability by having additional matrix background information, resulting in improved prediction performance.

3.5. Optimization of the image threshold segmentation process

During the NPAS process, as explained by Göksel *et al.* [1], sample migration and analyte separation from matrix compounds occur on the SERS substrate due to a wicking effect.

Considering this, only a portion of the SERS substrates contains the relevant spectral information related to MTX, and the distribution of MTX on the SERS substrate is mainly localized in the migration region. This distribution can also vary slightly from sample to sample. Appropriate image threshold segmentation methods could thus be used to solely include pixels containing relevant information related to MTX in the constructed PLSR models. Including less noise would, as a result, benefit the models in terms of MTX concentration prediction performance due to improved signal-to-noise ratios.

Different percentages (top 10 %, 20 %, 30 %, 40 % and 100 %) of image threshold segmentation were investigated to build PLSR models with 128 spiked commercial serum samples, and the performance of these models were compared by predicting the MTX concentration in 3 patient samples (Fig. 2 and Figure S6). In Table 6, it can be seen that when the top 30 % or top 20 % pixels (group 3 and 4, respectively) of spectral intensity were used for threshold segmentation, the results of training and prediction were the best, probably due to the most effective noise reduction. In both cases, all patient samples were predicted accurately, with slightly better performances of the top 20 % model (Table 6 group 4 and Fig. 2). On the contrary, including more pixels in the model increased the noise and deteriorated prediction performances (Table 6 group 1 and 2, and Figure S6A and B). Moreover, further

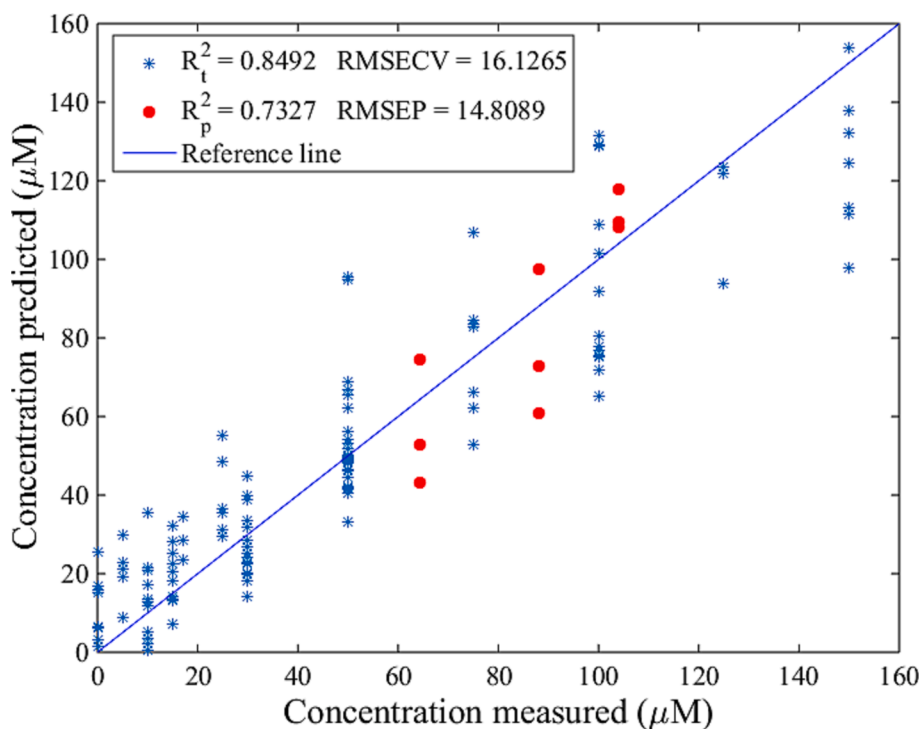


Fig. 2. Scatter plots of the PLSR model constructed with an image threshold segmentation of top 20 % pixels. The blue stars represent the samples used to build the models while the red circles represent the predicted 24 h MTX therapy patient samples.

Table 6

PLSR model results based on different percentages of image threshold segmentation for the prediction of MTX concentration in 3 patient samples ($n = 3$).

Group	Image threshold segmentation percentage (%)	Reference conc. (μM)	Predicted conc. (μM)	SD (μM)	RSD (%)	PE (%)	R_t^2	R_p^2	RMSECV (μM)	RMSEP (μM)
1	100	104	106.1	4.51	4.25	2.05	0.79	0.75	19.08	21.93
		64.3	42.9	3.18	7.4	-33.24				
		88	66.2	21.84	32.98	-24.75				
2	40	104	112.7	11.03	9.79	8.36	0.84	0.69	16.78	18.57
		64.3	53.3	16.3	30.58	-17.12				
		88	69.8	10.96	15.7	-20.67				
3	30	104	112.8	8.28	7.34	8.46	0.84	0.75	16.53	15.28
		64.3	55.5	13.44	24.23	-13.75				
		88	77.1	13.31	17.26	-12.41				
4	20	104	111.8	4.27	3.82	7.55	0.85	0.73	16.13	14.81
		64.3	56.9	13.14	23.11	-11.58				
		88	77.1	15.2	19.72	-12.38				
5	10	104	99.83	0.57	0.57	-4.01	0.86	0.46	15.54	16.34
		64.3	64.00	12.30	19.22	-0.47				
		88	89.13	25.10	28.17	1.29				

decreasing the amount of data included in the model also resulted in performance degradation (Table 6 group 5, and Figure S6D), which could be explained by the too-low amount of data included in the models.

A clear improvement was also noticed compared to the univariate model built (Group 3 in Table S1), with reduced SD (10.9 μM vs 13.2 μM), RSD (15.6 % vs 20.8 %), and PE (10.5 % vs -22.6 %).

The limits of detection and quantification (LOD and LOQ) of MTX using this data analysis strategy, for the model built with the top 20 %, were found to be 5.7 μM (LOD) and 17.0 μM (LOQ). The calculated LOD and LOQ values, using the here developed and the optimized multivariate method, based on multiple spectral features, proved to be lower than the ones obtained with the univariate model previously published [1], where the LOD and LOQ were 8 μM and 26.5 μM , respectively.

These results show that the data analysis method developed in this study based on the combination of standard normal variable transformation, SG smoothing, GA and image threshold segmentation (SNVT-

SG-GA-ITS-PLSR) can predict the concentration of MTX in patient samples based on a model solely constructed with spiked serum samples.

4. Conclusion

The developed data analysis strategy for SERS maps, based on the combination of SNVT-SG-GA-ITS-PLSR, enabled accurate prediction of MTX concentration in patient samples based on chemometrics models constructed with only commercial serum samples. Notably, SNVT baseline correction and GA screening method applied for data pre-processing showed better performance than other pre-processing and selection methods such as MSC, FD, detrend, CARS or SI. During optimization of the image threshold segmentation, we found that using the top 20 % pixels of spectra intensity led to better prediction performance.

Not only were the prediction results superior to the previously reported univariate, single spectral feature-based, data analysis method, with lower RSD (15.6 % vs 20.8 % from multivariate model with top 20

% pixels vs univariate model, respectively), and PE (10.5 % vs 22.6 % in average, respectively), but the samples used for model construction were also more readily available (commercial serum instead of patient samples). In addition, avoiding constructing calibration models in each patient serum represents a time- and resource-saving benefit. The developed method is relatively user-friendly, could be automated, and the constructed models could be updated with upcoming new samples to make it more robust. Moreover, the method can quickly complete data extraction, analysis and prediction (2 min).

In conclusion, the main scientific advance of this work is the combination of multivariate data analysis with SERS mapping. The developed and optimized data analysis approach proved to be accurate, rapid, easy to implement for quantifying MTX in patient serum by fully exploiting the information contained in SERS maps and could be implemented in SERS-based assays in real-life settings. We believe that the presented methodology could be adapted to other quantitative SERS-based assays.

CRedit authorship contribution statement

Peihuan He: Conceptualization, Methodology, Software, Writing – original draft, Visualization. **Elodie Dumont:** Conceptualization, Methodology, Writing – original draft, Visualization. **Yaman Göksel:** Conceptualization, Methodology, Writing – review & editing. **Roman Slipets:** Software, Writing – review & editing. **Kjeld Schmiegelow:** Resources, Writing – review & editing. **Quansheng Chen:** Writing – review & editing, Supervision. **Kinga Zor:** Conceptualization, Methodology, Writing – review & editing, Supervision, Project administration. **Anja Boisen:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgement

This work was supported by the Chinese Scholarship Council (CSC) fund (Grant No. 202008320551), the Danish National Research Foundation (DNRF122), Villum Fonden (Grant no. 9301) for Intelligent Drug delivery and sensing Using microcontainers and Nanomechanics (IDUN), BioInnovation Institute Foundation for Therapeutic drug monitoring (Grant No. NNF20SA0063552), the Danish Cancer Society (Grant No. R257-A14720), the Danish Childhood Cancer Foundation (Grant Nos. 2019-5934 and 2020-6759), and the National Natural Science Foundation of China (Grant No. 31871945).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.saa.2023.123536>.

References

- [1] Y. Göksel, K. Zor, T. Rindzevicius, B.E. Thorhauge Als-Nielsen, K. Schmiegelow, A. Boisen, Quantification of methotrexate in human serum using surface-enhanced raman scattering—toward therapeutic drug monitoring, *ACS Sensors* 6 (2021) 2664–2673, <https://doi.org/10.1021/acssensors.1c00643>.
- [2] J. Cailletaud, C. De Bleye, E. Dumont, P.Y. Sacré, L. Netchacovitch, Y. Gut, M. Boiret, Y.M. Ginot, P. Hubert, E. Ziemons, Critical review of surface-enhanced Raman spectroscopy applications in the pharmaceutical field, *J. Pharm. Biomed. Anal.* 147 (2018) 458–472, <https://doi.org/10.1016/j.jpba.2017.06.056>.
- [3] M. Viehrig, S.T. Rajendran, K. Sanger, M.S. Schmidt, T.S. Alstrom, T. Rindzevicius, K. Zór, A. Boisen, Quantitative SERS assay on a single chip enabled by electrochemically assisted regeneration: A Method for Detection of Melamine in Milk, *Anal. Chem.* 92 (2020) 4317–4325, <https://doi.org/10.1021/acs.analchem.9b05060>.
- [4] L. Jiang, M.M. Hassan, S. Ali, H. Li, R. Sheng, Q. Chen, Evolving trends in SERS-based techniques for food quality and safety: A review, *Trends Food Sci. Technol.* 112 (2021) 225–240, <https://doi.org/10.1016/j.tifs.2021.04.006>.
- [5] H. Wei, S.M. Hossein Abtahi, P.J. Vikesland, Plasmonic colorimetric and SERS sensors for environmental analysis, *Environ. Sci. Nano* 2 (2015) 120–135, <https://doi.org/10.1039/c4en00211c>.
- [6] J. Langer, D.J. de Aberasturi, J. Aizpurua, R.A. Alvarez-Puebla, B. Auguie, J. J. Baumberg, G.C. Bazan, S.E.J. Bell, A. Boisen, A.G. Brolo, J. Choo, D. Cialla-May, V. Deckert, L. Fabris, K. Faulds, F. Javier Garcia de Abajo, R. Goodacre, D. Graham, A.J. Haes, C.L. Haynes, C. Huck, T. Itoh, M. Käll, J. Kneipp, N.A. Kotov, H. Kuang, E.C. Le Ru, H.K. Lee, J.F. Li, X.Y. Ling, S.A. Maier, T. Mayerhöfer, M. Moskovits, K. Murakoshi, J.M. Nam, S. Nie, Y. Ozaki, I. Pastoriza-Santos, J. Perez-Juste, J. Popp, A. Pucci, S. Reich, B. Ren, G.C. Schatz, T. Shegai, S. Schlücker, L.L. Tay, K. George Thomas, Z.Q. Tian, R.P. van Duyne, T. Vo-Dinh, Y. Wang, K.A. Willets, C. Xu, H. Xu, Y. Xu, Y.S. Yamamoto, B. Zhao, L.M. Liz-Marzán, Present and future of surface-enhanced Raman scattering, *ACS Nano* 14 (2020) 28–117, <https://doi.org/10.1021/acsnano.9b04224>.
- [7] P. Liu, R. Wu, H. Li, J. Huang, Z. Liu, X. Geng, Y. Xiong, S. Ai, Exploring variables optimization methods to screen surface-enhanced Raman spectroscopy characteristic peaks for rapid detection of difenoconazole pesticides in tea, *Vib. Spectrosc.* 123 (2022), 103448, <https://doi.org/10.1016/j.vibspec.2022.103448>.
- [8] H. Pan, W. Ahmad, T. Jiao, A. Zhu, Q. Ouyang, Q. Chen, Label-free Au NRs-based SERS coupled with chemometrics for rapid quantitative detection of thiabendazole residues in citrus, *Food Chem.* 375 (2022), 131681, <https://doi.org/10.1016/j.foodchem.2021.131681>.
- [9] T. Jiao, M. Mehedi Hassan, J. Zhu, S. Ali, W. Ahmad, J. Wang, C. Lv, Q. Chen, H. Li, Quantification of deltamethrin residues in wheat by Ag@ZnO NFs-based surface-enhanced Raman spectroscopy coupling chemometric models, *Food Chem.* 337 (2021), 127652, <https://doi.org/10.1016/j.foodchem.2020.127652>.
- [10] Y. Xu, F.Y.H. Kutsanedzie, M.M. Hassan, H. Li, Q. Chen, Synthesized Au NPs@silica composite as surface-enhanced Raman spectroscopy (SERS) substrate for fast sensing trace contaminant in milk, *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.* 206 (2019) 405–412, <https://doi.org/10.1016/j.saa.2018.08.035>.
- [11] X.M. Lin, Y. Cui, Y.H. Xu, B. Ren, Z. Q. Tian, Surface-enhanced raman spectroscopy: Substrate-related issues, *Anal. Bioanal. Chem.* 394 (2009) 1729–1745, <https://doi.org/10.1007/s00216-009-2761-5>.
- [12] C. Zong, M. Xu, L.J. Xu, T. Wei, X. Ma, X.S. Zheng, R. Hu, B. Ren, Surface-Enhanced Raman Spectroscopy for Bioanalysis: Reliability and Challenges, *Chem. Rev.* 118 (2018) 4946–4980, <https://doi.org/10.1021/acs.chemrev.7b00668>.
- [13] S.E.J. Bell, N.M.S. Sirimuthu, Quantitative surface-enhanced Raman spectroscopy, *Chem. Soc. Rev.* 37 (2008) 1012–1024, <https://doi.org/10.1039/b705965p>.
- [14] E. Kozhina, S. Bedin, A. Martynov, S. Andreev, A. Piryazev, Y. Grigoriev, Y. Gorbunova, A. Naumov, Ultrasensitive Optical Fingerprinting of Biorelevant Molecules by Means of SERS-Mapping on Nanostructured Metasurfaces, *Biosensors* 13 (2023) 46, <https://doi.org/10.3390/bios13010046>.
- [15] S. Fornasaro, C. Beleites, V. Sergio, A. Bonifacio, Data analysis in SERS diagnostics, in: A. Fales (Ed.), *SERS for Point-Of-care and Clinical Applications*, Elsevier, 2022, 1–51, <https://doi.org/10.1016/B978-0-12-820548-8.00002-3>.
- [16] S. Rostami, K. Zór, D.S. Zhai, M. Viehrig, L. Morelli, A. Mehdiinia, J. Smedsgaard, T. Rindzevicius, A. Boisen, High-throughput label-free detection of Ochratoxin A in wine using supported liquid membrane extraction and Ag-capped silicon nanopillar SERS substrates, *Food Control* 113 (2020), 107183, <https://doi.org/10.1016/j.foodcont.2020.107183>.
- [17] O. Durucan, T. Rindzevicius, M.S. Schmidt, M. Matteucci, A. Boisen, Nanopillar Filters for Surface-Enhanced Raman Spectroscopy, *ACS Sensors* 2 (2017) 1400–1404, <https://doi.org/10.1021/acssensors.7b00499>.
- [18] L. Morelli, L. Seriola, F.A. Centorbi, C.B. Jendresen, M. Matteucci, O. Ilchenko, D. Demarchi, A.T. Nielsen, K. Zór, A. Boisen, Injection molded lab-on-a-disc platform for screening of genetically modified: E. coli using liquid-liquid extraction and surface enhanced Raman scattering, *Lab Chip* 18 (2018) 869–877, <https://doi.org/10.1039/c7lc01217a>.
- [19] O. Durucan, K. Wu, M. Viehrig, T. Rindzevicius, A. Boisen, Nanopillar-Assisted SERS Chromatography, *ACS Sensors* 3 (2018) 2592–2598, <https://doi.org/10.1021/acssensors.8b00887>.
- [20] A. Kamińska, T. Szymorski, E. Witkowska, E. Kijeńska-Gawrońska, W. Świączkowski, K. Niciński, J. Trzcinańska-Danielewicz, A. Girstun, Detection of circulating tumor cells using membrane-based sers platform: A new diagnostic approach for ‘liquid biopsy’, *Nanomaterials* 9 (2019) 366, <https://doi.org/10.3390/nano9030366>.
- [21] A. Matschulat, D. Drescher, J. Kneipp, Surface-enhanced Raman scattering hybrid nanoprobe multiplexing and imaging in biological systems, *ACS Nano* 4 (2010) 3259–3269, <https://doi.org/10.1021/nn100280z>.
- [22] X. Li, T. Yang, S. Li, D. Wang, Y. Song, K. Yu, Different classification algorithms and serum surface enhanced Raman spectroscopy for noninvasive discrimination of gastric diseases, *J. Raman Spectrosc.* 47 (2016) 917–925, <https://doi.org/10.1002/jrs.4924>.
- [23] A. Zhu, S. Ali, Y. Xu, Q. Ouyang, Z. Wang, Q. Chen, SERS-based Au@Ag NPs Solid-phase substrate combined with chemometrics for rapid discrimination of multiple

- foodborne pathogens, *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.* 270 (2022), 120814, <https://doi.org/10.1016/j.saa.2021.120814>.
- [24] N.E. Dina, A.M.R. Gherman, V. Chiş, C. Sărbu, A. Wieser, D. Bauer, C. Haisch, Characterization of Clinically Relevant Fungi via SERS Fingerprinting Assisted by Novel Chemometric Models, *Anal. Chem.* 90 (2018) 2484–2492, <https://doi.org/10.1021/acs.analchem.7b03124>.
- [25] Y. Zhu, S. Liu, M. Li, W. Liu, Z. Wei, L. Zhao, Y. Liu, L. Xu, G. Zhao, Y. Ma, Preparation of an AgNPs@Polydimethylsiloxane (PDMS) multi-hole filter membrane chip for the rapid identification of food-borne pathogens by surface-enhanced Raman spectroscopy, *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.* 267 (2022), 120456, <https://doi.org/10.1016/j.saa.2021.120456>.
- [26] S. Shakeel, H. Nawaz, M.I. Majeed, N. Rashid, M.R. Javed, A. Tariq, B. Majeed, A. Sehar, S. Murtaza, N. Sadaf, G. Rimsha, I. Amin, Surface-enhanced Raman spectroscopic analysis of centrifugally filtered blood serum samples of hepatitis C patients, *Photodiagn. Photodyn. Ther.* 39 (2022), 102949, <https://doi.org/10.1016/j.pdpdt.2022.102949>.
- [27] S. Fornasaro, A. Bonifacio, E. Marangon, M. Buzzo, G. Toffoli, T. Rindzevicius, M. S. Schmidt, V. Sergio, Label-Free Quantification of Anticancer Drug Imatinib in Human Plasma with Surface Enhanced Raman Spectroscopy, *Anal. Chem.* 90 (2018) 12670–12677, <https://doi.org/10.1021/acs.analchem.8b02901>.
- [28] J. Zhu, A.A. Agyekum, F.Y.H. Kutsanedzie, H. Li, Q. Chen, Q. Ouyang, H. Jiang, Qualitative and quantitative analysis of chlorpyrifos residues in tea by surface-enhanced Raman spectroscopy (SERS) combined with chemometric models, *Lwt.* 97 (2018) 760–769, <https://doi.org/10.1016/j.lwt.2018.07.055>.
- [29] L. Morelli, F.A. Centorbi, O. Ilchenko, C.B. Jendresen, D. Demarchi, A.T. Nielsen, K. Zór, A. Boisen, Simultaneous quantification of multiple bacterial metabolites using surface-enhanced Raman scattering, *Analyst* 144 (2019) 1600–1607, <https://doi.org/10.1039/c8an02128g>.
- [30] G. Soufi, E. Dumont, Y. Göksel, R. Slipets, R.A. Raja, K. Schmiegelow, H. Bagheri, A. Boisen, K. Zor, Discrimination and quantification of methotrexate in the presence of its metabolites in patient serum using SERS mapping, assisted by multivariate spectral data analysis, *Biosens. Bioelectron.: X* 14 (2023), 100382, <https://doi.org/10.1016/j.biosx.2023.100382>.
- [31] Y. Chen, C.H. Moran, Z. Tan, A. Lake Wooten, J.A. O'Sullivan, Robust analysis of multiplexed SERS microscopy of Ag nanocubes using an alternating minimization algorithm, *J. Raman Spectrosc.* 44 (2013) 703–709, <https://doi.org/10.1002/jrs.4241>.
- [32] C.M. Galloway, E.C. Le Ru, P.G. Etchegoin, An iterative algorithm for background removal in spectroscopy by wavelet transforms, *Appl. Spectrosc.* 63 (2009) 1370–1376, <https://doi.org/10.1366/000370209790108905>.
- [33] J. Zhu, W. Ahmad, Y. Xu, S. Liu, Q. Chen, M.M. Hassan, Q. Ouyang, Development of a novel wavelength selection method for the trace determination of chlorpyrifos on Au@Ag NPs substrate coupled surface-enhanced Raman spectroscopy, *Analyst* 144 (2019) 1167–1177, <https://doi.org/10.1039/c8an02086h>.
- [34] H. Li, M. Mehed Hassan, J. Wang, W. Wei, M. Zou, Q. Ouyang, Q. Chen, Investigation of nonlinear relationship of surface enhanced Raman scattering signal for robust prediction of thiabendazole in apple, *Food Chem.* 339 (2021), 127843, <https://doi.org/10.1016/j.foodchem.2020.127843>.
- [35] A. Bonifacio, Label-free surface-enhanced Raman scattering for clinical applications, In: Y. Wang (Ed.), *Principles and clinical diagnostic applications of surface-enhanced Raman spectroscopy*, Elsevier, 2022, 125–170. <https://doi.org/10.1016/B978-0-12-821121-2.00001-9>.
- [36] H. Li, Y. Liang, Q. Xu, D. Cao, Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration, *Anal. Chim. Acta* 648 (2009) 77–84, <https://doi.org/10.1016/j.aca.2009.06.046>.
- [37] R. Leardi, A. Lupiáñez González, Genetic algorithms applied to feature selection in PLS regression: How and when to use them, *Chemom. Intell. Lab. Syst.* 41 (1998) 195–207, [https://doi.org/10.1016/S0169-7439\(98\)00051-3](https://doi.org/10.1016/S0169-7439(98)00051-3).
- [38] R. Leardi, Application of genetic algorithm-PLS for feature selection in spectral data sets, *J. Chemom.* 14 (2000) 643–655, [https://doi.org/10.1002/1099-128X\(200009/12\)14:5/6<643::AID-CEM621>3.0.CO;2-E](https://doi.org/10.1002/1099-128X(200009/12)14:5/6<643::AID-CEM621>3.0.CO;2-E).
- [39] H. Shinzaw, B. Li, T. Nakagawa, K. Maruo, Y. Ozaki, Multi-objective genetic algorithm-based sample selection for partial least squares model building with applications to near-infrared spectroscopic data, *Appl. Spectrosc.* 60 (2006) 631–640, <https://doi.org/10.1366/00037020677670576>.
- [40] L. Nørgaard, A. Saudland, J. Wagner, J.P. Nielsen, L. Munck, S.B. Engelsen, Interval partial least-squares regression (iPLS): A comparative chemometric study with an example from near-infrared spectroscopy, *Appl. Spectrosc.* 54 (2000) 413–419, <https://doi.org/10.1366/0003702001949500>.
- [41] P.Y. Sacré, C. De Bleye, P.F. Chavez, L. Netchacovitch, P. Hubert, E. Ziemons, Data processing of vibrational chemical imaging for pharmaceutical applications, *J. Pharm. Biomed. Anal.* 101 (2014) 123–140, <https://doi.org/10.1016/j.jpba.2014.04.012>.
- [42] R. Leardi, Genetic algorithms in chemometrics and chemistry: A review, *J. Chemom.* 15 (2001) 559–569, <https://doi.org/10.1002/cem.651>.
- [43] H. Zhang, X. Hu, L. Liu, J. Wei, X. Bian, Near infrared spectroscopy combined with chemometrics for quantitative analysis of corn oil in edible blend oil, *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.* (270 2022), 120841, <https://doi.org/10.1016/j.saa.2021.120841>.
- [44] G. Shen, X. Kang, J. Su, J. Qiu, X. Liu, J. Xu, J. Shi, S.R. Mohamed, Rapid detection of fumonisin B1 and B2 in ground corn samples using smartphone-controlled portable near-infrared spectrometry and chemometrics, *Food Chem.* 384 (2022), 132487, <https://doi.org/10.1016/j.foodchem.2022.132487>.