POLITECNICO DI TORINO
Repository ISTITUZIONALE

Surface Enhanced Raman Spectroscopy for Quantitative Analysis: Results of a Large-Scale European Multi-Instrument Interlaboratory Study

Original
Surface Enhanced Raman Spectroscopy for Quantitative Analysis: Results of a Large-Scale European Multi-Instrument Interlaboratory Study / Fornasaro, S., Alsamad, F., Baia, M., Batista De Carvalho, L.A.E., Beleites, C., Byrne, H.J., Chiadò, Alessandro., Chis, M., Chisanga, M., Daniel, A., Dybas, J., Eppe, G., Falgayrac, G., Faulds, K., Gebavi, H., Giorgis, F., Goodacre, R., Graham, D., La Manna, P., Laing, S., et al.. - In: ANALYTICAL CHEMISTRY. - ISSN 0003-2700. - 92:5(2020), pp. 4053-4064. [10.1021/acs.analchem.9b05658]

Availability:
This version is available at: 11583/2851238 since: 2020-11-05T22:39:59Z

Publisher:
American Chemical Society

Published
DOI:10.1021/acs.analchem.9b05658

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
GENERICO -- per es. Nature : semplice rinvio dal preprint/submitted, o postprint/AAM [ex default]

(Article begins on next page)

11 September 2026

Surface Enhanced Raman Spectroscopy for quantitative analysis: results of a large-scale European multi-instrument interlaboratory study.

Stefano Fornasaro[a], Fatima Alsamad[b], Monica Baia[c], Luís A.E. Batista de Carvalho[d], Claudia Beleites[e], Hugh J. Byrne[f], Alessandro Chiadò[g], Mihaela Chis[c], Malama Chisanga[h], Amuthachelvi Daniel[i], Jakub Dybas[j], Gauthier Eppe[k], Guillaume Falgayrac[l], Karen Faulds[o], Hrvoje Gebavi[m], Fabrizio Giorgis[g], Royston Goodacre[n], Duncan Graham[o], Pietro La Manna[p], Stacey Laing[o], Lucio Litti[q], Fiona M. Lyng[i], Kamilla Malek[j], Cedric Malherbe[k], Maria P.M. Marques[d,r], Moreno Meneghetti[q], Elisa Mitri[a], Vlasta Mohaček-Grošev[m], Carlo Morasso[s], Howbeer Muhamadali[n], Pellegrino Musto[p], Chiara Novara[g], Marianna Pannico[p], Guillaume Penel[l], Olivier Piot[b], Tomas Rindzevicius[t], Elena A. Rusu[c], Michael S. Schmidt[u], Valter Sergo[a,v], Ganesh D Sockalingum[b], Valérie Untereiner[b], Renzo Vanna[s], Ewelina Wiercigroch[j], Alois Bonifacio[a*].

[a] Raman Spectroscopy Lab, Department of Engineering and Architecture, University of Trieste, P.le Europa 1, 34100 Trieste, Italy; [b] Université de Reims Champagne-Ardenne, BioSpecT-EA7506, UFR de Pharmacie, 51 rue Cognacq-Jay, 51097 Reims, France; [c] Faculty of Physics, Babes-Bolyai University, M. Kogalniceanu 1, 400084 Cluj-Napoca, Romania; [d] Molecular-Physical Chemistry R&D Unit, Department of Chemistry, University of Coimbra, 3004-535 Coimbra, Portugal; [e] Chemometrix GmbH, Södeler Weg 19, 61200 Wölfersheim, Germany; [f] FOCAS Research Institute, Technological University Dublin, Kevin Street, Dublin 8, Ireland; [g] Department of Applied Science and Technology, Politecnico di Torino, C.so Duca degli Abruzzi 24, 10129 Torino, Italy; [h] School of Chemistry, Manchester Institute of Biotechnology, University of Manchester, Manchester, UK, M1 7DN; [i] Radiation and Environmental Science Centre, FOCAS Research Institute, Technological University Dublin, Kevin Street, Dublin 8, Ireland; [j] Faculty of Chemistry, Jagiellonian University, ul. Gronostajowa 2, 30-387 Krakow, Poland; [k] Mass Spectrometry Laboratory (MSLab), MolSys RU, University of Liège, Liège, Belgium; [l] Univ. Lille, Univ. Littoral Côte d'Opale, EA 4490 - PMOI, F-59000 Lille, France; [m] Centre of Excellence for Advanced Materials and Sensing Devices, Division of Materials Physics, Rudjer Boskovic Institute, Bijenicka c. 54, 10000 Zagreb, Croatia; [n] Department of Biochemistry, Institute of Integrative Biology, University of Liverpool, Liverpool, UK, L69 7ZB; [o] Bionanotechnology Research Section, Department of Pure and Applied Chemistry, University of Strathclyde, 99 George Street, Glasgow, G1 1RD, United Kingdom; [p] Institute on Polymers, Composites and Biomaterials, National Research Council of Italy, via Campi Flegrei, 34, Pozzuoli, NA 80078, Italy; [q] Nanostructures & Optics Laboratory, Department of Chemical Sciences, University of Padova, Via Marzolo 1 - 35131, Padova, Italy;[r] Department of Life Sciences, University of Coimbra, 3000-456 Coimbra, Portugal; [s] Nanomedicine and Molecular Imaging Lab, Istituti Clinici Scientifici Maugeri IRCCS, Via Maugeri 4, 27100 Pavia, Italy; [t] Technical University of Denmark, Dep. of Health Technology, Ørsteds Plads, Building 345C, DK-2800 Kgs. Lyngby, Denmark; [u] Silmeco ApS, Kenny Drews Vej 101, 2450 Copenhagen, Denmark; [v] Faculty of Health Sciences, University of Macau, SAR Macau, China.

ABSTRACT: Surface-enhanced Raman scattering (SERS) is a powerful and sensitive technique for the detection of fingerprint signals of molecules and for the investigation of a series of surface chemical reactions. Many studies introduced quantitative applications of SERS in various fields and several SERS methods have been implemented for each specific application, ranging in performance characteristics, analytes used, instruments, and analytical matrices. In general, very few methods have been validated according to international guidelines. As a consequence, the application of SERS in highly-regulated environments is still considered risky and the perception of a poorly reproducible and insufficiently robust analytical technique has persistently retarded its routine implementation. Collaborative trials are a type of interlaboratory study (ILS) frequently performed to ascertain the quality of a single analytical method. The idea of an ILS of quantification with SERS arose within the framework of Working Group 1 (WG1) of the COST Action BM1401 Raman4Clinics in an effort to overcome the problematic perception of quantitative SERS methods. Here we report the first interlaboratory SERS study ever conducted, involving 15 laboratories and 44 researchers. In this study we tried to define a methodology to assess the reproduc-

ibility and trueness of a quantitative SERS method, and to compare different methods. In our opinion, this is a first important step toward a "standardization" process of SERS protocols, not proposed by a single laboratory but by a larger community.

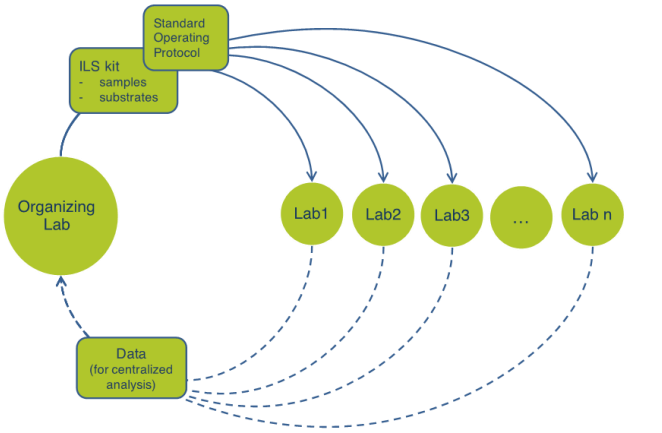
Surface-enhanced Raman scattering (SERS) is a powerful and sensitive technique for the detection of fingerprint signals of molecules and for the investigation of a series of surface chemical reactions.[1] Several monographs and reviews describe the mechanisms of SERS, confirming that metal (mostly silver and gold) nanostructures can generate a strong local electromagnetic field upon illumination with light having a wavelength capable of exciting localized surface plasmons. It is generally agreed that this electromagnetic mechanism (EM), as well as the chemical mechanism (CM), occurring when a chemical bond is formed between the metal and the adsorbed analyte, lead to the Raman signal enhancement of those analytes located close to or directly adsorbed onto the metal surface.[2] In spite of its established sensitivity, SERS applied to quantitative analysis is still very challenging,[3-6] and there is no general consensus on the key factors affecting performance.[7] A number of studies have previously been designed to address some important issues regarding signal (enhancement) variability in SERS studies.[8-11] Guicheteau et al.[10] undertook an extensive, collaborative study within the US to design and implement an evaluation protocol for SERS active surfaces enabling the definition of an enhancement value by which different substrates can be directly compared. However, there are other factors that cannot be easily disentangled from the substrate-related issues, such as the characteristics of the laser light used for excitation, the protocol for sample preparation (i.e. the way the analyte is put in contact with the metal surface) or the type of approaches used for data pre-processing and regression analysis. Muehlethaler et al.[11] undertook a systematic study of such different aspects of the analytical procedure, in an attempt to qualitatively validate the SERS technique for forensic purposes, albeit in a single laboratory environment. Independently, many studies introduced quantitative applications of SERS in various fields, such as quantification of biomarkers, drugs and related metabolites in biofluids, [12-22] or the determination of pesticides or toxins in foodstuffs or other biological samples.[23-25] Several SERS substrates have been designed and tested for each specific application. Thus, among the published methods, there is a wide range in performance characteristics, analytes used, instruments, and analytical matrices.[6,26,27] In general, very few methods have been validated according to international guidelines.[28] In the literature, the evaluation of figures of merit (when performed) has been limited, as the validation protocols of the analytical method included only one equipment/laboratory. The robustness of results is seldom assessed. As a consequence, the application of SERS in highly-regulated environments is still considered risky and the perception of a poorly reproducible and insufficiently robust analytical technique has persistently retarded its routine implementation outside academia. On the other hand, guidelines concerning the validation of analytical procedures, as detailed in official documents (regulatory or normative),[29] do not explicitly cover the application of SERS, or even related techniques such as normal Raman spectroscopy. Collaborative trials (also called method performance studies) are a type of interlaboratory study (ILS) frequently performed to ascertain the performance (generally expressed as repeatability and reproducibility) of a single analytical method.[30] As a part of the full method validation process, collaborative trials are very powerful tools to prove that an analytical method is indeed doing what it is supposed to do, independent of the laboratory in which the test is performed.[31] Various examples can be found in the literature, especially for chromatographic techniques, but so far, to the best of our knowledge, such data have never been published for quantitative SERS, even if some examples can be found for qualitative methods[11]. In this context, the use of arbitrarily defined criteria based on the experience of laboratories is a viable option.[32] The idea of an ILS of quantification with SERS arose within the framework of Working Group 1 (WG1) of the COST Action BM1401 Raman4Clinics[33] in an effort to overcome the problematic perception of quantitative SERS methods by addressing the following two questions:

i) given the simplest conditions (i.e. a well-defined, well-known analyte in a simple matrix such as a buffered aqueous solution), can a quantitative SERS method be consistently implemented by different laboratories?

ii) if different SERS methods are used to quantify the same analyte, which is the best way to compare them?

These two general questions need to be further clarified. First, one must clearly define what a "SERS method" is. SERS signals of the same analyte strongly depend on the type of the metal surface and on the choice of the laser excitation wavelength.[34] Therefore, the "complete" definition of a SERS method should take into account both a specific metal nanostructure (i.e. the substrate) and a specific laser wavelength (e.g. colloidal Ag excited at 785 nm, for brevity we use "cAg@785nm") as well as further working conditions. In this way, a method is completely described by the related "standard operating procedure" (SOP); ours are detailed within the accompanying Supplementary Information. Second, the above question (i) actually consists of two distinct aspects: reproducibility and trueness (the meaning of both terms and their relation to accuracy, according to ISO 5725, are described in the method section). In other words: when a SERS method is applied by different labs using different instrumental set-ups, how similar to each other are the results obtained (reproducibility)? And how "close" to the truth are the results obtained (trueness)? We would like to stress that, at this stage, the reproducibility (also referred to as "precision") is more important than the trueness, since reproducibility is usually

considered the main concern for SERS methods. Once reproducibility has been assessed, trueness can be considered as well, as both are aspects of the overall accuracy for a method (see Table 1). In the light of these clarifications, the second question (ii) can be better re-phrased as: can one compare two or more SERS methods to assess which is the most reproducible?



METHODS TESTED	colloidal substrates		solid substrates	
	Ag	Au	Ag	Au
excitation in the green (532 or 514 nm)	**cAg@532**	-	**sAg@532**	-
excitation in the NIR (785 nm)	**cAg@785**	**cAu@785**	**sAg@785**	**sAu@785**

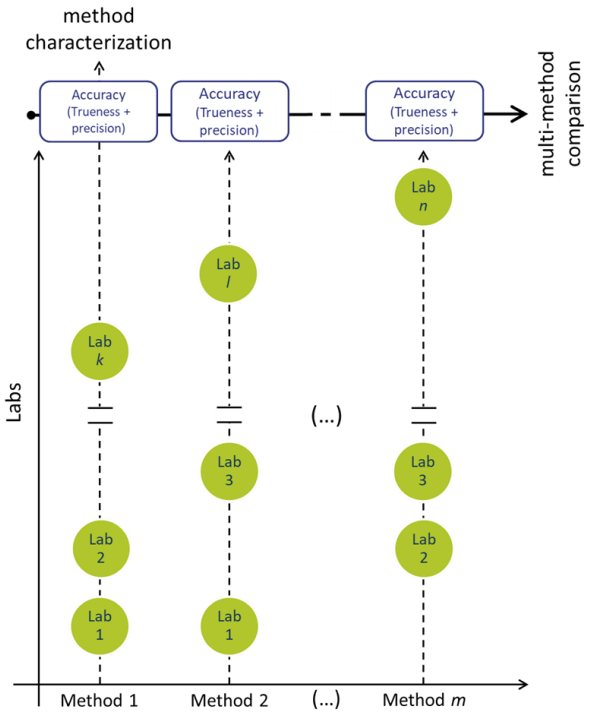


Figure 1. Organization structure of the Raman4Clinics WG1 interlaboratory study (ILS).

To answer these questions, we set up an ILS in which six different SERS methods, involving Ag and Au plasmonic nanostructures (both colloidal and solid substrates), have been considered for the determination of adenine concentrations, chosen as the standard analyte (see Experimental Section for the justification of this choice). On one hand, we are well aware that the results of such a study are bound to be specific to the systems considered (i.e. adenine on very specific substrates), so that no extrapolation can yet be made to other substrates or analytes. We also expect the results to have a broader significance going beyond the system studied, especially as far as the methodology proposed to assess and compare SERS methods is concerned.

EXPERIMENTAL SECTION

After the voluntary enrolment of the participating institutions from the membership of Raman4Clinics, the ILS started by collectively defining the study design, which was conclusively decided during a COST-Raman4Clinics WG1 meeting in Trieste in March 2018. The discussion eventually led to a study design in which 6 SERS methods are tested, and each method is independently evaluated by up to eight laboratories on the basis of the same SOP, using the same centrally provided materials for sample preparation. Samples and substrates were sent from the ILS organizing laboratory (OL, University of Trieste) to all participants, who sent all the data collected back to the OL for centralized data analysis (Figure 1). The results were combined to estimate the reproducibility obtained using the same protocol in different laboratories (along with trueness and accuracy) by calculating the corresponding Figures of Merit (FoMs, see Table 1 for a complete description). An SOP was prepared by the ILS OL and shared and agreed among the participants. The SOP detailed all the experimental procedures to be carried out by each ILS participant, detailing how to prepare the samples for analysis, how to perform a measurement and which format aimed to export data is used. All laboratories were given a chance to provide feedback on the SOP, and revisions were introduced into a final document. For brevity, only the essential elements of the methodology are summarized here; for a more detailed description along with complete formulae for all calculations, see Supporting Information (Section S1, "Standard Operating Procedure").

Each participant received a kit containing the necessary materials to prepare the samples and the SERS substrates necessary to perform the measurements, as detailed in the SOP. To ensure homogeneity, the kits were all assembled by the OL, using the same reagents and materials, and shipped to all ILS participants (Figure 1). For each experiment, all the participants prepared one calibration set (to build the regression model) and one test set (to validate the regression model and to compare the results) of samples (Figure 2). They performed SERS measurements using their own setups and instruments and sent the raw spectral data back to the OL. Only the OL knew the concentration of the test samples. The period for active participation was from July to October 2018 (see Supplementary Figure S1).

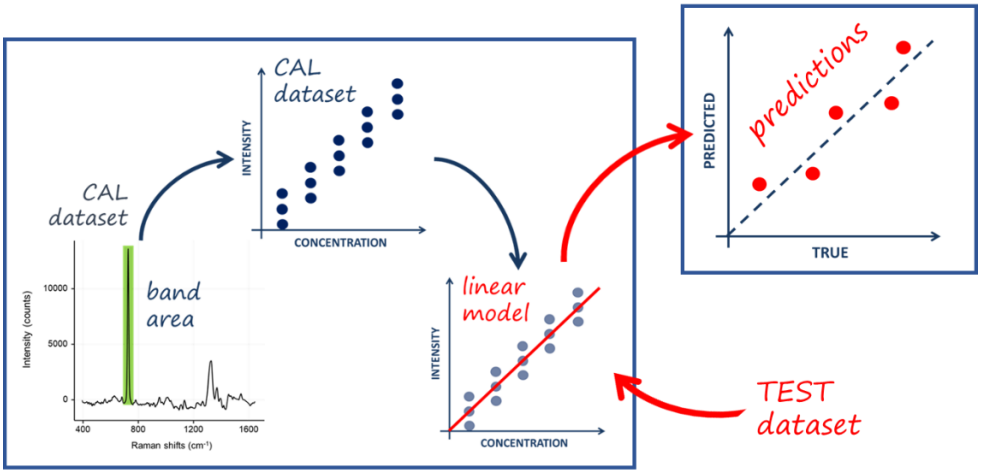


Figure 2. Single experiment validation scheme.

Samples. All chemicals were acquired, aliquoted and shipped by the OL. The selection of the analyte and of the SERS substrates to be used by the participants was mainly based on experience within the SERS community, the feasibility of performing experiments and the availability of substrates. Selection criteria for the reference analyte included: the ability to be detected by all the tested methods (and thus to adsorb on both Ag and Au substrates); presence of intense, well-characterized bands in the SERS spectrum; stability to light and temperature (to avoid complications during shipping and sample preparation); absence of toxicity (to simplify international shipping and sample handling); absence of tautomers at the measurement conditions (to ensure a SERS spectrum as simple as possible); absence of thiol groups in the molecule (which lead to a very specific and strong metal-sulfur covalent bond, not present in many analytes of interest such as drugs); absence of electronic transitions in resonance with the excitation wavelengths used in the study (to allow the use of different excitation wavelengths while ruling out resonance Raman effects); commercial availability at a reasonable cost. Eventually, after testing a shortlist of several substances, the choice fell on adenine as the standard analyte, while the choice of the matrix fell on pH 7.4 phosphate buffer (0.01 M), to ensure a constant pH and considering the availability of easily shipped buffer tablets ready to be dissolved. It should be stressed that many reasons behind the choice of adenine as analyte are exclusively practical (e.g. stability, non-toxicity, affinity for both Ag and Au). It is very likely that other analytes would have been a better choice in terms of performance, but lacked other characteristics which were deemed as necessary given the available resources. The reader is referred to Section S1, part 3, for a detailed description of the sample preparation.

SERS methods. The discussion on the selection of the SERS substrates to be used involved organizational aspects, such as total number of the samples, sustainability and reproducibility of metal nanostructures synthesis, and number of participants. There are several criteria to classify SERS substrates (e.g. top-down or bottom-up synthetic methods, surface characteristics, etc.), but most of them fall into one of two broad categories: colloidal or non-colloidal (usually referred to as solid substrates) substrates. Everyone involved agreed that both these types of substrates should be included in the study. All participants were given the option of contributing with their own substrates, but eventually only one type of colloidal substrate and one type of solid substrate were offered as available for such large number of experiments. As colloidal substrates, naked Ag and Au nanoparticles obtained by laser ablation synthesis in solution[35] were provided by the University of Padova. Silmeco provided their commercially available Ag and Au solid substrates,[36] based on metal-coated silicon nanopillars. The six SERS methods considered for this ILS are reported in Figure 1. The 785 nm and 532 nm excitation were selected as the most commonly available among ILS participants. One participant had a 514 nm laser instead of a 532 nm, but, on the basis of the previous experience within the working group, such difference has been considered as negligible, especially as non-resonant molecules were analyzed.

SERS measurements. Details on how to perform the measurements can be found in the SOP (Section S1, part 4.1 and 4.2), but we think it is particularly relevant to briefly report here some details about the number of measurements performed for each method. For colloidal substrates, 3 different batches of colloids for each metal were used, and participants were asked to collect, for each sample, 3 replicates using the 3 different batches (i.e. 3 spectra, one for each batch). For solid substrates, the official indication from the producer was to acquire a full map of several tens of spectra for each substrate. However, this would have translated into a substantial amount of work for each participant. Instead of a full map, the final version of the SOP prescribed the collection of 3 spectra (from random locations) for each substrate, and to use 3 different substrates for each sample, for a total of 9 spectra for each sample (calibration or test).

Instruments. Six different models of Raman instruments were used in this ILS, from three different manufacturers (Figure 3). The instruments used were all calibrated according to the finalized protocol (Section S1, part 4.3). Given the variety of set-up characteristics used in the study, we decided to leave the choice of instrumental parameters such as laser power and acquisition time to each participant, as far as the signal-to-noise ratio of the sample containing the smallest amount of analyte was acceptable, suggesting upper thresholds of the laser power density and proper optical magnification (10x or 20x) for each method, to minimize potential sample photo-degradation (see Supplementary Material, Section S1).

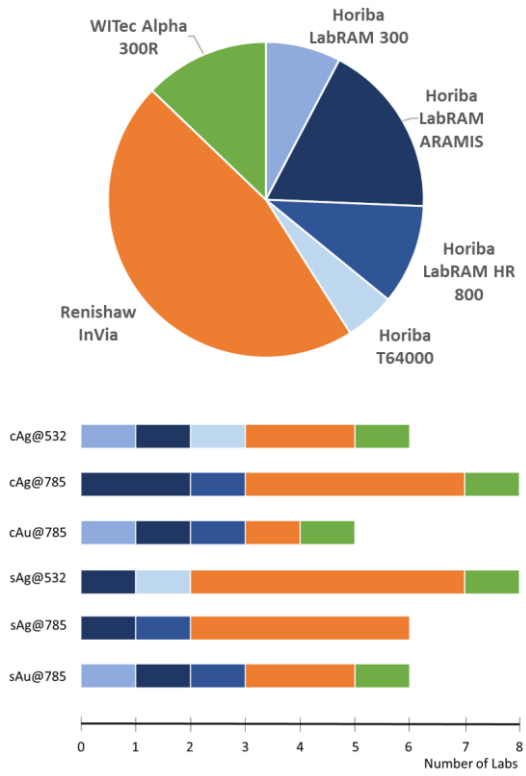


Figure 3. Instruments used in the study.

DATA ANALYSIS

All statistical analysis and strategic decisions related to data analysis were entirely the responsibility of the OL; this was decided by all participants to be the most objective way of assessing reproducibility in the different laboratories. Calculations were performed within the R software environment (version 3.4.3) for statistical computing and graphics, building on the packages *hyperSpec*,[37] and *dplyr*,[38] on a commercially available workstation (Intel Core i7–4770, 4 cores 3.40 GHz, 32 GB DDR3-RAM). In-house developed R scripts were used for visualization and further processing.

Once all raw spectral data were collected, they were processed according to the following multi-step procedure. Raw data were first inspected for data integrity (e.g. missing data, spectral artefacts), and then grouped and processed by single experiment. The second step involved the pre-processing of the spectral data (e.g., smoothing, downsampling, selection of spectral range, baseline correction and normalization), and the building of a calibration curve by means of an inverse least squares regression. The Supporting Information (Section S2, "Data Analysis Protocol") lists all relevant setting parameters necessary to reproduce the analysis in a software-independent manner, and provides detailed information describing the protocol used in this step. Briefly, the inverse calibration approach was used under the assumption that the uncertainty coming from the preparations of spiked standards was negligible if compared with the random variability of each instrument. The calibration function was obtained using the calibration standards, relating the response (integrated area of SERS intensity between 715 and 750 cm^{-1}) and the concentration (Figure 2).

Visual inspection of the calibration and residual plots, together with examination of the regression statistics obtained from each calibration curve (r^2, F-test), was used as a system suitability check to make an overall assessment of the reliability of the data. If the fitness-for-purpose of the curve was judged to be satisfactory, each calibration model was further validated by plotting the predicted values versus the nominal concentrations of the test samples. A good calibration leads to observations falling close to a 45-degree straight line ($y = x$ equality line). The results were visually examined by looking at the dispersion at each concentration level in the predicted versus reference values plot. These profiles are the visual decision tools that allowed us to evaluate the presence of possible different levels of precision for the considered methods, and recognize regions with different levels of prediction accuracy between different laboratories.

To characterize and compare the performance of different SERS methods, during the next step of analysis, the residuals generated from multi laboratory predictions were grouped by method, and a set of performance measures, expressed as specific FoMs, was computed by taking into account the relationships between accuracy, trueness and reproducibility (see Table 1). According to the ISO definition (ISO 5725),[39] the overall accuracy of a method, is considered as a global entity with two components, trueness and precision, representing the systematic and the random components of the total error, respectively.[40] In the case of ILS, the precision is more appropriately expressed as reproducibility, defined as the difference between repeated measurements when between laboratories variations are included in each replicate. In this study, the root mean squared error of prediction (RMSEP) takes account of simultaneous combination of the random and the systematic parts of the error that occur between the different laboratories, including errors from sample preparation, measurement, and calibration model. The basic model is:

$$RMSEP^2 \cong SEP^2 + BIAS^2$$

where the RMSEP is the estimate of the accuracy, the standard error of performance (SEP) accounts for the interlaboratory reproducibility (residual variance) and BIAS is the estimate of the trueness.[41] It is worth noting that, when one compares RMSEP, BIAS and SEP obtained from different analyte ranges, re-scaling can be advantageous,

for example if the considered analytical protocols require different dilution steps. Here, we normalized RMSEP, BIAS and SEP values using the range of the concentration values of the calibration references.

Table 1. A summary of the Figures of Merit used for method characterization

Characteristic	Description	FoM	Interpretation
Accuracy	Closeness of agreement between measurement results and the accepted reference values.	$RMSEP = \sqrt{\sum_{i=1}^{N} \frac{(\hat{x}_i - x_i)^2}{N}}$	Total prediction error
Trueness	Difference between the expected measurement results and the accepted reference values	$BIAS = \sum_{i=1}^{N} \frac{\hat{x}_i - x_i}{N}$	Average of the residuals; Systematic component of the total error (i.e. constant offset)
Precision	Closeness of agreement between independent measurement results obtained with the same method in different measurement facilities with different operators using different equipment	$\mathrm{SEP}=\sqrt{\frac{\sum_{i=1}^{n}(\hat{x}_i - \mathrm{BIAS} - x_i)^2}{N-1}}$	Width of residuals distribution; Random component of the total error

$\hat{x}_i$ and x_i are the predicted and assigned reference values, respectively, for the test sample *i*, and *N* is the number of samples in the TEST set.

*Accuracy = reproducibility + trueness ($RMSEP^2 \cong SEP^2 + BIAS^2$), as from ISO 5725 (ref. 39).

RESULTS AND DISCUSSION

The first notable achievement of the current study was the high number of participants throughout the European SERS community who accepted to share their expertise (creation and adoption of the consensual SOP) and experimental data (the first study to compare the reproducibility of different SERS methods). Fifteen laboratories from 11 European countries participated in the study, using 6 different models of Raman instruments from 3 manufacturers. Detailed technical information on each instrumental configuration that may lead to identification of the participating laboratories will not be disclosed. An aggregated summary is presented in Table 2.

Table 2. ILS in numbers.

Laboratories involved:	15
Researchers involved:	44
European countries involved:	11
SERS protocols tested:	6
Single SERS substrates used:	1080
Metal colloids used:	488 mL
Spectra delivered:	3694
Spectra analyzed:	3516

Of the 48 expected datasets, 41 (i.e. 85.4%) were uploaded by the participating laboratories, a total of 3694 spectra. The failure to upload data in the specified time window by 3 participants was mainly due to problems with the instrumentation or misjudgment of the time and effort required for carrying out the experiments. After data inspection, cAg@532 and cAu@785 data from laboratory P10 were rejected, because spectral features of parafilm, used as a hydrophobic substrate for liquid samples prepared from colloidal substrates (see SOP in Supplementary Materials, section S1),), were found to strongly interfere with the adenine bands, impeding any data analysis. This problem may occur in case of an incorrect focusing on the sample, and could have been prevented by introducing a warning in the SOP, or by using more expensive substrates such as CaF_2 slides, which were not available to all participants. After this first step, 39 datasets (3516 spectra), corresponding to 81.3% of the datasets planned, were further pre-processed (see Supplementary Figure S2) and analyzed.

After pre-processing, the specific ring-breathing mode of adenine observed in the range between 715 and 750 cm^{-1} appeared well resolved and nearly superimposable for spectra of samples of the same concentration. A specific example (method cAu@785) was selected and is shown in Figure 4 for illustration. The slight deviations among the spectra on the left side of Figure 4 can be attributed to variations in the laser power density, optics and detectors responsivities, as produced by different instrumental set-ups used for the measurements. In fact, the peculiarities of each set-up prevented the use of perfectly homogeneous measurement settings, and only a set of thresholds for maximum laser power densities have been suggested (see section S1 of Supplementary Materials). The pre-processing used for data analysis, which prescribed an extended multiplicative scatter correction (EMSC) of the spectra (see section S2 of Supplementary Materials, DAP), made possible the comparison between different datasets over the same concentration levels (Figure 4, on the right).

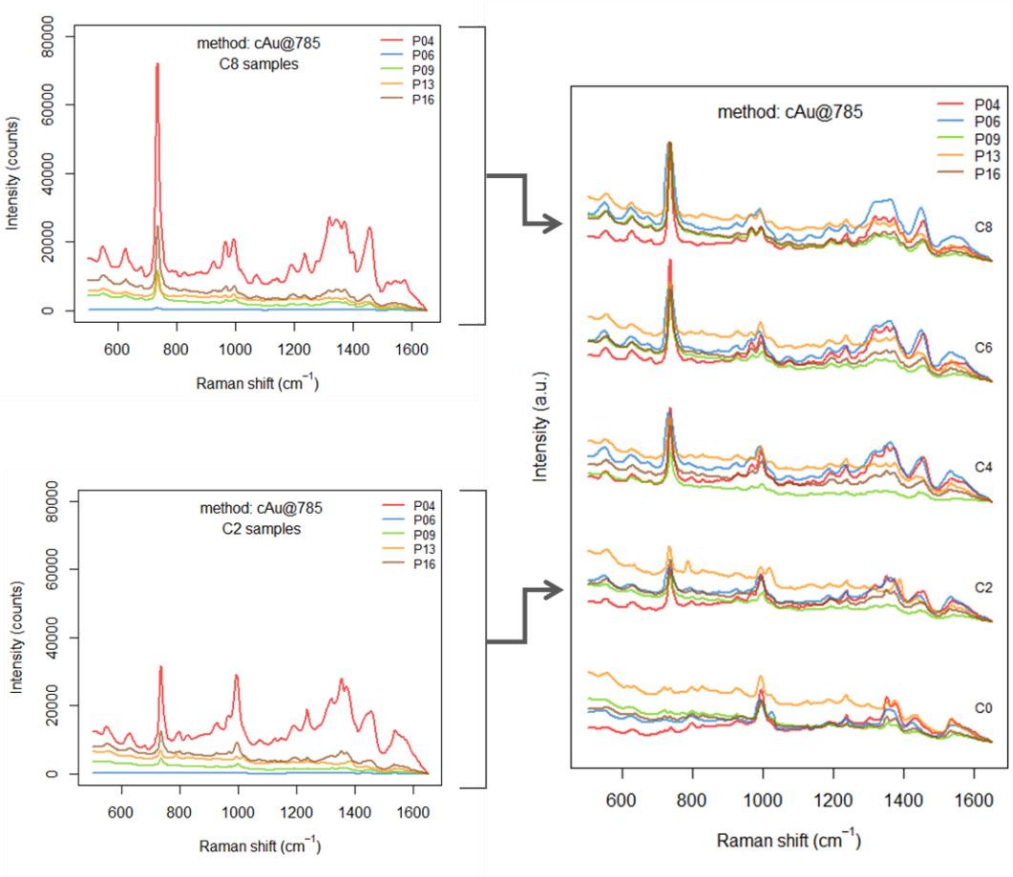


Figure 4. Comparison of spectra collected by different partecipants (P04, P06, P09, P13, and P16) at 5 concentration levels (C0, C2, C4, C6, C8) for cAu@785 method (see Supplementary Information, Section S4, Table S1 for the actual concentrations and more details)On the left, spectra are shown before pre-processing. On the right, spectra are shown after preprocessing, offset for clarity.

The quantification of the SERS vibrational signature of adenine has been conducted by integrating the area for the region 715-750 cm^{-1} in the spectra from each training set, and using it for the construction of the pertinent linear regression models by inverse least square regression. Performance and uncertainty of 39 calibration curves were analyzed for all laboratories. A simple linear model (straight line) fitted on the training data was selected because it is the most commonly accepted (and the most widely used) for other physico-chemical analytical methods.[42] The main requirement for this kind of model to be valid is that the computed values be sufficiently free of random errors to obtain a relationship able to give results that are proportional to the analyte concentration within a given range. Four calibration curves (cAg@532, cAg@785, and cAu@785 from laboratory P14, sAu@785 from laboratory P06) were rejected because of very low quality of the linear fit (r^2 lower than 0.6, p-value for the F-test higher than 0.01 (See Supplementary Figure S3). The final dataset was then composed of 35 curves, with different number of laboratories for each method. Figure 5 shows the flowchart of the data evaluation strategy and the results obtained after the selection process. The entire dataset can be found as Supplementary Material, section S4 "Tables", sorted by methods and laboratory. Experimental data were deposited in the Zenodo data repository (https://zenodo.org/) and are publicly available (Access Number: 3572359).

Although ILS are usually designed to characterize just one method, we wanted to compare the reproducibility of more methods, each involving the use of the most commonly employed substrate types (i.e. colloidal and solid) and laser wavelengths (i.e. visible and near-infrared). Considering the number of participants and methods, the definition of an acceptable workload for each lab led to a design in which each method was going to be tested by a different group of laboratories. Collaborative trials are also very time-consuming (guidelines suggest at least 8 valid results from different laboratories for each method). Although the study has been planned to achieve such numbers (i.e. 8 labs for each method), we eventually managed to acquire and use only 35 datasets, of the 48 expected (i.e. 72.9%), leading to less than 8 datasets for each method (Figure 5) and this may limiting the strength of the obtained results. Preliminary trials with fewer laboratories, however, are still considered useful and suggested by many international guidelines,[43] especially for setting up a clearly written SOP that includes system suitability checks to be evaluated for errors and ambiguities before the actual collaborative trial starts. Thus, the results obtained in this

study are to be considered as somewhat preliminary, and will hopefully encourage other collaborative trials on SERS.

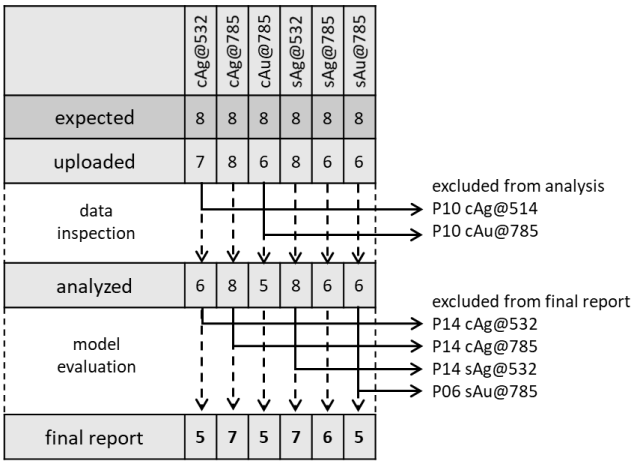


Figure 5. Data selection workflow.

Single method characterization. The goal of calibration models is to predict the analyte concentration in an unknown (chemical) sample from instrument responses. The results from the predictions of the 35 selected calibration curves are summarized in Figure 6, presented as method-wise reference versus predicted value plots. Five adenine concentrations, over four different ranges, are covered. A visual inspection of this plot qualitatively describes the accuracy of each method. Although it appears that there is room for improvement, the predicted values were, in most cases, consistent. At least one laboratory for five of the considered methods obtained excellent results. The best performances were achieved for both colloids and solid substrates, within the cAg@785 (P01) and sAg@785 (P04) methods.

The spread in the adenine predictions at each concentration level represents the variability among the laboratories. In three cases (cAg@532, cAu@785, sAg@785) the predictions show a certain level of homoscedasticity or, in other words, the prediction error seemed independent of the concentration. In the other three cases (cAg@785, sAg@532, sAu@785), the spread increased as the concentration increased. More insights can be obtained from Figure S10 of the Supplementary Material: in the three methods with a common internal variance, the median prediction error at each concentration levels lie on the diagonal. Interestingly, many perfectly good curves in the cAg@785 method exhibit symptoms of slightly non-linearity (P06, P08, P13, P18), with the errors at the right end tending to curve away from the equality line. It should be noted that the instrumental differences were not influential at all, since they were performed with different instruments (cf. Figure 3).

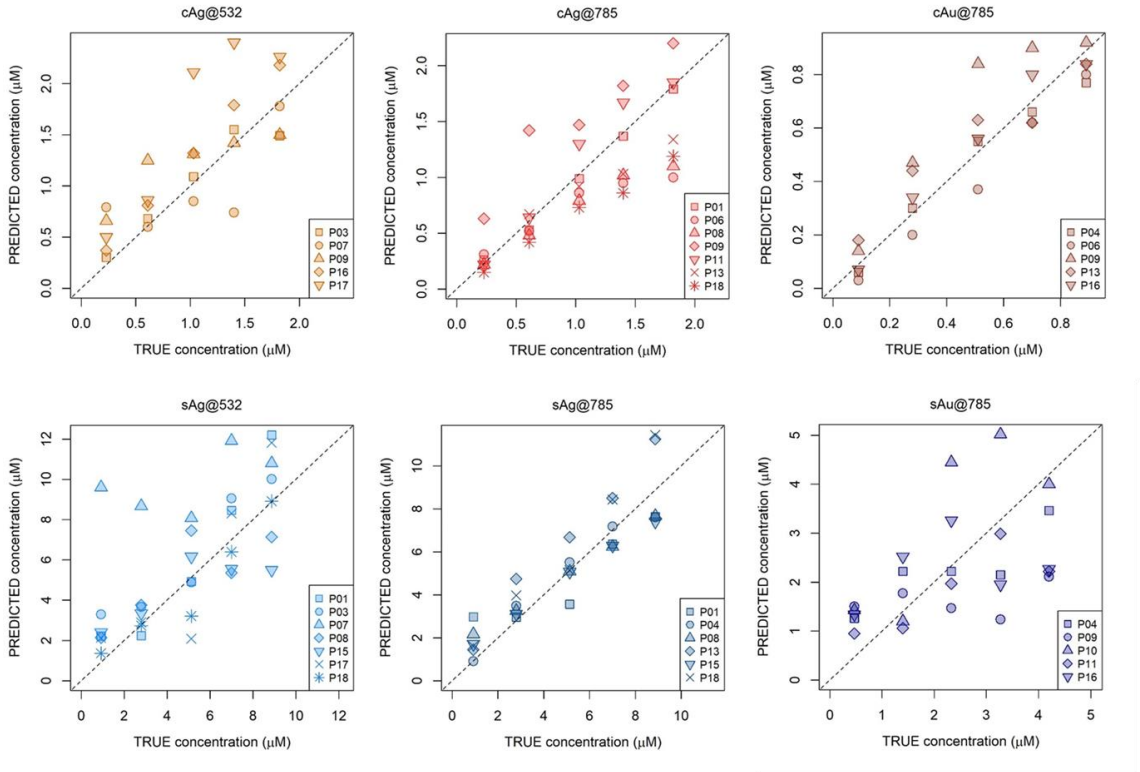


Figure 6. Plot of estimated values against the actual values for the test set samples for different methods by different participants (P01-P18); the dotted line is the line of equality. For a complete representation of the uncertainty of each individual prediction see Figures S4-9 in supplementary material.

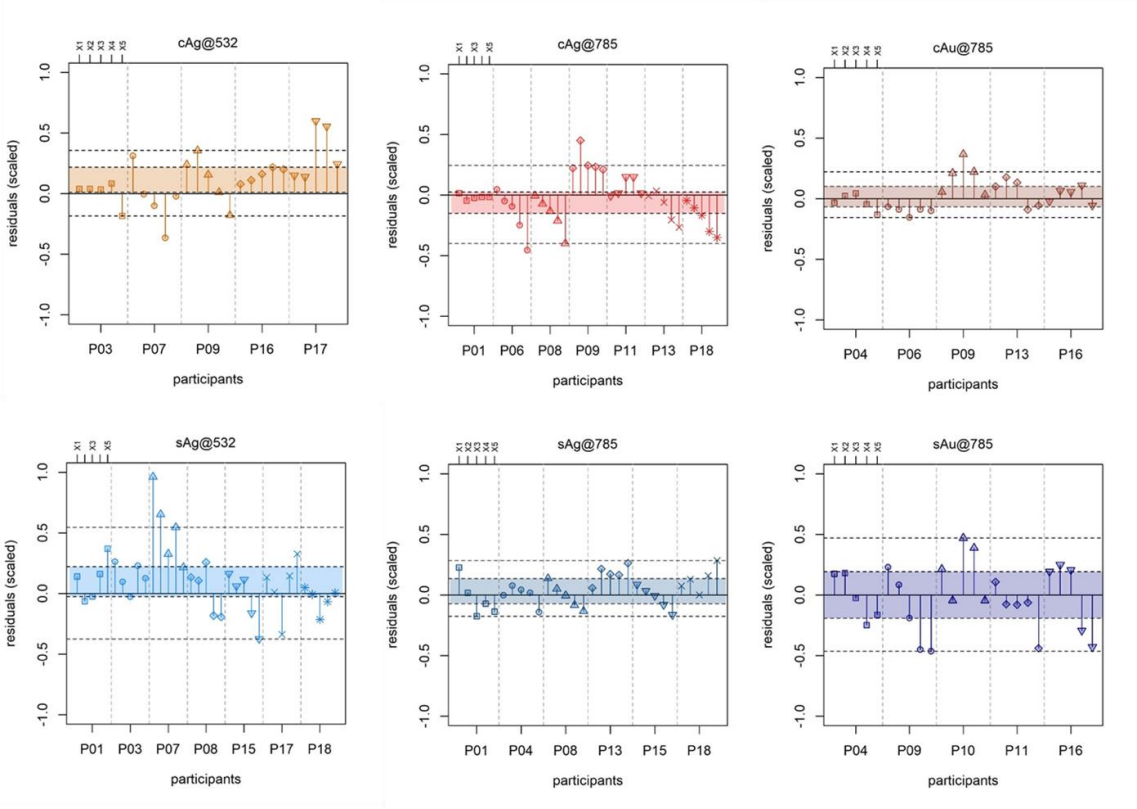


Figure 7. Residual plots for the TEST sets of the six SERS methods. Concentration levels of the TEST sets (X1-X5, for actual concentrations see Supplementary Information, Table S1) are aggregated according to each participant (P01-P18). The limits of the colored areas are the upper and lower quartiles, so each area spans the interquartile range (IQR) for each method; The two dashed lines outside the colored areas range to the extreme data point (< 1.5 × IQR). Residuals values were rescaled using the range of the reference values in the test set data to enable comparison between calibrations obtained with different ranges.

The residual plots in Figure 7 summarize the results from another perspective by focusing on the normalized prediction errors (i.e. residuals): each group represents a different laboratory, whereas the points within each group represent the five concentration levels for the test sets (X1-X5). In addition, colored areas are shown, corresponding to the interquartile range (IQR) and two control lines (< 1.5 × IQR), calculated over the entire set of residuals for each method. These plots are a visual decision tool that allows the evaluation of the discrepancies between different laboratories without the need for tests of significance. The larger the IQR, the bigger is the data deviation, indicating the poorer the performance. Moreover, severe outliers, are immediately identified. However, the outliers were not excluded from the FoMs calculation because such rejection would have only artificially improved the appearance of the data, but do nothing in terms of avoiding future instances of outlying results.

Comparison among methods. To gain more insight into the consistency of quantitative results obtained from different SERS methods, we computed a set of performance measures, expressed as specific FoMs calculated from the residuals from the validation samples. A summary is provided in Table 3.

Table 3. Figures of merit for different SERS methods.

Method	*N*	RMSEP	SEP	BIAS
cAg@532	25	24 %	21 %	11 %
cAg@785	35	19 %	19 %	-4 %
cAu@785	25	13 %	13 %	3 %
sAg@532	35	29 %	27 %	11 %
sAg@785	30	13 %	12 %	4 %
sAu@785	25	28 %	29 %	-2 %

Normalization (as range scaling) was carried out to compare different methods; non-normalized RMSEP, BIAS and SEP values are available in Supplementary Table S8 (see Methods for details). *N* is the number of residuals for each method.

The general behavior of the residuals for each SERS method is depicted in Figure 8. For all the considered methods, the residuals have a distribution that looks roughly normal in shape and centered close to zero. A failure to center on zero is described as a bias, and the size of the mean error is the BIAS value, calculated as the average difference between predicted and reference samples in the validation set. The width of the distribution is described by the SEP value, calculated as square root of the quadratic sum of the error of the predicted *versus* the reference value,

once the predicted value has been corrected for bias (Table 1).

A high BIAS means a low trueness of the methods. The main concern here is with the possible importance of the calculated BIAS, since it is the nature of spectroscopic data to present some variation occurring between measurements due to randomly distributed noise. A Student's *t* test at the 99% confidence level verified whether these deviations could be considered random noise, as usual and expected, or whether the deviations were larger than expected by random chance alone. Only two methods (both employing the 532 nm laser source, cAg@532, and sAg@532) demonstrated significant BIAS in the selected range.

In terms of reproducibility, two methods, i.e. sAg@785 and cAu@785 appeared as the most reliable, with narrower distribution of residuals and SEP values of 12% and 13%, respectively.

Interestingly, the inverse of the (normalized) SEP is often used as a quality threshold for model performance.[44,45] For 1/SEP > 4, the calibration is considered acceptable for sample screening; for 1/SEP > 10, the calibration is considered acceptable for quality control; and for 1/SEP > 15, the calibration is good for quantification. In this context, all methods could be considered acceptable for screening purposes, with the exception of sAg@532 and sAu@785 (1/SEP = 3.72 and 3.49, respectively). Considering the fact that this is the first collaborative trial on quantitative SERS conducted on a range of different instrumental set-ups, the fact that two methods, one using a colloidal substrate (i.e. cAu@785) and one using a solid substrate (i.e. sAg@785) achieved a 1/SEP value close to the limit set for quality control (i.e. 1/SEP > 10) is significant.

Although the primary focus of this study is on reproducibility, the RMSEP, used as a practical measure of accuracy, enabled us to characterize a SERS method by a single acceptability criterion, consistent with the ISO definition. The lower the RMSEP, the better the method is. The RMSEP typically express how well a calibration, on average, will predict new samples. This study, however, was conducted to compare the performance of a whole analytical process, from sample preparation to the stability of the instrument, for which the calibration is just a part. As expected from the low values of BIAS, the overall accuracy as indicated by the RMSEP is reflecting the SEP values, indicating two methods (i.e. cAu@785 and sAg@785) as the most accurate.

It must be noted that the RMSEP of a method, as calculated in this study, includes all the uncertainty contributions from different laboratories, depending essentially on the design of the experiments presented in the SOP. Since the RMSEP is calculated from all test samples, thus averaging over different laboratories, it does not directly provide an uncertainty for future measurements by a single lab, but it has been used here exclusively to compare the accuracy of different methods.

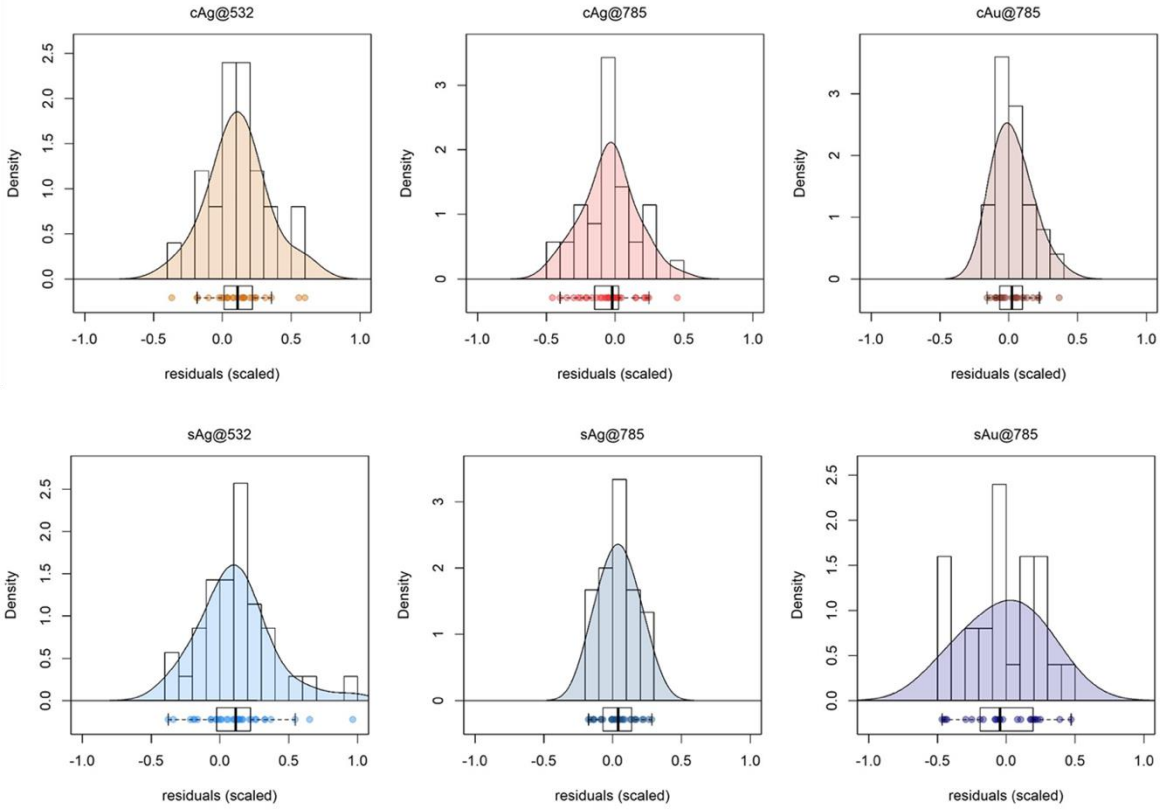


Figure 8. Probability density function (PDF) and boxplots of the normalized residuals for the six SERS methods.

Limitations and possible improvements. As with all first steps, it is still imperfect, amendable, and somewhat limited in scope. The overall results in themselves are to be strictly considered as limited to prescribed methods using specific substrates to quantify adenine, and should not be extended to SERS in general or to other methods or analytes. The aim was not even to quantify adenine (which was chosen out of necessity, having desirable characteristics of stability, non-toxicity, etc. see Methods section), but to assess how different results were obtained among different laboratories, and to be able to compare the performance of different methods. In a way, this study is more about methodology than performance. Moreover, it had the merit of having fostered active collaboration among tens of spectroscopists all over Europe, in an effort to reach a consensus on how to evaluate SERS experiments performed by different laboratories. Since SERS is increasingly being used by individual laboratories for quantitative applications, this issue is clearly extremely relevant to the SERS community, as proved by a very recent review written by a panel of international researchers that addressed this issue by proposing some recommendations, in terms of "good analytical practice", to increase the comparability of quantitative SERS results obtained by different laboratories.[6] In spite of these positive findings, even the SERS methods tested with the lowest SEP do not yet satisfy the strict reproducibility requirements for a quantitative analytical method (1/SEP > 15). However, there is space for improvement: the use of internal standards, whether these are isotopologues or through use of standard addition (as reviewed in [7]), could help to decrease the intrinsic variability due to the enhancement substrates. Although not within the goals of this study, it may be of interest to devote future effort to understanding the mechanistic rationale underlying the differing responses of the different substrates. For solid substrates, a larger dataset including maps instead of single measurements could also improve the data. The use of non-linear models could take into account deviations from linearity which appeared in many calibration datasets, thus improving the predictions.

CONCLUSIONS

In this first SERS study involving several laboratories, we tried to define a methodology to assess the reproducibility and trueness of a quantitative SERS method, and to compare different methods. In our opinion, this is a first important step toward a "standardization" process of SERS protocols, not proposed by a single laboratory but by a larger community. This study addressed two questions: can a quantitative SERS method consistently be used by different labs? And if different SERS methods are used to quantify the same analyte, which is the best way to compare them? Based on the results obtained, we suggest that indeed a SERS method can be consistently used by different labs, provided that the method is very well defined (with a detailed SOP that all participants agree to follow). The methods tested provided varying results in terms of reproducibility, but the best ones proved to be reasonably reproducible, with average SEP as low as 12% and 13%, which is promising considering the fact that different instruments were used over a wide time frame, with different set-up and acquisition parameters (laser power, acquisition time, etc.). These results are valid within the framework of the system we proposed to use to compare different methods, considering RMSEP, SEP and BIAS values in Table 3. Using these tools, one can effectively compare different SERS methods to assess which one is the more reproducible and accurate. The present study is a starting point, and should ideally stimulate other groups of SERS researchers to set up similar studies for other analytes, substrates and methods. The next step with respect to the present kind of study should be the evaluation of each source of experimental uncertainty (e.g. substrates, instruments, and operators) for the best performing methods, as already suggested for qualitative SERS methods.[11] Future SERS studies should possibly focus their effort on a single method, rather than many, to reach more easily a significant number of laboratories.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.

- Supplementary Experimental Section (S1 "Standard Operating Procedure"; S2 "Data Analysis Protocol"; S3 "Supplementary Figures"; S4 "Supplementary Tables")
- The dataset is available at DOI: 10.5281/zenodo.3572359

AUTHOR INFORMATION

Corresponding Author

* E-mail: abonifacio@units.it

Author Contributions

S. Fornasaro and A. Bonifacio led the ring trial and developed the experimental design, the SOP and the DAP and lead the discussion of all authors to finalize these parts of the study. C. Beleites contributed to the experimental design and to parts of the SOP and of the DAP. T. Rindzevicius and M.S. Schmidt prepared and made available all the solid SERS substrates used in the study, M. Meneghetti and L. Litti prepared and made available all the colloidal SERS substrates used in the study. S. Fornasaro and E. Mitri prepared the training and test samples, shipped samples and substrates and received the submitted spectra. All other authors including M. Meneghetti, L. Litti, S. Fornasaro and E. Mitri performed SERS measurements. S. Fornasaro and A. Bonifacio performed the centralized data analysis, wrote the DAP and wrote the initial version of the manuscript, which was then amended according to the remarks, suggestions and comments by all the other co-authors.

ACKNOWLEDGMENT

This article is based upon work from COST Action Raman4Clinics – Raman-based applications for clinical diagnostics (BM1401), supported by COST (European Cooperation in Science and Technology) – www.cost.eu. The authors would like to thank the COST Action Grant Holder, the Leibniz Institute of Photonic Technology, and in particular Prof. Jürgen

Popp (as the Grant Holder Scientific Representative) and Gabriele Hamm for the help and support in the organization of the Raman4Clinics events related to the present study. S.F. acknowledges support from IRCCS Burlo Garofolo, Trieste. L.A.B.C. and M.P.M.M. acknowledge Portuguese Foundation for Science and Technology (Project UID/Multi/00070/2019). C.N, A.C., and F.G. acknowledge the Piedmont Region- Italy in the frame of the projects POR FESR "DEFLECT" (ID nr 10808-178675264) and "Food Drug Free" (ID nr 12866-2074-64512). T.R. acknowledges that his work was financially supported by the Danish National Research Foundation (Project DNRF122) and Villum Fonden (Grant No. 9301).

REFERENCES

(1) Langer, J.; Jimenez de Aberasturi, D.; Aizpurua, J.; Alvarez-Puebla, R. A.; Auguie, B.; Baumberg, J. J.; Bazan, G. C.; Bell, S. E. J.; Boisen, A.; Brolo, A. G.; Choo, J.; Cialla-May, D.; Deckert, V.; Fabris, L.; Faulds, K.; Garcia de Abajo, F. J.; Goodacre, R.; Graham, D.; Haes, A. J.; Haynes, C. L.; Huck, C.; Itoh, T.; Kall, M.; Kneipp, J.; Kotov, N. A.; Kuang, H.; Le Ru, E. C.; Lee, H. K.; Li, J. F.; Ling, X. Y.; Maier, S.; Mayerhoefer, T.; Moskovits, M.; Murakoshi, K.; Nam, J. M.; Nie, S.; Ozaki, Y.; Pastoriza-Santos, I.; Perez-Juste, J.; Popp, J.; Pucci, A.; Reich, S.; Ren, B.; Schatz, G. C.; Shegai, T.; Schlucker, S.; Li-Lin, T.; Thomas, K. G.; Tian, Z. Q.; Van Duyne, R. P.; Vo-Dinh, T.; Wang, Y.; Willets, K. A.; Xu, C.; Xu, H.; Xu, Y.; Yamamoto, Y. S.; Zhao, B.; Liz-Marzan, L. M. Present and Future of Surface Enhanced Raman Scattering, *ACS Nano* **2019**.

(2) Graham, D.; Goodacre, R.; Arnolds, H.; Masson, J. F.; Schatz, G.; Baumberg, J.; Kim, D. H.; Aizpurua, J.; Lum, W.; Silvestri, A.; de Nijs, B.; Xu, Y.; Di Martino, G.; Natan, M.; Schlucker, S.; Wuytens, P.; Bruzas, I.; Kuttner, C.; Hardy, M.; Chikkaraddy, R.; Martin Sabanes, N.; Delfino, I.; Dawson, P.; Gawinkowski, S.; Bontempi, N.; Mahajan, S.; Reich, S.; Hourahine, B.; Bell, S.; Krolikowska, A.; Porter, M.; Keeler, A.; Kamp, M.; Fountain, A.; Fasolato, C.; Giorgis, F.; Otero, J. C.; Matricardi, C.; Van Duyne, R.; Lombardi, J.; Deckert, V.; Velleman, L. Theory of SERS enhancement: general discussion, *Faraday Discuss* **2017**, *205*, 173-211.

(3) Sackmann, M.; Materny, A. Surface enhanced Raman scattering (SERS)—a quantitative analytical tool?, *Journal of Raman Spectroscopy* **2006**, *37*, 305-310.

(4) Bell, S. E. J.; Stewart, A. In *Surface Enhanced Raman Spectroscopy*, Schlücker, S., Ed.; Wiley-VCH Verlag GmbH & Co. KGaA, 2010, pp 71-86.

(5) Aitchison, H.; Aizpurua, J.; Arnolds, H.; Baumberg, J.; Bell, S.; Bonifacio, A.; Chikkaraddy, R.; Dawson, P.; de Nijs, B.; Deckert, V.; Delfino, I.; Di Martino, G.; Eremina, O.; Faulds, K.; Fountain, A.; Gawinkowski, S.; Gomez Castano, M.; Goodacre, R.; Gracie, J.; Graham, D.; Guicheteau, J.; Hardwick, L.; Hardy, M.; Heck, C.; Jamieson, L.; Kamp, M.; Keeler, A.; Kuttner, C.; Langer, J.; Mahajan, S.; Martin Sabanes, N.; Murakoshi, K.; Porter, M.; Schatz, G.; Schlucker, S.; Tian, Z.; Tripathi, A.; Van Duyne, R.; Vikesland, P. Analytical SERS: general discussion, *Faraday Discuss* **2017**, *205*, 561-600.

(6) Schlucker, S.; Bell, S.; Charron, G.; Cortes, E.; Kneipp, J.; Lamy de la Chapelle, M.; Langer, J.; Prochazka, M.; Tran, V. Towards Reliable and Quantitative SERS: from Key Parameters to Good Analytical Practice, *Angewandte Chemie* **2019**.

(7) Goodacre, R.; Graham, D.; Faulds, K. Recent developments in quantitative SERS: Moving towards absolute quantification, *TrAC Trends in Analytical Chemistry* **2018**, *102*, 359-368.

(8) Novara, C.; Dalla Marta, S.; Virga, A.; Lamberti, A.; Angelini, A.; Chiadò, A.; Rivolo, P.; Geobaldo, F.; Sergo, V.; Bonifacio, A.; Giorgis, F. SERS-Active Ag Nanoparticles on Porous Silicon and PDMS Substrates: A Comparative Study of Uniformity and Raman Efficiency, *J Phys Chem C* **2016**, *120*, 16946-16953.

(9) Fisk, H.; Westley, C.; Turner, N. J.; Goodacre, R. Achieving optimal SERS through enhanced experimental design, *Journal of Raman Spectroscopy* **2016**, *47*, 59-66.

(10) Guicheteau, J. A.; Farrell, M. E.; Christesen, S. D.; Fountain, A. W.; Pellegrino, P. M.; Emmons, E. D.; Tripathi, A.; Wilcox, P.; Emge, D. Surface-Enhanced Raman Scattering (SERS) Evaluation Protocol for Nanometallic Surfaces, *Applied Spectroscopy* **2013**, *67*, 396-403.

(11) Muehlethaler, C.; Leona, M.; Lombardi, J. R. Towards a validation of surface-enhanced Raman scattering (SERS) for use in forensic science: repeatability and reproducibility experiments, *Forensic Science International* **2016**, *268*, 1-13.

(12) Novara, C.; Chiadò, A.; Paccotti, N.; Catuogno, S.; Esposito, C. L.; Condorelli, G.; De Franciscis, V.; Geobaldo, F.; Rivolo, P.; Giorgis, F. SERS-active metal-dielectric nanostructures integrated in microfluidic devices for label-free quantitative detection of miRNA, *Faraday Discussions* **2017**, *205*, 271-289.

(13) Yuen, C.; Zheng, W.; Huang, Z. Low-level detection of anticancer drug in blood plasma using microwave-treated gold-polystyrene beads as surface-enhanced Raman scattering substrates, *Biosensors and Bioelectronics* **2010**, *26*, 580-584.

(14) Andreou, C.; Hoonejani, M. R.; Barmi, M. R.; Moskovits, M.; Meinhart, C. D. Rapid Detection of Drugs of Abuse in Saliva Using Surface Enhanced Raman Spectroscopy and Microfluidics, *ACS Nano* **2013**, *7*, 7157-7164.

(15) Fornasaro, S.; Bonifacio, A.; Marangon, E.; Buzzo, M.; Toffoli, G.; Rindzevcius, T.; Schmidt, M. S.; Sergo, V. Label-Free Quantification of Anticancer Drug Imatinib in Human Plasma with Surface Enhanced Raman Spectroscopy, *Analytical Chemistry* **2018**.

(16) Fornasaro, S.; Marta, S. D.; Rabusin, M.; Bonifacio, A.; Sergo, V. Toward SERS-based point-of-care approaches for therapeutic drug monitoring: the case of methotrexate, *Faraday Discussions* **2016**.

(17) Panikar, S. S.; Ramírez-García, G.; Sidhik, S.; Lopez-Luke, T.; Rodriguez-Gonzalez, C.; Ciapara, I. H.; Castillo, P. S.; Camacho-Villegas, T.; Rosa, E. D. l. Ultrasensitive SERS Substrate for Label-Free Therapeutic-Drug Monitoring of Paclitaxel and Cyclophosphamide in Blood Serum, *Analytical Chemistry* **2018**, *91*, 2100-2111.

(18) Turzhitsky, V.; Zhang, L.; Horowitz, G. L.; Vitkin, E.; Khan, U.; Zakharov, Y.; Qiu, L.; Itzkan, I.; Perelman, L. T. Picoanalysis of Drugs in Biofluids with Quantitative Label-Free Surface-Enhanced Raman Spectroscopy, *Small* **2018**, *14*, 1802392.

(19) Litti, L.; Ramundo, A.; Biscaglia, F.; Toffoli, G.; Gobbo, M.; Meneghetti, M. A surface enhanced Raman scattering based colloid nanosensor for developing therapeutic drug monitoring, *J Colloid Interface Sci* **2019**, *533*, 621-626.

(20) Muhamadali, H.; Watt, A.; Xu, Y.; Chisanga, M.; Subaihi, A.; Jones, C.; Ellis, D. I.; Sutcliffe, O. B.; Goodacre, R. Rapid Detection and Quantification of Novel Psychoactive Substances (NPS) Using Raman Spectroscopy and Surface-Enhanced Raman Scattering, *Front Chem* **2019**, *7*, 412.

(21) Shende, C.; Farquharson, A.; Brouillette, C.; Smith, W.; Farquharson, S. Quantitative Measurements of Codeine and Fentanyl on a Surface-Enhanced Raman-Active Pad Test, *Molecules* **2019**, *24*.

(22) Sivashanmugan, K.; Squire, K.; Tan, A.; Zhao, Y.; Kraai, J. A.; Rorrer, G. L.; Wang, A. X. Trace Detection of Tetrahydrocannabinol in Body Fluid via Surface-Enhanced Raman Scattering and Principal Component Analysis, *ACS Sens* **2019**, *4*, 1109-1117.

(23) Lee, K.-M.; Herrman, T. J.; Bisrat, Y.; Murray, S. C. Feasibility of Surface-Enhanced Raman Spectroscopy for Rapid Detection of Aflatoxins in Maize, *J Agr Food Chem* **2014**, *62*, 4466-4474.

(24) Yaseen, T.; Pu, H.; Sun, D.-W. Fabrication of Silver-coated Gold Nanoparticles to Simultaneously Detect Multi-class Insecticide Residues in Peach with SERS Technique, *Talanta* **2018**.

(25) Hassan, M. M.; Li, H.; Ahmad, W.; Zareef, M.; Wang, J.; Xie, S.; Wang, P.; Ouyang, Q.; Wang, S.; Chen, Q. Au@Ag nanostructure based SERS substrate for simultaneous determination of pesticides residue in tea via solid phase extraction coupled multivariate calibration, *LWT* **2019**, *105*, 290-297.

(26) Jaworska, A.; Fornasaro, S.; Sergo, V.; Bonifacio, A. Potential of Surface Enhanced Raman Spectroscopy (SERS) in Therapeutic Drug Monitoring (TDM). A Critical Review, *Biosensors (Basel)* **2016**, *6*.

(27) Ricci, M.; Trombetta, E.; Castellucci, E.; Becucci, M. On the SERS quantitative determination of organic dyes, *Journal of Raman Spectroscopy* **2018**, *49*, 997-1005.

(28) McLaughlin, C.; MacMillan, D.; McCardle, C.; Smith, W. E. Quantitative analysis of mitoxantrone by surface-enhanced resonance Raman scattering, *Anal Chem* **2002**, *74*, 3160-3167.

(29) Ellison, S. L. R.; Williams, A. *Eurachem/CITAC guide: Quantifying Uncertainty in Analytical Measurement*, Third ed.; Eurachem, 2012.

(30) Heyden, Y. V.; Smeyers-Verbeke, J. Set-up and evaluation of interlaboratory studies, *Journal of Chromatography A* **2007**, *1158*, 158-167.

(31) Hund, E.; Massart, D. L.; Smeyers-Verbeke, J. Inter-laboratory studies in analytical chemistry, *Analytica Chimica Acta* **2000**, *423*, 145-165.

(32) Szewczak, E.; Bondarzewski, A. Is the assessment of interlaboratory comparison results for a small number of tests and limited number of participants reliable and rational?, *Accreditation and Quality Assurance* **2016**, *21*, 91-100.

(33) Krafft, C.; Popp, J. Raman4Clinics: the prospects of Raman-based methods for clinical application, *Anal Bioanal Chem* **2015**, *407*, 8263-8264.

(34) Alvarez-Puebla, R. A. Effects of the Excitation Wavelength on the SERS Spectrum, *The journal of physical chemistry letters* **2012**, *3*, 857-866.

(35) Amendola, V.; Meneghetti, M. Controlled size manipulation of free gold nanoparticles by laser irradiation and their facile bioconjugation, *Journal of Materials Chemistry* **2007**, *17*, 4705.

(36) Schmidt, M. S.; Hübner, J.; Boisen, A. Large Area Fabrication of Leaning Silicon Nanopillars for Surface Enhanced Raman Spectroscopy, *Advanced Materials* **2012**, *24*, OP11-OP18.

(37) Beleites, C.; Sergo, V. hyperSpec: a package to handle hyperspectral data sets in R, *R package version 0.99-20180627* **2018**.

(38) Wickham, H.; François, R.; Henry, L.; Mueller, K. dplyr: A Grammar of Data Manipulation, *R package version 0.8.3* **2019**.

(39) ISO 5725-1:1994. Accuracy (trueness and precision) of measurement methods and results — Part 1: General principles and definitions, **1994**.

(40) Westgard, J. O.; Carey, R. N.; Wold, S. Criteria for judging precision and accuracy in method development and evaluation, *Clin Chem* **1974**, *20*, 825-833.

(41) Næs, T.; Isaksson, T.; Fearn, T.; Davies, T. *A User-Friendly Guide to Multivariate Calibration and Classification*, 2nd ed.; IM Publications: Chichester, UK, 2017.

(42) Hubert, P.; Nguyen-Huu, J. J.; Boulanger, B.; Chapuzet, E.; Chiap, P.; Cohen, N.; Compagnon, P. A.; Dewe, W.; Feinberg, M.; Lallier, M.; Laurentie, M.; Mercier, N.; Muzard, G.; Nivet, C.; Valat, L.; Rozet, E. Harmonization of strategies for the validation of quantitative analytical procedures. A SFSTP proposal--part II, *J Pharm Biomed Anal* **2007**, *45*, 70-81.

(43) Horwitz, W. Protocol for the design, conduct and interpretation of method-performance studies: Revised 1994 (Technical Report), *Pure and Applied Chemistry* **1995**, *67*, 331-343.

(44) Martens, H.; Næs, T. *Multivariate calibration*; Wiley, 1989.

(45) AACC International Approved Methods of Analysis. Method 39-00.01 Near-Infrared Methods--Guidelines for Model Development and Maintenance, **2009**.

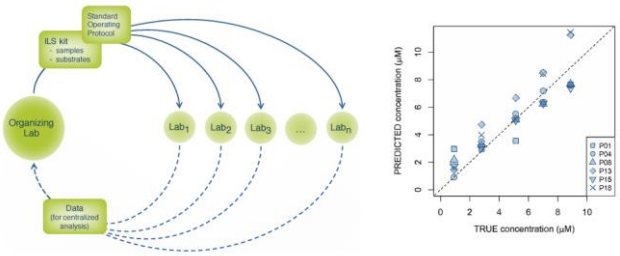



Table of Contents graphic