Comparability of Raman Spectroscopic Configurations: A Large Scale Cross-Laboratory Study

Journal:	*Analytical Chemistry*
Manuscript ID	ac-2020-02696n.R2
Manuscript Type:	Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Guo, Shuxia; Friedrich-Schiller-Universitat Jena, Institut für Physikalische Chemie Beleites, Claudia; chemometrix GmbH; Leibniz-Institut für Photonische Technologien Neugebauer, Ute; Friedrich-Schiller-Universitat Jena, Institut für Physikalische Chemie; Leibniz-Institut für Photonische Technologien; Jena University Hospital, Center for Sepsis Control and Care Abalde-Cela, Sara; International Iberian Nanotechnology Laboratory, Life Science Afseth, Nils Kristian; Nofima AS, Norwegian Institute of Food, Fisheries and Aquaculture Research Alsamad, Fatima; Reims Champagne-Ardenne University Anand, Suresh; National Institute of Optics National Research Council Araujo-Andrade, Cuauhtemoc; Barcelona Institute of Science and Technology, ICFO-Institut de Ciencies Fotoniques Aškrabić, Sonja; University of Belgrade, Institute of Physics Belgrade Avci, Ertug; Yeditepe Universitesi, Genetics and Bioengineering Department, Faculty of Engineering Baia, Monica; Babes-Bolyai University Faculty of Physics Baranska, Malgorzata; Uniwersytet Jagiellonski w Krakowie, Faculty of Chemistry Baria, Enrico; University of Florence, Department of Physics; University of Florence European Laboratory for Non-Linear Spectroscopy Batista de Carvalho, Luis; Universidade de Coimbra Unidade de I&D Química-Física Molecular, Department of Chemistry de Bettignies, Philippe; HORIBA ABX SAS Bonifacio, Alois; Universita degli Studi di Trieste, Department of Materials and Natural Resources Bonnier, Franck; University of Tours Faculty of Pharmacy Brauchle, Eva; University of Tübingen, NMI Natural and Medical Sciences Institute; University of Tübingen, Department of Women's Health, Research Institute of Women's Health Byrne, Hugh; Technological University Dublin - Dublin City Center Campus, Focas Research Institute Chourpa, Igor; Universite Francois-Rabelais de Tours, UFR de Pharmacie Cicchi, Riccardo; National Institute of Optics National Research Council; University of Florence European Laboratory for Non-Linear Spectroscopy Cuisinier, Frederic; Universite de Montpellier, BioNano Culha, Prof. Dr. Mustafa; Yeditepe Universitesi, Genetics and Bioengineering Department, Faculty of Engineering Dahms, Marcel; Friedrich-Schiller-Universitat Jena, Institut für Physikalische Chemie; Jena University Hospital, Center for Sepsis Control

	and Care David, Catalina; HORIBA ABX SAS Duponchel, Ludovic; University of Lille, LASIR Lab Duraipandian, Shiyamala; Technological University Dublin, Radiation and Environmental Science Centre, FOCAS Research Institute El-Mashtoly, Samir; Ruhr-Universitat Bochum, Center for Protein Diagnostics Ellis, David; University of Manchester, School of Chemistry, Manchester Institute of Biotechnology Eppe, Gauthier; University of Liege, Mass Spectrometry Laboratory, MolSys Research Unit Falgayrac, Guillaume; Universite de Lille 2 Faculte de Chirurgie Dentaire, Research Gamulin, Ozren; Sveuciliste u Zagrebu Medicinski fakultet, Department of Physics and Biophysics Gardner, Benjamin; University of Exeter, Physics and Astronomy Gardner, Peter; University of Manchester, School of Chemical Engineering and Analytical Science Gerwert, Klaus; Ruhr-Universitat Bochum, Lehrstuhl fur Biophysik Giamarellos-Bourboulis, Evangelos J.; General University Hospital Attikon, 4th Department of Internal Medicine Gizurarson, Sveinbjorn; University of Iceland Faculty of Pharmaceutical Sciences Gnyba, Marcin; Department of Metrology and Optoelectronics, Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology Goodacre, Royston; University of Liverpool, Department of Biochemistry, Institute of Integrative Biology Grysan, Patrick; Luxembourg Institute of Science and Technology, Materials Research and Technology Guntinas-Lichius, Orlando; Jena University Hospital, ENT Department Helgadottir, Helga ; University of Iceland Faculty of Pharmaceutical Sciences Grošev, Vlasta; Institut Ruđer Bošković Kendall, Catherine; Gloucestershire Royal Hospital, Biophotonics Research Group Kiselev, Roman; St Jude Children's Research Hospital, Structural Biology Kölbach, Micha; Renishaw GmbH Krafft, Christoph; Leibniz-Institut für Photonische Technologien Krishnamoorthy, Sivashankar; Luxembourg Institute of Science and Technology (LIST), Materials Research and Technology (MRT) Kubryck, Patrick; Renishaw GmbH Lendl, Bernhard; Technische Universitat Wien, Institute of Chemical Technologies and Analytics Loza-Alvarez, Pablo; Barcelona Institute of Science and Technology, ICFO-Institut de Ciencies Fotoniques Lyng, Fiona; Technological University Dublin, FOCAS Research Institute Machill, Susanne; TU Dresden, Bioanalytical Chemistry Malherbe, Cedric; Universite de Liege, Analytical Inorganic Chemistry Marro, Monica; ICFO - The Institute of Photonic Sciences, Marques, Maria Paula; Universidade de Coimbra, Dep Life Sciences Matuszyk, Ewelina; Jagiellonian University, Faculty of Chemistry Morasso, Carlo; Istituti Clinici Scientifici Maugeri SpA SB IRCCS Moreau, Myriam; Univ. Lille, CNRS, UMR 8516 - LASIRE Muhamadali, Howbeer; University of Liverpool, Department of Biochemistry, Institute of Integrative Biology Mussi, Valentina; National Research Council, CNR, Istitute for Complex Systems, ISC Notingher, Ioan; University of Nottingham, School of Physics and Astronomy Pacia, Marta; Jagiellonian Centre of Experimental Therapeutics (JCET),

	Jagiellonian University, Pavone, Francesco S.; University of Florence European Laboratory for Non-Linear Spectroscopy Penel, Guillaume; Laboratoire PMOI Petersen, Dennis ; Ruhr-Universitat Bochum, Department of Biophysics, Faculty of Biology and Biotechnology Piot, Olivier; Universite de Reims Champagne-Ardenne, EA 7506 BioSpecT Rau, Julietta; Istituto di Struttura della Materia, Consiglio Nazionale delle Ricerche Richter, Marc; Renishaw GmbH Rybarczyk, Maria; Department of Metrology and Optoelectronics, Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology Salehi, Hamideh; University Montpellier 1, LBN Schenke-Layland, Katja; Eberhard Karls Universitat Tubingen, Dept. for Women's Health, Research Institute for Women's Health Schlücker, Sebastian; University of Duisburg-Essen, Department of Chemistry Schosserer, Markus; University of Natural Resources and Life Sciences Vienna, Department of Biotechnology, Institute of Molecular Biotechnology Schuetze, Karin; CellTool GmbH Sergo, Valter; Universita degli Studi di Trieste, Engineering and Architecture Sinjab, Faris; University of Nottingham, School of Physics and Astronomy Smulko, Janusz; Politechnika Gdanska, Metrology and Optoelectronics Sockalingum, Ganesh; Unité MéDIAN CNRS UMR7369, Pharmacy Dept. Stiebing, Clara; Leibniz Institute of Photonic Technology, Stone, Nick; University of Exeter, Physics and Astronomy Untereiner, Valérie; Université de Reims Champagne-Ardenne, PICT-URCA Vanna, Renzo; Istituti Clinici Scientifici Maugeri SpA SB IRCCS Wieland, Karin; Technische Universitat Wien, Institute of Chemical Technologies and Analytics Popp, Juergen; Leibniz-Institut für Photonische Technologien; Friedrich-Schiller-Universitat Jena, Institut für Physikalische Chemie; Jena University Hospital, Center for Sepsis Control and Care Bocklitz, Thomas; Friedrich-Schiller-Universitat Jena, Institute for Physical Chemistry and Abbe Center of Photonics

Comparability of Raman Spectroscopic Configurations: A Large Scale Cross-Laboratory Study

Shuxia Guo[1,2], Claudia Beleites[2,3], Ute Neugebauer[1,2,4], Sara Abalde-Cela[5], Nils Kristian Afseth[6], Fatima Alsamad[7], Suresh Anand[8], Cuauhtemoc Araujo-Andrade[9], Sonja Aškrabić[10], Ertug Avci[11], Monica Baia[12], Malgorzata Baranska[13,14], Enrico Baria[15,16], Lois A. E. Batista de Carvalho[17], Philippe de Bettignies[18], Alois Bonifacio[20], Franck Bonnier[21], Eva Maria Brauchle[22,23], Hugh J. Byrne[24], Igor Chourpa[21], Riccardo Cicchi[8,16], Frederic Cuisinier[25], Mustafa Culha[11], Marcel Dahms[1,2,4], Catalina David[18], Ludovic Duponchel[19], Shiyamala Duraipandian[24,26], Samir F. El-Mashtoly[27,28], David I. Ellis[29], Gauthier Eppe[30], Guillaume Falgayrac[31, 32], Ozren Gamulin[33, 34], Benjamin Gardner[35], Peter Gardner[29,36], Klaus Gerwert[27,28], Evangelos J. Giamarellos-Bourboulis[37], Sveinbjorn Gizurarson[38], Marcin Gnyba[39], Royston Goodacre[40], Patrick Grysan[41], Orlando Guntinas-Lichius[42], Helga Helgadottir[38], Vlasta Mohaček Grošev[34,43], Catherine Kendall[44], Roman Kiselev[2,45], Micha Kölbach[46], Christoph Krafft[2], Sivashankar Krishnamoorthy[41], Patrick Kubryck[46], Bernhard Lendl[47], Pablo Loza-Alvarez[9], Fiona M. Lyng[24,26], Susanne Machill[48], Cedric Malherbe[30], Monica Marro[9], Maria Paula M. Marques[17,49], Ewelina Matuszyk[14], Carlo Francesco Morasso[50], Myriam Moreau[19], Howbeer Muhamadali[40], Valentina Mussi[51], Ioan Notingher[52], Marta Z. Pacia[14], Francesco S. Pavone[15,16], Guillaume Penel[31,32], Dennis Petersen[28], Olivier Piot[7,53], Julietta V. Rau[54, 55], Marc Richter[46], Maria Krystyna Rybarczyk[56], Hamideh Salehi[25], Katja Schenke-Layland[22,23], Sebastian Schlücker[57], Markus Schosserer[58], Karin Schütze[59], Valter Sergo[20,60], Faris Sinjab[52], Janusz Smulko[39], Ganesh D Sockalingum[7,53], Clara Stiebing[2], Nick Stone[35], Valérie Untereiner[53], Renzo Vanna[50], Karin Wieland[47], Jürgen Popp[1,2], Thomas Bocklitz[1,2,*]

1. Institute of Physical Chemistry and Abbe Center of Photonics, University Jena, 07743 Jena, Germany
2. Leibniz Institute of Photonic Technology Jena, Member of Leibniz Health Technologies, 07745 Jena, Germany
3. Chemometrix GmbH, Södeler Weg 19, 61200 Wölfersheim, Germany
4. Center for Sepsis Control and Care, Jena University Hospital, Am Klinikum 1, D-07747 Jena, Germany
5. International Iberian Nanotechnology Laboratory (INL), Avda Mestre José Veiga, 4715-310 Braga, Portugal
6. Nofima, Norwegian Institute of Food, Fisheries and Aquaculture Research, NO-9291 Tromsø, Norway
7. Université de Reims Champagne-Ardenne, EA 7506 BioSpecT, 51 rue Cognacq-Jay, Reims, 51096 CEDEX, France
8. National Institute of Optics, National Research Council, 50019 Sesto Fiorentino, Italy
9. ICFO-Institut de Ciencies Fotoniques, The Barcelona Institute of Science and Technology, 08860 Castelldefels, Barcelona, Spain
10. Institute of Physics Belgrade, University of Belgrade, Studentski trg 1, Beograd, Serbia
11. Genetics and Bioengineering Department, Faculty of Engineering, Yeditepe University, Kayışdağı, 34755 Ataşehir/İstanbul, Turkey
12. Faculty of Physics, Babes-Bolyai University, Strada Mihail Kogălniceanu 1, Cluj-Napoca 400084, Romania
13. Faculty of Chemistry, Jagiellonian University, 2 Gronostajowa Str., 30-387 Krakow, Poland
14. Jagiellonian Centre for Experimental Therapeutics (JCET), Michała Bobrzyńskiego 14, 30-348 Kraków, Poland
15. Department of Physics, University of Florence, Piazza di San Marco, 4, 50121 Firenze FI, Italy
16. European Laboratory for Non-linear Spectroscopy, Via Nello Carrara, 1, 50019 Sesto Fiorentino FI, Italy
17. University of Coimbra, “Molecular Physical Chemistry” R&D Unit, Department of Chemistry, 3004-535 Coimbra, Portugal
18. HORIBA France SAS, 231 Rue de Lille, 59650 Villeneuve-d'Ascq, France
19. Univ. Lille, CNRS, UMR 8516 - LASIRE – LAboratoire de Spectroscopie pour les Interactions, la Réactivité et l'Environnement, F-59000 Lille, France
20. Raman Lab, Dept. Engineering and Architecture, University of Trieste, Via Alfonso Valerio, 6/1, 34127 Trieste TS, Italy
21. University of Tours, Faculty of pharmacy, EA6295 NanoMédicaments et Nanosondes, 60 Rue du Plat d'Étain, 37000 Tours, France

22. NMI Natural and Medical Sciences Institute at the University of Tübingen, Markwiesenstraße 55, 72770 Reutlingen, Germany
23. Department of Women's Health, Research Institute of Women's Health, Eberhard Karls University Tübingen, Geschwister-Scholl-Platz, 72074 Tübingen, Germany
24. FOCAS Research Institute, Technological University Dublin, City Campus, Aungier St, Dublin, D02 HW71, Dublin 8, Ireland
25. LBN, University Montpellier, 641 Av. du Doyen Gaston Giraud, 34000 Montpellier, France
26. School of Physics & Clinical & Optometric Sciences, Technological University Dublin, City Campus, Kevin Street, Saint Peter's, Dublin 2, D08 X622, Ireland
27. Center for Protein Diagnostics (ProDi), Ruhr University Bochum, Gesundheitscampus 4, 44801 Bochum, Germany
28. Department of Biophysics, Faculty of Biology and Biotechnology, Ruhr University Bochum, Universitätsstraße 150, 44801 Bochum, Germany
29. Manchester Institute of Biotechnology, School of Chemistry, University of Manchester, M1 7DN, Manchester, UK
30. University of Liege, Mass Spectrometry Laboratory, MolSys Research Unit, Place du 20 Août 7, 4000 Liège , Belgium
31. Univ. Lille F-59000 Lille, France, Univ. Littoral Côte d'Opale, F-62300 Boulogne-sur-Mer, France, ULR 4490, MABLab, Marrow Adiposity and Bone Lab.
32. CHU Lille, 2 Avenue Oscar Lambret, F-59000 Lille, France
33. Department of Physics and Biophysics, School of Medicine, University of Zagreb, Šalata 3, 10000 Zagreb, Croatia
34. Centre for Advanced Materials Science, Bijenička 54, 10000 Zagreb, Croatia
35. Physics and Astronomy, College of Engineering, Mathematics and Physical Sciences, Exeter, EX4 4Q, UK
36. Department of Chemical Engineering and Analytical Science, School of Engineering, The University of Manchester M1 3AL, UK
37. 4th Department of Internal Medicine, ATTIKON University Hospital, 1 Rimini Str, 12462 Athens, Greece
38. Faculty of Pharmaceutical Sciences, University of Iceland, Reykjavik, Iceland
39. Gdańsk University of Technology, Faculty of Electronics, Telecommunications and Informatics, Gabriela Narutowicza 11/12, 80-233 Gdańsk, Poland
40. Department of Biochemistry, Institute of Integrative Biology, University of Liverpool, Liverpool L69 7ZB, UK,
41. Materials Research and Technology, Luxembourg Institute of Science and Technology, 5, rue Bommel, 4940 Luxemburg, Belgium
42. Department of Otorhinolaryngology, Jena University Hospital, Bachstraße 18, 07743 Jena, Germany
43. Ruđer Bošković Institute, Bijenička 54, 10000 Zagreb, Croatia
44. Biophotonics Research Unit, Gloucestershire Hospitals NHS Foundation Trust, Leadon House, Great Western Rd, Gloucester GL1 3NN, UK
45. St. Jude Children's Research Hospital, Memphis, 262 Danny Thomas Pl, Memphis, TN 38105, USA
46. Renishaw GmbH, Karl-Benz-Straße 12, 72124 Pliezhausen, Germany
47. Institute of Chemical Technologies and Analytics, TU Wien, 1040 Wien, Austria
48. Chair of Bioanalytical Chemistry, TU Dresden, 01062 Dresden, Germany
49. University of Coimbra, Department of Life Sciences, 3000-456 Coimbra, Portugal
50. Istituti Clinici Scientifici Maugeri IRCCS, Via Salvatore Maugeri, 10, 27100 Pavia PV, Italy
51. National Research Council, Institute for Microelectronics and Microsystems (IMM-CNR), Via del Fosso del Cavaliere, 100, 00133 Roma RM, Rome, Italy
52. School of Physics and Astronomy, University of Nottingham, Nottingham NG7 2RD, UK
53. Université de Reims Champagne-Ardenne, PICT, 9 Boulevard de la Paix, 51100 Reims, France
54. Istituto di Struttura della Materia, Consiglio Nazionale delle Ricerche (ISM-CNR), Via del Fosso del Cavaliere, 100-00133 Rome, Italy
55. Sechenov First Moscow State Medical University, 119991, Moscow, Trubetskaya 8, build. 2, Russian Federation
56. Gdansk University of Technology, Chemical Faculty, 11/12 Narutowicza St., 80-233 Gdansk, Poland
57. Faculty of Chemistry, University of Duisburg-Essen, Universitaetsstr. 5, 45141 Essen, Germany

58. Institute of Molecular Biotechnology, Department of Biotechnology, University of Natural Resources and Life Sciences, Gregor-Mendel-Straße 33, 1180 Vienna, Austria
59. CellTool GmbH, 82327 Tutzing, Germany
60. Faculty of Health Sciences, University of Macau, Macau SAR China

Authors, except SG, CB, UN, JP, TB, are sorted alphabetically.
Corresponding author: *thomas.bocklitz@uni-jena.de

Abstract

The variable configuration of Raman spectroscopic platforms is one of the major obstacles to establish Raman spectroscopy as a valuable physicochemical method within real-world scenarios such as clinical diagnostics. For such real world applications like diagnostic classification, the models should ideally be usable to predict data from different setups. Whether it is done by training a rugged model with data from many setups or by a primary-replica strategy where models are developed on a 'primary' setup and the test data is generated on 'replicate' setups, this is only possible if the Raman spectra from different setups are consistent, reproducible and comparable. However, Raman spectra can be highly sensitive to the measurement conditions and they change from setup to setup even if the same samples are measured.

Although increasingly recognised as an issue, the dependence of the Raman spectra on the instrumental configuration is far from fully understood and great effort is needed to address the resulting spectral variations and to correct for them. To make the severity of the situation clear, we present a round robin experiment investigating the comparability of 35 Raman spectroscopic devices with different configurations in 15 institutes within 7 European countries from the COST (European Cooperation in Science and Technology) action Raman4clinics. The experiment was developed in a fashion that allows various instrumental configurations ranging from highly confocal setups to fibre-optic based systems with different excitation wavelengths. We illustrate the spectral variations caused by the instrumental configurations from the perspectives of peak shifts, intensity variations, peak widths, and noise levels. We conclude this contribution with recommendations that may help to improve the inter-laboratory studies.

Introduction

Raman spectroscopy is known as a non-invasive, label-free technology with high selectivity, and thus it has been increasingly applied to biological studies [1-3]. These studies include forensics [4, 5], diagnostics [6-9], metabolism research [10-12], microbiology [13, 14], clinical pharmacology [15-17], and food science [18, 19]. Most of these studies are yet at the proof-of-concept stage and are performed using a single or multiple similar Raman spectrometer(s). Whilst the Raman spectra from any sample contain the vibrational fingerprint information of the molecules within the sample, they unfortunately also contain fingerprints of the analytical setup, e.g. Raman spectrometer, itself [20]. Therefore, the same sample can lead to different Raman spectra if measured on multiple setups, in different conditions, or at different times. To be clear, we will henceforth refer to 'setup' as an indication of all measurement related effects, be these temporal drifts, variations of measurement conditions, or instrumental configurations (*viz*. laser sources, spectral resolution).

Subtle differences in the Raman setup can degrade the reproducibility of the Raman spectroscopic signals, which makes it almost impossible to quantitatively compare

measurements from different setups and this is highly detrimental to cross-setup data analysis. The setup-dependence is very likely to have a larger influence in biological applications, in which the spectral contributions of interest can be very subtle in the acquired Raman signal, and thus are more easily masked. If Raman spectroscopy is expected to be utilized in e.g. clinical scenarios, it is likely that one or more 'primary' setups hold a large database of Raman spectral characteristics of e.g. some disease states. Other Raman setups in different laboratories act as 'replicate' setups and send data to the 'primary' setup for testing against this 'primary' model/database. If the Raman setups are significantly different, these statistical models are likely to fail to correctly predict the newly measured samples on the 'replicate' setups. Building up individual statistical models for each setup could be a solution to this problem, but this is not likely to be acceptable in real-world scenarios, as the models would also be setup-dependent and this would further hamper the generalisation of Raman spectroscopy.

A more feasible and attractive approach is to remove the setup-induced spectral variations from the spectral 'database'. This is the task of pre-processing in general. The most straightforward way to remove the confounding influence of a setup is spectrometer calibration, including wavenumber and intensity corrections[20-22]. Spectrometer calibration corrects for the influence of a setup on the Raman signals by retrieving the correct values from the measurements of known standard materials. In many cases, an interpolation is needed to deal with the different nominal spectral resolutions of different setups[23]. The extent to which an instrument software takes care of the spectrometer calibration varies greatly across manufacturers; so does the possibility to access raw data which allows the users to apply their own calibration functions from a known basis. Spectrometer calibration does decrease the setup-dependence but rarely removes it completely, due to the inaccuracy in the estimated instrumental response function. The remaining setup-derived spectral variations can still negatively affect the reliability of data analysis. In addition, the standard materials used for calibration may differ across laboratories (especially for biological materials), which can introduce additional spectral variations among different setups.

In addition to robust spectrometer calibration, standard-free approaches such as warping methods, in which the unwanted spectral variations are removed by aligning all spectra against a given reference spectrum, can also be used[24]. However, as a purely data-driven procedure, it does not necessarily give meaningful results, as these procedures may remove both sample- and setup-induced variations without distinction. In this context, methods based on bilinear modelling, such as replicate EMSC (extended multiplicative signal correction) and ASCA (ANOVA-simultaneous component analysis), make it possible to estimate the setup-induced spectral differences and remove them from the data while preserving the sample-induced variations[25-27]. This seems to be a more reliable mechanism, although it requires the experiments to be well designed. Such method is also limited if the sample- and setup-induced spectral variations are not independent of each other.

Considering all the issues stated above, setup-dependence is still a significant concern and a major challenge to Raman spectroscopy being translated into real-world applications. To investigate this analytical challenge, we designed a round robin experiment with researchers from approximately 50 European institutes, which was initiated within the COST action Raman4Clinics. Briefly, aliquots of the same samples were prepared in one partner laboratory and sent to other laboratories, in which data were collected using Raman spectrometers (see Table S1) from various manufacturers, with various laser sources, different spectral

resolutions (i.e., pixel size) and optical configurations ranging from highly confocal to fibre-optic probes (and hence different numerical aperture). We note that, of course, there can be spectral differences due to resonance excitation and therefore, for the present study, we have chosen Raman setups and samples absent of resonance Raman contributions. In the end, data from 35 setups of 15 institutes were returned to a single laboratory for unified data processing and assessments. The comparability of the setups was evaluated from the perspectives of the peak positions, peak intensities, peak widths as well as noise levels. We report these differences and provide recommendations for future cross-laboratory studies. Ultimately, the translation of Raman spectroscopy as a bioanalytical protocol to clinical applications will require regulatory certification, as per, for example the European Medicine Agency "Guideline on bioanalytical method validation", which details requirements of accuracy and precision [28]. We hope that this study will trigger more unified efforts from the Raman community to come to some harmonisation on handling setup-dependence, for instance, rules of setup standardization, certification of devices to be used in clinics, etc. In the end, we hope to help establish Raman spectroscopy as a reliable tool for real-world applications.

Experimental and Methods

Raman spectroscopy

Details of the experimental design are given in supporting information and the spectral data are available *via* an open data repository (https://doi.org/10.5281/zenodo.4152953). Briefly, we received data from 35 setups for the analysis in this contribution. The information of Raman setups is summarised in Table S2, which includes the manufacturer, the source wavelength as well as the nominal spectral resolution (i.e., pixel size). Raman spectra of the NeAr glow lamp and different substances were measured, including agar, gelatine, paracetamol, polystyrene, and cyclohexane. The number of replicate spectra obtained using each setup from the different substances is given in Table 1. The spectral data from setup ID25 were excluded from the analysis, due to recording errors in the wavenumber axis (see Figure S1).

Table 1. The number of spectra for each substance from each setup. The setup ID was sorted according to the excitation wavelength.

Setup ID	Wl./nm	#NeAr	#agar	#gelatine	#paracetamol	#polystyrene	#cyclohexane
ID01	514	10	--	--	10	10	10
ID02	514	--	--	10	10	--	--
ID03	515	711	--	--	1053	709	361
ID04	532	200	10	300	30	300	100
ID05	532	151	135	116	121	11	177
ID06	532	--	--	--	--	20	--
ID07	532	30	135	10	80	12	--
ID08	532	--	--	--	--	120	--
ID09	532	--	--	10	--	--	
ID10	532	11	12	12	13	11	--
ID11	532	10	12	10	--	--	--
ID12	532	--	12	10	12	--	12
ID13	532	200	70	190	10	100	200
ID14	532	1000	60	100	985	900	500

ID15	780	30	22	--	30	--	--
ID16	785	99	--	60	150	20	--
ID17	785	10	--	10	10	10	10
ID18	785	--	--	--	--	--	30
ID19	785	--	--	10	--	--	--
ID20	785	40	--	100	40	--	40
ID21	785	10	10	10	10	10	10
ID22	785	10	10	10	10	10	--
ID23	785	12	11	--	20	30	50
ID24	785	--	--	875	20	--	50
ID25	--	--	--	--	--	--	--
ID26	785	200	58	93	4139	208	50
ID27	785	100	100	100	100	100	100
ID28	785	--	10	--	--	--	--
ID29	785	10	--	10	10	--	10
ID30	785	11	10	11	12	11	--
ID31	785	251	76	348	164	151	26
ID32	785	10	--	--	10	10	10
ID33	785	--	--	--	--	601	814
ID34	785	--	23	10	17	--	14
ID35	785	--	--	--	--	21	--

Spectral pre-processing

The spectra from the NeAr lamp were normalized with respect to their maxima without any additional processing. All spectra from the five substances (i.e., agar, gelatine, paracetamol, polystyrene, and cyclohexane) were subjected to the same pre-processing steps. The details can be found in the supporting information. Briefly, any cosmic spikes were removed *via* a comparison between every pair of Raman spectra of the same substance. A wavenumber calibration was performed based on the spectra of the standard material paracetamol. Thereafter, all spectra were interpolated to an equidistant wavenumber grid of 1 cm^{-1}. This is followed by a sensitive nonlinear iterative peak (SNIP[29]) clipping algorithm to remove any broad baseline artefacts. In the end, a vector normalisation (l_2 norm) was performed. Note, a wavenumber calibration was conducted on 23 of the instruments where paracetamol measurements were reported.

Spectral characterisation

In order to assess any variations between the different setups, the pre-processed Raman spectra were analysed to investigate the following four aspects: signal-to-noise ratio (SNR), peak shift, peak width, and specific peak ratios. These four metrics were chosen to benchmark analytical sensitivity, the reproducibility of peak positions, the spectral resolution, and the reproducibility of relative intensities of a measurement. We consider these four properties to be the most important for establishing whether there are variations between the different Raman platforms. The definition of these characteristics and their calculation is summarised in the following.

Signal-to-noise ratio (SNR): The SNR can be used to determine the lowest signal that can be reliably detected, by comparing an average signal to the noise level. Herein, the SNR was calculated in two ways using the samples of agar and gelatine, which featured the highest

noise among all measured substances in this study. The mean spectra of these two substances from each measurement are given in Figures S2-S3, along with the standard deviation, as shadows. In the first case, we calculated the SNR for each spectrum individually as the ratio between the mean Raman intensity and the standard deviation of the estimated noise (see Eq. (S3.1). The noise (I_n) was estimated as the difference between the spectrum and the output of a strong Savitzky–Golay (S-G) smoothing ($p = 2, n = 31$) of the spectrum[30]. Herein the parameters p and n represent the degree of the polynomial function and the window size of the S-G filter, respectively. An example of spectrum before and after smoothing together with the estimated noise is given in Figure S8. As the noise could be upper-biased for the spectral region with sharp Raman bands, the SNR tends to be under-estimated in this case. In the second case, we calculated the SNR as the ratio between the mean and standard variation of the Raman intensities at a given wavenumber over 10 randomly selected Raman spectra (see Eq. (S3.2)), in which the peak intensities of the Raman bands around 846 and 1451 cm^{-1} were used for agar and gelatine, respectively. These two Raman bands were selected to be sufficiently intense and well separated from other bands.

Peak shift: The peak shift characterises the accuracy of the wavenumber axis (*x*-axis, spectral axis) of a setup. It was defined as the deviation of a measured peak position from its theoretical position. To do so, we employed three substances having well-defined Raman bands: paracetamol, polystyrene, and cyclohexane. The mean spectra and standard deviations of the three substances for each setup are shown in Figure S4-S6 (a). The Raman bands used in our calculation are highlighted in Figure S4-S6 (b). Only the bands within the fingerprint region were considered, as the CH stretching region was not measured on all setups. To start, we fitted each of these Raman bands by a Gaussian peak according to Eq. (1), because Gaussian fits have been proven to work well for solids, powders, gels or resins samples[31]. The spectral region used for the fit was defined by $\tilde{\nu}_0 \pm 10$ cm^{-1}, where $\tilde{\nu}_0$ denotes the theoretical peak position. The parameter μ of the resulting Gaussian peak was considered to be the measured peak position of the Raman band being fitted and the peak shift of this Raman band was determined by: $\Delta\tilde{\nu} = \mu - \tilde{\nu}_0$. The final result of one single spectrum was determined as the mean absolute value of all peak shifts from the multiple Raman bands within this spectrum, which indicate the mean absolute deviation in the wavenumber axis of a measurement.

$$\mathrm{I}(\tilde{\nu}) = \mathrm{A} \cdot \exp\left(-\frac{(\tilde{\nu} - \mu)^2}{2\sigma^2}\right) \qquad \text{Eq. (1)}$$

Peak width: The peak width is considered as a straightforward metric to characterise the spectral resolution of a given setup and can be benchmarked by the full-width-at-half-maximum (FWHM). Herein, we employed the results of σ from the peak fit described in the previous section and obtained the FWHM according to Eq. (2). The calculation was performed on both the NeAr emission as well as the Raman spectra of the three substances, i.e., paracetamol, polystyrene, and cyclohexane. The measured NeAr emission on different setups was visualized in Figure S7. To make the determination more precise, the FWHM was only calculated for the best-defined peak in each case; i.e., no significant shoulder peaks exist and the peak does not overlap with its neighbouring peaks. According to this criterion, we employed the emission at 626.56 and 878.2 nm for the NeAr measurement with laser sources of 514/532 and 785 nm, respectively; the peak at 1169, 1602, and 1028 cm^{-1} was used for paracetamol, polystyrene, and cyclohexane, respectively.

$$\mathrm{FWHM} = 2\sigma \cdot \sqrt{2\ln(2)} \qquad \text{Eq. (2)}$$

Peak ratio: The peak ratio is defined as the ratio between the areas of two given Raman bands. It quantifies the consistency of the relative Raman intensity across setups and it is an important criterion for a reproducible Raman spectroscopic data analysis, either quantitative or qualitative. In particular, the area of a given Raman band was calculated *via* a simple integration of the peak region according to Eq. (3). We did not additionally subtract the baseline offset during this integration, as the spectra were already baseline corrected. The terms k_l and k_r denote the start and the end index of the band. These values were determined automatically as the positions around a band where the Raman intensity starts to increase (i.e., $I(\tilde{\nu}_{k_l-1}) > I(\tilde{\nu}_{k_l})$) or stops decreasing (i.e., $I(\tilde{\nu}_{k_r}) < I(\tilde{\nu}_{k_r+1})$), respectively. We based our calculation on the spectra of cyclohexane and obtained three peak ratios for each spectrum: the area of the bands at 1028, 1266, and 1444 cm^{-1} with respect to the area of the band at 801 cm^{-1}. The four bands are well-defined and do not show significant shoulders or overlap with neighbouring bands, which helps to precisely estimate the peak area.

$$\mathrm{A} = \sum_{i=k_l}^{k_r} I(\tilde{\nu}_i) \qquad \text{Eq. (3)}$$

Results and Discussion

We present here the results of assessment of the above-mentioned metrics calculated for different setups and substances. As our major focus in this study is the cross-setup comparability, we will mainly compare the results of different setups with each other but not specifically consider their reference values given in the literature. For such reference information, the interested readers are kindly referred to refs.[32-34].

The SNR results using a single-spectrum and 10 replicate spectra are given in Figure 1 and Figure S9, respectively. The colour shades represent setups with different source wavelengths. The meta-information of the datasets is provided in the legend, including the identification number (ID) of the setup, the type of the data, and the excitation wavelength for each measurement. The two data types, represented by '*raw*' and '*ical*', differed in whether an intensity calibration was done by the setup itself or not, respectively. In particular, the SNR of each measurement was plotted as a single bar containing the information of the mean and standard deviation over the multiple spectra from the same measurement (the number of replicates for each experimental setup is provided in Table 1). The red dash line marks the global mean of the SNR over all measurements, while the grey dash lines show 1 to 3 times the global standard deviation of SNR over all measurements. The minimal mean SNRs for agar and gelatine, given by the blue dash line, are 3.8 and 5.8, respectively. That is to say, the SNR is sufficient for all setups, which is reasonable considering a preliminary SNR filtering during the data upload (see experimental design in supporting information). This can be concluded as well from the SNR calculated from multiple spectra, as is shown in Figure S9.

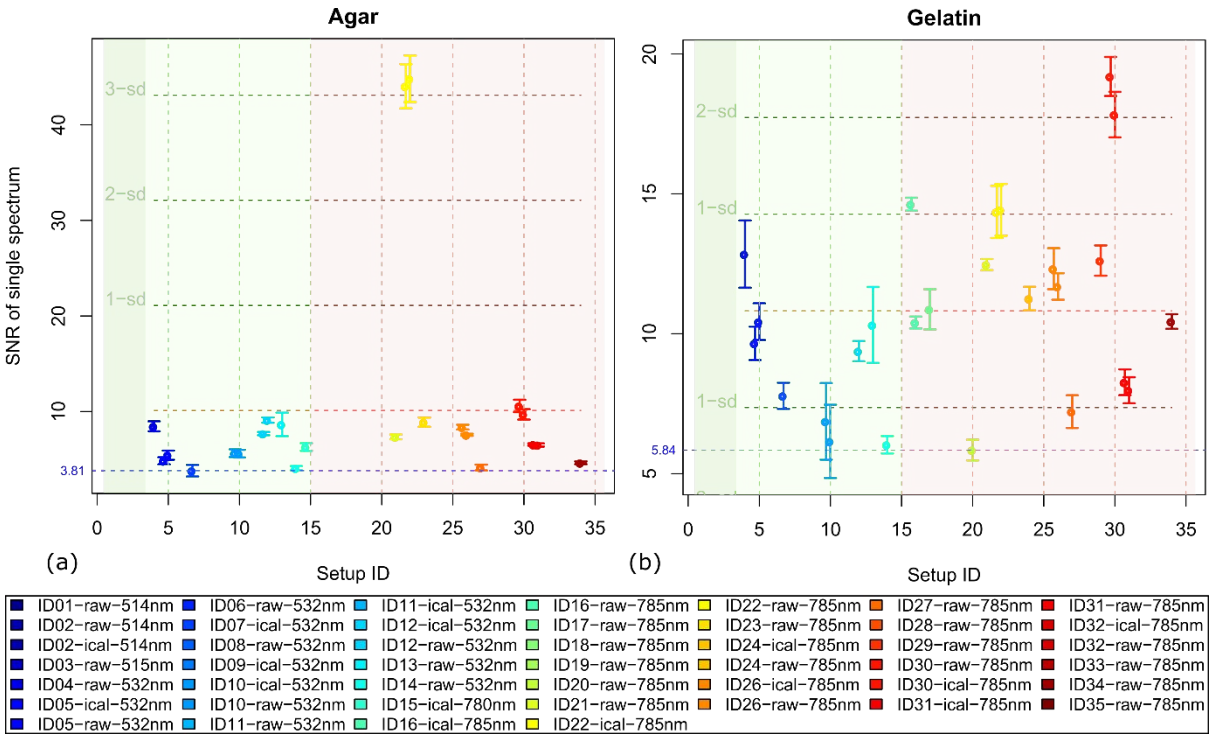


Figure 1. SNR of agar (a) and gelatine (b): The colour shades represent the setups with different source wavelengths. Each bar represents the measurement of a substance on a given setup. The bar contains the mean and standard deviation value of the SNR for each setup. The average and standard deviation of all setups are given by the red and grey dash lines, respectively. The blue dash line marks the minimal SNR of all setups. The identification of the setup, the type of data, and the source wavelength are given in the legend. In particular, the data type '*ical*' and '*raw*' denotes the cases where the data were exported from the setup with (*ical*) and without (*raw*) intensity calibration, respectively.

The results of the peak shift calculations are plotted in Figure 2 and Figure S10, following the same structure as in Figure 1. It is implicit in the definition that, the smaller the absolute value of the peak shift, the better the setup. An ideal measurement would give a zero peak shift; i.e., peaks from the measurement match perfectly with the theoretical values. As the ideal case is not easily achievable in reality, wavenumber calibration is important to minimise the peak shifts. To do so, we applied paracetamol as the standard material and calculated a calibration function based on the deviation between the measured and the theoretical positions for well-defined bands. This wavenumber calibration function was thereafter applied to the spectra of paracetamol, polystyrene, and cyclohexane. As shown in Figure 2 and Figure S10, the peak shifts were reduced closer to zero after the calibration, compared to before calibration. To make the conclusion clearer, we summarised the statistics of the peak shifts in Table 2. Therein, $\overline{\Delta\tilde{\nu}}$ denotes the mean peak shift of each setup. The factors sd and sd_0 represent the standard deviation over all measurements and standard deviation for a single measurement, respectively. In particular, the decrease of $\overline{\Delta\tilde{\nu}}$ demonstrates an improved setup-independence after the wavenumber calibration, although we note that there are also setups that are (extremely) negatively affected (ID05, ID14), or not improved (ID31). A generally better reproducibility within the same measurement was observed after wavenumber calibration, evidenced by a reduced maximal sd_0. The increase in the mean and standard deviation of $\overline{\Delta\tilde{\nu}}$ after calibration for polystyrene and cyclohexane will be explained in the next paragraph.

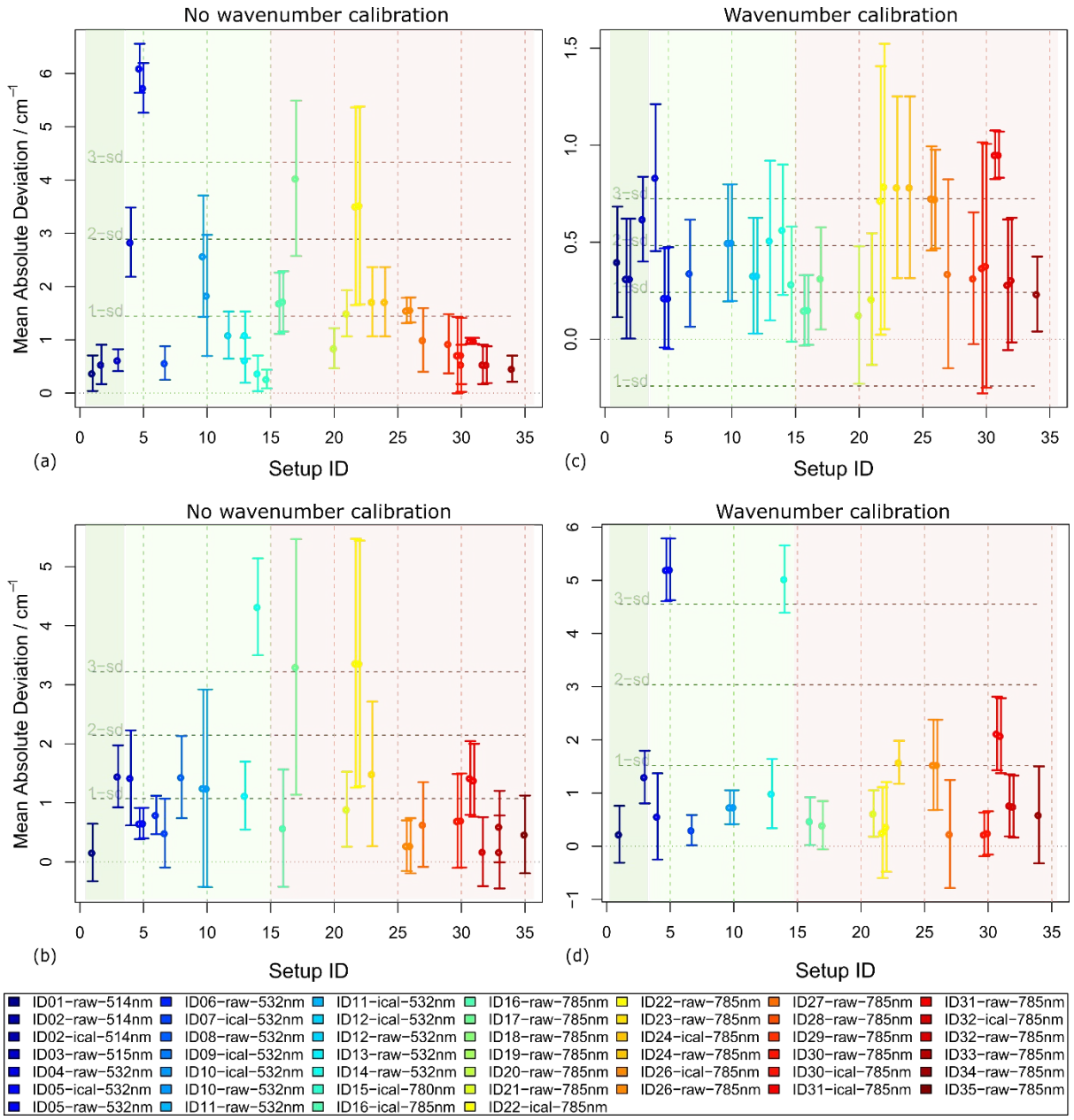


Figure 2. Results of peak shifts for paracetamol (a, c) and polystyrene (b, d), without and with wavenumber calibration. The wavenumber calibration for polystyrene was only calculated if there is paracetamol spectrum measured on the same setup. Therefore, a lower number of setups are shown for the calibrated results for polystyrene than without wavenumber calibration.

Table 2. Statistics of peak shifts, summarised from the results of Figure 2.

Substance		$\mathbf{mean}(\overline{\Delta\tilde{\nu}})$	$\boldsymbol{sd}$	$\mathbf{max}(\boldsymbol{sd_0})$
Paracetamol	No calibration	1.58 cm^{-1}	1.44 cm^{-1}	1.86 cm^{-1}
	Calibrated	0.45 cm^{-1}	0.24 cm^{-1}	0.73 cm^{-1}
Polystyrene	No calibration	1.20 cm^{-1}	1.07 cm^{-1}	2.16 cm^{-1}
	Calibrated	1.31 cm^{-1}	1.52 cm^{-1}	1.02 cm^{-1}
Cyclohexane	No calibration	1.33 cm^{-1}	1.79 cm^{-1}	1.37 cm^{-1}

	Calibrated	1.93 cm^{-1}	2.16 cm^{-1}	0.73 cm^{-1}

As it is apparent in the figures, the improvement due to the wavenumber calibration was not the same for the three substances: it was the most significant for paracetamol but much less so for polystyrene and cyclohexane. As is shown in Table 2, the mean of $\overline{\Delta\tilde{\nu}}$ was decreased for paracetamol but not for the other two substances. This is easy to explain, as the calibration function of a setup was 'learnt' from the spectra of paracetamol and 'generalized' to the other substances. This function is supposed to capture the best the peak shift of the Raman spectra of paracetamol and less so for the other substances. The performance of the wavenumber calibration can also be degraded due to other issues. For example, the peak shifts can increase after wavenumber calibration if the spectra of paracetamol measured on the specific instrument do not truly reflect the peak shift of spectra to be calibrated. This was seen for the setup 'ID05', for which the spectra of paracetamol exhibited significantly larger peak shifts than those of polystyrene and cyclohexane. The wavenumber calibration in this scenario introduced additional errors in the peak positions and led to increased peak shifts (Figure 2 and Figure S10). Another possibility that makes the wavenumber calibration invalid was seen from setup 'ID14', for which the peak shifts in paracetamol were much smaller than the shifts in the spectra of polystyrene and cyclohexane. Therefore, the calibration function could not represent the peak shifts well and thus the wavenumber calibration hardly reduced the peak shifts (Figure 2 and Figure S10). These two scenarios may happen if the standard material (e.g. paracetamol) and the samples to be calibrated (e.g. polystyrene and cyclohexane in this study) are measured under different conditions. It could also be the case that the setup bears short-term instability, leading to variations among spectra of standard material and samples. This could be a significant concern in applying Raman spectroscopy. However, we will not go into details of this issue as we did not have sufficient data from this respect.

An additional interesting observation is that sd, which benchmarks the comparability across setups, was increased by wavenumber calibration for polystyrene and cyclohexane. There are two likely reasons, to our assessment. First, the two cases with improper calibration as mentioned previously contributed to the increase in sd. Second, the variations in the spectra of paracetamol may be passed to the calibration procedure and add to the variation in the calibrated spectra. With all these observations, it is obvious and fair to say that the peak shifts remain a significant issue for the cross-setup reproducibility in Raman spectroscopy. Wavenumber calibration, as was already shown in[21], cannot completely remove the setup-induced shifts in the wavenumber axis of Raman spectroscopy.

The results of the FWHM are given in Figure 3(a-b) for paracetamol and polystyrene, respectively. Similar results are shown in Figure S11 (a-b) for cyclohexane and the NeAr lamp, respectively. The cross-setup variations are clearly seen from all results. In addition, the setup 'ID22' was detected as an 'outlier', having significantly larger FWHM (i.e., lower spectral resolution), which turned out to differ from others as it was a fibre-optics based setup with low nominal spectral resolution (see table S1). In addition, the intensity calibration, which was undertaken in some setups, did not significantly influence the peak widths. This can be seen from the similar results for all the encircled pairs in the plots, which correspond to the results from spectra of the same setup, with and without the intensity calibration.

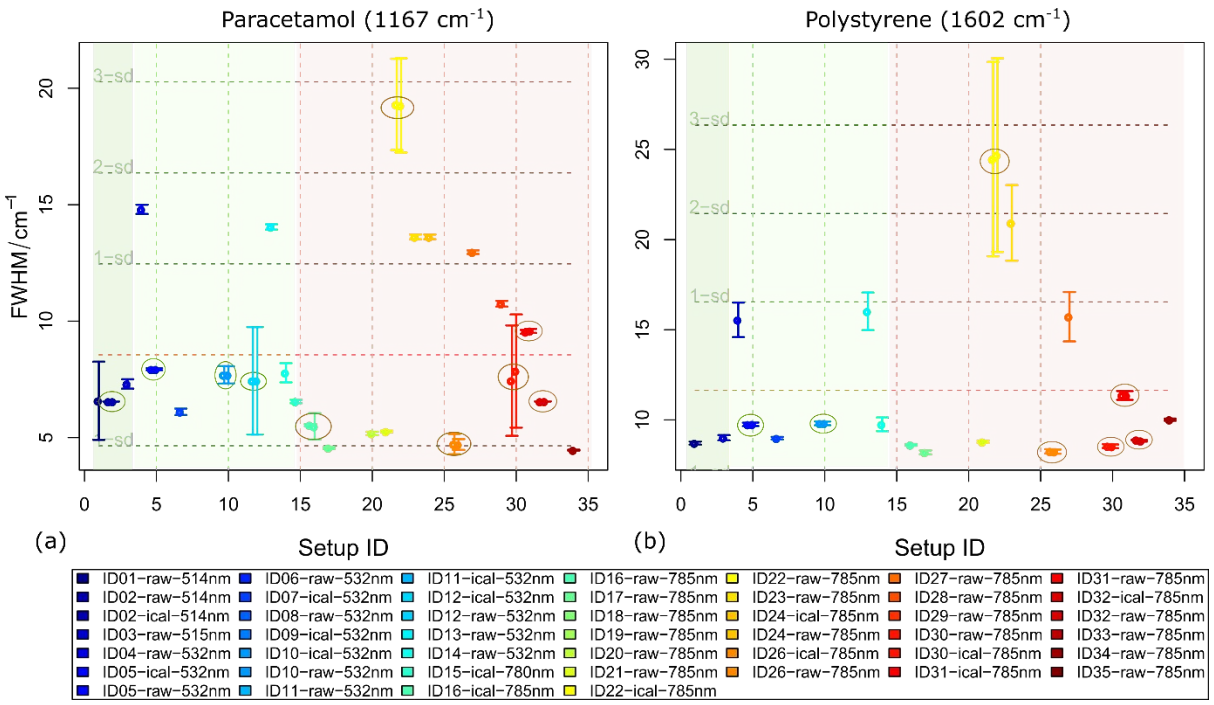


Figure 3. Results of FWHM calculated from paracetamol (a) and polystyrene (b) based on selected peaks. The setups from different source wavelengths were marked by the coloured shades. Pairs of spectra with and without intensity calibration were highlighted by circles. The variations among different setups are clearly seen from the results.

To check the influence of nominal spectral resolution (i.e., pixel size in wavenumber) on the estimation of the peak shift and peak width, we visualized additionally the results of the peak shift and the FWHM from all setups with respect to their nominal spectral resolutions. As shown in Figure S12-13, no correlation can be concluded between the spectral resolution and the peak shift or peak width. That justifies our estimation of these two spectral characteristics, which were not systematically biased by the differences in the nominal spectral resolution (i.e., pixel size).

The results of the peak ratios are given in Figure 4 and Figure S14. The curved brown arrows indicate data pairs from the same setup, with (*ical*) and without (*raw*) intensity calibration. In general, the peak ratios varied largely across setups. Additionally, there was no clear trend as to how the peak ratios change with the source wavelengths. Rather, however, the intensity calibration did not necessarily bring the peak ratios closer to the global mean; i.e., this correction did not improve the cross-setup comparability. In contrast, the influence of the intensity calibration seemed to be rather arbitrary. These facts very likely indicate the inaccuracy of the estimated intensity response function.

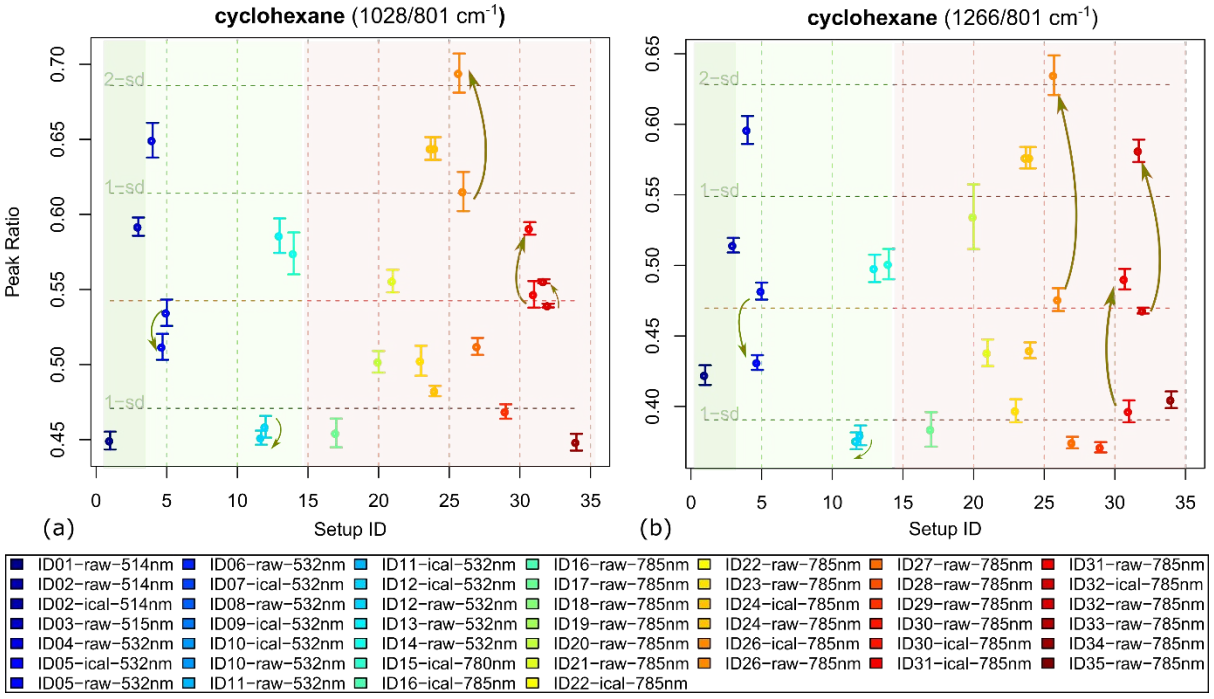


Figure 4. Ratios of peak intensities at (a) 1028 and (b) 1266 cm^{-1} to the peak intensity at 801 cm^{-1}. The colour shades represent the setups with different source wavelengths. The curved arrows in brown show the paired results without and with intensity calibration, pointing to the results with intensity calibration. It is clear that the peak ratios varied largely among measurements, and the setup intern intensity calibration does not improve the situation.

Table 3. Statistics of the peak ratios summarised from Figure 4 and Figure S14 for the setups with different source wavelengths (514, 532, and 785 nm). The terms $\mathrm{mean}(\bar{r})$, sd, and $\max(sd_0)$ represent the global average, global standard deviation, and the maximal standard deviation of the peak ratios, respectively. The variations in the relative intensities tend to increase between two peaks with longer distance along the wavenumber axis.

Raman bands (cm^{-1})	$\mathbf{mean}(\bar{r})$			$\boldsymbol{sd}/\mathbf{mean}(\bar{r})$			$\mathbf{max}(\boldsymbol{sd_0})/\mathbf{mean}(\bar{r})$		
	514 nm	532 nm	785 nm	514 nm	532 nm	785 nm	514 nm	532 nm	785 nm
1028/801	0.52	0.54	0.55	0.19	0.13	0.13	0.01	0.03	0.02
1266/801	0.47	0.47	0.47	0.14	0.17	0.18	0.02	0.02	0.05
1444/801	0.53	0.54	0.52	0.11	0.24	0.29	0.03	0.03	0.13

For clearer interpretation, we additionally summarised the statistics of the peak ratios in Table 3, corresponding to the results of the three source wavelengths: 514, 532, and 785 nm. Therein, $\bar{\boldsymbol{r}}$ and $\boldsymbol{sd_0}$ represent the mean and standard deviation of each setup, while $\mathrm{mean}(\bar{\boldsymbol{r}})$ and sd represent the global mean and standard deviation. Taking $\mathrm{mean}(\bar{\boldsymbol{r}})$ as an approximation of the 'true' peak ratio, the cross- and within-setup variations can be benchmarked by $\max(\boldsymbol{sd_0})/\mathrm{mean}(\bar{\boldsymbol{r}})$ and $sd/\mathrm{mean}(\bar{\boldsymbol{r}})$, respectively. On this basis, a comparison among the three ratios showed that the cross- and within-setup variations increased when the two peaks involved are further apart in the spectra. That is to say, the variations in the relative peak intensities tend to increase between two points with longer distance along the wavenumber axis. This is understandable, as the spectrometer response function has a stronger influence if the two peaks are further apart.

With all the results shown above, we would like to stress the profound influence of the setup-dependence in Raman spectroscopy. The routinely applied spectrometer calibration clearly does not remove all setup-related effects on Raman spectra when the same chemical substance is measured on different Raman platforms. Inadequate calibration can be for multiple reasons: (i) it is difficult to duplicate the measurement conditions exactly between the standard material and real samples; (ii) Raman spectra can be angle and relative orientation dependent if the standard is crystalline, and the setup (laser/grating) is polarisation sensitive; (iii) there may be influence from contaminations if the standard material is not refreshed regularly – for this particular experiment, the latter reason is clearly not the case, as one laboratory supplied all materials for analysis to all participating laboratories; (iv) the full technical details of built-in (automatic) calibrations are normally inaccessible and thus it is hard to ensure that exactly the same calibration methods are used by the different manufacturers or setups. The application of different calibration algorithms certainly constitutes an additional source of setup-dependence.

Based on all facts and reasons stated above, computational strategies to remove the setup-induced spectral variations are urgently needed. As a first attempt to solve this issue, we have developed recently model transfer methods to deal with cross-setup variations in Raman spectroscopy[26, 35, 36]. However, a long-term effort is needed to completely solve the issue of setup-dependence, and this can only be achieved if cross-laboratory comparisons are undertaken, such as those reported in the present study and indeed elsewhere for surface-enhanced Raman scattering[37].

Conclusions

This contribution reports a large-scale and cross-laboratory round robin Raman experiment designed by researchers from approximately 50 institutes in Europe. The data were submitted by 15 different institutes and measured with 35 Raman setups, which were from 5 different manufacturers and configured with multiple laser sources. We have presented the results with respect to four key metrics to assess Raman performance across-setups. These include the SNR, the peak shift, the peak width, and the peak ratio for spectra of a common set of five substances measured on different setups. In this way, we could quantify and analyse the cross-setup comparability and variability in Raman spectroscopy and we could verify the appropriateness to implement spectrometer calibration. More inspiring studies and efforts are certainly needed to standardize Raman spectroscopy and hence push Raman techniques closer to real-world applications. A detailed investigation on each of the different sources that harm the spectral reproducibility can be a good start. In addition, we recommend the following points which we believe are worthy for future investigations.

- First, we suggest that the manufacturers make spectrometer calibration a by-default included module and make the full technical details of the correction explicit and open access. This built-in process is, on one hand, more likely to achieve ‘duplicated’ conditions for the real samples and the standard material, and on the other hand, makes calibration easier. In addition, by applying the same calibration procedure, the variations introduced by in-house written calibration programs in different labs are reduced. Moreover, further investigations of the temporal reliability of instruments and calibration are needed. Any short-term setup instabilities should be understood and well controlled for the development of clinical applications.
- Second, we recommend that the manufacturers provide access to the real raw data,

before any processing steps are applied, so that it is more feasible to unravel data that are closer to the 'physical truth'. We would like to encourage both researchers and manufacturers to work together on standard operating procedures for verifying instrument calibration and performance.

- Third, we encourage researchers and scientists to make their data openly available and actively contribute to establishing larger databases. This has been done in other data rich communities (e.g. ref.[38]) and would be a vast invaluable resource to build statistical models that are tolerant to unwanted spectral differences, like the spectral variations between seemingly identical analytical setups. In this way, reproducible predictions for different measurements from the same sample can be achieved. A larger database can also provide a better reference used for the spectral alignment approaches.
- Finally, we advocate a broader cooperation on the same scientific question in order to come up with 'global' solutions and reduce the variations in the setups used in different research groups answering the same question.

Acknowledgements

This publication is based upon work from COST Action BM1401 Raman4Clinics, supported by COST (European Cooperation in Science and Technology). COST is a funding agency for research and innovation networks (www.cost.eu). Our Actions help connect research initiatives across Europe and enable scientists to grow their ideas by sharing them with their peers. This boosts their research, career and innovation.

The authors would like to thank MD for preparing gelatine and agar gels. RK portioned, packaged and shipped the samples with the help of CS and M. Spreemann who did the customs declarations for samples shipped outside the EU. The authors highly thank M. Bedoni, L. Colombo, H. Dehghani, A. Mourka, W. Nahm, A. Pifferi, and L. Spinelli, who have actively contributed to the parallel discussion session 'Open Data' at the workshop 'Performance Assessment and Standardization in Biophotonics' taking place on 12.09.2019 in Brussels. The ideas and suggestions listed in the conclusion of this manuscript include a collective contribution from this parallel discussion. L.A.B.C. and M.P.M.M. acknowledge the Portuguese Foundation for Science and Technology (Project UIDB/00070/2020). S.K. and P.G. acknowledge the funding from National Research Fund of Luxembourg (FNR) *via* the project PLASENS (C15/MS/10459961). The scholarship from China scholarship council (CSC) for S.G is highly acknowledged. This project was supported by the BOKU Core Facilities Multiscale Imaging.

Associated Content

Supporting Information

- **Table S1** Acceptable signal-to-noise ratio for specific band integral intensities across the at least 10 submitted spectra.
- **Table S2** Information of Raman spectroscopic devices within the trial.
- **Figure S1**. Error in the measurement of the setup ID25.
- **Figure S2**. Mean spectra and standard deviation spectra of agar.
- **Figure S3**. Mean spectra and standard deviation spectra of gelatine.
- **Figure S4**. Mean and standard deviation spectra, the representative Raman bands of paracetamol.
- **Figure S5**. Mean and standard deviation spectra, the representative Raman bands of

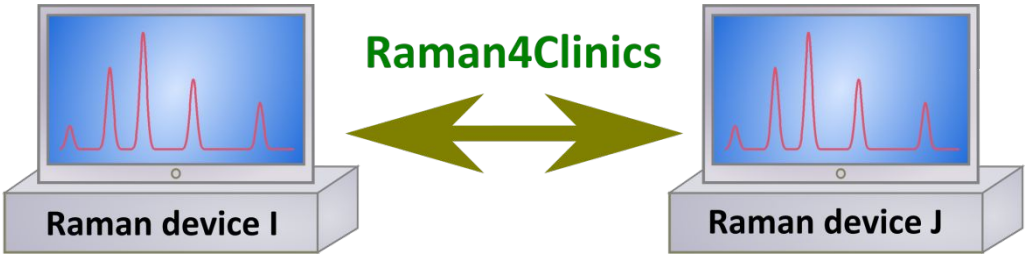
polystyrene.

- **Figure S6**. Mean and standard deviation spectra, the representative Raman bands of cyclohexane.
- **Figure S7**. Mean spectra and standard deviation spectra of the NeAr lamp.
- **Figure S8**. Example of estimated SNR
- **Figure S9**. Results of SNR for agar and gelatine.
- **Figure S10**. Results of peak shifts for cyclohexane without and with wavenumber calibration.
- **Figure S11**. Results of FWHM calculated from cyclohexane and NeAr.
- **Figure S12**. Mean absolute deviation with respective to the nominal spectral resolution of different setups.
- **Figure S13**. FWHM with respective to the nominal spectral resolution.
- **Figure S14**. Results of the peak ratio between the cyclohexane peaks at 1444 and 801 cm^{-1}.

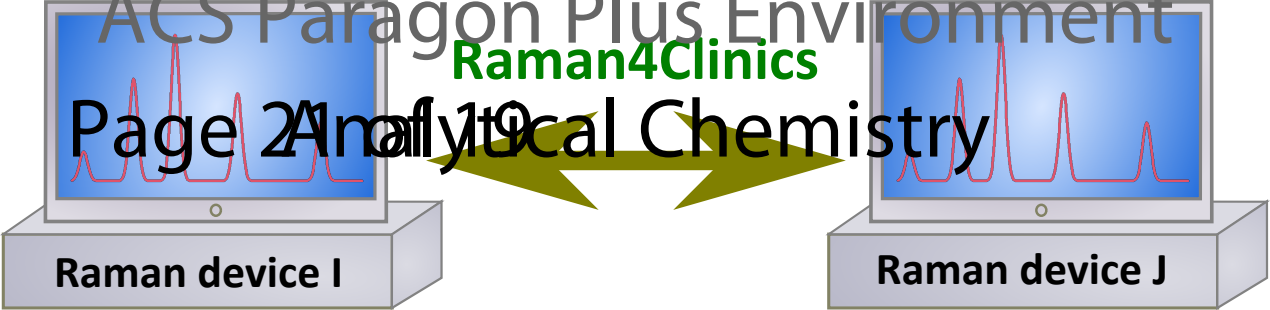
References

1. Bocklitz, T. W.; Guo, S.; Ryabchykov, O.; Vogler, N.; Popp, J. r., Raman based molecular imaging and analytics: a magic bullet for biomedical applications!? *Analytical chemistry* **2016,** *88* (1), 133-151.
2. Butler, H. J.; Ashton, L.; Bird, B.; Cinque, G.; Curtis, K.; Dorney, J.; Esmonde-White, K.; Fullwood, N. J.; Gardner, B.; Martin-Hirsch, P. L., Using Raman spectroscopy to characterize biological materials. *Nature protocols* **2016,** *11* (4), 664.
3. Cheng, J.-X.; Xie, X. S., Vibrational spectroscopic imaging of living systems: An emerging platform for biology and medicine. *Science* **2015,** *350* (6264), aaa8870.
4. de Oliveira Penido, C. A. F.; Pacheco, M. T. T.; Lednev, I. K.; Silveira Jr, L., Raman spectroscopy in forensic analysis: identification of cocaine and other illegal drugs of abuse. *Journal of Raman Spectroscopy* **2016,** *47* (1), 28-38.
5. Doty, K. C.; Lednev, I. K., Raman spectroscopy for forensic purposes: recent applications for serology and gunshot residue analysis. *TrAC Trends in Analytical Chemistry* **2018,** *103*, 215-222.
6. Marro, M.; Taubes, A.; Abernathy, A.; Balint, S.; Moreno, B.; Sanchez‐Dalmau, B.; Martínez‐Lapiscina, E. H.; Amat‐Roldan, I.; Petrov, D.; Villoslada, P., Dynamic molecular monitoring of retina inflammation by in vivo Raman spectroscopy coupled with multivariate analysis. *Journal of biophotonics* **2014,** *7* (9), 724-734.
7. Neugebauer, U.; Bocklitz, T.; Clement, J.; Krafft, C.; Popp, J., Towards detection and identification of circulating tumour cells using Raman spectroscopy. *Analyst* **2010,** *135* (12), 3178-3182.
8. Stremersch, S.; Marro, M.; Pinchasik, B. E.; Baatsen, P.; Hendrix, A.; De Smedt, S. C.; Loza‐Alvarez, P.; Skirtach, A. G.; Raemdonck, K.; Braeckmans, K., Identification of Individual Exosome‐Like Vesicles by Surface Enhanced Raman Spectroscopy. *Small* **2016,** *12* (24), 3292-3301.
9. Vogler, N.; Bocklitz, T.; Subhi Salah, F.; Schmidt, C.; Bräuer, R.; Cui, T.; Mireskandari, M.; Greten, F. R.; Schmitt, M.; Stallmach, A., Systematic evaluation of the biological variance within the Raman based colorectal tissue diagnostics. *Journal of biophotonics* **2016,** *9* (5), 533-541.
10. Aljakouch, K.; Lechtonen, T.; Yosef, H. K.; Hammoud, M. K.; Alsaidi, W.; Kötting, C.; Mügge, C.; Kourist, R.; El‐Mashtoly, S. F.; Gerwert, K., Raman microspectroscopic evidence for the metabolism of a tyrosine kinase inhibitor, neratinib, in cancer cells. *Angewandte Chemie International Edition* **2018,** *57* (24), 7250-7254.
11. Marro, M.; Nieva, C.; de Juan, A.; Sierra, A., Unravelling the metabolic progression of breast cancer cells to bone metastasis by coupling Raman spectroscopy and a novel use of MCR-ALS algorithm. *Analytical chemistry* **2018,** *90* (9), 5594-5602.
12. Van Nest, S. J.; Nicholson, L. M.; Pavey, N.; Hindi, M. N.; Brolo, A. G.; Jirasek, A.; Lum, J. J., Raman spectroscopy detects metabolic signatures of radiation response and hypoxic fluctuations in non-small cell lung cancer. *BMC cancer* **2019,** *19* (1), 474.
13. Hanson, C.; Barney, J. T.; Bishop, M. M.; Vargis, E., Simultaneous isolation and label‐free identification of bacteria using contactless dielectrophoresis and Raman spectroscopy. *Electrophoresis* **2019,** *40* (10), 1446-1456.
14. Stöckel, S.; Meisel, S.; Elschner, M.; Rosch, P.; Popp, J., Identification of Bacillus anthracis via Raman spectroscopy and chemometric approaches. *Analytical chemistry* **2012,** *84* (22), 9873-9880.

15. Enejder, A. M.; Koo, T.-W.; Oh, J.; Hunter, M.; Sasic, S.; Feld, M. S.; Horowitz, G. L., Blood analysis by Raman spectroscopy. *Optics letters* **2002,** *27* (22), 2004-2006.
16. Schut, T. C. B.; Puppels, G. J.; Kraan, Y. M.; Greve, J.; Van Der Maas, L. L.; Figdor, C. G., Intracellular carotenoid levels measured by Raman microspectroscopy: comparison of lymphocytes from lung cancer patients and healthy individuals. *International journal of cancer* **1997,** *74* (1), 20-25.
17. Zakel, S.; Rienitz, O.; Güttler, B.; Stosch, R., Double isotope dilution surface-enhanced Raman scattering as a reference procedure for the quantification of biomarkers in human serum. *Analyst* **2011,** *136* (19), 3956-3961.
18. Jiang, Y.; Sun, D.-W.; Pu, H.; Wei, Q., Surface enhanced Raman spectroscopy (SERS): A novel reliable technique for rapid detection of common harmful chemical residues. *Trends in Food Science & Technology* **2018,** *75*, 10-22.
19. Su, W. H.; Sun, D. W., Fourier transform infrared and Raman and hyperspectral imaging techniques for quality determinations of powdery foods: a review. *Comprehensive Reviews in Food Science and Food Safety* **2018,** *17* (1), 104-122.
20. McCreery, R. L., *Raman spectroscopy for chemical analysis*. John Wiley & Sons: 2005; Vol. 225.
21. Bocklitz, T.; Dörfer, T.; Heinke, R.; Schmitt, M.; Popp, J., Spectrometer calibration protocol for Raman spectra recorded with different excitation wavelengths. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **2015,** *149*, 544-549.
22. Dörfer, T.; Bocklitz, T.; Tarcea, N.; Schmitt, M.; Popp, J., Checking and improving calibration of Raman spectra using chemometric approaches. *Zeitschrift für Physikalische Chemie* **2011,** *225* (6-7), 753-764.
23. Mann, C. K.; Vickers, T. J., Instrument-to-instrument transfer of Raman spectra. *Applied spectroscopy* **1999,** *53* (7), 856-861.
24. Bloemberg, T. G.; Gerretzen, J.; Lunshof, A.; Wehrens, R.; Buydens, L. M., Warping methods for spectroscopic and chromatographic signal alignment: a tutorial. *Analytica chimica acta* **2013,** *781*, 14-32.
25. Afseth, N. K.; Kohler, A., Extended multiplicative signal correction in vibrational spectroscopy, a tutorial. *Chemometrics and Intelligent Laboratory Systems* **2012,** *117*, 92-99.
26. Guo, S.; Kohler, A.; Zimmermann, B.; Heinke, R.; Stöckel, S.; Rösch, P.; Popp, J. r.; Bocklitz, T., Extended multiplicative signal correction based model transfer for raman spectroscopy in biological applications. *Analytical chemistry* **2018,** *90* (16), 9787-9795.
27. Smilde, A. K.; Jansen, J. J.; Hoefsloot, H. C.; Lamers, R.-J. A.; Van Der Greef, J.; Timmerman, M. E., ANOVA-simultaneous component analysis (ASCA): a new tool for analyzing designed metabolomics data. *Bioinformatics* **2005,** *21* (13), 3043-3048.
28. Agency, E. M., Guideline on bioanalytical method validation. *21 July* **2011**.
29. Miroslav, M., Peaks: Peaks. **2012**.
30. Press, W. H.; Teukolsky, S. A., Savitzky‐Golay smoothing filters. *Computers in Physics* **1990,** *4* (6), 669-672.
31. Bradley, M., Curve fitting in Raman and IR spectroscopy: basic theory of line shapes and applications. *Thermo Fisher Scientific, Madison, USA, Application Note* **2007,** *50733*.
32. E1840-96, A., Standard guide for Raman shift standards for spectrometer calibration. ASTM International: West Conshohocken, PA, 2014.
33. E2529-06, A., Standard Guide for Testing the Resolution of a Raman Spectromete. ASTM International: West Conshohocken, PA, 2014.
34. E2911-13, A., Standard Guide for Relative Intensity Correction of Raman Spectrometers. ASTM International: West Conshohocken, PA, 2013.
35. Guo, S.; Heinke, R.; Stöckel, S.; Rösch, P.; Bocklitz, T.; Popp, J., Towards an improvement of model transferability for Raman spectroscopy in biological applications. *Vibrational Spectroscopy* **2017,** *91*, 111-118.
36. Guo, S.; Heinke, R.; Stöckel, S.; Rösch, P.; Popp, J.; Bocklitz, T., Model transfer for Raman‐spectroscopy‐based bacterial classification. *Journal of Raman Spectroscopy* **2018,** *49* (4), 627-637.
37. Fornasaro, S.; Alsamad, F.; Baia, M.; Batista de Carvalho, L. s. A.; Beleites, C.; Byrne, H. J.; Chiadò, A.; Chis, M.; Chisanga, M.; Daniel, A., Surface Enhanced Raman Spectroscopy for quantitative analysis: results of a large-scale European multi-instrument interlaboratory study. *Analytical Chemistry* **2020,** *92* (5), 4053-4064.
38. Salek, R. M.; Neumann, S.; Schober, D.; Hummel, J.; Billiau, K.; Kopka, J.; Correa, E.; Reijmers, T.; Rosato, A.; Tenori, L., COordination of Standards in MetabOlomicS (COSMOS): facilitating integrated metabolomics data access. *Metabolomics* **2015,** *11* (6), 1587-1597.



TOC graphic


Raman device I
Raman device J