Investigating pre-analytical requirements for serum and plasma based infrared spectro-diagnostic

Lila Lovergne[1,2], Jean Lovergne[2], Pascaline Bouzy[1], Valérie Untereiner[1,3], Marc Offroy[1], Roselyne Garnotel[†1,4], Gérard Thiéfin[†1,5], Matthew J. Baker[†2*], and Ganesh D. Sockalingum[†1*]

[1]Université de Reims Champagne-Ardenne, EA7506-Biospectroscopie Translationnelle (BioSpec*T*), UFR de Pharmacie, 51 rue Cognacq-Jay, 51096 Reims Cedex, France.

[2]WESTChem, Department of Pure and Applied Chemistry, Technology and Innovation Centre, University of Strathclyde, Glasgow, G1 1RD, UK.

[3]Plateforme en Imagerie Cellulaire et Tissulaire (PICT), Université de Reims Champagne-Ardenne, 51 rue Cognacq-Jay, 51096 Reims Cedex, France.

[4]CHU de Reims, Laboratoire de Biochimie-Pharmacologie-Toxicologie, Hôpital Maison Blanche, 51092 Reims, France.

[5]CHU de Reims, Service d'Hépato-Gastroentérologie, Hôpital Robert Debré, 51092 Reims Cedex, France.

*Corresponding authors:

ganesh.sockalingum@univ-reims.fr, matthew.baker@strath.ac.uk

† Authors contributed equally to this work and project supervision

Keywords: FTIR spectroscopy, biofluids, serum, plasma, pre-analytical requirements, standardisation, water interference correction

Abstract

Infrared spectroscopy is a rapid, easy-to-operate, label-free and therefore cost-effective technique. Many studies performed on biofluids (e.g. serum, plasma, urine, sputum, bile, cerebrospinal fluid) have demonstrated its promising application as a clinical diagnostic tool. Given all these characteristics, infrared spectroscopy appears to be an ideal candidate to be implemented into the clinics. However, before considering its translation, a clear effort is needed to standardise protocols for biofluid spectroscopic analysis. To reach this goal, careful investigations to identify and track errors that can occur during the pre-analytical phase is a crucial step. Here, we report for the first time, results of investigations into pre-analytical factors that can affect the quality of the spectral data acquired on serum and plasma, such as the impact of long-term freezing time storage of samples as well as the month- to-month reproducibility of the spectroscopic analysis. The spectral data discrimination has revealed to be majorly impacted by a residual water content variation in serum and plasma dried samples.

1. Introduction

Infrared spectroscopy applied to biofluids presents several advantages, primarily as these are common media easily collectable and routinely analysed in clinical settings. Spectroscopy can provide diagnostic results within minutes, giving access to a full molecular profile of a patient sample rather than current techniques targeting a specific biomarker [1,2]. Therefore, spectroscopy could be an excellent competitive alternative to expensive and time-consuming methods [3–5]. Serum based spectroscopy has proved its potential as a clinical diagnostic tool in a number of proof-of-concept studies [6]. However, these reports have also highlighted the lack of consensus towards a standardised protocol regarding sample handling and preparation methodology in order to perform a spectroscopic analysis [7]. It is known that the pre-analytical phase is the greatest source of sample variance [8–11]. Related erroneous results are a waste of time and money but could also be detrimental to patient's health [12] and to the translation of promising spectroscopic techniques to the clinic. It is therefore crucial to identify and track errors that can occur in the pre-analytical phase.

Due to a high contribution of water in IR spectra, the most common protocol for the analysis of biofluids is the drying of drop deposits. It has been shown by optical and spectroscopic assessments that this dry-drop procedure, referenced in the literature as the coffee-ring effect

[13], is not homogenous [14–18]. In order to decrease the impact of this phenomenon, a study has recommended to perform a High Throughput-Fourier Transform Infrared (HT-FTIR) spectroscopy macroanalysis on a film composed of picoliter drops obtained by an automated sampling in combination with vacuum drying [19]. However, no IR spectroscopic study has investigated the influence of such inhomogeneous deposits on spectral data classification in the context of a patient study. The only spectroscopic study that dealt with this effect is a micro-Raman study showing that analysis of spectra acquired in centre and periphery of dried serum drops compared to freeze-dried serum spectra, has shown no significant difference in classification performances of patients with and without hepatocellular carcinoma [20].

A study by Attenuated Total Reflectance (ATR)-FTIR spectroscopy has reported that spectra acquired on pure serum drop cracks suffer from spectral distortions resulting in an abnormal amide I/ II band shape and baseline [16]. Using HT-FTIR spectroscopy in the transmission mode, we have previously shown the necessity to dilute serum samples to avoid amide bands spectral saturation, providing visually dried drops with greater homogeneity [21]. Moreover, one averaged spectrum is acquired over the whole surface of a silicon plate well. This ensures, with good reproducibility, the collection of a representative spectrum of the entire serum sample contrary to spectra collected in a single point mode. Other sample preparation parameters that can impact on the results of the spectroscopic analysis have been assessed such as the use of solvents to dilute the samples. Normal saline (NaCl 0.9%) is preferable to distilled water to avoid protein precipitation [22]. Lithium-heparin anti-coagulant tube is recommended for plasma obtention as spectra of plasma collected in ethylenediaminetetraacetic acid (EDTA) tubes presented interference peaks compared to serum spectra [22]. An ATR-FTIR study has reported a significant contribution of sodium citrate, EDTA and lithium heparin anticoagulants to dried plasma and whole blood spectra compared to wet samples. A better detection at low range of malaria parasitemia in red blood cells was achieved for samples prepared in lithium heparin tubes [23]. Our previous study has shown that serum and plasma samples were affected mainly in the protein region when fresh samples were compared to the first freeze-thaw cycle [22]. However, no further changes were observed with additional freeze-thaw cycles. Operator dependence and the day-to-day reproducibility assessment of the HT-FTIR spectroscopy technique have not highlighted any issue [22]. In the present study, modalities of the sample (serum and plasma) storage such as the use of different aliquot volumes and different tube capacities have been assessed. Long-term freezing time storage and month-to-month reproducibility has been spectroscopically evaluated to pursue the aim of standardising

protocols for a clinical translation. This study has highlighted a variation in residual water content of serum and plasma dried drops, correlated to the relative humidity recorded during experiments. Different approaches to correct these biofluid spectra for water interferences are discussed.

2. Material and Method

2.1 Blood-derived product samples

This study was performed on sera and plasma from 3 healthy subjects, obtained from routine biochemical check-up at the biochemistry laboratory of Reims university hospital. Freshly collected blood was let clot at room temperature (30 min) and centrifuged at 3000 rpm for 15 min to obtain serum. Blood collected in tubes containing lithium heparin as anticoagulant, was centrifuged at 3000 rpm for 15 min to obtain plasma. Samples were aliquoted *via* 3 different modalities, 50 µL in 0.2 mL or 1.5 mL capacity plastic tubes and 100 µL in 1.5 mL capacity plastic tubes to assess the impact of aliquot volume and tube capacity during a long-term storage. All aliquots were stored at -80°C until experiments were performed.

2.2 Experimental design

Each month during 9 months, 2 serum and 2 plasma sample aliquots per condition described above (volumes, tubes) were thawed and 3-fold diluted with normal saline (Miniversol, Aguettant, France) as previously reported [21,22]. For each aliquot, ten instrumental replicates of 5 µL were deposited onto a silicon plate (384 wells) and left to dry at room temperature (Figure 1) or under vacuum drying conditions for 90 minutes. In both cases, the temperature and relative humidity (RH) were recorded for each experiment M1-M9 (Table 1). Spectra of normal saline were also acquired as a reference with the same spectral acquisition parameters (Figure S1).

2.3 HT-FTIR spectral acquisition parameters

Spectra were collected with a high-throughput module (HTS-XT) coupled to a FTIR-spectrometer (Tensor 27, Bruker Optics GmbH, Ettlingen, Germany). The spectral acquisition was performed in the transmission mode (OPUS v.6.5 software, Bruker Optics GmbH) in the 4000-400 cm^{-1} range with a spectral resolution of 4 cm^{-1} and 32 co-added scans. For each sample, a background was recorded and automatically removed to obtain the final absorbance

spectrum. A spectral quality test (QT) as previously described [11,13] was applied to each spectrum, which resulted in 1.4% of serum spectra and 0.7% of plasma spectra that were discarded. All raw spectra that passed the quality test are presented in Figure S2. However, it has been observed that among them 3.2% of serum and 0.3% of plasma spectra, all from experiments performed with a RH above 50%, had a non-strictly increasing absorbance in the spectral region 3700-3320 cm^{-1} (Figure S3). This translates a high content of water in the sample replicates, and this level has been chosen as a maximum threshold. These spectra were computationally discarded based on their 2nd derivative form exhibiting peaks not present in other spectra at 3530, 3510, 3484, 3450 and 3420 cm^{-1}.

2.4 Spectral data analysis

Following the QT, spectral data were pre-processed using the OPUS v.6.5 software (Bruker Optics GmbH). Spectra were reduced to the spectral range 4000-800 cm^{-1}, a rubberband baseline correction [25] was applied prior to a vector normalisation. Then, serum and plasma spectral reproducibility between aliquots from each subject and the impact of volumes and tube capacity during long-term storage was assessed by Principal Component Analysis (PCA) [26]. To circumvent the confounding effect of residual water in the serum and plasma sample spectra, an Extended Multiplicative Signal Correction (EMSC) model [27], including a normal saline constituent spectrum, was tested on the spectral data based on the methodology described by Afseth *et al*. [28]. The mean spectrum of all normalised absorbance spectra was defined as the reference spectrum and the spectrum of normal saline was used as the interference spectrum to be removed from the dataset.

3. Results and discussion

3.1 Serum and plasma inter-aliquot spectral reproducibility

Once a month and over 9 months, 2 aliquots of serum and 2 aliquots of plasma samples (different volumes/tubes) per experiment were thawed and 3-fold diluted with normal saline. The spectral reproducibility of the 20 serum or plasma instrumental replicates deposited was assessed by PCA using the spectral ranges 4000-800 cm^{-1} (Figure 2). For all 3 subjects and each experimental condition (aliquot volume, tube capacity), the results did not show any clustering. Spectra were mixed, indicating no issue in the spectral reproducibility and therefore in sample handling before spectral analysis.

3.2 Impact of aliquot volume and tube capacity on spectral reproducibility

Careful considerations must be taken while using different types of plastic tubes for the biofluid storage due to the different adsorption characteristics of the tube wall that can impact the molecular content [29] and the possibility of a polymer contamination [30]. Potential spectral variability due to the use of different tube capacities and aliquot volumes were therefore tested at each time point using PCA over the spectral range 4000-800 cm^{-1}. Again, the comparison of the spectra of the different volumes/tubes per healthy subject showed no evidence of a specific clustering of the serum (Figure 3) and plasma (Figure S4) spectra. From these observations, it can be concluded that neither the volume of sample stored nor the storage tube capacity has an impact on spectral reproducibility.

3.3 Longitudinal study of serum and plasma spectral reproducibility over a 9-month period

In the first instance, we have compared serum and plasma spectra from the 3 subjects at each time point using PCA. Figure 4 shows that the spectra from the 3 subjects were well separated and the cluster pattern was reproducible over the 9 time points, with the spectra from subject 1 well distinct from those of subjects 2 and 3. Similar results were obtained with plasma samples from the same subjects (Figure S5). These results are consistent with conclusions of proteomics studies observing no significant changes in serum and plasma long-term storage at -80°C [31–33].

In addition to the above analysis, the longitudinal reproducibility was investigated using PCA by comparing all serum normalised absorbance spectra (Figure 5a) of the 3 subjects over the 9 months. The differentiation expected between the sera samples of the 3 subjects (as found in Figure 4 for each time point) is not observed. The data of the sera are mixed but however a cluster pattern shift is visible for spectra from each subject (Figure 5b). When the spectra are labelled timepoint-wise (M1-M9) instead of subject-wise, the scatter plot (Figure 5c) clearly shows that the data are separated based upon the experiment dates and that the shift is correlated to a gradient of RH recorded during the experiments (Table 1). This result suggests that there is a difference in water content in the instrumental replicates of the serum samples responsible for this shift between the time points. To confirm this hypothesis, PC loadings of the data were analysed. From the PC1 loading (Figure 5d), responsible for the data separation according to

the RH, a major negative band can be seen at 3470 cm^{-1} attributable to water OH bond vibrations [34]. Of interest, similar results were obtained from the plasma samples (Figure S6).

3.4 Modification of sample drying protocol

In an attempt to experimentally remove the spectral variance caused by water OH bond vibrations, serum and plasma samples were dried under vacuum conditions (90 min) in order to achieve a controlled and constant low RH prior to the spectral acquisition and compared with the same samples dried at RT (90 min) at different RH. Spectra were analysed using PCA. Figure 6a shows that the spectra of the 3 subject samples obtained under vacuum drying are still separated according to the time points (Figure 6b) following the same pattern as RT-dried samples although both datasets can be distinguished showing that the vacuum drying has an effect on the samples especially at 1655 cm^{-1} (amide I band). This result suggests that the samples were rehydrated before and/or during the spectral acquisition at the same RH recorded at room temperature despite the presence of a desiccant in the spectrometer. Similar results were obtained from plasma samples (Figure S7).

3.5 Water correction by PC removal and by EMSC modeling

From the PCA scatter plot (Figure 5c), it was found that PC1 is responsible for the data separation and is directly related to the RH recorded during the serum analysis. Indeed, PC1 loading (Figure 5d) shows a strong contribution of water at 3470 cm^{-1}. In order to remove the residual water interference from the serum and plasma spectral data, the PC1 contribution was removed from the normalised absorbance data and spectra were reconstructed with all remaining PCs. The original PCA contained all PCs (100% of the explained variance). The spectra obtained after correction are presented in Figure 7a where a less pronounced impact of the shoulder at 3600-3320 cm^{-1} is noticed. A PCA was then performed; the 3 different sera can be distinguished along the PC2 (Figure 7b) and the data are no longer separated following the RH recorded during experiments (Figure 7c). PC2 loading (Figure 7d), responsible for the serum sample separation, no longer exhibits a dominant water contribution. This correction approach allows to reduce the mean standard deviation (Table 2) by a factor of 16 over the spectral range 3600-3320 cm^{-1}, as well as by a factor of 2.5 over the spectral ranges 1720-1600 cm^{-1} (amide I band) and 1600-1500 cm^{-1} (amide II band). The “streaking” effect observed (e.g. Figure 5bc) potentially due to some residual baseline effects was removed by correcting the variation of residual water.

To take this analysis further, an EMSC approach integrating a constituent spectrum of normal saline (Figure S1) was also sought to correct the data [28]. The results are displayed in Figure 8. It can be observed in the 3600-3320 cm^{-1} region that this approach does not fully correct the water contribution (Figure 8a) even if the standard deviation was reduced by a factor 6 (Table 2). PCA results show that subject spectra are not distinguishable (Figure 8b) and that the data are separated based upon the date of analysis (Figure 8c). The PC1 loading exhibits remaining water spectral features in the 3600-3320 cm^{-1} region (Figure 8d). As the EMSC approach is improving the standard deviation of the spectra in the water region, a combination of an EMSC correction and PC1 (Figure 8d) removal was evaluated. The standard deviation of the spectra (Figure 9a) for the 3 zones assessed (Table 2), is comparable to the PC1 removal strategy (Figure 7). It can be observed that the data are now separated based upon subject spectral signatures (Figure 9b) and not on the RH recorded during experiments (Figure 9c). In addition, the separation appears better using the combination of EMSC and PC1 removal; the replicates inter-spectral variance are decreased, suggesting that other spectral regions than those controlled in Table 2 were affected by this correction compared to the PC1 removal approach alone (Figure 7). However, the PC2 loading features (Figure 9d) responsible for the subject-wise discrimination are similar to those obtained previously by removing only the PC1 (Figure 7d). Similar results were obtained with the plasma samples (Figures S8-S10, Table S1) although the sample separation was less pronounced. Contrary to serum, plasma also contains coagulation proteins due to the use of anticoagulant (here heparin). As water interacts with proteins, it is possible that dried plasma samples are more prone to water interferences. The lack of separation between plasma compared to serum samples could be explained by the fact that the correction applied is more efficient in removing interferences related to free water molecules rather than bound water molecules (amide I band).

4. Conclusions

This study has allowed to estimate and identify some pre-analytical phase errors for biofluid spectroscopy applications that have not been hitherto described. Our findings show that a serum/plasma long-term freezing period (-80°C) of storage has no apparent impact on the spectral signal of the samples analysed. Moreover, the use of different capacity plastic tubes for the biofluid storage has no influence on the spectra. However, while evaluating the month-to-month reproducibility, it was observed that the sample drying at room temperature resulted in a difference in the sample residual water content that was dependent on the environmental

relative humidity. Additional devices such as purging systems could be implemented to better control the ambient air humidity during sample drying and analysis but such implementation increases the total cost of the spectral analysis. Another approach is to directly correct the spectral data for different water content contribution using chemometrics. From the different methods evaluated, the best data correction performance was the combination of an EMSC model including a saline constituent spectrum and the removal of the PC carrying variance due to water interference. However, it must be noted that this methodology has been developed on a simple model of 3 samples to understand the spectral variations induced by environmental factors. Whether this approach can be efficiently applied to larger datasets with higher inter-individual variability remains to be evaluated.

In conclusion, to perform biofluid spectroscopy aimed at clinical translation, careful experiment design and most importantly pre-analytical phase error evaluation and tracking are mandatory to ensure targeting the correct molecular information rather than environmental interferences.

Supporting Information

Additional supporting information can be found in the online version of this article at the publisher's website.

Acknowledgments: The authors would like to acknowledge the Defence and Science Technology Laboratory (Dstl, UK) and the Direction Générale de l'Armement (DGA, France) as well as the Société d'Accélération du Transfert de Technologie (SATT) NORD for research fundings. The URCA PICT-IBiSA Technological Platform is also gratefully acknowledged for technical support.

References

[1] E. P. Diamandis, *J. Natl. Cancer Inst.* **2010**, *102*, 1462.

[2] P. D. Wagner, S. Srivastava, *Transl. Res.* **2012**, *159*, 343.

[3] A. L. Mitchell, K. B. Gajjar, G. Theophilou, F. L. Martin, P. L. Martin-Hirsch, *J. Biophotonics* **2014**, *7*, 153.

[4] K. Spalding, R. Board, T. Dawson, M. D. Jenkinson, M. J. Baker, *Brain Behav.* **2016**, *6*, e00502.

[5] E. Gray, H. J. Butler, R. Board, P. M. Brennan, A. J. Chalmers, T. Dawson, J. Goodden, W. Hamilton, M. G. Hegarty, A. James, M. D. Jenkinson, D. Kernick, E. Lekka, L. J. Livermore, S. J. Mills, K. O'Neill, D. S. Palmer, B. Vaqas, M. J. Baker, *BMJ Open* **2018**,

8, e017593.

[6] M. J. Baker, S. R. Hussain, L. Lovergne, V. Untereiner, C. Hughes, R. A. Lukaszweski, G. Thiéfin, G. D. Sockalingum, *Chem. Soc. Rev.* **2016**, *45*, 1803.

[7] H. J. Byrne, M. Baranska, G. J. Puppels, N. Stone, B. Wood, K. M. Gough, P. Lasch, P. Heraud, J. Sulé-Suso, G. D. Sockalingum, *Analyst* **2015**, *140*, 2066.

[8] P. Bonini, M. Plebani, F. Ceriotti, F. Rubboli, *Clin. Chem.* **2002**, *48*, 691.

[9] E. Schleicher, *Anal. Bioanal. Chem.* **2006**, *384*, 124.

[10] M. Plebani, *Ann. Clin. Biochem.* **2010**, *47*, 101.

[11] M. Plebani, *Clin. Chem. Lab. Med.* **2016**, *54*, 1881.

[12] S. F. Green, *Clin. Biochem.* **2013**, *46*, 1175.

[13] R. D. Deegan, O. Bakajin, T. F. Dupont, G. Huber, S. R. Nagel, T. A. Witten, *Nature* **1997**, *389*, 827.

[14] J. Filik, N. Stone, *Analyst* **2007**, *132*, 544.

[15] F. Bonnier, F. Petitjean, M. J. Baker, H. J. Byrne, *J. Biophotonics* **2014**, *7*, 167.

[16] C. Hughes, M. D. Brown, G. Clemens, A. Henderson, G. Monjardez, N. W. Clarke, P. Gardner, *J. Biophotonics* **2014**, *7*, 180.

[17] K. A. Esmonde-White, F. W. L. Esmonde-White, M. D. Morris, B. J. Roessler, *Analyst* **2014**, *139*, 2734.

[18] J. M. Cameron, H. J. Butler, D. S. Palmer, M. J. Baker, *J. Biophotonics* **2018**, *11*, 1.

[19] J. Ollesch, S. L. Drees, M. H. Heise, T. Behrens, T. Brüning, K. Gerwert, *Analyst* **2013**, *138*, 4092.

[20] I. Taleb, G. Thiéfin, C. Gobinet, V. Untereiner, B. Bernard-Chabert, A. Heurgué, C. Truntzer, P. Hillon, M. Manfait, P. Ducoroy, G. D. Sockalingum, *Analyst* **2013**, *138*, 4006.

[21] L. Lovergne, G. Clemens, V. Untereiner, R. A. Lukaszweski, G. D. Sockalingum, M. J. Baker, *Anal. Methods* **2015**, *7*, 7140.

[22] L. Lovergne, P. Bouzy, V. Untereiner, R. Garnotel, M. J. Baker, G. Thiéfin, G. D. Sockalingum, *Faraday Discuss.* **2016**, *187*, 521.

[23] M. Martin, D. Perez-Guaita, D. W. Andrew, J. S. Richards, B. R. Wood, P. Heraud, *Analyst* **2017**, *142*, 1192.

[24] X. Zhang, G. Thiéfin, C. Gobinet, V. Untereiner, I. Taleb, B. Bernard-Chabert, A. Heurgué, C. Truntzer, P. Ducoroy, P. Hillon, G. D. Sockalingum, *Transl. Res.* **2013**, *162*, 279.

[25] S. Wartewig, *IR and Raman Spectroscopy: Fundamental Processing*, WILEY-VCH

Verlag GmbH & Co. KGaA, Weinheim, **2003**.

[26] H. Abdi, L. J. Williams, *Wiley Interdiscip. Rev. Comput. Stat.* **2010**, *2*, 433.

[27] H. Martens, E. Stark, *J. Pharm. Biomed. Anal.* **1991**, *9*, 625.

[28] N. K. Afseth, A. Kohler, *Chemom. Intell. Lab. Syst.* **2012**, *117*, 92.

[29] A. Perret-Liaudet, M. Pelpel, Y. Tholance, B. Dumont, H. Vanderstichele, W. Zorzi, B. ElMoualij, S. Schraen, O. Moreaud, A. Gabelle, E. Thouvenot, C. Thomas-Anterion, J. Touchon, P. Krolak-Salmon, G. G. Kovacs, A. Coudreuse, I. Quadrio, S. Lehmann, *Clin. Chem.* **2012**, *58*, 787.

[30] S. Lista, F. Faltraco, H. Hampel, *Prog. Neurobiol.* **2013**, *101–102*, 18.

[31] B. L. Mitchell, Y. Yasui, C. I. Li, A. L. Fitzpatrick, P. D. Lampe, *Cancer Inform.* **2005**, *1*, 98.

[32] M. E. Hassis, R. K. Niles, M. N. Braten, M. E. Albertolle, E. H. Witkowska, C. A. Hubel, S. J. Fisher, K. E. Williams, *Anal. Biochem.* **2015**, *478*, 14.

[33] C. Ellervik, J. Vaught, *Clin. Chem.* **2015**, *61*, 914.

[34] S. Y. Venyaminov, F. G. Prendergast, *Anal. Biochem.* **1997**, *248*, 234.

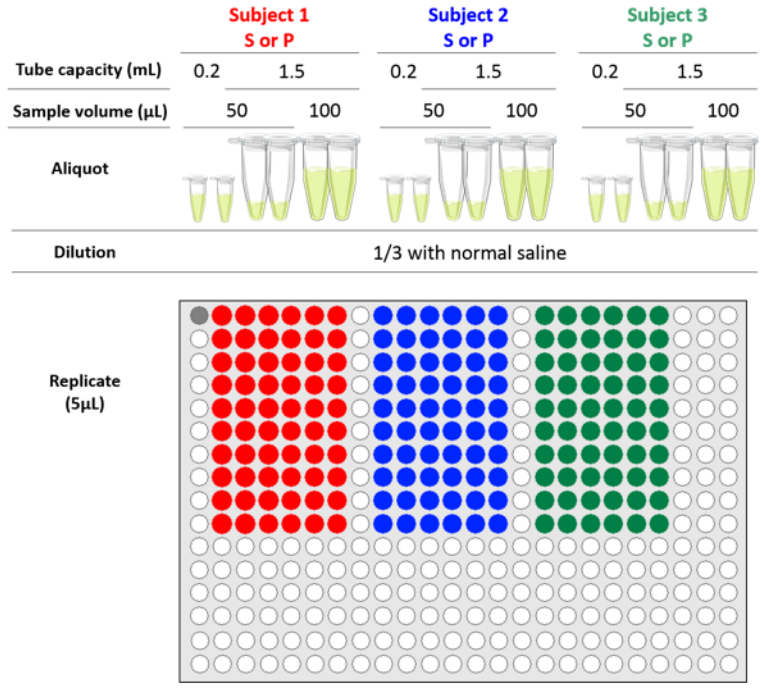


Figure 1. Schematic of the experimental study design. Serum (S) or plasma (P) replicates from 50 or 100 µL aliquots stored in 0.2 or 1.5 mL capacity plastic tubes were deposited onto an IR-transparent silicon plate (384 wells).

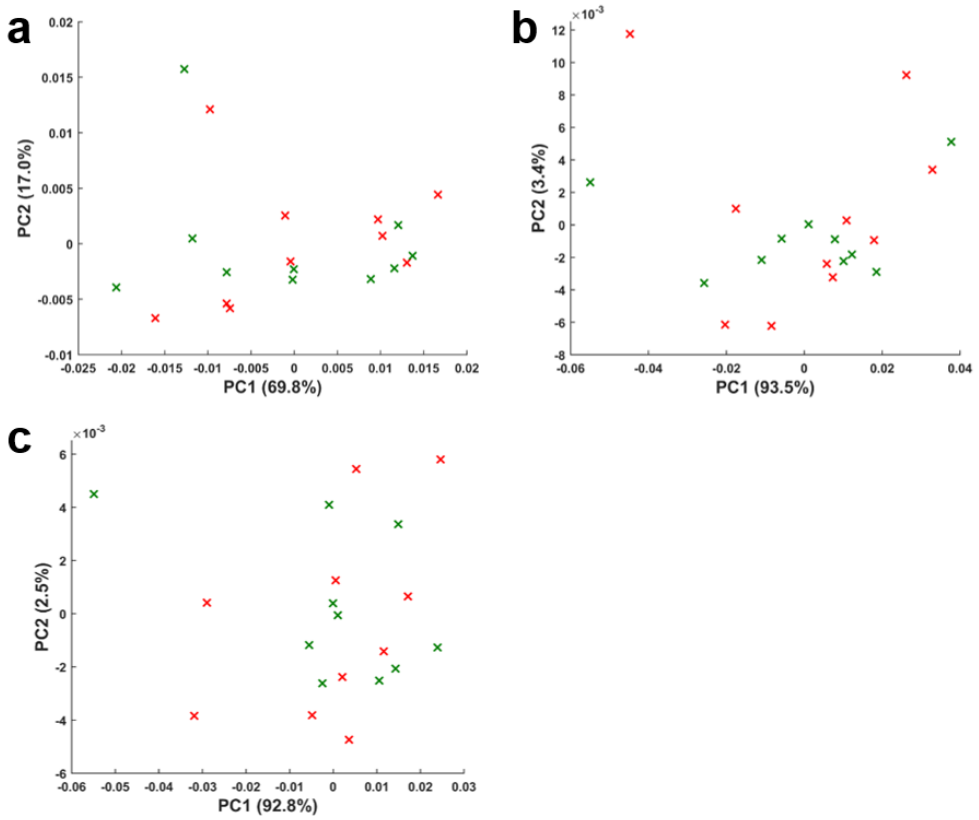


Figure 2. Assessment of inter-aliquot reproducibility. PCA of 50 µL of serum spectra over the spectral range 4000-800 cm^{-1} from subjects 1 (a), 2 (b) and 3 (c) stored 1 month in 0.2 mL capacity plastic tube. Instrumental replicate spectra of two aliquots for each subject are represented in red and green respectively.

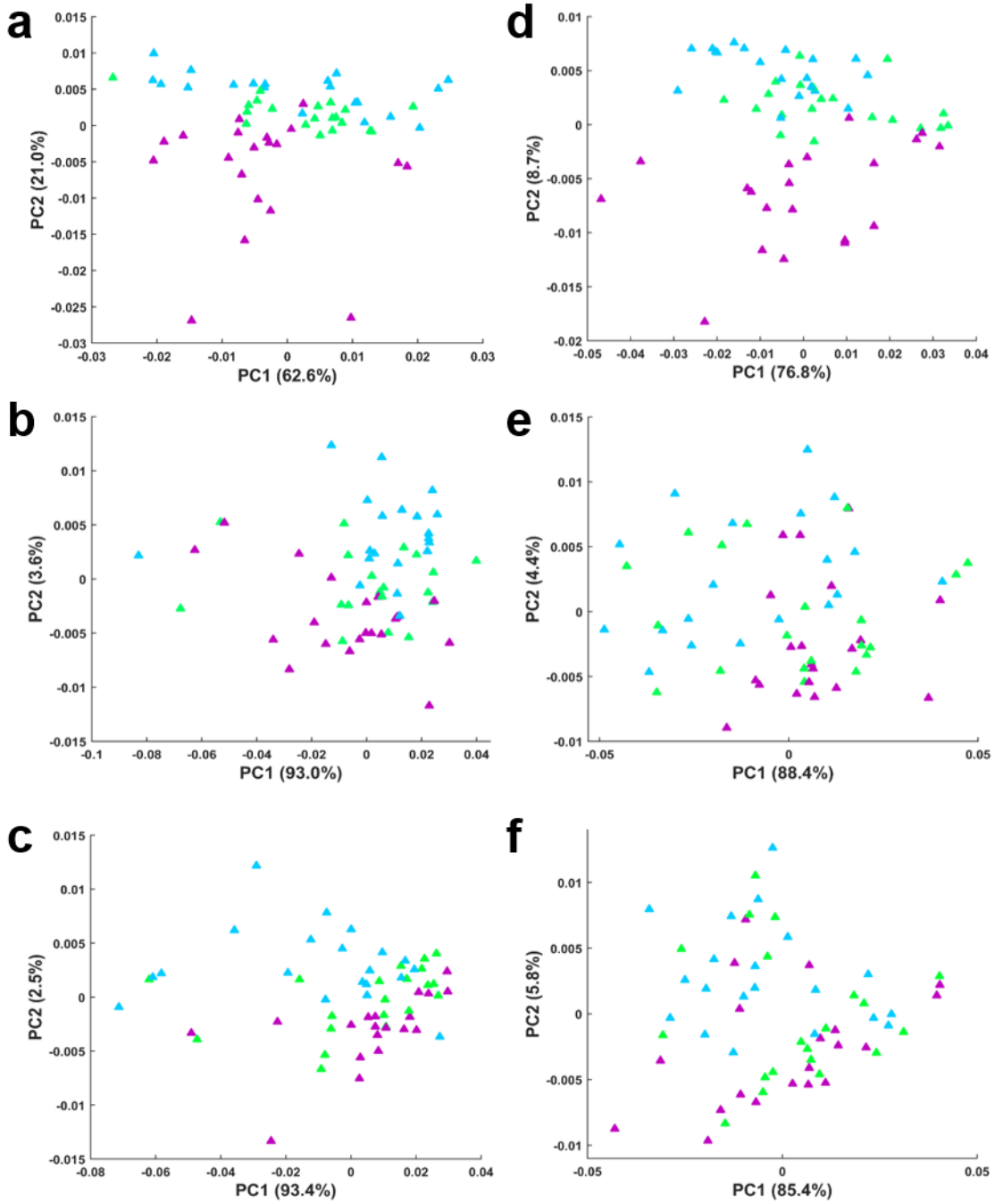


Figure 3. Impact of aliquot volume and tube capacity. PCA of serum spectra from subjects 1 (a, d), 2 (b, e), 3 (c, f) analysed after 1 month (a-c) and 6 months (d-f) of storage. Spectra of 50 µL of serum stored in 0.2 mL capacity tubes are represented by purple triangles. Spectra of 50 µL or 100 µL of serum stored in 1.5 mL capacity tubes are represented by green and blue triangles respectively.

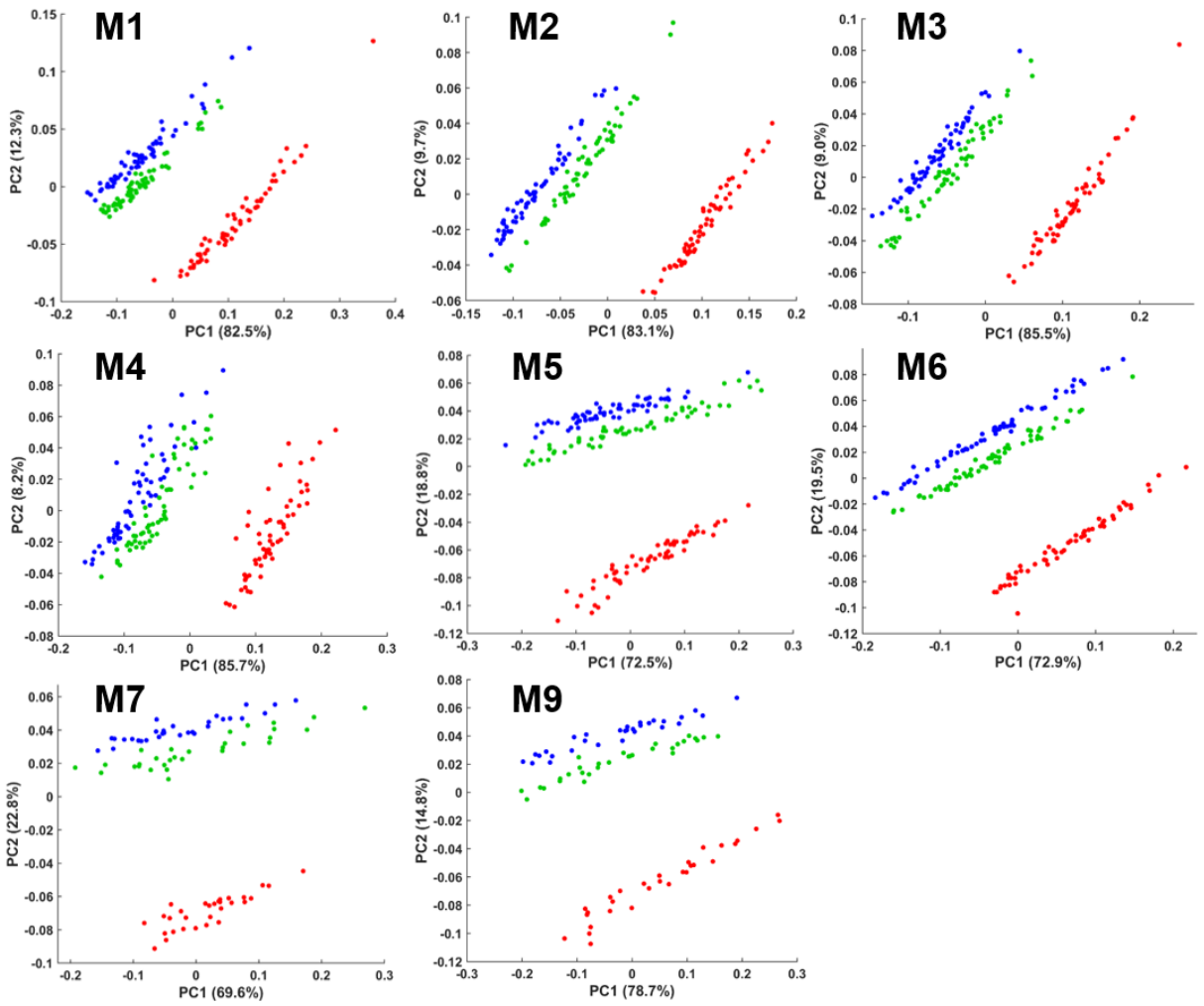


Figure 4. Assessment of the impact of a long-term storage. PCA performed on 3 subject serum spectra S1 (red), S2 (blue), S3 (green) in the spectral range 4000-800 cm^{-1} over a 9-month experiment (M1-M9). M8 was not included as the majority of the serum spectra from subjects 1 and 2 were discarded.

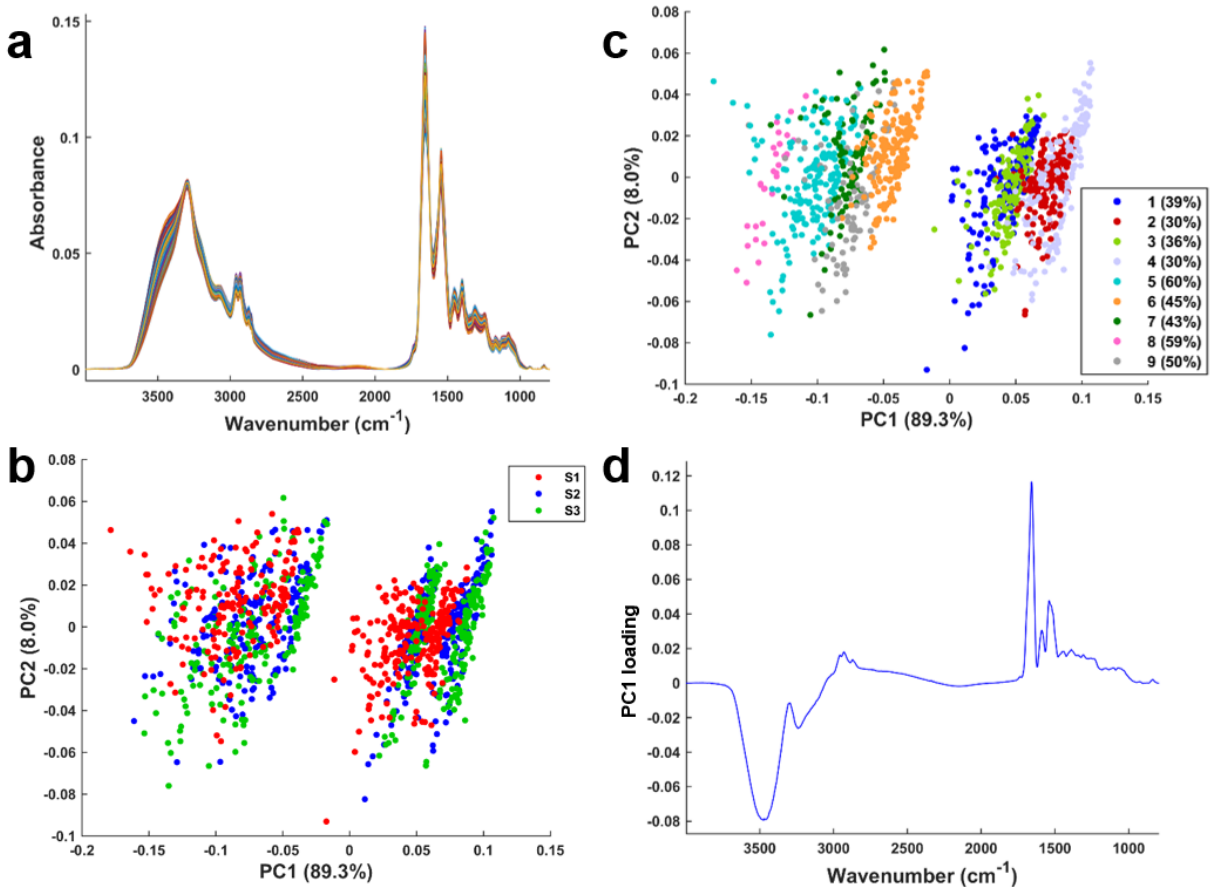


Figure 5. Month-to-month spectral data collection reproducibility. **a)** All normalised absorbance serum spectra of 3 subjects (S1, S2, S3). **b)** PCA scatter plot of the serum spectra presented subject-wise. **c)** PCA scatter plot of the same serum spectra presented timepoint-wise and highlighting associated relative humidity (%). **d)** PC1 loading from (b) and (c).

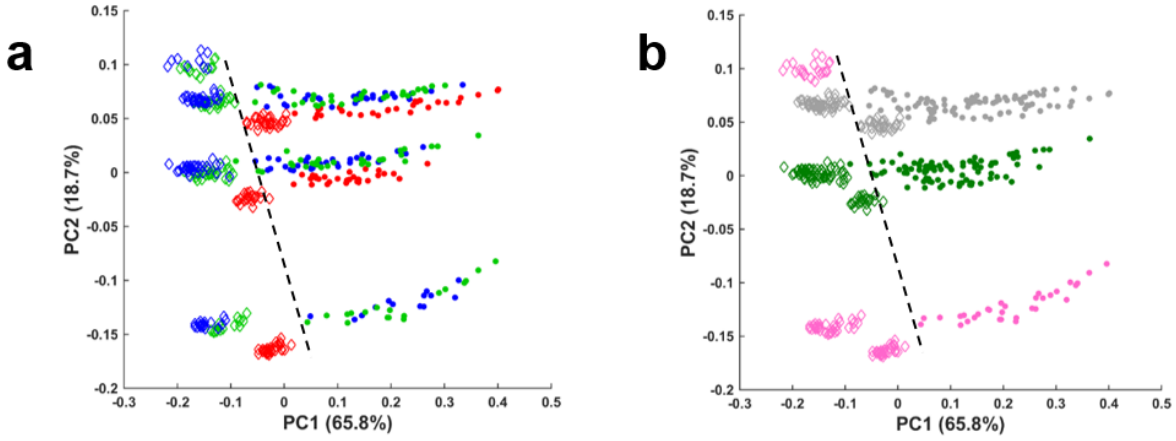


Figure 6. Serum spectral collection experimental correction by sample vacuum drying. **a)** PCA scatter plot of the 3 serum spectra (red, blue, green) obtained for air-dried samples at room temperature (dots) or for vacuum-dried samples (diamonds). **b)** PCA scatter plot of the same serum spectra highlighting the 3 experiments at a relative humidity of 43% (green), 50% (grey) and 59% (pink).

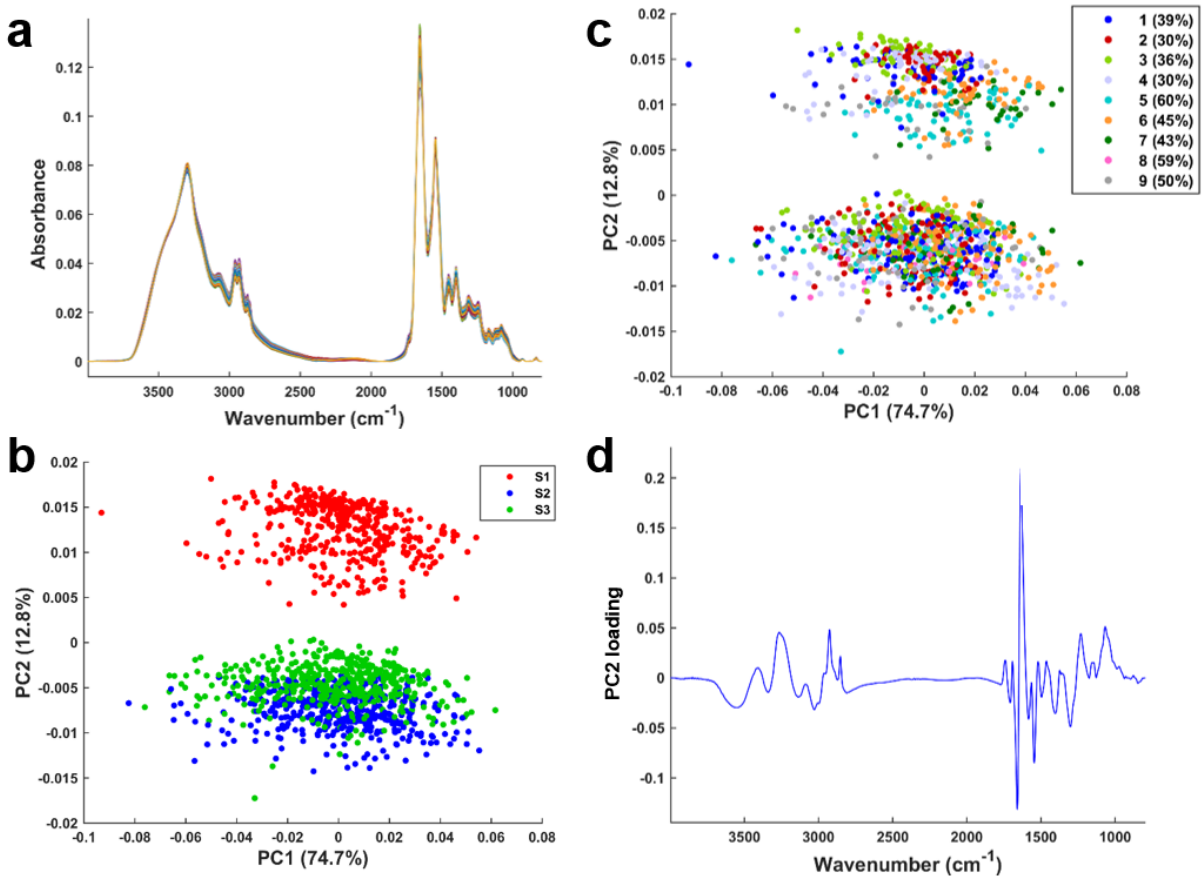


Figure 7. Spectral data water correction by PC removal. **a)** Reconstructed normalised absorbance serum spectra of the 3 subjects using all PCs except PC1 from Figure 5d. **b)** PCA scatter plot of reconstructed serum spectra shown subject-wise. **c)** PCA scatter plot of same reconstructed serum spectra shown timepoint-wise and highlighting associated relative humidity (%). **d)** PC2 loading from (b) and (c).

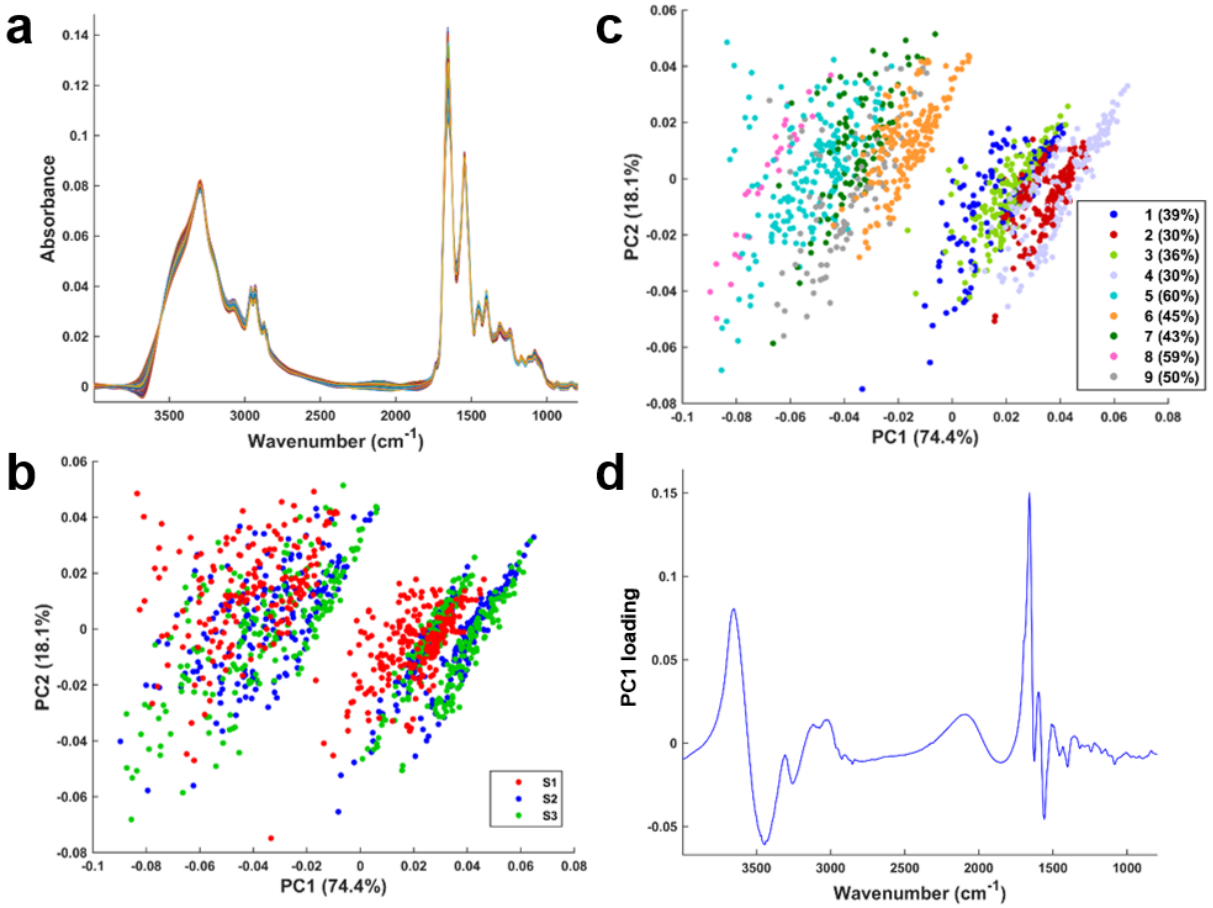


Figure 8. EMSC model for correcting water interference. **a)** EMSC-corrected normalised absorbance serum spectra using a normal saline constituent spectrum. **b)** PCA scatter plot of serum spectra displayed subject-wise. **c)** PCA scatter plot of the same serum spectra displayed timepoint-wise and highlighting the associated relative humidity (%). **d)** PC1 loading from (b) and (c).

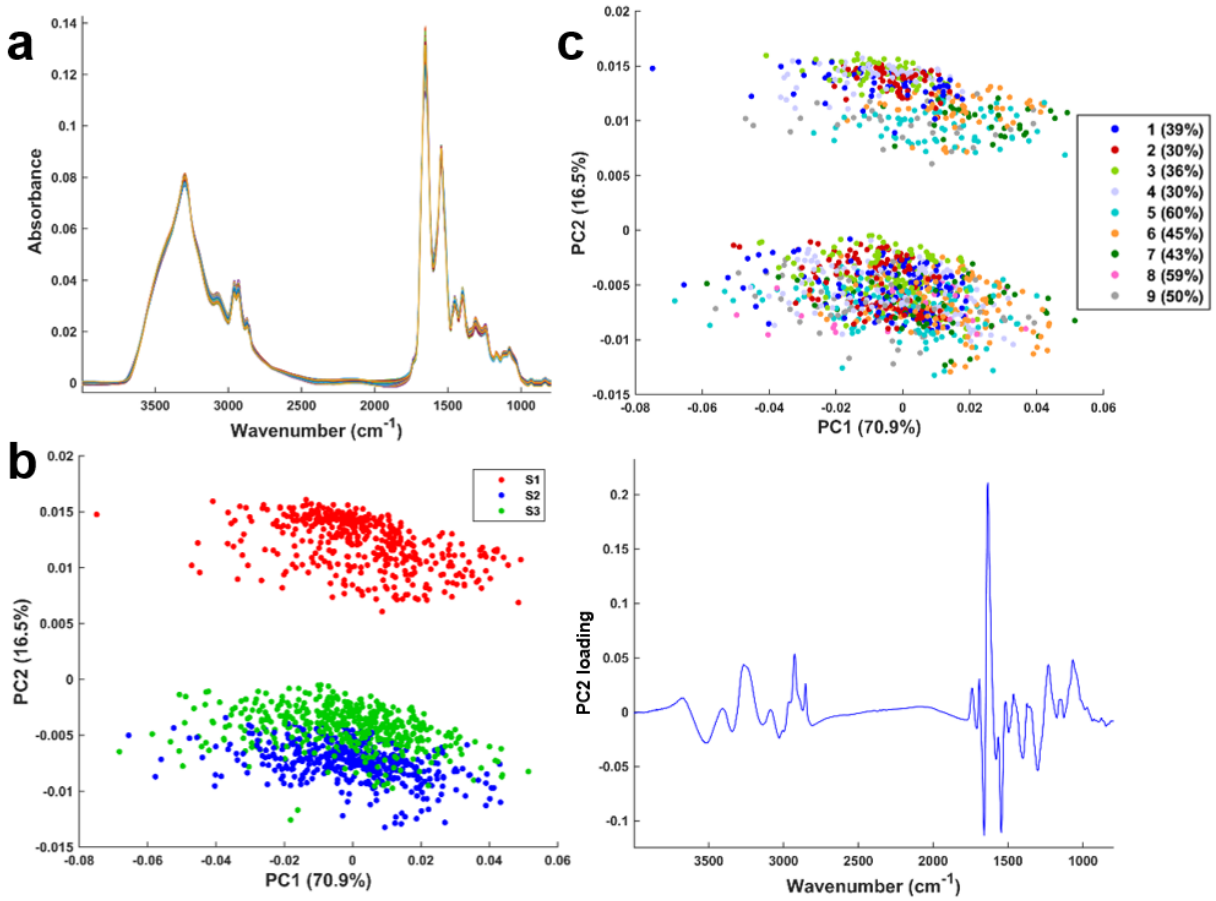


Figure 9. Combining EMSC model with normal saline constituent spectrum and PC removal. **a)** Normalised absorbance serum spectra after an EMSC correction using constituent normal saline spectrum and without the PC1 from Figure 8d. **b)** PCA scatter plot of serum spectra shown subject-wise. **c)** PCA scatter plot of the same serum spectra shown timepoint-wise and highlighting the associated relative humidity (%). **d)** PC2 loading from (b) and (c).

Table 1. Temperature and Relative Humidity (RH) recorded during the different experiments performed.

	RT drying		Vacuum drying	
Experiment	Temperature (°C)	RH (%)	Temperature (°C)	RH (%)
1	21.4	39	NA	NA
2	21.9	30	NA	NA
3	22.6	36	NA	NA
4	21.8	30	NA	NA
5	23.9	60	NA	NA
6	24.4	45	NA	NA
7	25.9	43	23.5	14
8	25.0	59	23.6	8
9	22.5	50	22.7	8

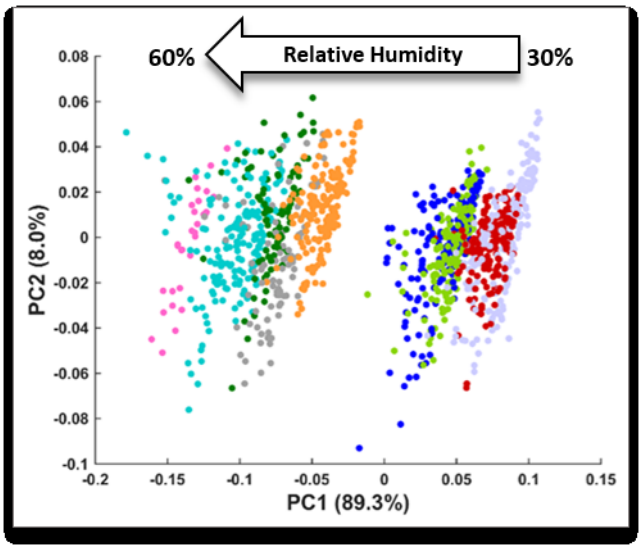
RT: room temperature

Table 2. Mean standard deviation (STD) calculated on 3 spectral regions according to different serum spectral water correction methods.

Correction	Mean STD		
	3600-1320 cm^{-1}	1720-1600 cm^{-1}	1600-1500 cm^{-1}
No correction	$4.47e^{-03}$	$3.84e^{-03}$	$2.71e^{-03}$
PC removal	$2.80e^{-04}$	$1.53e^{-03}$	$1.09e^{-03}$
$EMSC_{NS}$	$8.03e^{-04}$	$2.70e^{-03}$	$1.30e^{-03}$
$EMSC_{NS}$+ PC removal	$3.70e^{-04}$	$1.40e^{-03}$	$9.55e^{-04}$

NS: normal saline

Graphical Abstract



Principal component analysis plot showing the separation of infrared spectra of the same serum samples based on the relative humidity recorded during a 9-month experiment. This result demonstrates the importance of tracking potential pre-analytical phase errors to ensure spectral data collection reproducibility and to avoid the introduction of bias in the data analysis interpretation.

Supporting Information

Investigating pre-analytical requirements for serum and plasma based infrared spectro-diagnostic

Lila Lovergne[1,2], Jean Lovergne[2], Pascaline Bouzy[1], Valérie Untereiner[1,3], Marc Offroy[1], Roselyne Garnotel[†1,4], Gérard Thiéfin[†1,5], Matthew J. Baker[†2*], and Ganesh D. Sockalingum[†1*]

[1]Université de Reims Champagne-Ardenne, EA7506-Biospectroscopie Translationnelle (BioSpec*T*), UFR de Pharmacie, 51 rue Cognacq-Jay, 51096 Reims Cedex, France.

[2]WESTChem, Department of Pure and Applied Chemistry, Technology and Innovation Centre, University of Strathclyde, Glasgow, G1 1RD, UK.

[3]Plateforme en Imagerie Cellulaire et Tissulaire (PICT), Université de Reims Champagne-Ardenne, 51 rue Cognacq-Jay, 51096 Reims Cedex, France.

[4]CHU de Reims, Biochimie-Pharmacologie-Toxicologie, Hôpital Maison Blanche, 51092 Reims, France.

[5]CHU de Reims, Service d'Hépato-Gastroentérologie, Hôpital Robert Debré, 51092 Reims Cedex, France.

*Corresponding authors:

ganesh.sockalingum@univ-reims.fr, matthew.baker@strath.ac.uk

[†] Authors contributed equally to this work and project supervision

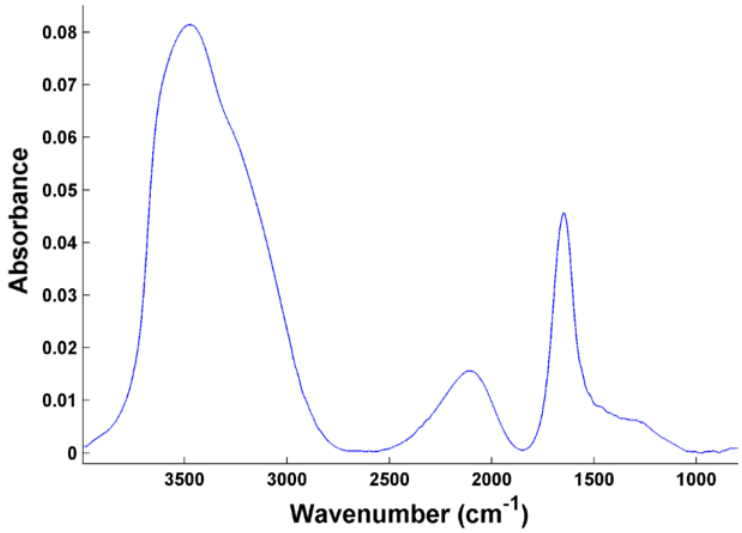


Figure S1. Normal saline normalised absorbance spectrum.

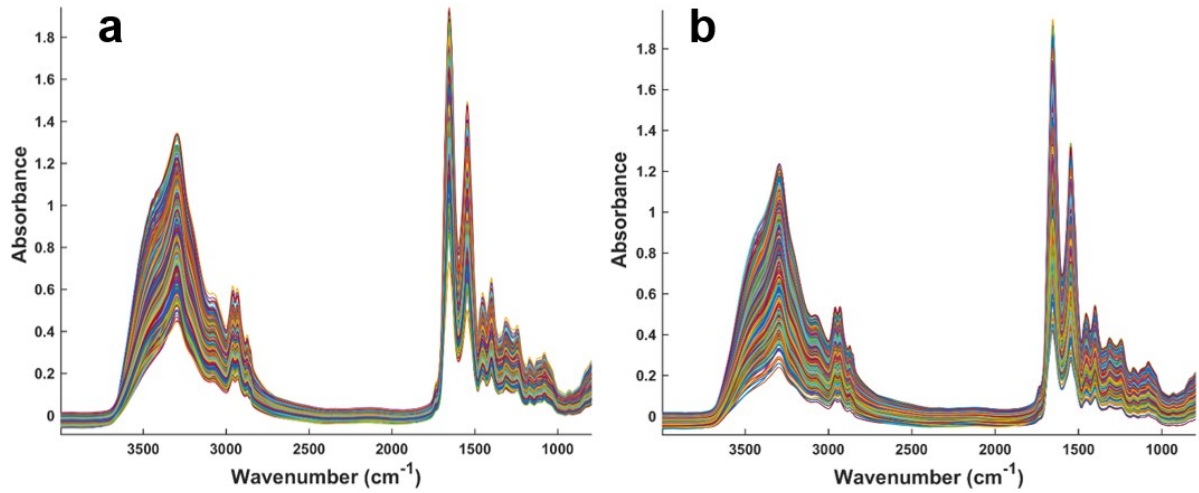


Figure S2. All positive quality tested serum (a) and plasma (b) raw spectra.

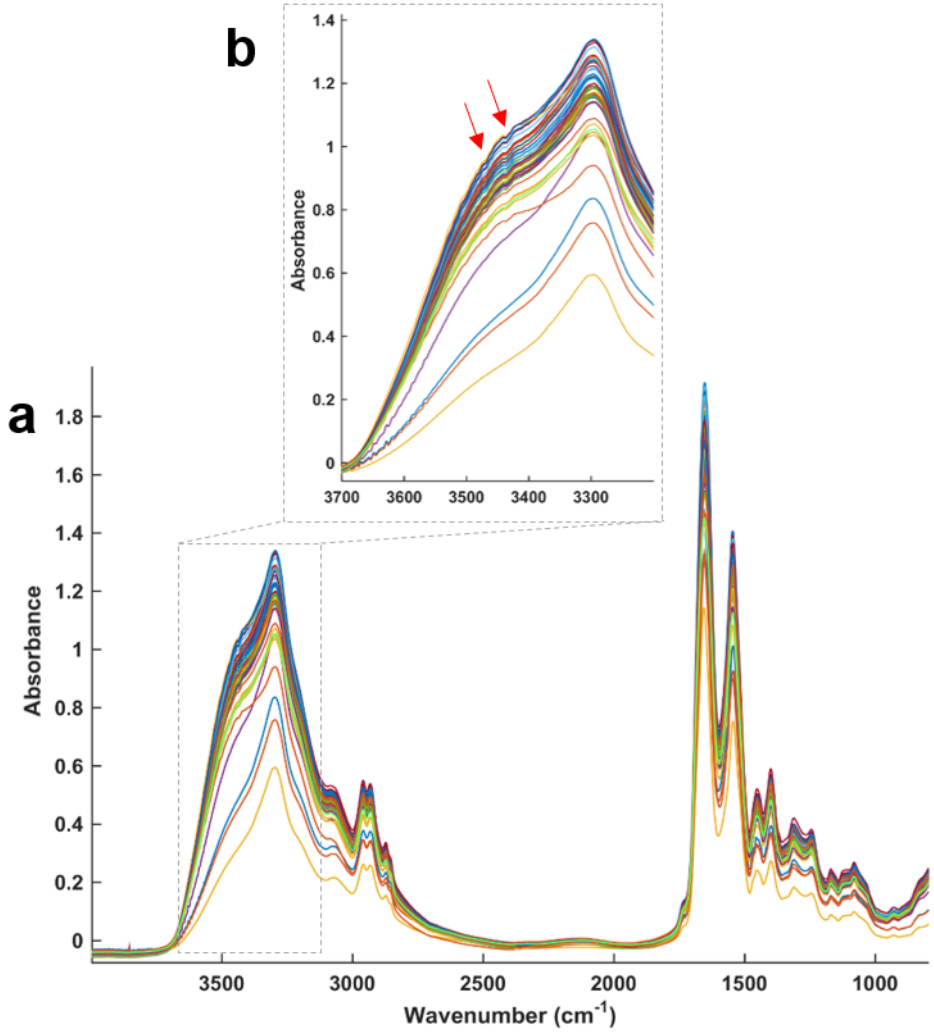


Figure S3. a) Spectra removed from the study. **b)** Zoom in the spectral region 3700-3200 cm^{-1}, red arrows indicate a non-strictly increasing absorbance.

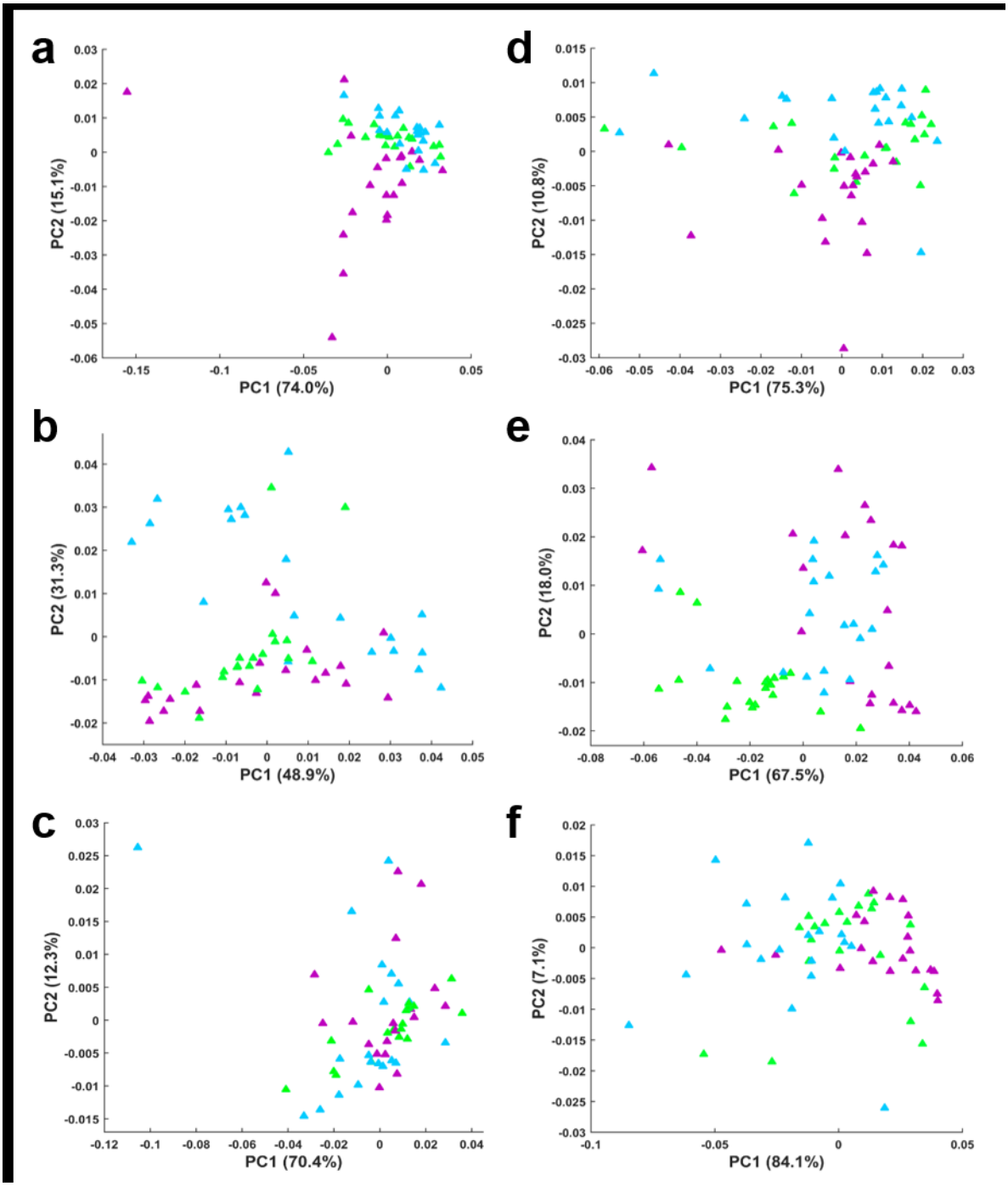


Figure S4. Impact of aliquot volume and tube capacity. PCA of plasma spectra from subjects 1 (a, d), 2 (b, e), 3 (c, f) analysed after 1 month (a-c) and 6 months (d-f) of storage. Spectra of 50 µL of plasma stored in 0.2 mL capacity tubes are represented by purple triangles. Spectra of 50 µL or 100 µL of plasma stored in 1.5 mL capacity tubes are represented by green and blue triangles respectively.

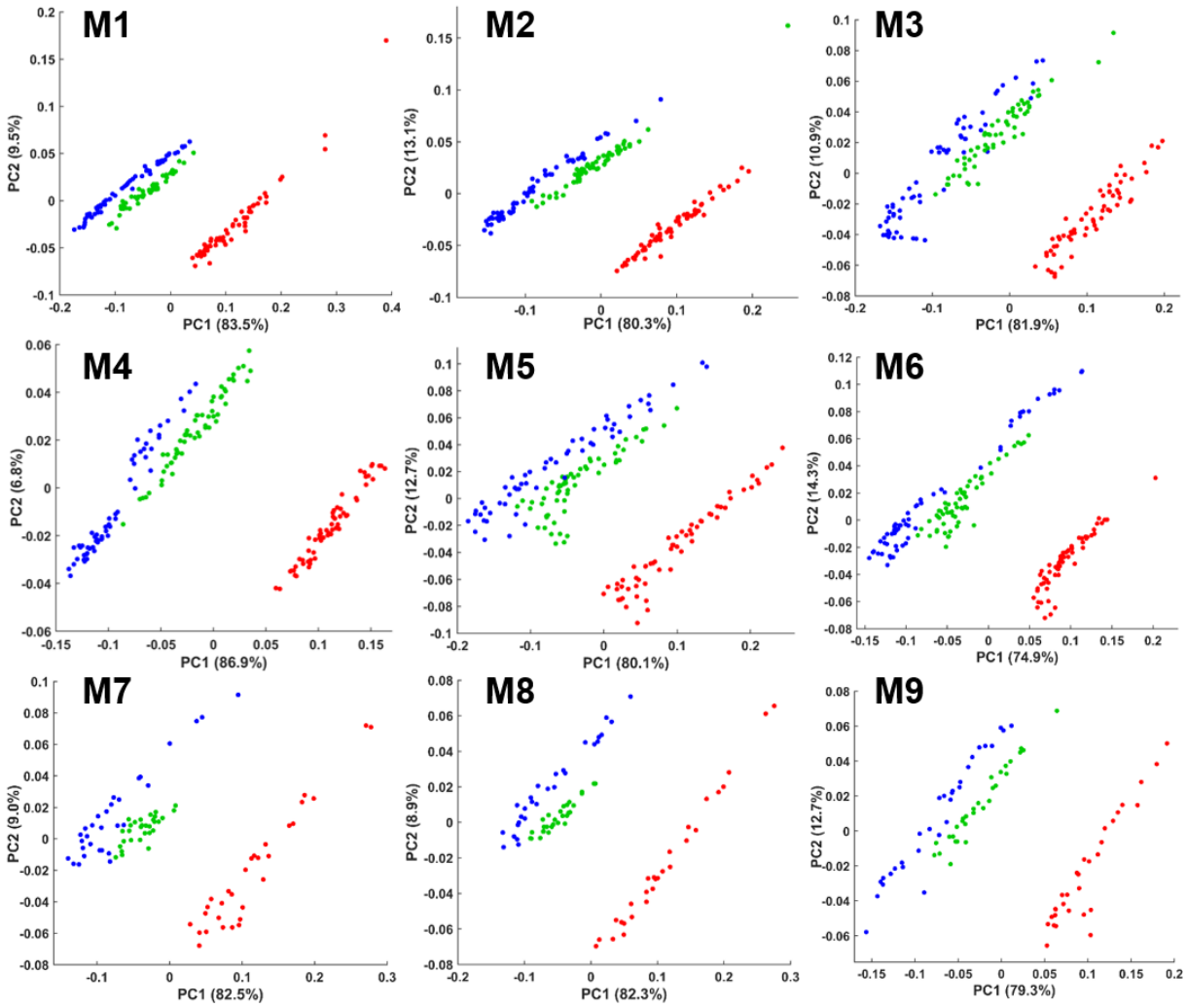


Figure S5. Assessment of the impact of a long-term storage. PCA performed on 3 subject plasma spectra P1 (red), P2 (blue), P3 (green) in the spectral range 4000-800 cm^{-1} over a 9-month experiment (M1-M9).

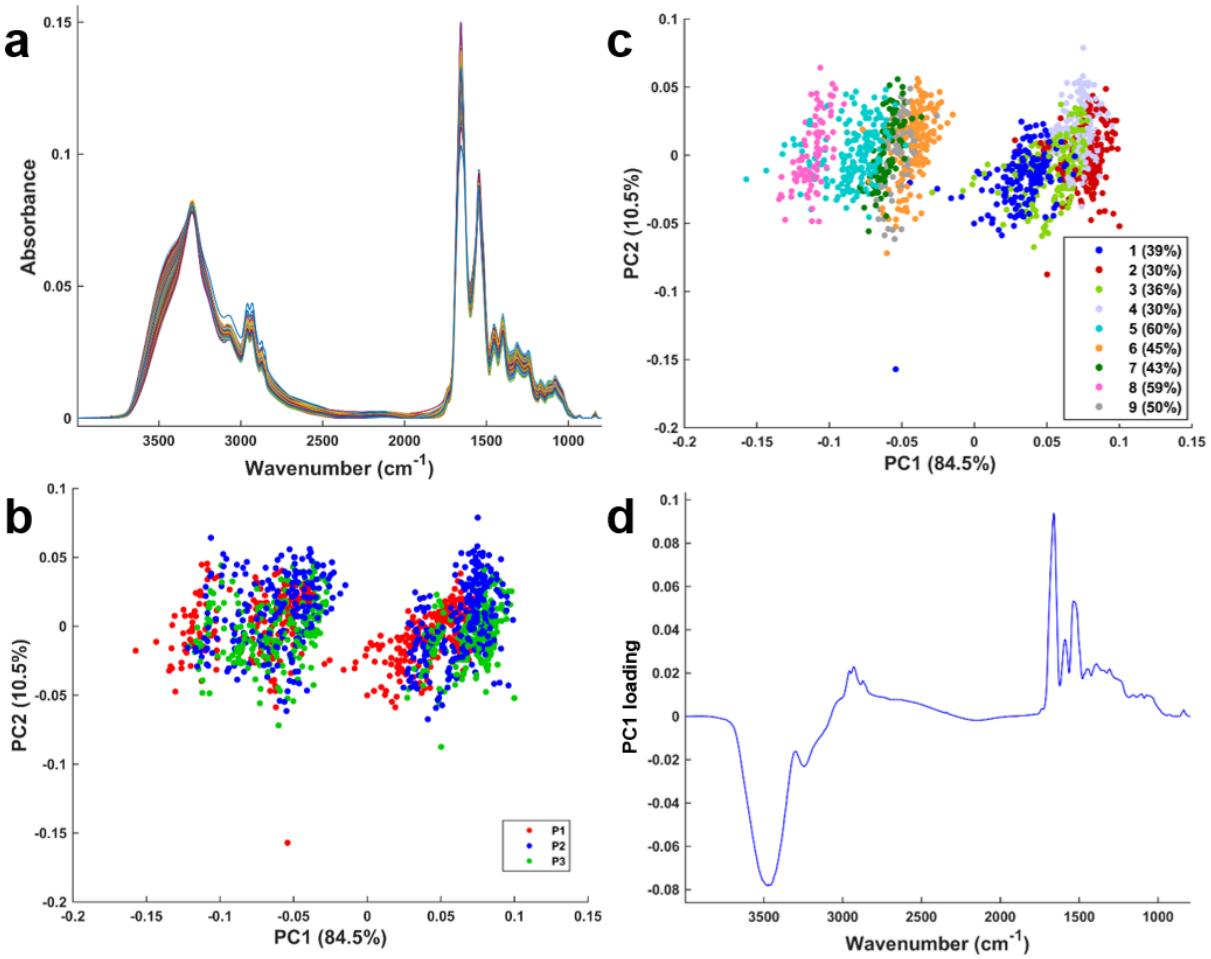


Figure S6. Month-to-month spectral data collection reproducibility. **a)** All normalised absorbance plasma spectra of 3 subjects (P1, P2, P3). **b)** PCA scatter plot of the plasma spectra presented subject-wise. **c)** PCA scatter plot of the same plasma spectra presented timepoint-wise and highlighting associated relative humidity (%). **d)** PC1 loading from (b) and (c).

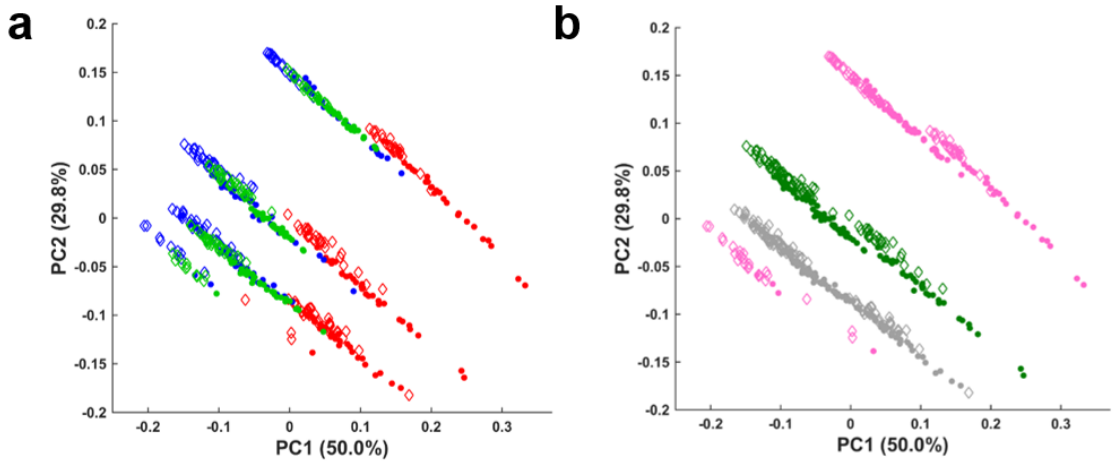


Figure S7. Plasma spectral collection experimental correction by sample vacuum drying. **a)** PCA scatter plot of the 3 plasma spectra (red, blue, green) obtained for air-dried samples at room temperature (dots) or for vacuum-dried samples (diamonds). **b)** PCA scatter plot of the same plasma spectra highlighting the 3 experiments at a relative humidity of 43% (green), 50% (grey) and 59% (pink).

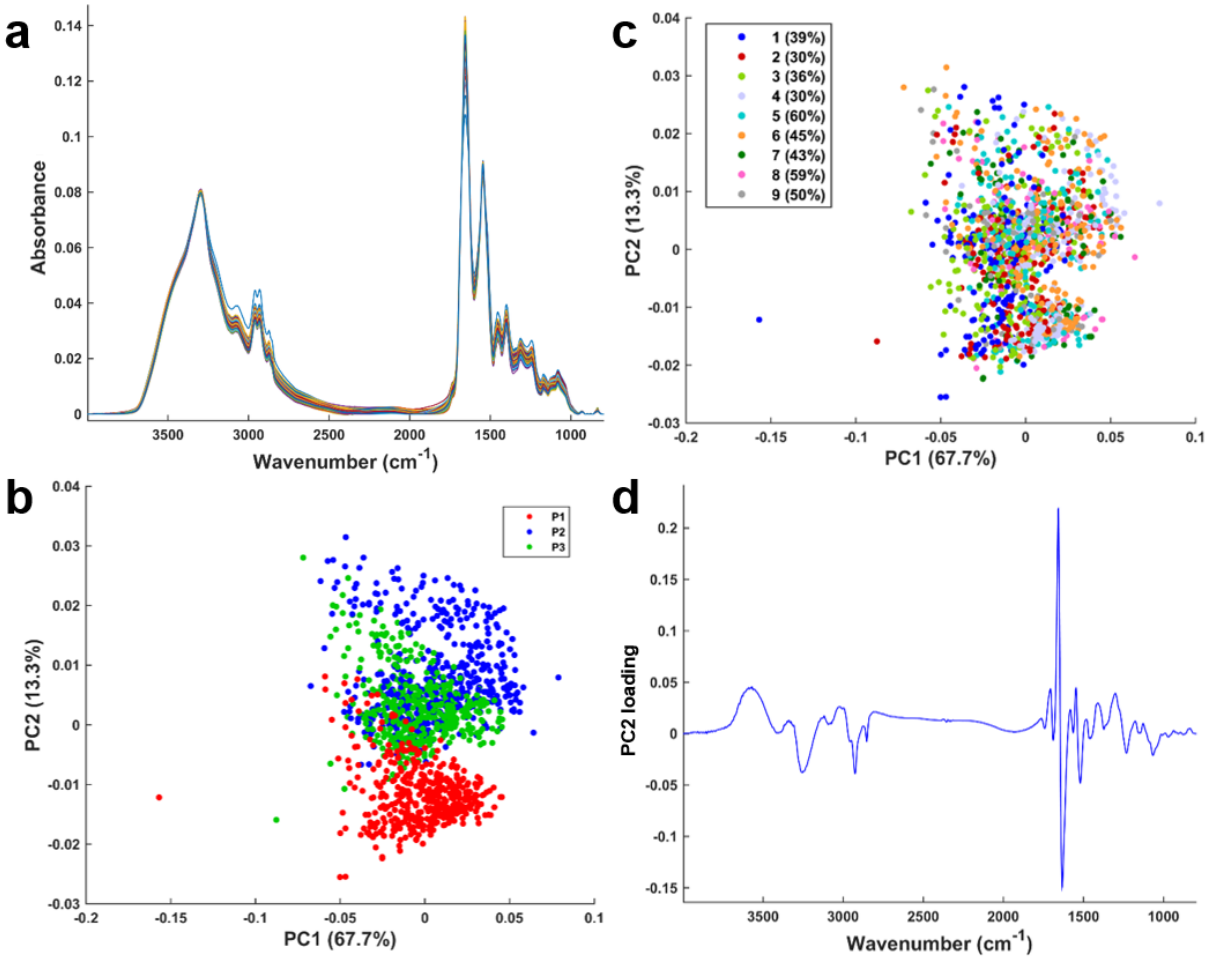


Figure S8. Spectral data water correction by PC removal. **a)** Reconstructed normalised absorbance plasma spectra of the 3 subjects using all PCs except PC1 from Figure S6d. **b)** PCA scatter plot of reconstructed plasma spectra shown subject-wise. **c)** PCA scatter plot of same reconstructed plasma spectra shown timepoint-wise and highlighting associated relative humidity (%). **d)** PC2 loading from (b) and (c).

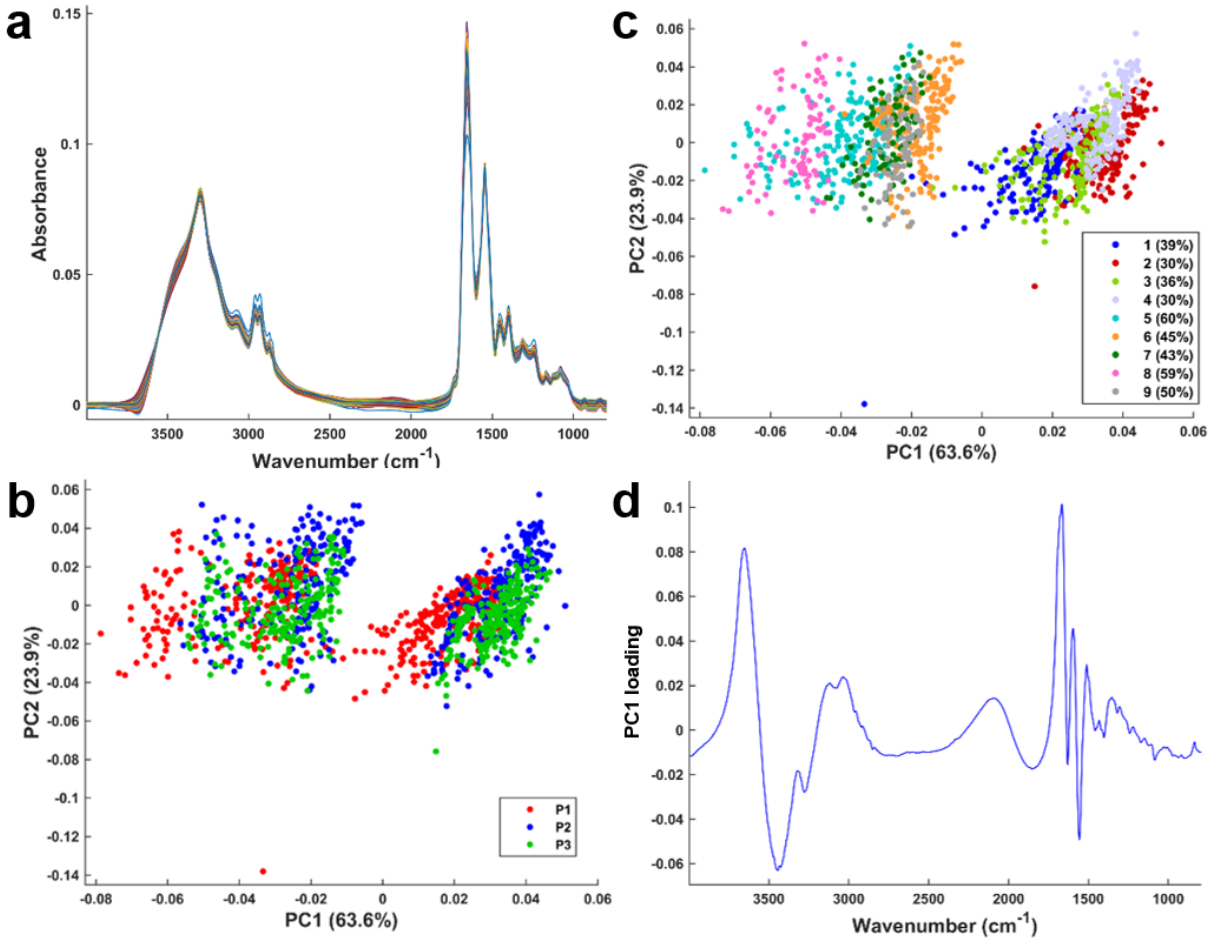


Figure S9. EMSC model for correcting water interference. **a)** EMSC-corrected normalised absorbance plasma spectra using a normal saline constituent spectrum. **b)** PCA scatter plot of plasma spectra displayed subject-wise. **c)** PCA scatter plot of the same plasma spectra displayed timepoint-wise and highlighting the associated relative humidity (%). **d)** PC1 loading from (b) and (c).

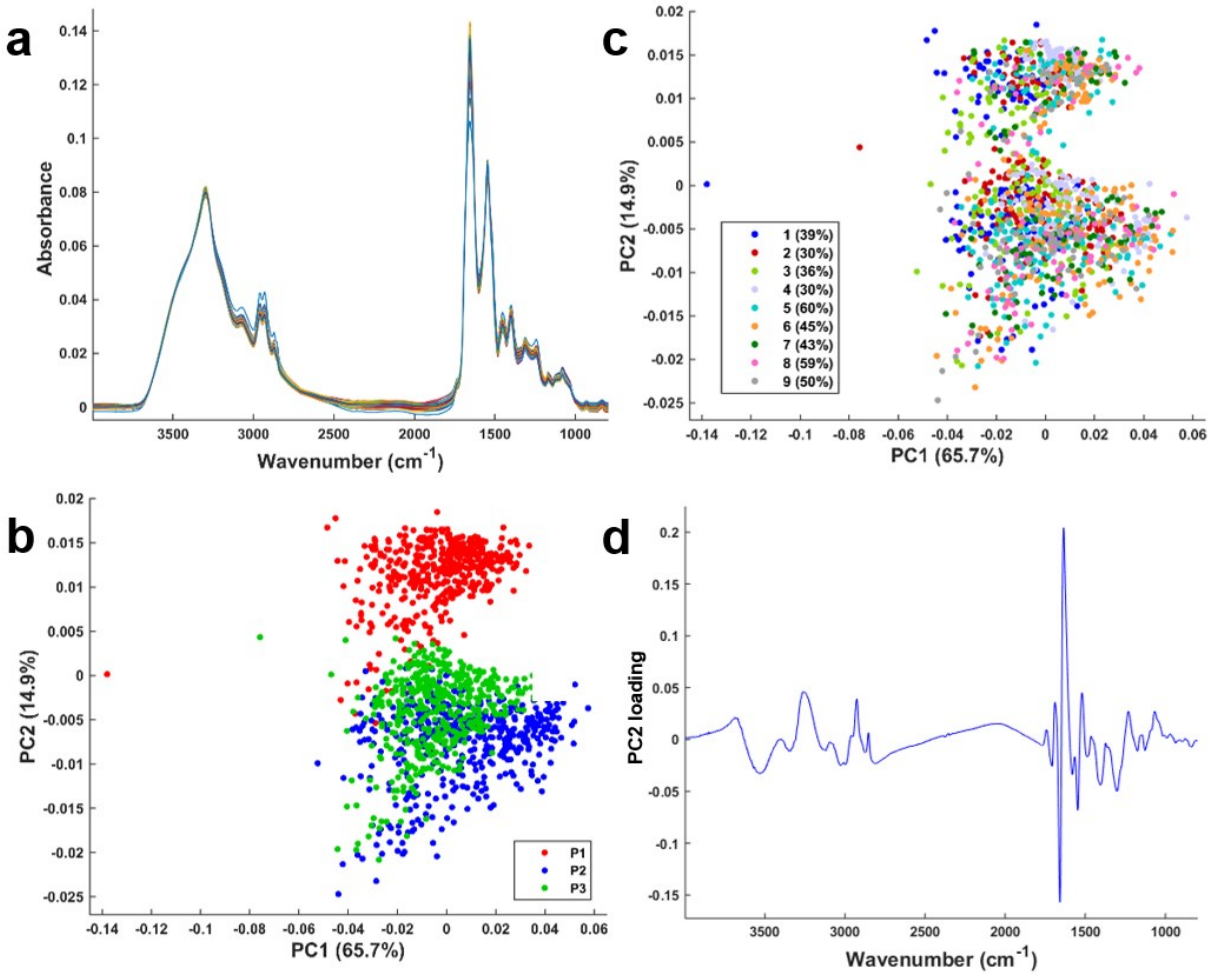


Figure S10. Combining EMSC model with normal saline constituent spectrum and PC removal. **a)** Normalised absorbance plasma spectra after an EMSC correction using constituent normal saline spectrum and without the PC1 from Figure S9d. **b)** PCA scatter plot of plasma spectra shown subject-wise. **c)** PCA scatter plot of the same plasma spectra shown timepoint-wise and highlighting the associated relative humidity (%). **d)** PC2 loading from (b) and (c).

Table S1. Mean standard deviation (STD) calculated on 3 spectral regions according to different plasma spectral water correction methods.

Correction	Mean STD		
	3600-1320 cm^{-1}	1720-1600 cm^{-1}	1600-1500 cm^{-1}
No correction	$4.13e^{-03}$	$3.46e^{-03}$	$2.79e^{-03}$
PC removal	$4.76e^{-04}$	$1.59e^{-03}$	$9.87e^{-04}$
$EMSC_{NS}$	$8.27e^{-04}$	$2.40e^{-03}$	$1.31e^{-03}$
$EMSC_{NS}$ + PC removal	$2.96e^{-04}$	$1.50e^{-03}$	$8.52e^{-04}$

NS: normal saline