Electronic Supplementary Information (ESI)

Predictive modelling of water contact angle of polymeric surfaces using Attenuated Total Reflection – Fourier Transform Infrared (ATR-FTIR) chemical imaging and Partial Least Squares Regression (PLSR)

S. Mukherjee[a], J.Á. Martínez-González[a,b,1], D.P. Dowling[c], A.A. Gowen[a]

[a]*School of Biosystems and Food Engineering, University College Dublin, Belfield, Dublin 4, Ireland.*

[b]*School of Chemical and Bioprocess Engineering, University College Dublin, Belfield, Dublin 4, Ireland*

[c]*School of Electrical, Electronic and Mechanical Engineering, University College Dublin, Belfield, Dublin 4, Ireland*

Contents

---

[1] Corresponding author: jose.martinez-gonzalez@ucd.ie

S.1 Outlier Removal (example Polytetrafluoroethylene or PTFE)

First, we proceed to visualise the dry spectra for PTFE collected from three separate replicates and specific spectral regions as shown in Figure S.1.

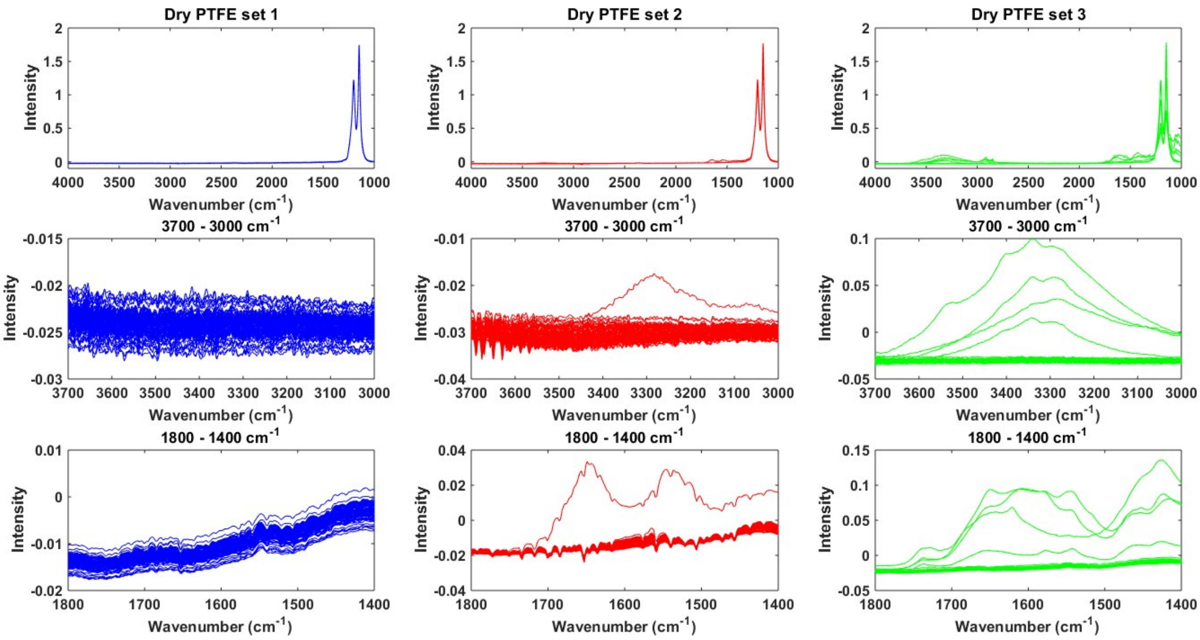

Figure S.1 This figure shows the dry PTFE spectra from 3 replicates or sets (top panel) and spectral regions of interest (3700-3000cm^{-1}, 1800 - 1400 cm^{-1}) in the lower panels.

The outlier in these cases, appear to have different spectral profiles, in comparison to the rest of the spectra, specifically for Set 2 & Set 3. To visualise the outlier spatially, we proceed to apply Principal Component Analysis (PCA[1]) on to the 3 sets of dry PTFE in the spectral range of 3700 – 3000 cm^{-1}, and then visualise the first Principal Component score image as well as a histogram to identify the outlier pixel(s). We can see the results for this in Figure S.2. In this case, only a single pixel (green) appears to be very different from the other pixels (PC 1 score image; Dry PTFE; Set 2), and the same is highlighted in the spectral window (Dry PTFE; Set 2: 3700 – 3000 cm^{-1}). Based on the histogram, we choose a value of '0.04' (indicated by a dotted red line, in Figure S.2) to mask out Dry PTFE; Set 2, but apply the binary mask to the entire spectrum, to remove 1 pixel. A similar process is applied to Dry PTFE; Set 3, but we remove 4 pixels, based on the score image and the histogram using a score threshold value of '0.1' as seen in Figure S.3.

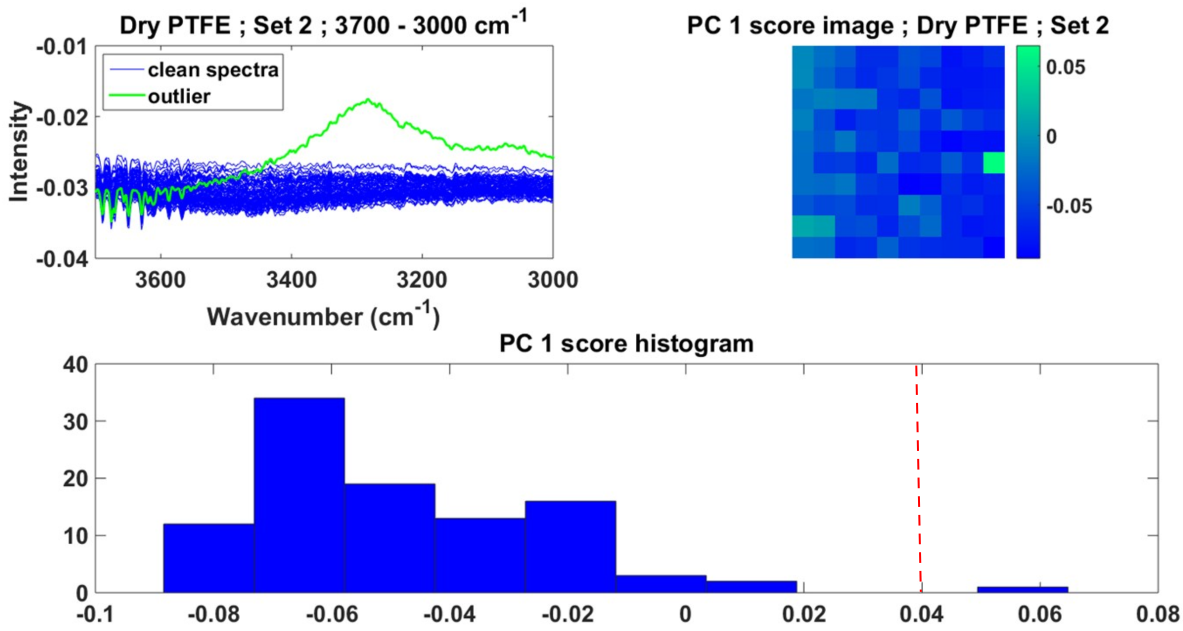

Figure S.2 Outlier removal for dry PTFE Set 2 (top left) using the PC 1 score image (top right) and a PC 1 score value histogram (lower).

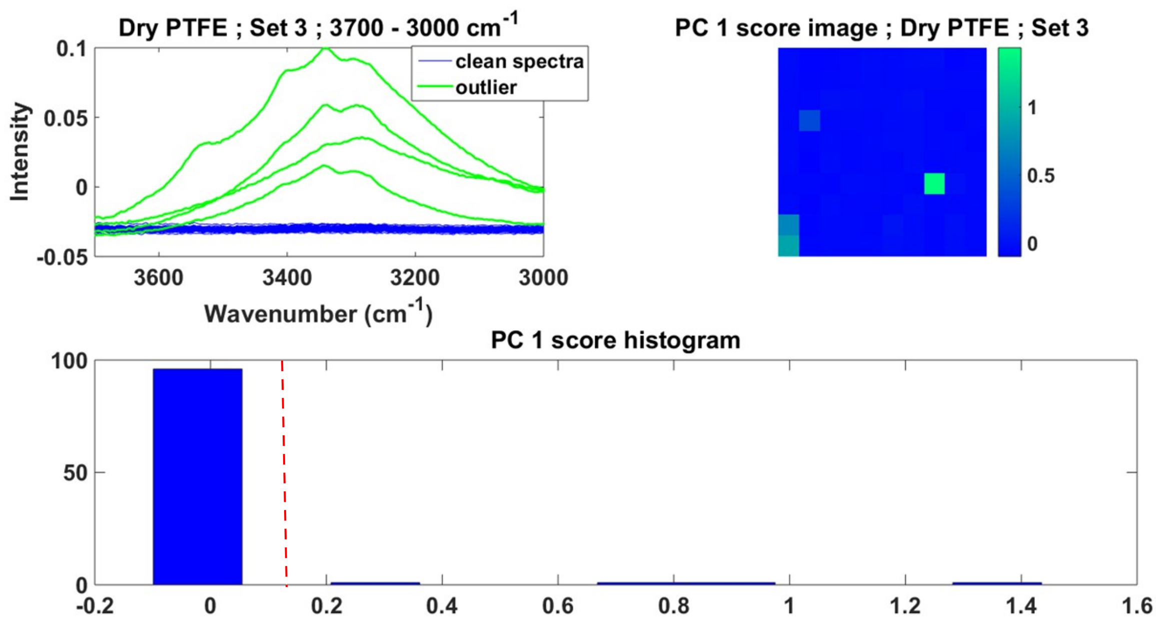

Figure S.3 Outlier removal for dry PTFE Set 3 (top left) using the PC 1 score image (top right) and a PC 1 score value histogram (lower).

After masking out the outliers from Set 2 and Set 3, we plot the cleaned data (Figure S.4), and it can be seen see that approach is useful for identifying the outlier spectrum.

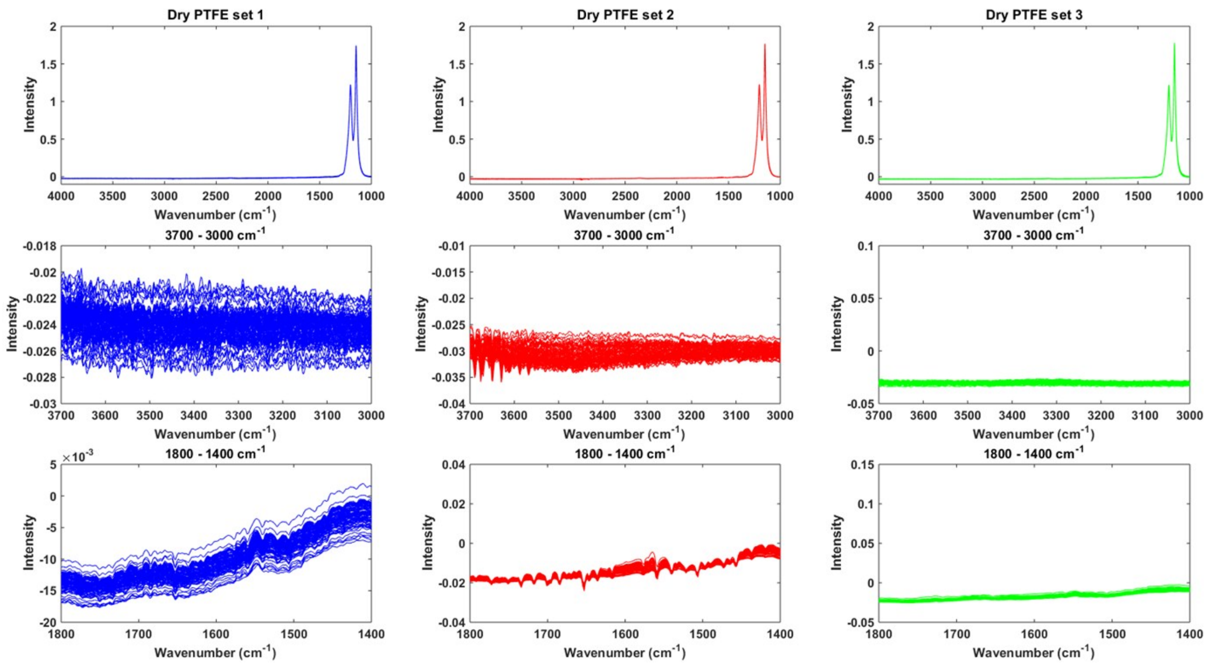

Figure S.4 Cleaned dry PTFE spectra for all sets along with spectral subsets.

A similar approach was used to mask out spectra from the wet PTFE spectra. For the wet spectra, we see outliers in Set 1 (4 pixels) & Set 3 (3 pixels) as per Figure S.5.

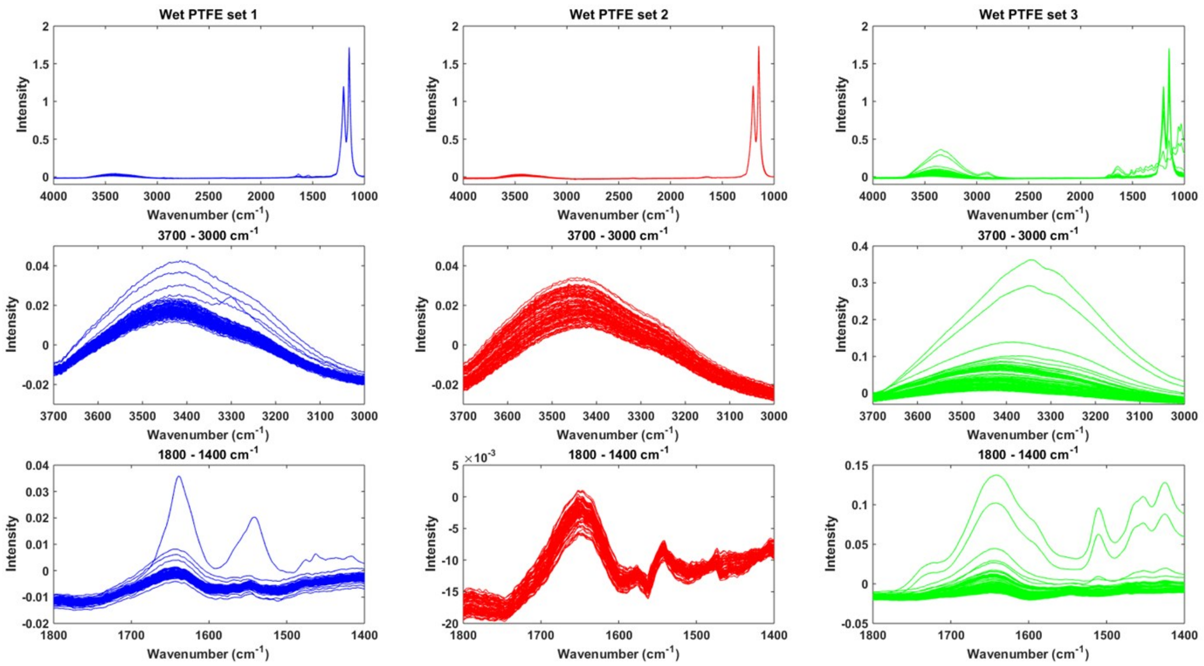

Figure S.5 Wet PTFE spectra for all sets along with spectral subsets displaying spectral outliers for Set 1 and Set 3.

A score threshold value of ‘0.025’ and ‘1’ (see Figure S.6 and Figure S.7) were used to mask out the outlier for Wet PTFE ; Set 1 & Set 3 respectively.

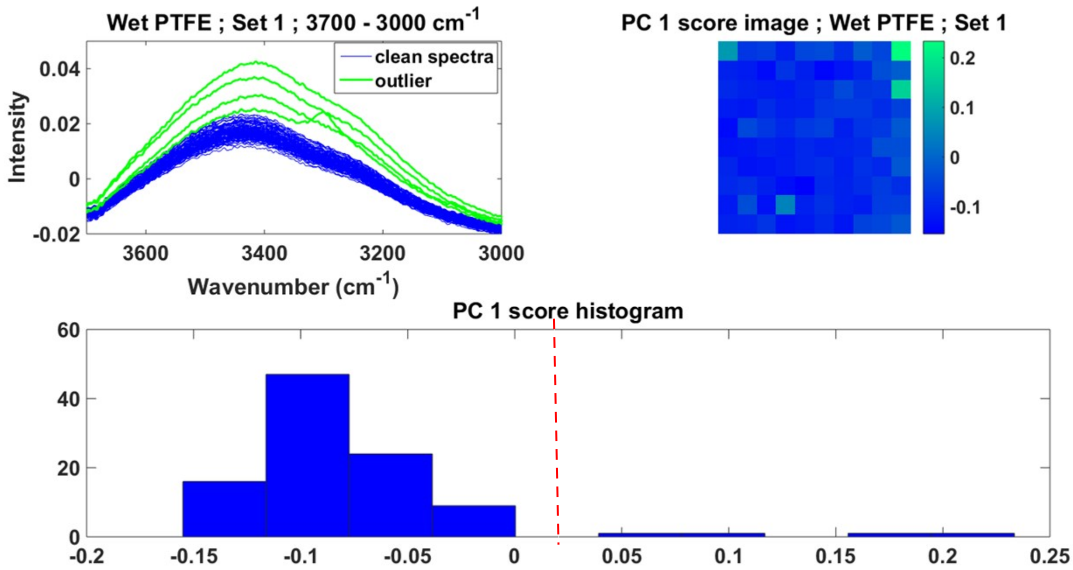

Figure S.6 Outlier removal for wet PTFE Set 1 (top left) using the PC 1 score image (top right) and a PC 1 score value histogram (lower).

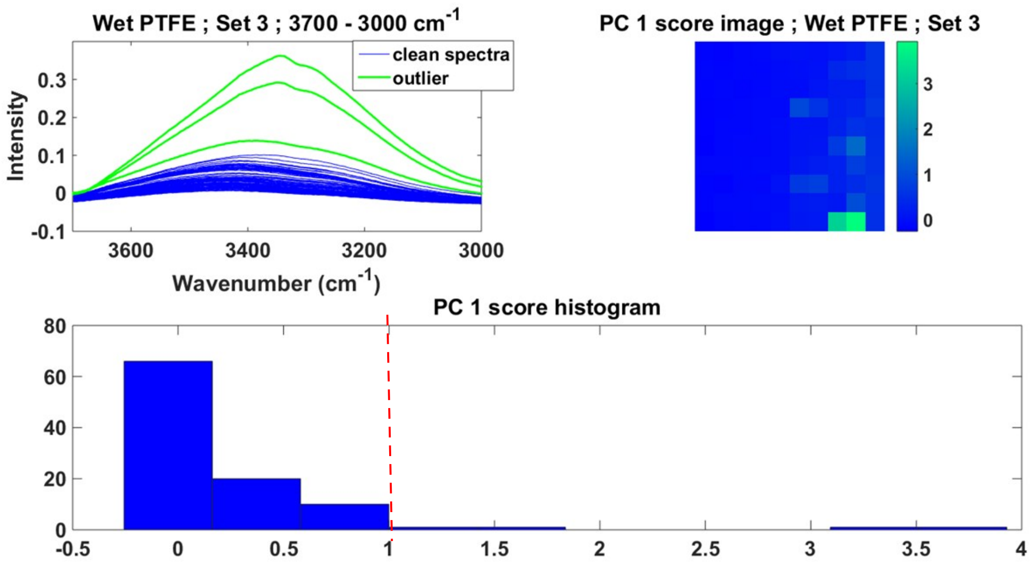

Figure S.7 Outlier removal for wet PTFE Set 3 (top left) using the PC 1 score image (top right) and a PC 1 score value histogram (lower).

The resultant cleaned spectra in shown in Figure S.8,

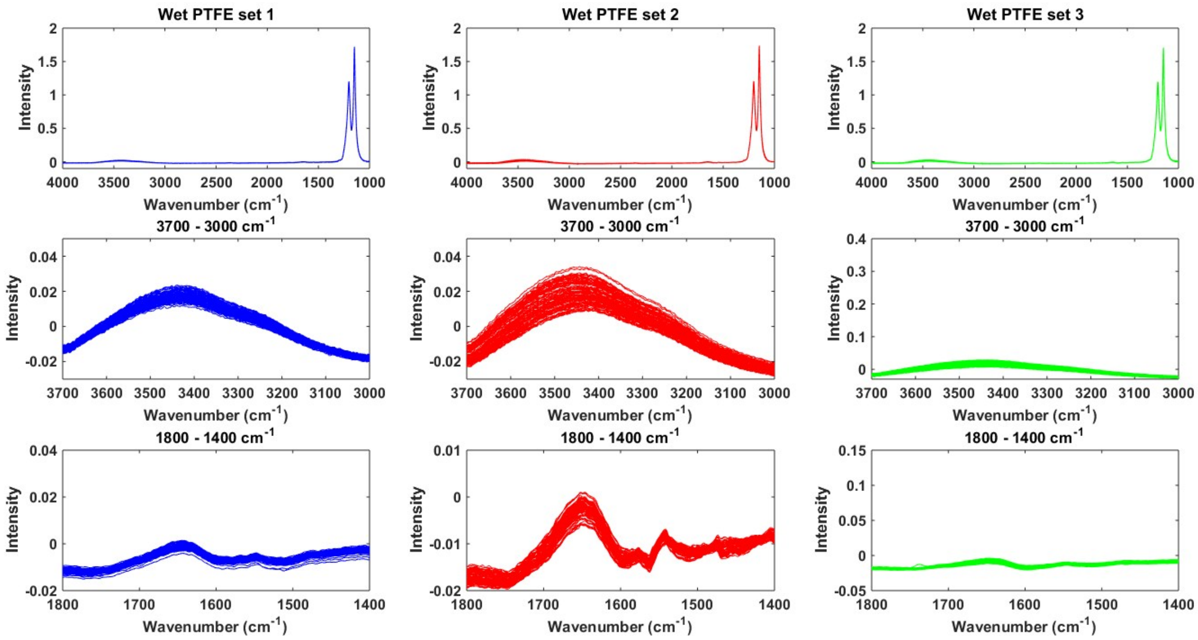

Figure S.8 Cleaned dry PTFE spectra for all sets along with spectral subsets.

As with the dry spectra, application of PCA to the wet PTFE sets (3700-3000cm^{-1}) along with histogram based thresholding allows us to remove a few spatial outliers, before proceeding with any further analysis. This process was applied to all polymers analysed, except for EVAL, which has a broad band in the 3700-3000cm^{-1} region.

S.2 ATR –FTIR spectrum of DI (deionised) water

The ATR-FTIR spectrum of DI water was collected by dipping the ATR crystal into a drop of DI water at 22°C. An air background was used, while the system was purged continuously. All other scan parameters are the same as mentioned in Section 2.2.4 of the main article, except complete ATR pressure wasn't applied, rather the ATR tip was suspended into the drop of DI water. The absorbance spectra of DI water is presented in Figure S.9. The main spectral features observable are the OH bending vibration (ν_B) [2] near 1640 cm^{-1} and the OH stretching vibration (ν_S)[3] near 3700-3000 cm^{-1}.

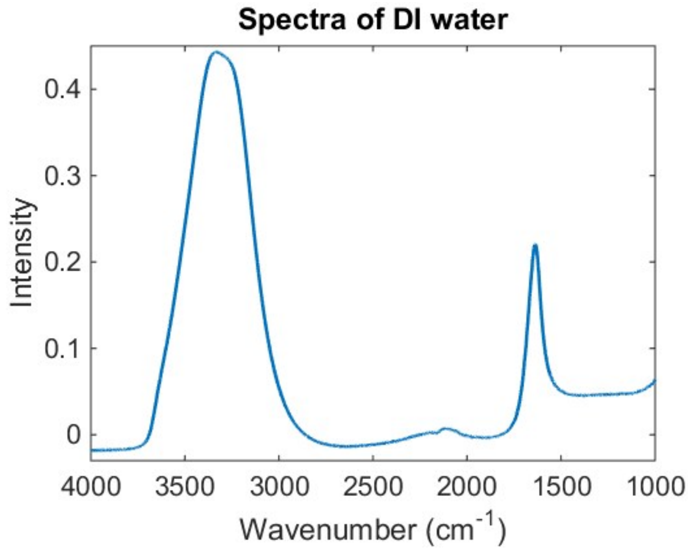

Figure S.9 The absorbance spectrum of DI water collected using ATR-FTIR.

S.3 Energy-dispersive X-ray spectroscopy (EDX) measurements of PTFE blocks.

Energy-dispersive X-ray spectroscopy measurements were collected for the surface roughened PTFE blocks, to rule out SiC contamination from the wet grinding process. 4 EDX measurements were taken for each block at different voltages (5kV, 10kV and 30kV), to have a balance between 'less bulk penetration' and 'greater sensitivity at the surface' as determined by the in-house expert. The results for each block, with respective parameters and elemental quantification results are presented in the following figures below. Each subplot of the images shown represents a measurement taken approximately a few millimetres spatially apart from the other on the PTFE blocks. None of the modified PTFE blocks showed any SiC contamination, however, the unmodified PTFE block did show a small oxide layer.

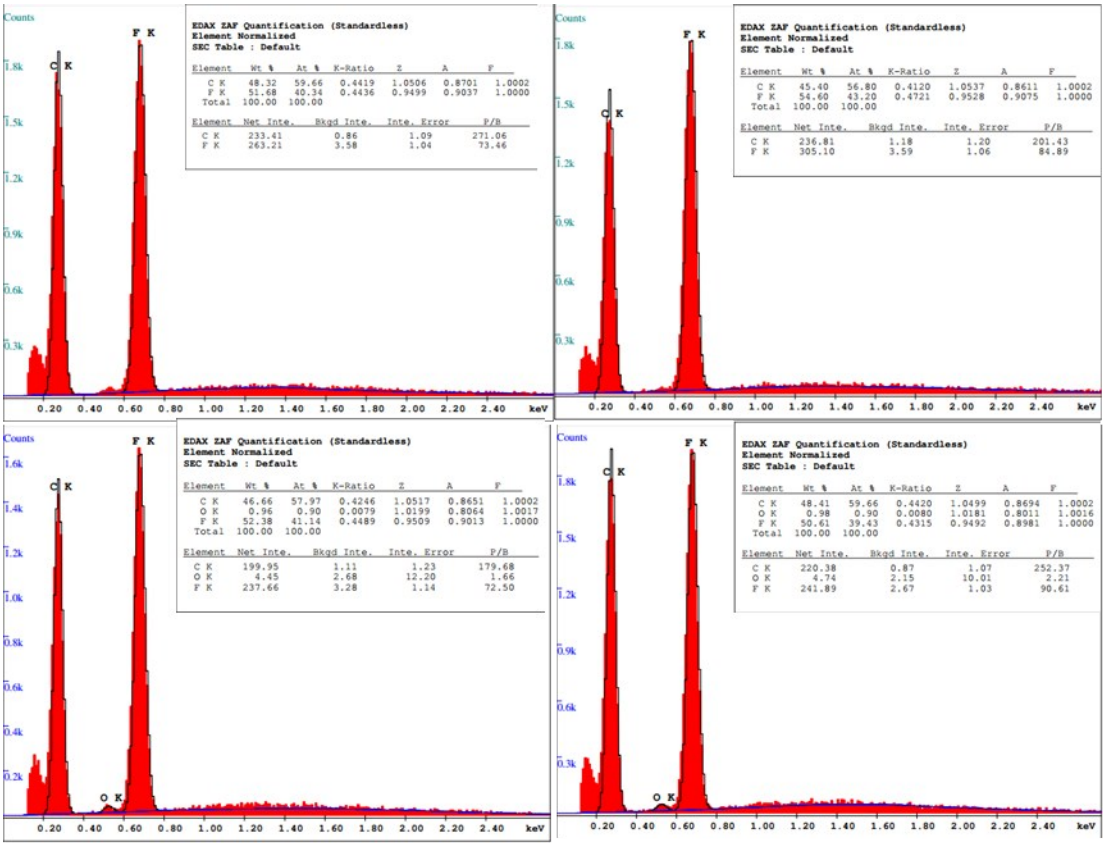

Figure S.10 EDX measurements for the unmodified PTFE surface. A tiny oxygen layer is seen in some regions of this block as suggested by lower section of this figure. No SiC contamination is observed as expected.

S.4 PLSR Modelling for films and PTFE blocks

With the aim of creating a more generally applicable Model for prediction of CA, the calibration data for "Model B wet" and "Model C" were combined and a new PLSR Model,"Model D" was built. The performance indicators for "Model D" are listed in **Table** S1. The measured and predicted CA values and the regression vector are shown in **Figure**s **S11** and **S12** respectively. The predicted CA values for the combined (i.e. films and block) validation set appear to be quite close to their measured values and this is reflected in the reasonable Model performance indicators in **Table** S1. The $RMSE_{CV}$ and $RMSE_{P}$ both increase slightly to 13° when compared to Models C and B wet, while the RPD_{C} and R^2_{C} similar or lower (2 and 0.79). The regression vector for "Model D", show some features in common with "Model B wet" and "Model C" (**Figure S12**, **Figure 6**) i.e. at 1030, 1151, 1653, 2918, 2850 cm^{-1} (see assignments in **Table** 4). However, the profile of the regression vector around the OH stretch region (near 3650 cm^{-1}) is different for "Model D" when compared to "Model B wet" and "Model C". Rather than one peak in the regression vector in that region, 2 smaller ones are observed.

Table S1. Partial least squares regression (PLSR) performance indicators for prediction of CA wet spectra. We create two new different types of Model: take account the best wet Model type, Set B (2,3) to generate: Type D that considers all the spectral features of polymeric films, silica, glass and PTFE blocks; Type E that considers the roughness and all the spectral features of polymeric films, silica, glass and PTFE blocks; The mean and standard deviation of the performance indicators was calculated over the calibration Model. (nLV = number of latent variables, RMSE=root mean square error, RPD=residual predictive deviation, R^2= coefficient of determination, and the subscripts CV=cross validation, P=prediction, C=calibration).

Model Type	Calibration	Validation	nLV	$RMSE_{CV}$	$RMSE_P$	RPD_C	RPD_P	R^2_C	R^2_P
C	(g4000, unmod, g80)	g180	4	11	6	2	1	0.76	-1.07
D	Set B (2,3) + (g4000, unmod, g80)	Set B (1) + (g180)	4	13	13	2	2	0.78	0.79
E	Set B (2,3) + (g4000, unmod, g80) + *Roughness*	Set B (1) + (g180) + *Roughness*	4	12	6	3	4	0.84	0.95
F	Set B (2,3) (g4000, unmod, g80)	Set B (1) + (g180)	4	5	4	6	7	0.97	0.98

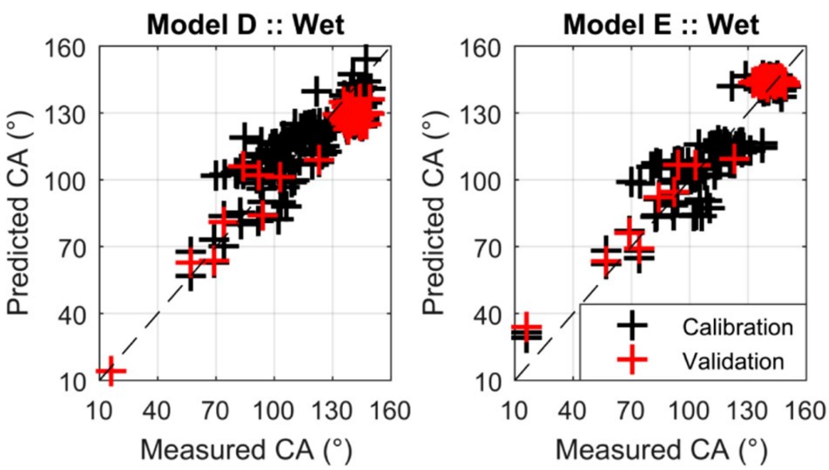

Figure S11 Comparison of measured and predicted CA values for calibration (black) and validation (red) sets for models D(left) and E (right).

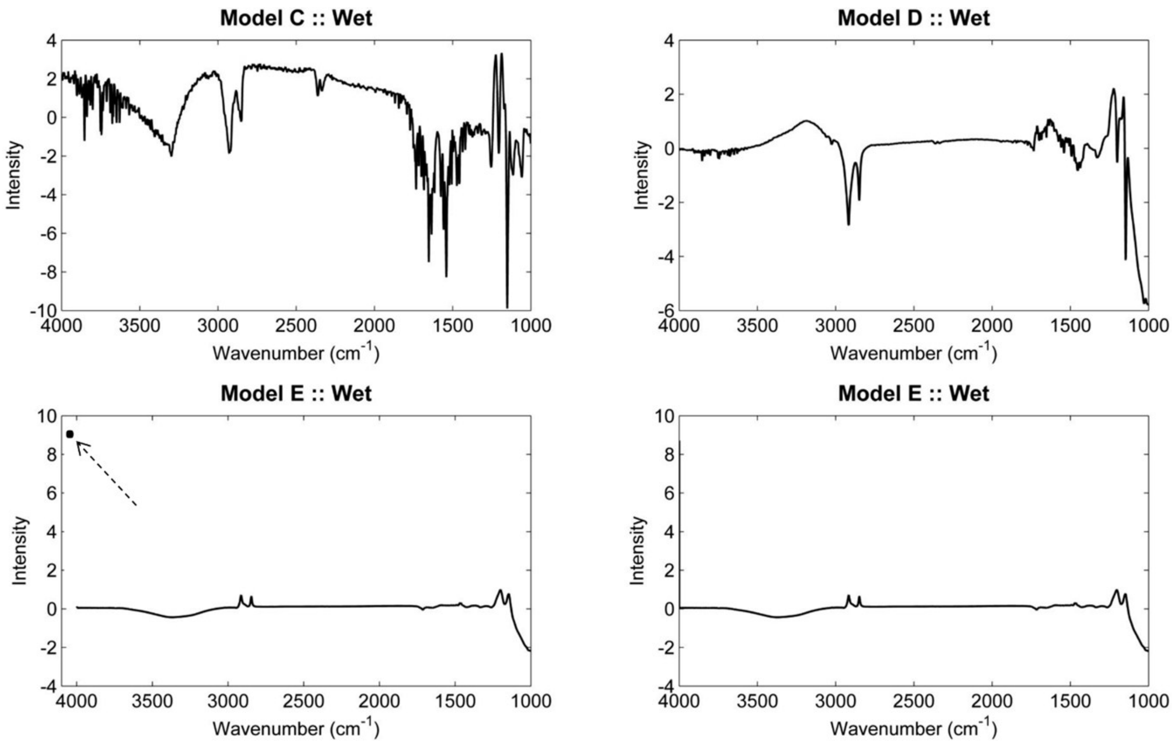

Figure S12 Regression calculated vectors for models type C (upper left), D (upper right) and E (down). The roughness variable is indicated in model E (lower left) as a separate black dot from the spectral variable, i.e. wavenumber to avoid ambiguity. A model E subset without the roughness variable is presented (lower right) for a better comparison with the previous models.

To further test this model, it was applied to the test image and the predicted CA map is shown in **Figure** S13. “Model D” could not distinguish well between the coating and glass region, underpredicting the coating while over predicting the glass region. Therefore, our aim is developing a more general model was not achieved by this approach.

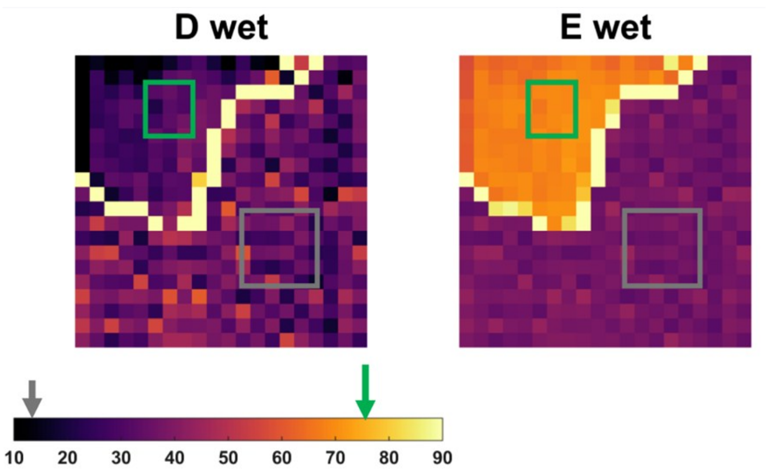

Figure S13. CA prediction maps obtained for the test sample after applying the PLSR models where D, E wet refer to Table 4. The regions marked on the prediction maps refer to coating (green square) and glass (grey square) regions selected for calculating predicted CA, mean error and RMSE presented in Table 8. Reference CA values for glass (grey) and for coating (green) are marked with an arrow in the legend.

Since the measured ATR spectra are mainly sensitive to differences in surface chemistry (see **Figure** 3 and 4) and not to the surface roughness, we decided to include surface roughness as a predictor variable our PLSR Model and call this approach "Model E". Although the roughness of the PTFE blocks was measured (**Table** 3), the transparency of the film samples made it difficult to obtain robust roughness measurements. For this reason, the roughness variable of the films was set to 0. Although this is not strictly true, this was the best compromise available at this point. The results of this approach can be seen in **Figure S10** and **Table S1**. Including the roughness improved model performance slightly in terms of the model performance indicators RPD and R^2, while maintaining the same number of latent variables as for "Model D" and halving the prediction error (6° vs 13°). When applied to the test image (**Figure** S13) the differences between Models C, D and E are very clear, with "Model E" predicting the coating CA very well ($76^{o} \pm 5$ vs $67^{o} \pm 2$) while over predicting the glass CA ($16^{o} \pm 1$ vs $37^{o} \pm 3$). **Table S2** shows the RMSE of each model applied to the test image, clearly "Model E" results in the lowest overall error. When considering the regression vector for "Model E", it shares many features in common with that for "Model B wet" and exhibits the largest weighting for the roughness variable. (**Figure S12**, **Figure** 6).

Table S2. Measured CA, mean predicted CA, mean error and RMSE as calculated from regions shown in **Figure S13** for each Model type.

			D			**E**	
	Actual	Predicted	Mean Error	RMSE	Predicted	Mean Error	RMSE
Coating	76±5	22 ± 9	54.39	55.07	67 ± 2	-9.44	9.68
Glass	16±1	34 ± 9	21.25	23.17	37 ± 3	24.13	24.29

References

1 S. Wold, K. Esbensen and P. Geladi, Principal component analysis, *Chemom. Intell. Lab. Syst.*, 1987, **2**, 37–52.

2 G. Socrates, *Infrared and Raman characteristic group frequencies*, John Wiley & Sons, Chichester ; New York, Third Edit., 2004.

3 L. J. Bellamy, *The Infrared Spectra of Complex Molecules*, 1980.