Quantification of titanium dioxide (TiO2) anatase and rutile polymorphs in binary mixtures by Raman spectroscopy: an interlaboratory comparison

Supplementary Information

Alessio Sacco[1], Luisa Mandrile[1], Li-Lin Tay[2], Nobuyasu Itoh[3], Ankit Raj[4], Alberto Moure[5], Adolfo del Campo[5], Jose F. Fernandez[5], Keith R. Paton[6], Sebastian Wood[6], Hyuksang Kwon[7], Tehseen Adel[8], Angela R. Hight Walker[8], Erlon Henrique Martins Ferreira[9], Ralf Theissmann[10], Thomas Koch[10], Andrea Mario Giovannozzi[1], Chiara Portesi[1] and Andrea Mario Rossi[1]

[1] Quantum Metrology and Nanotechnology Department, Istituto Nazionale di Ricerca Metrologica (INRiM), Torino, Italy
[2] Metrology Research Centre, National Research Council Canada, Ottawa, ON Canada
[3] National Metrology Institute of Japan (NMIJ), National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan
[4] Department of Applied Chemistry, National Yang Ming Chiao Tung University, Hsinchu, Taiwan
[5] Instituto de Cerámica y Vidrio, CSIC, C/Kelsen 5, Madrid, Spain
[6] National Physical Laboratory (NPL), Teddington, Middlesex, United Kingdom
[7] Korea Research Institute of Standards and Science (KRISS), Daejeon 34113, Republic of Korea
[8] Physical Measurement Laboratory, National Institute of Standards and Technology (NIST), Gaithersburg, MD, United States of America
[9] National Institute of Metrology, Quality and Technology (Inmetro), Rio de Janeiro, Brazil
[10] Research Services, KRONOS INT. Inc., Leverkusen, Germany

E-mail: a.sacco@inrim.it

S1. Participants and experimental setups

Table S1 — Experimental setups and parameters employed by the participants.

Participant code	Raman spectrometer manufacturer and model	Excitation wavelength (nm)	Objective magnification	Objective numerical aperture	Laser power (mW)	Exposure time (number of exposures × single exposure time)	Sample identification number
A	WITec™ Alpha 300 S	532	50× LWD	0.55	0.8	1 × 1 s	1
B	Horiba LabRam HR Evolution	532	10×	0.25	0.44	1 × 0.5 s	2
C	Horiba LabRam HR Evolution	532	10×	0.25	0.6	1 × 3 s	3
D	WITec™ Alpha 300 A/R	532	10×	0.30	0.5	1 × 1 s	4
E	Thermo Scientific™ DXR™xi	532	4×	0.10	0.5	50 × 0.2 s	5
F	Horiba LabRam HR 800	457	10×	0.25	0.13	25 × 0.2 s	6
G	WITec™ AR300	532	5×	0.12	0.5	1 × 1 s	7
H	Renishaw® inVia™	514.5	20×	0.4	0.4	10 × 0.2 s	8
I	Renishaw® inVia™	532	50× LWD	0.5	0.5	1 × 0.5 s	3
J	BWTEK i-Raman®	532	20×	0.4	0.3	2 × 1 s	7
K	Custom-made (Teledyne Princeton Instruments SP2500i spectrograph and Andor DU970N-BV CCD)	532	20×	0.25	1.0	1 × 12 s	9

S2. Participants training datasets, latent variables and PLS models

The following figures in this section show the spectra belonging to the training datasets of the participants before and after preprocessing, the latent variables employed in the respective final PLS regression models with their explained variance, and the PLS regressions themselves depicted with the calibration (training) and test (prediction) data points along with model performance metrics.

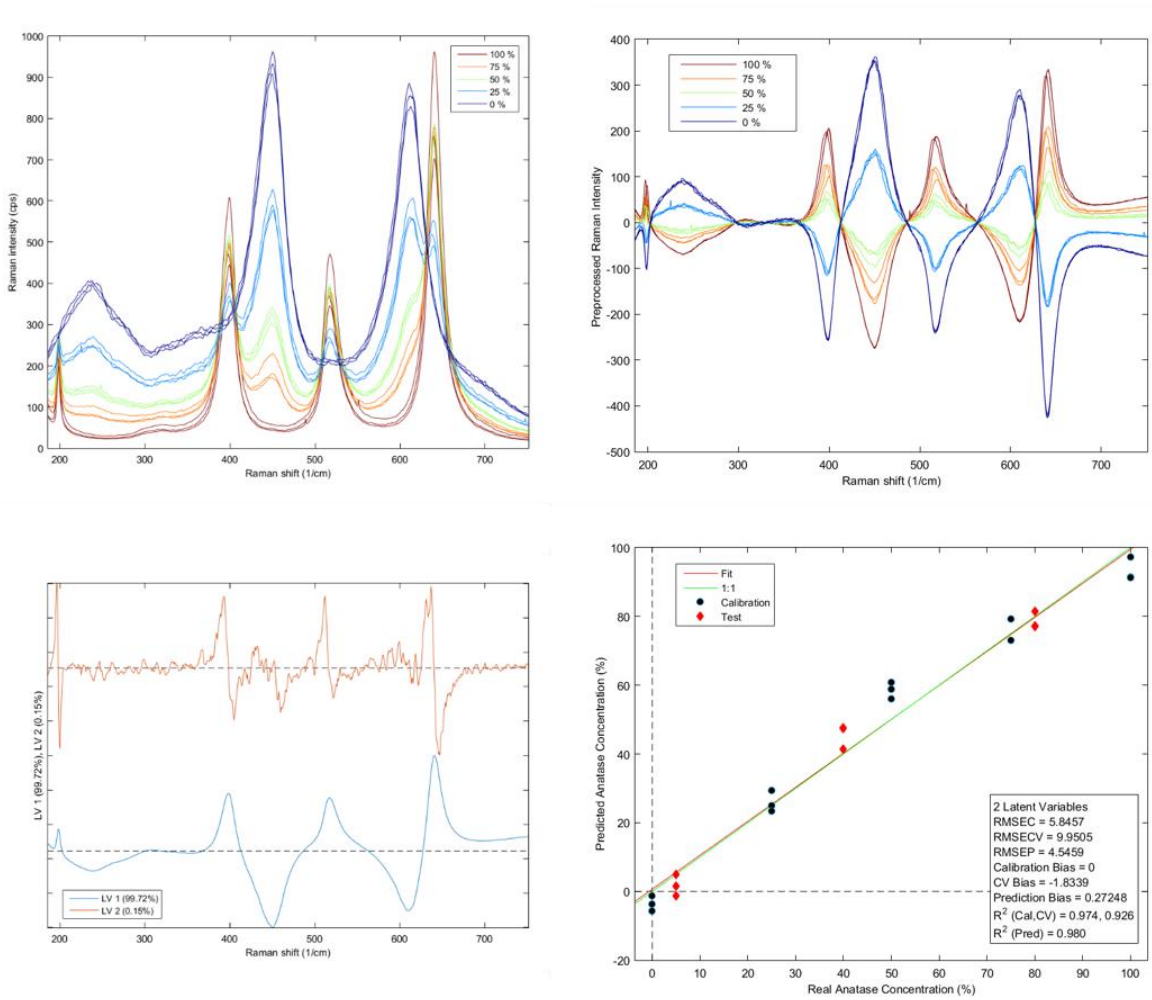

Figure S1-A — Data and PLS model pertaining participant A. Raw (top left) and preprocessed (top right) training dataset. Percentages in the legends correspond to the anatase ratio in the sample; latent variables employed in the model (bottom left), with relative explained variance in parentheses; PLS regression model (bottom right), showing the regression line in red, training dataset as black points, and prediction dataset as red points.

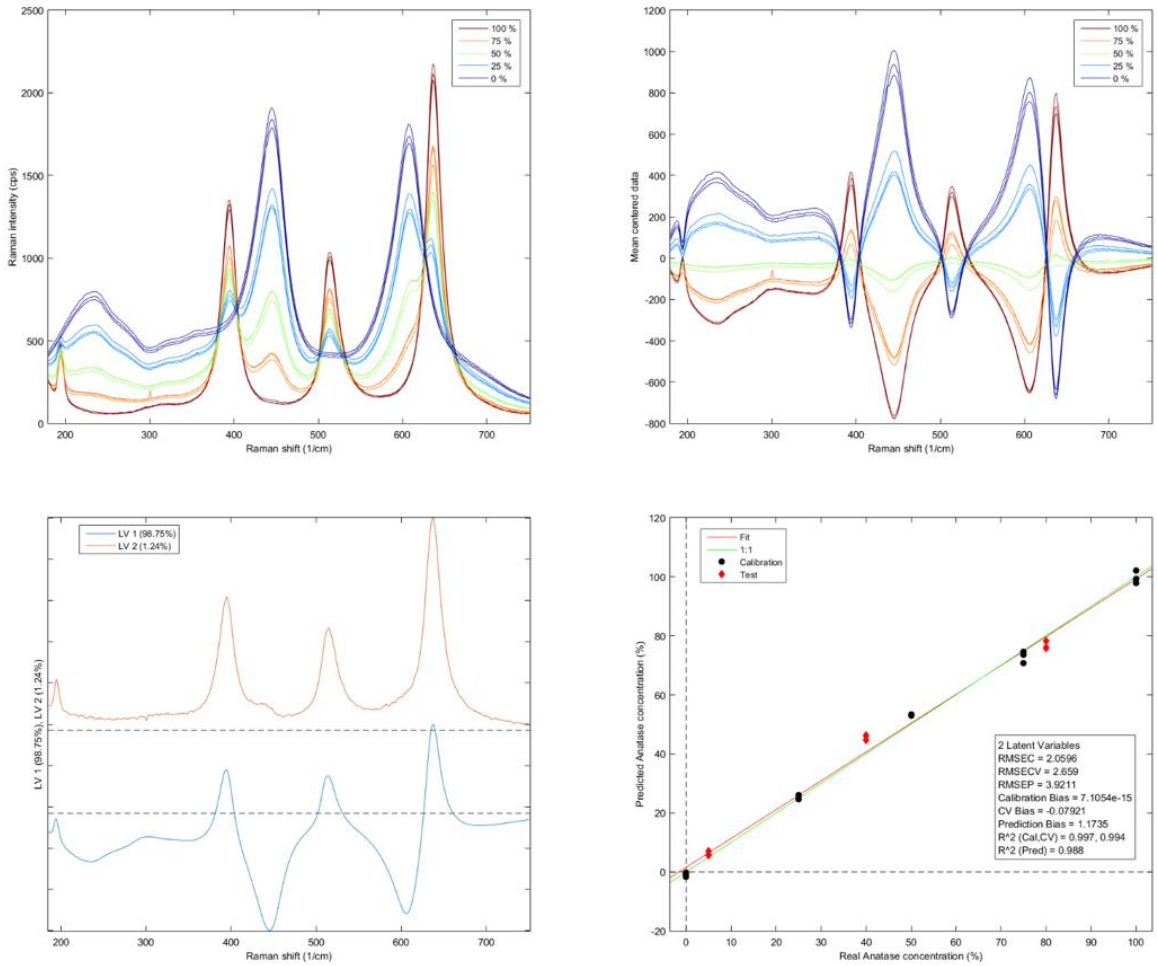

Figure S1-B — Data and PLS model pertaining participant B. Raw (top left) and preprocessed (top right) training dataset. Percentages in the legends correspond to the anatase ratio in the sample; latent variables employed in the model (bottom left), with relative explained variance in parentheses; PLS regression model (bottom right), showing the regression line in red, training dataset as black points, and prediction dataset as red points.

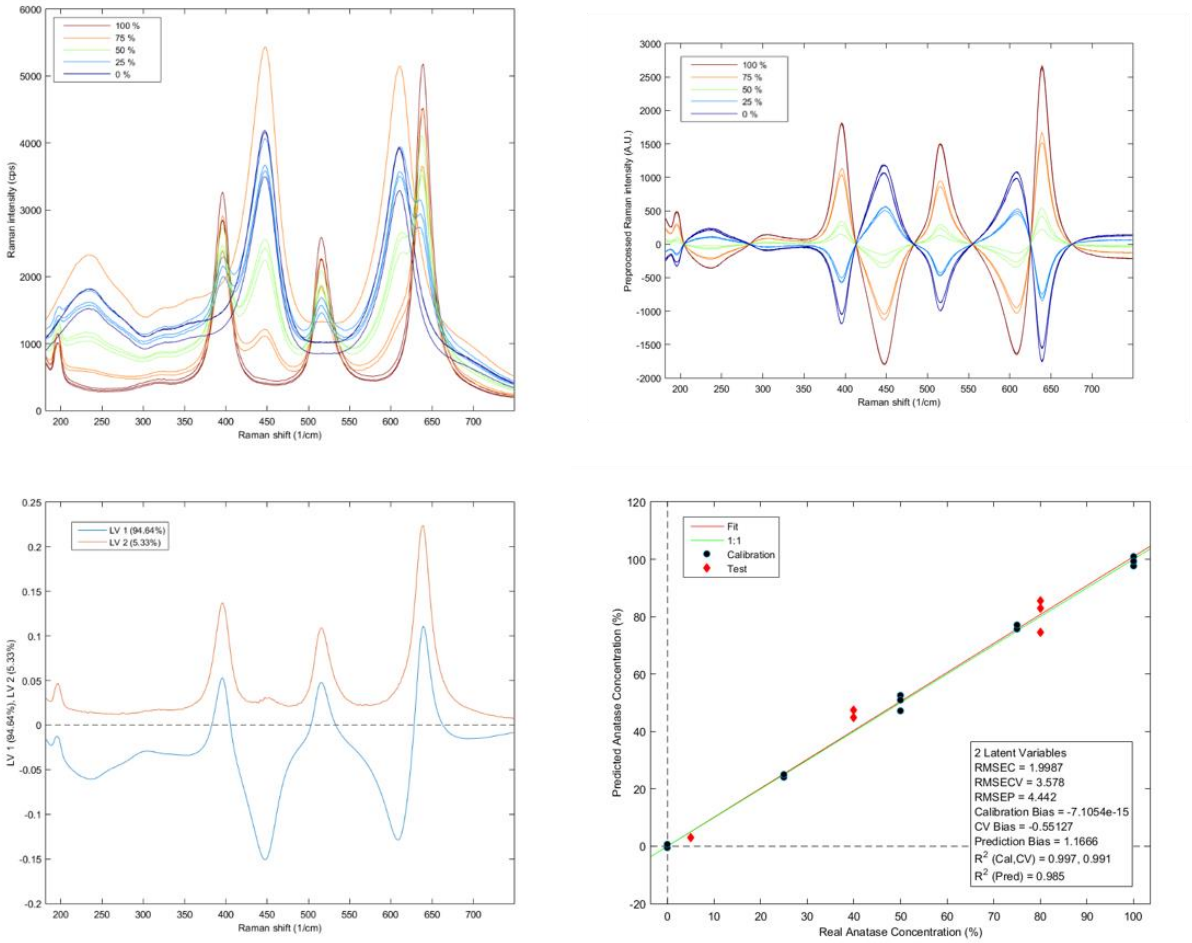

Figure S1-C — Data and PLS model pertaining participant C. Raw (top left) and preprocessed (top right) training dataset. Percentages in the legends correspond to the anatase ratio in the sample; latent variables employed in the model (bottom left), with relative explained variance in parentheses; PLS regression model (bottom right), showing the regression line in red, training dataset as black points, and prediction dataset as red points.

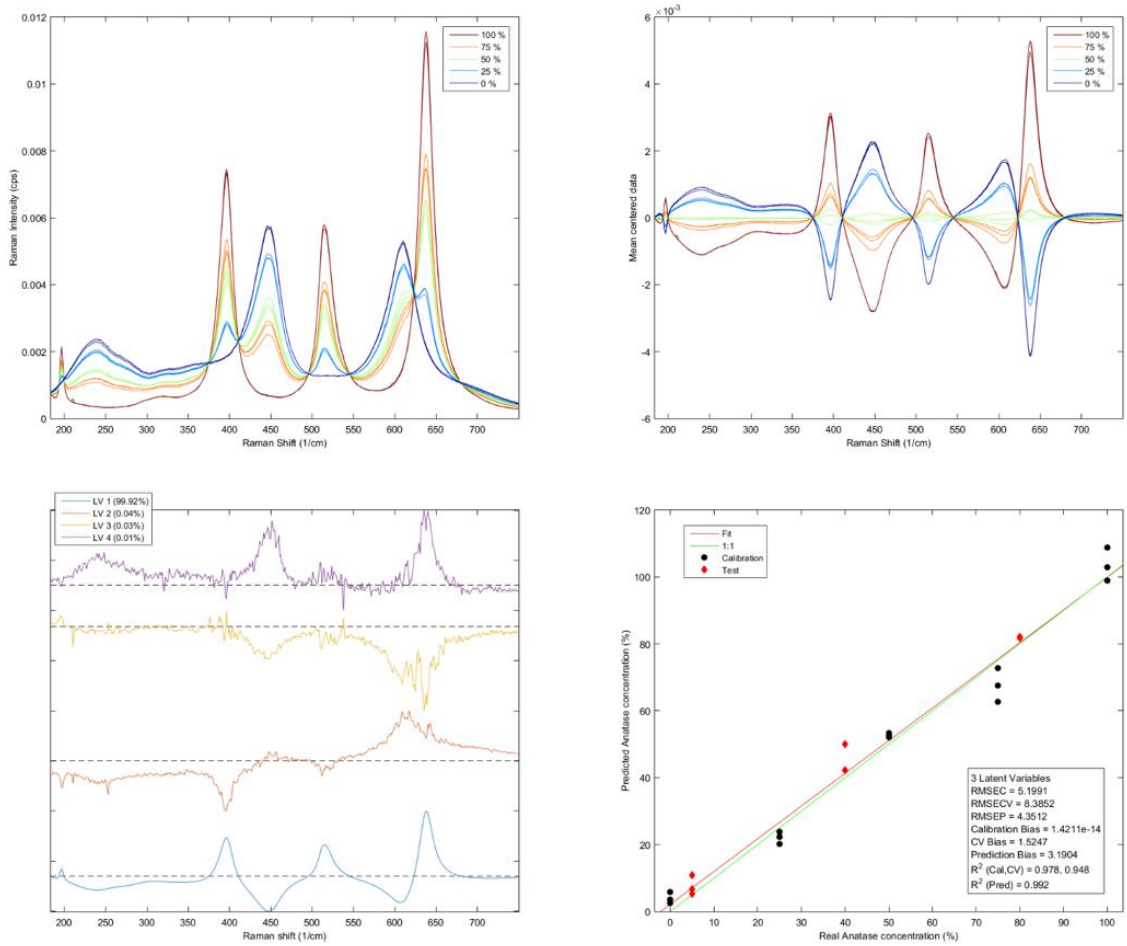

Figure S1-D — Data and PLS model pertaining participant D. Raw (top left) and preprocessed (top right) training dataset. Percentages in the legends correspond to the anatase ratio in the sample; latent variables employed in the model (bottom left), with relative explained variance in parentheses; PLS regression model (bottom right), showing the regression line in red, training dataset as black points, and prediction dataset as red points.

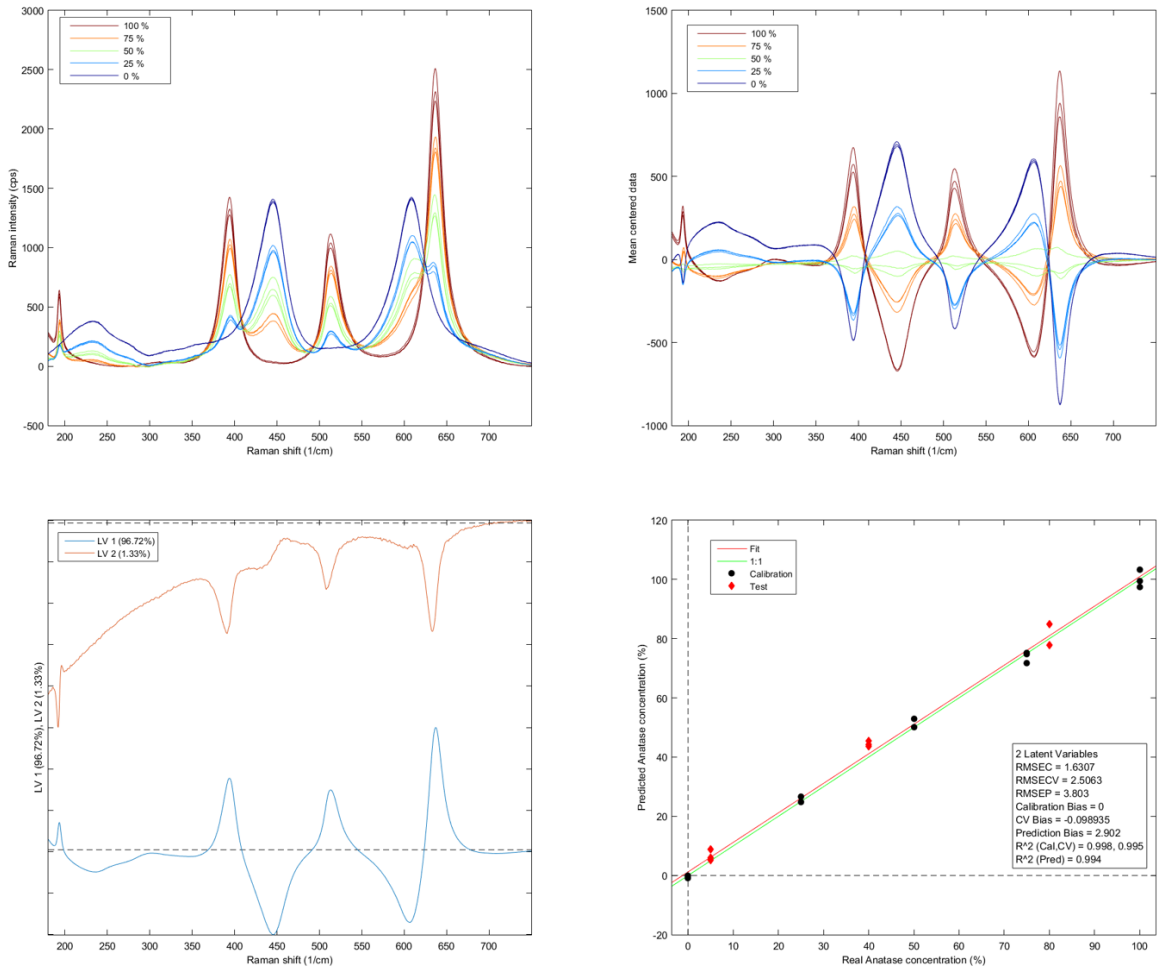

Figure S1-E — Data and PLS model pertaining participant E. Raw (top left) and preprocessed (top right) training dataset. Percentages in the legends correspond to the anatase ratio in the sample; latent variables employed in the model (bottom left), with relative explained variance in parentheses; PLS regression model (bottom right), showing the regression line in red, training dataset as black points, and prediction dataset as red points.

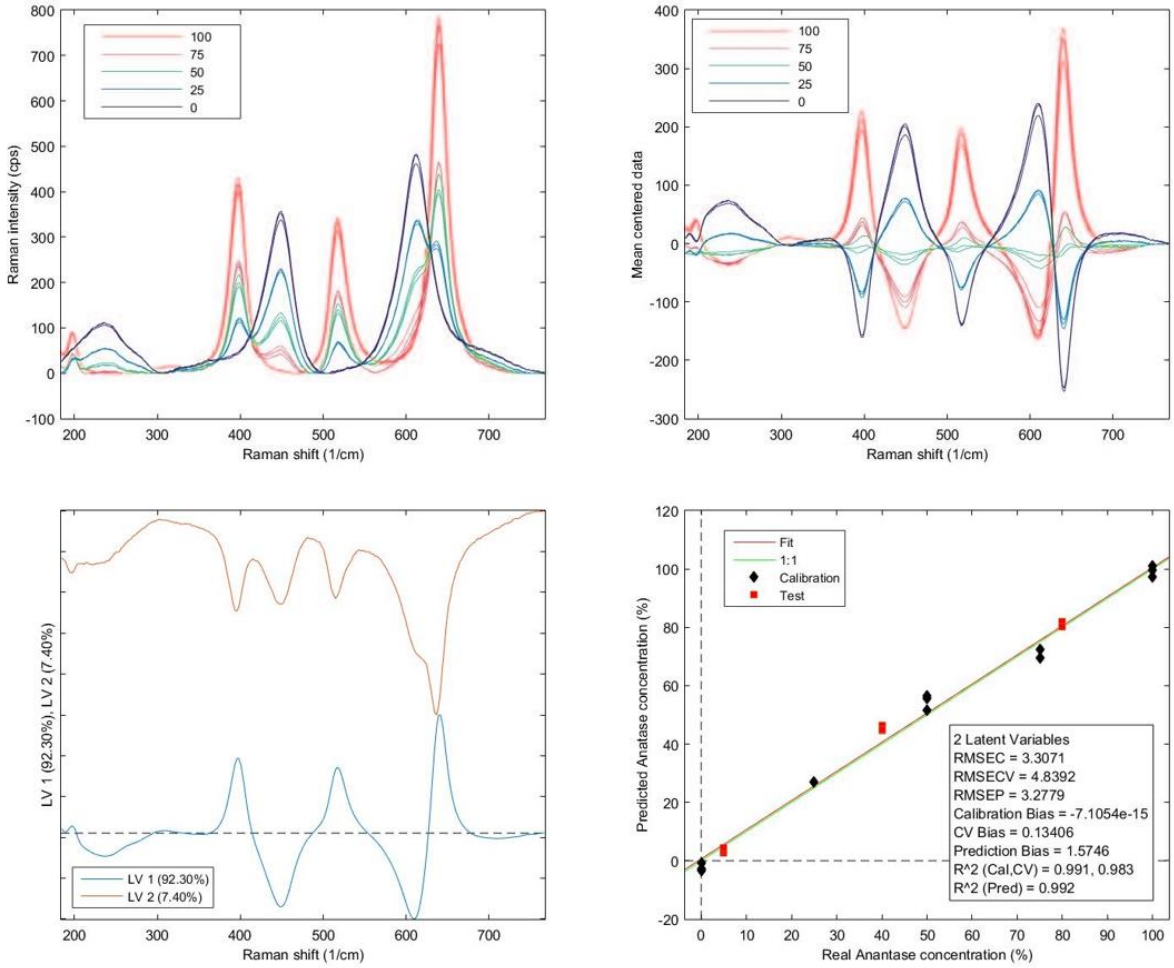

Figure S1-F — Data and PLS model pertaining participant F. Raw (top left) and preprocessed (top right) training dataset. Percentages in the legends correspond to the anatase ratio in the sample; latent variables employed in the model (bottom left), with relative explained variance in parentheses; PLS regression model (bottom right), showing the regression line in red, training dataset as black points, and prediction dataset as red points.

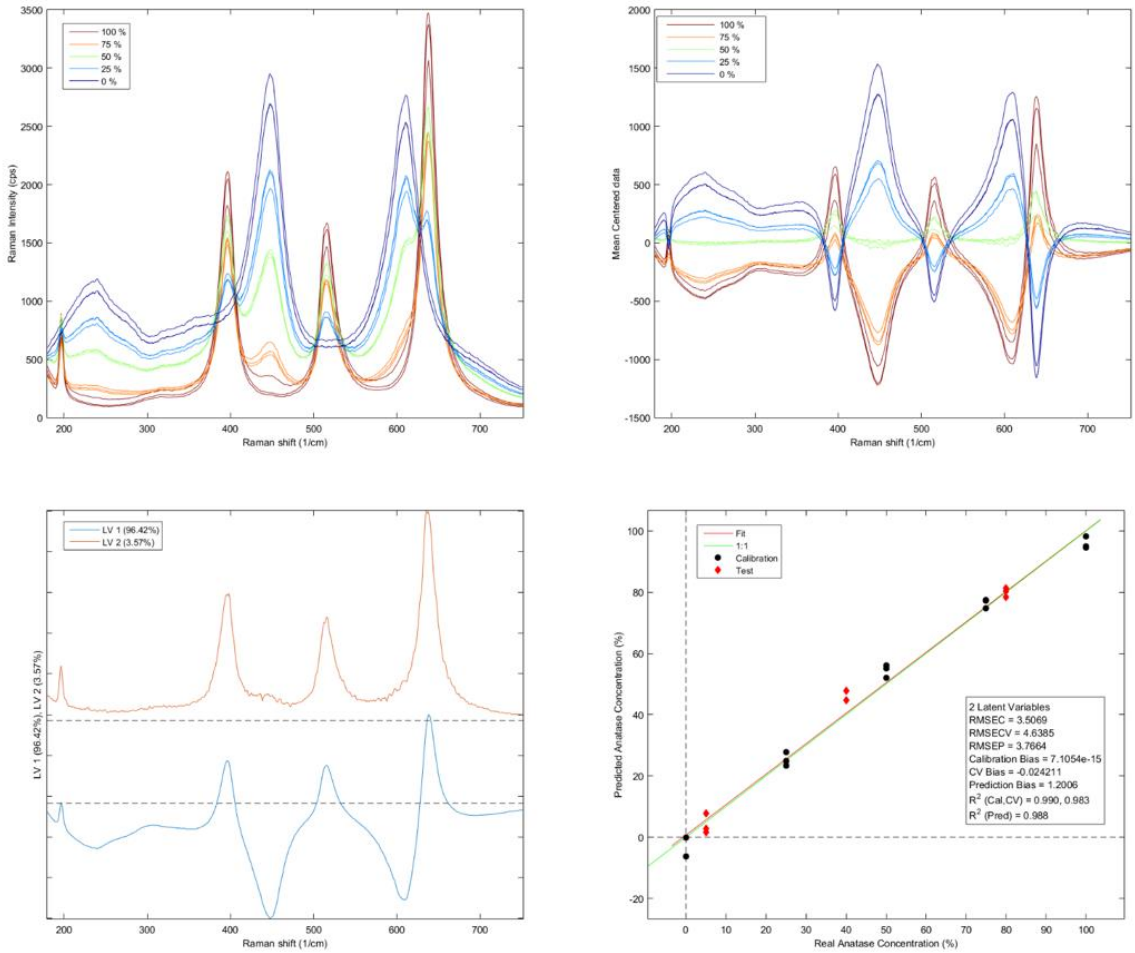

Figure S1-G — Data and PLS model pertaining participant G. Raw (top left) and preprocessed (top right) training dataset. Percentages in the legends correspond to the anatase ratio in the sample; latent variables employed in the model (bottom left), with relative explained variance in parentheses; PLS regression model (bottom right), showing the regression line in red, training dataset as black points, and prediction dataset as red points.

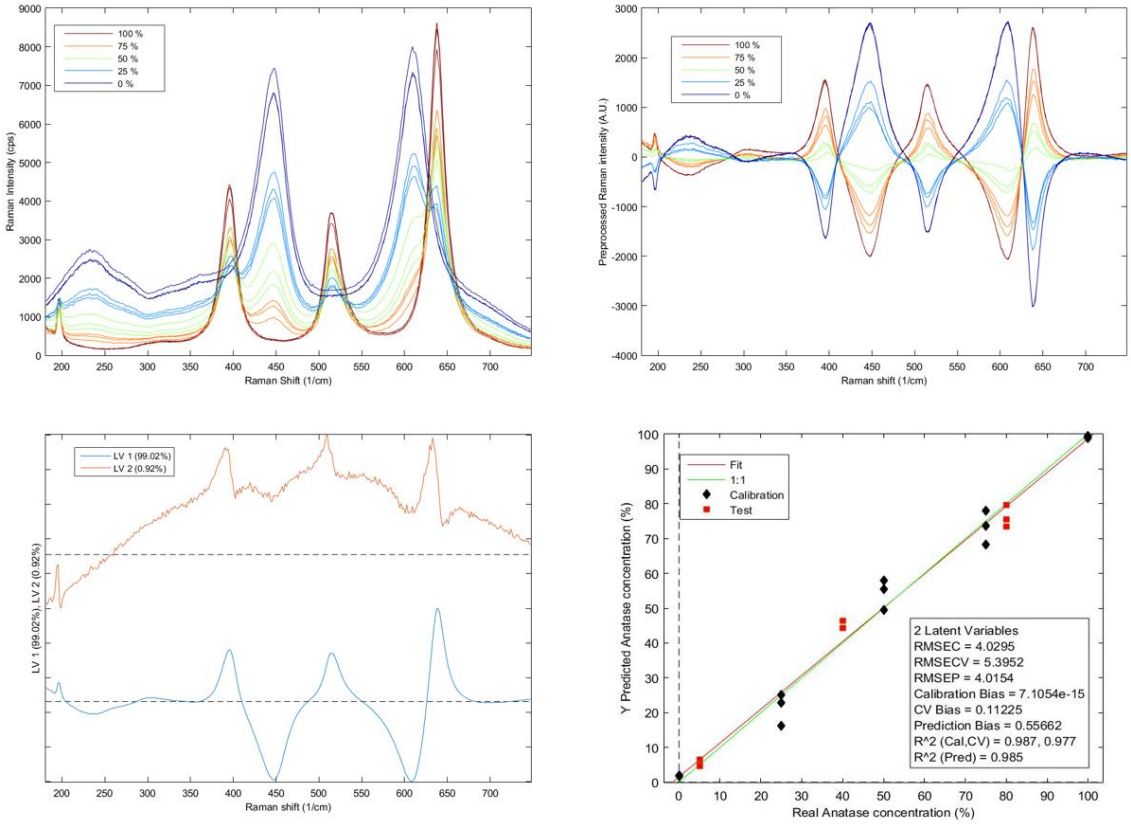

Figure S1-H — Data and PLS model pertaining participant H. Raw (top left) and preprocessed (top right) training dataset. Percentages in the legends correspond to the anatase ratio in the sample; latent variables employed in the model (bottom left), with relative explained variance in parentheses; PLS regression model (bottom right), showing the regression line in red, training dataset as black points, and prediction dataset as red points.

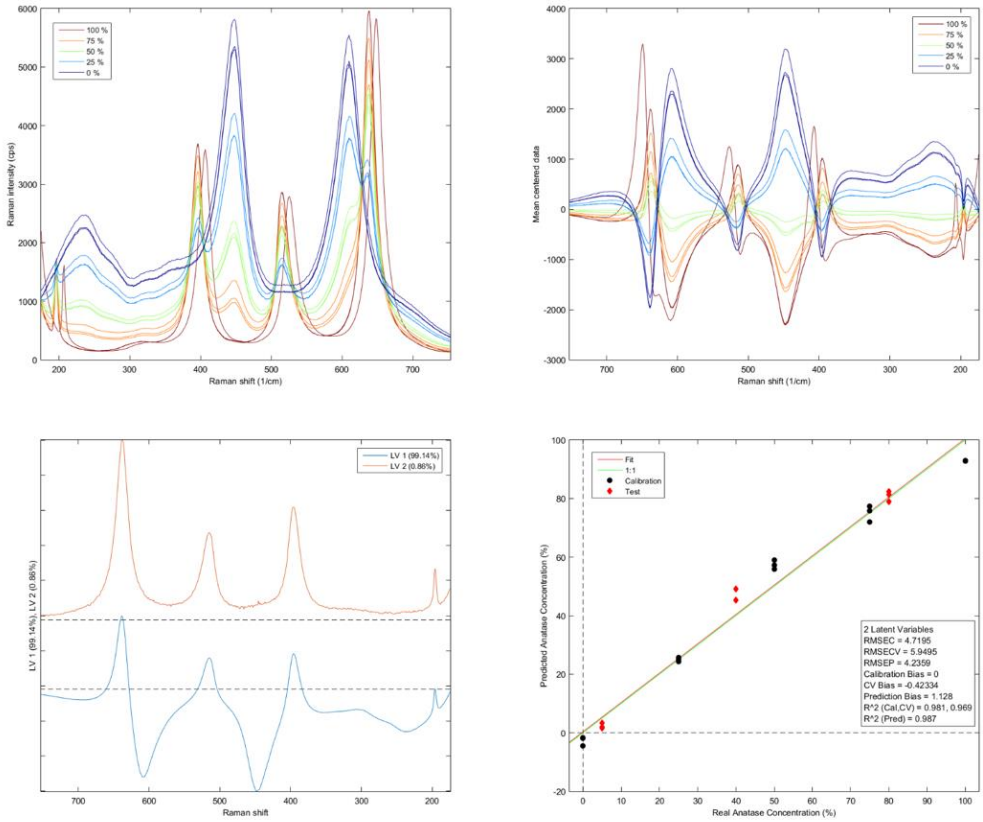

Figure S1-I — Data and PLS model pertaining participant I. Raw (top left) and preprocessed (top right) training dataset. Percentages in the legends correspond to the anatase ratio in the sample; latent variables employed in the model (bottom left), with relative explained variance in parentheses; PLS regression model (bottom right), showing the regression line in red, training dataset as black points, and prediction dataset as red points.

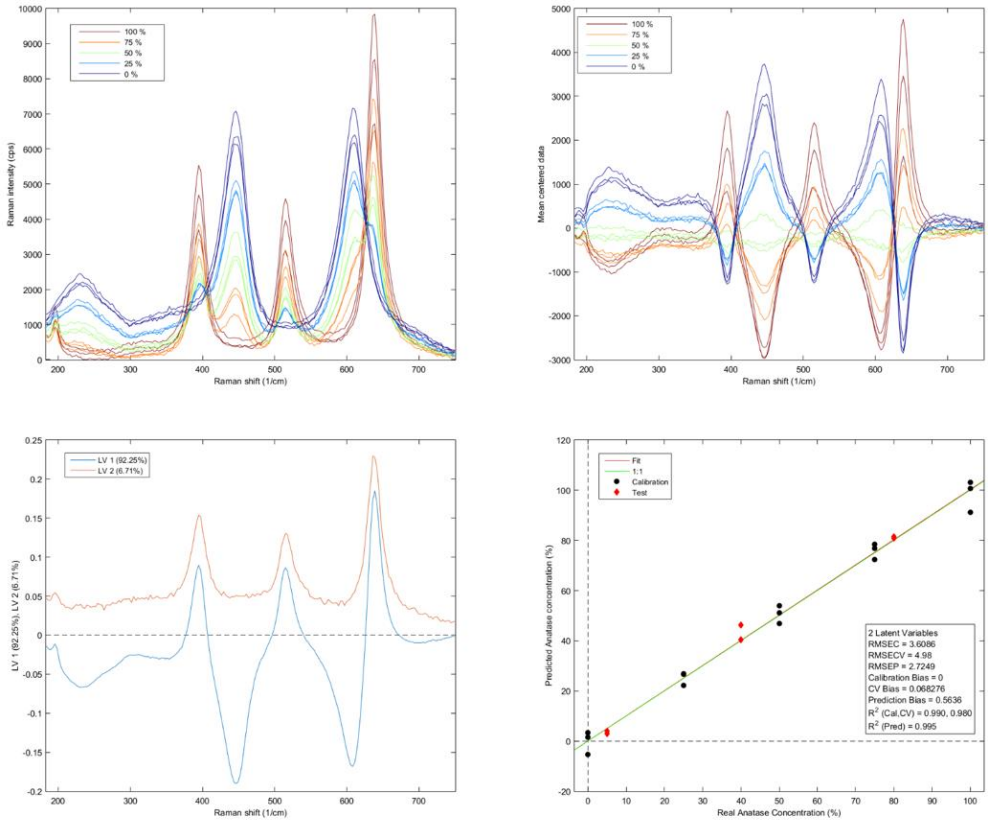

Figure S1-J — Data and PLS model pertaining participant J. Raw (top left) and preprocessed (top right) training dataset. Percentages in the legends correspond to the anatase ratio in the sample; latent variables employed in the model (bottom left), with relative explained variance in parentheses; PLS regression model (bottom right), showing the regression line in red, training dataset as black points, and prediction dataset as red points.

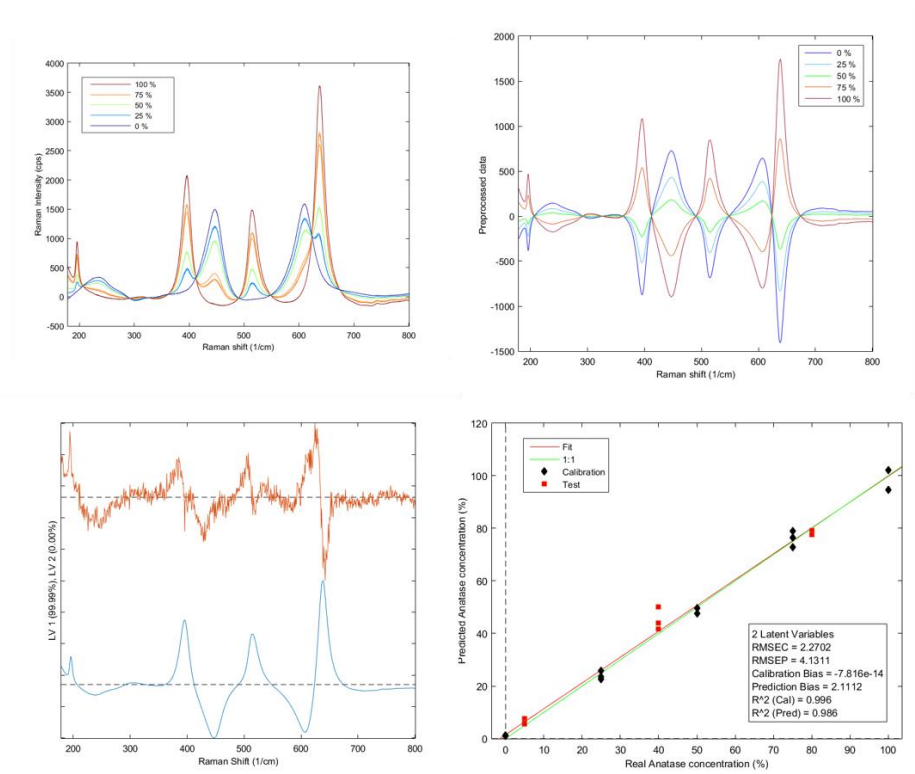

Figure S1-K — Data and PLS model pertaining participant K. Raw (top left) and preprocessed (top right) training dataset. Percentages in the legends correspond to the anatase ratio in the sample; latent variables employed in the model (bottom left), with relative explained variance in parentheses; PLS regression model (bottom right), showing the regression line in red, training dataset as black points, and prediction dataset as red points.

S3. Figures of merit for training, cross validation and prediction

In Table S2, RMSEC, RMSECV, RMSEP and R^2 values for model training, cross validation, and prediction phases for each model are reported. These were calculated as explained in section 2.8 of the main manuscript.

Table S2 — Figures of merit for model training, cross validation, and prediction for all participants.

Participant code	RMSEC	R^2 (calibration)	RMSECV	R^2 (cross validation)	RMSEP	R^2 (prediction)
A	5.85	0.974	9.95	0.926	3.27	0.980
B	2.06	0.997	2.66	0.994	2.22	0.988
C	2.00	0.997	3.58	0.991	3.55	0.985
D	5.20	0.978	8.39	0.948	4.24	0.992
E	1.63	0.998	2.51	0.995	2.82	0.994
F	3.31	0.991	4.84	0.983	1.18	0.992
G	3.51	0.990	4.64	0.983	2.17	0.988
H	4.03	0.987	5.39	0.977	2.78	0.985
I	4.72	0.981	5.95	0.969	2.46	0.987
J	3.61	0.990	4.98	0.980	2.25	0.995
K	2.27	0.996	3.17	0.988	3.02	0.986

Table S3 shows the overall discrepancies and standard uncertainties values, and the limit of detection (LOD) and limit of quantification (LOQ) for anatase and for rutile, for each model. These were calculated as explained in section 2.8 of the main manuscript.

Table S3 — Calculated discrepancies, standard uncertainties, LODs and LOQs of the models from all datasets.

Participant code	Discrepancy /%mass	Standard uncertainty /%mass	LOD (anatase) /%mass	LOD (rutile) /%mass	LOQ (anatase) /%mass	LOQ (rutile) /%mass
A	1.7	2.5	5.4	9.4	18.0	19.1
B	1.7	0.8	0.7	4.6	2.2	14.4
C	0.8	2.2	0.5	6.4	1.8	8.2
D	0.7	2.8	4.6	6.5	8.1	20.4
E	1.6	1.9	1.0	4.6	3.3	11.6
F	0.8	0.8	3.2	2.4	10.7	7.6
G	0.8	1.8	0.1	5.8	0.2	7.3
H	1.4	1.4	2.0	1.4	2.6	2.8
I	1.4	1.6	0.3	7.4	1.0	8.0
J	1.5	1.2	5.3	3.7	11.7	12.3
K	0.8	2.6	1.6	0.1	2.4	0.5