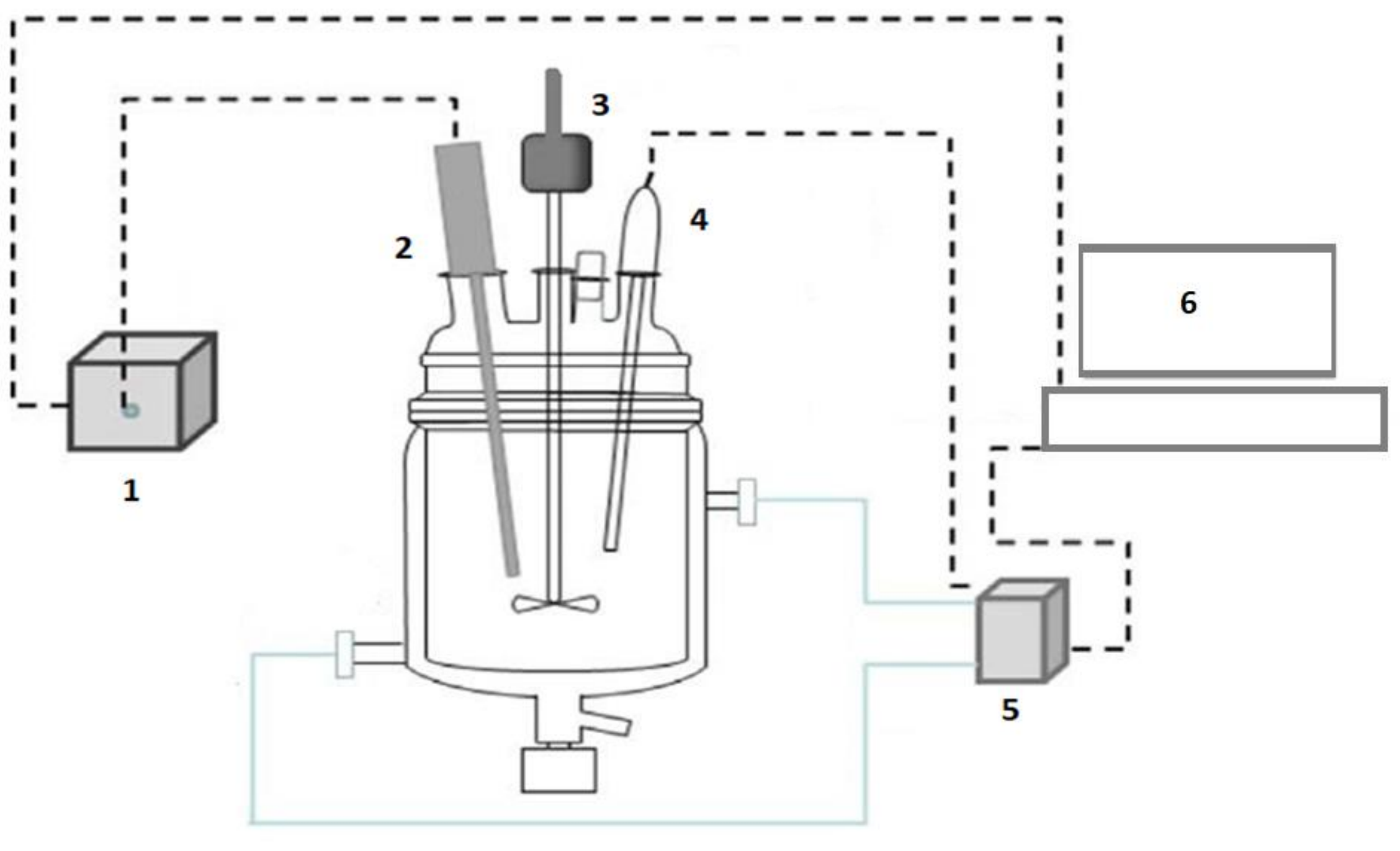
COMPARISON OF LINEAR CALIBRATION MODELS FOR PREDICTING GLYCINE CONCENTRATION

Matea Gavran, Josip Sacher, Nenad Bolf, Željka Ujević Andrijić
University of Zagreb, Faculty of Chemical Engineering and Technology, Zagreb, Croatia
*mgavran@fkit.hr

INTRODUCTION

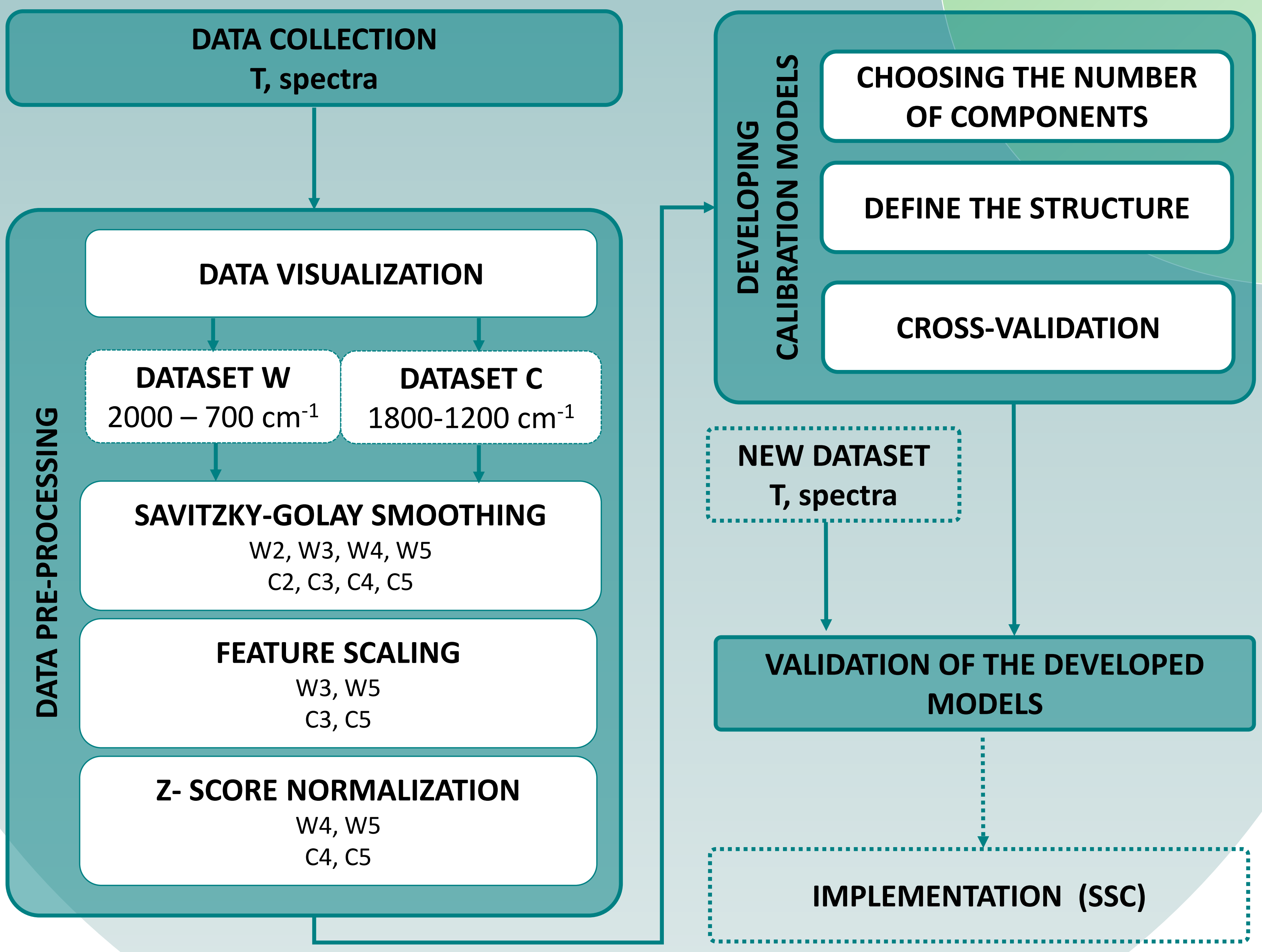
Accurate measurement of solution concentration is of great importance for crystallization processes since supersaturation is prerequisite for nucleation and crystal growth. The most commonly used method for in-situ concentration measurements is using ATR-FTIR spectroscopy. Calibration is necessary to convert absorbance data into the solution concentration. Partial Least Squares Regression and Principal Component Regression are methods to model a response variable (concentration) when there are a large number of predictors such as absorbances.

METODOLOGY



Experimental set-up: (1) FTIR; (2) ATR-FTIR probe; (3) Stirrer; (4) Temperature sensor; (5) Thermostat; (6) Computer for data aquisition

Temperature and spectra was collected continuously in-situ for 10 different concentrations.



DATA COLLECTION
T, spectra

DATA PRE-PROCESSING

DATA VISUALIZATION

DATASET W
2000 – 700 cm^{-1}

DATASET C
1800-1200 cm^{-1}

SAVITZKY-GOLAY SMOOTHING
W2, W3, W4, W5
C2, C3, C4, C5

FEATURE SCALING
W3, W5
C3, C5

Z- SCORE NORMALIZATION
W4, W5
C4, C5

DEVELOPING CALIBRATION MODELS

CHOOSING THE NUMBER OF COMPONENTS

DEFINE THE STRUCTURE

CROSS-VALIDATION

NEW DATASET
T, spectra

VALIDATION OF THE DEVELOPED MODELS

IMPLEMENTATION (SSC)

PLSR and PCR models were developed on 10 different datasets.

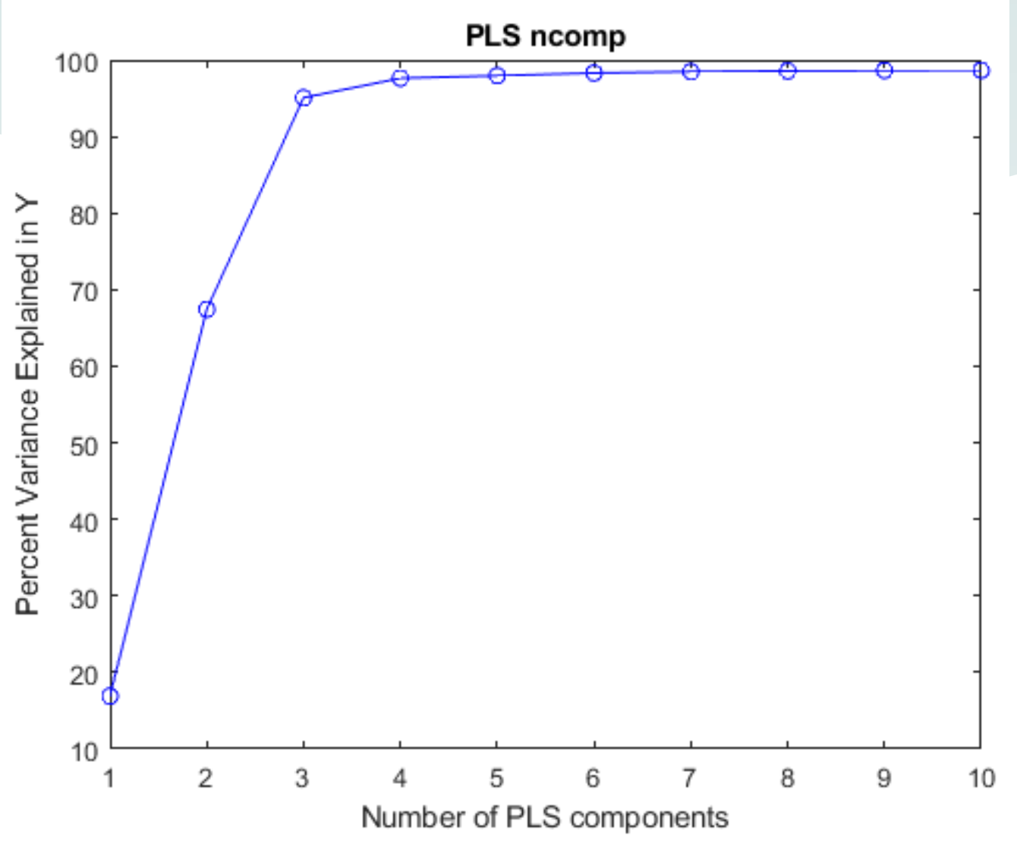
To compare the performance of the models, both PLSR and PCR were fitted with the same number of components.

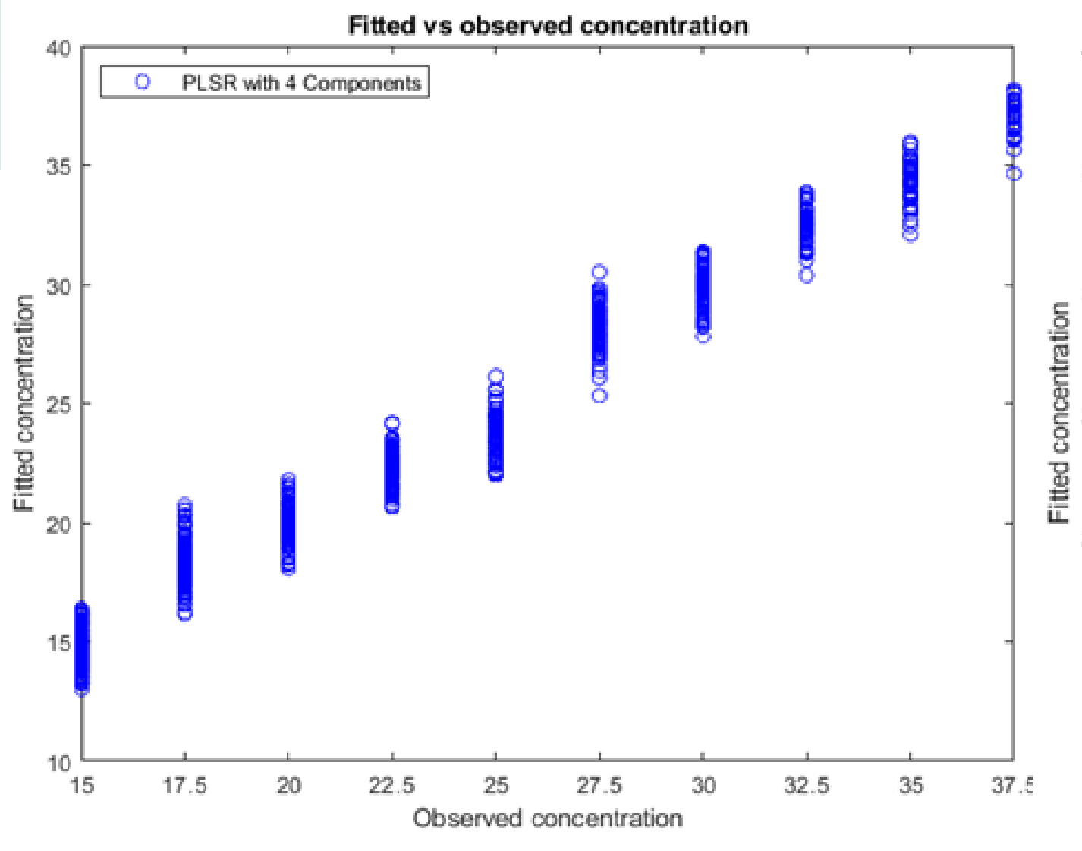
Models were validated on a new dataset.

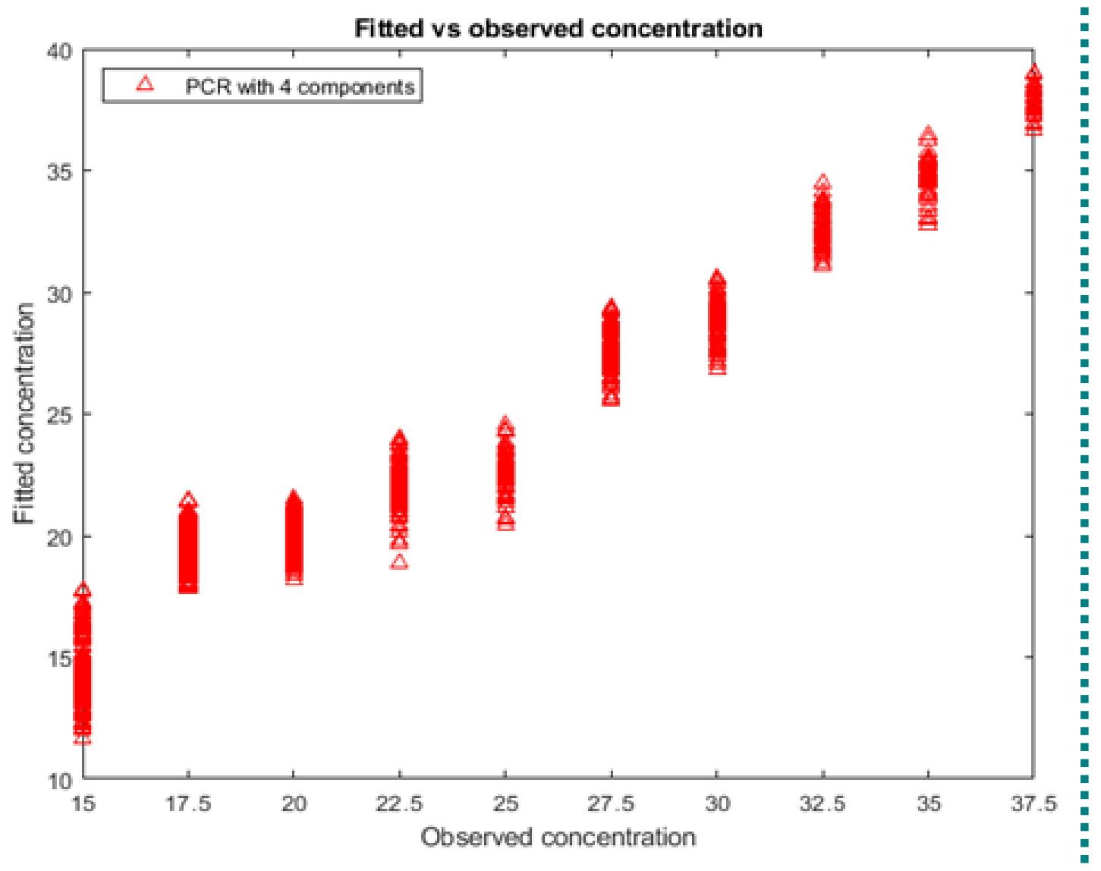
RESULTS

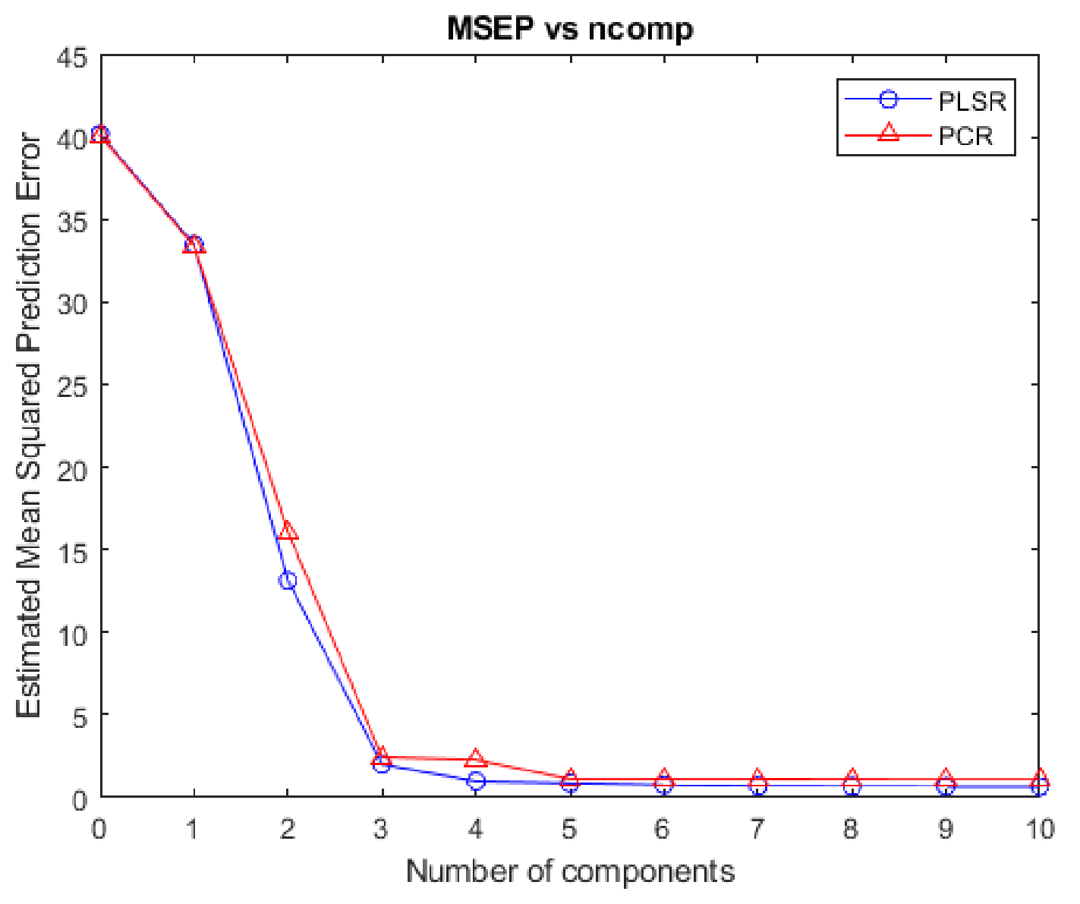
DATASET	NUMBER OF COMPONENTS	PLSR			PCR		
		MSEP	R^2	VALIDATION R^2	MSEP	R^2	VALIDATION R^2
W1	9	1.8938	0.9567	0.9587	31.5119	0.2326	0.1209
W2	6	1.6573	0.9594	0.9654	2.2380	0.9449	0.9553
W3	2	1.8432	0.9563	0.9456	1.8479	0.9536	0.9440
W4	4	1.0817	0.9740	0.6174	1.8171	0.9551	0.6150
W5	2	1.7351	0.9568	0.6170	1.8479	0.9537	0.6168
C1	**4**	**0.9832**	**0.9780**	**0.9790**	**2.2696**	**0.9439**	**0.9544**
C2	4	1.0543	0.9696	0.9719	1.9787	0.9510	0.9584
C3	5	0.8766	0.9785	0.9786	1.1177	0.9724	0.9669
C4	5	0.9157	0,9775	0.6258	1.3481	0,9666	0.6222
C5	5	0.9218	0.9775	0.6008	1.3463	0.9666	0.6067

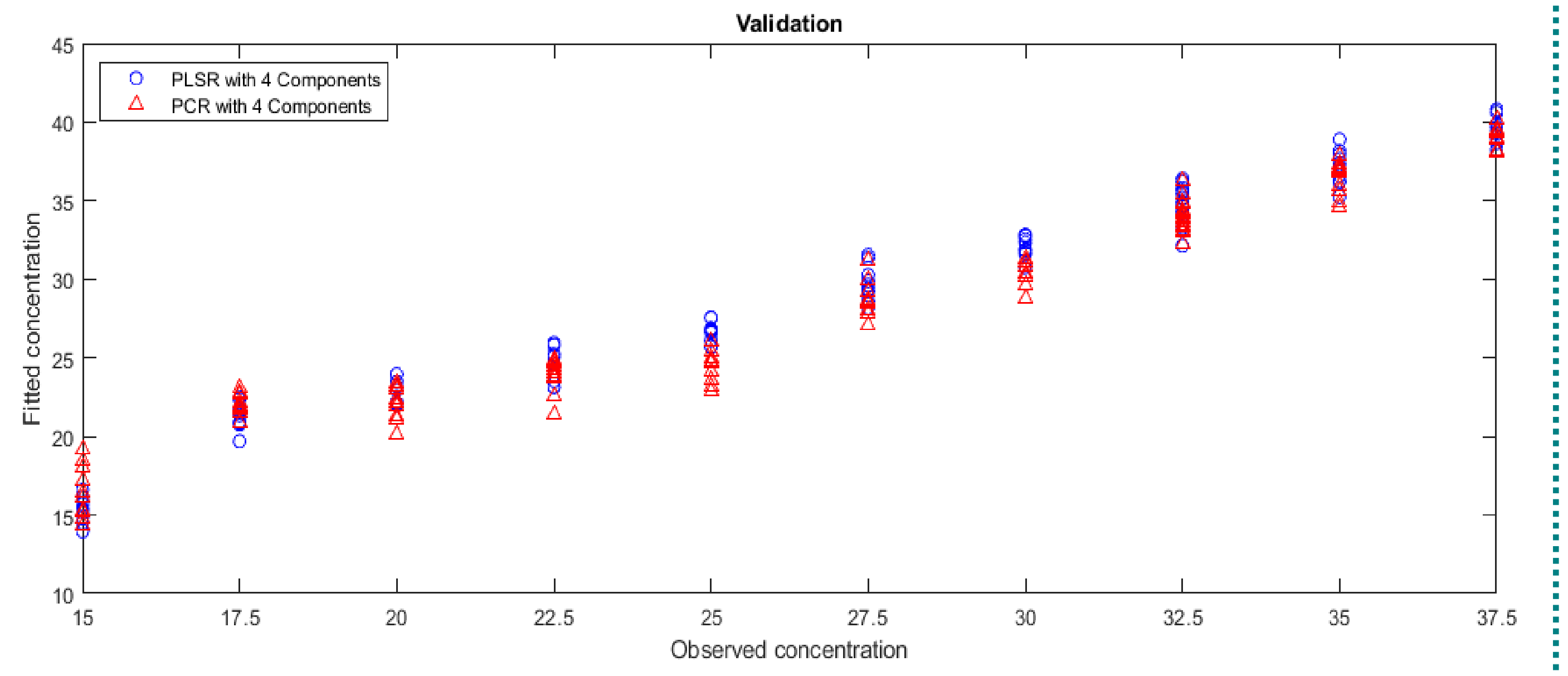
Results for the dataset C1 are presented graphically.


PLS ncomp


Fitted vs observed concentration


Fitted vs observed concentration


MSEP vs ncomp


Validation

PCR creates components without considering the response variable and thus PLSR has better predictive performance in comparison to PCR with the same number of components.

ACKNOWLEDGEMENTS

This research is funded by European Structural and Investment Funds, grant number KK.01.1.1.07.0017 (CrystAPC – Crystallization Advanced Process Control).