

Table S5. The partial least squares regression (PLS-R) models for the determination of THCA concentration where the sample size (n = 36) has been reduced. Calibration:Validation, (n = 27 : n = 9).

Model	Region (nm) ¹	Scatter Correction ²	Derivative ^{2a}	N ³	RMSEC ⁴	R ² _{Cal} ⁵	RMSECV ⁶	R ² _{CV} ⁷	RMSEP ⁸	Pred Bias ⁹	R ² _{Pred} ¹⁰
PLS-R	950 – 1650	DT, SNV and MC	2, 2, 5	36	19.66	0.80	33.42	0.46	30.62	8.03	0.59

¹nm: wavelength in nanometres.

²Preprocessing parameters. DT: detrend; SNV: standard normal variate; MC: mean centering;

^{2a}Derivative pre-treatment: the first digit is the polynomial order, the second digit is the derivative order and the third digit is the data point gap which the derivative is calculated.

³N: number of unique samples.

⁴RMSEC: root mean standard error of calibration.

⁵R²_{Cal}: coefficient of determination of calibration.

⁶RMSECV: root mean standard error of cross validation.

⁷R²_{CV}: coefficient of determination of cross validation.

⁸RMSEP: root mean standard error of prediction.

⁹Pred Bias: calculated prediction bias.

¹⁰R²_{Pred}: coefficient of regression of measured data vs predicted data