

**Table S4.** The partial least squares regression (PLS-R) models for the determination of THCA concentration where the water peak region (1400 – 1450 nm) from the NIR spectra has been removed.

| Model | Region (nm) <sup>1</sup>     | Scatter Correction <sup>2</sup> | Derivative <sup>2a</sup> | N <sup>3</sup> | RMSEC <sup>4</sup> | R <sup>2</sup> <sub>Cal</sub> <sup>5</sup> | RMSECV <sup>6</sup> | R <sup>2</sup> <sub>CV</sub> <sup>7</sup> | RMSEP <sup>8</sup> | Pred Bias <sup>9</sup> | R <sup>2</sup> <sub>Pred</sub> <sup>10</sup> |
|-------|------------------------------|---------------------------------|--------------------------|----------------|--------------------|--------------------------------------------|---------------------|-------------------------------------------|--------------------|------------------------|----------------------------------------------|
| PLS-R | 950 – 1400, -<br>1450 - 1650 | DT, SNV and MC                  | 2, 2, 5                  | 264            | 24.06              | 0.70                                       | 25.32               | 0.67                                      | 26.38              | -0.38                  | 0.58                                         |

<sup>1</sup>nm: wavelength in nanometres.

<sup>2</sup>Preprocessing parameters. DT: detrend; SNV: standard normal variate; MC: mean centering;

<sup>2a</sup>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.

<sup>3</sup>N: number of unique samples.

<sup>4</sup>RMSEC: root mean standard error of calibration.

<sup>5</sup>R<sup>2</sup><sub>Cal</sub>: coefficient of determination of calibration.

<sup>6</sup>RMSECV: root mean standard error of cross validation.

<sup>7</sup>R<sup>2</sup><sub>CV</sub>: coefficient of determination of cross validation.

<sup>8</sup>RMSEP: root mean standard error of prediction.

<sup>9</sup>Pred Bias: calculated prediction bias.

<sup>10</sup>R<sup>2</sup><sub>Pred</sub>: coefficient of regression of measured data vs predicted data; Permutation testing (n = 50) using Wilcoxon test returned p-value < 0.05.