Supplementary Information for

Rapid quantification of goat milk adulteration with cow milk using Raman spectroscopy and chemometrics

Wangfang Li[a], Wei Huang[a], Desheng Fan[a], Xuhui Gao[a], Xian Zhang[a], Yaoyong Meng[a,b]*, Timon Cheng-yi Liu[c]

[a]MOE Key Laboratory of Laser Life Science & Laboratory of Photonic Chinese Medicine, College of Biophotonics, South China Normal University, Guangzhou 510631, China

[b]Analysis and Testing Center, South China Normal University, Guangzhou 510631, China

[c]Laboratory of Laser Sports Medicine, South China Normal University, Guangzhou 510631, China

*Corresponding authors: yaoyongmeng@aliyun.com

Determination of optimal A

To improve the fit of the PLSR model to the original data, we increase the number of PLS components, but this may also reduce its predictive ability due to noise information, which is known as over-fitting. Here, we apply the leave-one-out method for cross-validation, RMSE criterion based on the above cross-validation results are proposed to determine the optimal number of PLS components (A) and prevent the over-fitting problem effectively. For RMSE criterion, the root mean square error (RMSE) is calculated according to equation following, and the optimal A is obtained when the RMSE value is the smallest.[1]

$$RMSE_{CV} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}}$$

$$RMSE = RMSE_{CV} \times \sqrt{\frac{n}{n-A-1}}$$

n is the number of samples in the training set, y_i and $\hat{y}_i$ are the reference value and the predicted value, respectively.

Reference

1. S. Li, T. T. Ng and Z. P. Yao, *Food Chemistry*, 2021, **334**, 127601.