

# **A perspective demonstration on the importance of variable selection in inverse calibration for complex analytical systems**

Yi-Zeng Liang, Yong-Huan Yun, and Qing-Song Xu

*College of Chemistry and Chemical Engineering, Central South University, Changsha 410083, P. R. China. E-mail: [yizeng\\_liang@263.net](mailto:yizeng_liang@263.net)*

*School of Mathematics and Statistics, Central South University, Changsha 410083, P. R. China. E-mail: [dasongxu@gmail.com](mailto:dasongxu@gmail.com)*

Two chemical modeling spaces, say component spectral space and measured variable space, are firstly defined, respectively. From this point of view, classical calibration and inverse calibration can be two kinds of multivariate calibration in chemical modeling. It is worth noting that the intrinsic difference between these two calibration models is not fully investigated. The net analyte signal (NAS) proposed by Lorber<sup>1</sup> based on orthogonal projections can be regarded as the theoretic summary for classic calibration. Also, the tensor calibration for high dimensionally linear data is its natural extension. However, in the case of complex analytical systems, NAS cannot be well defined in inverse calibration due to the existence of uninformative and/or interfering variables. Therefore, application of the NAS cannot improve the predictive performance for this kind of calibration, since it is essentially a technique based on the full-spectrum. From our perspective, variable selection can significantly improve the predictive performance through removing uninformative and/or interfering variables. Although the need for variable selection in the inverse calibration model has already been experimentally demonstrated, it has not aroused so much attention. In this study, we first clarify the intrinsic difference between these two calibration models and then use a new perspective to intrinsically prove the importance of variable selection in the inverse calibration model for complex analytical systems. In addition, we have experimentally validated our viewpoint through the use of one UV dataset and two generated near infrared (NIR) datasets.