XIX CAC 2024
CHEMOMETRICS IN ANALYTICAL CHEMISTRY
Santa Fe, Argentina
9th - 12th september 2024

Regression building made efficient: preprocessing selection and model evaluation

Ezenarro J, Schorn-García D, Busto B, Boqué R

Universitat Rovira i Virgili. ChemoSens group, Department of Analytical Chemistry and Organic Chemistry, Campus Sescelades, Edifici N4, C/Marcel·lí Domingo 1, Tarragona, 43007, Spain

jokin.ezenarro@urv.cat

Chemometrics plays a crucial role in building reliable models for analytical chemistry, particularly using spectroscopic data. Traditionally, this process requires significant expertise to manage data preprocessing and model evaluation effectively, which even for a trained analyst can be time-consuming and suboptimal, even more when it is based on pure trial-and-error. To paliate this process, as other approaches proposed in literature, we introduce two innovative tools that streamline the chemometric analysis of spectroscopic data regression models: the ProSpecTool [1] for automated preprocessing selection and the J-Score [2] for enhanced model performance evaluation.

The ProSpecTool addresses common distortions on spectroscopy such as noise and scatter-effects, quantifying and correcting them through an objective criteria-based system that mimics expert decision-making, helping the analyst search for optimal preprocessing methods for their data. This tool integrates seamlessly with MATLAB, providing a user-friendly interface for analysts. In this toolbox also the J-Score is introduced, a new composite metric, evaluates regression model performance by joining several key parameters into a single indicator, facilitating the comparison of different models, ultimately for the selection of the optimal dimensionality (number of latent variables) and preprocessing methods.

The ProSpecTool demonstrated its ability to select appropriate preprocessing methods that align with those an experienced analyst would choose, thereby standardizing the preprocessing step and reducing reliance on user expertise. The J-Score was validated against traditional metrics using multiple datasets, showing its robustness and reliability in identifying the most suitable models for different types of spectroscopic data. Ultimately, the integration of the ProSpecTool and the J-Score into the analyst's workflow significantly enhances the efficiency of spectroscopic analyses. These tools not only save time and reduce the skill barrier for model optimization but also improve the reproducibility and generalizability of chemometric models, paving the way for broader applications in analytical chemistry.

References

[1] Ezenarro J, Schorn-García D, Busto O, Boqué R, Chemometrics and Intelligent Laboratory Systems, 2024, 246, 105096.

[2] Ezenarro J, Schorn-García D, Aceña L, Mestres M, Busto O, Boqué R, Chemometrics and Intelligent Laboratory Systems, 2023, 240, 104883.