Chemometric Identification of Reaction Mechanisms from In-Situ Raman Spectra

P. Adrian Baum[1*], Jose Delgado[2], Martin Gerlach[1], Sébastien Leveneur[3], Christof Hamel[1]

1 Otto-von-Guericke University Magdeburg, Chair of Chemical Process Engineering, Universitätsplatz 2, 39106 Magdeburg, Germany;2 INSA Rouen Normandie, Laboratory for Chemical Process Safety, 685 avenue de l'Université, Rouen F-76801, France;3 IRCELYON, UMR5256, CNRS-Université de Lyon, Villeurbanne F-69626, France;

**Corresponding author: philipp.baum@ovgu.de*

Highlights

- Method combines PCA, MCR-ALS and peak separation for spectral data.
- Library of PCA score patterns to identify reaction mechanisms in complex spectra.
- Reaction network analysis and kinetic modelling

1. Introduction

Recent advances in the programming capabilities of large language models and the availability of open-source or commercial tools have significantly simplified and democratized the automated analysis of reaction data. This development offers new opportunities, since established chemometric analysis methods are not yet fully adapted in the chemical engineering community. However, AI generated code may contain hidden implementation errors and available software depends on the correct application by the user. Any approach must produce physically interpretable results. Therefore, every workflow should be validated with well-known benchmark data.

In this work, we present a combined method of **principal components analysis (PCA), multivariate curve resolution (MCR-ALS)** and physics-informed peak decomposition. This method is validated with a **synthetic Raman dataset**.

In-situ Raman data of the **acid-catalyzed esterification of glycerol** are then used to demonstrate the approach. The esterification of glycerol is industrially relevant, as esters are either a desired product in the production of glycerides or act as an essential intermediate in the hydrochlorination of glycerol [1].

2. Methods

Experiments were performed in a heat flow calorimeter (EasyMax 102 HFCal, 400 mL, Mettler Toledo) with in-situ Raman spectroscopy (ReactRaman 802l, Mettler Toledo, 10 s resolution, 3100 wavenumber channels). Investigated parameters include temperature (100 – 150°C), ratio of glycerol to adipic acid, and catalyst type (p-Toluenesulfonic acid and H_2SO_4). Adipic acid and catalyst were dosed in separate perturbation steps, therefore each experiment provides both calibration data and information on different reaction networks. The combination of in-situ Raman spectroscopy and reaction calorimetry thus yielded an information-dense dataset.

Raman bands for specific functional groups of the pure component are established in the literature. However, they overlap in the reaction mixture, forming clusters. These clusters were fitted as a sum of Gaussian functions, integrated, and compared with transient profiles obtained by PCA and MCR-ALS.

Synthetic time series of Raman spectra were generated. Starting with an arbitrary stoichiometry matrix, the user defines the kinetic parameters, initial concentrations and visibility of each component in the spectra. Individual spectra are constructed as a linear combination of pure component spectra, which in turn are represented as sums of Gaussian peaks, scaling linearly with the corresponding component's concentration. Optionally, a component can be modeled by a concentration-dependent baseline shift to mimic fluorescence. Random noise is superimposed on the summed spectra to reproduce experimental uncertainty.

3. Results and discussion

Using synthetic datasets, we established a **library of characteristic PCA score trajectories** for well-known basic reaction types.

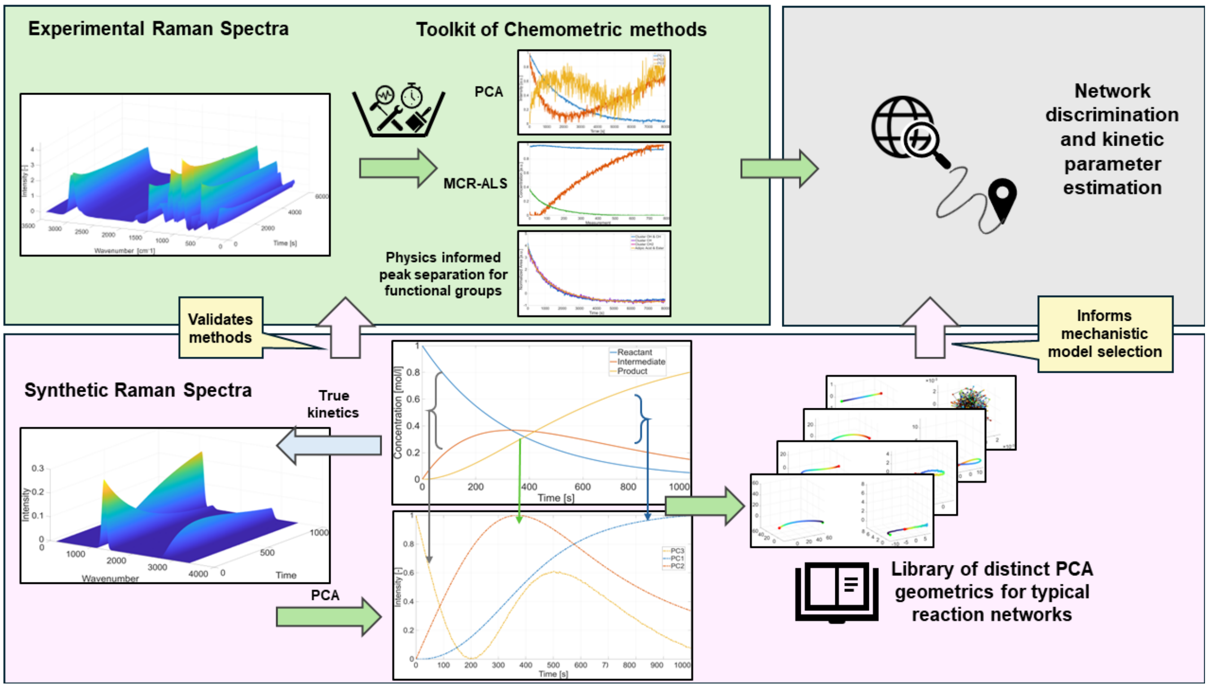

Figure 1: The experimental data are automatically pre-treated, analyzed via chemometric methods and yield concentration profiles for subsequent kinetic parameter estimation. Synthetic data are used both to validate the implementation of the method and to compile a library of geometric patterns for reaction mechanism identification.

The library demonstrates that, while PCA score kinetics do not reproduce true concentration profiles, distinct geometrical patterns indicate the underlying reaction mechanism. Preprocessing was shown to have a significant influence on these patterns. The developed method was validated. First, with the synthetic Raman spectra, and subsequently applied to experimental data. Analysis of the catalyzed dimerization of pure glycerol suggests a reaction with one or maximal two intermediates, with the first step being very fast. Adding adipic acid to glycerol reveals a PCA pattern consistent with a one-intermediate-reaction for the esterification.

These intermediates are not directly visible in individual Raman spectra. The functional groups monitored in the physics-informed peak separation belong to the reactants, therefore they display an exponential decline. The PCA-based analysis allows identification of such hidden intermediates through their geometric signatures in score space.

4. Conclusions

The proposed method provides a physically interpretable chemometric tool in Matlab® to extract kinetic information from complex in-situ Raman data. The compiled library of PCA-score trajectories informs reaction network identification. Furthermore, the synthetic spectra allow quantitative validation of the Matlab® tool and ensure that studies of experimental data can be traced back to a numerical foundation.

The approach demonstrates how classical chemometrics can be integrated into automated kinetic modeling workflows while maintaining physical interpretability.

References

[1] A. Sandid, V. Spallina, J. Esteban, *Fuel Processing Technology* **2024**, *253*, 108008. DOI: https://doi.org/10.1016/j.fuproc.2023.108008

Keywords

Chemometrics, mechanistic reaction-network analysis, PCA, Raman spectroscopy