A Hybrid Gaussian Fitting Approach for Multivariate Curve Resolution in Diffuse Reflective UV-Vis Spectroscopy

Reyhane Bahrami Ziabari[a,b], Mathias Sawall[a], Tatiana Ortroshchenko[b], Christoph Kubis[b], Evgenii V. Kondratenko[b], Klaus Neymeyr[a,b]

[a] *Institut für Mathematik, Universität Rostock, Ulmenstrasse 69, 18057 Rostock, Germany*
[b] *Leibniz-Institut für Katalyse, Albert-Einstein-Strasse 29a, 18059 Rostock, Germany*

Abstract

Spectroscopic data analysis is used to study multicomponent systems in catalyzed reactions, but overlapping spectra make it hard to separate the individual spectra of the different chemical species and their contribution to the overall spectrum. Multivariate Curve Resolution methods often produce non-unique solutions due to so-called rotational ambiguity. In this work, we introduce a hybrid approach to improve the analysis and interpretation of spectra series formed by a modest number of overlapping Gaussians. The method combines Multivariate Curve Resolution with Gaussian peak modeling to obtain physically meaningful spectral profiles and reduce the ambiguity of feasible solutions. Gaussian parameters (amplitude, center, and width) are optimized using nonlinear least-squares fitting, and clustering is applied to group components with similar behavior. This approach is applied to both model and experimental datasets from Diffuse Reflectance Ultraviolet-Visible spectroscopy (DR UV-Vis). The results show that the hybrid approach improves spectral reconstruction, reduces ambiguity within the Area of Feasible Solutions, and provides clearer chemical interpretation.

Keywords: Hybrid chemometric modeling, Multivariate Curve Resolution, Area of Feasible Solutions, Gaussian function fitting, *k*-means clustering

1. Introduction

Spectroscopic data analysis plays a central role in modern chemistry and physics because it helps us to understand what chemical species are present, how much of them there are, and how their concentration change over time. However, experimental spectra often have significant overlap between the absorption bands of different chemical species,which makes it difficult to directly obtain useful information from the measured data. To deal with this difficulty, mathematical decomposition techniques such as Multivariate Curve Resolution (MCR) have been developed. In MCR, the measured spectral dataset D is decomposed into concentration profiles C and pure component spectra S, under chemical or physical constraints [1, 2]. However, the decomposition into concentration profiles and pure component spectra is typically not unique, and many mathematically valid solutions exist. This inherent ambiguity naturally leads to the concept of Areas of Feasible Solutions (AFS). This concept comes from the early work of Lawton and Sylvestre [3], who showed that in the decomposition of two-component spectral data, instead of a unique solution there exists a range of feasible solutions (solution bands), and only under certain conditions can the pure components be uniquely determined. Borgen and Kowalski [4] further developed this idea by providing a geometric representation of the feasible solution space. In their classical theoretical work, they investigated the boundaries of feasible solutions for three-component systems. Later, Golshan, Abdollahi, and Maeder [5] proposed numerical approaches to calculate the boundaries of the AFS for datasets. The polygon inflation algorithm and ray-casting with their ability to handle experimental date are the basis of AFS-based numerical methods implemented in FACPACK, a software package for the computation of AFS sets for two-, three-, and four-component systems [6, 7, 8].

Although MCR provides interpretable decompositions, the obtained spectral profiles often still show overlapping or unresolved features, especially in situation where the DR UV-Vis absorption spectra are known to be formed by linear combination of a modest number of overlapping Gaussians. To improve this, Gaussian fitting can be used to describe pure component spectra with physically meaningful parameters such as peak positions, amplitudes, and

widths. Combining MCR with Gaussian modeling creates a hybrid approach that uses MCR-based decomposition and accurate Gaussian peak fitting.

The concept of optimizing the parameters of analytical peak models through nonlinear least-squares fitting has played a crucial role in chromatographic data analysis [9, 10, 11, 12], in UV-Vis Spectroscopy [13, 14] and others. The fundamental idea is to determine model parameters such as amplitude (A), mean (μ), and standard deviation (σ) by minimizing the sum of squared residuals (SSR) between the experimental data and the fitted model function

$$SSR = \sum_i (y_i - f(x_i; A, \mu, \sigma, \ldots))^2.$$

This mathematical approach ensures that the model function $f(x)$ fits the data by simultaneously optimized all parameters.

In addition to decomposition approaches, clustering methods can also be used to analyze spectral and concentration data. One of the most common clustering algorithms is k-means, originally introduced by Stuart P. Lloyd [15]. The k-means algorithm groups data into k clusters such that similar data points are placed in the same group and the variation within each cluster is minimized [16]. In this work, k-means is applied to cluster the reconstructed concentration profiles associated with the fitted Gaussian components in order to identify components with similar behavior. The Gaussian functions within each cluster are then averaged to form representative spectra, which are used to reconstruct the data. The quality of the clustering-based reconstruction is evaluated by comparison with the original spectra.

In this study, we introduce a hybrid MCR-Gaussian method to better analyze spectroscopic data. The method is first tested on model data and then applied to experimental DR UV-Vis spectra of the carbon deposits formed during the non-oxidative isobutane dehydrogenation to isobutene over a Y_2O_3-stabilized ZrO_2(YZrOx) catalyst at 550°C. The spectra were collected in a time-resolved manner. The same steps are used in both cases so the results can be compared directly. The results show that the proposed approach helps to reduce spectral ambiguity and provides a clearer physical interpretation of the data.

1.1. Model dataset

We apply the proposed hybrid MCR–Gaussian fitting approach on a model-based generated spectral dataset. This dataset contains 50 spectra, each defined on a grid of 1000 uniformly spaced wavelength points between 300 and 800 nm. Each spectrum represents a mixture of three chemical components with distinct Gaussian-shaped absorption bands and time-dependent concentration profiles.

The dataset is based on the standard bilinear mixture model

$$D = C\,S^T,$$

where $D \in \mathbb{R}^{k\times n}$ denotes the measured data matrix, $C \in \mathbb{R}^{k\times q}$ contains the concentration profiles of the three pure components, and $S \in \mathbb{R}^{n\times q}$ represents their corresponding pure component spectra. Each row of D corresponds to one mixture spectrum at a specific time step. Here, k denotes the number of time points, n the number of wavelength variables, and q the number of components. This noise-free dataset enables a controlled evaluation of spectral decomposition algorithms.

Each pure spectrum $S_j(t)$ is generated as a sum of Gaussian peaks

$$S_j(t) = \sum_{i=1}^{p_j} A_{ij} \exp\left(-\frac{(t-\mu_{ij})^2}{2\sigma_{ij}^2}\right),$$

where A_{ij}, μ_{ij}, and σ_{ij} denote the amplitude, center, and width of the i-th Gaussian peak of the j-component, respectively. The spectrum of the first component consists of four bands centered at 320, 365, 410, and 460 nm. The second component's spectrum is characterized by bands at 520, 575, and 630 nm. The spectrum of the third component contains two broad bands at 700 and 760 nm. Together, these form nine distinct Gaussian functions distributed across the spectral range.

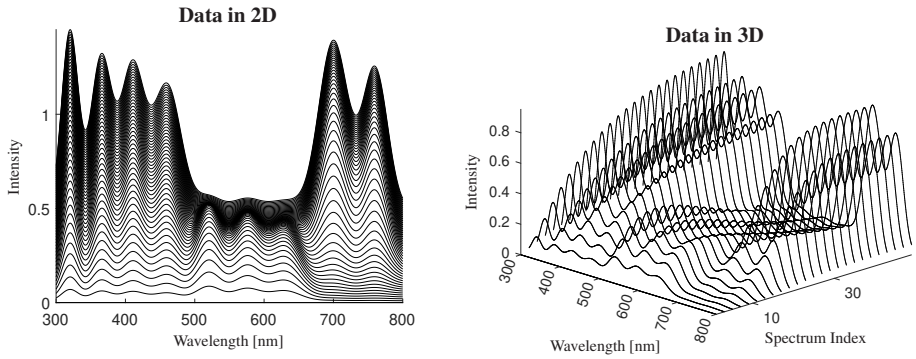


Figure 1: Visualization of the model dataset in two-dimensional (2D) (left) and three-dimensional (3D) (right) representations. The dataset contains 50 noise-free spectra sampled at 1000 wavelength points uniformly distributed between 300 and 800 nm.

The concentration profiles of the compounds, which are characterized by the above spectra, are defined analytically as

$$
\begin{aligned}
C_1(t) &= 1 - \exp(-3t), \\
C_2(t) &= 4.0\, t\, \exp(-4t), \\
C_3(t) &= \frac{1}{1 + \exp(-12(t - 0.4))} - \frac{1}{1 + \exp(4.8)},
\end{aligned}
$$

for time $t \in [0, 1]$, and the third profile $C_3(t)$ is scaled to unit amplitude. The concentration profiles are synthetically generated and do not correspond to a specific experimental kinetic model. They are designed to provide controlled and distinct temporal behaviors for method validation.

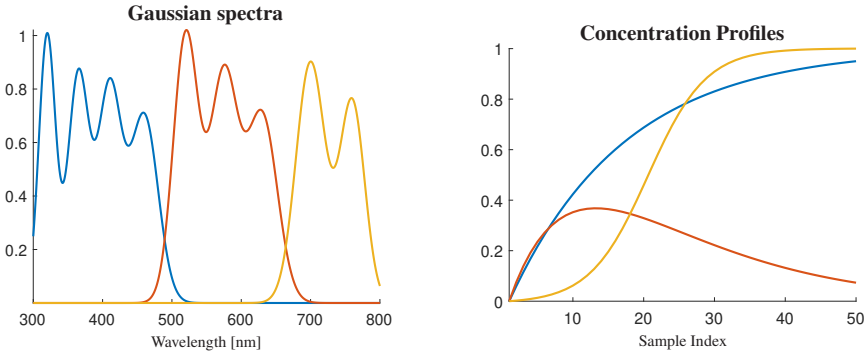


Figure 2: Left: Pure Gaussian spectra of the three components. Right: Their corresponding concentration profiles. The first component (blue) has four peaks at shorter wavelengths, the second (red) shows three peaks in the middle region, and the third (yellow) has two peaks at longer wavelengths.

1.2. Experimental dataset

To evaluate the applicability of the proposed method to real catalytic systems, in situ DR UV-Vis spectra are collected during isobutane dehydrogenation over YZrOx at 823 K using a fixed-bed flow reactor [17]. Along with gas-phase products, carbon-containing species are formed on the catalyst surface as byproducts under the reaction conditions, leading to catalyst deactivation. Such deposits can be detected by DR UV-Vis spectroscopy [18, 19, 20]. From both fundamental and practical perspectives, understanding the mechanism of coke formation is essential for the design of new catalysts and the improvement of existing ones. Carbon-containing species can be formed from the feed alkane and/or the desired product. As their exact nature is usually unclear, their respective DR UV-Vis spectra are also

unknown and can overlap, which makes evaluating them challenging. The experimental dataset of the DR UV-Vis spectra is represented as a matrix $D \in \mathbb{R}^{k\times n}$, where k denotes the number of time-resolved spectra and n the number of wavelength channels spanning the 300–900 nm range. Each spectrum is expressed using the Kubelka–Munk function

$$F(R) = \frac{(1-R)^2}{2R},$$

where R denotes the reflectance of the catalyst.

Figure 3 shows the experimental dataset. In the 2D plot, the y-axis represents F(R), showing a strong overlap between spectral features. In the 3D plot, the z-axis also corresponds to F(R) and illustrates how the optical response changes smoothly over time. These results show the need for methods that can separate overlapping spectra while keeping the interpretation physically meaningful.

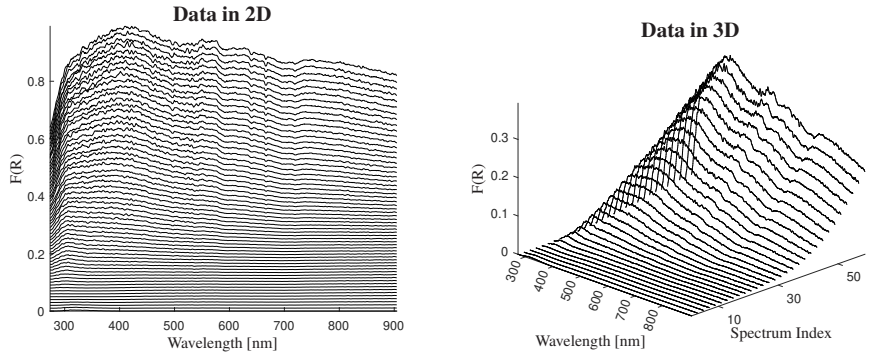


Figure 3: The experimental DR UV-Vis dataset is shown as stacked two-dimensional (2D) traces on the left and a three-dimensional (3D) line representation on the right The dataset consists of 60 time-resolved spectra recorded during the *i*-BDH reaction (*i.e.*, isobutane dehydrogenation: $i\text{-}C_4H_{10} \rightleftharpoons i\text{-}C_4H_8 + H_2$) over a yttria-stabilized zirconia (YZrOx) catalyst. The wavelength range spans 300–900 nm with uniform sampling. The 3D view highlights the smooth time-on-stream evolution, while the 2D stacked plot illustrates the pronounced spectral band overlap that motivates the use of decomposition methods.

2. SVD-based decomposition and mathematical representation

The spectral dataset D is analyzed using Singular Value Decomposition (SVD), $D = U\Sigma V^\top$, where U and V contain the left and right singular vectors, respectively, and Σ is a diagonal matrix of singular values. The SVD helps to identify the number of meaningful components in the data and the level of noise. Following the formulation of [21], any (nonnegative) factorization of a rank-s matrix D of the form $D = XY^\top$, can be written in terms of the SVD basis as

$$D = (U\Sigma T^{-1})(TV^\top),$$

where $T \in \mathbb{R}^{s\times s}$ is an invertible matrix. The elements of T parameterize the family of all possible factorizations, including the subset of chemically meaningful nonnegative ones. Accordingly, the concentration and spectral profiles can be written as $C = U\Sigma T^{-1}$ and $S = VT$, respectively.

According to the Perron–Frobenius theorem, if both $D^\top D$ and $DD^\top$ are nonnegative and irreducible, then the first right and left singular vectors, $V(:,1)$ and $U(:,1)$, can be chosen to have strictly positive entries. This implies that any nontrivial linear combination of the remaining singular vectors, $V(:,2:s)$ or $U(:,2:s)$, must contain at least one negative component. Consequently, the first singular vector defines a unique positive direction in the feasible subspace, while the remaining singular vectors span the space of possible rotations of the factorization. The inherent *rotational ambiguity* of $D = CS^\top$ can be interpreted geometrically as the set of all transformations T that preserve the nonnegativity of both factors, i.e., $C = U\Sigma T^{-1} \geq 0$ and $S = VT \geq 0$. Each admissible transformation T thus corresponds to one valid decomposition, and together these transformations define a compact, bounded region in the parameter space of T. This region is known in MCR as the Area of Feasible Solutions (AFS).

Figures 4 and 5 show the Singular Value Decomposition (SVD) of the model and experimental datasets. Figure 5 shows that the experimental dataset contains three dominant absorbing species. Accordingly, three spectral components are used in the later analyses.

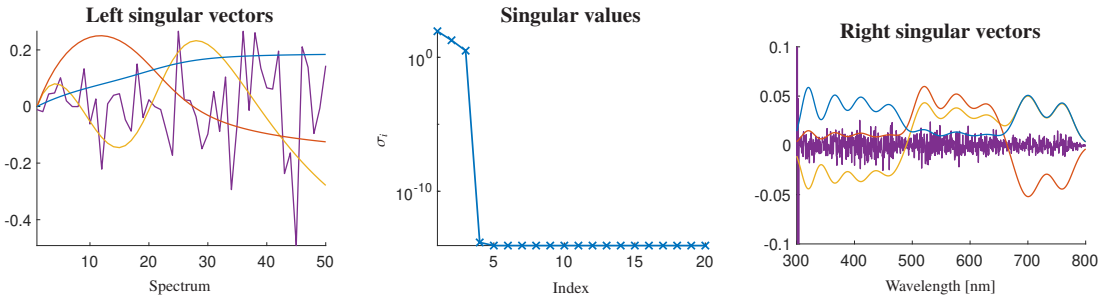


Figure 4: Singular Value Decomposition (SVD) of the model dataset. Left: first four left singular vectors (U), representing temporal concentration profiles. Middle: singular values (σ_i), indicating three dominant components. Right: first four right singular vectors (V), representing the main spectral features.

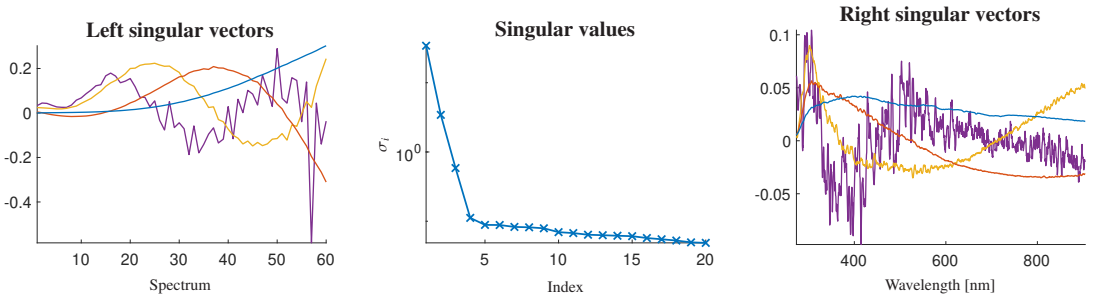


Figure 5: Singular Value Decomposition (SVD) of the experimental dataset. Left: first four left singular vectors (U), representing temporal concentration patterns. Middle: singular values (σ_i), indicating three dominant components above the noise level. Right: first four right singular vectors (V), representing the main spectral features.

3. Classical decomposition into Gaussians

3.1. Gaussian fitting and spectral reconstruction

Our goal is to approximate the given spectral dataset by linear expansion in terms of Gaussian functions. The measured data matrix D is approximated by a bilinear model

$$D_{kj} \approx \sum_{l=1}^{z} C_{kl} \exp\left(-\frac{(x_j - a_l)^2}{b_l^2}\right),$$

where z denotes the number of Gaussian used in the decomposition. The parameters a_l and b_l denote the center and width of the l-th Gaussian component, respectively. The coefficients C_{kl} describe the contribution to the k-th spectrum. The dataset is reconstructed by linearly combining the Gaussian with their corresponding amplitudes C_{kl}.

The optimal parameters $\{a_l, b_l, C\}$ are obtained by minimizing the sum of squared residuals (SSR) between the measured data and the reconstructed model

$$\min_{a_l, b_l, C} \|D - \hat{D}\|_F^2, \qquad \hat{D}_{kj} = \sum_{l=1}^{z} C_{kl}\, \delta_l(x_j).$$

Here, $\|\cdot\|_F$ denotes the Frobenius norm. MATLAB's `lsqnonlin` solver is used to perform the nonlinear least-squares minimization, with physically meaningful bounds applied to all parameters. The accuracy of the reconstruction is measured by the relative Frobenius error, which is defined as

$$E_{\text{rel}} = \frac{\|D - \hat{D}\|_F^2}{\|D\|_F^2}. \tag{1}$$

Since the model dataset is noise-free, an ideal reconstruction corresponded to $E_{\text{rel}} \approx 0$. Figure 6 shows representative results for $z = 8$ and $z = 9$. In particular, the model with $z = 9$ provides an excellent reconstruction, closely matching the original spectra (black) in the right panel.

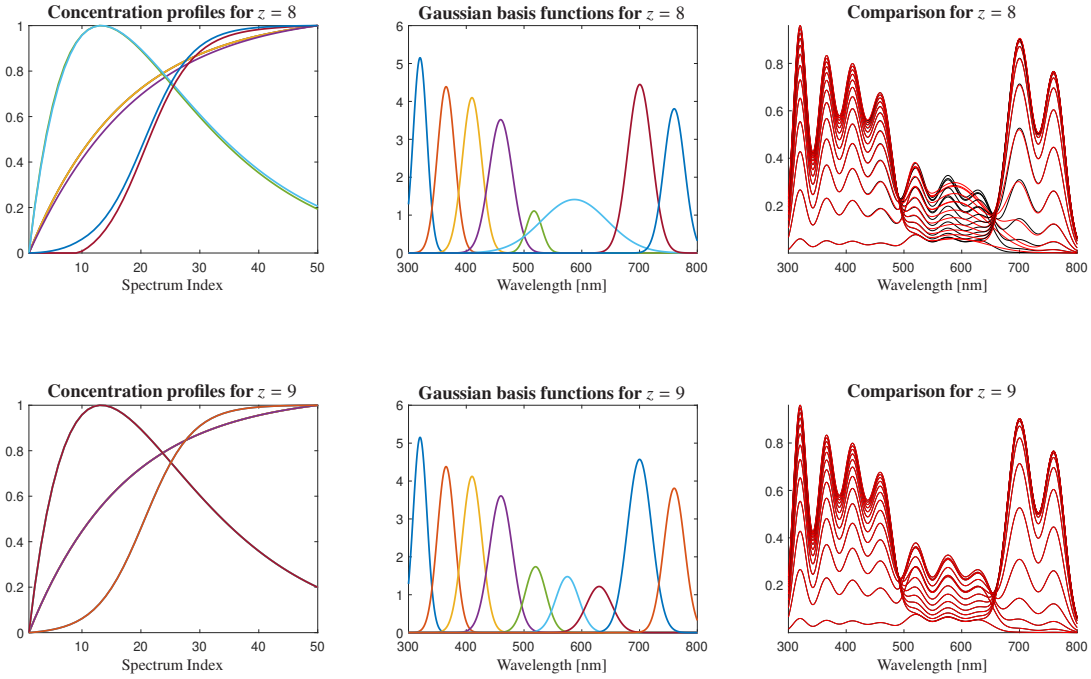


Figure 6: Gaussian decomposition of the model dataset with $z = 8$ and $z = 9$ Gaussian components. The left panels show the concentration profiles, the middle panels show the Gaussian functions, and the right panels show the reconstructed spectra (red) compared with the original spectra (black). For clarity of visualization, only a subset of the reconstructed spectra is displayed. In the upper-left panel ($z = 8$), eight concentration profiles are present; however, due to overlap, only five curves are distinguishable. In the lower-left panel ($z = 9$), nine concentration profiles are plotted, but because of stronger overlap, only three curves are visible.

Figure 7 shows the relative Frobenius reconstruction error (Eq. 1) for different numbers of Gaussian components applied to the experimental dataset. For each value of z, the dataset is reconstructed using the fitted Gaussian components, and the Frobenius norm of the difference between the original data and the reconstructed one is computed. This error is then normalized by the Frobenius norm of the original dataset to obtain the relative reconstruction error. Based on Figure 7, Gaussian decompositions with $z = 7$ and $z = 8$ are applied for the experimental dataset, as shown in Figure 8.

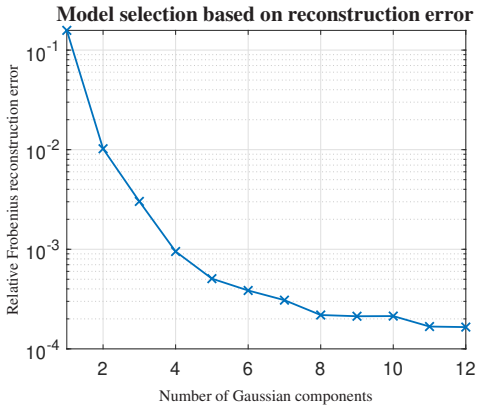


Figure 7: Relative Frobenius reconstruction error as a function of the number of Gaussian components z for the experimental dataset. It decreases quickly from $z = 1$ to about $z = 6$, then changes only slightly up to $z = 9$, and becomes almost constant for $z > 8$.

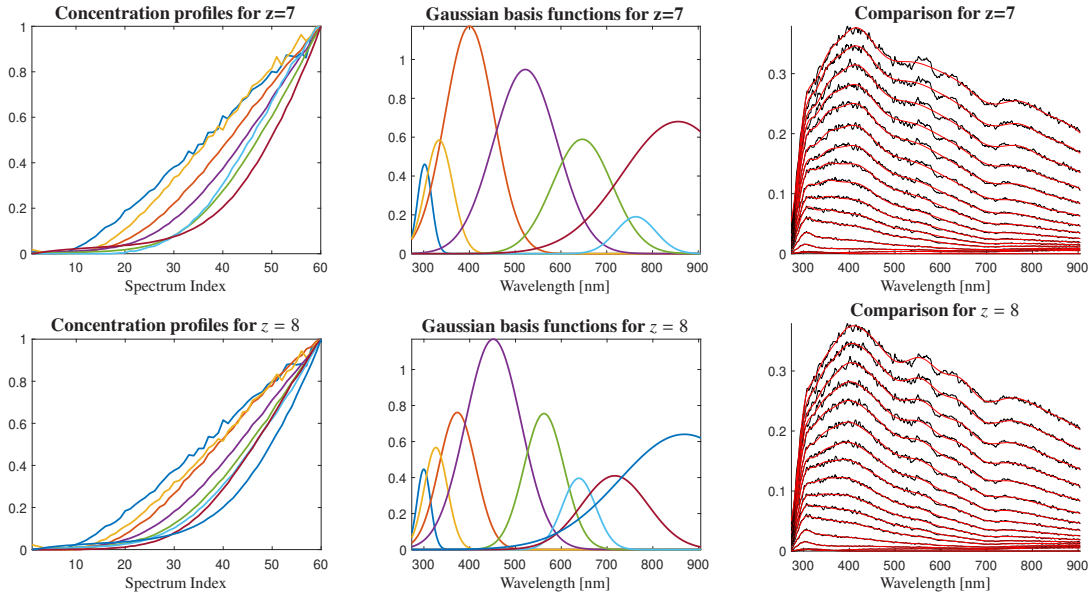


Figure 8: Gaussian decomposition of the experimental dataset with $z = 7$ and $z = 8$ Gaussian components. The left panels show the concentration profiles, the middle panels show the Gaussian functions, and the right panels show the reconstructed spectra (red) compared with the original spectra (black). For clarity of visualization, only a subset of the reconstructed spectra is displayed.

The results obtained from the model dataset show that the method can reconstruct overlapping spectral components under ideal noise-free conditions. The reconstruction error decreases close to zero for appropriate choices of z, indicating that the method successfully recovers the underlying structure of the data.

For the experimental dataset, although perfect reconstruction cannot be expected due to noise and physical complexity, the Frobenius error analysis shows that increasing the number of Gaussian components up to approximately $z = 8$ leads to a significant reduction in reconstruction error, while further increases lead to little improvement. Moreover, the comparison between reconstructed and original spectra confirms that the spectral features are well reproduced.

3.2. Clustering-based reconstruction of spectral data

After decomposing both datasets into multiple Gaussian functions, we analyze the concentration profiles and see that several of them have similar behavior. This suggests that some of these curves likely represent related components

and can therefore be grouped together into the same cluster. Data clustering aims to group n objects into K clusters such that similar objects are placed in the same group. The k-means algorithm, also known as Lloyd's algorithm [15], is an iterative method that divides a dataset of n data points into K clusters. Each data point is assigned to exactly one cluster, which is represented by its centroid, and the number of clusters K is fixed before the algorithm starts. Let

$$X = \{x_i\}_{i=1}^{n} \subset \mathbb{R}^d$$

be a set of n points in a d-dimensional space, which are to be grouped into K clusters

$$C = \{C_k\}_{k=1}^{K}.$$

Cluster assignment is done in a way that minimizes the sum of squared distances between the points in each cluster and the mean of that cluster. Let μ_k denote the centroid (mean) of cluster C_k. The squared error associated with cluster C_k is defined as

$$J(C_k) = \sum_{x_i \in C_k} \|x_i - \mu_k\|^2.$$

The overall objective of the k-means algorithm is to minimize the total within-cluster sum of squared errors over all K clusters [16]

$$J(C) = \sum_{k=1}^{K} \sum_{x_i \in C_k} \|x_i - \mu_k\|^2.$$

Application to Gaussian components. For further analysis of the fitted Gaussian components, the reconstructed concentration profiles are grouped using the k-means clustering algorithm. The aim is to find Gaussian components with similar behavior over time or across the spectra. Based on the SVD analysis presented earlier, both the model and experimental datasets are found to contain three dominant chemical components. Thus, three clusters are chosen to represent three chemically meaningful groups. The Gaussian functions in each cluster are averaged to form a representative spectrum, and the data are then reconstructed. The clustered reconstruction is evaluated by comparing it with the original data.

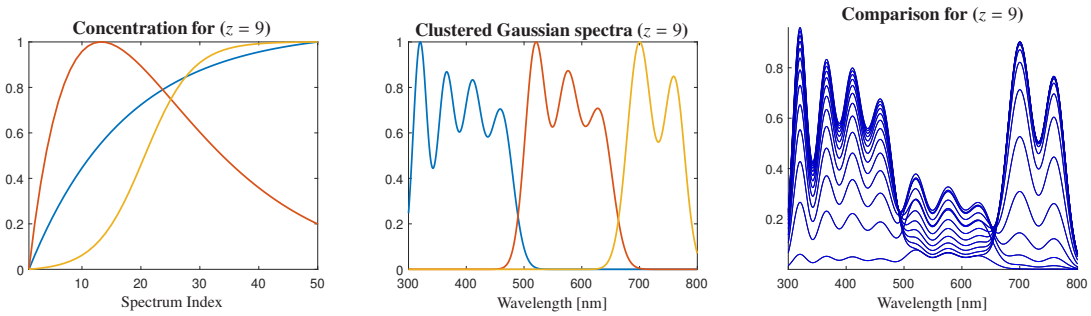


Figure 9: Clustering-based reconstruction for the model dataset using $z = 9$ fitted Gaussian functions group into $k = 3$ clusters. From left to right, the figure shows the concentration profiles, the cluster-averaged Gaussian basis functions, and a comparison between the reconstructed (blue) and original (black) spectra. For clarity of visualization, only a subset of the reconstructed spectra is displayed.

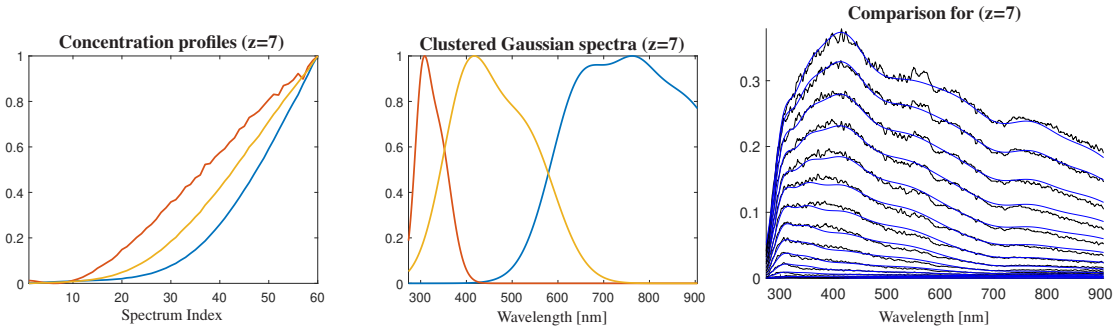


Figure 10: Clustering-based reconstruction for the experimental dataset using $z = 7$ fitted Gaussian functions grouped into $k = 3$ clusters. From left to right, the figure shows the concentration profiles, the cluster-averaged Gaussian basis functions, and a comparison between the reconstructed (blue) and original (black) spectra. For clarity of visualization, only a subset of the reconstructed spectra is displayed.

4. FACPACK–AFS Analysis

FACPACK is an easy-to-use software for the computation of nonnegative multi-component factorizations and for the numerical approximation of the AFS [8]. For a three-component system, the AFS is a subset of the 2D plane. To compute the AFS numerically one can use the triangle enclosure method and the polygon inflation method. FACPACK uses the polygon inflation method. We analyze both the model and experimental datasets using FACPACK. This results in three feasible regions (AFS_1, AFS_2, AFS_3), as shown in Figure 11.

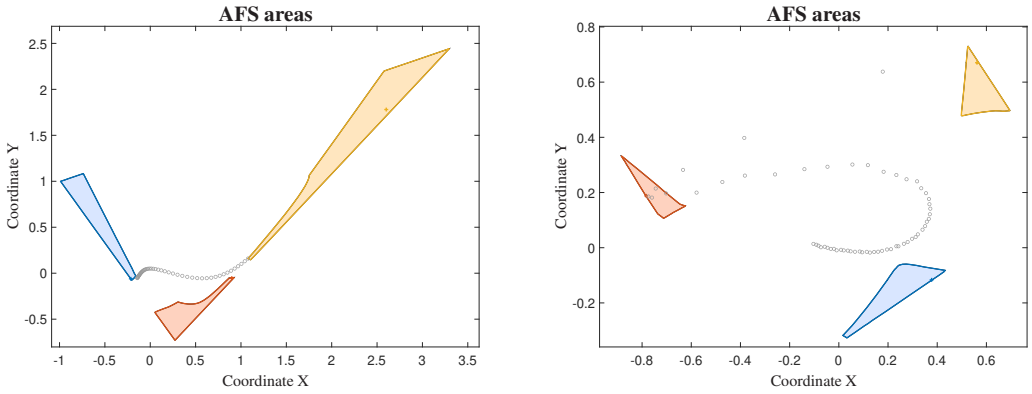


Figure 11: Area of Feasible Solutions (AFS) obtained using the FACPACK *Polygon Inflation* module for the model (left) and experimental (right) datasets. In both cases, the feasible domain is divided into three polygonal regions corresponding to the three species.

4.1. Gaussian Fitting of Feasible AFS Solutions

Each point in an AFS corresponds to a valid spectral profile that follows the bilinear model. By sampling points within each AFS region, we obtain a set of spectra that represents the full range of feasible solutions.

For every spectrum in a given AFS, Gaussian functions with different numbers of components ($z = 1, 2, 3, 4$) are fitted, and the optimal number of Gaussians in each case is determined by minimizing the relative Frobenius reconstruction error. Within each AFS region, the spectrum achieving the lowest reconstruction error is selected as the best-fitting solution.

The corresponding results (including the original and reconstructed spectra, the individual Gaussian components, and the position of the selected grid point within the AFS) are presented in the following figures for both the model and experimental datasets.

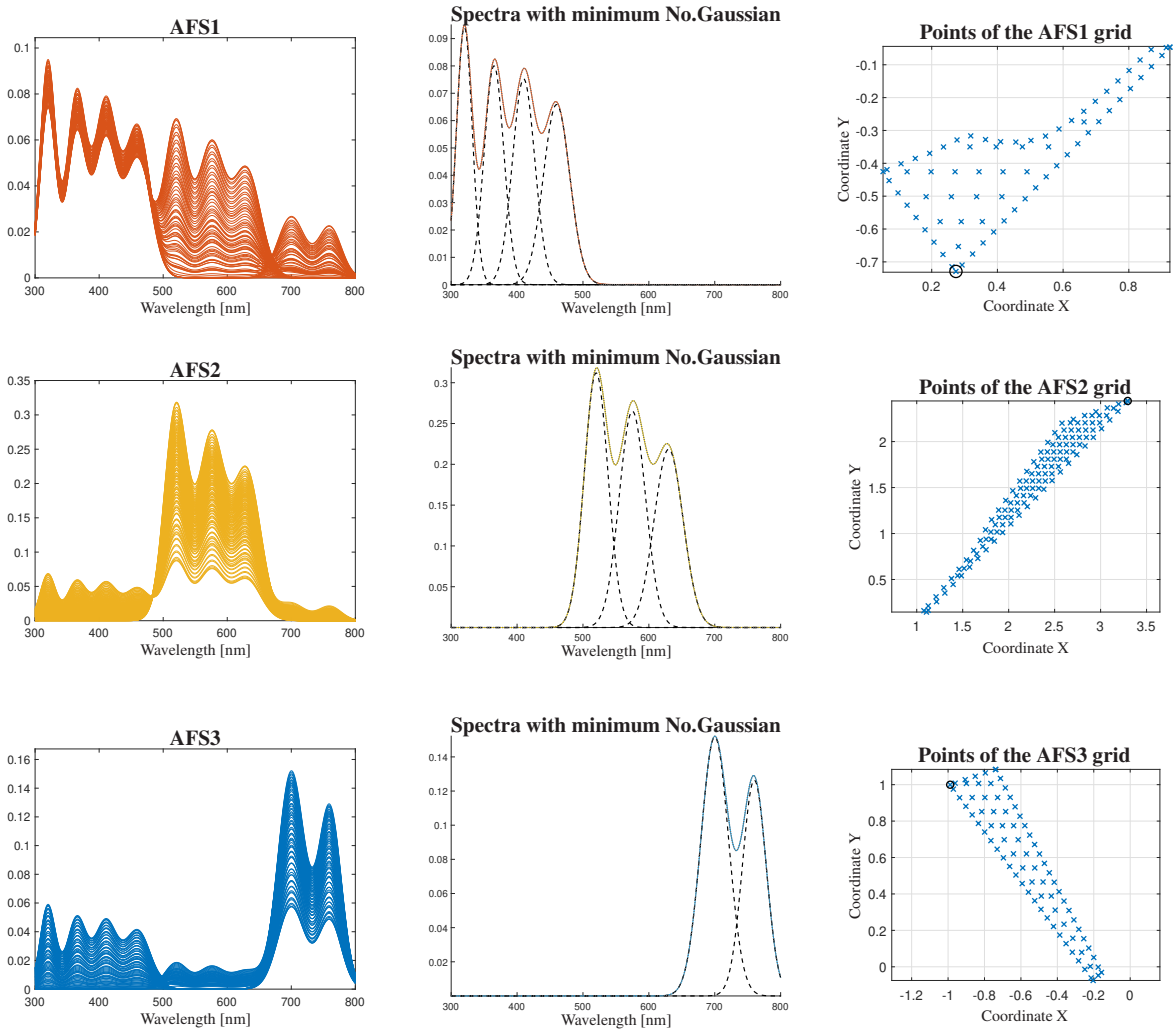


Figure 12: Results for the selected AFS region (model dataset). The left panel shows the spectra from sampled grid points within the feasible region. The middle panel shows the Gaussian decomposition of the best-fitting spectrum (solid: original spectrum; dashed: Gaussian components). The right panel shows the sampled grid points in the AFS plane, with the circled point indicating the solution with the lowest reconstruction error.

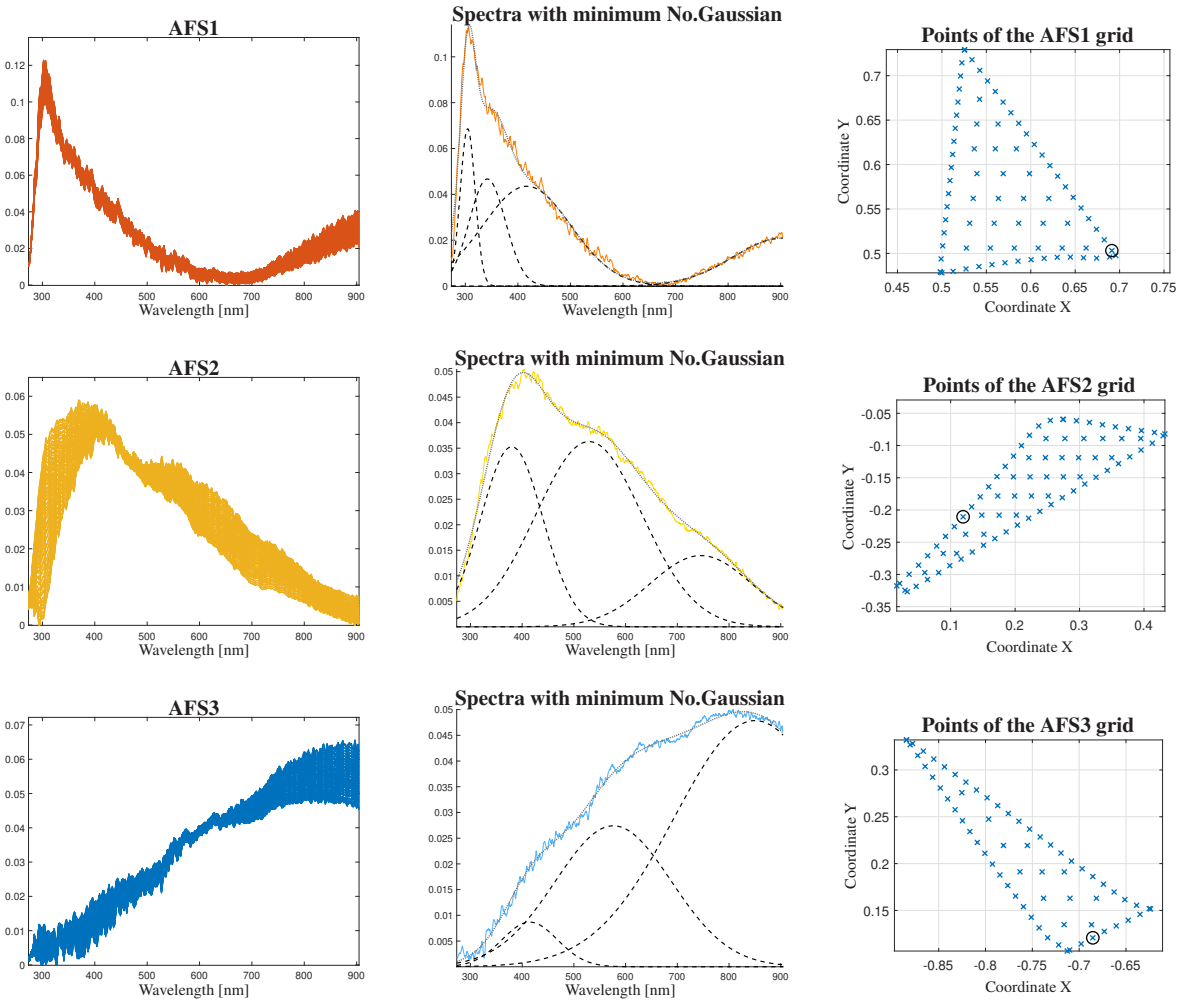


Figure 13: Results for the selected AFS region (experimental dataset). The left panel shows the spectra from sampled grid points within the feasible region. The middle panel shows the Gaussian decomposition of the best-fitting spectrum (solid: original spectrum; dashed: Gaussian components). The right panel shows the sampled grid points in the AFS plane, with the circled point indicating the solution with the lowest reconstruction error.

Comparison of Fit Quality Across AFS Regions. To compare the fit quality across the feasible regions, color maps of the relative reconstruction error is generated for each AFS region. Blue areas indicate better fits, while red areas indicate poorer fits. Figures 14 and 15 show how the number of Gaussian components (z = 1-4) affects the fit quality in each region.

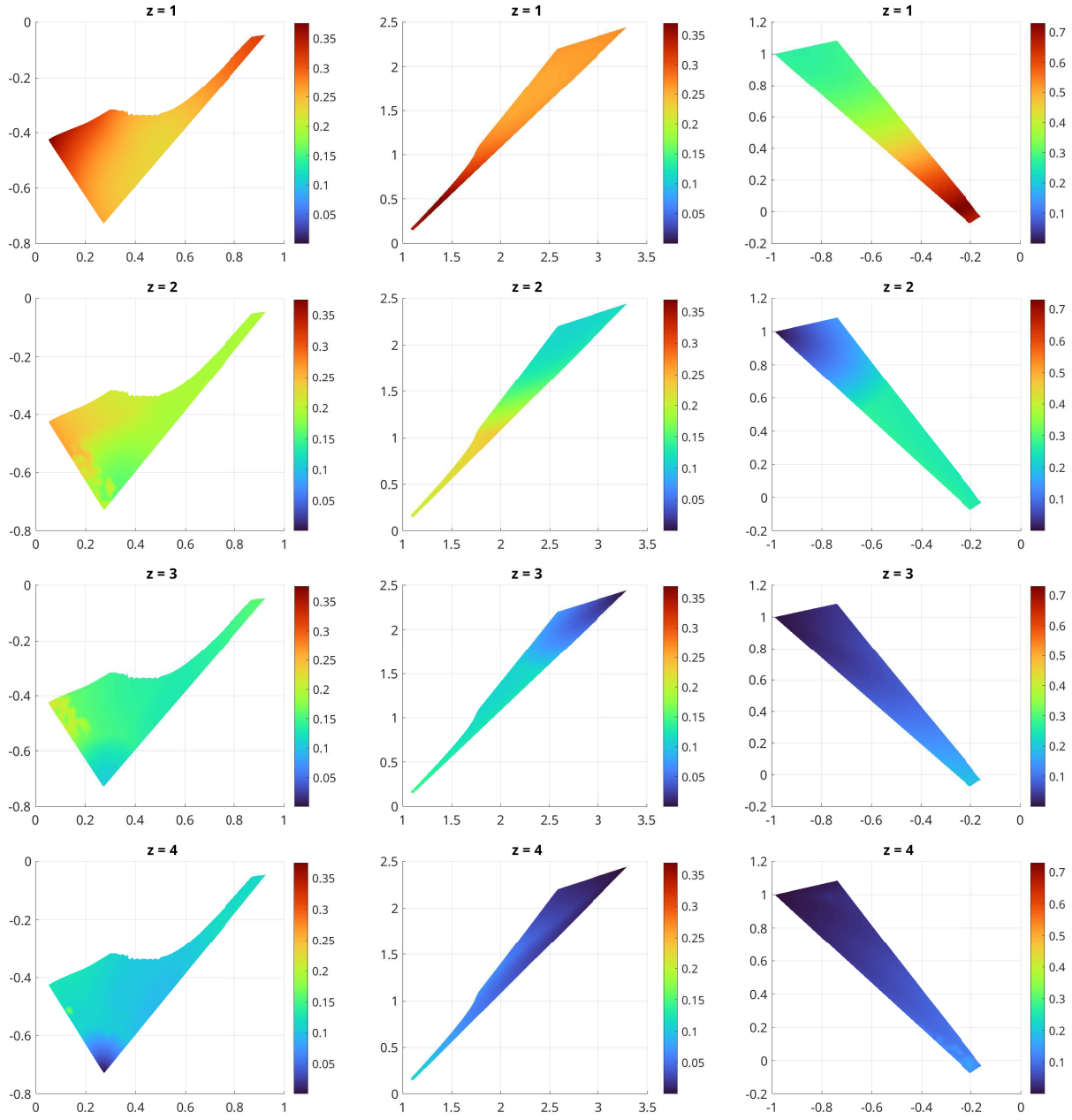


Figure 14: Reconstruction error maps for the **model dataset**. Each figure column corresponds to one AFS region: the first column corresponds to **AFS1**, the second column to **AFS2**, and the third column to **AFS3**. Within each column, the number of Gaussian functions increases from $z = 1$ (top) to $z = 4$ (bottom). Dark blue indicates better fits, while dark red indicates poorer fits.

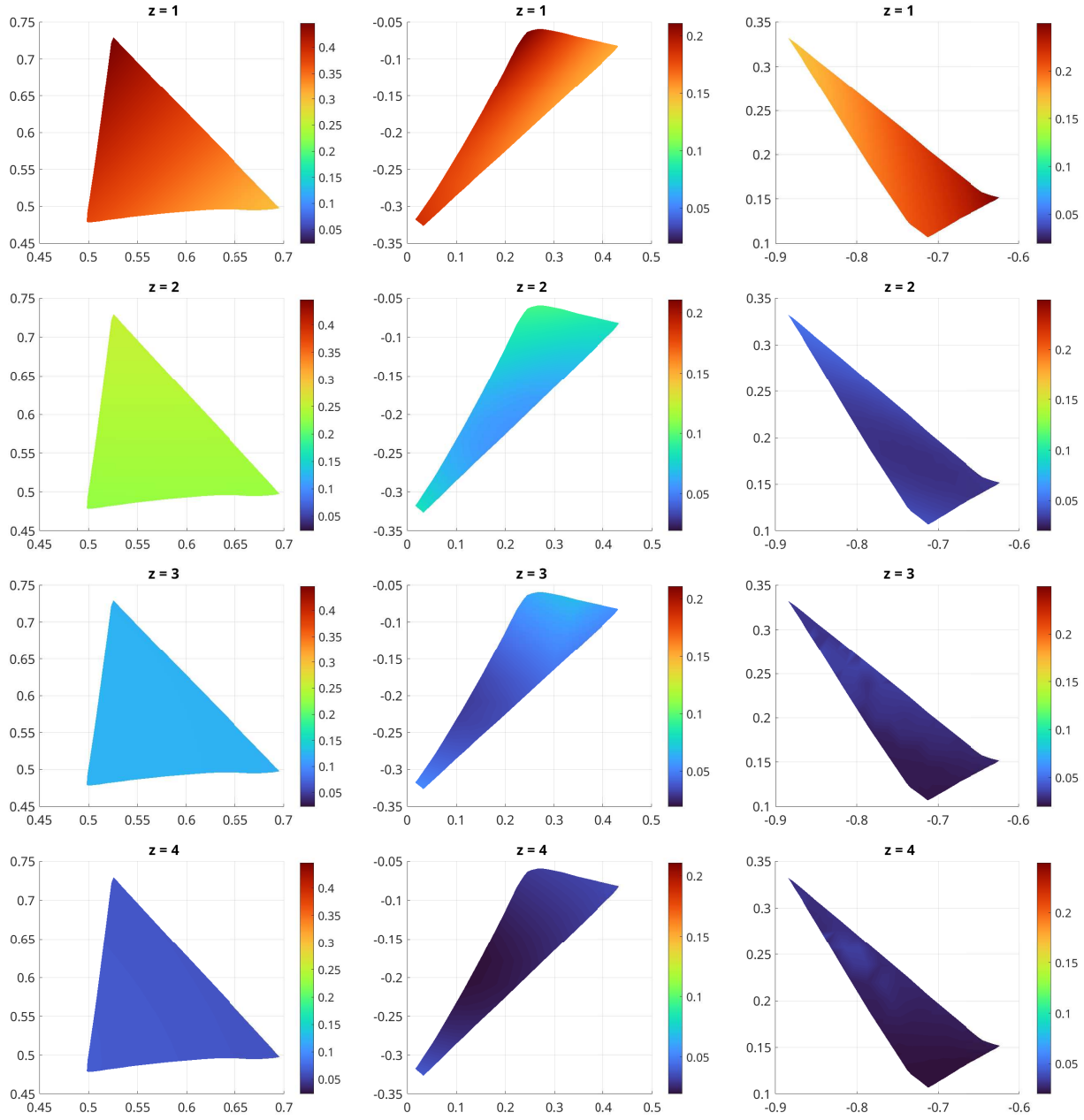


Figure 15: Reconstruction error maps for the **experimental dataset**. Each figure column corresponds to one AFS region: the first column corresponds to **AFS1**, the second column to **AFS2**, and the third column to **AFS3**. Within each column, the number of Gaussian functions increases from $z = 1$ (top) to $z = 4$ (bottom). Dark blue indicates better fits, while dark red indicates poorer fits..

4.2. Gaussian fitting of FACPACK-derived spectra

Using the FACPACK module *Duality & Area of Feasible Solutions*, three pure component spectra are extracted from the estimated spectral profiles S. These spectra are arranged into a matrix $S \in \mathbb{R}^{3\times n}$, where each row corresponds to one spectral component that follows the bilinear model. Each spectral component is fitted with a variable number of Gaussian functions to obtain physically meaningful spectral shapes. The number of Gaussian peaks for each component is chosen according to the complexity of the spectrum. After optimizing the Gaussian parameters, the fitted spectra are assembled again into the matrix S. The corresponding concentration profiles, denote by $\hat{C}$, are then

computed using the Moore-Penrose pseudoinverse of S. The reconstructed dataset is obtained as $\hat{D} = \hat{C}S$.

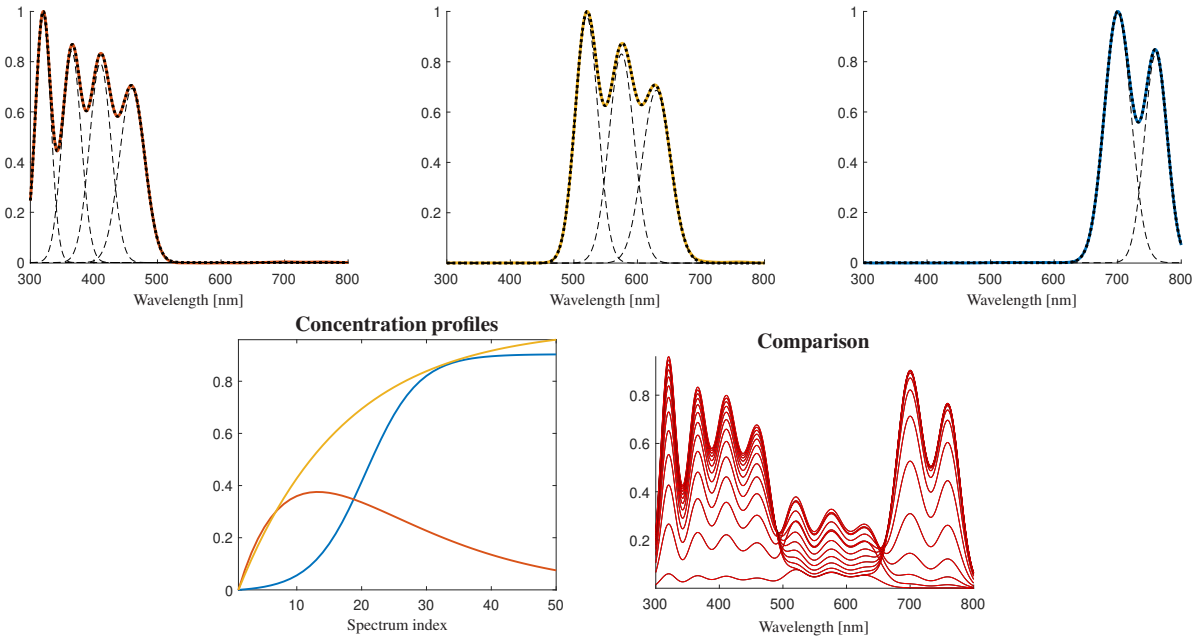
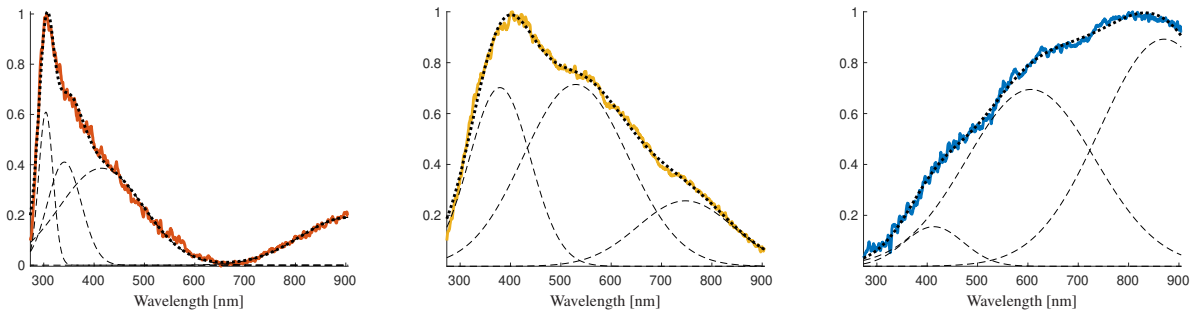


Figure 16: Gaussian parametrization of the three pure component spectra extracted from the model dataset using the FACPACK module (Duality & Area of Feasible Solutions). Each recovered pure spectrum is approximated by a sum of Gaussian basis functions in order to obtain physically interpretable and smooth spectral profiles. The first, second, and third components are modeled using $z = 4$, $z = 3$ and $z = 2$ Gaussian functions, respectively. Dashed curves represent the individual Gaussian contributions, while the solid colored curves denote the resulting composite fitted spectra. The bottom-left panel shows the estimated concentration profiles $\hat{C}$ obtained via the Moore-Penrose pseudoinverse of the fitted spectral matrix S. The bottom-right panel compares the reconstructed dataset $\hat{D} = \hat{C}S$ (red) with the original spectra (black), demonstrating excellent agreement. The total relative Frobenius reconstruction error is 4.31×10^{-7}, confirming that the Gaussian-based spectral representation provides an almost exact approximation of the model data. For clarity of visualization, only a subset of the reconstructed spectra is displayed.



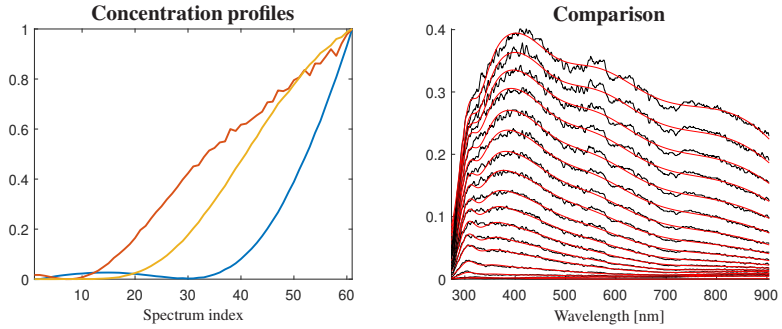


Figure 17: Gaussian parametrization of the three pure component spectra extracted from the experimental dataset using FACPACK module (Duality & Area of Feasible Solutions). Each recovered pure spectrum is approximated by a sum of Gaussian basis functions in order to obtain physically interpretable and smooth spectral profiles. The first, second, and third components are modeled using $z = 4$, $z = 3$ and $z = 3$ Gaussian functions, respectively. Dashed curves represent the individual Gaussian contributions, while the solid colored curves denote the resulting composite fitted spectra. The bottom-left panel shows the estimated concentration profiles $\hat{C}$ obtained via the Moore-Penrose pseudoinverse of the fitted spectral matrix S. The bottom-right panel compares the reconstructed dataset $\hat{D} = \hat{C}S$ (red) with the original spectra (black), demonstrating good agreement. The total relative Frobenius reconstruction error is 7.84×10^{-4}, confirming that the Gaussian-based spectral representation provides an accurate approximation of the experimental data. For clarity of visualization, only a subset of the reconstructed spectra is displayed.

The Gaussian parametrization of the three extracted pure component spectra indicates the validity and interpretability of the model. The very close agreement between the reconstructed dataset and the original spectra shows that the model describes the structure of the data very well. This implies that the spectral matrix S represents nearly all meaningful variance of the dataset without significant loss of information.

Moreover, the fact that each recovered pure spectrum can be approximated by a limited number of Gaussian basis functions indicates that the extracted components are meaningful rather than just mathematical results of the decomposition procedure.

5. Conclusion

In this work, we introduce a hybrid spectral decomposition approach that combines the modeling of Multivariate Curve Resolution (MCR) with the physical interpretability of Gaussian peak modeling. In this approach, each pure component spectrum is represented as a sum of Gaussian functions whose parameters are optimized using nonlinear least-squares fitting. Furthermore, k-means clustering is applied to group Gaussian components with similar behavior, providing clearer interpretability of the obtained profiles.

The method is applied to both model and experimental datasets. For the model dataset, where the components are known, the method achieved reconstruction error close to zero, showing that it can recover overlapping spectral features. For the experimental dataset of DR UV-Vis spectra of carbon-containing species formed on the surface of YZrOx at 550°C during the non-oxidative dehydrogenation of isobutane, the method successfully resolved overlapping absorption bands. These results indicate that the method improves both reconstruction quality and the interpretability of the extracted spectral and concentration profiles.

FACPACK-based AFS analysis is used to study the rotational ambiguity in MCR. By linking feasible mathematical solutions to physically interpretable Gaussian models, the method provides a clear connection between mathematical analysis and experimental chemistry.

These results demonstrate that the approach is capable of resolving strongly overlapping DR UV-Vis spectra. This capability opens the possibility of using the method to support kinetic and mechanistic analysis of the formation of surface carbon-containing species in different heterogeneous reactions. In particular, it may enable spectral components to be linked to physically meaningful surface species and their temporal evolution under reaction conditions to be analyzed.

References

[1] A. de Juan, J. Jaumot, and R. Tauler. Multivariate Curve Resolution (MCR). solving the mixture analysis problem. *Anal. Methods*, 6:4964–4976, 2014.

[2] M. Maeder and Y. M. Neuhold. *Practical data analysis in chemistry*. Elsevier, 2007.
[3] W. H. Lawton and E. A. Sylvestre. Self modeling curve resolution. *Technometrics*, 13(3):617–633, 1971.
[4] O. Borgen and B. R. Kowalski. An extension of the multivariate component-resolution method to three components. *Anal. Chim. Acta.*, 174:1–26, 1985.
[5] A. Golshan, H. Abdollahi, and M. Maeder. Resolution of rotational ambiguity for three-component systems. *Anal. Chem.*, 83(3):836–841, 2011.
[6] M. Sawall, C. Kubis, D. Selent, A. Börner, and K. Neymeyr. A fast polygon inflation algorithm to compute the area of feasible solutions for three-component systems I: Concepts and applications. *J. Chemom.*, 27:106–116, 2013.
[7] M. Sawall and K. Neymeyr. A fast polygon inflation algorithm to compute the area of feasible solutions for three-component systems II: Theoretical foundation, inverse polygon inflation and facpack implementation. *J. Chemom.*, 28:633–644, 2014.
[8] M. Sawall and K. Neymeyr. *How to compute the Area of Feasible Solutions, A practical case study and users' guide to FAC-PACK*, volume in Current Applications of Chemometrics, ed. by M. Khanmohammadi, chapter 6, pages 97–134. Nova Science Publishers, New York, 2014.
[9] W. F. Maddams. The scope and limitations of curve fitting. *Appl. Spectrosc.*, 34(3):245–267, 1980.
[10] V. B. Di Marco and G. G. Bombi. Mathematical functions for the representation of chromatographic peaks. *J. Chromatogr. A.*, 931(1):1–30, 2001.
[11] B. G. M. Vandeginste and L. De Galan. Critical evaluation of curve fitting in infrared spectrometry. *Anal. Chem.*, 47:2124–2132, 1975.
[12] X. Q. Zhang, J. B. Zheng, and H. Gao. Comparison of wavelet transform and Fourier self-deconvolution (FSD) and wavelet FSD for curve fitting. *Analyst*, 125:915–919, 2000.
[13] L. Antonov and D. Nedeltcheva. Resolution of overlapping UV–Vis absorption bands and quantitative analysis. *Chem. Soc. Rev.*, 29(3):217–227, 2000.
[14] F. M. Avcu and M. Karakaplan. Finding exact number of peaks in broadband UV–Vis spectra using curve fitting method based on evolutionary computing. *Journal of the Turkish Chemical Society A: Chemistry*, 7(1).
[15] S. Lloyd. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129–137, 1982.
[16] A. K. Jain. Data clustering: 50 years beyond k-means. *Pattern Recognition Letters*, 31(8):651–666, 2010.
[17] T. P. Otroshchenko, V. A. Kondratenko, U. Rodemerck, D. Linke, and E. V. Kondratenko. Non-oxidative dehydrogenation of propane, n-butane, and isobutane over bulk ZrO_2-based catalysts: effect of dopant on the active site and pathways of product formation. *Catal. Sci. Technol.*, 7:4499–4510, 2017.
[18] M. L. Bols, J. Ma, F. Rammal, D. Plessers, X. Wu, S. Navarro-Jaén, A. J. Heyer, B. F. Sels, E. I. Solomon, and R. A. Schoonheydt. In situ uv–vis–nir absorption spectroscopy and catalysis. *Chem. Rev.*, 124(5):2352–2418, 2024.
[19] S. Han, T. Otroshchenko, D. Zhao, H. Lund, N. Rockstroh, T. H. Vuong, J. Rabeah, U. Rodemerck, D. Linke, M. Gao, G. Jiang, and E. V. Kondratenko. The effect of ZrO_2 crystallinity in $CrZrO_x/SiO_2$ on non-oxidative propane dehydrogenation. *Appl. Catal. A: Gen.*, 590:117350, 2020.
[20] T. Zhang, T. P. Otroshchenko, V. A. Kondratenko, S. Bartling, and E. V. Kondratenko. Regulation of lattice oxygen reactivity of ZrO_2 to promote efficient chemical looping oxidative dehydrogenation of ethane. *Nat. Commun.*, 16:8556, 2025.
[21] K. Neymeyr and M. Sawall. On the set of solutions of the nonnegative matrix factorization problem. *SIAM J. Matrix Anal. Appl.*, 39:1049–1069, 2018.