Conference on Modelling Fluid Flow (CMFF'25)
The 19th International Conference on Fluid Flow Technologies
Budapest, Hungary, August 26-August 29, 2025

An Open Workflow for Unsupervised Clustering of Fluid-Particle Flows into Compartments

Michael Mitterlindner[2], Daniel Berger[3], Maximilian Graber[3], Regina Kratzer[4], Markus Reichhartinger[3], Stefan Radl[1]

[1] Corresponding Author. Institute of Process and Particle Engineering, Graz University of Technology. Inffeldgasse 13/III, Graz, Austria. Tel.: +43 316 873 - 30412, E-mail: radl@tugraz.at
[2] Institute of Process and Particle Engineering, Graz University of Technology. Inffeldgasse 13/III, Graz, Austria. E-mail: mitterlindner@tugraz.at
[3] Institute of Automation and Control, Graz University of Technology. Inffeldgasse 21/B/I, Graz, Austria. E-mail: markus.reichhartinger@tugraz.at
[4] Institute of Biotechnology and Biochemical Engineering, Graz University of Technology. Petersgasse 10-12/I, Graz, Austria. E-mail: regina.kratzer@tugraz.at

ABSTRACT

We present a workflow for the unsupervised clustering of fluid–particle flows into distinct compartments based on open-source tools. Our aim is to enhance the understanding of complex multiphase flow systems - prevalent in many chemical and biochemical reactors - by using models consisting of such compartments. The proposed methodology integrates computational fluid dynamics (CFD) simulations with unsupervised machine learning (i.e., clustering algorithms) to identify coherent flow regions (i.e., compartments) without prior labelling. Our workflow is fully automated and designed for reproducibility, with a modular structure that allows for an easy adaptation to various flow systems. The results demonstrate that our approach can successfully capture essential flow features and partition the domain into meaningful compartments. This facilitates direct use in compartment models, which reduces computational costs in larger-scale simulations. Our findings suggest that unsupervised machine learning algorithms are mature enough to simplify complex multiphase systems in a largely automated fashion, making them a valuable tool for both academic research and industrial applications.

Keywords: CFD, multiphase, compartment modelling, bioreactor, clustering, unsupervised learning

NOMENCLATURE

C	$[-]$	cluster
Da	$[-]$	Damköhler number
H	$[m]$	channel height
J_C	$[-]$	clustering objective for the k-means algorithm
$J_{\dot{V}}$	$[-]$	optimization objective flow rate adaption
L	$[m]$	channel length
S	$[\frac{kmol}{m^3 s}]$	additional sources or sinks
V_R	$[m^3]$	volume of compartment
c	$[kmol/m^3]$	concentration of species
c^*	$[-]$	dimensionless concentration of species
k	$[s^{-1}]$	reaction rate constant
$k_L a$	$[s^{-1}]$	volumetric mass transfer coefficient
n_c	$[-]$	number of compartments
n_k	$[-]$	number of clusters
n_x	$[-]$	number of datapoints
$n_{cells,min}$	$[-]$	minimum number of CFD cells required per cluster
r	$[\frac{kmol}{m^3 s}]$	reaction rate
v_1	$[m/s]$	velocity of the moving top wall
$v_{\max}$	$[m/s]$	maximum velocity of the flow profile
$\dot{V}$	$[m^3/s]$	flow rate
$\underline{\underline{\dot{V}}}$	$[m^3/s]$	flow rate matrix
$\underline{\underline{\dot{V}}}'$	$[m^3/s]$	optimized flow rate matrix
$\underline{v}$	$[m/s]$	velocity vector with the components v_1 and v_2
$\underline{x}$	$[-]$	feature vector from data matrix
α	$[-]$	volume fraction
λ	$[-]$	regularization parameter
μ	$[-]$	cluster average
v_{thresh}	$[-]$	threshold for reassignment condition based on volume
τ_R	$[s]$	residence time
$\underline{\xi}_1^*$	$[-]$	dimensionless axial coordinate
$\underline{\xi}_2^*$	$[-]$	dimensionless transverse coordinate

C	[−]	set of clusters
C_i	[−]	cluster subset of $\mathcal{X}$
$\mathcal{X}$	[−]	data set

Subscripts and Superscripts

in	at the inlet
out	at the outlet
φ	for each phase

1. INTRODUCTION

Accurate prediction of large-scale multi-species gas-liquid bioreactor performance is essential for advancing biotechnological applications. For example, *Cupriavidus necator* bacteria within multiphase bioreactors may produce valuable products such as the polymer Polyhydroxybutyrate (PHB) [1], or food proteins. These systems often involve two-phase flows, in which the dissolution of key gases like O_2, CO_2, and H_2 significantly impacts reactor performance. However, the complexity of these systems requires advanced modelling techniques to capture the interplay between flow dynamics and chemical reactions.

Despite progress in computational fluid dynamics (CFD) and compartment modelling, several gaps in our scientific understanding of such reactors remain. Unsupervised machine learning algorithms can assist in finding patterns in the large datasets generated by CFD simulations. For instance, Laborda et al. [2] used unsupervised clustering on CFD and experimental data to identify flow regimes in bioreactors. While their custom k-means algorithm ensured spatial continuity and optimal compartment count, the study focused solely on regime identification without integrating chemical reaction networks (CRNs), or predicting reactor performance. Le Nepvou de Carfort et al. [3] proposed a fully automated CFD-based method for generating 3D compartment models using structured grids. Although their approach enabled real-time bioreactor simulations, it was constrained to Cartesian grids and lacked validation for non-ideal geometries. Savage et al. [4] combined CFD, high-dimensional design spaces, and Bayesian optimisation to improve reactor geometries with respect to flow behaviour and mixing characteristics. However, their work focused on the reactor design and did not address (bio-)chemical reaction kinetics, or propose a method to create to compartment models. Savarese et al. [5] developed a CRN generation method using unsupervised clustering and graph algorithms for combustion in CFD simulations. While effective for predicting NO_x emissions, their approach was limited to single-phase systems and consequently an application to multiphase bioreactors or biochemical reaction networks cannot be realized. Other studies, such as Tajsoleiman et al. [6] and Delafosse et al. [7], introduced automated CFD-based zoning and compartment models for stirred bioreactors. However, these methods were either geometry-specific, lacked spatial continuity enforcement, or did not integrate biochemical reactions or a fully automated zoning algorithm.

1.1. Goals of our Contribution

Given the limitations of CFD simulations in modelling detailed chemical reactions, our present study aims to identify meaningful compartments, and subsequently solve the equations that govern the resulting compartment model (CM). By leveraging analytical solutions, we validate the CM for flows between two infinite plates, including single-phase flow with reactions and two-phase flow with mass transfer.

Therefore, we utilise the codebase from Savarese et al. [5] and developing it further into our own tool box called CLARA (CLustering AlgoRithm Austria). To perform the clustering and compartment generation, simulation data is needed. We added two new simulation tool options from which the data can be read by CLARA. These tools are (i) OpenFOAM®, a classic open-source CFD tool, and (ii) SimVantage®, a commercial bioreactor CFD simulation tool developed by the SimVantage GmbH. The SimVantage® software uses the Lattice Boltzmann method (LBM) for simulating complex flows, including multi-phase and multi-component systems. With the integration of these tools, we have two powerful options for simulating complex flows in bioreactors, allowing us to accurately capture the dynamics of gas-liquid interactions therein. Our clustering and compartmentalization tool, CLARA, will soon be available as an open-source tool.

Our long-term vision is to develop a fast and reliable tool for optimising bioreactor reactor design, akin to the approach presented in [4], which showed that it is crucial to scan the design space and find the "best" geometrical design of a device. The tool will be useful for the process optimization of the gas-fermentation bioreactor as shown in [1]. Due to regulatory restrictions and safety issues, which arise due to the highly explosive medium (hydrogen and oxigen), it is vital to have an excellent understanding of transport phenomena in bioreactors to avoid explosive gas mixtures. A secondary vision is to develop a physics-informed model capable to be used for the design of model-based automatic control algorithms, and which can be exploited to optimise reactor performance during operation. This is particularly important for bioreactors, as they are complex systems with many variables that need to be controlled.

2. METHODS

In this section, we describe the methods used to achieve the goals outlined in the previous chapter. We begin by discussing the integration of OpenFOAM® and SimVantage®, followed by a detailed explanation of the clustering process, including feature selection and graph reassignment. We then present the compartment model equations and the optimisation process for inconsistent flow rates.

Specifically, the overall clustering process involves reading in field values, calculating volume flow rates over faces, and selecting relevant features for the core clustering step. Once the clustering is performed, graph reassignment ensures spatial continuity of the clusters. After the clustering is completed, we will call these clusters compartments as each represents a theoretical compartment in a reactor. The next step is to calculate the compartment attributes such as volume, flow rate matrix, temperature and, for multiphase systems, the volumetric mass transfer coefficient ($k_L a$). If mass balance inconsistencies are detected, an optimisation step is applied to correct them. Finally, boundary conditions for the compartment model (CM) are selected, including flow rates and concentrations, which are then passed to the CM solver. In order to ensure the quality of the code, analytical solutions are used to verify the results. Specifically, we test the performance of the CM with simple chemical reactions and mass transfer scenarios.

2.1. Integration of OpenFOAM® and SimVantage®

The integration of OpenFOAM® and SimVantage® with CLARA is achieved through a custom Python interface. This interface allows for seamless communication between the two software packages (i.e., OpenFOAM® or SimVantage® and CLARA), enabling the transfer of data and results. For the LBM method (which is used in SimVantage®), face fluxes must be calculated as a post processing step to obtain the volumetric flow rates as these are not used in a classical LBM-based simulation. This is especially critical for moving boundaries, such as the stirrer, where some cells are partly solid and partly fluid.

2.2. Clustering

To perform clustering, we must carefully select the features that represent the underlying physics of the system. These features are derived from CFD simulations and typically include flow-relevant quantities such as velocity, residence time, or turbulence intensity. It is recommended not to use concentration as a feature, since we simulate only until the flow reaches a quasi-steady state, long before the concentration field has fully evolved. This is particularly relevant in bioreactors, where bio-reactions involving bacteria often take several days to produce measurable outputs due to the inherently slow biological processes. Simulating such long timescales with CFD alone would be computationally prohibitive. Instead, we simulate the flow dynamics until a statistically steady state is reached and then use the compartment model (CM) to predict long-term concentration behaviour. This approach maintains accuracy while reducing computational cost. To enhance the clustering, it is also useful to include statistical measures such as the mean and standard deviation of time-dependent features, allowing the method to capture the unsteady or turbulent characteristics of the flow.

The core clustering step uses the k-means algorithm, which partitions the dataset into n_k distinct, non-overlapping clusters. The dataset consists of n_x feature vectors $\mathcal{X} = \{\underline{x}_1, \ldots, \underline{x}_{n_x}\}$. Each cluster $C_i \subseteq \mathcal{X}$, with $i = 1, \ldots, n_k$, is a subset of the data. The algorithm assigns each feature vector $\underline{x}_j$ to one of the clusters via the mapping $\underline{x}_j \mapsto C_i$. Each cluster has a centroid $\underline{\mu}_i$, defined as the mean of all feature vectors in that cluster $\underline{\mu}_i = \frac{1}{|C_i|} \sum_{\underline{x}_j \in C_i} \underline{x}_j$. where $|C_i|$ denotes the number of elements in C_i.

Each vector is then assigned to the cluster with the nearest centroid (in terms of Euclidean distance). The centroids are recalculated iteratively until the assignments no longer change, indicating convergence. The k-means algorithm minimizes the following objective function:

$$J_C = \sum_{i=1}^{n_k} \sum_{\underline{x}_j \in C_i} \|\underline{x}_j - \underline{\mu}_i\|^2. \tag{1}$$

To ensure cluster connectivity, a graph reassignment algorithm is applied. This algorithm checks each cluster for disconnected components and reassigns them based on a connectivity criteria (e.g. spatially next cluster). Additionally, thresholds can be defined to trigger reassignment. For example, a compartment must contain a minimum number of CFD cells (see Eq. 2) or a minimum volume (see Eq. 3). The reassignment conditions are:

$$\texttt{reassign}(C_i) \iff |C_i| < n_{cells,min} \tag{2}$$

$$\texttt{reassign}(C_i) \iff \sum_{\underline{x}_j \in C_i} V(\underline{x}_j) < \nu_{\text{thresh}} \cdot V_{\text{total}} \tag{3}$$

After clustering is completed, each cluster is converted into a compartment, as compartments can hold additional attributes. The number of clusters n_k is therefore equal to the number of compartments n_c.

2.2.1. Mass Conservation and Flow Balance

For a network of compartments, the flow rate matrix $\underline{\underline{\dot{V}}}$ represents the flow rates between the compartments via a two-dimensional matrix. These flow rates are derived by summing the volume flows extracted from the simulation data. Let $\underline{\dot{V}}_{\text{in}}$ and $\underline{\dot{V}}_{\text{out}}$ represent the external inflows and outflows, respectively, for each compartment. The system must satisfy mass conservation, which is expressed mathematically in Eq. 4:

$$\sum_{j=1}^{n_c} \dot{V}_{ij} - \sum_{j=1}^{n_c} \dot{V}_{ji} + \dot{V}_{\text{in}_i} - \dot{V}_{\text{out}_i} = 0 \quad \forall i \tag{4}$$

In this equation, $\dot{V}_{ij}$ represents the flow rate from compartment i to compartment j, while $\dot{V}_{ji}$ denotes

the flow rate from compartment j to compartment i. The terms $\dot{V}_{\text{in}_i}$ and $\dot{V}_{\text{out}_i}$ correspond to the external inflow and outflow rates for compartment i, respectively. Finally, n_c is the total number of clusters in the system. This equation ensures that the net flow into each compartment, considering both internal and external flows, is zero. It is important to note that for two-phase systems, achieving mass conservation can be challenging, as the mass transfer between the phases must be considered.

2.2.2. Optimisation of Flow Rates

To address inconsistencies in flow rates, an optimisation process is applied. The objective is to adjust the flow rate matrix $\underline{\underline{\dot{V}}}$ to minimise deviations from the initial values while ensuring mass conservation. The optimisation problem is defined by the objective function in Eq. 5:

$$J_{\dot{V}} = \sum_{i,j}^{n_c} (\dot{V}'_{ij} - \dot{V}_{ij})^2 + \lambda \sum_{i,j}^{n_c} \dot{V}'^2_{ij} \tag{5}$$

In this equation, $\dot{V}'_{ij}$ represents the optimised flow rates, which are adjusted to satisfy mass conservation while minimising deviations from the initial flow rates $\dot{V}_{ij}$. The parameter λ is a regularization term that penalizes large flow rates, thereby preventing excessive deviations and ensuring a more stable solution. A rule of thumb for estimating this regularization term is $\lambda \approx \frac{1}{||\underline{\underline{\dot{V}}}||^2}$. This approach ensures that the optimisation process maintains the integrity of the flow distribution across the clusters.

The optimisation is subject to the mass balance constraint, as shown in Eq. 6:

$$\sum_{j=1}^{n_c} \dot{V}'_{ij} - \sum_{j=1}^{n_c} \dot{V}'_{ji} + \dot{V}_{\text{in}_i} - \dot{V}_{\text{out}_i} = 0 \quad \forall i \tag{6}$$

This constraint ensures that the optimised flow rates satisfy mass conservation for each cluster. The terms in Eq. 6 are defined similarly to those in Eq. 4.

The optimisation problem is solved using Sequential Least Squares Quadratic Programming (SLSQP), an iterative method suitable for constrained optimisation. The solution is bounded to ensure positive flow rates. Additionally, flow rates which are under a certain threshold will be kept/set to zero, so that it is ensured that spatially not connected compartments remain not connected. This approach ensures minimal adjustments to the flow rates while maintaining the integrity of the flow distribution across the clusters.

2.3. Compartment attributes

The next step is to calculate the compartment attributes like volume, temperature, and also for a multiphase system the $k_L a$ value. For a two phase system the volume of the phase is used as the compartment volume ($\alpha_\varphi \cdot V$).

2.4. Compartment Model Equations

In the context of a bioreactor, k-Means can be used to cluster different regions based on certain characteristics, such as concentration or reaction rates, to form different compartments which can then be solved. This is very similar to a Chemical Reactor Network (CRN), or as we call it, the compartment model.

The compartment model (CM) simplifies a complex system by dividing it into compartments, each representing a distinct region or phase. The CM equations describe the mass balance of species within each compartment, accounting for inflows, outflows, reactions, and other processes. The governing equation for the liquid phase is given in Eq. 7:

$$\begin{aligned}\frac{dc_{\varphi,i,j}}{dt} =& \frac{\sum_k^{n_c} \dot{V}_{\varphi,i,k\to j} c_{\varphi,i,k} - \sum_k^{n_c} \dot{V}_{\varphi,i,j\to k} c_{\varphi,i,j}}{V_{R,\varphi,j}} \\ &+ \frac{\dot{V}_{\text{in}_{\varphi,i,j}} c_{\text{in}_{\varphi,i,j}} - \dot{V}_{\text{out}_{\varphi,i,j}} c_{\text{out}_{\varphi,i,j}}}{V_{R,\varphi,j}} + r_{\varphi,i,j} + S_{\varphi,i,j}\end{aligned} \tag{7}$$

In this equation, $c_{\varphi,i,j}$ represents the concentration of species i in compartment j. The terms $\dot{V}_{\varphi,i,k\to j}$ and $\dot{V}_{\varphi,i,j\to k}$ denote the flow rates of species i into and out of compartment j, respectively. These flow rates account for the transport of species between compartments. The external inflow and outflow rates of species i for compartment j are represented by $\dot{V}_{\text{in}_{\varphi,i,j}}$ and $\dot{V}_{\text{out}_{\varphi,i,j}}$, respectively. The volume of compartment j is denoted by $V_{R,\varphi,j}$, which is used to normalize the mass balance equation. The reaction rate of species i in compartment j is represented by $r_{\varphi,i,j}$, which accounts for the chemical reactions occurring within the compartment. Finally, S_φ represents additional sources or sinks, such as mass transfer from other phases. This term ensures that all external contributions to the species concentration are included in the model.

The CM equations are solved numerically using standard ODE solvers provided by the SciPy Python library. This approach allows for an accurate approximation of the system's dynamic behavior and supports efficient integration of the compartment model under various flow and reaction conditions.

2.5. Verification

To test the implementation and results of the compartment model, we verify it against an analytical solution. This analytical solution represents flow between two infinite plates with a simple first-order chemical reaction. The verification process ensures that the CM accurately predicts the concentration profiles and flow dynamics under controlled conditions.

The analytical solutions for plug flow, Couette flow, and Poiseuille flow provide benchmarks for evaluating the CM's performance. These solutions are derived based on non diffusive steady-state assumptions and specific velocity profiles. By com-

paring the CM results to these analytical solutions, we can assess the accuracy and reliability of the model. Additionally, the verification process helps to identify potential limitations or areas for improvement in the CM implementation.

2.5.1. Analytical Solution for Plug Flow

For a first-order chemical reaction in a plug flow between two infinite plates, the concentration profile can be derived under steady-state conditions. The velocity profile is uniform ($v_1 = v_{\max}, v_2 = 0$), and the concentration decreases exponentially along the flow direction. Specifically, the resulting concentration profile is given by:

$$c^*(\xi_1^*) = e^{-Da\xi_1^*} \tag{8}$$

where $Da = \frac{kL}{v_{\max}}$ is the Damköhler number, and $\xi_1^* = \frac{\xi_1}{L}$ is a dimensionless axial coordinate.

2.5.2. Analytical Solution for Couette Flow

In Couette flow, the velocity profile is linear and given by $v_1 = v_{\max}\frac{\xi_2}{L}, v_2 = 0$, where $v_{\max}$ is the velocity of the moving top wall, ξ_2 is the dimensionless transverse coordinate, and L is the characteristic length. The mean residence time can be approximated as $\tau_R = \frac{L}{\bar{v}}$, and for a linear velocity profile we get $\bar{v} = \frac{v_{\max}}{2}$. Substituting this, the mean residence time becomes $\tau_R = \frac{L}{2v_{\max}}$. Therefore, the Damköhler number is calculated as $Da = \frac{2kL}{v_{\max}}$. The concentration profile for a first-order reaction is expressed as:

$$c^*(\xi_1^*, \xi_2^*) = e^{-Da\frac{\xi_1^*}{\xi_2^*}} \tag{9}$$

where $\xi_1^* = \xi_1/L$ and $\xi_2^* = \xi_2/H$ are the dimensionless axial and transverse coordinate, respectively.

2.5.3. Analytical Solution for Poiseuille Flow

In a Poiseuille flow, the velocity profile is parabolic and given by $v_1 = v_{\max}\left(1 - \left(\frac{\xi_2}{H/2}\right)^2\right), v_2 = 0$, where $v_{\max}$ is the maximum velocity at the centerline, ξ_2 is the dimensionless transverse coordinate, and H is the channel height. The mean velocity for this profile is $v_1 = 2/3 v_{\max}$. Using this, the Damköhler number is evaluated as $Da = k\frac{3L}{2v_{\max}}$. The concentration profile for a first-order reaction is given by:

$$c^*(\xi_1^*, \xi_2^*) = e^{-Da\frac{\xi_1^*}{1-\xi_2^{*2}}} \tag{10}$$

where ξ_1^* and ξ_2^* are the dimensionless axial and transverse coordinate, respectively.

3. RESULTS

The results of this study demonstrate the capability of the compartment model (CM) to accurately reproduce the reaction dynamics of a simple tubular reactor. The Fig. 1 shows the concentration profile of a reactor with Poiseuille flow and a Darmköhler number of 1.

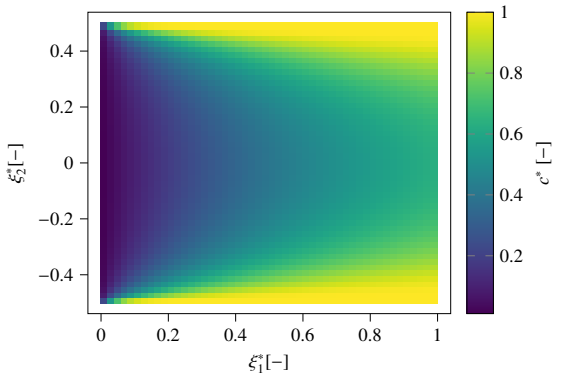

Figure 1. Analytical concentration profile for Poiseuille flow for a simple first-order chemical reaction.

3.1. Importance of Feature Selection

Feature selection plays a crucial role in the clustering process, as it directly determines how the compartments are formed and how well they capture the relevant dynamics of the system. To illustrate this, we compare two clustering results, each using six compartments: one based on concentration as the feature, see Fig. 2, and the other based on velocity, see Fig. 3. The clustering based on concentration yields a significantly lower error of 16% compared to 29% when using velocity, indicating that it provides a better representation of the system in this case. However, while the performance is better, concentration is very difficult to obtain from the simulation. In practical CFD setups, especially those limited to the flow field, the full concentration profile may not be available or may require prohibitively long simulation times to resolve accurately. This highlights a key trade-off: the most informative features may not be accessible in realistic scenarios. Moreover, in this

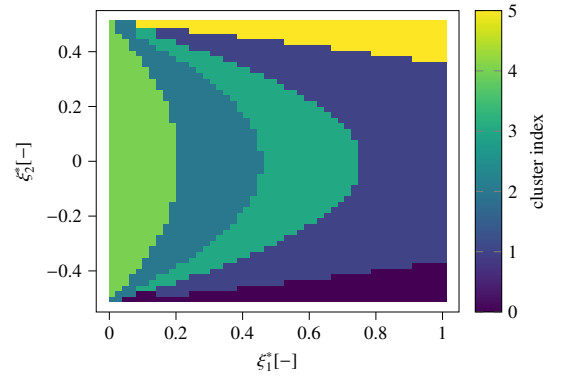

Figure 2. Clustering solution with six clusters and the concentration as feature vector.

simplified analytical test case, only a limited number of features are available, which further complicates the selection process. The lack of diverse or rich fea-

ture data restricts the clustering performance and the ability to tailor compartments to specific physical behaviours. Therefore, while the example shows that using concentration leads to better clustering quality, it also underlines the challenges in feature availability and extraction, particularly when applying this method to more complex or computationally expensive systems.

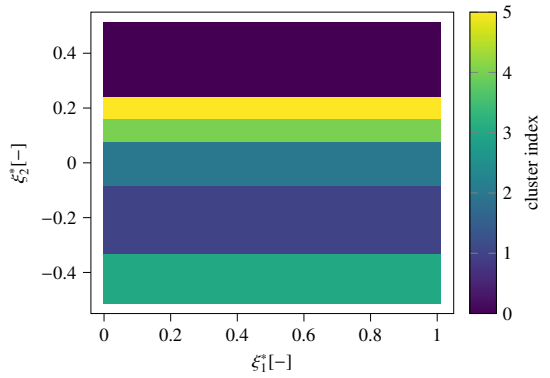

Figure 3. Clustering solution with six clusters and the velocity as feature vector.

3.2. Compartment Model Performance

Fig. 4 illustrates the absolute error of the predicted total mole flow compared to the analytical solutions for three different flow profiles: plug flow, Couette flow, and Poiseuille flow. These profiles serve as benchmarks for evaluating the CM's performance. For this comparison, the feature vectors used in the clustering included both the velocity and the concentration field.

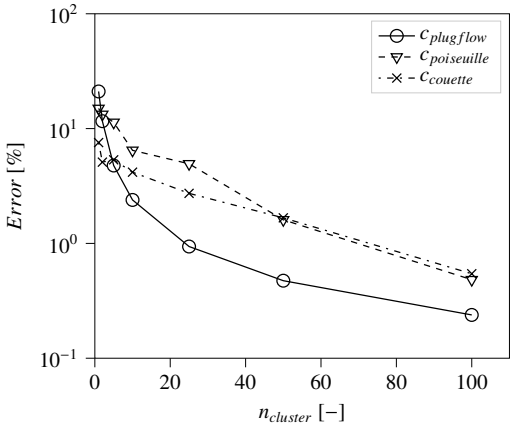

Figure 4. Absolute error of the CM solution compared to the analytical solution for plug flow, Couette flow, and Poiseuille flow.

The results show that as the number of compartments increases, the CM solution converges to the analytical solution, i.e., the error vanishes. This convergence highlights the ability of the CM to capture the essential flow and reaction characteristics of the system. For all three flow profiles, the error decreases significantly with an increasing number of compartments, demonstrating the robustness of the CM approach.

The plug flow results exhibit the lowest error, as the uniform velocity profile simplifies the flow dynamics. In contrast, the Couette and Poiseuille flows, which involve more complex velocity profiles, show slightly higher errors. These results were obtained without any advanced feature selection for clustering, meaning the compartments were generated based solely on basic flow properties. Despite this, the CM performed well, suggesting that even a straightforward clustering approach can yield meaningful results. Future work could explore the impact of incorporating additional features, such as turbulence intensity or residence time, to further enhance the accuracy and efficiency of the CM.

4. CONCLUSION

The results presented in our paper demonstrate the effectiveness of the compartment model (CM). We established a seamless, automated workflow by integrating the OpenFOAM® and SimVantage® flow solvers with our in-house tool called CLARA (CLustering AlgoRithm Austria). These CFD tools provide flow field data, which CLARA uses for unsupervised clustering via k-means and graph-based reassignment to identify spatially coherent compartments that mechanistically represent the reactor's flow characteristics. Based on these compartments, CLARA calculates compartment volumes, in- and outflows, and inter-compartment flow rates to generate the CM. An optimization step ensures mass conservation. This enables an efficient and mechanistic prediction of concentration fields and reactor behaviour using reduced-order models.

The verification of the CM against analytical solutions for plug flow, Couette flow, and Poiseuille flow confirmed its accuracy and reliability. The corresponding CMs successfully captured the concentration profiles and flow dynamics under various conditions, with errors below 1% for all flow profiles. These results highlight the potential of the CM as a fast and reliable tool for optimising reactor design and performance.

Future work will focus on extending the CM to handle more complex two phase system and complex chemical reactions. Additionally, efforts will be made to improve the feature selection with intelligent systems, enabling its application to larger-scale systems. The development of a physical-informed model for reactor control will also be explored, with the goal of optimising reactor performance under dynamic operating conditions.

ACKNOWLEDGEMENTS

This work is funded by Graz University of Technology through the lead project "DigiBioTech".

REFERENCES

[1] Lambauer, V., Permann, A., Petrášek, Z., Subotić, V., Hochenauer, C., Kratzer, R., and Reichhartinger, M., 2023, "Automatic Control of Chemolithotrophic Cultivation of Cupriavidus necator: Optimization of Oxygen Supply for Enhanced Bioplastic Production", *Fermentation*, Vol. 9 (7), p. 619, URL https://www.mdpi.com/2311-5637/9/7/619.

[2] Laborda, V. P. I., Puiman, L., Groves, T., Haringa, C., and Nielsen, L. K., 2025, "Unsupervised learning bioreactor regimes", *Computers & Chemical Engineering*, Vol. 194, p. 108891, URL https://linkinghub.elsevier.com/retrieve/pii/S0098135424003090.

[3] Le Nepvou De Carfort, J., Pinto, T., and Krühne, U., 2024, "An Automatic Method for Generation of CFD-Based 3D Compartment Models: Towards Real-Time Mixing Simulations", *Bioengineering*, Vol. 11 (2), p. 169, URL https://www.mdpi.com/2306-5354/11/2/169.

[4] Savage, T., Basha, N., McDonough, J., Krassowski, J., Matar, O., and Del Rio Chanona, E. A., 2024, "Machine learning-assisted discovery of flow reactor designs", *Nature Chemical Engineering*, Vol. 1 (8), pp. 522–531, URL https://www.nature.com/articles/s44286-024-00099-1.

[5] Savarese, M., Cuoci, A., De Paepe, W., and Parente, A., 2023, "Machine learning clustering algorithms for the automatic generation of chemical reactor networks from CFD simulations", *Fuel*, Vol. 343, p. 127945, URL https://linkinghub.elsevier.com/retrieve/pii/S0016236123005586.

[6] Tajsoleiman, T., Spann, R., Bach, C., Gernaey, K. V., Huusom, J. K., and Krühne, U., 2019, "A CFD based automatic method for compartment model development", *Computers & Chemical Engineering*, Vol. 123, pp. 236–245, URL https://linkinghub.elsevier.com/retrieve/pii/S0098135418308950.

[7] Delafosse, A., Collignon, M.-L., Calvo, S., Delvigne, F., Crine, M., Thonart, P., and Toye, D., 2014, "CFD-based compartment model for description of mixing in bioreactors", *Chemical Engineering Science*, Vol. 106, pp. 76–85, URL https://linkinghub.elsevier.com/retrieve/pii/S0009250913007690.