Systems & Control Transactions

Original Research Article
Peer Reviewed Conference Proceeding
ESCAPE 36 - European Symposium on Computer Aided Process Engineering
Sheffield, UK. 21-24 June 2026

PSE PRESS

Accelerating Efficient Dimethyl Ether Synthesis through Machine Learning-Based Process Optimization

Mitra Jafari[a], Jefferson Santos da Silva[b], Wilson Sousa Mercês Neto[b], Lucas Fonseca Couto[b], Bogdan Dorneanu[a], Karen Valverde Pontes[b], Harvey Arellano-Garcia[a]*

[a] 1 Department of Process and Plant Technology, Brandenburg University of Technology Cottbus-Senftenberg, Cottbus
[b] Programa de Pós-Graduação em Energia e Ambiente (PGENAM), Escola Politécnica, Universidade Federal da Bahia, Salvador, Brazil
* Corresponding Author: arellano@b-tu.de

ABSTRACT

Dimethyl ether (DME) is a promising clean fuel and chemical intermediate, yet its synthesis from synthesis gas remains highly sensitive to both catalyst formulation and operating conditions. In this work, a data-driven framework is developed that combines machine learning surrogate modeling with multi-objective optimization to support systematic decision-making in DME synthesis. The novelty lies in the systematic comparison of different optimization approaches applied to an identical machine learning surrogate model for DME synthesis, thereby highlighting their respective strengths and limitations as decision-support tools under limited-data conditions. A dataset compiled from published literature includes catalyst composition, preparation methods, physicochemical descriptors, and operating conditions, with CO conversion and DME selectivity as performance indicators. After data preprocessing, feature analysis using correlation analysis and principal component analysis (PCA) is applied to explore dominant trends and interactions. Several supervised machine learning models are trained and benchmarked, and the best-performing model is selected as a surrogate for optimization. Two complementary multi-objective optimization strategies are then applied to the same surrogate model: a fuzzy-enhanced NSGA-II algorithm and a constrained multi-objective Bayesian optimization approach. Gradient boosting is found to provide the most reliable predictive performance among the tested models. Both optimization strategies identify similar catalyst formulations as optimal, while differences emerge in the recommended operating conditions. The comparative analysis highlights how different optimization paradigms influence compromise solutions when balancing CO conversion and DME selectivity.

Keywords: Dimethyl Ether, Machine Learning, Optimization, Modelling, Syngas

INTRODUCTION

The transition toward sustainable and flexible fuel production pathways has intensified interest in synthesis gas (CO/H_2) conversion processes [1]. Among the available options, dimethyl ether (DME) has attracted considerable attention due to its high cetane number, clean combustion behaviour, and compatibility with existing fuel infrastructure. DME can be produced either via a two-step route involving methanol synthesis followed by dehydration or through a direct, one-step process using bifunctional catalysts. Despite extensive experimental efforts, achieving high CO conversion and DME selectivity simultaneously remains challenging because of the strong coupling between catalyst preparation method, and operating conditions [2, 3].

From a process engineering perspective, DME synthesis constitutes a high-dimensional and strongly nonlinear optimization problem. Reaction temperature, pressure, feed composition, space velocity, and catalyst characteristics interact in complex ways that are difficult to capture using mechanistic models, particularly when kinetic data are scarce or inconsistent across studies. As a result, experimental trial-and-error approaches remain common, but are time-consuming and resource-intensive.

Machine learning (ML) techniques provide an alternative route by enabling the construction of surrogate

models that directly learn input-output relationships from experimental data [4]. When applied to catalytic systems, ML models can capture nonlinear trends and interactions that are not easily accessible through simplified kinetic expressions [5, 6].

Multi-objective optimization methods are particularly relevant for processes such as DME synthesis, where improvement in CO conversion may come at the expense of DME selectivity and vice versa. Evolutionary algorithms such as NSGA-II are widely used to generate Pareto fronts representing trade-offs between competing objectives [7], while fuzzy decision-making methods enable the selection of a single compromise solution from these fronts. However, these approaches typically treat surrogate model predictions as deterministic and do not explicitly account for uncertainty [8]. Bayesian optimization [9] offers a complementary paradigm by combining probabilistic surrogate models with acquisition functions that guide the search toward promising regions of the design space while quantifying uncertainty [10, 11]. Although Bayesian optimization has been applied successfully in catalyst and reaction optimization, its performance relative to evolutionary multi-objective methods in the context of DME synthesis has not been systematically assessed.

Motivated by this gap, the present work compares two surrogate-based multi-objective optimization strategies: a fuzzy-enhanced NSGA-II algorithm and a constrained multi-objective Bayesian optimization approach, applied to the same machine learning model trained on literature data for DME synthesis. By focusing on methodological comparison rather than introducing new experimental results, this study aims to clarify how different optimization frameworks influence the identification of optimal solution and compromise operating conditions under limited data scenarios.

METHODOLOGY

Data Collection and Preprocessing

The dataset used in this study is compiled from published experimental literature on DME synthesis from synthesis gas (CO and H_2 mixture). It includes operating conditions and catalyst descriptors commonly reported across studies, such as reaction temperature, pressure, H_2/CO feed ratio, space velocity, catalyst composition, and preparation-related parameters. CO conversion and DME selectivity are selected as the output variables, as they directly reflect process performance and selectivity toward the desired product.

Prior to model development, the dataset is screened for outliers to reduce the influence of inconsistent or erroneous entries. Categorical variables describing catalyst type and preparation method are encoded numerically. Prior to analysis, all variables are normalized using min-max scaling, which linearly transforms each feature to the range [0, 1] based on its minimum and maximum values. This step mitigates scale differences and prevents features with larger magnitudes from dominating the analysis. Missing values are imputed using automated imputation procedures in JMP. Model training and validation are performed using k-fold cross-validation to ensure robust performance assessment given the limited dataset size.

To explore relationships between descriptors and performance indicators, Spearman correlation analysis is applied to identify monotonic trends. In parallel, principal component analysis (PCA) is used to examine the structure of the dataset and to assess how catalyst-related and operating condition-related variables contribute to overall variance.

Machine Learning Surrogate Model

Several supervised ML algorithms, such as XGBoost, Gradient Boosting, CatBoost, Random Forest, and Support Vector Regression (SVR), are trained to predict CO conversion and DME selectivity from the input variables. Model performance is evaluated using standard regression metrics, including the coefficient of determination (R^2) and the root mean square error (RMSE). The model demonstrating the best balance between predictive accuracy and generalization capability is selected as the surrogate model for subsequent optimization.

Fuzzy-Enhanced NSGA-II Multi-Objective Optimization

Multi-objective optimization is formulated with the goal of simultaneously maximizing CO conversion and DME selectivity. The NSGA-II evolutionary algorithm [12] is employed to generate a set of non-dominated solutions representing trade-offs between these objectives. Since NSGA-II is inherently a minimization algorithm, the negative values of the predicted objectives are minimized. This algorithm is implemented using the Pymoo library in Python [13].

To select a single compromise solution from the Pareto front, a fuzzy logic-based decision-making approach is applied [14]. This method integrates relative preferences between objectives into a unified metric, enabling the identification of a solution that best balances CO conversion and DME selectivity. The fuzzy-optimal solution is subsequently compared with experimentally reported data by identifying the closest real datapoint in terms of Euclidean distance in feature space.

Multi-Objective Bayesian Optimization

A constrained multi-objective Bayesian optimization strategy based on the ParEGO framework [15, 16] is also applied. In this approach, the two objectives are combined into a scalarized objective function using predefined weights, and a Gaussian process model is trained to approximate this scalarized response. An Expected

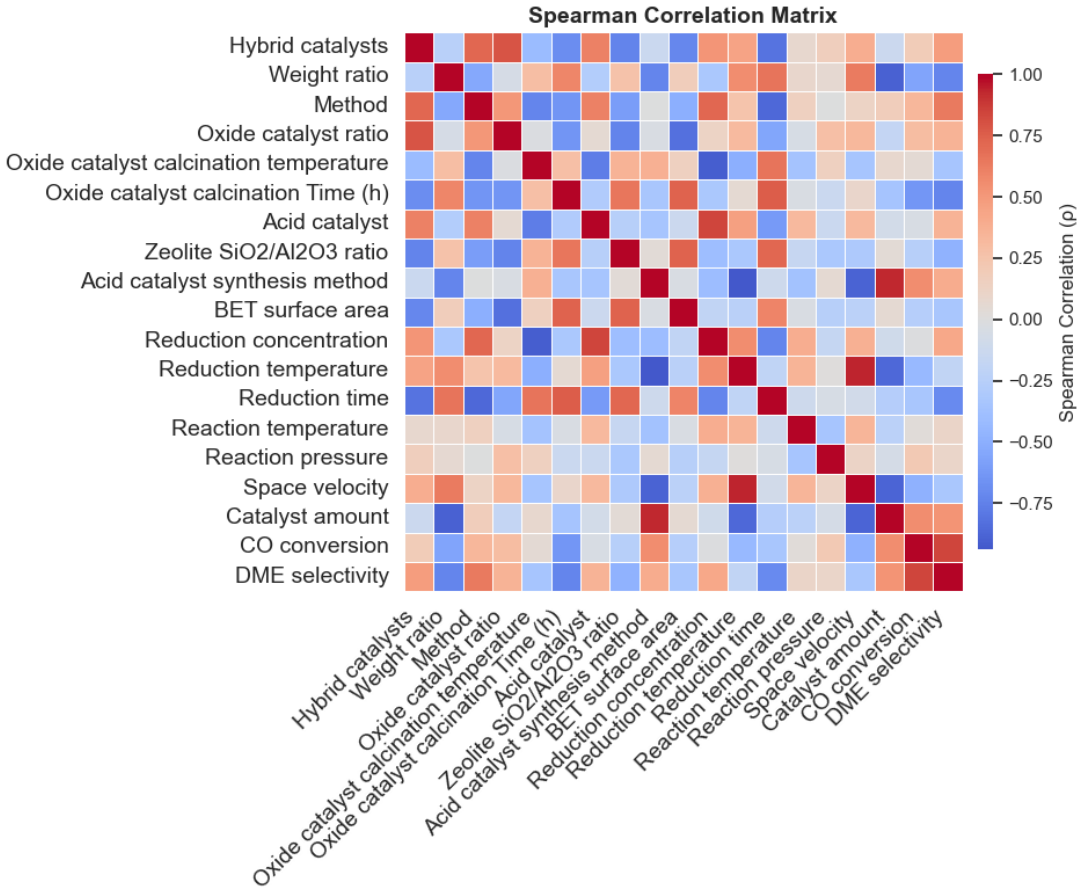

Figure 1. Spearman Correlation Analysis.

improvement acquisition function guides the selection of new candidate points while respecting process constraints [11, 17]. After completion of the optimization loop, the best compromise solution is identified based on the scalarized objective values. Presenting the optimized descriptors in their original physical units enables direct comparison with both experimental data and the fuzzy-enhanced NSGA-II results.

RESULTS

Data Preparation and Feature Analysis

The final dataset consists of 52 datapoints with 17 input features describing catalyst formulation, preparation, and operating conditions. Initial preprocessing ensures consistency across studies and prepares the data for ML analysis. After preprocessing, the data span CO conversions between 30 and 55%, and DME selectivity in the range of roughly 45-75%, reflecting the diversity of catalytic systems reported in the literature.

Spearman correlation analysis (Figure 1) provides an initial, quantitative indication of descriptor relevance. Catalyst-related, particularly synthesis method and zeolite SiO_2/Al_2O_3 ratio, show stronger monotonic correlations with DME selectivity, whereas operating variables such as space velocity exhibit higher correlation with CO conversion. However, several expected effects, such as the influence of temperature and pressure, are not fully captured, highlighting the importance of nonlinear modelling approaches.

PCA reveals further clarifies the structure of the dataset. As shown in Figure 2(a), the first four principal components account for approximately 90% of the total variance. The loading plot in Figure 2(b) indicates that the first two principal components are dominated by catalyst formulation and preparation descriptors, while PC3 and PC4 are more strongly associated with operating conditions such as reaction temperature, pressure, and space velocity. This separation quantitatively supports the need for modelling approaches capable of handling interactions between catalyst-related and operational variables.

Surrogate Model Performance

Table 1 summarizes the predictive performance of the evaluated ML models. The reduction in R^2 and increase in RMSE from training to validation are attributed

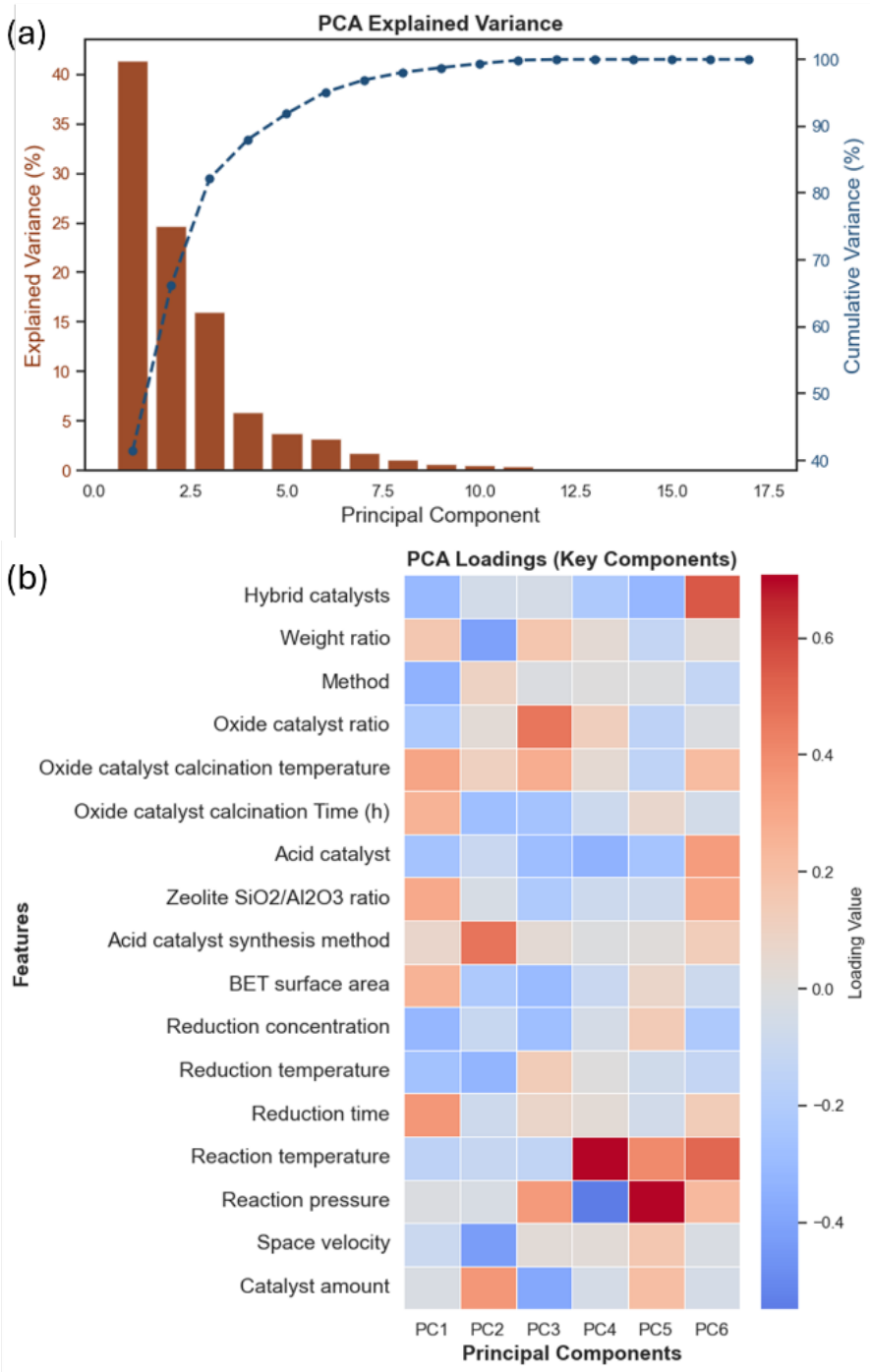

Figure 2. (a) Cumulative explained variance by PCA component, (b) PCA component loading.

to the generalization gap, indicating that models fit the training data more closely than unseen data, with some degree of overfitting. This effect can be reduced through improved hyperparameter tuning, regularization, and increasing dataset size.

While XGBoost and Gradient Boosting achieve similarly high training accuracy (R^2 of 0.99 for both CO conversion and DME selectivity), differences become apparent during validation. Gradient Boosting achieves R^2 of 0.83 for CO conversion and 0.89 for DME selectivity, with RMSE values of 0.01038 and 0.01258, respectively. In comparison, XGBoost shows lower generalization performance, particularly for CO conversion (validation R^2=0.74).

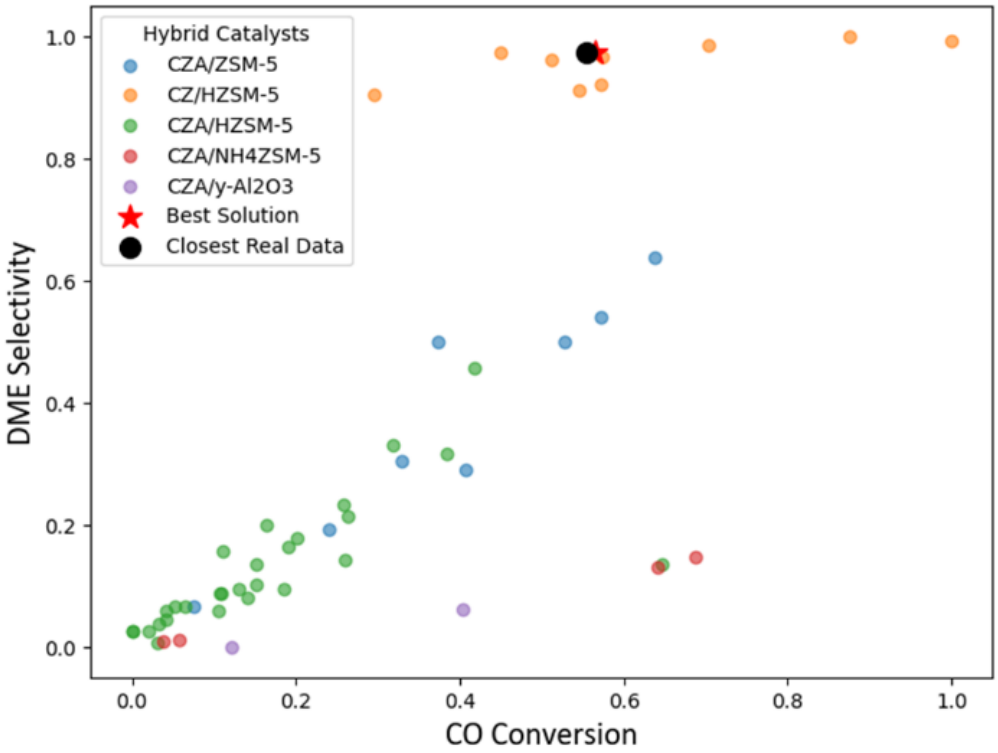

Figure 3. Comparison of the best solution and real datapoints.

Table 1: Predictive performance of the predictive models.

Training			
Method	**Parameter**	**R^2**	**RMSE**
XGBoost	CO Conversion	0.99	0.00002
	DME Selectivity	0.99	0.00001
Gradient Boosting	CO Conversion	0.99	0.00024
	DME Selectivity	0.99	0.00029
CatBoost	CO Conversion	0.97	0.00157
	DME Selectivity	0.98	0.00138
Random Forest	CO Conversion	0.96	0.00236
	DME Selectivity	0.85	0.01000
SVR	CO Conversion	0.83	0.01000
	DME Selectivity	0.91	0.01071
Validation			
Method	**Parameter**	**R^2**	**RMSE**
XGBoost	CO Conversion	0.74	0.01668
	DME Selectivity	0.82	0.01400
Gradient Boosting	CO Conversion	0.83	0.01038
	DME Selectivity	0.89	0.01258
CatBoost	CO Conversion	0.78	0.01332
	DME Selectivity	0.83	0.11604
Random Forest	CO Conversion	0.75	0.01547
	DME Selectivity	0.87	0.01360
SVR	CO Conversion	0.66	0.02115
	DME Selectivity	0.74	0.02619

These results indicate that Gradient Boosting provides the best balance between accuracy and generalization for the available dataset. Consequently, it is selected as the surrogate model for subsequent multi-objective optimization. The performance gap between training and validation metrics also reflects the limited size and heterogeneity of the literature-derived dataset, reinforcing importance of cautious interpretation of surrogate predictions.

Fuzzy-Enhanced NSGA-II Optimization Results

Using the Gradient Boosting surrogate, the fuzzy-enhanced NSGA-II algorithm generates a Pareto front representing the trade-offs between CO conversion and DME selectivity. The fuzzy decision-making step identifies a single compromise solution, which is highlighted in Figure 3 alongside the experimentally reported datapoints. The selected fuzzy-optimal solutions correspond to a CZ/HZSM-5 catalyst prepared via a core-shell method, operated at a reaction temperature of approximately 261°C and a pressure of 37 bar. The surrogate predicts a CO conversion of about 53% and a DME selectivity close to 70%. As shown in Table 2, the closest experimental datapoint exhibits very similar performance values (CO conversion close to 51% and DME selectivity of approximately 70%), lending credibility to the optimized solution.

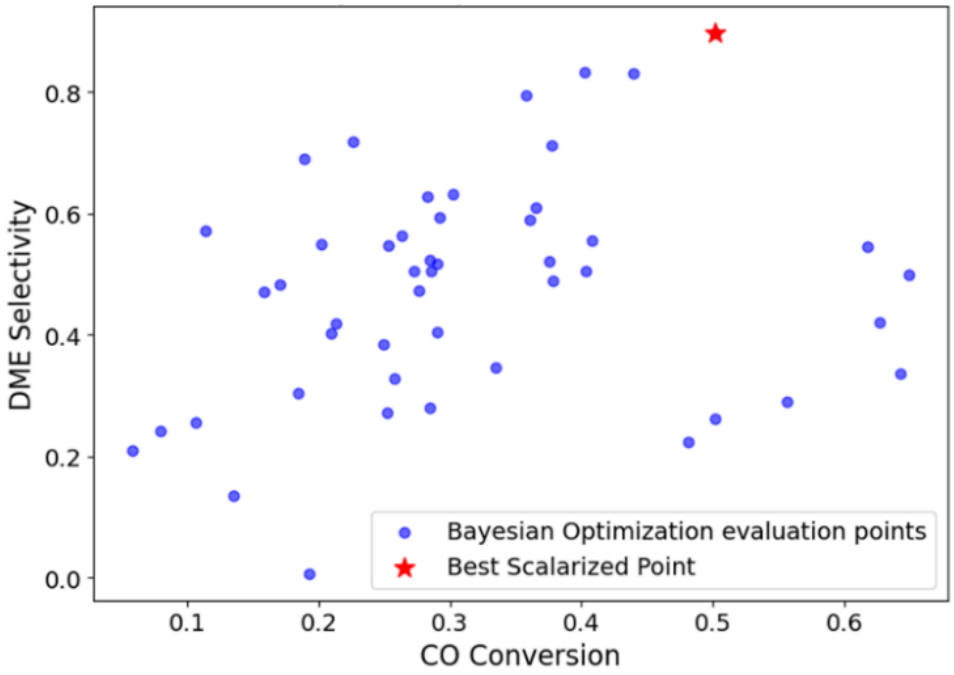

Figure 4. Constrained multi-objective Bayesian optimization.

Table 2: Feature comparison of the best solution and closest real data point.

Feature	Best Fuzzy	Closest real point
Hybrid catalysts	CZ/HZSM-5	
synthesis method	Core-shell	
Oxide catalyst ratio	69:31:0	
Oxide catalyst calcination temperature (°C)	350.8	354.9
Oxide catalyst calcination Time (h)	4.0	67.7
Zeolite SiO2/Al2O3ratio	22.0	
surface area(m2/g)	60.4	67.0
Reduction gas concentration (H2 %)	90.0	100
Reduction temperature (°C)	270	270
Reduction time (h)	1.7	1.0
Reaction temperature (°C)	261	260
Reaction pressure (bar)	37	40
Space velocity (h-1)	14181	6000
Catalyst amount (g)	0.7	0.5
CO conversion (%)	53	51
DME selectivity (%)	70	70

A detailed comparison in Table 2 shows that catalyst formulation and preparation parameters are closely aligned between the fuzzy-optimal solution and the closest real datapoint. In contrast, differences are more pronounced in operating conditions, particularly space velocity and reaction pressure. These deviations suggest that the optimization framework primarily exploits operational flexibility to improve performance while remaining consistent with experimentally validated catalyst formulations.

Bayesian Optimization Results

The constrained multi-objective Bayesian optimization results are illustrated in Figure 4, where each evaluated point represents a surrogate-predicted combination of CO conversion and DME selectivity. The best scalarized solution, indicated by the red star, reflects the optimal compromise according to the equal-weighted objective function.

As summarized in Table 3, Bayesian optimization also identifies CZ/HZSM-5 catalyst prepared by a core-shell method as optimal, confirming consistency with the evolutionary optimization results. However, the recommended operating conditions differ, with a lower reaction temperature (240°C) and a higher space velocity. The predicted performance at this point is close to 47% CO conversion and 65% DME selectivity, which is lower than the fuzzy-enhanced NSGA-II solution. The difference can be attributed to the scalarization and uncertainty-aware nature of Bayesian optimization, which tends to favour conservative solutions when model uncertainty is present. While this approach efficiently explores the design space, it may sacrifice peak performance in favour of robustness.

Table 3: Feature comparison of the best solution from multi-objective Bayesian optimization.

Feature	Best Bayesian optimized point
Hybrid catalysts	CZ/HZSM-5
synthesis method	Core-shell
Oxide catalyst ratio	69:31:0
Oxide catalyst calcination temperature (°C)	360.00
Oxide catalyst calcination Time (h)	4.0
Zeolite SiO2/Al2O3ratio	27.0
surface area(m2/g)	151
Reduction gas concentration (H2 %)	24
Reduction temperature (°C)	261
Reduction time (h)	2.71
Reaction temperature (°C)	240
Reaction pressure (bar)	38
Space velocity (h-1)	11100
Catalyst amount (g)	0.9
CO conversion (%)	46.6
DME selectivity (%)	64.5

Comparative Assessment

A direct comparison of results in Tables 2 and 3 show that both optimization strategies converge toward similar catalyst formulations, indicating that catalyst-related decisions are robust across optimization paradigms. In contrast, operating conditions vary more significantly between methods, leading to differences of up to 6-7 percentage points in predicted CO conversion and approximately 5-6 percentages points in DME selectivity.

This quantitative comparison highlights the central contribution of the present study: different surrogate-based multi-objective optimization strategies, when applied to the same dataset and predictive model, can lead to distinct yet plausible optimal operating conditions. Such differences are particularly relevant for decision-making under limited data availability, where the choice of optimization framework directly influences recommended process conditions.

CONCLUSION

This study presents a surrogate-based multi-objective optimization framework for DME synthesis from syngas that integrates ML with two distinct optimization strategies. Gradient Boosting is identified as a reliable surrogate model capable of capturing nonlinear relationships within a limited experimental dataset. Both fuzzy-enhanced NSGA-II and multi-objective Bayesian optimization successfully identify feasible and competitive operating conditions. While the two approaches agree closely on catalyst formulations, they differ in their recommended operating conditions and predicted performance levels. These differences highlight the importance of selecting an optimization strategy that aligns with the desired balance between exploration, exploitation, and decision-making under uncertainty. These results demonstrate that comparative analysis of optimization paradigms applied to the same surrogate model and dataset provides valuable insight into the robustness and limitations of data-driven decision-support tools for complex catalytic processes such as DME synthesis. Future studies will include targeted experimental validation of the optimized conditions identified by both optimization strategies, with particular emphasis on assessing robustness under realistic process variability. Extending the framework to include additional objectives, such as catalyst stability, by-product formation, or energy efficiency, would enable a more comprehensive assessment of industrial relevance.

ACKNOWLEDGEMENTS

This research was funded by: Project LOCATE - Engineering Low Carbon Technologies for Green Dimethyl Ether, Funding: German Academic Exchange Service (Deutscher Akademischer Austauschdienst - DAAD) under the Programm des Projektbezogenen Personenaustauschs Brasilien CAPES ab 2025 (Project number 57754630)

AUTHOR IDENTIFIERS

Jafari M: 0009-0004-2714-3500
Dorneanu B: 0000-0003-3553-6625
Pontes K: 0000-0002-8208-0854
Arellano-Garcia H: 0000-0002-8297-0232

REFERENCES

1. Amoo CC, Xing C, Tsubaki N, Sun J. Tandem reactions over zeolite-based catalysts in syngas conversion. ACS Cent. Sci. 8:1047-1062 (2022). https://doi.org/10.1021/acscentsci.2c00434
2. Livescu A, Navar R, Mangalindan JR, Mahnaz F, Ge Y, Shetty M, Yang X. Catalysts for clean energy: a review on current progress for the catalyzed recycling of CO2 into dimethyl ether. Top Catal 69:1633-1654 (2024). https://doi.org/10.1007/s11244-024-01913-z
3. Peinado C, Liuzzi D, Sluijter SN, Skorikova G, Boon J, Guffanti S, Groppi G, Rojas S. Review and perspective: next generation DME synthesis technologies for the energy transition. Chemical Engineering Journal 479:147494 (2024). https://doi.org/10.1016/j.cej.2023.147494

4. Das M, Ghosh A, Sunoj RB. Advances in machine learning with chemical language models in molecular property and reaction outcome predictions. J Comput Chem 45:1160-1176 (2024). https://doi.org/10.1002/jcc.27315
5. Singh S, Sunoj RB. Molecular machine learning for chemical catalysis: prospects and challenges. Acc. Chem. Res. 56:402-412 (2023). https://doi.org/10.1021/acs.accounts.2c00801
6. Shim E, Tewari A, Cernak T, Zimmerman PM. Machine learning strategies for reaction development: toward the low-data limit. J. Chem. Inf. Model. 63:3659-3668 (2023). https://doi.org/10.1021/acs.jcim.3c00577
7. Bian K, Priyadarshi R. Machine learning optimization techniques: a survey, classification, challenges, and future research issues. Arch Computat Methods Eng : (2024). https://doi.org/10.1007/s11831-024-10110-w
8. Yazdinejad A, Dehghantanha A, Parizi RM, Epiphaniou G. An optimized fuzzy deep learning model for data classification based on NSGA-II. Neurocomputing 522:116-128 (2023). https://doi.org/10.1016/j.neucom.2022.12.027
9. Garnett R. Bayesian optimization. Cambridge University Press (2023). https://doi.org/10.1017/9781108348973
10. Ramos MC, Michtavy SS, Porosoff MD, White AD. Bayesian optimization of catalysts with in-context learning. arXiv preprint arXiv:230405341. 2023;
11. Kumar A, Pant KK, Upadhyayula S, Kodamana H. Multiobjective bayesian optimization framework for the synthesis of methanol from syngas using interpretable gaussian process models. ACS Omega 8:410-421 (2022). https://doi.org/10.1021/acsomega.2c04919
12. Ma H, Zhang Y, Sun S, Liu T, Shan Y. A comprehensive survey on NSGA-II for multi-objective optimization and applications. Artif Intell Rev 56:15217-15270 (2023). https://doi.org/10.1007/s10462-023-10526-z
13. Blank J, Deb K. Pymoo: multi-objective optimization in python. IEEE Access 8:89497-89509 (2020). https://doi.org/10.1109/access.2020.2990567
14. Santiago A, Dorronsoro B, Nebro AJ, Durillo JJ, Castillo O, Fraire HJ. A novel multi-objective evolutionary algorithm with fuzzy logic based adaptive selection of operators: FAME. Information Sciences 471:233-251 (2019). https://doi.org/10.1016/j.ins.2018.09.005
15. Nakamura K, Otani N, Koike T. Multi-objective bayesian optimization of optical glass compositions. Ceramics International 47:15819-15824 (2021). https://doi.org/10.1016/j.ceramint.2021.02.155
16. Nakamura K, Otani N, Koike T. Multi-objective bayesian optimization of optical glass compositions. Ceramics International 47:15819-15824 (2021). https://doi.org/10.1016/j.ceramint.2021.02.155
17. Shields BJ, Stevens J, Li J, Parasram M, Damani F, Alvarado JIM, Janey JM, Adams RP, Doyle AG. Bayesian reaction optimization as a tool for chemical synthesis. Nature 590:89-96 (2021). https://doi.org/10.1038/s41586-021-03213-y