Surrogate-Based Optimization of the OPEX of a Modular Plant for Biogas Conversion to Methanol Using the MADS Algorithm

Luis Felipe Sánchez Martínez[a], Andrea Galeazzi[a], Flavio Manenti[a,*]

[a]*CMIC Department "Giulio Natta", Politecnico di Milano, Piazza Leonardo da Vinci 32, Milano, 20133, Italy*

flavio.manenti@polimi.it

Abstract

The present work studies the potential of surrogate models for the global optimization of complex chemical processes. In particular, a modular plant for the conversion of biogas to methanol is considered. The Aspen HYSYS simulation of this plant was run 480 times, which ensured the even distribution of points in the input space. The evenness of this design of experiments was evaluated using a discrepancy measurement called the Mixture Discrepancy. With the simulation data, some of the most widely used surrogate models such as regression models and the Kriging Gaussian process were trained. The most accurate model for the prediction of each output variable was selected and used for the optimization of the OPEX. The optimization complemented the trained surrogate models with the Mesh Adaptive Direct Search (MADS) algorithm. For this purpose, the open-access computational implementation of the MADS algorithm called NOMAD was used. With the surrogate-based optimization, the computational times were reduced an 88% with respect to the simulation-based optimization. In addition, the accuracy of the surrogate model was paramount, as an average 0.75% prediction error was found. Consequently, the models proved sufficient for optimizing the studied process, resulting in a 22.2% reduction in the OPEX.

Keywords: surrogate model, optimization, MADS, NOMAD

1. Introduction

The mathematical complexity of a chemical process increases rapidly with the number of involved unit operations or the presence of intricate thermodynamics or kinetics (McBride and Sundmacher, 2019). To tackle these complexities while maintaining computational accuracy, surrogate models, a form of supervised machine learning, have gained relevance (Alizadeh et al., 2020). These mathematically simpler models require data from the process to be trained. Consequently, the quality of the surrogate is related to the quality of the extracted data (McBride and Sundmacher, 2019).

Considering the high accuracy presented by surrogate models in several applications (Alizadeh et al., 2020; Pishkari et al., 2023), this study presents the development of a surrogate model for a modular plant for the production of methanol from biogas. A one-shot space-filling design using the maximin-optimized LHS complemented with discrepancy measurements is adopted to guarantee uniform distribution and unbiased representation of the domain. The surrogate models trained were selected considering the most commonly used in chemical engineering. These models include polynomial regressions, regression trees and support vector machines, previously used by Galeazzi et

al. (2023) for surrogate modeling of an amine-washing section of a plant. Additionally, the Kriging Gaussian process, known for its effectiveness in managing complex mathematical systems, was incorporated. The most accurate surrogate model was then used to optimize the Operational Expenditures (OPEX) through a black-box derivative-free optimization powered by the NOMAD open-access software (Le Digabel, 2011).

2. Problem statement and process description

The objective of the present work is to optimize the OPEX of a modular plant which produces methanol from biogas. This module has been installed in a heat and power plant in Italy, and corresponds to the industrial validation of the BIGSQUID process, an innovative technology licensed by the Politecnico di Milano (Fedeli and Manenti, 2022; Negri et al., 2022). The BIGSQUID (BIoGaS to liQUID) process comprises five main steps: biogas upgrading to remove H_2S, biogas reforming to produce syngas, syngas purification to reach the synthesis requirements, methanol synthesis from syngas, and methanol purification (See Figure 1).
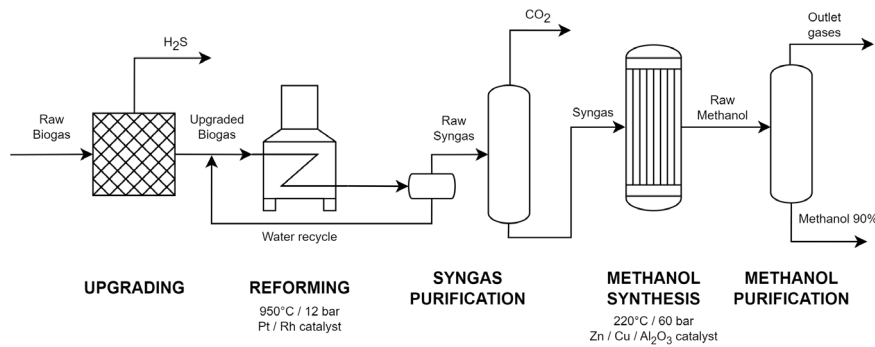

Figure 1: Overview of the BIGSQUID process to produce methanol from raw biogas.

For this process, a rigorous Aspen HYSYS simulation was developed by Fedeli and Manenti (2022) for the SuPER team of the Politecnico di Milano. Even if the number of unit operations shown in Figure 1 seems small and manageable, the presence of two recycle streams in the simulation, which involve critical process variables, considerably increased its complexity. The average simulation time was found to be around 2.77 seconds, while the convergence rate obtained from preliminary experiments was 60%. In a first attempt to optimize the process, the Aspen HYSYS optimizer failed. This failure condition was maintained even when different initial points were tested. This lack of convergence may be correlated to the low convergence rate of the simulation, the high correlation of the variables and their complex interactions. For the mentioned reasons, a different optimization approach should be considered. To avoid increasing the complexity of the problem through complex optimization software such as GAMS, the surrogate approach for the optimization presented an interesting alternative. The surrogate models were trained from real data extracted from the simulation using a computer with a processor 11th Gen Intel® Core™ i9-11900 @ 2.50GHz and 32 GB of RAM.

3. Data generation

The input variables for the rigorous simulation and the surrogate model were selected and delimited considering their physical, chemical, and process-related constraints. The variables are shown in Figure 1, and their boundaries are presented in Table 1.

Table 1: Input variables to the black-box model of the BIGSQUID process.

Stream	Description	Variable	Nominal value	Minimum value	Maximum value
BIOGAS	Raw biogas	F (kg/h)	600.0	500.0	700.0
		x_{CO2}	0.46	0,4	0,5
		x_{CH4}	0.53	0,45	0,7
DEMIWATER	Make-up in reforming recycle	F	10	9	25
2cc	Compressed BIOGAS	P (bar)	16.0	12.0	20.0
BRProduct	Reformer outlet	T (C)	950.0	800.0	1000.0
5b	BRPRoduct after cooling	T (C)	5.0	5.0	25.0
CompressedSyngas	Purified syngas	P (bar)	60.0	50.0	70.0
WATER	Water for syngas purification	F (kg/h)	5000.0	4500.0	6000.0
CLEANWATER	Make-up for syngas purification recycle	T (C)	10.0	5.0	25.0
TOREACTOR	Synthesis reactor inlet syngas	T (C)	250.0	220.0	270.0
7	Synthesis reactor outlet	T (C)	10.0	5.0	25.0

Data from Table 1 was used in a Design of Experiments (DoE) based on the maximin-optimized LHS, one of the most widely implemented DoE methods in the chemical industry (McBride and Sundmacher, 2019). The number of samples was set on 480, which was the lowest number of samples to minimize the Mixture Discrepancy, an accurate metric for the evaluation of the space-filling characteristics of a DoE (Zhou et al., 2013). These samples were introduced to the Aspen HYSYS simulation of the BIGSQUID. With the simulation data, the dataset for the training of the surrogate models was obtained.

4. Surrogate modeling

The input/output relation of the variables was determined by training 10 regression models, previously considered by Galeazzi et al. (2023), plus the Kriging Gaussian process. All the models were trained once for each output variable. The accuracy of the models was determined using the average Mean Absolute Error (MAE) of a 5 k-fold cross-validation. As the MAE was calculated over normalized data, the reported MAE is a representation of the relative error of the models. Equations 1 and 2 show the data normalization process and the MAE calculation, respectively. The output variables, their description inside the process and the best surrogate model for its prediction using the MAE as evaluation metric is presented in Table 2. In addition, Figure 2 presents the parity plots of the output variables, which confront the simulated (real) and predicted values for each variable.

$$x_{new} = \frac{|x - x_{min}|}{|x_{max} - x_{min}|} \tag{1}$$

$$MAE = \frac{\sum_i |y_i^{real} - y_i^{predicted}|}{n} \tag{2}$$

Table 2. Best Surrogate Models

Stream	Description	Variable	Best model	MAE (%)
METHANOL	Outlet methanol stream	F (kg/h)	Kriging	1,00%
		xMeOH	Kriging	1,37%
Qreactor	Reformer heat requirement	Q (kW)	Kriging	1,59%
QP100	Reformer make-up recycle pump	Q (kW)	Linear Regression	0,00%
Qk2	Syngas compressor before purirfication	Q (kW)	Kriging	0,57%
Duty	Synthesis reactor cooling requirement	Q (kW)	Kriging	1,00%
Q104	Cooling of refrigerant	Q (kW)	Kriging	1,61%
Q107	Cooling of refrigerant	Q (kW)	Kriging	1,33%
K100	First stage compressor of raw biogas	Q (kW)	Third Order Polynomial Regression	0,01%
K100-1	Second stage compressor of raw biogas	Q (kW)	Third Order Polynomial Regression	0,03%
QP-02	Reformer recycle pump	Q (kW)	Kriging	1,17%
P101	Syngas purification recycle pump	Q (kW)	Second Order Polynomial Regression	0,00%
Q03	Syngas cooling during purification	Q (kW)	Kriging	0,27%
			Average	0,77%

5. Optimization

The objective function for optimization is the OPEX of the system described as:

$$\min OPEX = \sum_{i=1}^{11} C_i Q_i$$

$$s.t.: l_j \leq x_j \leq u_j \forall j \in \{1, \ldots, 12\}$$

$$10.0 - METHANOL.F \leq 0$$

$$0.90 - METHANOL.x_{MeOH} \leq 0$$

(3)

Where i represents each of the 11 output energetic streams of Table 2, C_i represents the cost of each energetic stream and Q_i its heat flow. The cost of each energetic stream depends on the utility required and was extracted from Turton et al. (2012). Moreover, j represents each of the 12 input variables of Table 1 delimited by its lower l_j and upper u_j boundaries. Streams BIOGAS and CLEANWATER have not been considered for optimization as they are disturbances of the process and not optimization variables. Finally, the required flowrate and purity of methanol in the plant, which correspond to 10 kg/h and 90% respectively, were set as the constraints for the METHANOL stream.

By implementing the trained surrogate models and the derivative-free open-access black-box optimizer NOMAD, the optimization process was repeated. In this case the

optimization time was 86 seconds, which represents a reduction of 88% in the computational time with respect to the original optimization with the simulator, which required around 700 seconds. The results were validated by introducing the found optimal values to the Aspen HYSYS simulation. Nevertheless, the objective function and the constraints were not respected. To determine the cause of the mismatch, a sensitivity analysis with the surrogated model was proposed. The results of the sensitivity analysis showed a high sensitivity of the variable DEMIWATER, which monotonically (and almost linearly) increased the total OPEX. Considering this behavior, the upper boundary of the variable was re-set to 15 kg/h in Table 1.

With the new boundaries for DEMIWATER, the surrogate modelling methodology presented in this work was repeated. In this case, the validated results were satisfactory, and reduced the OPEX 22.2% compared to the base case. Table 3 presents a summary of the results from both optimizations, before and after the sensitivity analysis compared to the nominal case. As an important remark, Figure 2 and Table 2 present the results of the surrogate model trained after the sensitivity analysis.
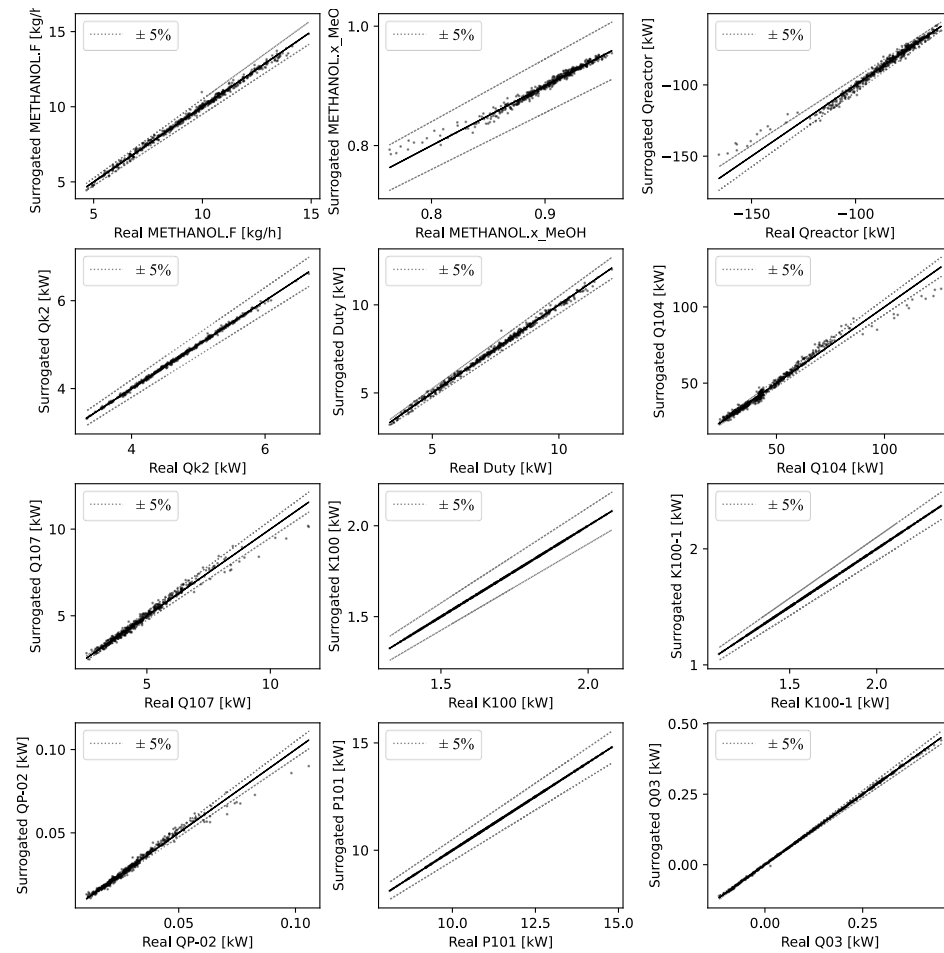

Figure 2. Parity plots of the output variable of the BIGSQUID process after the surrogate modeling process. Dotted lines delimit the zone with a prediction error below ± 5%.

Table 3. Results for the optimization of the BIGSQUID process.

Variable	UoM	Original	Before Sensitivity	After Sensitivity
BRProduct.T	°C	950.0	849.8	800.0
DEMIWater.F	kg/h	10.0	9.0	9.0
CompressedSyngas.P	bar	60.0	66.9	59.0
TOREACTOR.T	°C	250.0	265.0	255.9
7.T	°C	10.0	5.1	20.3
5b.T	°C	5.0	5.2	25.0
2cc.P	bar	16.0	12.0	12.0
WATER.F	kg/h	5000.0	4664.0	4500.0
OPEX Surrogate	$/y	22195	14929	17286
OPEX HYSYS	$/y	23569	20195	18329
OPEX reduction	%	0.0%	14.3%	22.2%

6. Conclusions

In this work, the surrogate-based optimization of a complete industrial process was successfully done. The methodology proposed allowed the training of highly accurate surrogate models for the surrogate-based optimization of a biogas to methanol modular plant. With the study, the values of the input variables which reduced the OPEX of the process of the biogas modular plant by 22.2% were found and validated with the rigorous simulation. The presented methodology ensures the complete consideration of the input domain by implementing a space-filling design, allows the determination of critical variables via surrogate-based sensitivity analysis, reduces the computational times for the optimization of the process and ensures the determination of an accurate optimum of the process by implementing the MADS algorithm by NOMAD.

References

R. Alizadeh, J.K. Allen, F. Mistree, 2020. Managing Computational Complexity Using Surrogate Models: A Critical Review. Res. Eng. Des. 31, 275–298.

M. Fedeli, F. Manenti, 2022. Assessing Process Effectiveness with Specific Environmental and Economic Impact of Heat, Power & Chemicals (HPC) Option as Future Perspective in Biogas. Clean. Chem. Eng. 2, 100016.

A. Galeazzi, K. Prifti, C. Cortellini, A. Di Pretoro, F. Gallo, F. Manenti, 2023. Development of a surrogate model of an amine scrubbing digital twin using machine learning methods. Comput. Chem. Eng. 174, 108252.

S. Le Digabel, 2011. Algorithm 909: NOMAD: Nonlinear Optimization with the MADS Algorithm. ACM Trans. Math. Softw. 37, 44:1-44:15.

K. McBride, K. Sundmacher, 2019. Overview of Surrogate Modeling in Chemical Process Engineering. Chem. Ing. Tech. 91, 228–239.

F. Negri, M. Fedeli, M. Barbieri, F. Manenti, 2022. A versatile modular plant for converting biogas into advanced biofuels. Invent. Discl. 2, 100008.

R. Pishkari, M. Fechtner, T. Keßler, A. Kienle, 2023. Optimization of Simulated Moving Bed Chromatographic Processes using Surrogate Models, in: Kokossis, A.C., Georgiadis, M.C., Pistikopoulos, E. (Eds.), Computer Aided Chemical Engineering, 33 European Symposium on Computer Aided Process Engineering. Elsevier, pp. 343–348.

R. Turton, R.C. Bailie, W.B. Whiting, J.A. Shaeiwitz, D. Bhattacharyya, 2012. Analysis, synthesis, and design of chemical processes, Fourth edition. ed, Prentice Hall international series in the physical and chemical engineering sciences. Prentice Hall, Upper Saddle River, N.J. Munich.

Y.-D. Zhou, K.-T. Fang, J.-H. Ning, 2013. Mixture discrepancy for quasi-random point sets. J. Complex. 29, 283–301.