Anton A. Kiss, Edwin Zondervan, Richard Lakerveld, Leyla Özkan (Eds.)
Proceedings of the 29th European Symposium on Computer Aided Process Engineering
June 16th to 19th, 2019, Eindhoven, The Netherlands. © 2019 Elsevier B.V. All rights reserved.

A model identification approach for the evaluation of plant efficiency

Keivan Rahimi-Adli[a,b,*], Patrick D. Schiermoch[a], Benedikt Beisheim[a,b], Simon Wenzel[b] and Sebastian Engell[b]

[a]*INEOS Manufacturing Deutschland GmbH, Alte Str. 201, 50769 Köln, Germany*
[b]*Process Dynamics and Operations Group, Department of Biochemical and Chemical Engineering, TU Dortmund University, Emil-Figge-Str. 70, 44227 Dortmund, Germany*
keivan.rahimi-adli@ineos.com

Abstract

Regulations and the public expectations on improving efficiency, reducing the carbon footprint and lowering the environmental impact drive the process industry towards improved operation and the development of new technologies. The efficiency of an existing production plant depends on a variety of factors like capacity utilisation, raw material quality, ambient temperature or operational performance. Identifying the influence of these factors on the performance of the plant helps to take suitable measures to drive it towards a more efficient operation. One approach to assess the resource efficiency potential of a plant is the comparison of the actual performance with the best possible operation under the given circumstances. This work presents a surrogate modelling approach for the identification of the best possible operation based on historical data. The surrogate model is compared to a more detailed rigorous model and advantages and possible shortcomings of the surrogate approach are discussed based on real production data at INEOS in Köln.

Keywords: Surrogate models, plant efficiency, energy management systems, model detail level, model identification

1. Introduction

The process industry is constantly developing methods for the evaluation of their resource consumption and the identification of possible improvement potentials. In cases where there are no structural changes of the process planned, on the one end this can be achieved by implementing Advanced Process Control (APC) solutions, which is time consuming and expensive. On the other end, there are well-trained operators who are often able to realize a significant fraction of the APC saving potential, if information on the magnitude of the performance gaps is accessible and the experience about how to improve the operational efficiency can be used. The international standard ISO50001:2011 demands the use of energy performance indicators (EnPI) which have to be compared with an energy baseline. The idea of the Best Demonstrated Practice (BDP), as shown in Fig. 1, is similar and provides the operators with a performance reference model which represents the most resource efficient and stable production at a specific instance of non-influenceable

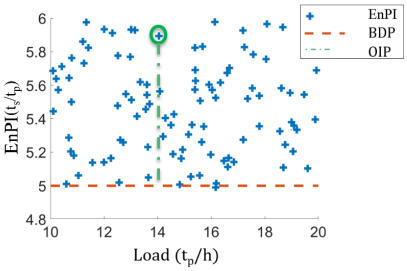


Figure 1: Illustrative example of the BDP concept

circumstances, as e.g. plant load. By comparing the real-time EnPI, which is defined as the specific steam consumption at a particular point in time in this example, with its BDP, the operational improvement potentials (OIP) can be identified (NAMUR WG 4.17, 2017).

2. BDP Model

The level of detail of the BDP models can vary from rigorous models to simple input-output models. The accuracy and the extrapolation capability of rigorous models are advantageous. However, the time and effort needed to develop these models is usually high. On the other end of the scale are input-output models based on linear regression which cannot always represent the complexities of the chemical processes properly.
An alternative to these two types of models are surrogate models (see e.g. Cozad et al. (2014)). In this modelling approach the process data is used to fit a model to the observations. The development of such models requires considerably less effort compared to the development of rigorous models and they are able to capture the plant behaviour better than linear regression models. Furthermore, the evaluation of the surrogate models is easier which makes them attractive for large scale optimisation (Søndergaard, 2003). On the downside, the validity of the surrogate models is limited to the range of the data that is used for their fitting and the results of the model for extrapolated data must be used with care.

2.1. Surrogate modelling procedure

This section briefly discusses the approach to surrogate modelling of the BDP curves that has been developed and implemented at INEOS in Köln. A detailed description of the procedure can be found in Beisheim et al. (2018). As depicted in Fig. 2, the method consists of five general steps. The first step is the acquisition of measurement data. The goal of this step is to collect representative and reliable measurements of the performance of the plant, where aspects as e.g. removal of gross errors, stationarity, selection of suitable sampling times and removal of abnormal operation windows have to be considered.

The second step is the preprocessing of the data. The goal of this step is to remove outliers, classify the data and to standardise. For the latter mean centring and unit variance scaling is used. Data standardisation is useful for the clustering of the data, which is explained in the next step.

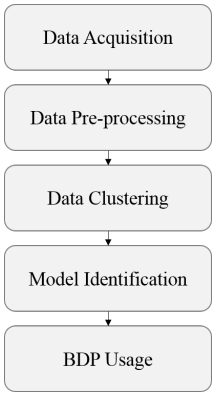


Figure 2: Steps of the BDP modelling procedure

The third step of the method is data clustering. The goal of this step is to reduce the large amount of the measurement data into a much smaller number of a few representatives of different operating regimes, which are then used for model fitting. In this work, the kmeans++ algorithm (Arthur and Vassilvitskii, 2007) is applied, which is an extension of the kmeans algorithm (MacQueen et al., 1967). The distance metric used for clustering is the Euclidean distance of the data from the cluster centres and therefore it is sensitive to the magnitude of the data and is prone to the assignment of a higher influence on a variable with a higher magnitude. As the important factor to consider in this work is the effect of the variation of the non-influenceable factors on the resource efficiency, the data is pre-processed by the subtraction of the mean of each variable and dividing the values by the standard deviation. The kmeans++ algorithm requires the number of the clusters as an input. The clustering algorithm assigns a centre to each cluster that represents the average of the points in that cluster. As the goal of the BDP model is to calculate the most efficient operational domain, the cluster centres are not used as representative values. Instead, a percentile analysis for each cluster

is performed:

$$\mathbf{r}_j = \frac{1}{\left|\mathscr{R}_j\right|} \sum_{\mathbf{x}\in\mathscr{R}_j} \mathbf{x} \tag{1}$$

$$\mathbf{x} \in \mathscr{R}_j \quad \forall \quad P_{j,n} \leq \mathrm{EnPI}(\mathbf{x}) \leq P_{j,m} \tag{2}$$

$$\mathscr{R}_j \subseteq \mathscr{X} \tag{3}$$

$$\mathrm{EnPI}_{\mathscr{R}_j} = \frac{1}{\left|\mathscr{R}_j\right|} \sum_{\mathbf{x}\in\mathscr{R}_j} \mathrm{EnPI}(\mathbf{x}), \tag{4}$$

where $\mathscr{R}_j$ denotes the set of points which are assigned to the cluster j, $\left|\mathscr{R}_j\right|$ is its cardinality and $\mathbf{r}_j$ is the cluster centre. $P_{j,n}$, $P_{j,m}$ are the lower and and upper percentile bounds for cluster j. These bounds are used to select achievable good operation points as representatives, $\mathrm{EnPI}_{\mathscr{R}_j}$, for each cluster. The chosen percentiles are tuning factors and can be modified for each case.

The surrogate model development is an adapted ALAMO approach (Cozad et al., 2014). It generates simple and accurate models from simulated or experimental data. In order to overcome the drawbacks of linear regression models, ALAMO selects a combination of transformed inputs using simple basis functions that fit the responses with an acceptable accuracy. The set of basis functions are defined by the user and ALAMO selects the most suitable ones and their respective parameters as a result of an optimisation. The details of the implemented adapted version of the ALAMO approach can be found in Beisheim et al. (2018). The model is fitted by solving an optimisation problem formulated as:

$$\min_{\beta,y} \quad \sum_{i=1}^{N} e_i \tag{5}$$

$$\text{s.t.} \quad e_i \geq z_i - \sum_{j\in\mathscr{B}} \beta_j X_{ij} \quad i = 1,\ldots,N \tag{6}$$

$$e_i \geq \sum_{j\in\mathscr{B}} \beta_j X_{ij} - z_i \quad i = 1,\ldots,N \tag{7}$$

$$\sum_{j\in\mathscr{B}} y_j = B \tag{8}$$

$$-U_j(1-y_j) \leq \sum_{i=1}^{N} X_{ij}\left(z_i - \sum_{j\in\mathscr{B}} \beta_j X_{ij}\right) \leq U_j(1-y_j) \qquad j \in \mathscr{B} \tag{9}$$

$$\beta^l y_j \leq \beta_j \leq \beta^u y_j \qquad j \in \mathscr{B} \tag{10}$$

$$y_j = \{0,1\} \qquad j \in \mathscr{B}, \tag{11}$$

where z_i are the values of the responses, X is the matrix of the transformed inputs and β is the coefficient vector of X. $\mathscr{B}$ is the set of the basis functions and B is the number of the maximum allowed terms of the model. The binary variables y_j are equal to 1 if their respective basis function is selected. β^u, and β^l are the upper and lower bounds for the coefficient vector. N is the number of the samples and the indices i and j correspond to the samples and the basis functions respectively. The optimisation is done repeatedly for increasing values of B and the suitable level of complexity is decided based on the modified corrected Akaike Information Criterion ($\mathrm{AIC_c}$). Eq. 9 is an additional constraint, which as described in Beisheim et al. (2018) uses the relaxed bounds U_j to convert the problem formulation into an MILP.

2.2. Application to production data at INEOS in Köln

In this section, the surrogate BDP modelling approach introduced in Sec. 2.1 is applied to two sections of the ethylene oxide production plant at INEOS in Köln and the results are compared to

the results of a rigorous model. All of the plots and information in this section are presented in the scaled space due to the confidentiality of the data.

2.2.1. Surrogate approach

The algorithm described in Sec. 2.1 is used with $[1, x^{\pm[1,2,3]}, \exp(x)]$ as basis functions. The first example is a section of the plant which processes the products from the ethylene oxide reactor. The upper and lower percentile limits, $P_{j,n}$ and $P_{j,m}$ are defined as 5 and 10 respectively. As a result of discussions with plant personnel, the production load ($\dot{m}_p$) is defined as the only non-influenceable factor and is used to calculate the resource efficiency of the specific steam consumption ($\text{EnPI}_{s,1}$). The number of the clusters is defined as 16. The algorithm results in a model with two basis functions:

$$\text{EnPI}_{s,1} = -3.1902\,\dot{m}_p^2 + 1.4843\exp(\dot{m}_p) \qquad (12)$$

Fig. 3 depicts the results of the model fitting. Each data cloud (shown by coloured points) represents a cluster and the circles and the + signs represent the cluster centres and the percentile centres that are used for model fitting. As mentioned in Sec. 2, extrapolation of the model should be avoided. Therefore, the range of validity of the model is defined as the range between the minimum and maximum of the percentile centre values, and the data outside of this range are shown with dark diamonds.

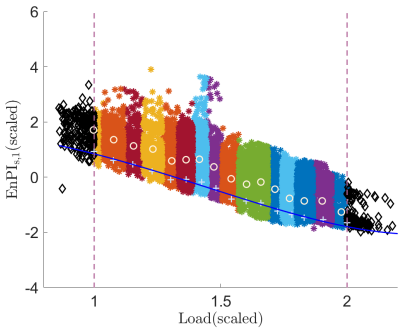


Figure 3: Surrogate BDP model of the product processing section of the ethylene oxide plant

The second investigated section is the ethylene oxide reactor. Both the main and the side reactions are exothermic and the produced heat is used for steam generation. The heat of reaction of the side reaction is significantly higher than the heat produced by the main reaction. Therefore, the reduction of the catalyst selectivity increases the steam production. Considering this fact, the load of the reactor ($\dot{m}_{p,r}$) and the catalyst selectivity (S) which is available as a measurement, are used as non-influenceable factors for modelling the resource efficiency of the steam production ($\text{EnPI}_{s,2}$). As the number of non-influenceable factors in this case is higher than the previous one, the number of the clusters is increased to 30. The other modification in this case is the choice of the percentile limits. Considering that this case is about the modelling of the production of steam, so higher values are preferable, the BDP is defined as the highest possible amount of steam production. Therefore the lower and upper percentile limits are defined as 90 and 95 respectively. The algorithm identifies a model, shown in Fig. 4, using three basis functions as:

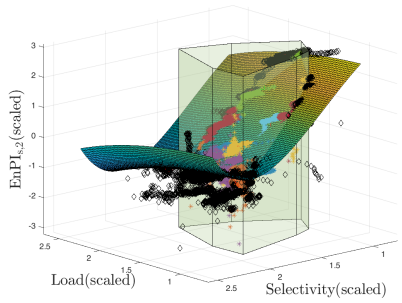


Figure 4: Surrogate BDP model of the steam production of the ethylene oxide reactor

$$\text{EnPI}_{s,2} = -0.0669\,\dot{m}_{p,r}^3 - 5.7478\,S^3 + 3.0651\exp(S). \qquad (13)$$

A conservative method for the definition of the domain of validity of regression models in higher dimensions is the "Convex Hull" of the input variables (Brooks et al., 1988). This method is used here to calculate the validity bounds of the model. The importance of the validity range is more obvious in this example compared to the previous model. It can be seen that outside of the convex hull, the model diverges from the data and should not be trusted.

2.2.2. Rigorous approach

In this section, a more detailed model of the ethylene oxide reactor is developed that is based on physico-chemical principles and measurement data. This model is then used to make a comparison between the detailed model and the proposed surrogate modelling approach and to identify the advantages and shortcomings of each. The model is based on the energy balance around the reactor as follows:

$$\frac{\partial Q}{\partial t} = \bar{C}_{p,in}\dot{n}_{in}T_{in} - \bar{C}_{p,out}\dot{n}_{out}T_{out} - \dot{Q}_{cool} + \sum_i(-\Delta H_{R,i}r_i) - \dot{Q}_{loss}, \tag{14}$$

where the left hand side of the equation is set to zero due to the assumed stationarity of the process. The first two terms on the right hand side of the equation represent the amount of the heat entering and leaving the reactor with the material streams. $\bar{C}_{p,j}$, $\dot{n}_j$, T_j denote average specific heat capacity of the gas mixture, mole flow and temperature at position j. $\sum_i(\Delta H_{R,i}r_i)$ represents the produced heat of reaction, $\dot{Q}_{cool}$ is the heat transferred from the reactor to the cooling fluid inside the jacket and $\dot{Q}_{loss}$ is the amount of heat that is transferred to the environment.

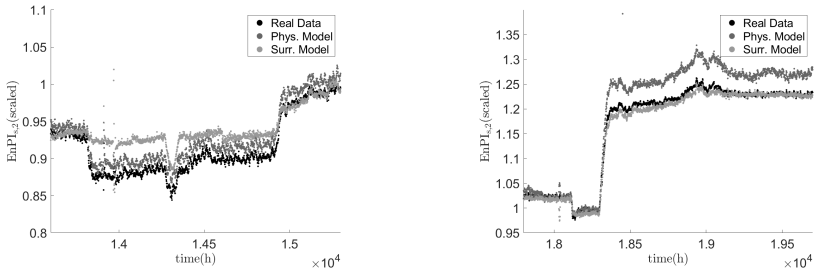
$\bar{C}_p$ is calculated as the weighted average of the values $C_{p,j}$ of the different components using the volumetric percentage values available in the plant, where the dependencies of the values of $C_{p,j}$ of the components on the temperature are determined based on the Shomate equation (Linstrom and Mallard, 2018). The molar flow $\dot{n}_{in}$ is not measured, but can be calculated based on the volumetric flow rate and the ideal gas law. The situation for $\dot{n}_{out}$ is more complex, as there is no direct flow measurements at the output of the reactor. However, using the composition measurements and the change of the volumetric percentage of an inert gas present in the mixture, this flow can be estimated.
The values of $\Delta H_{R,i}$ for each reaction are known and $\sum_i(-\Delta H_{R,i}r_i)$ is calculated using this information together with the selectivity measurement and the estimation of the ethylene conversion based on its volumetric percentage at the inlet and outlet flows. $\dot{Q}_{cool}$ is assumed to be equal to the amount of the heat required for the production of steam.

Considering the above mentioned assumptions, the value of $\dot{Q}_{loss}$ can be calculated from eq. 14. $\dot{Q}_{loss}$ can also be described as:

$$\dot{Q}_{loss} = k(T_{cf} - T_{amb}), \tag{15}$$

where T_{cf} and T_{amb} are equal to the cooling fluid and ambient temperature, and k is the heat transfer coefficient. Assuming that k remains constant for an unchanged setup (physical structure, cooling fluid etc.), its value can be estimated based on eq.15.

Eq. 14 is rearranged and $\dot{Q}_{cool}$, and thus the amount of steam produced is calculated based on the energy balance. Fig. 5 depicts the real steam production together with the values that are calculated based on the surrogate BDP model and the detailed model at two different time windows. The x-axis and the y-axis represents the time and the resource efficiency of steam production $\text{EnPI}_{s,2}$.
From Fig. 5a it can be seen that the surrogate model predicts a higher steam production rate compared to the real data and the detailed model. As the result of a deeper investigation it became clear that the conversion of ethylene in the reactor is lower in this period, which is not considered by the surrogate model and thereby deviates from reality. It is expected that the addition of conversion as a new non-influenceable factor to the surrogate model improves the accuracy of its prediction. Results from a second interesting time window are presented in Fig. 5b. Here it can be seen that the amount of produced steam predicted by the detailed model is significantly higher after a period of time. The reason is, that the heat exchanger that is used for steam production has a limit for the heat removal. Therefore, as a result of the decrease of the catalyst selectivity and the increase of the amount of heat produced by the reaction, the heat removal system is not able to convert all of the heat into steam, and this heat is lost in an extra heat exchanger against a large stream of cooling



(a) Comparison 1 (b) Comparison 2

Figure 5: Comparison of the physical and the surrogate BDP models against production data

water. This limitation is identified with the help of the rigorous model, which is not possible with a surrogate BDP model.

3. Conclusions

In this work a surrogate and a more detailed modelling approach were used to calculate the BDP of production plants. It was observed that the surrogate modelling approach has a good performance in fitting the BDP data and requires a considerably lower modelling effort compared to more rigorous models. The proposed surrogate modelling approach is flexible and does not require a large effort. The time spent for the development of the surrogate model of the reactor was approximately 20 % of time that was needed for the development of the rigorous model. However, the range of trustworthiness of surrogate models is limited to the range of the observed data. Therefore, the performance of the models cannot be trusted beyond the data used for fitting them. Also, as discussed above for the situation depicted in Fig. 5b, deviations of the observed BDP from the possible best performance cannot be detected using a surrogate model and improvements of the plant that realize the potential cannot be deduced.

Acknowledgment

The project leading to this publication has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 723575 (CoPro) in the framework of the SPIRE PPP.

References

D. Arthur, S. Vassilvitskii, 2007. k-means++: The advantages of careful seeding. In: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms. Society for Industrial and Applied Mathematics, pp. 1027–1035.

B. Beisheim, K. Rahimi-Adli, S. Krämer, S. Engell, 2018. Energy performance analysis of continuous processes using surrogate models. Submitted to Energy, Manuscript No. EGY-D-18-07030.

D. G. Brooks, S. S. Carroll, W. A. Verdini, 1988. Characterizing the domain of a regression model. The American Statistician 42 (3), 187–190.

A. Cozad, N. V. Sahinidis, D. C. Miller, 2014. Learning surrogate models for simulation-based optimization. AIChE Journal 60 (6), 2211–2227.

P. Linstrom, W. Mallard, 2018. Nist chemistry webbook, nist standard reference database number 69, national institute of standards and technology data.

J. MacQueen, et al., 1967. Some methods for classification and analysis of multivariate observations. In: Proceedings of the fifth Berkeley symposium on mathematical statistics and probability. Vol. 1. Oakland, CA, USA, pp. 281–297.

NAMUR WG 4.17, 2017. Resource efficiency indicators for monitoring and improving resource efficiency in processing plants. NAMUR Recommendation 162, NAMUR, NAMUR, Leverkusen, Germany.

J. Søndergaard, 2003. Optimization using surrogate models-by the space mapping technique. Ph.D. thesis, Technical University of Denmark.