A hybrid approach for advanced monitoring and forecasting of fouling with application to an ethylene oxide plant

Joel Sansana1, Ricardo Rendall2, Ivan Castillo3, Luc de Bruijne2, Jonathan Huggins2, Ailene Phillips2, and Marco Reis4

1University of Coimbra Faculty of Sciences and Technology
2The Dow Chemical Company
3Dow Chemical Company
4University of Coimbra

May 3, 2023

Abstract

Fouling in heat exchangers leads to increased pressure drop, associated with higher energy consumption, utility costs, and CO$_2$ emissions. However, other effects can also take place, threatening process operations and safety. This is the case of ethylene oxide operations, where unplanned outages and decomposition events pose significant safety risks. Therefore, the development of a framework for advanced monitoring and forecasting of heat exchanger fouling is both important and opportune to improve the reliability and safety of the operation. We propose a hybrid approach, where knowledge-based feature generation is integrated with data-driven methods, to forecast a key performance indicator (KPI) that acts as a fouling surrogate. The forecasting model can predict one-month ahead with an accuracy of $R^2 = 0.7$. We also show that long-term forecasting is possible with this model, which can be applied to optimize maintenance scheduling. The solution can be extended to other situations, where fouling takes place.
A hybrid approach for advanced monitoring and forecasting of fouling with application to an ethylene oxide plant

Joel Sansana\textsuperscript{a}, Ricardo Rendall\textsuperscript{b}, Ivan Castillo\textsuperscript{b}, Luc de Bruijne\textsuperscript{b}, Jonathan Huggins\textsuperscript{b}, Ailene Phillips\textsuperscript{b}, Marco S. Reis\textsuperscript{a}.

\textsuperscript{a}Univ Coimbra, CIEPQPF, Department of Chemical Engineering, Rua Sílvio Lima, Pólo II – Pinhal de Marrocos, 3030-790 Coimbra, Portugal
\textsuperscript{b}The Dow Chemical Company

Abstract

Fouling in heat exchangers leads to increased pressure drop, associated with higher energy consumption, utility costs, and CO\textsubscript{2} emissions. However, other effects can also take place, threatening process operations and safety. This is the case of ethylene oxide operations, where unplanned outages and decomposition events pose significant safety risks. Therefore, the development of a framework for advanced monitoring and forecasting of heat exchanger fouling is both important and opportune to improve the reliability and safety of the operation. We propose a hybrid approach, where knowledge-based feature generation is integrated with data-driven methods, to forecast a key performance indicator (KPI) that acts as a fouling surrogate. The forecasting model can predict one-month ahead with an accuracy of $R^2 = 0.7$. We also show that long-term forecasting is possible with this model, which can be applied to optimize maintenance scheduling. The solution can be extended to other situations, where fouling takes place.

Keywords: Fouling; Ethylene Oxide; Feature engineering; Machine learning; Monitoring; Forecasting;

1. Introduction

Heat exchanger fouling in ethylene oxide (EO) plants is a major long-drawn-out problem that can lead to unplanned outages and temperature excursions or decomposition events, therefore constituting a major risk for operations and safety \cite{1, 2}. Fouling of the shell & tubes heat exchanger that preheats the reactor inlet and cools down the reactor outlet stream, reduces energy recovery, and increases pressure drop. The increase in pressure drop eventually leads to added expenditures in utilities (steam) to keep the gas flowing at the target design levels. Furthermore, fouling deposits originate localized heat transfer dead zones where hotspots can form leading to strong temperature excursions and decomposition events \cite{3}. All these effects are detrimental to the plant of operation and should be managed as closely as possible to secure the stability, economy, and safety of the
process. However, in most circumstances, fouling is not directly observable or measured, and the only alternative available is to infer it from the analysis of collected data or field measurements, usually in a rather \textit{ad hoc} fashion, depending on the personnel experience.

Fouling is a process that consists of the deposition, accumulation, and aging of suspended solids or insoluble salts on the surface of heat exchangers \cite{4, 5}. This phenomenon increases surface thickness and decreases conductivity, thus increasing heat transfer resistance \cite{6}. Fouling alone has a meaningful economic impact on US manufacturing costs, estimated that the effects reached 25 basis points of the country’s gross domestic product in the 1980s and early 1990s \cite{5}. In the dairy industry, 80\% of the total operating costs involved are due to fouling \cite{7} while in an oil refinery, 6\% of the total processed crude oil is used by the process itself, a number exacerbated by the presence of fouling. Furthermore, considering that industry is responsible for 23\% of \(\text{CO}_2\) emissions in the US \cite{8}, reducing energy expenditure helps companies achieve their environmental, social, and corporate governance (ESG) goals.

Even though under certain circumstances fouling can be measured online using specific sensors, this is not the most common scenario in the industry. However, its effects are felt and can be captured indirectly using other most commonly available sensors. Literature on this subject matter can be segmented into two major classes of approaches: one based on monitoring fouling dynamics using its indirect effects \cite{9}; and the other based on digital twins of heat exchangers based on first-principles models that incorporate detailed fouling dynamics aspects \cite{10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21}.

The latter approach heavily depends on first-principles knowledge of the heat exchanger system and on modeling expertise on the part of the practitioner. These models are distributed parameters models that include wall models, fouling deposition and aging, as well as, removal dynamics. Often, these are extensions of classical fouling models such as Kern & Seaton \cite{22} and Ebert & Panchal \cite{23}.

Indirect methods, on the other hand, usually rely on pressure drop measurements obtained with pressure sensors placed in inlet and outlet streams \cite{24, 25, 26, 27, 28}. This quantity is very important because excessive pressure can cause damage to the heating equipment. However, it is relatively insensitive to the deposition of thin layers and it also works better in plate heat exchangers in comparison with tubular arrangements. It is worth noting that the pressure drop varies a lot with flow changes that may not occur as a consequence of fouling. Therefore, it is advisable to compute a rolling average in order to smooth the pressure drop signal \cite{24} or do an aggregate of the pressure per batch if one does not operate in a continuous process \cite{25}. Pressure drop as a surrogate for fouling achieves good results for fouling detection \cite{24}, as well as, fouling monitoring using a partial least squares (PLS) regression model \cite{25}. Furthermore, when a fouling model based on pressure drop can be included in a state space formulation \cite{27} to help decision-making regarding maintenance scheduling \cite{26}.

The monitoring of temperature is also presented as an efficient way for online fouling detection. \cite{29} used wavelets to clean temperature measurement signals and compared the resulting temperature profiles in the absence of fouling to the ones in the presence of fouling. Nevertheless, a detectable temperature change does not happen until a critical thickness of deposit develops, making the approach insensitive to the initial stages of fouling phenomena \cite{30, 24}. Finally, monitoring of heat transfer parameters is also used as a surrogate for fouling. These set of parameters
include the amount of heat transferred, the overall heat transfer coefficient \[31\] and the fouling resistance \[32, 33, 34, 35, 9, 36\]. A selected set of literature examples are discussed below, and a table summarizing surrogate measurements for fouling can be found in \[9\].

The goal of this work is to develop a heat exchanger fouling model based on the analysis of historical data and to explore its application to monitoring, forecasting, and optimization of process and maintenance activities. For such, the following aspects are explicitly addressed:

- Selection of key performance indicators (KPIs) for heat exchanger fouling monitoring;
- Validation of KPIs for the past events and setting the associated performance limits;
- Identification of regions of increased fouling rates by correlating KPIs with process conditions;
- Forecasting heat exchanger performance (one month ahead)

The hybrid approach developed in the present work consists of a mechanistic-based feature extraction stage that allows the subsequent utilization of off-the-shelf machine learning methods to forecast fouling in a gas-to-gas heat exchanger. The remainder of the paper is organized as follows: Section 2 introduces several mechanistic-based fouling KPIs and the machine learning methods used to predict them. In Section 3, an industrial case study from The Dow Chemical Company is described. This case study provides a challenging scenario for testing the proposed hybrid methodology under real operating conditions. Section 4 presents the results and discussion of the method, including the mechanistic feature generation and forecasting models construction. Lastly, Section 5 summarizes the conclusions of this work and offers some perspectives for future work.

2. Hybrid approach for monitoring and forecasting fouling

An overview of the several components developed to build the proposed hybrid approach is provided in this section, including the associated methodological details. Since we aim at exploring available data-driven predictive methods as part of the methodology, a first challenge is to derive the target response to be predicted. As the fouling phenomena is not directly observable or measured, a surrogate feature or fouling KPI must first be extracted or derived. For such, mechanistic knowledge is leveraged to obtain a KPI that is related to heat transfer phenomena, and that is sensitive enough to map the evolution of fouling over time. Given its mechanistic origin, the KPI will be also interpretable to subject matter experts and plant operators, facilitating the integration of the hybrid approach \[37, 38\] in daily activities of process management. This is the topic of Section 2.1. With the fouling KPI available, the next step is to integrate it with process operation data, to derive a forecasting model capable of predicting fouling behavior over time. Among other applications, this predictive model can then be used to anticipate hotspot events some time before they happen, allowing for timely preventive action to take place. Data-driven model building is covered in Section 2.2.

2.1. Mechanistic feature generation: fouling KPI

The main challenge for monitoring and modeling fouling is that it is not usually measured online directly. Therefore, a surrogate measurement or fouling KPI is necessary to indirectly
assess the state of fouling and to incorporate such information to derive the predictive model. In the proposed hybrid approach, the KPI is knowledge-based, i.e., derived from the mechanistic understanding of the system, expressed through mathematical equations involving the relevant dimensionless numbers, and mainly regarding the description of heat transfer phenomena [39].

Below, we present a rundown of the possible KPIs to indirectly measure fouling evolution. The heat duty ($\dot{Q}$) is a measure of how much energy is transferred from a stream,

$$\dot{Q} = \dot{m} C_P \Delta T,$$

where, $\dot{m}$ is the mass flow, $C_P$ is the heat capacity at constant pressure and $\Delta T$ is the temperature difference between the inlet and outlet streams. $C_P$ is computed as the average of $C_P$s at inlet temperature and outlet temperature,

$$C_P = \frac{1}{N} \left( \sum_{i} x_i C_P^{i,\text{in}} + \sum_{i} x_i C_P^{i,\text{out}} \right),$$

for each $i^{th}$ chemical component in the stream. $C_P^i$ is computed using the Shomate equation,

$$C_P^\omega = A + B T + C T^2 + D T^3 + E T^{-2},$$

where $T$ is temperature and the coefficients are specific for each component and retrieved from the NIST database. With the heat duty estimated, it is straightforward to compute the overall heat transfer coefficient ($U$) as follows,

$$U = \frac{\dot{Q}}{A_t \Delta T_{lm}},$$

where $A_t$ is the cross section area of the tube,

$$A_t = n_t \pi D L,$$

and $\Delta T_{lm}$ is the logarithmic mean temperature difference, defined by

$$\Delta T_{lm} = \frac{\Delta T_1 - \Delta T_2}{\log \left( \frac{\Delta T_1}{\Delta T_2} \right)}.$$

with $\Delta T_1$ being the temperature difference on the right side of the heat exchanger, and $\Delta T_2$ the temperature difference on the left side of the heat exchanger.

Equations (1) and (4) can be manipulated to get a ratio between the temperature difference and the logarithmic mean temperature difference,

$$\frac{\Delta T}{\Delta T_{lm}} = \frac{U A_t}{\dot{m} C_P}.$$

This dimensionless ratio represents the proportion between stream enthalpy and heat transfer driving force. In this work, we have also included other dimensionless numbers that could be affected by fouling and therefore be used as surrogate quantities for monitoring its progress. They
have also the desirable characteristic of being interpretable and familiar to most process engineers: the Reynolds number, the Prandtl number, the Nusselt number and the Biot number. The Reynolds number \( (Re) \) is the ratio of inertial forces to viscous forces within a fluid that is subjected to relative internal movement due to different fluid velocities, and is defined as

\[
Re = \frac{\rho u D}{\mu},
\]

where, \( \rho \) is the stream density, \( u \) is the fluid velocity and \( \mu \) is the stream dynamic viscosity. The Prandtl number is defined as the ratio of momentum diffusivity to thermal diffusivity. It can be computed as,

\[
Pr = \frac{C_P \mu}{k},
\]

where, \( k \) is the thermal conductivity. The Nusselt number is defined as the ratio of convective to conductive heat transfer across a boundary. There are a number of ways of computing it, depending on the specific assumptions taken. In this work, we used the Dittus-Boelter equation \[39],

\[
Nu = 0.023 \cdot Re^{0.4} \cdot Pr^{0.3}
\]

This expression is based on the assumption of smooth pipes, which in industrial applications may not be fully valid. However, since we are looking for a fouling surrogate or KPI, our major focus is on capturing correctly the trend instead of predicting the actual nominal values. Finally, the Biot number is defined as the ratio of the thermal resistances inside of a body and at the surface of a body. It is calculated by,

\[
Bi = \frac{hL}{k},
\]

where, \( h \) is the film coefficient, and \( L \) is the characteristic length.

These dimensionless parameters form the prospective set of fouling KPIs. Their relative merits as inferential indicators of fouling are discussed in Section [4.1]

2.2. Data-driven forecasting model

The classical Multiple Linear Regression (MLR) approach was included in this study, given the characteristics of the dataset (see Section [3] and the need to interpret the model. MLR has a well-established body of knowledge and has been widely used together with a variety of variable selection methods to mitigate the effects of sparsity (the existence of many predictors that are not relevant) and collinearity (the occurrence of correlated predictors that disturb the estimation process based on ordinary least squares, OLS) \[40,41,42,43,44,45,46\]. The MLR method also presents optimal estimation properties, provided that there are no collinearity issues and the model structure is well postulated; this is secured by the Gauss-Markov theorem that proves that under certain assumptions, MLR produces unbiased estimates that have the smallest variance among all possible linear estimators. The basic MLR model structure is,
\[ Y = \beta_0 + \sum_{j=1}^{p} \beta_j X_j + \epsilon \]  

(12)

where \( p \) is the number of predictors, and \( \beta_j \) is the regression coefficient associated with predictor \( X_j \). MLR assumes that only the response variable carries a sizeable error, which is additive, independent, and has constant variance. The regression coefficients are found by least squares fitting [47],

\[
\hat{\beta}_{MLR} = \text{minimize}_{\beta=[\beta_0...\beta_p]^T} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2
\]

(13)

s.t.  \( y_i = \beta_0 + \sum_{j=1}^{p} \beta_j \times x_{i,j} \),  \( i = 1 : n \)

(14)

where \( \hat{\beta}_{MLR} \) is a vector containing the regression coefficients, \( n \) is the number of observations, \( y_i \) is the \( i^{th} \) observed response value, \( \hat{y}_i \) is the respective estimated response using the MLR model, and \( x_{i,j} \) the \( i^{th} \) observation of variable \( X_j \). As mentioned above, MLR faces problems when predictors present high levels of collinearity, because the estimation of the regression coefficients becomes unstable (high variance). To tackle this issue, we will select the most relevant predictors informed by the knowledge-based variable selection, as interpretation and transparency are priority aspects in our analysis.

3. **Industrial case study: Ethylene Oxide (EO) production**

In this section, we present an overview of the ethylene oxide production process, as it is presented in [48]. Then, some specifics on the dataset are referred together with a clarification of the challenges and goals established for this case study.

3.1. **Overview of the Ethylene Oxide process**

Ethylene Oxide is industrially produced by the direct oxidation of ethylene with air or oxygen. The oxygen-based process is the most used and is the technology used in the selected case study. A scheme of the section of the process in study is shown in Figure 1. The reaction happens in a continuous multitubular reactor, where the tubes filled with catalysts are surrounded by a coolant that can be water or, more commonly, a high-boiling hydrocarbon mixture. The coolant removes the reaction heat and allows for temperature control. The heat removal system must cope with the rising reactor temperatures necessary to deal with catalyst aging. When the coolant temperature reaches a maximum, the process must stop and the catalyst changed. After leaving the reactor, the gas outlet is used to preheat the reactor inlet gases in a shell & tubes heat exchanger. In our investigation, we are focusing on the fouling that takes place in the aforementioned heat exchanger.

The heat exchanger is a typical gas-to-gas shell & tubes that operates continuously. During operation, fouling is constantly developing on the tube side, forming a hard polymer layer that becomes visible when the equipment is open for maintenance. The mechanism that leads to this
polymer formation is unknown, but its effects are felt in operation in the form of pressure drop and loss of heat transfer capacity. Furthermore, the fouled polymer forms dead zones where hotspots may develop. The hotspots are potentially dangerous as EO at very high temperatures decomposes, which is a major safety concern for the plant.

3.2. Dataset description

The dataset used is composed of plant measurements of gas flows, various temperatures, pressure sensors, and also the dimensions of the heat exchanger. The data goes back three years encompassing a full catalyst cycle, at an hourly rate.

This dataset was enriched with the information provided by the plant operators, namely regarding the hotspots that were historically detected. Operators are able to detect potential hotspots and take measures to avoid decomposition. The detection method and the measurements are proprietary. We call these labels hotspot events.

4. Results

In this section, we present results for various aspects of the proposed hybrid framework for monitoring and forecasting fouling, applied to the EO heat exchanger. We start with the computation of the prospective set of KPIs that may serve as surrogates for the evolving fouling phenomenon. We will present the two best KPIs identified in our study, namely: the Reynolds number (Re) and the ratio $\frac{NTU}{T_{m}}$ on the tube side. Then, these KPIs are correlated with process conditions to find out what operating conditions cause more fouling. Furthermore, the forecasting model results are shown, as well as the workflow to achieve the best performing model (filtering to improve...
the signal-to-noise ratio and variable selection). The model’s performance is estimated for a one-month ahead forecast as well as for long-term forecasting. Finally, a discussion is provided on how the fouling forecasting model can be effectively used for process/maintenance optimization and decision-making.

4.1. Key Performance Indicators (KPI)

Fouling cannot usually be measured directly, and can only be inferred through a combination of readily available data measurements and engineering *a priori* knowledge. In this work, the fouling surrogate is given the name of KPI. All the KPIs presented in Section 2 were applied to our case study (see Appendix A), and below we present the results for the two that yielded the most consistent trends, according to the history of events: the Reynolds number (Re) and the ratio $\frac{\Delta T}{\Delta T_{lm}}$ on the tube side.

4.1.1. KPI 1: Reynolds number

The Reynolds number is one of the KPIs that represented fouling well enough to match the events recorded in the process. Figure 2 shows the evolution of the Reynolds number over time. It can be noticed that the highest value of the Reynolds Number (>80,000) is recorded at the beginning of the time series, right after heat exchanger cleaning. As time progresses, this KPI starts to decrease until it reaches 72,000. This precedes the first hotspot event in the 8th month. Through the adjustment of process conditions, the KPI achieves a higher value of over 75,000 which remains stable for almost a year until in the 18th month another downward trend becomes visible, reaching the threshold again by the end of the 24th month. Again, this coincides with a new hotspot event on the same month. From the 25th month on, although multiple adjustments were made to the process, hotspot events kept occurring (30th month, 35th month, and 37th month) due to the high amount of fouling already present on the heat exchanger’s tubes. This condition is also paralleled by a low Reynolds number.

Figure 2: Evolution of the Reynolds number in the heat exchanger since its cleanup (blue); the red dashed line represents the threshold from where events are likely to start happening; the purple dots are recorded hotspot events in the heat exchanger.
Note that this fouling KPI has a natural downward trend. Any monotonic trend with the progress of fouling would be acceptable, as long as it parallels the progress of fouling and is consistent with the known events. This is clearly observed for the Reynolds number. The value of the Reynolds KPI is not just a function of how much fouling is clogging the tubes of the heat exchanger, but also of the cycle gas flow and the fluid properties that result from temperature and pressure swings in the process. The knowledge encoded in this KPI is substantial and allows for setting a fixed hotspot event threshold which is a valuable information for process operation.

4.1.2. KPI 2: ratio $\frac{\Delta T}{\Delta T_{lm}}$

As for the previous KPI, the ratio $\frac{\Delta T}{\Delta T_{lm}}$ on the tube side of the heat exchanger also represented fouling well enough to match the events recorded in the process, with the advantage of being easier to compute and requiring less information. Figure 3 shows the evolution of the ratio $\frac{\Delta T}{\Delta T_{lm}}$. Contrary to the Reynolds number, this KPI increases with time and remains fairly monotonic, albeit noisy, for similar process conditions. Its lowest value of 3.70 is recorded at the beginning of the time series, right after heat exchanger cleaning. As time progresses, this KPI increases until it reaches 3.87 in the 8th month, which precedes the first hotspot event. Through the adjustment of process conditions, the KPI achieves a lower value around 3.75. In the 18th month, the KPI starts to increase, reaching the threshold again by the end of the 24th month. As expected, that coincides with a new hotspot event. From the 25th month on, although multiple adjustments were made to the process events kept coming up due to the high amount of fouling already present on the heat exchanger’s tubes.

![Figure 3: Evolution of the ratio $\frac{\Delta T}{\Delta T_{lm}}$ on the tube side in the heat exchanger since its cleanup (blue); the red dashed line represents the threshold from where events start to happen; the purple dots are recorded events in the heat exchanger.](image-url)

Overall, the two KPIs show the expected behavior over time and are well-aligned with the known landmarks where hotspot events took place. Therefore, they are well suited to monitor fouling and the effect it has on the heat exchanger’s operation. From prior knowledge, we know that the Reynolds number reflects the flow conditions in the heat exchanger while the ratio $\frac{\Delta T}{\Delta T_{lm}}$ gives us information about heat transfer phenomena. Thus, both KPIs were selected and used.
in tandem to monitor the evolution of fouling, although historically, they seem to convey similar information.

4.2. Impact of process conditions on KPIs

At this stage, the identification of two relevant KPIs already allows to monitor fouling over time and take reactive actions to sudden changes in their behavior that make the indicators closer to the operational limits. However, we aim to develop a proactive approach to fouling management, for which a predictive methodology is required. This will be covered in Section 4.3. Before constructing a model, let us first relate fouling with process conditions. In other words, we want to better understand which process conditions lead to a higher or lower fouling rate, as a basis to establish prescriptive measures that operators can consider or follow. This information is valuable to adjust process conditions in order to avoid hotspot events even when it is not possible to stop the process and clean up the heat exchanger. By avoiding the events, process safety is safeguarded.

Figure 4 shows the KPI (ratio $\frac{\Delta T}{\Delta T_{lm}}$) colored in four distinct parts. We first look at the period from month 1 to 12, colored in yellow and green. Here, the goal is to understand what causes the KPI to "reset" to a lower value after the event. By understanding the operator response and its effects, it can be replicated in the future to mitigate operational issues with the heat exchanger. Then, we compare two periods in the period from month 19 to 24, colored in red and sea green. The first period corresponds to a stable period for fouling evolution, where the fouling rate was null. In the second period, fouling starts to increase until it causes an event at the end of the period.

Figure 4: Evolution of the ratio $\frac{\Delta T}{\Delta T_{lm}}$ on the tube side in the heat exchanger since heat exchanger cleanup. The four colored periods correspond to different regions of interest to study how process conditions affect the fouling rate.

In the first period, the cause for the KPI to lower was an increase in the cycle gas flow. In Figure 5a, we see that the flow was steadily decreasing due to fouling, but in the 8th month it was increased by plant operators, which increased heat transfer efficiency. In Figure 5b, we are able to relate that an increase in EO production leads to an increase in fouling.

From this analysis, we can conclude that the flow conditions and the production rate have a major influence on fouling behavior. These findings are well in line with prior knowledge of heat transfer in heat exchangers. From [1], we know that larger flows favor heat transfer. On the
production rate side, we can only infer what is causing fouling. The finding that higher production rates cause higher fouling rates is consistent with a hypothesis that EO may act as a reactant for the fouling formation reaction. However, this can only be confirmed by analyzing fouling deposits once the heat exchanger is opened.

4.3. Forecasting model

The KPIs used for monitoring can be also used as target responses to build a forecasting model. Below, we present the main steps to develop a forecasting model capable of aiding in decision-making regarding how to operate systems undergoing fouling. We start by filtering the KPI signal in order to make it smoother and cleaned up from spurious stochastic events and process/measurement noise, i.e., to increase the signal-to-noise ratio of the KPIs. Then the knowledge acquired during the stage reported in the previous section was applied to conduct variable selection, and a model to predict the filtered KPIs, one month ahead, was developed. Long-term forecasting was also covered as a possible application. The one-month ahead forecasting window was decided together with the process subject matter experts and in general, it depends on the process requirements and business needs. The main stages of this workflow are briefly described in the following subsections.

4.3.1. Filtering

The use of a filter to smooth out the signal removes high-frequency dynamics and noise, making the analysis of short-time windows not informative or adequate. In this work, we applied a moving average filter to smooth the KPI signal, a 30-day moving average (MA), as shown in Figure 6.

4.3.2. Variable selection

As referred to in Section 1, the high number of features encountered in the chemical process, the industry can represent a problem when building a robust and well-performing data-driven model, especially when only a few of them are relevant to predict the target response (sparsity) or, when there are strong redundancies (collinearity), even though this last aspect does not pose significant challenges to some class of methods, such as latent variable or penalized regression
methods. It is therefore imperative to remove noisy and irrelevant features before model building. Although, there are a number of data-driven techniques for variable selection (filter, wrapping, or embedded methods), in this work we will take advantage of the engineering knowledge available and in particular, the process insights addressed in section 4.2.

As referred in Section 4.2, cycle gas flow and EO production rate are well correlated with the KPIs. The input tubes temperature is also included because of its dependence upon the catalyst cycle. Furthermore, as the goal is to derive a forecasting model, the model should be composed of terms that reflect the evolution of the dynamic of the fouling phenomenon over time. As such, the introduction of lags is an important decision in feature selection. The correlation of the KPI with its two first lagged versions (a lag regards a one month period) was included in the model.

4.3.3. One-month ahead forecasting

The final forecasting model was obtained by applying MLR. It is a linear regression model with the following form,

$$KPI_{k+1\ month} = \beta_0 + \beta_1 \cdot KPI_k + \beta_2 \cdot KPI_{k-1\ month} + \beta_3 \cdot \text{Flow}_k + \beta_4 \cdot EO_{total,k} + \beta_5 \cdot T_{in,k} + \epsilon,$$  

(15)

where, KPI is the fouling surrogate measurement, Flow is the cycle gas flow, $EO_{total}$ is the cumulative production of ethylene oxide and $T_{in}$ is the input tubes temperature. Both of the KPIs were tested for this model. In Figure 7, we show the results for the Reynolds number, as it yielded the best forecasting performance among the two. The model’s parameters were identified by least squares using data from the first twelve months as the training set.

The model performance was tested on data from the second twelve months, always predicting the KPI one month ahead. The performance metrics computed are $R^2$, RMSE and P-Bias. The $R^2_{test}$ is 0.698, the RMSE is 978 and the P-Bias is $-0.14\%$. Overall, the forecasting prediction is able to follow the general downward trend of the true KPI, as shown in Figure 7 even though some short-term oscillatory behavior was not captured.
4.3.4. Long-term forecasting

For some business decisions, a one-month ahead prediction may not be long enough. Therefore, we tested the ability of the model to go beyond the one-month horizon. One should note, that as the forecasting was built to predict over a specific time horizon, going beyond that is bound to produce poorer estimates, certainly worse than an $R^2$ of about 0.7. However, it was still possible to get a general trend from the model as shown below.

Considering the present forecasting scenario and equation (15), one is left with the following problem: how to use the predictor for long-term forecasting if elements of the predictor (cycle gas flow, cumulative production of EO, and input tubes temperature) are unknown in the future?. The cumulative production of EO is understood to be a process decision guided by market or industrial constraints. With this reasoning, in this work, we test how the model behaves assuming a constant production rate. The considered rates, after scaling take the values of -0.6, 0.9, and 2.4.

For the other two predictor variables, we leverage relations between them the KPI and the EO production rate. Cycle gas flow can be related to the KPI by the following relationship,

\[ \text{Flow}_k = \beta_6 + \beta_7 \cdot \text{KPI}_k + \epsilon, \]  

(16)

while the input tubes temperature is related to the cumulative production of EO by,

\[ T_{in,k} = \beta_8 + \beta_9 \cdot \text{EO}_{total,k} + \epsilon. \]  

(17)

The results of applying these approximations can be seen in Figure 8.

In order to test the long-term forecasting approach, forecasts were made from the 13th month to the 43th month (one for each production rate), and superimposed on top of the true KPI (see Figure 9). It is worth noting that the furthest we look into the future, the more uncertain the forecast will be. Nevertheless, we can observe that the forecast follows the KPIs general trend quite accurately. The influence of changing the production rate is obvious as well. At 2.4 production rate, the KPI threshold is reached on the 33th month. At 0.9 production rate, it is reached on the 37th month, four months later. And at -0.6 production rate, it is reached on the 43th month, adding
six months to the latter production rate. This analysis shows the potential of using the model to aid the decision of adjusting the production to prevent events that compromise process safety and the plant’s ability to keep producing. The nature of the decision is both interesting for process optimization and maintenance planning, as further discussed next.

4.4. Suggestions for operation optimization and maintenance planning

The period of operation of the heat exchanger is an important driver for improving the EO process reliability, safety, and profitability. As stated before, the hotspot events or low cycle gas events lead to plant "trips" or short periods without production. Their periods of halted production albeit short can be translated into major disturbances and thousands or millions of dollars in losses in the commodity business. Their anticipation and proper management are therefore of utmost importance and is one application envisioned for the forecasting models developed. In fact, two scenarios are considered by the decision-maker:
1. Reduce production rate enough to lower fouling rate, thus extending the heat exchanger's "lifetime" until the scheduled plant maintenance;
2. Stop the process, clean the heat exchanger, and resume production.

Therefore, one should perform a risk-based analysis and optimization to decide which of the two options offers the best perspective at a given time. The developed forecasting model is instrumental for conducting this analysis at two distinct levels: first, it allows plant operators to know that an event may be close and to trigger timely actions; secondly, the model can be used to simulate scenarios where the production rate is reduced, and conduct an optimization study that eventually will lead to a decision of reducing the production or stopping and cleaning the heat exchanger.

5. Conclusions

Fouling in heat exchangers can lead to unplanned outages and temperature excursion/decomposition events, which constitute major safety risks for EO Operations. The heat exchanger considered in this study operates under conditions that make it susceptible to fouling; some potential hotspot events were reported during the last catalyst cycle. To improve the safety and reliability of the operation, we lay down a hybrid approach to monitor and forecast fouling in the heat exchanger.

Knowledge-based feature generation allowed the monitoring the evolution of fouling over time and enables the use of off-the-shelf data-driven forecasting methods. Among the KPIs tested, two were selected to act as fouling surrogates. In the end, we advise the use in tandem of the Reynolds number and the ratio $\frac{\Delta T}{\Delta T_{lm}}$ as KPIs because the first one reflects the flow conditions in the heat exchanger while the ratio provides information on the heat transfer side. The developed forecasting model was capable to predict one-month ahead with a testing accuracy of $R^2 = 0.7$.

Furthermore, we showed that long-term forecasting is also possible with this model, always with the caveat that the furthest we look into the future, the more uncertainty the forecast carries. Nevertheless, the model can still be applied for process optimization and maintenance scheduling.

In the future, we believe it will be opportune to further develop the optimization framework by including uncertainty quantification on the inputs and in the model. In this way, more robust decisions can be made, that take into account process safety, reliability, and profitability.

Data Availability and Reproducibility Statement

The Supplementary Material provides access to the numerical data pertaining to Figures 2 to 9 and A.10 in a compressed file format (.zip). By utilizing these data, it is possible to replicate the corresponding figures and the models, thereby ensuring the reproducibility of the study’s results.

Acknowledgements

The authors acknowledge support from the Fundação para a Ciência e Tecnologia (FCT, Portugal), project reference UIDB/00102/2020.
Appendix A. Other KPIs

In this section, the KPIs described in Section 1 but not used in the case study are showed for the reader to be able to compare them with the two selected KPIs.

![Figure A.10: Evolution of a) the heat transferred; b) the overall heat transfer coefficient; c) the Nusselt number; d) the Biot number; in the heat exchanger since its cleanup; the red dashed line represents the threshold from where events are likely to start happening; the purple dots are recorded hotspot events in the heat exchanger.](image)

References


