12th IFAC International Symposium on
Advanced Control of Chemical Processes
July 14-17, 2024. Toronto, Canada

A Hybrid Modeling Approach to Predict Pollutant Scrubber Remaining Useful Life

Juan M. Venegas[A], Zhenyu Wang[B], Andrew Athon[C], Ivan Castillo[D*]

[A] – Dow Chemical Company, Silicones Process R&D, Midland, MI, USA 48640
[B] - Dow Chemical Company, Energy and Climate Technology Center, Lake Jackson, TX, USA 77566
[C] - Dow Chemical Company, LHC Technology Center, Oyster Creek, TX, USA 77541
[D] - Dow Chemical Company, TES Chemometrics, AI, & Statistics, Lake Jackson, TX, USA 77566
** - Corresponding author, phone: 989-238-7164, e-mail: castillo2@dow.com*

Abstract: Chemical plants require reliable systems for pollutant abatement. These processes often operate under cyclic abatement and regeneration cycles over extended periods of time. Throughout this period, the abatement systems experience a multitude of phenomena that may degrade performance in a fashion that is challenging to predict by first-principle models. These complex phenomena offer an opportunity to leverage data-driven models. To improve their predictive ability, data driven models can be complemented with physics-based information that constrains modeling results. In this contribution, we describe a hybrid modeling approach where physics-derived features are developed to enable data-driven models to effectively predict the performance of real pollutant abatement systems in the Dow Chemical Company.

Keywords: hybrid modeling, pollution abatement, predictive modeling, batch analysis, Partial Least Squares, Remaining Useful Life.

1. INTRODUCTION

Chemical plant operations frequently feature impurity abatement or scrubbing unit operations. Chemical neutralization, adsorption, catalytic transformation, selective oxidation, amongst others (Taoufik et al., 2022). These processes are frequently needed to maintain downstream performance, meet safety requirements, or ensure regulatory compliance. These abatement systems require high reliability to avoid unplanned plant shutdowns (P.K. et al., 2021), and high performance to remove pollutants that are often in the sub-percent range.

A particular challenge in scrubbing operations is their often-complex performance degradation over cyclical operation. Abatement and regeneration cycles enable long-term use of these units, but impart significant stress (*e.g,* thermal, chemical, mechanical) on the process. In turn, degraded pollutant removal performance is frequently observed with extended scrubber age. The complexity of this degradation stems from the various factors that may influence it. For example, sorbent pores may contain chemical buildup, or a thermal treatment may sinter catalytic metals to large, less reactive particles. Thus, degradation may be challenging to predict based on physical phenomena (Abbasi et al., 2022; Sansana et al., 2023).

Modeling and predicting such an abatement process could be accomplished in many ways. Of course, if we possessed all physical properties and approximated all dynamics related to abatement and regeneration cycles, the system could be fully described from first principles. This has been achieved for idealized systems at the lab scale (Haghpanah et al., 2013; Jacobs et al., 2021). This approach requires in-depth knowledge of the physical process taking place, a quantifiable amount of pollutant to remove, and tools to quantify the impact of transport phenomena on a process. If successfully leveraged, physics-based modeling can provide additional information on a process, such as variables that are not directly measured. These could be the rate of saturation of a liquid by a pollutant or the coverage of a catalyst particle by adsorbates, for example. With these first-principle models, one could improve process operations to optimize a given performance indicator. This rigor, however, may lead to modeling challenges when process conditions change unexpectedly, or a process experiences unquantified changes (*e.g.* solids attrition due to temperature cycling.). Thus, this modeling approach may pose difficulties for plant operations that often operate in environments with limited information or with unexpected variations.

To mitigate these challenges, empirical modeling approaches can predict the performance of systems outside of ideal environments (Bogojeski et al., 2021; Bradley et al., 2022). For example, time-dependent catalytic activity decay can be estimated from observed temperature history, pressure drop, and change in process performance (Bogojeski et al., 2021). Thus, an estimate for remaining useful life of the catalyst can affect maintenance activities. The accuracy of these data-driven models relies heavily on data quality. However, not all the variables required to build an accurate data-driven model are always measured in an industrial plant, *e.g.* quantity of pollutant to remove in this use case. Thus, employing first-principle models to estimate such variables is needed to seal the gap. In this contribution, we explore a hybrid modeling approach to exploit the advantages of first-principle and data-driven models to predict long-term performance and remaining useful life of a pollutant abatement process in operation at Dow. Beyond the challenges with degradation already described, this plant possesses multiple scrubbing units that operate in parallel. Thus, the performance of a single unit is correlated with its parallel partner. We use a physics-based approach to supplement available data features from plant data. In turn, this supplemented data set is then explored to identify the key process metrics needed to predict the long-term evolution of scrubber capacity.

2. PROPOSED MODELING APPROACH

In the development of a modeling approach to predict the abatement capacity of impurities in scrubbing units, there were several key factors that had to be addressed: a) the features available to quantify capacity; b) the decay in capacity as a function of process history and c) the parallel operation of scrubbers with potentially different process histories. In this section, we describe the series of steps necessary to address these factors and their subsequent implementation in a quantitative prediction tool. This tool combines a physics-based feature engineering step followed by empirical analysis, as summarized in Figure 1.

2.1 Case Study Description

Scrubbing operations often have strict requirements on downstream impurity content. This may be motivated by process reliability or quality specifications or environmental concerns. As such, scrubbers may operate individually, in series, or in parallel to meet impurity capture needs based on processing volumes needed or ultimate impurity targets. In this particular case study, a plant utilizing parallel scrubbers was selected. To protect Dow's confidential information and trade secrets, all process information, model specifics, and results have been masked.

Figure 2 shows a high-level overview of the process under consideration. Three scrubber units filled with pollutant capture media are utilized, with two operating in parallel at any given point. The third unit is undergoing a regeneration cycle (*e.g.* steam treatment, solvent wash, thermal heating).

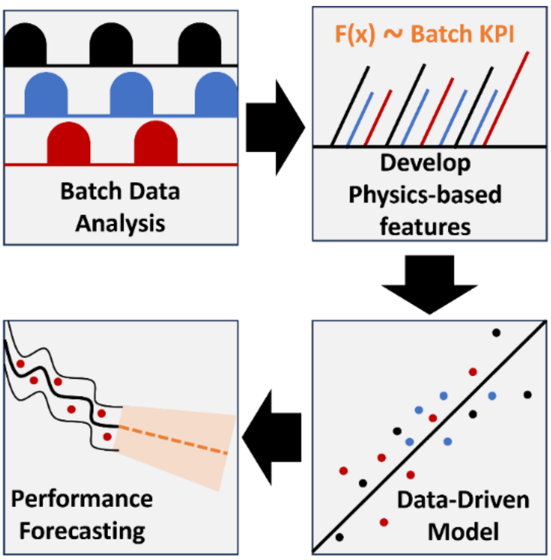

Figure 1 – Graphical overview of model development steps.

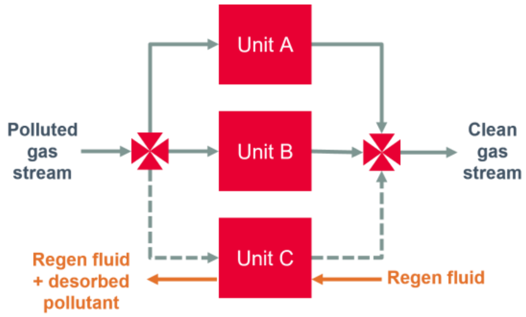

Figure 2 – Overview of case study process. Dashed arrows denote scrubber undergoing regeneration.

The plant operations require that a regenerated scrubber is always available to put online upon reaching scrubber capacity of one of the two operational units. Thus, if the capacity of a scrubber is too low, it may be saturated prior to the completion of the regeneration cycle. This plant provides sensor readings of the upstream fluid flow rate, temperature, and pressure. Downstream of the scrubbers we have pressure, temperature and impurity readings. Within each scrubber, we have inlet/outlet temperature and inlet pressure readings available. The operation-regeneration cycles are controlled by the process control system, and thus we also have data available on the process state of each unit. Each 'cycle' of a scrubber was started upon completion of the regeneration phase and was ended upon scrubber saturation and pollutant breakthrough. Notably, the flow rate to each on-line scrubber is not available, with the feed rate determined by the pressure drop across the scrubber media. Additionally, no composition analyzer is available upstream of the scrubbers, thus impurity levels are not quantified. In this type of operation, the macro-scale dynamics such as flow rate, pressure drop, and temperature evolution may be readily described *via* first principles. On the other hand, the details of the impurity removal process itself, as well as the capacity during operation, is a complex process that may not be modeled as readily *via* first principles (Hanif et al., 2020; Jacobs et al., 2021; Taoufik et al., 2022).

2.2 Physics-Informed Feature Development

To quantify the capacity of the three scrubbers during operation, estimation of the impurity feed composition and subsequent feed rate into each scrubber in operation is necessary. This was done *via* a physics-based approach. First, we will describe the quantization of the impurity. In this particular plant, the gaseous feed stream, m_{gas}, is expected to be at its saturation point with the impurity, based on upstream processes. As such, the impurity composition, x_i, is determined by vapor-liquid equilibrium and primarily determined by temperature, T. Thus, using the Antoine Equation, we can estimate the vapor pressure of the impurity (P_i) in the process stream using known Antoine parameters, A, B, C (1-2). Next, the vapor pressure is normalized by the inlet pressure (P_{gas}) into the scrubbers to estimate the molar fraction of impurity in the feed stream. Subsequent unit conversions coupled with the known process feed rate provide the total impurity feed rate into the scrubbers (5). Until scrubber saturation, all pollutant is expected to be scrubbed. Thus, (5) is also equivalent to the scrubber abatement rate. In (1-5), m, n, and MW denote mass flow, molar flow, and molecular weight, respectively.

$$\log_{10} P_i = A - \frac{B}{T + C} \quad (1)$$

$$x_i = \frac{P_i}{P_{gas}} \quad (2)$$

$$N_{gas} = \frac{m_{gas}}{MW_{gas}} \quad (3)$$

$$N_i = \frac{N_{gas} * x_i}{1 - x_i} \quad (4)$$

$$m_i = \frac{N_i * MW_i}{1000} \quad (5)$$

With the impurity feed rate, m_i, known, the next feature to be estimated is the effective flow rate to each of the scrubbers during parallel operation. As mentioned previously, the flow rate to each is not metered or controlled and is determined by pressure drop across the pollutant capture media. As each scrubber unit has an independent process history, we expect their total capacity and pressure drop to be different. At the same time, a scrubber's real-time performance is inherently correlated to the specific scrubber operating in parallel, as each pair will feature different flow distributions. In this implementation, we aim to relate the pair-wise performance of the scrubbers by using a physics-based approach and draw an analogy between hydraulic flow (Darcy's law) and electrical circuit resistance, R, (Ohm's law) frameworks:

$$I = \frac{V}{R} \sim \Delta P \, \frac{k}{\mu L} = Q$$
$$\rightarrow \frac{\Delta P}{R_{flow}} = Q \quad (6)$$

This relationship is then used in parallel flow networks, such as the one involved in the scrubber plant, to relate overall pressure drop, ΔP and feed flow rate, Q, to the individual flow (Q_i) across each scrubber. We note that the pressure drop across both scrubbers should be the same in a parallel flow network.

$$\frac{1}{R_{flow}} = \frac{1}{R_1} + \frac{1}{R_2} \quad (7)$$

$$Q_i = \frac{\Delta P}{R_i} \quad (8)$$

From (8), we obtain the individual flow across a scrubber, and can convert this to mass flow to use in (3-5) to calculate impurity mass flow rate. This framework is a first approximation, as gas flow features compression that makes this analogy incomplete. A more elaborate model would require implementation of equations such as the Ergun equation. This situation, however, would include a great deal of uncertainty on individual variables. These variables require information on scrubber material properties, which are likely affected by the scrubbing cycle itself due to degradation, incomplete scrubber use, and fouling. As such, we will continue with the resistance analogy, which reduces the number of parameters we would need to estimate in the system.

With a known impurity feed rate into the process and a framework to split the flow between the two active scrubbers, the final aspect to define the system was quantifying the relative resistance parameters (R_1, R_2) used in (7-8). *A priori* determination of these parameters is once again complex and may require process knowledge beyond feasibility. As such, we propose a simple relationship based on relative scrubber media age:

$$R_{older} = \alpha * R_{newer} \quad (9)$$

$$where \ \alpha = -0.5 * \frac{Age_{newer}}{Age_{older}} + 1.5 \quad (10)$$

The slope and constant of (10) were selected such that at comparable scrubber ages, the flow would be split nearly evenly between them. As the ratio approximates zero (*i.e.* maximum age gap between scrubbers), the newer scrubber receives 60% of total flow. This 60/40 split is an estimate based on process experience and unit manufacturer recommendations. Larger flow splits would likely be noted as excessive pressure drops in the system. We note that scrubbing processes that operate under different conditions such as liquid-phase extraction, filters, or condensers may experience different pressure drop-flow relationships and will require appropriate consideration. Under scenarios where the physical phenomena are difficult to reduce to a model, a physics-inspired approach may not be possible and purely machine learning approaches may be necessary.

With the relative resistance and the physics-based impurity flow rate estimated, we now have a fully defined system to predict the impurity mass flowrate into each scrubber during the operation cycle. The cycle time for each scrubber was quantified by monitoring the total time a scrubber was online. This time was then multiplied by the estimated impurity mass flow rate to calculate the total impurity abated by the scrubber, assuming the scrubber's on-line phase ended at breakthrough (*e.g.* when a threshold level of impurity was detected in downstream processes). This calculation provided an estimated value for the scrubber's capacity during a particular cycle, allowing us to monitor capacity evolution as a function of process variables. Thus, our target metric of scrubber capacity was obtained leveraging physics-based approaches. We emphasize that these physics-based features can only be developed from already-completed cycles due to the process data required (e.g. cycle time, relative age of scrubbers) and thus provides limited direct predictive power. The estimated capacity, however, is used to enable the development of a data-driven forecasting model to predict the evolution of scrubber capacity over its lifetime.

2.3 Data-driven Model Development

As the process operating conditions change over time, we start with a predictive batch Partial Least Squares (PLS) (Kourti et al., 1995) model using SIMCA (Eriksson et al., 2013) to model the dynamics of pollutant removal. With the information that time-varying variables do not impact predictive performance, we then developed a simpler Ordinary Least Squares (OLS) model using Python's Stats Models package (Seabold & Perktold, 2010) PLS was selected due its capability of handling potential variable multicollinearity and retaining a level of explainability. OLS was used upon simplification to further enhance the explainability of the model, which is valuable to aid in model adoption in plant operations.

We used 85% of the available batches (n = 106) for model training and 15% (n = 20) was used for validation. The test set was taken as a random sample from the entire dataset and not used in any step of the training process. Model development involved creating batch identifiers for each scrubber and reprofiling the batches to cycle times rather than absolute process time. This reprofiling enabled the unfolding of process features within a batch for potentially improved modeling ability (Kourti et al., 1995). Thus, each batch contained time independent variables (*e.g.* scrubber age at start of online cycle, batch-average metrics) and unfolded batch-time dependent variables (*e.g.* gas flow rate at batch time *t*, pressure drop, temperature). Care was taken to not include variables that would encode information on estimated cycle capacity, such as cumulative pollutant uptake, to avoid biasing the predictive model. In turn, this would allow for future predictions to be carried out with no *a priori* knowledge of the pollutant feed rate into the scrubbers. The combination of unfolded batch-time dependent and time independent features leads to a large amount of features for each batch, potentially thousands of features. As such, analysis of variable of importance in projection (VIP) was done to simplify the model without compromising predictive ability (Bui et al., 2022; Lu et al., 2014).

With a trained predictive model available, the scrubber capacity is forecasted by assuming nominal conditions of the input variables based on historical data or scrubber age-dependent estimates (Figure 3). Age-independent data was taken as historical averages from the available data. Age-dependent data was generated by estimating their evolution as a function of each scrubber's age through linear regression. The age range to forecast used in the model is a user input based on historical scrubber media lifetimes and/or manufacturer recommendations.

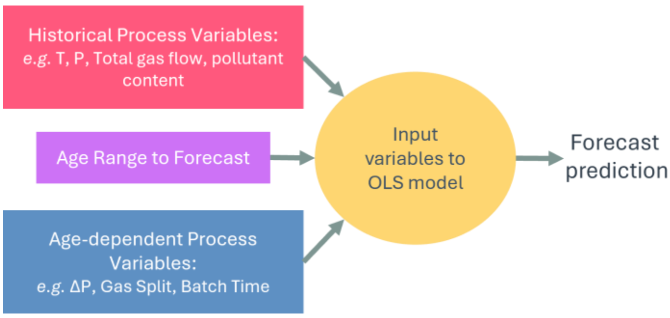

Figure 3 – Graphical representation of feature engineering for forecasting model.

3. CASE STUDY RESULTS

3.1 Physics-based feature generation – Scrubber capacity

Within the available dataset, the three scrubbers in the plant underwent one capture media changeover. These changeovers are staggered to avoid having low-performing scrubbers operating simultaneously. As such, the three scrubbers will show variation in their batch uptake, depending on their sequence and age.

Figure 4 shows representative batch uptakes for the three scrubbers. We note that one scrubber (Unit C) is operating near the manufacturer nominal capacity (Pollutant uptake = 1), which agrees with the fact that this unit had its scrubbing media changed closest to the beginning of the available dataset. The rated manufacturer capacity is not used anywhere in our model development, so its accurate prediction suggests that our physics-based features are able to represent realistic process features without specifically encoding them into the model.

One key finding during model development was variation in scrubber cycle finishing criteria. We find that in the first half of the dataset, most batch cycles correlated with a rapid rise in downstream pollutant concentration. This increase in pollutant is evidence of scrubber saturation and pollutant breakthrough. In contrast, the second half of the dataset shows few batch cycles terminating with a pollutant breakthrough. As such, we hypothesize that, during the second half of the dataset, plant operations were conservative and based their cycles on time rather than pollutant media saturation. This change in operation can be seen in Figure 5, where the distribution of batch cycle times in the second half of the data set are biased towards the same batch time, in contrast to the more evenly distributed cycle times in the first half of the dataset. Thus, training data was limited to the first half of the dataset to ensure that the scrubber capacity estimation was based on scrubber saturation rather than time.

3.2 Data-driven Model Development

As described in section 2.3, the PLS model to predict scrubber capacity during a cycle was first developed with both unfolded batch-time-dependent and independent features, amounting to 5938 variables per batch.

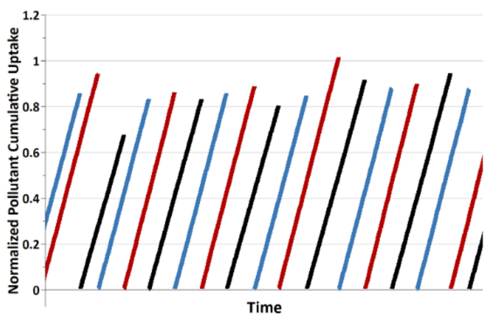

Figure 4 - Representative scrubber batch uptake from developed feature. Scrubbers are colored as Unit A (black), Unit B (blue), Unit C (red). Capacity normalized to manufacturer nominal capacity.

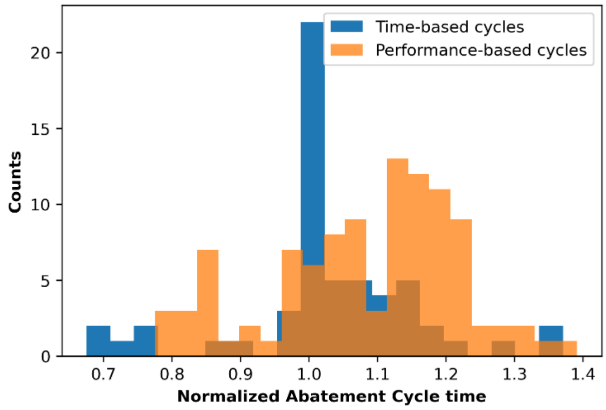

Figure 5 - Histogram of on-line cycle time in dataset. Time-based cycles do not indicate pollutant breakthrough at the end of a batch and were not used in training. Cycle time normalized to most frequent cycle time in data set.

This model uses 6 principal components with a $Q^2 = 0.71$ and $R^2 = 0.92$. Q^2 is an indication of model quality with new data not used during training (Eriksson et al., 2013) These metrics would be used as reference for model simplification. SIMCA provides variable importance metrics *via* its VIP score, shown in Figure 6 for all variables.

We see that all variables feature a sharp rise in importance relative to the bulk of its values. These higher importance variables corresponded to time-independent versions of each variable, such as their initial value and average. Thus, we expect that a predictive model without time-dependent features would be sufficient for predictions, while remaining relatively simple for eventual deployment. Further refinement of the PLS model *via* VIP analysis led to a model with 5 features and 3 principal components with a $R^2 = 0.91$ and $Q^2 = 0.90$. Throughout the model development, the scrubber media age was found to be the most significant feature to predict capacity, in line with expectation from process experience. Additional relevant features included average pollutant concentration, average pressure drop during a batch, feed mass flow rate, and batch temperature.

The relatively small subset of relevant features prompted us to assess their direct use in a model, rather than generating principal components in a PLS model. As such, an OLS model was developed using the direct feature values (11).

$$\begin{aligned} &Pollutant\ uptake\ at\ breakthrough\ (Y) \\ &= C_1[\ Scrubber\ Age] \\ &+ C_2[Pollutant\ Concentration] + C_3[Temp] \\ &+ C_4[\Delta P/Feed_{total}] \end{aligned} \tag{11}$$

We combined two correlated variables: ΔP and $Feed_{total}$ by taking their ratio and thus use only uncorrelated features in model predictions. We note that this new term represents the total resistance, R_{flow}, in our hydraulic resistance model (7). Table 1 shows a comparison of the original model and the simplified PLS and OLS models.

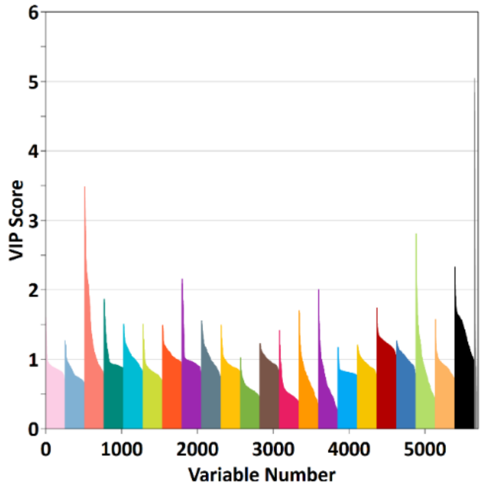

Figure 6 - VIP scores for full feature dataset, including time dependent and independent variables. Colors indicate the variables for each specific batch in the dataset, including unfolded variables and time independent variables.

Table 1 – Summary of PLS and OLS models evaluated.

Model	Variables	R^2	Q^2
Batch PLS – With unfolded variables	5938	0.92	0.71
Batch PLS – No unfolded variables	134	0.92	0.68
Batch PLS - Simplified	5	0.91	0.90
OLS – Test Data	5	0.92	N/A

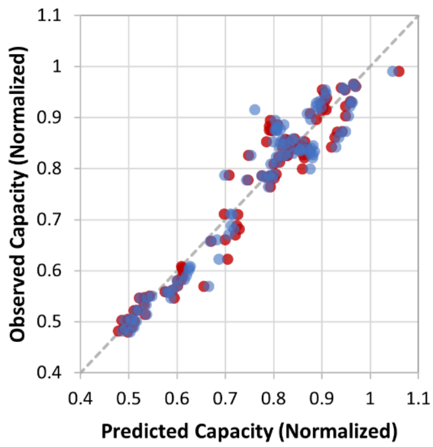

Figure 7 - Parity plot comparison of OLS model (red) and simplified PLS model (blue). Capacity normalized to manufacturer nominal capacity.

The OLS model was validated with our test set, showing an $R^2 = 0.92$ and a mean absolute error from the estimated capacity values of 5.4%, shown in Figure 8. All data fell within the estimated 95% prediction intervals of the OLS model. We included batches of the dataset that were time based in the test set to assess the model's performance.

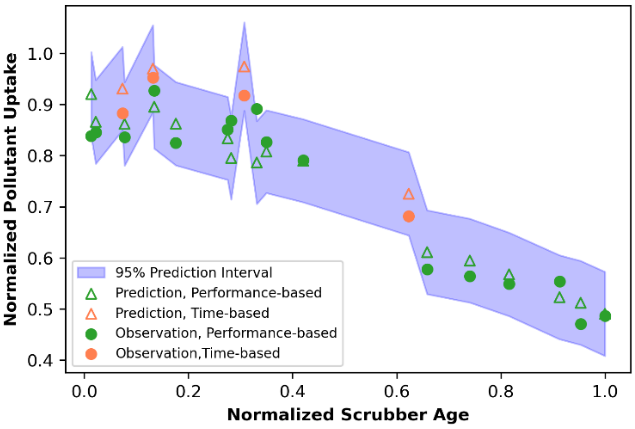

Figure 8 - Test set validation of OLS model. Scrubber age normalized to maximum age and capacity normalized to manufacturer nominal value.

Figure 8 shows that the performance-based batches show a variety of over and under-prediction, in contrast, the time-based batches show the model consistently overpredict their capacity. This overprediction is expected since the process cycles are being finished before full use of the scrubber capacity. Thus, our predictive model could also be used in analysis of process data to evaluate time-based process cycling and maximize their use if breakthrough is deemed an excessive risk during operations.

6. CONCLUSIONS

The combinations of physics-based feature development with a data-driven modeling methodology can enable the prediction of performance degradation in pollutant abatement processes. This is particularly relevant in scenarios where a direct measurement of a performance indicator, such as scrubbing capacity, is not available and its evolution is influenced by both single batch factors such as pollutant feed rate, and long-term effects such as scrubbing media fouling.

In the presented case study, the capacity of scrubber units was further influenced by the parallel operation of multiple units which featured different process history. A data-driven predictive model was able to predict the evolution of each individual scrubber whilst avoiding the need to estimate or measure all physical phenomena influencing unit capacity. Future efforts with this approach will involve deployment of the developed empirical models into multiple plants with different pollutant abatement configurations to understand how a general modeling approach may be leveraged across different plant configurations and abatement technologies.

REFERENCES

Abbasi, E., Moghaddam, M. R. A., & Kowsari, E. (2022). A systematic and critical review on development of machine learning based-ensemble models for prediction of adsorption process efficiency. *Journal of Cleaner Production*, *379*, 134588.

Bogojeski, M., Sauer, S., Horn, F., & Müller, K.-R. (2021). Forecasting industrial aging processes with machine learning methods. *Computers & Chemical Engineering*, *144*, 107123.

Bradley, W., Kim, J., Kilwein, Z., Blakely, L., Eydenberg, M., Jalvin, J., Laird, C., & Boukouvala, F. (2022). Perspectives on the integration between first-principles and data-driven modeling. *Computers & Chemical Engineering*, *166*, 107898.

Bui, L., Joswiak, M., Castillo, I., Phillips, A., Yang, J., & Hickman, D. (2022). A Hybrid Modeling Approach for Catalyst Monitoring and Lifetime Prediction. *ACS Engineering Au*, *2*(1), 17–26.

Eriksson, L., Byrne, T., Johansson, E., Trygg, J., & Vikström, C. (2013). *Multi- and Megavariate Data Analysis : Basic Principles and Applications* (Third revised edition, Vol. 1). MKS Umetrics AB Malmö, Sweden.

Haghpanah, R., Majumder, A., Nilam, R., Rajendran, A., Farooq, S., Karimi, I. A., & Amanullah, M. (2013). Multiobjective Optimization of a Four-Step Adsorption Process for Postcombustion CO2 Capture Via Finite Volume Simulation. *Industrial & Engineering Chemistry Research*, *52*(11), 4249–4265.

Hanif, M. A., Ibrahim, N., & Jalil, A. A. (2020). Sulfur dioxide removal: An overview of regenerative flue gas desulfurization and factors affecting desulfurization capacity and sorbent regeneration. *Environmental Science and Pollution Research*, *27*(22), 27515–27540.

Jacobs, J. H., Deering, C. E., Lesage, K. L., Stashick, M. J., & Marriott, R. A. (2021). Rapid Cycling Thermal Swing Adsorption Apparatus: Commissioning and Data Analyses for Water Adsorption of Zeolites 4A and 13X Over 2000 Cycles. *Industrial & Engineering Chemistry Research*, *60*(19), 7487–7494.

Kourti, T., Nomikos, P., & MacGregor, J. F. (1995). Analysis, monitoring and fault diagnosis of batch processes using multiblock and multiway PLS. *Journal of Process Control*, *5*(4), 277–284.

Lu, B., Castillo, I., Chiang, L., & Edgar, T. F. (2014). Industrial PLS model variable selection using moving window variable importance in projection. *Chemometrics and Intelligent Laboratory Systems*, *135*, 90–109.

P.K., R. M., Subbaiah, J., & Krithivasan, K. (2021). RF-LSTM-based method for prediction and diagnosis of fouling in heat exchanger. *Asia-Pacific Journal of Chemical Engineering*, *16*(5), e2684.

Sansana, J., Rendall, R., Castillo, I., Bruijne, L. de, Huggins, J., Phillips, A., & Reis, M. (2023). *A hybrid approach for advanced monitoring and forecasting of fouling with application to an ethylene oxide plant*.

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python. *9th Python in Science Conference*.

Taoufik, N., Boumya, W., Achak, M., Chennouk, H., Dewil, R., & Barka, N. (2022). The state of art on the prediction of efficiency and modeling of the processes of pollutants removal based on machine learning. *Science of The Total Environment*, *807*(Pt 1), 150554.