Systems & Control Transactions

Research Article - Peer Reviewed Conference Proceeding
ESCAPE 35 - European Symposium on Computer Aided Process Engineering
Ghent, Belgium. 6-9 July 2025
Jan F.M. Van Impe, Grégoire Léonard, Satyajeet S. Bhonsale,
Monika E. Polańska, Filip Logist (Eds.)

PSE PRESS

Soft-Sensor-Enhanced Monitoring of an Alkylation Unit via Multi-Fidelity Model Correction

Rastislav Fáber[a]*, Marco Vaccari[b], Riccardo Bacci di Capaci[b], Karol Ľubušký[c], Gabriele Pannocchia[b], and Radoslav Paulen[a]

[a] Faculty of Chemical and Food Technology, Slovak University of Technology in Bratislava, 812 37 Bratislava, Slovakia
[b] Department of Civil and Industrial Engineering, University of Pisa, 561 22 Pisa, Italy
[c] Slovnaft, a.s., 824 12 Bratislava, Slovakia
* Corresponding Author: rastislav.faber@stuba.sk.

ABSTRACT

Industrial process monitoring can benefit from utilizing historical data, providing insights for decision-making and operational efficiency. This study develops a soft-sensor-based approach leveraging multi-fidelity modeling to correct discrepancies between online sensors and laboratory analyses. A Gaussian process-based strategy is used to predict deviations between high-frequency low-fidelity sensor data and less frequent high-fidelity laboratory measurements. By exploring static and dynamic modeling frameworks, we assess their suitability for capturing process dynamics and addressing time-dependent variability. The multi-fidelity soft sensor noticeably improves predictive accuracy, outperforming high-fidelity and low-fidelity methods. This approach demonstrates applicability across various industrial settings where integrating diverse data sources enhances real-time process control and monitoring, reducing reliance on costly laboratory sampling.

Keywords: Machine Learning, Process Monitoring, Modelling, Information Management, Industry 4.0

INTRODUCTION

Industrial processes rely on effective monitoring and control to ensure efficiency, safety, and optimal performance. Accurate process data, acquired through sensors and analyzers, form the foundation for decision-making and optimization [1]. Laboratory measurements provide precise, high-fidelity (HF) data but are often infrequent due to cost or logistical constraints. In contrast, online sensors generate low-fidelity (LF) data at high frequencies but can vary in reliability and accuracy [2]. Differences in the quality and frequency of these data sources have motivated the development of advanced modeling approaches to improve process monitoring [3].

Soft sensors (SS) [4] have emerged as a key tool for improving process monitoring by using predictive models to estimate process variables based on online measurements. These models typically rely on periodic calibration with HF data to address sensor drift and inaccuracies. However, traditional calibration processes often use HF data solely for single-point corrections, leaving the potential of historical HF data to improve predictive accuracy over time. Multi-fidelity (MF) modeling [5] addresses this limitation by combining LF and HF data, enhancing predictive accuracy and improving the reliability of process monitoring systems.

MF modeling utilizes historical datasets to develop predictive models that mitigate the discrepancies between LF and HF data. Both static and dynamic models offer viable approaches for capturing the underlying behavior of the system. Static models [6] are well-suited for representing systems in relatively stable states but lack the ability to incorporate time-dependent behaviors, making them less appropriate for processes with significant temporal variability. In contrast, dynamic models [7] explicitly account for temporal dependencies, offering a more comprehensive representation of process variability at the cost of increased structural complexity.

Techniques, such as the Kalman filter [8], can be applied to both static and dynamic models to improve predictive accuracy. By estimating and compensating for potential unmeasured errors or biases in real time, these tools enhance the reliability of predictions, particularly in noisy, high-frequency datasets. While dynamic models

commonly integrate these methods to manage process variability, traditional approaches like bias correction have long been employed in static models to address long-term drifts or measurement inaccuracies.

Gaussian Process Regression (GPR) has demonstrated effectiveness in both single-fidelity and multi-fidelity modeling contexts, providing accurate predictions and quantifying uncertainties in high-dimensional industrial datasets [9]. Within this study, GPR serves as a key tool for fusing LF and HF data, correcting LF signals by predicting deviations and applying these corrections to approximate HF data in real time.

In this context, we aim to develop a MF modeling framework that enhances industrial process monitoring by integrating LF online sensor data with HF laboratory measurements. This approach addresses challenges such as sensor faults and measurement anomalies by utilizing a soft sensor (SS) that estimates true values, reducing reliance on frequent and costly laboratory sampling. By combining static and dynamic methods within a MF modeling framework, this work demonstrates how soft-sensor-based correction can improve predictive performance and reliability in process monitoring applications.

PROBLEM DEFINITION

This study considers two types of historical datasets: LF data obtained from online sensors, and HF data derived from laboratory measurements. LF data are collected continuously at a high sampling rate, offering detailed temporal information. In contrast, HF data are gathered less frequently and serve as reference points for calibrating and validating the LF sensors.

The time instances for LF measurements are represented as $\mathcal{T}_{LF} = \{t_1, t_2, \ldots, t_{n_{LF}}\}$, and for HF samples as $\mathcal{T}_{HF} = \{\tau_1, \tau_2, \ldots, \tau_{n_{HF}}\}$, with $\mathcal{T}_{HF} \subset \mathcal{T}_{LF}$. Typically, HF data are collected far less frequently than LF data, making $n_{HF} \ll n_{LF}$. The corresponding index sets for HF and LF data are denoted as $\mathrm{I}_{HF} \subset \mathrm{I}_{LF}$, respectively.

The objective is to create a predictive model that maps process variables $\boldsymbol{x} \in R^{n_x}$ to a target output $y \in R$. A linear model at the i-th time instant is expressed as:

$$\hat{y}_i = \boldsymbol{\theta}^T \boldsymbol{\varphi}_i + \epsilon_i, \tag{1}$$

where $\hat{y}_i$ is the predicted output, $\boldsymbol{\varphi}_i$ represents a vector of regressors, $\boldsymbol{\theta}$ is the parameter vector, and ϵ_i accounts for errors or unmodeled variability. Alternatively, a nonlinear model can be defined as:

$$\hat{y}_i = f(\boldsymbol{\theta}, \boldsymbol{\varphi}_i) + \epsilon_i, \tag{2}$$

where f is a nonlinear function parameterized by $\boldsymbol{\theta}$. The structure of the regressor vector $\boldsymbol{\varphi}_i$ depends on the chosen modeling approach:

a) Static model: $\boldsymbol{\varphi}_i = \boldsymbol{x}_i$, this approach relies solely on current input values.

b) Dynamic model: $\boldsymbol{\varphi}_i = (\boldsymbol{x}_{i-1}, \ldots, \boldsymbol{x}_{i-j}, y_{i-1}, \ldots, y_{i-k})$, this approach incorporates lagged inputs and outputs to account for temporal dependencies.

Here, j and k denote the number of lagged values included for the input variables $\boldsymbol{x}$ and output y, respectively.

Figure 1: Soft sensor development flowchart.

METHODOLOGY

This section outlines the methodology used to develop and compare static and dynamic modeling approaches (see Figure 1).

Preprocessing

Raw data, denoted as $\boldsymbol{X}_{raw}$ for input features and $\boldsymbol{y}_{raw}$ for the dependent variable, often contain noise, outliers, and missing values due to plant disturbances. Preprocessing addresses these issues to ensure robust model development [10]. Outliers are detected and replaced while preserving temporal dependencies in the dynamic model, as discussed in [11].

Feature Selection

Feature selection (FS) aims to identify a subset of relevant process variables, $\boldsymbol{\phi}$, enhancing model interpretability and reducing overfitting. Several methods guide feature selection, balancing statistical rigor with practical relevance informed by process knowledge. As detailed in [11], the following approaches were employed for both static and dynamic models:

a) Principal Component Analysis (PCA) reduces dimensionality by transforming the original variables into uncorrelated principal components (PCs) that maximize variance.

b) Partial Least Squares (PLS) Regression identifies latent variables that maximize covariance between the independent variables and the dependent variable.

c) LASSO Regression introduces an ℓ_1-norm penalty to enforce sparsity and enhance interpretability.

d) Stepwise Regression (SR) iteratively selects variables based on statistical significance, guided by metrics such as Akaike Information Criterion (AIC), Corrected Akaike Information Criterion (AICc), and the Bayesian Information Criterion (BIC).

Beyond statistical selection, domain knowledge is incorporated to refine the feature set. Process experts review the ranked variables and identify unsuitable candidates that could introduce misleading correlations. The next-best ranked alternative from the best-performing FS method is substituted. This approach ensures the model captures the most relevant process features while maintaining transferability to similar environments with minor adjustments based on site-specific expertise. These methods enhance model interpretability, streamline computational efficiency, and minimize the risk of overfitting [12].

Soft Sensor Training

The parameters of the model defined in Equation 1 are estimated by solving the following optimization problem:

$$\min_{\boldsymbol{\theta}} \frac{1}{2}||\boldsymbol{y}_l - \hat{\boldsymbol{y}}_l||_2^2 \equiv \min_{\boldsymbol{\theta}} \frac{1}{2}\Sigma_{i=1}^{n_l}(y_{l,i} - \hat{y}_{l,i})^2, \quad (7)$$

where $l \in \{LF, HF\}$ denotes whether the model corresponds to LF or HF data, and n_l is the number of samples in the respective dataset.

Multi Fidelity Soft Sensor Development

The MF SS integrates trained LF and HF models into a unified MF modeling approach. The predictor matrix $\boldsymbol{\phi}_{MF}$ is constructed as follows:

$$\boldsymbol{\varphi}_{MF,i} = (\boldsymbol{\varphi}_{LF,i}^T, \boldsymbol{\varphi}_{HF,i}^T, \boldsymbol{y}_{LF}^T, \hat{\boldsymbol{y}}_{LF}^T)^T. \quad (8)$$

This matrix combines features and predictions from both fidelity levels. An additional feature selection step can be conducted to further refine the input space.

Subsequently, the MF SS is trained by using a Gaussian Process (GP), which is defined as:

$$\hat{\boldsymbol{y}}_{MF} \sim \mathcal{GP}(m(\boldsymbol{\kappa}), k(\boldsymbol{\kappa}, \boldsymbol{\kappa}')), \quad (9)$$

where $\boldsymbol{\kappa}$ represents selected predictors derived from $\boldsymbol{\phi}_{MF}$. The mean function $m(\boldsymbol{\kappa})$ captures the expected output, while the covariance function, also referred to as the kernel $k(\boldsymbol{\kappa}, \boldsymbol{\kappa}')$, describes dependencies between data points.

In this model, the kernel is defined as:

$$k(\boldsymbol{\kappa}, \boldsymbol{\kappa}') = C.e^{-M} + e^{\frac{2sin^2 M}{\ell^2}} + \sigma^2.\delta_{\kappa,\kappa'}, \quad (10)$$

where $M = 1/2\,||\boldsymbol{\kappa} - \boldsymbol{\kappa}'||^2$. The kernel comprises four components (from left to right): a constant term C, a radial basis function (RBF) kernel to capture smooth variations, a periodic kernel to model sinusoidal behaviors, and a white noise kernel accounting for measurement noise through $\delta_{\kappa,\kappa'}$. The parameters include ℓ (length scale), determining the range of influence, and σ^2, representing noise variance

CASE STUDY

The alkylation process plays a pivotal role in refineries, producing high-octane branched isoparaffins, known as alkylate, which are crucial components of clean gasoline. This process involves the reaction of C_3–C_4 olefins with isobutane (i-C_4) under the influence of an acid catalyst. The reaction mechanism begins with the protonation of olefins, forming carbonium cations that subsequently react to generate C_8 isomers [13, 14].

Efficient alkylate production relies on maintaining an optimal reactant ratio. To ensure this, online analyzers are strategically deployed throughout the plant (Figure 2) to provide real-time measurements of key feed and recycle stream components. This enables precise process control. Over a six-month period, a comprehensive

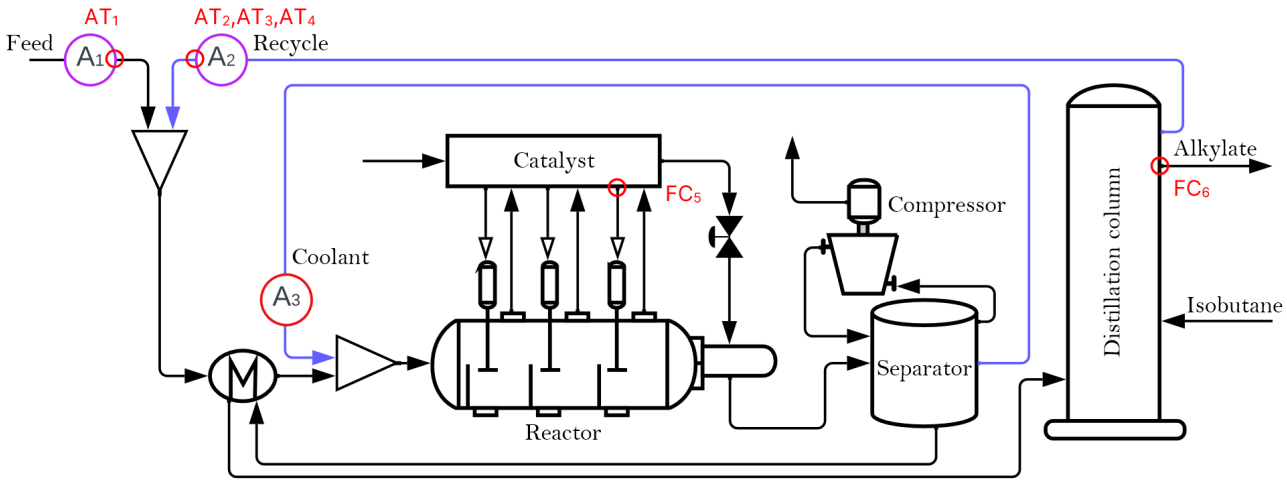

Figure 2: Simplified schematic of the alkylation unit, highlighting Analyzer A_3 (red) as well as the six selected variables used in soft sensor training.

dataset, $\boldsymbol{X}_{raw}$, was gathered, comprising more than 1085 process variables from online sensors, analyzers, resulting in $N \times 1085$ data points, while the laboratory samples correspond to a subset of these timestamps.

The study focuses on Analyzer A_3 and its corresponding laboratory reference data, $\boldsymbol{y}_{LF}$ and $\boldsymbol{y}_{HF}$. Analyzer A_3 monitors the i-C_4 concentration in the recycle stream but has demonstrated inconsistent performance. These inconsistencies often lead to an over-recycling of i-C_4, imposing additional downstream processing burdens, such as increased heating and treatment requirements. Therefore, improving the prediction of analyzer measurements is vital for optimizing production efficiency and minimizing by-products.

IMPLEMENTATION

We preprocess the raw input data $\boldsymbol{X}_{raw}$ and corresponding raw output data $\boldsymbol{y}_{raw}$ (see Figure 1). The resulting preprocessed data, $\boldsymbol{X}_{proc}$ and $\boldsymbol{y}_{proc}$, are then split into training and testing sets in a 60% to 40% ratio, using a training period spans four months, while the testing period covers two months. This selection is particularly relevant as seasonal variations during this timeframe can significantly influence process dynamics, including changes in the characteristics of the feedstream. Given the limited availability of HF data, chronological ordering was maintained to preserve time dependencies.

From the preprocessed dataset $\boldsymbol{X}_{proc}$, we identify the most relevant variables ($\boldsymbol{\phi}_i$) using the best-performing FS method outlined in the Feature Selection subsection. To address the limited HF data, we rely on the LF dataset during FS, as the sparsity of HF data may fail to capture key interactions and trends where the output variable is influenced by time-dependent changes across multiple inputs. By leveraging the more frequent LF measurements, we ensure that the selected features ($\boldsymbol{\phi}_{LF}$) reflect these dynamics more effectively, reducing the risk of missing correlations present in the data that would not be evident in $\boldsymbol{\phi}_{HF}$.

Both static HF and LF SSs ($\hat{\boldsymbol{y}}_{HF}$ and $\hat{\boldsymbol{y}}_{LF}$) are trained using the same set of features $\boldsymbol{\phi}_{LF}$, as described in Equation 1. These features are further refined with insider knowledge to enhance their operational relevance and consistency across both SSs. For the dynamic LF SS, we extend the feature set ($\boldsymbol{\phi}_{LF}$) by incorporating time-shifted versions determined by the model order to capture time-dependent system dynamics. We formulate the SS as a linear state-space model, and system identification is performed using the SIPPY (Systems Identification Package for Python) [15] library, employing the PARSIM-K robust identification method [16]. This approach, aimed for closed-loop data identification, ensures accurate representation of system dynamics with minimal number of parameters. The SS performance is further enhanced using information criteria such as AIC, AICc, and BIC for optimal model order selection. Additionally, a Kalman filter option of the PARSIM-K method is explored to refine the dynamic SS predictive accuracy.

We extend the LF models into MF framework by incorporating a GP model. This approach corrects deviations between LF and HOnline sensor ents, leveraging the strengths of both datas data r more accurate real-time monitoring of process variables (as illustrated in Figure 3). The MF SS effectively mitigates signal drift and compensates for the limited availability of HF samples.

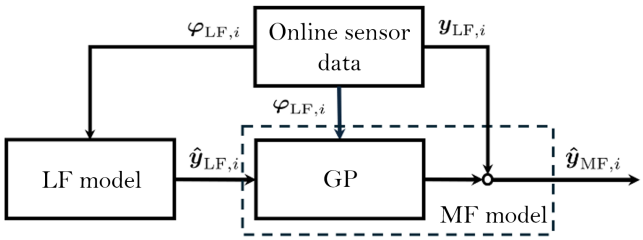

Figure 3: Workflow of the MF SS correction of the online analyzer A_3.

Table 1: RMSE comparison of the online analyzer against the static and dynamic LF prediction models.

Method	Training RMSE		Testing RMSE	
$\boldsymbol{y}_{LF}$ to $\hat{\boldsymbol{y}}_{LF,s}$	0.21		0.37	
$\boldsymbol{y}_{LF}$ to $\hat{\boldsymbol{y}}_{LF,d}$	w KF	w/o KF	w KF	w/o KF
	0.03	0.29	0.02	0.42

Table 2: RMSE comparison of the laboratory samples against the applied predictive approaches.

Comparison	Training RMSE	Testing RMSE
$\boldsymbol{y}_{HF}$ to $\boldsymbol{y}_{LF}$	0.63	0.80
$\boldsymbol{y}_{HF}$ to $\hat{\boldsymbol{y}}_{HF}$	0.18	0.91
$\boldsymbol{y}_{HF}$ to $\hat{\boldsymbol{y}}_{MF,s}$	0.22	0.46
$\boldsymbol{y}_{HF}$ to $\hat{\boldsymbol{y}}_{MF,d}$	0.31	0.38
$\boldsymbol{y}_{HF}$ to $\hat{\boldsymbol{y}}_{MF,d_{KF}}$	0.21	0.46

RESULTS

We standardize the dataset to ensure consistent variable scaling across all analyses. During FS, we prioritize the LF data, as it better captures the time-dependent relationships between input ($\boldsymbol{X}_{proc}$) and output ($\boldsymbol{y}_{proc}$). The sparse nature of $\boldsymbol{y}_{HF}$ can overlook these dynamic interactions, as it fails to capture the nuances of how $\boldsymbol{y}_{LF}$ responds to changes in $\boldsymbol{x}$ over time. We select SR as the most suitable feature selection method based on the results from [11]. Additionally, we refine these features with domain knowledge, improving their relevance and ensuring consistency across both static and dynamic models. The final feature set includes six critical variables, highlighted in Figure 2: four concentrations (AT_1 to AT_4) and

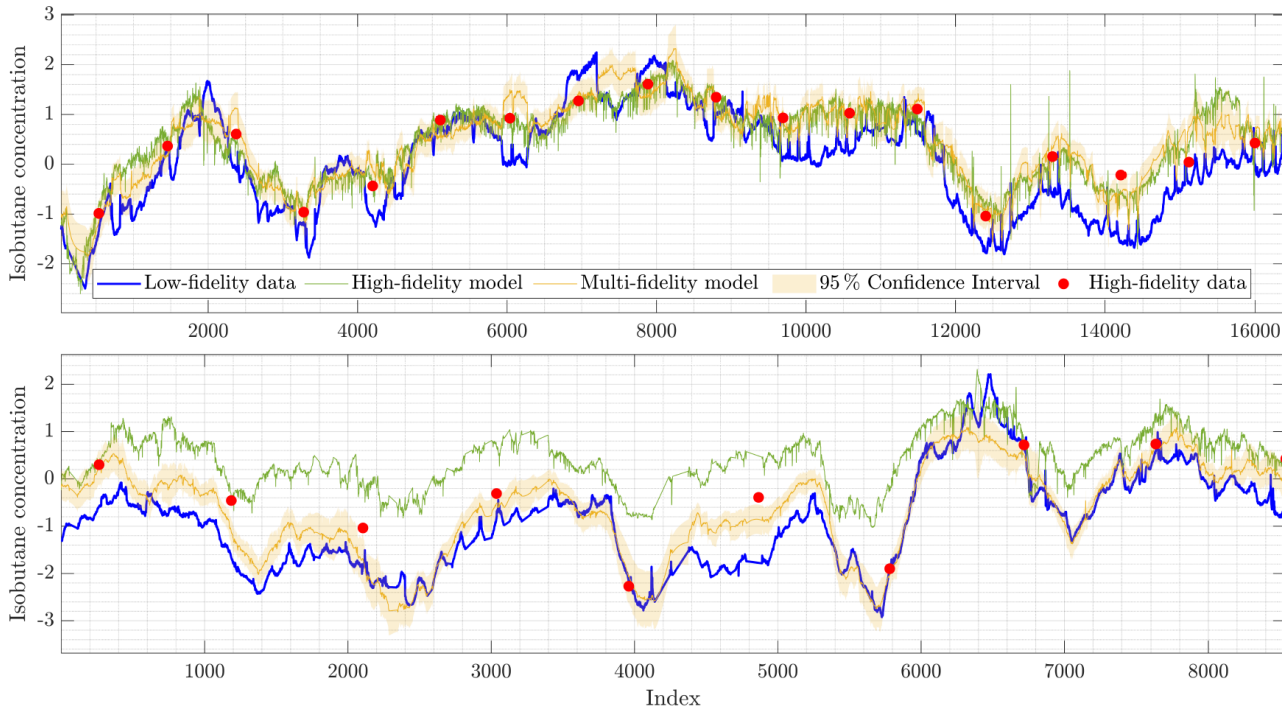

Figure 4: Predicted normalized i-C_4 concentration for the training set (top) and testing set (bottom) as time series.

two flow rates (FC_5 and FC_6).

We aim to predict the measurements from the online analyzer A_3. We train a static LF SS $\hat{\boldsymbol{y}}_{LF,s}$ with the six previously selected inputs. This approach achieves RMSE values of 0.21 during training and 0.37 during testing (Table 1). However, the alkylation unit operates with multiple recycles, where process streams re-enter the system at different stages, causing data variations. These recycles introduce complexities that static models cannot address. We develop a dynamic LF model $\hat{\boldsymbol{y}}_{LF,d}$ to address the temporal dependencies in the data. We optimize the SS complexity through system identification and select a 4th-order model to balance accuracy and computational efficiency based on AIC, AICc, and BIC criteria. We achieve a RMSE value of 0.29 for training and 0.42 for testing. To improve predictive accuracy, we integrate a Kalman filter into the $\hat{\boldsymbol{y}}_{LF,d}$, defined as $\hat{\boldsymbol{y}}_{LF,d_{KF}}$, significantly reducing RMSE values to 0.03 for training and 0.02 for testing. The Kalman filter proves instrumental in mitigating disturbances and improving reliability.

The performance metrics for all predictive approaches are summarized in Table 2, with laboratory data ($\boldsymbol{y}_{HF}$) compared against all SSs. The current online analyzer A_3 ($\boldsymbol{y}_{LF}$) exhibits RMSE values of 0.63 for training and 0.80 for testing, resulting in inaccurately capturing the real process trends.

The HF SS $\hat{\boldsymbol{y}}_{HF}$ (green line in Figure 4) achieves an RMSE of 0.18 on the training set, which increases to 0.91 on the testing set. This increased RMSE value highlights that both, the features selected, and the limited amount of HF data available likely hamper the SS ability to capture the full variability within the HF dataset; as further visual analysis suggests that the HF SS effectively captures the trends in the HF data.

To obtain the MF SS output, we train a GP to predict deviations of the LF analyzer data from the HF laboratory measurements, defined as $\boldsymbol{\Delta}_{HF,i} = \mathbf{y}_{HF,i} - \mathbf{y}_{LF,i}, \forall i \in \mathfrak{I}_{HF}$. This approach enables operators to monitor $\boldsymbol{\Delta}_{HF}$ and its dispersion, offering a practical means to assess the calibration of the online analyzer and identify potential deviations in its accuracy. The GP model can incorporate either the static LF model ($\hat{\boldsymbol{y}}_{LF,s}$) or the dynamic LF model ($\hat{\boldsymbol{y}}_{LF,d}$) as an additional input, with the latter having two variations: with ($\hat{\boldsymbol{y}}_{LF,d_{KF}}$) and without Kalman filtering. PCA is applied to reduce the dimensionality of the input variables, consisting of the original six process variables and the selected LF model, into four principal components (PCs). Depending on the chosen input, the MF SS predictions are denoted as $\hat{\boldsymbol{y}}_{MF,s}$, $\hat{\boldsymbol{y}}_{MF,d}$, or $\hat{\boldsymbol{y}}_{MF,d_{KF}}$.

The static MF SS ($\hat{\boldsymbol{y}}_{MF,s}$) achieved an RMSE of 0.22 during training and 0.46 during testing. Although this approach provides promising results during training, its higher testing RMSE indicates similar limitations in addressing the aforementioned system dynamics.

When selecting the most suitable additional input within the dynamic MF SS ($\hat{\boldsymbol{y}}_{MF,d}$), $\hat{\boldsymbol{y}}_{LF,s}$ achieves a lower RMSE when compared to $\boldsymbol{y}_{LF}$ (0.37 compared to 0.42 for testing, as shown in Table 1), it exhibits significant deviations from $\boldsymbol{y}_{HF}$. In contrast, $\hat{\boldsymbol{y}}_{LF,d}$ better aligns with the trends in $\boldsymbol{y}_{HF}$, as evidenced by smaller maximum and

minimum errors relative to $\boldsymbol{y}_{HF}$. This suggests that the dynamic SS, despite a higher RMSE against $\boldsymbol{y}_{LF}$, provides a more reliable representation of process dynamics.

The Kalman-filtered dynamic MF SS ($\hat{\boldsymbol{y}}_{MF,d_{KF}}$) amplifies certain LF anomalies, propagating biases that reduce the GPs ability to make unbiased predictions (RMSE = 0.46). Consequently, the non-Kalman dynamic MF SS ($\hat{\boldsymbol{y}}_{MF,d}$) enhances predictive accuracy while mitigating the influence of LF anomalies (RMSE = 0.38). As shown in Figure 4, the blue solid line represents the original LF data ($\boldsymbol{y}_{LF}$), the orange line denotes the MF SS predictions ($\hat{\boldsymbol{y}}_{MF,d}$), and the red markers indicate the HF laboratory data. The orange shaded region illustrates the 95% confidence interval of the GP model.

Ultimately, the most accurate MF SS ($\hat{\boldsymbol{y}}_{MF,d}$) accounts for LF discrepancies with GP-predicted corrections, achieving RMSE values of 0.31 for training and 0.38 for testing. These findings highlight the importance of time-dependent behavior, enabling us to capture system dynamics and adapt to temporal variations, which leads to improved performance over static soft sensors.

CONCLUSION

In this work, we compare static and dynamic modeling approaches within a multi-fidelity framework to improve the accuracy of online measurements. The static models ($\hat{\boldsymbol{y}}_{LF,s}$) fail to capture temporal dependencies, which limits their applicability in complex systems with time-varying dynamics. Dynamic models ($\hat{\boldsymbol{y}}_{LF,d}$), especially those incorporating Kalman filtering, effectively address these challenges, achieving an improvement in capturing the underlying process behavior. However, incorporating the Kalman filter into the MF SS increases the RMSE, negatively impacting performance in the MF framework. The most viable SS ($\hat{\boldsymbol{y}}_{MF,d}$) demonstrates a 52.50% enhancement in accuracy over the currently implemented online analyzer, significantly improving the alignment between online LF measurements and laboratory HF values.

Future work will focus on refining dynamic modeling techniques by addressing process non-linearities and further optimizing GP implementations to enhance accuracy and robustness.

ACKNOWLEDGEMENTS

This work is funded by the Slovak Research and Development Agency under the project APVV-21-0019, by the Scientific Grant Agency of the Slovak Republic under the grant 1/0263/25, and by the European Commission under the grant 101079342 (Fostering Opportunities Towards Slovak Excellence in Advanced Control for Smart Industries). RP acknowledges the financial support by the European Commission under the grant scheme NextGenerationEU project no. 09I03-03-V04-00530.

REFERENCES

1. Mondal PP, Galodha A, Verma VK, Singh V, Show PL, Awasthi MK, Lall B, Anees S, Pollmann K, Jain R, Review on machine learning-based bioprocess optimization, monitoring, and control systems. Bioresource Technology, 370, 128523. (2023)
2. Peherstorfer B, Willcox K, Gunzburger M, Survey of multifidelity methods in uncertainty propagation, inference, and optimization SIAM Rev., 60 (3), pp. 550-591. (2018)
3. Zhu J, Ge Z, Song Z, Gao F, Review and big data perspectives on robust data mining approaches for industrial process modeling with outliers and missing data. Ann. Rev. Control, 46, 107-133. (2018)
4. Sun YN, Qin W, Hu JH, Xu HW, Sun PZH, A causal model-inspired automatic feature-selection method for developing data-driven soft sensors in complex industrial processes. Engineering, 22, 82–93. (2023)
5. Giselle Fernández-Godino M, Review of multifidelity models. Advances in Computational Science and Engineering, 1(4), 351–400. (2023)
6. Zou KH, Tuncali K, Silverman SG, Correlation and simple linear regression. Radiology. 227(3). (2003)
7. Nelles,O, Nonlinear Dynamic System Identification. In: Springer, Cham. (2020)
8. Kalman RE, A new approach to linear filtering and prediction problems. J Basic Eng 82(1):35–45 (1960)
9. Rasmussen CE, Gaussian Processes in Machine Learning, 63–71. Springer Berlin Heidelberg. (2004)
10. Fáber R, Mojto M, Ľubušký K, Paulen R, From data to alarms: Data-driven anomaly detection techniques in industrial settings. In ESCAPE34 - PSE24. (2024)
11. Fáber R, Vaccari M, Bacci di Capaci R, Ľubušký K, Pannocchia G, Paulen R, Improving process monitoring via dynamic multi-fidelity modeling. 14th IFAC Symposium on Dynamics and Control of Process Systems, including Biosystems. (2025)
12. Li J, Cheng K, Wang S, Morstatter F, Trevino RP, Tang J, Liu H, Feature selection: A data perspective. ACM Comput. Surv., 50(6). (2017)
13. Pall Refineries: Application focus H_2SO_4 alkylation process description. Tech. report, Pall Corp. (2018)
14. Speight JG, The refinery of the future. Gulf Professional Publishing, Elsevier. (2020)
15. Armenise G, Vaccari M, Bacci di Capaci R, Pannocchia G, An open-source system identification package for multivariable processes. In UKACC 12th International Conference on Control,

152–157. (2018)
16. Pannocchia G, Calosi M, A predictor form parsimonious algorithm for closed-loop subspace identification. Journal of Process Control, 20(4), 517–524. (2010)