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

Integrated Data Analytics and Regression Techniques for Real-time Anomaly Detection in Industrial Processes

Rastislav Fáber * **Martin Mojto** * **Karol Ľubušký** ** **Radoslav Paulen** *

* *Faculty of Chemical and Food Technology, Slovak University of Technology in Bratislava, Bratislava, Slovakia*
** *Slovnaft, a.s., Bratislava, Slovakia*

Abstract: In this paper, we present a data-based monitoring approach designed for industrial data classification, aiming to minimize misclassifications of normal operations and to maximize the detection of anomalies and outliers. We make use of moving-horizon approaches and regression methods. Through evaluation of various algorithms on an industrial dataset, we showcase the effectiveness of the classification. As per our findings, effective detection can only be realized in conjunction of moving-horizon estimator with a regression model trained on historical measurements. The best prediction models consistently achieve accurate detection within the approved process tolerance, highlighting the efficacy of the proposed approach.

Keywords: Streaming data, Anomaly detection, Outlier detection, Regression

1. INTRODUCTION

In the industrial sector, precision is essential for ensuring efficiency, safety, and reliability. The critical task of real-time monitoring and validating process measurements relies heavily on anomaly detection to identify deviations, signaling potential malfunctions. As such, advanced algorithms are increasingly supplementing periodic laboratory measurements to uncover these anomalies within large datasets, thus reducing operational burdens and improving responsiveness.

Guided by established practices in anomaly detection (Wang and Liu, 2021), we process raw measurements to identify crucial anomalies, enabling timely interventions in operational settings. Our approach targets both outliers (sudden changes) and anomalies (deviating trends) to differentiate normal process variations from those altered by disturbances — a requirement for maintaining operational efficiency and safety. In this paper, "outliers" refer to isolated data points or clusters that appear and dissipate abruptly, breaking the usual temporal patterns. "Anomalies", however, describe sequences of data points that form trends significantly diverging from standard operations (Iglesias Vázquez et al., 2023).

The application of data analysis in anomaly detection across sectors, from industrial operations (Antonini et al., 2023; Inoue et al., 2017) to health monitoring (Raza et al., 2023), has seen a surge in popularity. Various methods, including filtering, deep learning, machine learning, regression, and clustering algorithms have been explored for their potential to enhance anomaly detection capabilities (Schmidl et al., 2022). Within this spectrum, filtering methods such as threshold or standard deviation filters (Afanasyev and Fedorova, 2019; Blázquez-García et al., 2021), and regression approaches have shown promise in detecting outliers in both univariate and multivariate time-series data. These methods, including the use of exponentially weighted moving averages (Carter and Streilein, 2012; Roberts, 1959) and autoregressive models, offer nuanced insights into system dynamics, overcoming the limitations tied to more straightforward univariate approaches (Yoon et al., 2022).

Mathematical models empower a deeper grasp of complex process dynamics. For example, a Kalman filter can exploit the connection with a first-principles model to estimate the state of a system and reduce the impact of uncertainties, based on a series of observations. The filter uses a recursive algorithm that minimizes the mean squared error between the predicted and actual state (Jin et al., 2022). Although powerful as a tool, the uncertainty of prediction accuracy of the underlying dynamic model must be reasonably well described. Moreover, the availability of a dynamic model is not common in industrial setup. A density-based clustering algorithm, such as DBSCAN (Ester et al., 1996), is recognized for its effectiveness in identifying outliers that lie alone in low-density regions. However, its application can be challenging in high-dimensional datasets due to its effectiveness being significantly reduced. A relevant approach could be the application of deep

⋆ E-mail address: rastislav.faber@stuba.sk (R. Fáber).
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 1/0691/21, and by the European Commission under the grant no. 101079342 (Fostering Opportunities Towards Slovak Excellence in Advanced Control for Smart Industries).

autoencoders, focusing on complex time-series data from production processes. Evaluating multiple types of autoencoders, the available research identifies long short-term memory and convolutional neural network-based models as particularly promising for enhancing productivity (Tziolas et al., 2022).

Despite the various methodologies explored, a gap remains in the application and integration of these methods in real-time industrial settings. Our study aims to bridge this gap by emphasizing simplicity and practicality, crucial for industrial implementation. We present a novel approach that integrates a moving-horizon methodology with regression techniques to enhance the detection of anomalies in real-time industrial datasets. This approach not only addresses the immediate need for operational anomaly detection but also contributes to the broader field of industrial digitalization, pushing beyond traditional practices to leverage extensive process data for real-time decision-making.

The paper is structured to describe the utilized methodology, beginning with pre-processing the raw industrial dataset (see Sec. 2.1), though application of the moving-horizon approach (Sec. 2.2) to establish effective outlier detection in process variables based on recent measurements, to the use of regression methodologies (see Sec. 2.3) for anomaly monitoring. A case study is introduced in Sec. 3 and results are shown in Sec. 4.

2. METHODOLOGY

Our objective is to detect outliers and anomalies within a one-dimensional (1D) process variable $y(t)$ with real-time streamed data. At time t, current and past values of y are available, along with measurements of other process variables $\mathbf{x}(t) \in \mathbb{R}^n$. This section details the data treatment for 1D and multi-dimensional (nD) data and the studied approaches for outlier/anomaly detection.

2.1 Data Treatment

Identifying systematic errors through visual inspection of time series is effective yet limited. To uncover errors not immediately visible, we employ data treatment methods designed to enhance error detection capabilities.

Three-standard-deviations rule. Under the assumption of normally distributed data, the sample mean $\hat{y}$ and standard deviation σ establish the 3-sigma interval as:

$$T = \hat{y} \pm 3\sigma. \tag{1}$$

Data points outside this interval ($y(t) \notin T$) are deemed outliers, with the expectation that 99.7% of data falls within this range. This rule's simplicity facilitates quick identification of significant deviations from the normality.

Minimum Covariance Determinant (MCD). MCD (Rousseeuw and Driessen, 1999) provides robust outlier detection for nD data. It uses the Mahalanobis distance:

$$d_{\mathrm{MCD},i} = \sqrt{(\boldsymbol{x}(t_i) - \hat{\boldsymbol{x}})^{\mathsf{T}}\, \mathbf{S}^{-1}\, (\boldsymbol{x}(t_i) - \hat{\boldsymbol{x}})}, \tag{2}$$

which evaluates the dissimilarity between a measurement $\boldsymbol{x}$ and the underlying probability distribution. Robustness is achieved by iteratively identifying subsets of historical data with the minimum determinant of the sample covariance matrix $\mathbf{S}$, thereby mitigating the influence of outliers. The core of the MCD algorithm involves the iterative selection of data subsets and calculation of $\hat{\boldsymbol{x}}$ and $\mathbf{S}$. A new subset is formed by selecting a specified number of observations with the smallest d_{MCD}. This process continues until the stabilization of the determinant of $\mathbf{S}$. An equivalent of (1) reads as:

$$T = \hat{\boldsymbol{x}} \pm \chi^2_{n,0.997}\mathbf{S}^{\frac{1}{2}}\boldsymbol{e}, \tag{3}$$

which uses the matrix square root of $\mathbf{S}$ and unit vector $\boldsymbol{e}$ and where $\chi^2_{n,0.997}$ is the quantile of the χ^2 distribution with n degrees of freedom and probability level of 99.7 %.

The techniques presented so far are global and may not be effective in capturing local anomalies. We consider the temporal aspect of the data in the following text.

2.2 Outlier Detection using Data Averaging

The idea is to capture the local behavior of the signal $y(t)$ over a window of size N. The choice of N determines whether the designed detector concentrates on local, temporary, or global deviations. One can compute the confidence interval within such a method as:

$$T_{t_i} = \hat{y}(t_i) \pm t_{N,0.997}\sqrt{\frac{\sigma^2}{N}}, \tag{4}$$

where $t_{N,0.997}$ is the inverse of Student's t distribution (Student, 1908) with N degrees of freedom and σ^2 is the variance corresponding to the monitored window. The calculation of the sample mean is the distinguishing feature of the studied methods as described below.

Historical Mean. This involves calculating the historical mean of $y(t)$ (N is in order of months or years). The detection rule consists in the use of (4).

Historical Mean of Differences. By calculating the mean of the differences between consecutive data points $\Delta y(t_i) = y(t_i) - y(t_{i-1})$, we capture the instantaneous variations in the measurements. Observations deviating from the interval (1) are flagged as outliers.

Moving Average (MA filter). By considering only a certain number of past data points, we compute an average that dynamically adapts to changes in the dataset. The filtered value is calculated as (Oppenheim, 1999):

$$\hat{y}(t_i) = \frac{1}{N}\sum_{j=0}^{N-1} y(t_{i-j}). \tag{5}$$

Measurements outside the interval (4) are identified as outliers, i.e. potential recent data inconsistencies.

Centered Moving Average (MA smoother). The centered moving average is calculated as:

$$\hat{y}(t_i) = \frac{1}{N}\sum_{j=-\lfloor (N-1)/2 \rfloor}^{\lfloor (N-1)/2 \rfloor} y(t_{i-j}). \tag{6}$$

While dealing with the streamed data, we do not posses the knowledge of future measurements. Consequently, this approach cannot be used for detection. Yet, it can be used to assess the past detection outcomes.

Moving Average with Prediction (MA predictor). Our methodology extends the traditional moving average to allow for a dynamic adjustment of the moving average, integrating additional information from other process variables. Specifically, the moving average is modified to include the predicted measurement difference at the current time step, as follows:

$$\hat{y}(t_i) = \frac{1}{N}\sum_{j=0}^{N-1} y(t_{i-j}) + \Delta\hat{y}(\boldsymbol{x}(t_i), \boldsymbol{x}(t_{i-1})), \qquad (7)$$

where $\Delta\hat{y}(\boldsymbol{x}(t_i))$, is derived from a regression model trained to predict the change in the measurement based on current and historical process variables. This model accounts for the trends within the process variables thereby enabling a more accurate prediction of future states. This adjustment allows the moving average to not only reflect past measurement trends but also to adapt based on current (and near future) trends given the process state indicated by process variables $\boldsymbol{x}$.

2.3 Anomaly Detection using Regression Methods

Regression models can be used to identify anomalies of the process variable based on the deviations of measurements relative to the model predictions $\hat{y}(\boldsymbol{x})$, e.g., using a linear model $\hat{y}(\boldsymbol{x}) = \boldsymbol{\beta}^\intercal \boldsymbol{x}$ with parameters $\boldsymbol{\beta}$. The regression model can also predict the relative change of the process variable (see Eq. (7)):

$$\Delta\hat{y}(\boldsymbol{x}(t_i), \boldsymbol{x}(t_{i-1})) = \boldsymbol{\beta}_\Delta^\intercal(\boldsymbol{x}(t_i) - \boldsymbol{x}(t_{i-1})). \qquad (8)$$

Ordinary Least Squares. The approach estimates model parameters by minimizing the sum of squared differences between the observed and predicted values as:

$$\min_{\boldsymbol{\beta}} \frac{1}{2}\sum_{i=1}^{n_t}(y(t_i) - \boldsymbol{\beta}^\intercal \boldsymbol{x}(t_i))^2, \qquad (9)$$

where n_t is the number of training data points.

Least Absolute Shrinkage and Selection Operator (LASSO). This method extends the regression by incorporating a penalty term (Santosa and Symes, 1986). This encourages model sparsity, allowing to effectively identify and shrink less relevant variables by solving:

$$\min_{\boldsymbol{\beta}} \frac{1}{2}\sum_{i=1}^{n_t}(y(t_i) - \boldsymbol{\beta}^\intercal \boldsymbol{x}(t_i))^2 + \lambda\|\boldsymbol{\beta}\|_1, \qquad (10)$$

where λ is a weight between the accuracy of the model training and the model overfitting. The magnitude of the ℓ_1-penalization element results in certain parameters being equal to zero at the optimum. The resulting model is then less complex, more robust and interpretable.

Principal Component Regression (PCR). Principal Component Analysis (PCA) can be applied to large, multi-dimensional datasets to enhance their interpretability (Pearson, 1901). By creating new uncorrelated variables (principal components) that maximize the variance, PCA helps to reduce the dimensionality of the data while minimizing information loss. Subsequently, OLS or LASSO can be applied to learn the model parameters in the latent space. The combination of PCA and LASSO (denoted further as PCA+LASSO) can leverage the strengths of both methods. PCA reduces dimensionality and LASSO adds sparsity to best represent the dependent variable prediction.

To evaluate the effectiveness of these approaches, we employ the Root Mean Square Error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{n_t}\sum_{i=1}^{n_t}(y(t_i) - \boldsymbol{\beta}^\intercal \boldsymbol{x}(t_i))^2}, \qquad (11)$$

as a quantitative measure. Substantial deviations from RMSE value may signal the presence of anomalous observations. Here, we suggest considering the $\pm 2 \times$ RMSE confidence interval over the 3-sigma counterpart, owing to the historical process variability (similar to historical mean approach for outlier detection) preventing to capture process trends accurately using simple model structure. The use of $\pm 3 \times \text{RMSE}$ could lead to minimal outlier detection, rendering the approach nonviable.

3. CASE STUDY

Oil refineries represent industrial complexes integrated with energy and material streams, where timely detection of anomalies improves the process sustainability and prevents environmental hazards.

3.1 Process Description

The alkylation process is an important step in refining operations, converting C_3–C_4 olefins into high-octane isoparaffins through a reaction with isobutane. These isoparaffins are critical for producing clean gasoline. The process utilizes cooling and mixing of the reactants before a catalytic reaction in the presence of sulfuric acid as:

$$\text{C−C=C−C (olefin)} \xrightarrow{\text{H}_2\text{SO}_4} \text{C−C−C}^+\text{−C} \qquad (12)$$

This operation spans various units, including reactors, distillation columns, compressors, and heat exchangers, each maintaining precise conditions, such as the isobutane-to-olefin ratio, for optimal reaction efficiency and product quality (Speight, 2020; Hommeltoft, 2001). We monitor every measured variable across all operational components. This extensive collection captures the full spectrum of influences on the process — mainly the concentration of isobutane $y(t)$. However, the intricacies of alkylation mean that upstream and downstream processes, as well as ancillary operations, can exert significant, albeit less direct, impacts on the measurements.

3.2 Problem Definition

In the current industrial practice, anomaly detection heavily relies on laboratory sampling and the expertise of the operating personnel. The integration of automated algorithms into the monitoring process presents an opportunity to significantly reduce the operational burden by alerting operators only when misbehavior occurs.

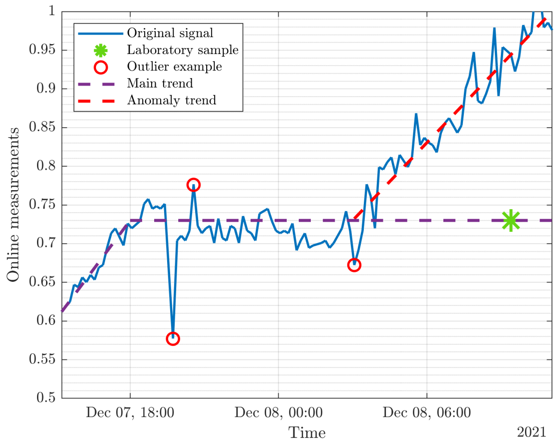

Fig. 1. An example of disruptions present on the concentration of isobutane $y(t)$.

We investigate a comprehensive industrial dataset comprising over 500 process variables. In the pre-processing phase, we employed several steps to ensure data quality and consistency: data cleaning, to identify and correct or remove any errors, inconsistencies, or inaccuracies; data standardization, to convert all data into a uniform range, facilitating comparison and analysis; and variable removal, to eliminate linear dependencies among the variables, simplifying the model without sacrificing its predictive power. These steps resulted in the selection of 377 independent entities each with 15,906 points (defining process variables matrix $\mathbf{X} \in \mathbb{R}^{377\times 15,906}$). The variable of interest, $y(t)$, is the concentration of isobutane, measured by an online analyzer at 15-minute intervals.

Within the dataset, we distinguish between two distinct disruptions: additive outliers, and slow gradual drifts, and as illustrated in Fig. 1. Additive outliers are characterized by abrupt, significant deviations from the recent trend with the subsequent restoration within the following observations. Slow gradual drifts represent a more subtle deviation, resulting in the most challenging-to-detect series of disruptions. In the scope of our work, we classify the sudden and isolated signal changes (additive outliers) as “outliers” and measurements that deviate from the main trend (slow gradual drifts) are tagged as “anomalies”, due to their more extended influence and potential to necessitate manual sensor recalibration.

4. RESULTS

A fair assessment of anomaly detection methods depends on the real occurrence of anomalies in the dataset. As this knowledge is unavailable and plant operators alone cannot reliably identify anomalous behaviour themselves even in historical data, we adopted the MA smoother (6) of order 7 to derive the ground truth (GT). This is grounded in the smoother’s ability to reduce noise while preserving significant shifts in the data, potentially indicative of anomalies. The 7$^{\text{th}}$ order was determined empirically, balancing the need for smoothing to capture genuine process variations without overly diluting potential anomalies. This approach identified 306 outliers among the 15,906 measurements.

4.1 Indicators Training

For outlier detection using filter-based approaches (Section 2.2), we systematically compare the efficacy of different window sizes applied to the signal — maximizing anomaly detection accuracy while minimizing false positives due to normal signal variations. We used cross-validation approach for this training, having 1,881 training and 750 testing measurements, respectively. This approach resulted in selection of an order of 7 for the MA filter, as described in Eq. (5). This choice aligns with the GT definition. Our analysis revealed that approximately 7.07% of the data points were identified as outliers within the training dataset. Within the testing set, containing 750 measurements, the outlier detection rate slightly increased to 7.87%. In practice, without the GT knowledge, the order would need to be determined by stepwise tuning to optimize the filter’s performance under varying process conditions.

Regression models were then applied, predicting isobutane concentration $\hat{y}(t) = \boldsymbol{\beta}^{\mathsf{T}}\boldsymbol{x}(t)$ and its time difference Eq. (8) after preprocessing the dataset, leveraging all techniques discussed in Section 2.3, including the MCD method (see Section 2.1) to ensure a clean basis for model training. The partitioning of the dataset into training and testing datasets followed a random indexing approach, maintaining an 80/20 split. The PCA, used for dimensionality reduction, identified seven principal components as the most informative, capturing around 62% of the variance. Additional components offered a minimal increase in total explained variance, indicating saturation in capturing dataset variability.Application of LASSO further refined our model selection process, using a defined threshold value.

The trained prediction models resulted in RMSE values spanning from 0.4031 (LASSO) to 0.6655 (PCA+LASSO), with OLS and PCR yielding intermediate values of 0.5243 and 0.6322, respectively. All methods achieved RMSE levels within $\pm 5\%$ of the isobutane concentration, aligning with the industrial standard for a representative prediction model. A detailed examination of the variables selected by these models highlighted the predictive superiority of features chosen by LASSO — namely, the concentration of n-butane in the recycle stream, the concentration of olefins in the feed, and the pressures within the olefin and recycle streams. Although OLS selected a reasonable array of process variables, PCA’s selection was deemed less appropriate due to its focus on less informative features such as compressor discharge, vibrations, and ventilation, a conclusion further supported by discussions with the industrial partners. This discrepancy could stem from PCA’s potential omission of relevant components, the nonlinear relationships between variables, or the high noise levels in the retained components. Thus, LASSO’s feature selection methodology proved more suitable for the dataset at hand, making it the preferred method for this analysis. Moreover, a comparison of the prediction models with

Table 1. Confusion matrix entries for implemented outlier detection methods.

Method	TP	TN	FP	FN
Historical mean (HM)	15601	0	306	0
HM of differences	15431	61	245	170
MA filter	14553	122	184	1048
MA predictor	14590	125	181	1011
Regression model	15175	15	291	426

laboratory measurements revealed notable similarities, warranting further investigation in future work.

For the model predicting the output difference, we compared each regression method and selected LASSO with the lowest value of RMSE = 0.037 (standardized data), with OLS and PCA+LASSO both being close second (RMSE = 0.038). The PCR approach resulted in RMSE = 0.178. The key process variables selected mostly align with those identified in prediction models (previous paragraph), with a predominant focus on variables measuring pressure and temperature differences.

4.2 Assessment of Applied Detection Methods

We assess the performance of trained outlier detectors by means of a confusion matrice. The correctly classified non-anomalous measurements are labeled as true positives (TP) while misclassifications as false positives (FP). The outliers being correctly identified are labeled as true negatives (TN) while misclassifications as false negatives (FN). Based on the ratio between TP and TN values in the GT labels, we deal with imbalanced datasets. Consequently, calculating the standard performance metrics (e.g., Recall, Precision, or Accuracy) does not result in a conclusive assessment. The metrics shown in Tab. 1 indicate the comparison of correctly classified TN and TP measurements with regards to the misclassifications (FP and FN).

The Historical Mean approach did not perform well, as it failed to detect any outliers (TN = 0). This was expected due to the varying range of the isobutane concentration over multiple operation points within the available data range. The Historical Mean of Differences correctly detected 97 % of normal operation points (TP = 15,431); however, in the case of outliers, it failed to identify over 80 % of the GT values (TN = 61). Using the MA, we were able to detect 40 % of the overall outliers (TN = 122); the filter has a more limited perspective and operates solely on historical data without foresight into upcoming measurements. The MA predictor increased the efficacy of outlier detection compared to the MA filter while reducing false predictions (FP from 1,048 to 1,011; FN from 184 to 181). The TP and TN predictions improved as well (TP from 14,553 to 14,590; TN from 122 to 125). These subtle adjustments achieved the best distribution of correctly classified data and they reflect the improved utilization of latent information, contributing to a marginal increase in overall accuracy. To evaluate the efficacy of the regression model, we used the $\pm 2 \times$ RMSE metric, correctly identifying 15 outliers (TN). Unlike filter-based approaches, the prediction model captures different dynamics, emphasizing its unique perspective on

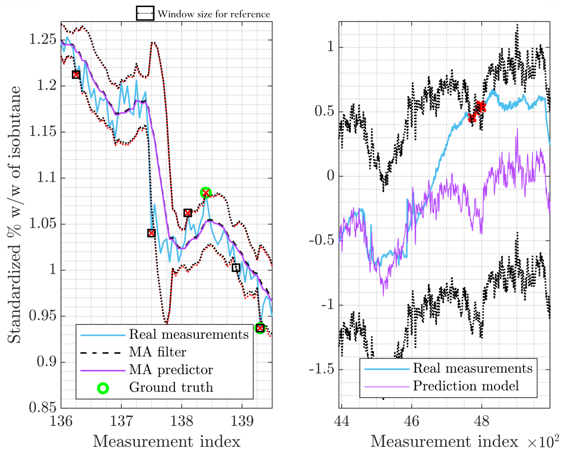

Fig. 2. Comparison of the outlier detection (left) and anomaly detection (right), where green circles represent the GT, red crosses/black squares represent the classified outliers.

anomalies beyond the expected range of the dependent variable. All identified outliers (TN+FN) indicate slow, gradual drifts (defined as anomalies) that require sensor calibration.

In Fig. 2, the performance of the MA filter and predictor is visualized for a selected testing period (left-hand plot). The MA filter's mean and confidence interval are depicted with black dashed lines ((5) and (4)), while the MA predictor's characteristics are shown in purple and red dashed lines, respectively (Eqs. (7) and (4)). This comparison highlights the MA predictor's improved accuracy in identifying anomalies (red crosses) compared to the MA filter (black squares), as it holds additional information about other independent variables $\mathbf{X}(t)$. This method has the potential to identify outliers that escaped detection by the MA smoother, however, it may result in more FNs (instances where normal behavior is flagged as anomalous), as indicated by green circles for the ground truth. The right-hand plot in Fig. 2 demonstrates the strength of regression models, specifically LASSO (purple), in identifying broader, systemic anomalies, such as slow gradual drifts in concentration values. Importantly, the effectiveness of such anomaly detection is dependent on the quality of the regression model structure. This comparison reveals the potential for each method in enhancing anomaly detection within the dataset.

5. CONCLUSION

In our study, we aimed to enhance outlier and anomaly detection in process variables by integrating moving-horizon filters and regression-based prediction methods. The MA predictor emerged as particularly effective within our industrial dataset, identifying 40 % of outliers. This approach capitalized on incorporating additional process variables and their time-differenced relationships during model training, showcasing its utility in complex industrial environments. However, the detection rate also highlights the challenges in capturing the full spectrum of outliers, underlining the need for further methodological advancements. While the MA predictor method demonstrated its merit, regression models revealed their strength in identifying long-term anomalies. Our future research will explore the potential of regression approaches in detecting anomalous measurements, including long-term level-shifts and slow gradual-shifts. Understanding static and dynamic aspects, considering linear and potentially non-linear sensor behaviours, will be crucial for further advances. We will focus on non-linear transformations and dynamic sensor characteristics, contributing to the development of a comprehensive and robust anomaly detection framework.

REFERENCES

Afanasyev, D.O. and Fedorova, E.A. (2019). On the impact of outlier filtering on the electricity price forecasting accuracy. *Applied Energy*, 236, 196–210.

Antonini, M., Pincheira, M., Vecchio, M., and Antonelli, F. (2023). An adaptable and unsupervised tinyml anomaly detection system for extreme industrial environments. *Sensors*, 23(4).

Blázquez-García, A., Conde, A., Mori, U., and Lozano, J.A. (2021). A review on outlier/anomaly detection in time series data. *ACM Comput. Surv.*, 54(3).

Carter, K.M. and Streilein, W.W. (2012). Probabilistic reasoning for streaming anomaly detection. In *IEEE Statistical Signal Processing Workshop*, 377–380.

Ester, M., Kriegel, H.P., Sander, J., and Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In E. Simoudis, J. Han, and U.M. Fayyad (eds.), *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*, 226–231. AAAI Press. CiteSeerX 10.1.1.121.9220.

Hommeltoft, S.I. (2001). Isobutane alkylation: Recent developments and future perspectives. *Applied Catalysis A: General*, 221(1), 421–428.

Iglesias Vázquez, F., Hartl, A., Zseby, T., and Zimek, A. (2023). Anomaly detection in streaming data: A comparison and evaluation study. *Expert Syst. Appl.*, 233, 120994.

Inoue, J., Yamagata, Y., Chen, Y., Poskitt, C.M., and Sun, J. (2017). Anomaly detection for a water treatment system using unsupervised machine learning. In *IEEE International Conference on Data Mining Workshops*, 1058–1065.

Jin, Z., Zhao, J., Ding, L., Chakrabarti, S., Gryazina, E., and Terzija, V. (2022). Power system anomaly detection using innovation reduction properties of iterated extended kalman filter. *Int. J. Electr. Power Energy Syst.*, 136, 107613.

Oppenheim, A.V. (1999). *Discrete-time signal processing*. Pearson Education India.

Pearson, K. (1901). Liii. on lines and planes of closest fit to systems of points in space. *Lond.Edinb.Dubl.Phil.Mag*, 2(11), 559–572.

Raza, A., Tran, K.P., Koehl, L., and Li, S. (2023). Anofed: Adaptive anomaly detection for digital health using transformer-based federated learning and support vector data description. *Eng. Appl. Artif. Intell.*, 121, 106051.

Roberts, S.W. (1959). Control chart tests based on geometric moving averages. *Technometrics*, 1(3), 239–250.

Rousseeuw, P. and Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41, 212–223.

Santosa, F. and Symes, W.W. (1986). Linear inversion of band-limited reflection seismograms. *SIAM J. Sci. Comput.*, 7(4), 1307–1330.

Schmidl, S., Wenig, P., and Papenbrock, T. (2022). Anomaly detection in time series: A comprehensive evaluation. *Proc. VLDB Endow.*, 15(9), 1779–1797.

Speight, J.G. (2020). *The refinery of the future*. Elsevier.

Student (1908). The probable error of a mean. *Biometrika*, 6(1), 1–25.

Tziolas, T., Papageorgiou, K., Theodosiou, T., Papageorgiou, E., Mastos, T., and Papadopoulos, A. (2022). Autoencoders for anomaly detection in an industrial multivariate time series dataset. *Eng. Proc.*, 18(1).

Wang, K. and Liu, X. (2021). An anomaly detection method of industrial data based on stacking integration. *Journal on Artificial Intelligence*, 3(1), 9–19.

Yoon, Y.S., Jeong, W., Kim, J., Seok, M., Park, J., Bae, J., Lee, K., and Lee, J.H. (2022). Development of inferential sensor and real-time optimizer for a vacuum distillation unit by recurrent neural network modeling of time series data. *Comput. Chem. Eng.*, 168, 108039.