Slovak Society of Chemical Engineering
Institute of Chemical and Environmental Engineering
Slovak University of Technology in Bratislava

PROCEEDINGS

52nd International Conference of the Slovak Society of Chemical Engineering SSCHE 2026

Hotel SOREA TRIGAN

Štrbské Pleso, Slovakia

May 26 - 29, 2026

Editors: Assoc. prof. Mário Mihaľ

ISBN: 978-80-8208-177-3, EAN: 9788082081773

Published by the Faculty of Chemical and Food Technology, Slovak University of Technology in Bratislava in Slovak Chemistry Library for the Institute of Chemical and Environmental Engineering; Radlinského 9, 812 37 Bratislava, 2026

Arbetová, P., Fáber, R., Ľubušký, K., Paulen, R.: A Practical Approach to Anomaly Detection in Industrial Data, Editors: Mihaľ, M., In *52nd International Conference of the Slovak Society of Chemical Engineering SSCHE 2026*, Štrbské Pleso, Slovakia, 2026.

A Practical Approach to Anomaly Detection in Industrial Data

Patrícia Arbetová[1], Rastislav Faber[1], Karol L'ubušký[2], and Radoslav Paulen[1]

[1]*Slovak University of Technology in Bratislava, 812 37 Bratislava, Slovakia*
[2]*Slovnaft, a.s., 824 12 Bratislava, Slovakia*
email: xarbetova@stuba.sk

Industrial processes are usually continuously monitored using two types of measurements, online sensors (low-fidelity measurements, LF) and laboratory analyses (high-fidelity measurements, HF). Low-fidelity measurements provide frequent but potentially noisy, inaccurate, or fault-affected data, while high-fidelity measurements are less frequent but are considered to accurately represent the true process values. Comparing LF and HF data may reveal big discrepancies between these measurements, which can indicate the presence of anomalies. Anomalies are patterns in a dataset that deviate from general trends and expected behaviour [1]. Their detection is critical for early identification of sensor faults, maintaining data quality and reliability of models. Corrupted measurements may result in failure to achieve desired process targets.

For the purpose of this work, data are used from the alkylation unit at Slovnaft, a.s. Since the available dataset contains a large number of process variables, we select most relevant features using correlation analysis and Partial Least Squares (PLS) regression. We use the resulting feature subset as input to anomaly detection models. To evaluate models performance, we compute confusion matrices using ground-truth anomalies and predicted anomalies. A multi-fidelity model, described in [2], is used to define the ground-truth anomalies. This model uses a simple residual-correction approach combining a dynamic LF predictor with a Gaussian Process correction learned from HF data. To define predicted anomalies we use Density-Based Spatial Clustering of Applications with Noise (DBSCAN), Ordinary Least Squares (OLS) regression, and Gaussian Process Regression (GPR). We optimize anomaly detection models parameters using the Nelder-Mead method, targeting the best balance between detection sensitivity and false alarm rate.

The results indicate that each of the considered methods provides a different perspective on the data structure and types of deviations. When comparing regression methods, OLS generally achieves better performance than GPR, suggesting that more complex nonlinear methods do not provide additional benefits for this type of data. The proposed framework contributes to improving measurement reliability and may support early detection of sensor faults.

Key words: Anomaly detection, Confusion matrix, Optimization, Soft sensors, Laboratory analyses

This work is funded by the Slovak Research and Development Agency under the project APVV-24-0007, and by the Scientific Grant Agency of the Slovak Republic under the grant 1/0263/25. We acknowledge the financial support provided by the European Union through the NextGenerationEU programme under the Recovery and Resilience Plan for Slovakia, projects no. 09I01-03-V05-00002, no. 09I01-03-V04-00024, and the Early Stage Grant no. 09I03-03-V05 (23-04-06-A).

References

[1] Singh, M., Singh, M., and Kaur, S. (2019). Issues and challenges in DNS based botnet detection: A survey. *Computers & Security*, 86, 28–52. doi: 10.1016/j.cose.2019.05.019.

[2] Fáber, R., Vaccari, M., Bacci di Capaci, R., L'ubušký, K., Pannocchia, G., and Paulen, R. (2025). Soft-Sensor-Enhanced Monitoring of an Alkylation Unit via Multi-Fidelity Model Correction. *Systems and Control Transactions*, 1389–1395. doi: 10.69997/sct.125527.