# Hybrid Technique for Sensor Fault Diagnosis in Natural-Gas Pipelines

Khadija Shaheen*, Apoorva Chawla*, Ferdinand Evert Uilhoorn†, Pierluigi Salvo Rossi*,
*Dept. Electronic Systems, Norwegian University of Science and Technology, Trondheim, Norway
†Dept. Gas Engineering, Warsaw University of Technology, Warsaw, Poland
Email: shaheen.khadija@ntnu.no; apoorva.chawla@ntnu.no; ferdinand.uilhoorn@pw.edu.pl; salvorossi@ieee.org

***Abstract*—Sensor fault diagnosis is crucial for the safe and reliable monitoring of natural gas pipelines. Even though model-based techniques have remarkable performance, they still face significant challenges such as non-linearity and external disturbances. In contrast, data-driven methods struggle to deal with unknown disturbances and uncertainties. To address these challenges, this paper proposes a hybrid fault diagnostic scheme that leverages the benefits of both model-based and data-driven techniques. First, it introduces a model-based fault estimation technique based on a partial-distributed ensemble Kalman filter (EnKF). It reduces the computational complexity by separating the non-linear computation from the distributed architecture. Furthermore, a data-driven method is developed to detect and isolate sensor faults based on the Gaussian mixture model (GMM). Experimental results considering different sensor fault conditions confirm the effectiveness of the proposed method.**

***Index Terms*—Hybrid, model-based, data-driven, fault detection and isolation, ensemble Kalman filter, Gaussian mixture model, natural gas pipeline.**

## I. Introduction

Sensor-based monitoring is crucial for ensuring system safety and reducing the risk of catastrophic failures in natural gas pipelines [1], [2]. However, sensors are susceptible to several errors and faults, making sensor-fault detection, isolation and accommodation (SFDIA) vital for efficient pipeline operation. SFDIA methods are generally classified into model-based methods [3], [4] and data-driven methods [5], [6]. The model-based techniques are effective for fault diagnosis when the system model is well-defined. However, their performance is compromised by factors like model non-linearity, external disturbances, and high dimensionality. Conversely, the data-driven methods rely on input-output mapping but face challenges with unknown disturbances and uncertainties. Therefore, combining model-based and data-driven methods can mitigate potential disturbances and enable fast and accurate fault detection [7].

Based on these insights, this paper presents a novel hybrid approach that integrates model-based and data-driven methods for sensor fault diagnosis in natural gas pipelines. A partially distributed ensemble Kalman filter (EnKF) is proposed to reduce computational complexity by separating the nonlinear computations from the local filters and delegating them to the main filter. The main filter handles time updates and information fusion, while the local filters manage measurement updates, reducing computational load. Further, a fault diagnosis approach based on the Gaussian mixture model (GMM) is developed to improve the estimation performance in the presence of sensor faults. GMMs capture complex data patterns and identify significant anomalies by incorporating features from the partially distributed architecture [8]. This integration significantly enhances the fault diagnosis capabilities, outperforming purely model-based or data-driven approaches.

The rest of the paper is organized as follows: Section II explains the system model; the proposed technique is developed in Section III; Section IV illustrates the simulation results, and conclusions are given in Section V.

## II. System Model

Natural gas pipelines under transient flow can be mathematically represented by a set of hyperbolic partial differential equations (PDEs) as

$$\frac{\partial \boldsymbol{x}}{\partial t} = -\boldsymbol{A}(\boldsymbol{x})\frac{\partial \boldsymbol{x}}{\partial s} - \boldsymbol{\zeta}(\boldsymbol{x}) \,, \tag{1}$$

where $s \in [0, L]$, $t \in [0, t_f]$ denote the space and the time with $L$, $t_f$ as the pipeline length and the time span, respectively [9], and the state vector $\boldsymbol{x} = [p, \dot{m}, T]^{\mathrm{T}}$ comprises of pressure ($p$), flow ($\dot{m}$), and temperature ($T$). The coefficient matrix $\boldsymbol{A}(\boldsymbol{x}) \in \mathbb{R}^{3\times3}$ and the vector $\boldsymbol{\zeta}(\boldsymbol{x}) \in \mathbb{R}^{3\times1}$ which represent the non-linear thermodynamic transformations are given in [9].

Further, the numerical method of lines (5-point, 4th-order finite difference method) is used to convert the system of PDEs in (1) into ordinary differential equations (ODEs) as

$$\frac{d\boldsymbol{x}(t)}{dt} = \boldsymbol{A}(\boldsymbol{x})\boldsymbol{D}\boldsymbol{x}(t) - \boldsymbol{\zeta}(\boldsymbol{x}, t) \,, \tag{2}$$

where $\boldsymbol{\zeta}(\boldsymbol{x}, t) \in \mathbb{R}^{3n\times1}$, $\boldsymbol{A}(\boldsymbol{x}) \in \mathbb{R}^{3n\times3n}$ represent the assembled vector and matrix, respectively. The state vector $\boldsymbol{x}$ becomes $\boldsymbol{x}(t) = [p_1(t), \ldots, p_n(t), \dot{m}_1(t), \ldots, \dot{m}_n(t), T_1(t), \ldots, T_n(t)]^T \in \mathbb{R}^{3n\times1}$ and the computational matrix $\boldsymbol{D}$ for spatial discretization is presented in [9]. The resulting ODEs are further solved using the 4th-order Runge-Kutta method to obtain the state-space model [4], which is subsequently used for sensor fault diagnosis in natural gas pipelines.

## III. Proposed Design

Fig. 1 illustrates all the steps involved in the proposed design: (1) the sensor measurements are initially grouped into several subgroups; (2) state estimation begins with the

[0]This work was partially supported by the Research Council of Norway under Project 311902 (SIGNIFY) within the framework IKTPLUSS.

2024 IEEE SENSORS | 979-8-3503-6351-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/SENSORS60989.2024.10784789

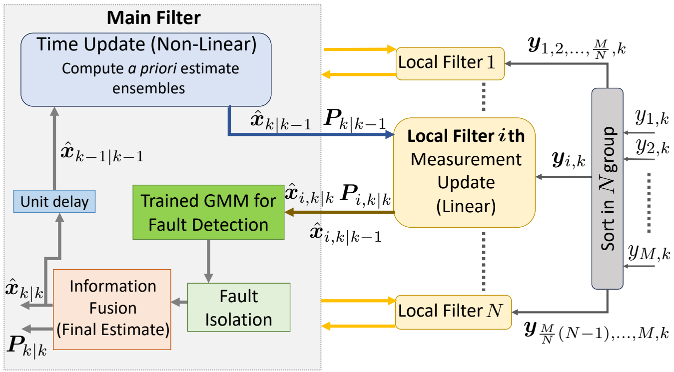


Fig. 1: Proposed architecture

non-linear time update in the main filter; (3) the local filters perform the linear measurement updates; (4) the fault detection mechanism based on the GMM is used to analyze the local state variance; (5) faulty state estimates are corrected with the non-faulty estimates; (6) the global estimates are finally computed in the information mixture.

Within the partial-distributed framework, the state space model for the main and the $i$th local filter can be given as

$$\boldsymbol{x}_k = \boldsymbol{f}(\boldsymbol{x}_{k-1}, \boldsymbol{u}_{k-1}) + \boldsymbol{w}_k \ ,$$
$$\boldsymbol{y}_{i,k} = \boldsymbol{h}_i(\boldsymbol{x}_k, \boldsymbol{u}_k) + \boldsymbol{v}_{i,k} \ , \tag{3}$$

where $\boldsymbol{f}(\cdot,\cdot) : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \to \mathbb{R}^{n_x}$ represents the nonlinear flow model. At the $k$th time step, the measurement vector for the $i$th local filter (where $i = 1, 2, \ldots, N$) and the state vector are denoted by $\boldsymbol{y}_{i,k} \in \mathbb{R}^{n_y \times 1}$ and $\boldsymbol{x}_k \in \mathbb{R}^{n_x \times 1}$, respectively. The input vector $\boldsymbol{u}_{k-1} \in \mathbb{R}^{n_u \times 1}$, given by $\boldsymbol{u}_{k-1} = [\boldsymbol{u}_{in}^{\mathrm{T}} \boldsymbol{u}_{bc,k-1}^{\mathrm{T}}]^{\mathrm{T}}$ contains both initial and boundary conditions. In our technique, we assume linear measurement model, i.e., $\boldsymbol{h}_i(\boldsymbol{x}_k, \boldsymbol{u}_k) = \boldsymbol{H}_i \boldsymbol{x}_k$, aiming to reduce the dimension of the state vector based on the dimension of each subgroup of measurements. Additionally, the process and the measurement noises are represented by $\boldsymbol{w}_k \in \mathbb{R}^{n_x \times 1} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_k)$ and $\boldsymbol{v}_{i,k} \in \mathbb{R}^{n_y \times 1} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}_{i,k})$, respectively.

The partial-distributed filtering architecture performs the non-linear state estimation as follows.

**Step 0: Initialization.** Initialize the state estimate $\hat{\boldsymbol{x}}_{0|0}$ based on the specific use case requirements. Define distinct subgroups of sensor measurements and the number of local filters.

**Step 1: Main filter.** An ensemble of samples of size $N_e$, $\{\hat{\boldsymbol{x}}_{k-1|k-1}^{(j)}, 1 \le j \le N_e\}$, is generated to represent the distribution $p(\boldsymbol{x}_{k-1}|\mathbb{Y}_{k-1})$, where $\mathbb{Y}_{k-1} = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_{k-1}\}$. The samples $\{\boldsymbol{w}_k^{(j)}, \ 1 \le j \le N_e\}$ are drawn from the Gaussian distribution $\mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_k)$. Using these ensembles, the *a priori* ensemble $\{\hat{\boldsymbol{x}}_{k|k-1}^{(j)}, 1 \le j \le N_e\}$, is produced as

$$\hat{\boldsymbol{x}}_{k|k-1}^{(j)} = \boldsymbol{f}\left(\hat{\boldsymbol{x}}_{k-1|k-1}^{(j)}, \boldsymbol{u}_{k-1}^{(j)}\right) + \boldsymbol{w}_k^{(j)} \ . \tag{4}$$

Next, the *a priori* state estimate and covariance matrix are calculated as $\hat{\boldsymbol{x}}_{k|k-1} = \frac{1}{N_e}\sum_{j=1}^{N_e} \hat{\boldsymbol{x}}_{k|k-1}^{(j)}$ and $\boldsymbol{P}_{k|k-1} = \frac{1}{N_e - 1}\boldsymbol{E}_{k|k-1}^{x}\left(\boldsymbol{E}_{k|k-1}^{x}\right)^T$, respectively, with $\boldsymbol{E}_{k|k-1}^{x} = \left[(\hat{\boldsymbol{x}}_{k|k-1}^{(1)} - \hat{\boldsymbol{x}}_{k|k-1}), \ldots, (\hat{\boldsymbol{x}}_{k|k-1}^{(N_e)} - \hat{\boldsymbol{x}}_{k|k-1})\right]$. These *a priori* estimates are then shared with $N$ linear local filters.

**Step 2: Local filters.** The linear Kalman filters are employed for measurement updates. The linear measurement update of the $i$th local filter is performed as

$$\boldsymbol{K}_{i,k} = \boldsymbol{P}_{k|k-1}\boldsymbol{H}_i^T\left(\boldsymbol{H}_i\boldsymbol{P}_{k|k-1}\boldsymbol{H}_i^T + \boldsymbol{R}_{i,k}\right)^{-1},$$
$$\hat{\boldsymbol{x}}_{i,k|k} = \hat{\boldsymbol{x}}_{k|k-1} + \boldsymbol{K}_{i,k}\left(\boldsymbol{y}_{i,k} - \boldsymbol{H}_i\hat{\boldsymbol{x}}_{k|k-1}\right),$$
$$\boldsymbol{P}_{i,k|k} = \left(\boldsymbol{I} - \boldsymbol{K}_{i,k}\boldsymbol{H}_i\right)\boldsymbol{P}_{k|k-1}. \tag{5}$$

**Step 3: Fault detection and isolation.** The state-variance vector $\boldsymbol{\xi}_k \in \mathbb{R}^{n_x \times 1}$ is employed for fault detection, whose $\ell$th element is defined as $\xi_k^{(\ell)} = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{x}_{i,k|k}^{(\ell)} - \frac{1}{N}\sum_{i=1}^{N}\left(\hat{x}_{i,k|k}^{(\ell)}\right)\right)^2$. This metric is effective only when a specific grouping of sensor measurements is utilized, i.e., a unique/non-repetitive set of sensor measurements is assigned to each local filter. For $M$ sensors and $N$ local filters, each local filter contains $M/N$ sensor measurements, where $M/N$ must be an integer. This grouping allows each local filter to generate an independent state estimate based on its assigned subset of sensor measurements [10]. A faulty measurement only affects the estimate of its respective local filter, while the remaining local filters generate non-faulty estimates. Consequently, the local state variance measures the variations in these independent local state vector estimates in the presence of faults.

To improve the evaluation of the state variance for diagnosing sensor faults, we utilize the GMM that estimates a combination of multiple Gaussian components for each element of the local state variance. The GMM is trained offline under non-faulty conditions using the expectation-maximization (EM) algorithm [8]. The algorithm for the offline training of the GMM for the $\ell$th entry of the error vector $\boldsymbol{\xi}_k$ consists of the following steps. (1) The normalization is performed as $\tilde{\xi}_k^{(\ell)} = \xi_k^{(\ell)} - \frac{1}{N}\sum_{\ell=1}^{n_x}\left(\xi_k^{(\ell)}\right)$, where $\tilde{\xi}_k^{(\ell)}$ represents the $\ell$th entry of the normalized vector $\tilde{\boldsymbol{\xi}}_k \in \mathbb{R}^{n_x \times 1}$. (2) The parameters for the $r$ Gaussian component, where $r = 1, \ldots, R$, including the mixing coefficient ($\pi_r^{(\ell)}$), mean ($\mu_r^{(\ell)}$), and standard deviation ($\sigma_r^{(\ell)}$) are initialized. (3) **E-Step:** For each time instant $k = 1, \ldots, N_k$ (where $N_k = t_f/\Delta t$ is the final time instant), the responsibilities are computed as $\gamma_{kr}^{(\ell)} = \frac{\pi_r^{(\ell)}\mathcal{N}(\tilde{\xi}_k^{(\ell)}|\mu_r^{(\ell)}, \sigma_r^{(\ell)})}{\sum_{m=1}^{R}\pi_m\mathcal{N}(\tilde{\xi}_k^{(\ell)}|\mu_m^{(\ell)}, \sigma_m^{(\ell)})}$, which represents the probability that the data point $\tilde{\xi}_k^{(\ell)}$ belongs to the $r$th Gaussian component. (4) **M-Step:** The parameters $\{\pi_r^{(\ell)}, \mu_r^{(\ell)}, \sigma_r^{(\ell)}\}_{r=1}^{R}$ are updated using the computed responsibilities as $\pi_r^{(\ell)} = \frac{1}{N_k}\sum_{k=1}^{N_k}\gamma_{kr}^{(\ell)}$, $\mu_r^{(\ell)} = \frac{\sum_{k=1}^{N_k}\gamma_{kr}^{(\ell)}\tilde{\xi}_k^{(\ell)}}{\sum_{k=1}^{N_k}\gamma_{kr}^{(\ell)}}$ and $\sigma_r^{(\ell)} = \frac{\sum_{k=1}^{N_k}\gamma_{kr}^{(\ell)}(\tilde{\xi}_k^{(\ell)} - \mu_r^{(\ell)})(\tilde{\xi}_k^{(\ell)} - \mu_r^{(\ell)})^T}{\sum_{k=1}^{N_k}\gamma_{kr}^{(\ell)}}$. Further, Steps (3) and (4) are repeated until convergence.

For online sensor fault diagnosis, the probability density function $p(\tilde{\xi}_k^{(\ell)}|\{\pi_r^{(\ell)}, \mu_r^{(\ell)}, \sigma_r^{(\ell)}\}_{r=1}^{R})$, obtained from GMM, is employed. An error metric $\boldsymbol{e}_k \in \mathbb{R}^{n_x \times 1}$ is used for detecting sensor faults such that its $\ell$th entry $e_k^{(\ell)}$ is defined as $e_k^{(\ell)} = \sum_{r=1}^{R}\pi_r^{(\ell)}\frac{1}{\sigma_r^{(\ell)}}\left(\mu_r^{(\ell)} - \tilde{\xi}_k^{(\ell)}\right)^2$. The fault is detected if the error metric $e_k^{(\ell)}$ exceeds a predefined threshold $\chi_\ell$. For fault isolation, it is crucial to determine the exact location of the faulty estimates. This can be achieved by examining the components of the error metric $\boldsymbol{e}_k$. Specifically, if the error metrics $e_k^{(\ell)}$, $e_k^{(p)}$ and $e_k^{(n)}$ surpass the threshold $\chi_\ell$, $\chi_p$ and $\chi_n$, respectively, it indicates faulty estimates corresponding to the sensor locations $\ell$, $p$, and $n$. Furthermore, the locations of the faulty estimates can be determined by the set $\mathcal{F} = \{l : e_k^{(\ell)} > \chi_\ell\}$. Identifying faulty estimate locations allows for their replacement with fault-free estimates determined through spatial interpolation [10]. Assuming a distributed spatio-temporal system with known non-faulty and faulty estimate locations, accurate estimates can be obtained through one-dimensional linear interpolation using neighboring non-faulty estimates across space.

**Step 4: Information fusion.** The final estimates are computed as $\boldsymbol{P}_{k|k}^{-1} = \sum_{i=1}^{N} \boldsymbol{P}_{i,k|k}^{-1}$, and $\hat{\boldsymbol{x}}_{k|k} = \boldsymbol{P}_{k|k} \sum_{i=1}^{N} \boldsymbol{P}_{i,k|k}^{-1} \hat{\boldsymbol{x}}_{i,k|k}$.

## IV. Simulation Results and Discussions

For numerical simulations, we analyze a high-pressure natural gas pipeline with sensors to measure pressure, temperature, and flow rate, as detailed in [10]. The pressure, temperature, and flow rate measurements $\boldsymbol{y}_{i,k}$ are generated by adding zero-mean additive Gaussian noise $\boldsymbol{v}_{i,k}$ having standard deviations $0.0005$ MPa, $1.5$ K and $2.5$ kgs$^{-1}$, respectively. We consider $M = 63$ sensors and $N = 3$ locals, with each local filter containing 21 sensor measurements. The ensemble size is $N_e = 120$ and GMM with three components ($R = 3$) is employed. The covariance matrices are initialization as $\boldsymbol{P}_{0|0} = \boldsymbol{I}_{63}$, $\boldsymbol{Q}_k = \sigma_{w,k}^2 \boldsymbol{I}_{63}$, with $\sigma_{w,k}^2 = 0.1\sigma_{v,i,k}^2$, $\sigma_{w,k}^2$ and $\sigma_{v,i,k}^2$ being process and measurement noise variances.

The estimation performance is evaluated in terms of the root mean square error (RMSE) as presented in Table I. The proposed method is compared against various baselines including model-based multi-sensor fault detection, isolation, and accommodation (MM-SFDIA) [10], fusing unscented Kalman filter (UKF) [3], [11] and classic UKF. Our proposed method performs similarly to fully distributed MM-SFDIA and fusing UKF, and exhibits higher estimation performance than the classic filter. Next, we evaluate the effectiveness of the proposed architecture in handling multiple simultaneous sensor faults, including bias and drift faults, as detailed in [10]. Fig. 2 demonstrates that the proposed technique and MM-SFDIA successfully detect and isolate the faulty sensors, providing reliable state estimates in the presence of faults due to their capability to handle multiple faults. Conversely, the fusion UKF fails to deliver reliable detection, isolation, and estimation.

Furthermore, we evaluate the detection performance of the proposed design using probabilities of detection and false alarm. The receiver operating characteristic (ROC) curves for various combinations of simultaneous weak bias and drift faults are shown in Fig. 3. These ROC curves demonstrate that the proposed technique achieves a high probability of detection and a low probability of false alarm compared to MM-SFDIA. Additionally, Fig. 4 presents the confusion matrices for the weak bias fault scenario to assess the isolation performance of our proposed architecture. The results demonstrate that the proposed design can accurately predict and isolate all faulty sensors, outperforming MM-SFDIA, which lacks GMM. Hence, the detection and isolation capabilities of the proposed technique have improved by incorporating GMM.

| Method | Pressure ($10^{-3}$MPa) | Temperature (K) | Flow Rate (kg/s) |
|---|---|---|---|
| Proposed | 0.4522 | 0.2372 | 0.8274 |
| MMSFDIA | 0.535 | 2.5235 | 2.9985 |
| Fusing UKF | 0.123 | 0.347 | 0.646 |
| UKF | 0.158 | 0.449 | 0.802 |

TABLE I: RMSE in the presence of measurement noise.

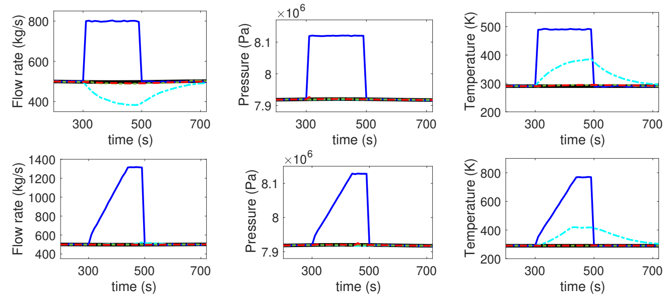


Fig. 2: State estimation in the presence of simultaneous faults. Actual/faulty values in black/blue, proposed method in red, MM-SFDIA in green, and fusing UKF in cyan.

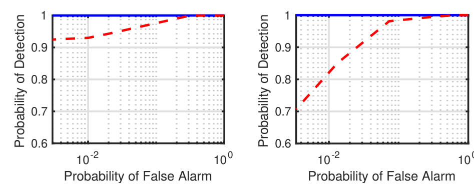


Fig. 3: ROC curves for the proposed technique (blue) and MM-SFDIA (red) under weak bias and weak drift faults.

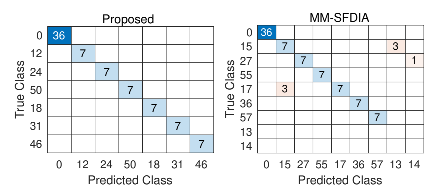


Fig. 4: Confusion matrix

## V. Conclusions

The proposed design introduced a novel hybrid technique for effective fault diagnosis in natural gas pipelines. It employed a novel partial-distributed EnKF framework, offloading the nonlinear computations from the local filters to the main filter, thus reducing computational complexity. Further, a new fault detection and isolation mechanism based on GMM is developed for handling multiple simultaneous sensor faults. This integration enhanced the fault detection and isolation capabilities while minimizing the computational overheads. Simulation results demonstrated that the proposed technique outperformed the popular benchmarks. Future work will focus on developing hybrid techniques for detecting simultaneous process-related (compressor) and sensor faults.

## REFERENCES


[1] M. A. Adegboye, W.-K. Fung, and A. Karnik, "Recent advances in pipeline monitoring and oil leakage detection technologies: Principles and approaches," *Sensors*, vol. 19, no. 11, p. 2548, 2019.

[2] A. Chawla, Y. Arellano, M. V. Johansson, H. Darvishi, K. Shaneen, M. Vitali, F. Finotti, and P. Salvo Rossi, "IoT-based monitoring in carbon capture and storage systems," *IEEE Internet of Things Magazine*, vol. 5, no. 4, pp. 106–111, 2022.

[3] A. Vafamand, B. Moshiri, and N. Vafamand, "Fusing unscented Kalman filter to detect and isolate sensor faults in DC microgrids with CPLs," *IEEE Transaction of Instrumentation and Measurement*, vol. 71, pp. 1–8, 2021.

[4] K. Shaheen, A. Chawla, F. E. Uilhoorn, and P. Salvo Rossi, "Sensor-Fault Detection, Isolation and Accommodation for Natural-Gas Pipelines Under Transient Flow," *IEEE Transactions on Signal and Information Processing over Networks*, 2024.

[5] H. Darvishi, D. Ciuonzo, E. R. Eide, and P. Salvo Rossi, "Sensor-fault detection, isolation and accommodation for digital twins via modular data-driven architecture," *IEEE Sensors Journal*, vol. 21, no. 4, pp. 4827–4838, 2021.

[6] H. Darvishi, D. Ciuonzo, and P. Salvo Rossi, "A machine-learning architecture for sensor fault detection, isolation, and accommodation in digital twins," *IEEE Sensors Journal*, vol. 23, no. 3, pp. 2522–2538, 2023.

[7] S. Chen, R. Yang, M. Zhong, X. Xi, and C. Liu, "A random forest and model-based hybrid method of fault diagnosis for satellite attitude control systems," *IEEE Transactions on Instrumentation and Measurement*, 2023.

[8] J. Luo and H. Zhu, "GMM-based distributed Kalman filtering for target tracking under cyber attacks," *IEEE Sensors Letters*, 2023.

[9] F. Uilhoorn, "Comparison of Bayesian estimation methods for modeling flow transients in gas pipelines," *Journal of Natural Gas Science and Engineering*, vol. 38, pp. 159–170, 2017.

[10] K. Shaheen, A. Chawla, F. E. Uilhoorn, and P. Salvo Rossi, "Model-based architecture for multi-sensor fault detection, isolation and accommodation in natural-gas pipelines," *IEEE Sensors Journal*, 2023.

[11] ——, "Model-based sensor-fault detection and isolation in natural-gas pipelines for transient flow," in *2023 IEEE SENSORS*. IEEE, 2023, pp. 1–4.