

Article

Data-Driven Fault Diagnosis in Water Pipelines Based on Neuro-Fuzzy Zonotopic Kalman Filters

Esvan-Jesús Pérez-Pérez ¹, Yair González-Baldizón ^{2,3}, José-Armando Fragoso-Mandujano ²,
Julio-Alberto Guzmán-Rabasa ⁴ and Ildeberto Santos-Ruiz ^{2,*}

¹ Research Group of Advanced Control Systems, Universitat Politècnica de Catalunya, Rambla Sant Nebridi 22, 08222 Terrassa, Spain; esvan.de.jesus.perez@upc.edu

² Turix-Dynamics Diagnosis and Control Group, I.T. Tuxtla Gutiérrez, Tecnológico Nacional de México, Carretera Panamericana S/N, Tuxtla Gutiérrez 29050, Mexico; yair.gonzalez@unach.mx (Y.G.-B.); jose.fm@tuxtla.tecnm.mx (J.-A.F.-M.)

³ Faculty of Accounting and Administration, Universidad Autónoma de Chiapas, Blvd. Belisario Domínguez Km. 1081, Tuxtla Gutiérrez 29050, Mexico

⁴ Turix-Dynamics Diagnosis and Control Group, I.T. Hermosillo, Tecnológico Nacional de México, Av. Tecnológico 115, Hermosillo 83170, Mexico; jaguzmanrabasa@gmail.com

* Correspondence: ildeberto.dr@tuxtla.tecnm.mx

Abstract: This work presents a data-driven approach for diagnosing sensor faults and leaks in hydraulic pipelines using neuro-fuzzy Zonotopic Kalman Filters (ZKF). The approach involves two key steps: first, identifying the nonlinear pipeline system using an adaptive neuro-fuzzy inference system (ANFIS), resulting in a set of Takagi–Sugeno fuzzy models derived from pressure and flow data, and second, implementing a neuro-fuzzy ZKF bench to detect pipeline leaks and sensor faults with adaptive thresholds. The learning phase of the neuro-fuzzy systems considers only fault-free data. Fault isolation is achieved by comparing zonotopic sets and evaluating a fault signature matrix. The method accounts for parametric uncertainty and measurement noise, ensuring robustness. Experimental validation on a hydraulic pipeline demonstrated high precision (up to 99.24%), recall (up to 99.20%), and low false positive rates (as low as 0.76%) across various fault scenarios and operational points.



Academic Editor: Leonardo Trujillo

Received: 15 October 2024

Revised: 11 December 2024

Accepted: 25 December 2024

Published: 30 December 2024

Citation: Pérez-Pérez, E.-J.; González-Baldizón, Y.; Fragoso-Mandujano, J.-A.; Guzmán-Rabasa, J.-A.; Santos-Ruiz, I. Data-Driven Fault Diagnosis in Water Pipelines Based on Neuro-Fuzzy Zonotopic Kalman Filters. *Math. Comput. Appl.* **2025**, *30*, 2. <https://doi.org/10.3390/mca30010002>

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: pipeline diagnosis; leak detection; water monitoring; neuro-fuzzy system; zonotopic Kalman filter; sensor faults

1. Introduction

Pipelines are essential for the distribution of liquids, gases, and hydrocarbons, facilitating the efficient and safe transport of these resources. However, sensor faults and leaks can cause infrastructure damage, economic losses, environmental contamination, and health risks. The lack of timely attention to these faults exacerbates their repercussions, highlighting the need to improve maintenance strategies [1]. Although significant efforts are made in preventive and corrective maintenance, these are not always sufficient to anticipate and mitigate all possible faults. Factors such as corrosion, aging, installation errors, natural events, and human activities present constant challenges [2]. Despite these hurdles, crafting a robust FD methodology remains an essential undertaking, albeit an intricate and costly one. This is particularly true in the context of multi-component systems where complexity scales exponentially, such as pressurized water pipelines.

Various fault detection methods in pipelines can be performed, including vibration analysis, leak detection based on interferometric fiber sensors, acoustic techniques, and the

filter diagonalization method, among others [3]. In [4], hybrid algorithms were used to identify faults using data fusion in two pharmaceutical manufacturing cases. The authors investigated the use of Raman spectra with collected data for multivariate statistical process control with a data-driven approach in Industry 4.0 contexts. Similarly, the methodology proposed in [5] reduced the amount of data needed for the learning process and achieved higher accuracy in the detection and classification of faults in photovoltaic systems. In [6], the authors exploited cooperation between cloud and edge servers. The cloud server employed a new approach that included feature selection based on a genetic algorithm to identify correlations and redundancies among features. A notable strategy was presented in [7], where an early time series classification algorithm was applied to support fault diagnosis in a hydraulic system with multiple interconnected components. The early classification model successfully balanced accuracy and timeliness.

In the field of hydraulic systems, ref. [8] presented a fault detection method based on machine learning to locate leak faults using data variation rates from flow meters and pressure sensors. The proposed method includes a delayed alert activation algorithm that periodically checks the time series of replacement flow to identify network leaks. In [9], the authors introduced evaluation metrics for classification using machine learning and deep learning to enhance the diagnosis of anomalies and defects in each component of the hydraulic system. Data were collected from IoT sensor data in the time domain. Similarly, the approach proposed by [10] combined a fault isolation system with an inference algorithm. The beneficial synergy effect was demonstrated and confirmed for the example of diagnosing a two-tank system.

In the realm of state observers, a fault diagnosis method using a discrete-time approach based on interval observers was proposed in [11] to enhance the integration of fault detection and isolation tasks. The results were applied to the urban sewer system of Barcelona. In [12], the authors addressed the problem of robust fault diagnosis in a four-tank hydraulic system in the presence of unknown input disturbances and sensor noise. Fault detection was investigated by generating residuals. Similarly, ref. [13] presented a hydraulic system as a case study. The methodology estimated faults in both sensors and actuators. The proposed approach included an observer that estimated induced faults, and the observer gains were calculated by solving linear matrix inequality constraints. An example using an extended Kalman filter was shown by [14], where an extended Kalman filter was used to detect and locate blockages in a pipeline. Additionally, observability and controllability in hydraulic networks were analyzed. Recently, ref. [15] proposed a methodology for fault diagnosis in hydraulic systems. Two features of the proposed technique were highlighted, namely, the reduction of redundant information and the improvement of diagnostic performance, with the reduced number of sensors simplifying the model and its complexity.

On the other hand, one of the challenges in pipeline systems is related to the uncertainty of parameters such as roughness, hydraulic friction, fluid turbulence, and measurement noise. Robust fault detection and identification schemes have been implemented to address these issues; for example, complex computational schemes using CNN [16], transformer neural network [17], and LSTM approaches have been explored [18]. In [19], the use of multilabel classification techniques was proposed to simultaneously predict pipe failures in water supply systems. Three models (discriminant analysis, logistic regression, and random forest) were analyzed, and different prediction periods were used. In [20], the authors proposed an adaptive constrained clustering approach for real-time fault detection in industrial systems. Their method employs a cyclical Adaptive Constrained Clustering algorithm that combines micro-clustering and macro-clustering steps. This dynamic process incorporates must-link and cannot-link constraints to classify nominal and non-nominal

working conditions, leveraging a mix of unlabeled data and limited labeled information. In [21], the authors developed a condition-based maintenance (CBM) scheme for detecting air leaks in pneumatic cylinders by monitoring piston speed profiles using Hall effect sensors. In addition, they employed machine learning techniques, including a majority voting ensemble of three models. The method achieved an accuracy of 88.4% in experimental tests. However, machine learning methods require prior knowledge of faults for detection and classification, which entails high computational costs. In the case of model-based methods, a refined mathematical model of the system is required, making it crucial to design methods that are robust to parametric uncertainty or mathematical model mismatches.

The present work introduces a data-driven method for detecting faults and leaks in hydraulic pipeline systems using neuro-fuzzy ZKF. Our approach begins with identification of the nonlinear pipeline system using an ANFIS, which generates Takagi–Sugeno fuzzy models from measured pressure and flow data. Note that this method does not require prior knowledge of the faults, as only fault-free data are used for training. Subsequently, a neuro-fuzzy ZKF bench is implemented to detect sensor faults and pipeline leaks through adaptive thresholds. The method accounts for parametric uncertainty arising from hydraulic friction and fluid turbulence as well as measurement noise, ensuring system robustness. Neuro-fuzzy ZKF observers are employed for fault detection, while fault isolation is achieved by comparing zonotopic sets and evaluating a fault signature matrix. Experimental validation in a hydraulic pipeline system demonstrates the method's high precision and reliability in fault and leak detection.

The rest of this document is organized as follows: Section 2 presents the preliminaries and notations used throughout the paper; Section 3 outlines the methodology for data-based fault diagnosis, describes the pipeline setup for industrial applications and system requirements, and details the ANFIS architecture as well as the design of the neuro-fuzzy ZKF observers; Section 4 presents the results obtained from the methodology under different operating conditions, using various metrics to evaluate the method's effectiveness; finally, Section 5 discusses the conclusions and outlines future work.

2. Preliminaries

Throughout this paper, $\mathbb{N}$ and $\mathbb{R}$ denote the set of natural and real numbers, respectively; $\mathbb{R}^m$ denotes the set of m -dimensional real vectors; the superscript T denotes the matrix transpose, whereas $\det(\cdot)$ denotes the matrix determinant; I_n is an $n \times n$ identity matrix; $\text{Tr}(\cdot)$ and $\text{Cov}(\cdot)$ denote the trace and covariance, respectively; and $\|\cdot\|_p$ indicates the p -norm, whereas $\|\cdot\|_F$ and $\|\cdot\|_\infty$ indicate the Frobenius and infinity norm, respectively.

Definition 1 (Zonotope [22]). A zonotope $\langle c, R \rangle \subset \mathbb{R}^m$ with center $c \in \mathbb{R}^m$ and generator matrix $R \in \mathbb{R}^{m \times p}$ is a particular form of polytope defined as the affine transformation of a unitary hypercube $[-1, 1]^p \subset \mathbb{R}^p$, as follows:

$$\langle c, R \rangle = \{c + R\xi : \|\xi\|_\infty \leq 1\}. \quad (1)$$

Definition 2 (Centered zonotope [23]). A centered zonotope is defined as $\langle R \rangle = \langle 0, R \rangle \subset \mathbb{R}^m$. For any permutation of the columns of R , the zonotope remains invariant.

Definition 3 (Interval hull [24]). The interval hull $b(R) \subset \mathbb{R}^{m \times m}$ of a given zonotope $\langle c, R \rangle \subset \mathbb{R}^m$ with $R \in \mathbb{R}^{m \times p}$ is the smallest aligned box in which $\langle c, R \rangle$ can be enclosed, i.e., $\langle c, R \rangle \subset \langle c, b(R) \rangle$. The term $b(R)$ is a diagonal matrix determined by the sum of the rows of R :

$$b(R)_{ii} = \sum_{j=1}^p |R_{ij}|. \quad (2)$$

Definition 4 (Reduction operator [24]). Given the zonotope $\langle c, R \rangle \subset \mathbb{R}^m$ with $c \in \mathbb{R}^m$ and $R \in \mathbb{R}^{m \times p}$, the reduction operator $\downarrow_q$, where $q \in \mathbb{N}$ specifies the maximum number of columns of R such that $\langle c, R \rangle \subset \langle c, \downarrow_q(R) \rangle$ consists of the following:

- Sort p columns of $R = [r_1, r_2, \dots, r_p]$ on the decreasing Euclidean norm, i.e., $\|r_j\| \geq \|r_{j+1}\|$.
- $\downarrow_q(R) = [R_{\downarrow}, b(R_{\uparrow})] \in \mathbb{R}^{m \times q}$, where $R_{\downarrow} = [r_1, \dots, r_{q-m}]$ and $R_{\uparrow} = [r_{q-m+1}, \dots, r_p]$

Definition 5 (F_W -radius [25]). Given the zonotope $\langle c, R \rangle \subset \mathbb{R}^m$ and the weighting matrix $W \in \mathbb{R}^{m \times m}$, s.t., $W = W^T \succ 0$. The F_W -radius is determined by computing the weighted Frobenius norm of R as follows:

$$\|\langle c, R \rangle\|_{F,W} = \sqrt{\text{Tr}(R^T W R)}, W = I_n. \quad (3)$$

Definition 6 (Covariation [25]). The covariation of a given zonotope $\langle c, R \rangle \subset \mathbb{R}^m$ is defined as $\text{Cov}(\langle c, R \rangle) = R R^T$.

Property 1. Representing the Minkowski sum and the linear image as $\oplus$ and $\odot$, respectively, s.t., M is a matrix of appropriate dimension; then, $\langle c_1, R_1 \rangle \oplus \langle c_2, R_2 \rangle = \langle c_1 + c_2, [R_1, R_2] \rangle$ and $M \odot \langle c, R \rangle = \langle M c, M R \rangle$.

3. Data-Driven Fault Diagnosis Approach for Pipeline System

This section presents a data-driven method for pipeline system fault diagnosis utilizing only the measurement data available during the operation process, as illustrated in Figure 1. The proposed approach employs ANFIS to model the inherent nonlinear dynamics of the pipeline. ANFIS combines the learning capability of neural networks with the interpretability of fuzzy logic, making it effective for identifying complex nonlinear behaviors in dynamic systems [26]. By training ANFIS with data from fault-free operation, the proposed method captures the pipeline's dynamics under normal conditions and generates Takagi–Sugeno fuzzy systems, allowing for accurate modeling of the pipeline's behavior across varying operating scenarios.

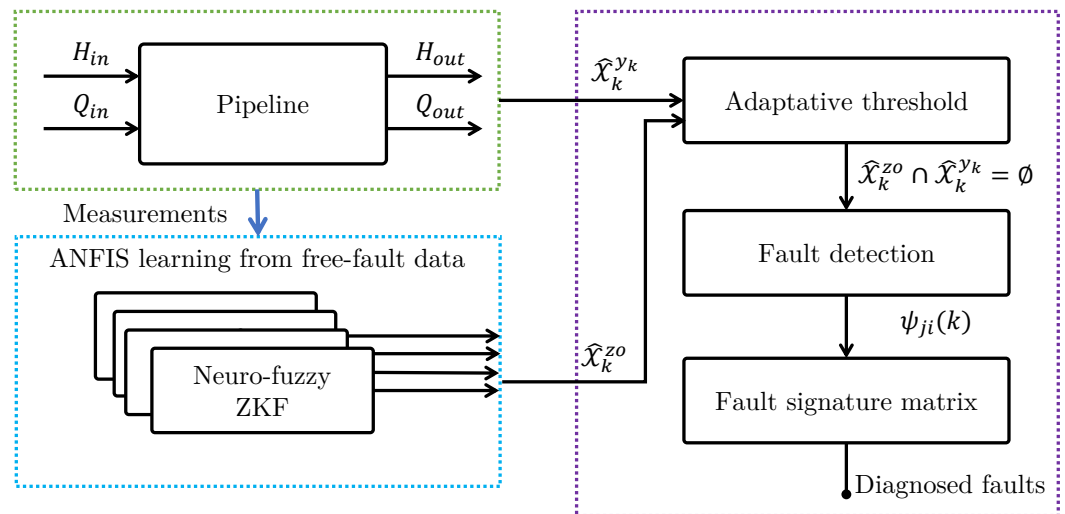


Figure 1. Diagram of the proposed method for fault diagnosis.

To improve fault detection, neuro-fuzzy ZKF with adaptive thresholds is applied. ZKF operates using deterministic set-based estimation, representing uncertainties and disturbances within bounded convex sets known as zonotopes. This approach ensures robust performance under uncertain conditions and provides a reliable mechanism for detecting faults in noisy environments [27]. The adaptive thresholds adjust dynamically to changes in operating conditions, enabling early detection of a variety of fault types with high sensitivity and precision.

Finally, fault isolation is achieved through a fault signature matrix analysis strategy that systematically correlates observed fault signatures with potential fault types. This efficient strategy facilitates the identification and classification of faults within the pipeline system, ensuring rapid and accurate fault diagnosis and enabling timely corrective actions.

3.1. Pipeline Setup for Industry Applications and System Requirements

The experimental setup consists of a hydraulic pipeline system designed to simulate and study various operational scenarios, including fault and leakage detection. The P&ID diagram of the system is shown in Figure 2. It features a single pipeline configuration that can be reconfigured into a branched network. The pipeline is constructed with 2" Schedule 80 PVC, providing an internal diameter of 0.052 m and sufficient durability for high-pressure applications.

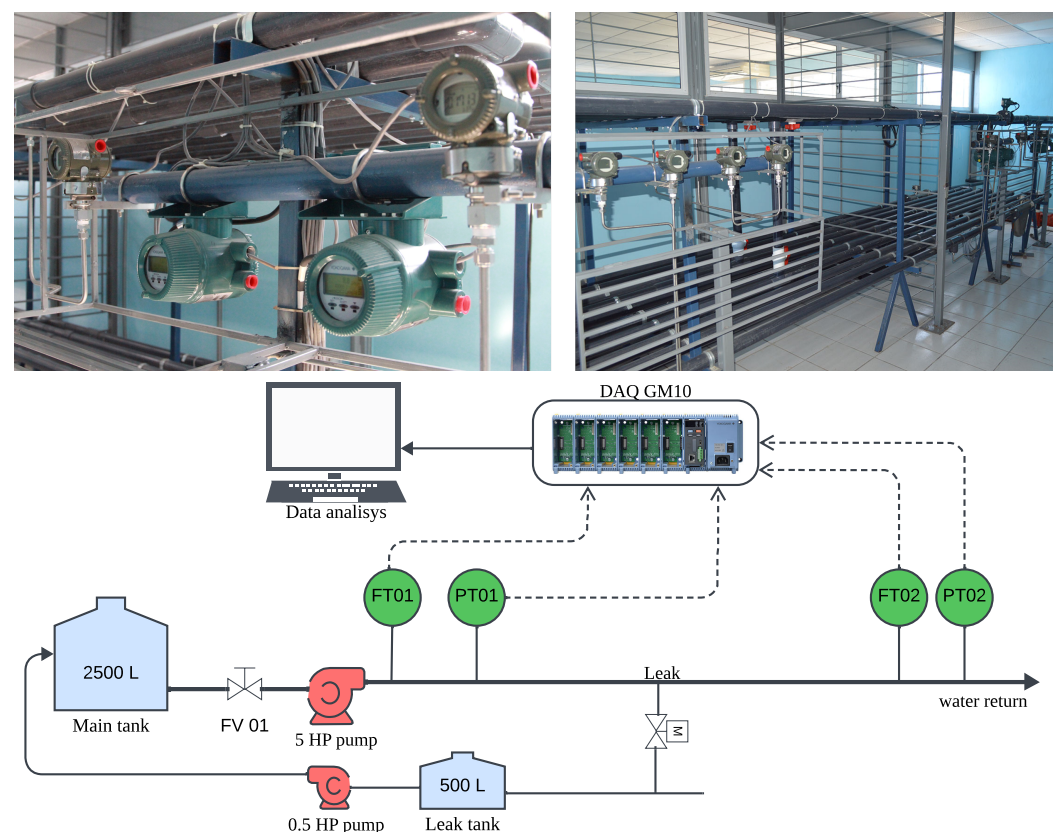


Figure 2. Physical hydraulic system (top) and P&ID Diagram (bottom).

The pipeline spans a total length of 115 m, and is arranged in a serpentine layout to optimize space and flow dynamics. Pressure (PT01, PT02) and flow (FT01, FT02) sensors are strategically installed at the inlet and outlet of the system to capture key operational parameters. These measurements are collected and stored using a Yokogawa GM10 data acquisition (DAQ) system, which communicates via the Modbus TCP/IP protocol over Ethernet.

To simulate different operating conditions, a Siemens Micromaster 420 variable frequency drive is employed to control the pumping power, with a capacity of up to 5 HP; additionally, a manually operated valve is included to simulate leak scenarios along the pipeline. The system components, including the sensors and their respective models, are summarized in Table 1.

This experimental setup provides a robust platform for evaluating various fault detection and diagnosis methodologies under controlled yet realistic conditions. Implementing the proposed method requires placement of sensors at the inlet and outlet of the pressurized pipeline. Additionally, a DAQ system and a computer are necessary to process the data collected from the sensors.

Table 1. Components and sensors in the experimental hydraulic pipeline setup.

Component	Sensor/Element and Model	Key Features
Pipeline	N/A	Internal diameter: 0.052 m, Material: PVC, High pressure resistance
Flow Sensors	Coriolis Mass Flow Meter (Yokogawa ROTAMASS)	Mass flow measurement, $\pm 0.1\%$ accuracy, wide range
Pressure Sensors	Gauge Pressure Sensor (Yokogawa EJA530E)	$\pm 0.055\%$ accuracy, 4 a 20 mA range, HART communication, DPharp sensor, IP67 rated
SCADA System	Data Acquisition Recorder (Yokogawa GM10)	Modbus TCP/IP communication, reliable data storage
Pump Control	Variable Frequency Drive (Siemens Micromaster 420)	Up to 5 HP, precise speed control
Valves	Manual	Flow control, configuration adjustments
Leak Valve	Manual	Simulates leaks, no integrated sensor

3.2. Flow Dynamics in a Pipeline

A model for fluid flowing through a horizontal pipeline is derived from the physical principles of momentum conservation and mass conservation (continuity equation) in terms of time t and a spatial variable z conveniently defined in the axial direction of the pipeline. The spatiotemporal dynamics of the flow are modeled by a set of hyperbolic partial differential equations [28], commonly referred to as the water hammer equations:

$$\frac{\partial H(z, t)}{\partial t} + \frac{b^2}{gA_r} \frac{\partial Q(z, t)}{\partial z} = 0, \quad (4)$$

$$\frac{1}{A_r} \frac{\partial Q(z, t)}{\partial t} + g \frac{\partial H(z, t)}{\partial z} + \frac{\lambda Q(z, t) |Q(z, t)|}{2dA_r^2} = 0, \quad (5)$$

where H is the pressure head [m], Q is the flow rate [m^3/s], b is the wave speed [m/s], g is the gravity acceleration [m/s^2], A_r is the cross-sectional area [m^2] of the pipeline, and d is its diameter [m]. To obtain accurate results with the above model, it is necessary to precisely know the friction factor λ and wave speed b , although the latter only influences transient calculations and not the steady-state case. A challenge in estimating these parameters is their dependence on temperature, and in the case of λ its nonlinear variation with flow rate.

For constant flow (steady-state) conditions, the friction factor can be determined from the Darcy–Weisbach equation:

$$\rho g \Delta H = \lambda \frac{\rho q^2}{2A_r^2 d} \Delta z \quad (6)$$

where Δz is the length traveled by the fluid and ΔH is the pressure drop over that length.

In this work, turbulent flow conditions were considered throughout the experiments, as determined by the Reynolds number (Re) calculations for the pipeline system. Under turbulent flow, the friction factor λ is a nonlinear implicit function of the Reynolds number and the internal roughness ε of the pipeline. The Swamee–Jain approximation was used to estimate λ , providing a practical and accurate approach for turbulent regimes. This dependence is provided by the Colebrook–Moody formula [29], although in calculations the explicit Swamee–Jain approximation is often used [30]:

$$\lambda = 0.25 \left[\log_{10} \left(\frac{\varepsilon}{3.7d} + \frac{5.74}{\text{Re}^{0.9}} \right) \right]^{-2}. \quad (7)$$

It should be noted that the partial differential Equations (4) and (5) model the spatial and temporal variations of the pressure and flow rate in the pipeline subject to certain constraints, namely, the uniform velocity distribution, one-dimensional flow, constant liquid density, and constant cross-sectional area. Continuous flow and no branches are assumed without considering the existence of lateral connections.

An explicit analytical solution for the partial differential Equations (4) and (5) is not known. The solution is often obtained numerically using finite differences or the method of characteristics [31], and two boundary conditions are required; these can be known pressures or flow rates at the pipeline ends.

For a pipeline of length L , considering a spatial partition with N sections and $N + 1$ points, $\{z_i\} := \{z_1 = 0, z_2 = z_1 + \Delta z_1, \dots, z_{N+1} = z_N + \Delta z_N = L\}$, and approximating the partial derivatives by first-order finite differences, a discretized version of (4) and (5) in the spatial variable z is obtained:

$$H(z_i, t) = H_i, \quad Q(z_i, t) = Q_i, \quad (8)$$

$$\frac{\partial H(z, t)}{\partial t} = \dot{H}_i, \quad \frac{\partial Q(z, t)}{\partial t} = \dot{Q}_i, \quad (9)$$

$$\frac{\partial H(z, t)}{\partial z} \approx \frac{\Delta H_i}{\Delta z_i} = \frac{H_i(t) - H_{i+1}(t)}{\Delta z_i}, \quad (10)$$

$$\frac{\partial Q(z, t)}{\partial z} \approx \frac{\Delta Q_i}{\Delta z_i} = \frac{Q_i(t) - Q_{i+1}(t)}{\Delta z_i}, \quad (11)$$

where i represents the considered spatial section.

Incorporating the approximations in (8)–(10) into (4) and (5), the dynamic model of the flow is expressed by

$$\dot{H}_{i+1} = a_2 \frac{Q_i(t) - Q_{i+1}(t)}{\Delta z_i}, \quad (12)$$

$$\dot{Q}_i = a_1 \frac{H_i(t) - H_{i+1}(t)}{\Delta z_i} + \mu Q_i(t) |Q_i(t)|, \quad (13)$$

where

$$a_1 = gA_r, \quad a_2 = \frac{b^2}{gA_r}, \quad \mu = \frac{-\lambda}{2dA_r}. \quad (14)$$

In Table 2, the most relevant parameters of the experimental pipeline are presented for nominal operating conditions.

These equations and parameterizations can be used to test methods for detecting sensor faults and leaks. The experimental plant provides four measured variables: the inlet pressure variable H_{in} is measured by sensor PT01, and the inlet flow variable Q_{in} is measured by FT01; at the outlet, sensors PT02 and FT02 measure the pressure H_{out} and flow Q_{out} , respectively. Additionally, faults can be induced in the sensors and leaks in

the pipeline. The faults are introduced as deviations in the four available measurements. Specifically, four types of synthetic faults are induced in the sensors, including a step-type bias of 10% with respect to the nominal value. The leak is created by actuating one of the leak valves in the experimental plant.

Table 2. Pipeline parameters for nominal operating conditions.

Parameter	Value
Pipe Length, L	115 m
Pipe Diameter, d	0.052 m
Cross-sectional area, A_r	0.0021 m ²
Relative Roughness, ε	2.47×10^{-4}
Fluid Density, ρ	995.736 kg/m ³
Kinematic Viscosity, ν	8.03×10^{-7} m ² /s
Wave Speed, b	422.754 m/s
Gravity Acceleration, g	9.81 m/s ²

3.3. Preparation of the Dataset for Training the Neuro-Fuzzy System

Experiments were carried out under various fault-free operating conditions, each lasting 600 s per experiment and having a sampling rate of 5 samples per second. It is important to note that each experiment differed due to the turbulent flow of pressurized water. The measurements contain noise, and no data preprocessing is required. To capture the nonlinearity of the system, the four variables H_{in} , H_{out} , Q_{in} , and Q_{out} were structured in the form of a regressor considering two instances of k . For each measured variable, an identified variable was considered in its regressor form, as shown in Table 3:

Table 3. System variables to be identified in regressive form.

Variable	Regressive Form
$\hat{H}_{in}(k)$	$(H_{in}(k), H_{in}(k-1), H_{in}(k-2), Q_{in}(k), Q_{out}(k))$
$\hat{H}_{out}(k)$	$(H_{out}(k), H_{out}(k-1), H_{out}(k-2), Q_{in}(k), Q_{out}(k))$
$\hat{Q}_{in}(k)$	$(Q_{in}(k), Q_{in}(k-1), Q_{in}(k-2), H_{in}(k), H_{out}(k))$
$\hat{Q}_{out}(k)$	$(Q_{out}(k), Q_{out}(k-1), Q_{out}(k-2), H_{in}(k), H_{out}(k))$

These regressive expressions function as inputs for the ANFIS networks; through the learning process, they can estimate the variables and develop Takagi–Sugeno fuzzy models to design the neuro-fuzzy ZKF observers.

3.4. System Identification-Based ANFIS Learning

The system identification process uses the regressive form of the variables presented in Table 3. The inputs to the ANFIS are structured for system identification as illustrated in Figure 3. It is important to mention that the training process uses fault-free data to ensure the accuracy and reliability of the model. Figure 3 shows the regressive variables $Q_{in}(k)$, $Q_{out}(k)$, $H_{in}(k-2)$, $H_{in}(k-1)$, and $H_{in}(k)$ used as inputs for the fuzzy inference system. The ANFIS utilizes these inputs to identify the system's behavior and estimate the output variable $\hat{H}_{in}(k)$.

The regressive expressions serve as inputs to ANFIS. Through training with fault-free data, these can accurately identify the estimated variables and develop Takagi–Sugeno models. These models are essential for designing zonotopic Kalman filters and enhancing the system's fault detection capabilities.

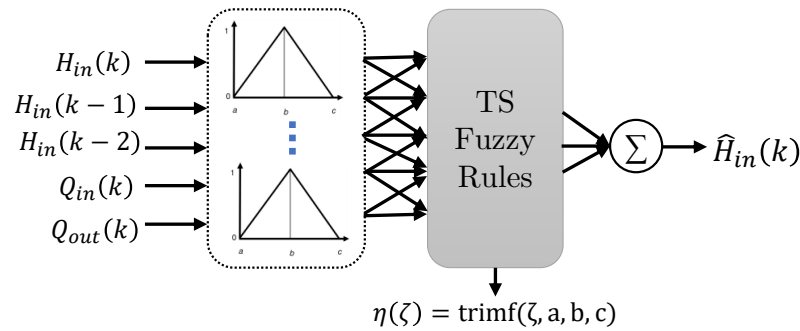


Figure 3. Structure of the ANFIS model for system identification using regressive inputs.

ANFIS is employed to approximate the nonlinear behavior of each variable of the pipeline. The input data for ANFIS are made up of each variable, as shown in Table 3, and are structured as follows:

$$\zeta = \begin{bmatrix} H_{in}(k) & H_{in}(k-1) & H_{in}(k-2) & Q_{in}(k) & Q_{out}(k) \end{bmatrix}^T. \quad (15)$$

Layer 1. This is the fuzzification layer, where membership functions (MF) are used for fuzzification. Each MF, denoted by $\eta(\cdot)$, has three parameters (a_{nf}, b_{nf}, c_{nf}) , and is defined as follows:

$$\eta_{nf}(\zeta_o) = \begin{cases} 0 & \text{if } \zeta_o \leq a_{nf} \\ \frac{\zeta_o - a_{nf}}{b_{nf} - a_{nf}} & \text{if } a_{nf} < \zeta_o \leq b_{nf} \\ \frac{c_{nf} - \zeta_o}{c_{nf} - b_{nf}} & \text{if } b_{nf} < \zeta_o \leq c_{nf} \\ 0 & \text{if } \zeta_o > c_{nf} \end{cases}, \quad \begin{matrix} \forall nf = 1, \dots, N_{MF}, \\ \forall o = 1, \dots, N_{\zeta}, \end{matrix} \quad (16)$$

where ζ stands for the vector of the ANFIS input variables (from now on referred to as the antecedent parameter) and N_{MF} signifies the number of membership functions for each antecedent parameter.

Layer 2. This layer formulates the rules using the preceding membership functions; in each of the $N_v = (N_{MF})^{N_{\zeta}} = 32$ nodes, there is a fixed node that multiplies the incoming signals and transmits the product:

$$\mu_i(\zeta) = \prod_{o=1}^{N_v} \eta_{nf}(\zeta_o), \quad \forall i = 1, \dots, N_v. \quad (17)$$

Layer 3. This is the normalization layer:

$$\bar{\mu}_i(\zeta) = \frac{\mu_i(\zeta)}{\sum_{i=1}^{N_v} \mu_i(\zeta)}, \quad \forall i = 1, \dots, N_v. \quad (18)$$

Layer 4. Known as the defuzzification or consequent layer, the if-then fuzzy rules of Takagi and Sugeno are used [32]:

$$\begin{aligned} \mathcal{R}_i : & \text{ IF } \zeta_1 \text{ is } \eta_{m1} \text{ AND } \dots \text{ AND } \zeta_{N_{\zeta}} \text{ is } \eta_{mN_{\zeta}} \\ & \text{ THEN } \bar{\mu}_i(\zeta; \alpha_i + \gamma_i), \quad \forall i = 1, \dots, N_v. \end{aligned} \quad (19)$$

Output 1. This layer calculates the final output by aggregating all incoming signals from the defuzzification layer:

$$\sum_{i=1}^{N_v} \bar{\mu}_i(\zeta_i \alpha_i + \gamma_i). \quad (20)$$

After training the ANFIS and calculating the normalized weights (18) and consequent parameters (19), the fuzzy TS representation is constructed. As an example, the variable $\hat{H}_{in}$ can be used, which is formulated as follows:

$$\hat{H}_{in}(k) = \sum_{i=1}^{N_v} \bar{\mu}_i(\zeta(k)) \left(\alpha_{1i} H_{in}(k) + \alpha_{2i} H_{in}(k-1) + \alpha_{3i} H_{in}(k-2) + \alpha_{4i} Q_{in}(k) + \alpha_{5i} Q_{out}(k) + \gamma_i \right). \quad (21)$$

The terms in (21) can be rearranged as

$$\hat{H}_{in}(k) = \sum_{i=1}^{N_v} \bar{\mu}_i(\zeta(k)) \left(\begin{bmatrix} \alpha_{1i}^1 & \alpha_{2i}^1 & \alpha_{3i}^1 \\ \alpha_{1i}^2 & \alpha_{2i}^2 & \alpha_{3i}^2 \\ \alpha_{1i}^3 & \alpha_{2i}^3 & \alpha_{3i}^3 \end{bmatrix} x + \begin{bmatrix} \alpha_{4i}^1 & \alpha_{5i}^1 \\ \alpha_{4i}^2 & \alpha_{5i}^2 \\ \alpha_{4i}^3 & \alpha_{5i}^3 \end{bmatrix} u + \begin{bmatrix} \gamma_i^1 \\ \gamma_i^2 \\ \gamma_i^3 \end{bmatrix} \right), \quad (22)$$

where $x = \begin{bmatrix} H_{in}(k) & H_{in}(k-1) & H_{in}(k-2) \end{bmatrix}^T$ represents the state vector and $u = \begin{bmatrix} Q_{in} & Q_{out} \end{bmatrix}^T$ represents the input vector. The superscript 1, 2, 3 denotes the output number for ANFIS. The polytopic structure is reformulated as a discrete-time state space representation:

$$\begin{aligned} x(k+1) &= \sum_{i=1}^{N_v} \bar{\mu}_i(\zeta(k)) (A_i x(k) + B_i u(k) + \gamma_i) \\ y(k) &= Cx(k), \end{aligned} \quad (23)$$

where N_v is determined by $N_v = (N_{MF})^{N_\zeta}$, $\bar{\mu}_i(\zeta(k))$ represents the premise functions, $A_i \in \mathbb{R}^{n_x \times n_x}$, $B_i \in \mathbb{R}^{n_x \times n_u}$, $\gamma_i \in \mathbb{R}^{n_x}$, and $C \in \mathbb{R}^{n_y \times n_x}$ denote the system matrices, and $y(k) \in \mathbb{R}^{n_y}$ denotes the computed output vector. It is important to note that the system is subject to uncertainties due to model mismatches caused by friction and turbulent flow. The matrices Ψ_i and Ω_i represent the uncertainties, whereas F_v accounts for noise.

$$\begin{aligned} x(k+1) &= \sum_{i=1}^{N_v} \bar{\mu}_i(\zeta(k)) ((A_i + \Psi_i)x(k) + (B_i + \Omega_i)u(k) + \gamma_i), \\ y(k) &= Cx(k) + F_\sigma \sigma(k). \end{aligned} \quad (24)$$

The values of the uncertain matrices are obtained from the error covariance matrix of all consequent neuro-fuzzy parameters, which is produced by the hybrid learning algorithm. To build the Ψ_i and Ω_i matrices, standard deviations are derived from the variances present in the covariance matrix; F_σ denotes the noise matrices with fixed dimensions and $\sigma(k) \in \mathbb{R}^{n_y}$ represents the measurement noise based on the sensor's additive noise in the pipeline system. The uncertain parameters can be approximated in a single term according to [33]. Thus, (24) is reformulated as follows:

$$\begin{aligned} x(k+1) &= \sum_{i=1}^{N_v} \bar{\mu}_i(\zeta(k)) (A_i x(k) + B_i u(k) + \gamma_i + E_i \delta(k)), \\ y(k) &= Cx(k) + F_\sigma \sigma(k) \end{aligned} \quad (25)$$

with

$$E_i \delta(k) = \Psi_i x(k) + \Omega_i u(k), \quad (26)$$

where E_i is the uncertainty distribution matrix of suitable dimensions and $\delta(k) \in \mathbb{R}^{n_x}$ is a vector that captures the impact of uncertainty. The next section outlines the procedure for designing the set-based neuro-fuzzy ZKF observer.

3.5. Design of a Neuro-Fuzzy ZKF Observer

The design of a neuro-fuzzy ZKF observer using a Takagi–Sugeno fuzzy system is obtained through ANFIS learning. The observer is based on the ZKF, which offers several advantages over the traditional Kalman filter (KF). This design is called the neuro-fuzzy ZKF observer due to its combination of neuro-fuzzy logic and robust zonotopic estimation. Unlike a KF, the proposed observer uses bounded uncertainties instead of probabilistic distributions, allowing it to handle large deterministic disturbances more effectively. Both techniques share a similar gain structure and use prediction–update cycles to refine state estimates; however, the neuro-fuzzy ZKF provides zonotopic sets that enclose all possible admissible system states, whereas a KF offers confidence ellipses based on state covariance.

The neuro-fuzzy ZKF does not require assumptions about the distribution of uncertainties, making it more flexible and robust against large deterministic disturbances. Additionally, it provides zonotopic sets that enclose all possible state values, offering explicit guarantees of state inclusion. This design leverages these advantages by combining the adaptability and flexibility of fuzzy systems with the robustness and precision of the ZKF, enhancing state estimation in complex dynamic systems.

To provide a simplified notation, the TS system (24) is reformulated as follows:

$$\begin{aligned} x(k+1) &= A_\omega x(k) + B_\omega u(k) + \gamma_\omega + E_\omega \delta(k), \\ y(k) &= Cx(k) + F_\sigma \sigma(k), \end{aligned} \quad (27)$$

where $A_\omega = \sum_{i=1}^{N_v} \bar{\mu}_i(\zeta(k)) A_i$, $B_\omega = \sum_{i=1}^{N_v} \bar{\mu}_i(\zeta(k)) B_i$, $\gamma_\omega = \sum_{i=1}^{N_v} \bar{\mu}_i(\zeta(k)) \gamma_i$, and $E_\omega = \sum_{i=1}^{N_v} \bar{\mu}_i(\zeta(k)) E_i$.

The uncertainties and noise are considered as a zonotopic representation, as follows:

$$\begin{aligned} \delta &\in \langle c_\delta, R_\delta \rangle, \\ \sigma &\in \langle c_\sigma, R_\sigma \rangle, \end{aligned} \quad (28)$$

where c_δ and c_σ denote the centers of the zonotopes that bound the uncertainty and noise, respectively, with their associated generator matrices $R_\delta \in \mathbb{R}^{n_x \times n_x}$ and $R_\sigma \in \mathbb{R}^{n_y \times n_y}$. The following assumption is considered when designing the observer.

Assumption 1. The uncertainties and noise in (28) are assumed to be bounded by a unit hypercube zonotope centered at the origin. Specifically, for all $k \geq 0$, we have $\delta \in [-1, 1]^{n_\delta} = \langle 0, I_{n_\delta} \rangle$ and $\sigma \in [-1, 1]^{n_\sigma} = \langle 0, I_{n_\sigma} \rangle$, where $I_{n_\delta} \in \mathbb{R}^{n_\delta \times n_\delta}$ and $I_{n_\sigma} \in \mathbb{R}^{n_\sigma \times n_\sigma}$ are identity matrices.

Considering the assumption that uncertainties and noise are bounded by the unit hypercube zonotope centered at the origin and that the initial state x_0 belongs to the set $\mathcal{X}_0^{zo} = \langle c_{k,0}^{zo}, R_{k,0}^{zo} \rangle$, where $c_{k,0}^{zo} \in \mathbb{R}^{n_x}$ represents the center and $R_{k,0}^{zo} \in \mathbb{R}^{n_x \times n_x}$ is a non-empty matrix containing the generator matrices of the initial zonotope $\mathcal{X}_0^{zo}$, the following neuro-fuzzy ZKF observer is structured.:

$$\hat{x}(k+1) = A_\omega \hat{x}(k) + B_\omega u(k) + \gamma_\omega(k) + E_\omega \delta(k) + L_\omega (y - Cx(k) - F_\sigma \sigma(k)) \quad (29)$$

where the vector $\hat{x}(k+1) \in \mathbb{R}^{n_x}$ denotes the estimated states and $L_\omega \in \mathbb{R}^{n_x \times n_y}$ represents the observer gains that need to be determined. The following proposition is essential for calculating the observer gains.

Proposition 1. *Given the system (27) and observer structure (29), the zonotope $\hat{\mathcal{X}}_k^{zo} = \langle c_{k+1}^{zo}, R_{k+1}^{zo} \rangle$ is recursively forward-predicted as follows:*

$$\begin{aligned} c_{k+1}^{zo} &= (A_\omega - L_\omega C) c_k^{zo} + B_\omega u_k + \gamma_k + L_\omega y_k \\ R_{k+1}^{zo} &= [(A_\omega - L_\omega C) \bar{R}_k^{zo}, E_\omega, -L_\omega F_\sigma] \\ \bar{R}_k^{zo} &= \downarrow_q (R_k^{zo}) \end{aligned} \quad (30)$$

such that the reduction operator $\downarrow_q$ satisfies $\bar{R}_k = \downarrow_q (R)$ and $\hat{x}(k) \in \langle c_k, R_k \rangle \subset \langle c_k, \bar{R}_k \rangle$.

Proof. Assuming $\hat{\mathcal{X}}_k^{zo} = \langle c_{k+1}^{zo}, R_{k+1}^{zo} \rangle$ at time k (true at $k = 0$) and $\downarrow_{q,W}$ preserving inclusion, $\hat{\mathcal{X}}^{zo} = \langle c_k, \bar{R}_k \rangle$; because $\delta = \langle 0, I_{n_\delta} \rangle$ and $\sigma \in \langle 0, I_{n_\sigma} \rangle$ (27), from (29) we have

$$\begin{aligned} \hat{x}(k+1) &= ((A_\omega - L_\omega C) \odot \langle c_k, R_k \rangle) \oplus (B_\omega \odot \langle u_k, 0 \rangle) \oplus (\langle \gamma_\omega, 0 \rangle) \\ &\quad \oplus (E_\omega \odot \langle 0, I_{n_\delta} \rangle) \oplus (L_\omega \odot \langle y_k, 0 \rangle) \oplus (-L_\omega F_\sigma \odot \langle 0, I_{n_\sigma} \rangle). \end{aligned} \quad (31)$$

Then, applying the zonotope property 1 yields (30). Hence, the proof is completed. $\square$

As highlighted in Proposition 1, the zonotopic state-bounding observer (30) is characterized by the zonotopic observer gain L_z at each time step k . According to [25,34], the size of the state-bounding zonotope $\hat{\mathcal{X}}^{zo} = \langle c_{k+1}^{zo}, R_{k+1}^{zo} \rangle$ can be minimized using its F -radius. The following theorem offers a method to calculate L_z for this purpose.

Theorem 1. *Consider the nonlinear TS fuzzy system (27) and its associated zonotopic observer (29). The size of the zonotope defined in (30) can be optimized by using the following observer gain:*

$$L_\omega = \Gamma_\omega \Phi_k^{-1} \quad (32)$$

with

$$\Gamma_\omega = A_\omega P_k C^T, \quad P_k = R_{x_k} R_{x_k}^T, \quad \Phi_k = C P_k C^T + F_\sigma F_\sigma^T.$$

Proof. According to [25], minimizing the F -radius and F_W -radius of a zonotope is equivalent to minimizing the trace of its covariance; therefore, minimizing the F -radius of the zonotope $\hat{\mathcal{X}}^{zo} = \langle c_{k+1}^{zo}, R_{k+1}^{zo} \rangle$ is equivalent to minimizing the trace of its covariance $P_{k+1} = R_{k+1}^{zo} R_{k+1}^{zo T}$, i.e.,

$$\mathcal{F} = \|R_{k+1}^{zo}\|_F^2 = \text{Tr}(R_{k+1}^{zo} R_{k+1}^{zo T}) = \text{Tr}(P_{k+1}), \quad (33)$$

where $\mathcal{F}$ denotes the Frobenius radius and P is the covariance of the zonotope matrix R_x . Similarly, minimizing the F_W -radius of the zonotope $\mathcal{X}_{k+1}$ is equivalent to minimizing the criterion, i.e.,

$$\mathcal{F}_W = \|R_{k+1}^{zo}\|_{F,W}^2 = \text{Tr}(W P_{k+1}), \quad (34)$$

where $\mathcal{F}_W$ denotes the weighted Frobenius radius and WP is the weighted function covariance of the zonotope matrix R_{x_k} . Note that W is any positive definite weighting matrix. It follows from (30) that

$$P_{k+1} = (A_\omega - L_\omega C)P_k(A_\omega - L_\omega C)^T + D_\omega + L_\omega D_\sigma L_\omega^T, \quad (35)$$

where $D_\omega = E_\omega E_\omega^T$ and $D_\sigma = F_\sigma F_\sigma^T$. From (35), the criterion $\mathcal{F}_W$ in (34) can be rewritten as

$$\mathcal{F}_W = \text{Tr}(W(A_\omega - L_\omega C)P_k(A_\omega - L_\omega C)^T + WD_\omega + WL_\omega D_\sigma L_\omega^T). \quad (36)$$

Then, the optimal value of the observer gain L_ω is determined such that

$$\frac{\partial \mathcal{F}_W}{\partial L_\omega} = 0. \quad (37)$$

Considering (36), this latter yields

$$\frac{\partial \mathcal{F}_W}{\partial L_\omega} = \frac{\partial}{\partial L_\omega} \text{Tr}(WL_\omega(CP_k C^T + D_\sigma)L_\omega^T) - 2\frac{\partial}{\partial L_\omega} \text{Tr}(WA_\omega P_k C^T L_\omega^T) = 0. \quad (38)$$

Developing Equation (38), we obtain

$$\frac{\partial}{\partial L_\omega} \text{Tr}(WL_\omega(CP_k C^T + D_\sigma)L_\omega^T) = 2WL_\omega(CP_k C^T + D_\sigma), \quad (39)$$

$$\frac{\partial}{\partial L_\omega} \text{Tr}(WA_\omega P_k C^T L_\omega^T) = WA_\omega P_k C^T. \quad (40)$$

Then, Equation (38) becomes

$$2WL_\omega(CP_k C^T + D_\sigma) - 2WA_\omega P_k C^T = 0. \quad (41)$$

Simplifying, we obtain

$$WL_\omega(CP_k C^T + D_\sigma) = WA_\omega P_k C^T. \quad (42)$$

Because W is a positive definite weighting matrix, we can multiply by W^{-1} and transpose

$$L_\omega(CP_k C^T + D_\sigma) = A_\omega P_k C^T. \quad (43)$$

Finally, we arrive at the desired equation:

$$L_\omega \Phi_k = A_\omega P_k C^T \quad (44)$$

which in turn leads to the observer gain expression in (32). $\square$

3.6. Neuro-Fuzzy ZKF Fault Detection Scheme

The fault detection and isolation process using the neuro-fuzzy ZKF involves estimating the pipeline variables with ANFIS and propagating disturbances and noise through a zonotopic observer. This process identifies faults by checking the intersection between the estimated zonotope and the measured strip at each time instant. If the intersection is empty, a fault is indicated. The results are stored in a Fault Signature Matrix (FSM), which aids in diagnosing the faults. The detailed steps of this fault detection scheme are outlined in Algorithm 1.

Algorithm 1 Zonotopic Fault Detection Scheme

- 1: **Input:** Pipeline variables estimated by ANFIS, as in Table 3
- 2: **Output:** Fault Signature Matrix (FSM)
- 3: **Initialize:**
- 4: Number of estimated variables of system s
- 5: Zonotopic observer parameters and disturbance bounds
- 6: **for** each time instant k **do**
- 7: **Estimate** the pipeline variables using ANFIS
- 8: **Propagate** uncertainty and noise through the zonotopic observer
- 9: Calculate the strip $\mathcal{X}_k^{y_k}$ for each measured variable:

$$\mathcal{X}_k^{y_k} = \{x(k) \in \mathbb{R}^{n_x} : |Cx(k) - y_s(k)| \leq F_\sigma\} \quad (45)$$

- 10: **Check** for the existence of a fault by verifying:

$$\hat{\mathcal{X}}_k^{zo} \cap \mathcal{X}_k^{y_k} = \emptyset \quad (46)$$

- 11: **if** $\hat{\mathcal{X}}_k^{zo} \cap \mathcal{X}_k^{y_k} = \emptyset$ **then**
- 12: A fault is indicated
- 13: Generate residuals $r_s(k)$
- 14: **Update** Fault Signature Matrix (FSM) as follows:

$$\psi_{s,j}(k) = \begin{cases} 0 & \text{if } r_s(k) \text{ is consistent (No fault)} \\ 1 & \text{if } r_s(k) \text{ is not consistent (Fault)} \end{cases} \quad (47)$$

- 15: **else**
- 16: No fault is indicated
- 17: **end if**
- 18: **end for**
- 19: **Return:** Fault Signature Matrix (FSM)

4. Results

This section presents the results of the proposed method for faults and leak detection in hydraulic pipeline systems using neuro-fuzzy ZKF. The method was evaluated in the experimental pipeline specifically configured for this purpose.

4.1. Experimental Data Setup

The experimental conditions for the hydraulic pipeline system are summarized in Table 4. Each operating condition was tested multiple times to ensure data consistency and reliability.

Table 4. Summary of experimental conditions.

Parameter	Details
Pump power frequency	60 Hz, 50 Hz, 40 Hz, 30 Hz
Number of repetitions per condition	5
Duration of each experiment	600 s
Sampling frequency	5 samples per second
Total samples per experiment	3000
Fault condition	free

All experiments were conducted under fault-free conditions to establish a baseline for model training and validation. The collected dataset was partitioned into two subsets to facilitate the development of the Adaptive Neuro-Fuzzy Inference System (ANFIS) models. Specifically:

- Training Set (80% of the dataset): Used to optimize the model's parameters during the training process.
- Validation Set (20% of the dataset): Employed to evaluate model performance and fine-tune the neuro-fuzzy parameters.

4.2. Identification Results Using ANFIS and ZKF Observers

The system variables H_{in} , H_{out} , Q_{in} , Q_{out} were identified using ANFIS, the structure of which is described in Section 3.4.

Subsequently, the neuro-fuzzy ZKF observers were implemented as described in Section 3.5. The results of the observers are graphically presented below, corresponding to an operating point of 60 Hz for the pump. Figure 4 shows the inlet pressure H_{in} under fault-free conditions, where the zonotopic observer (red and green lines) envelops the measured signal shown by the blue line. Similarly, Figure 5 shows the outlet pressure H_{out} under fault-free conditions, Figure 6 shows the inlet flow Q_{in} , and Figure 7 shows the outlet flow Q_{out} , all under fault-free conditions, with the zonotopic observer enveloping the measured signals.

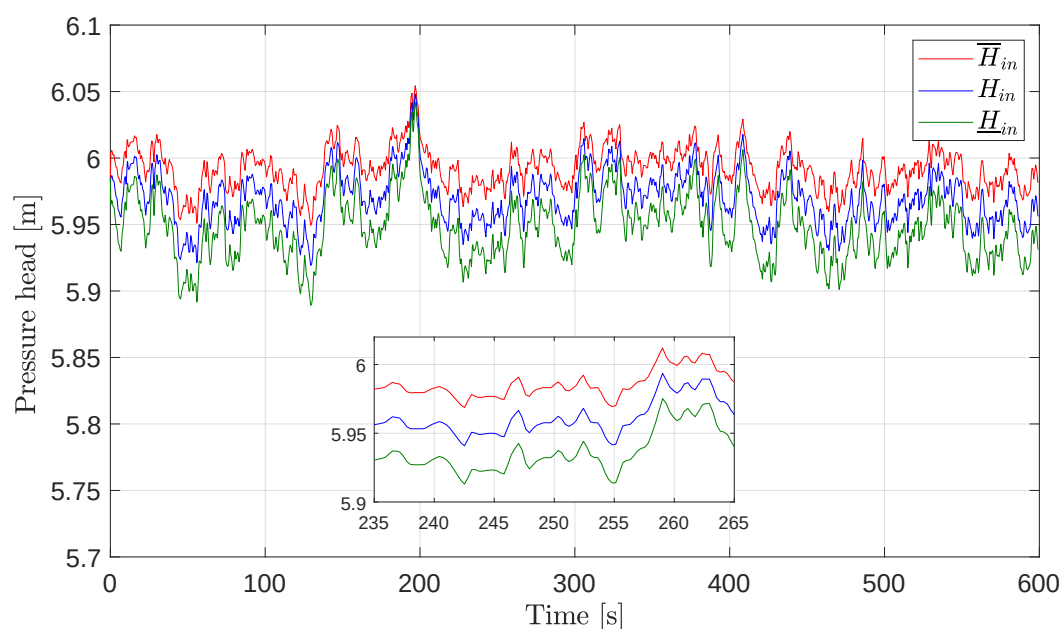


Figure 4. Inlet pressure H_{in} under fault-free conditions.

4.3. Evaluation Under Fault Conditions

To assess the effectiveness of the proposed method, various tests were conducted by intentionally inducing faults in the sensors and creating a leak in the pipeline. The specific faults we induced were as follows:

- Fault 1: A step-type bias of 10% with respect to the nominal value in the inlet pressure sensor (H_{in}).
- Fault 2: A step-type bias of 10% with respect to the nominal value in the outlet pressure sensor (H_{out}).
- Fault 3: A step-type bias of 10% with respect to the nominal value in the inlet flow sensor (Q_{in}).
- Fault 4: A step-type bias of 10% with respect to the nominal value in the outlet flow sensor (Q_{out}).
- Fault 5: A leak induced by actuating one of the leak valves in the pipeline.

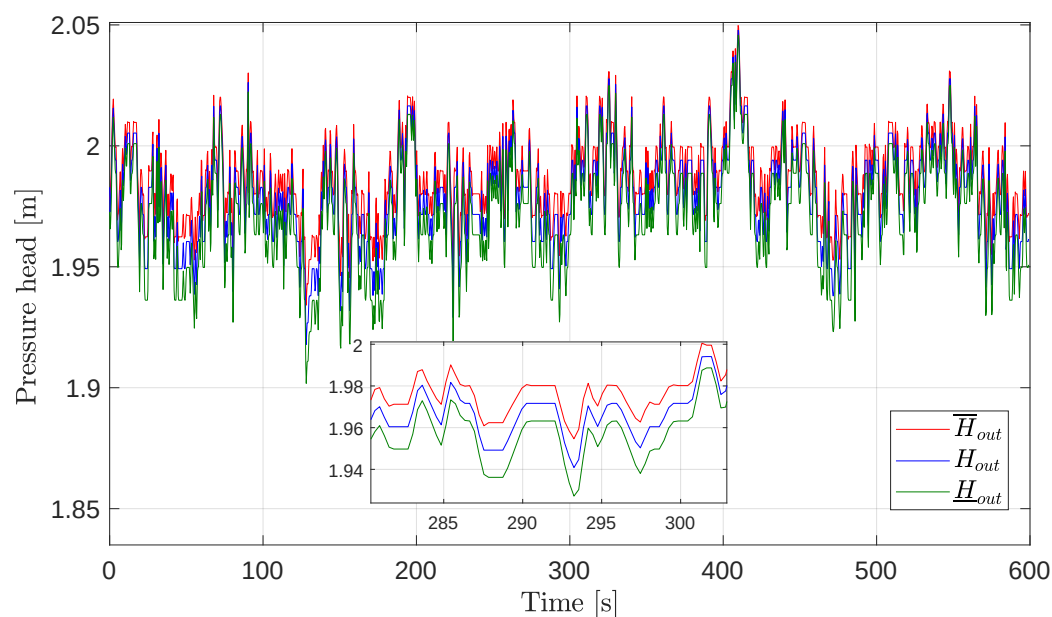


Figure 5. Outlet pressure H_{out} under fault-free conditions.

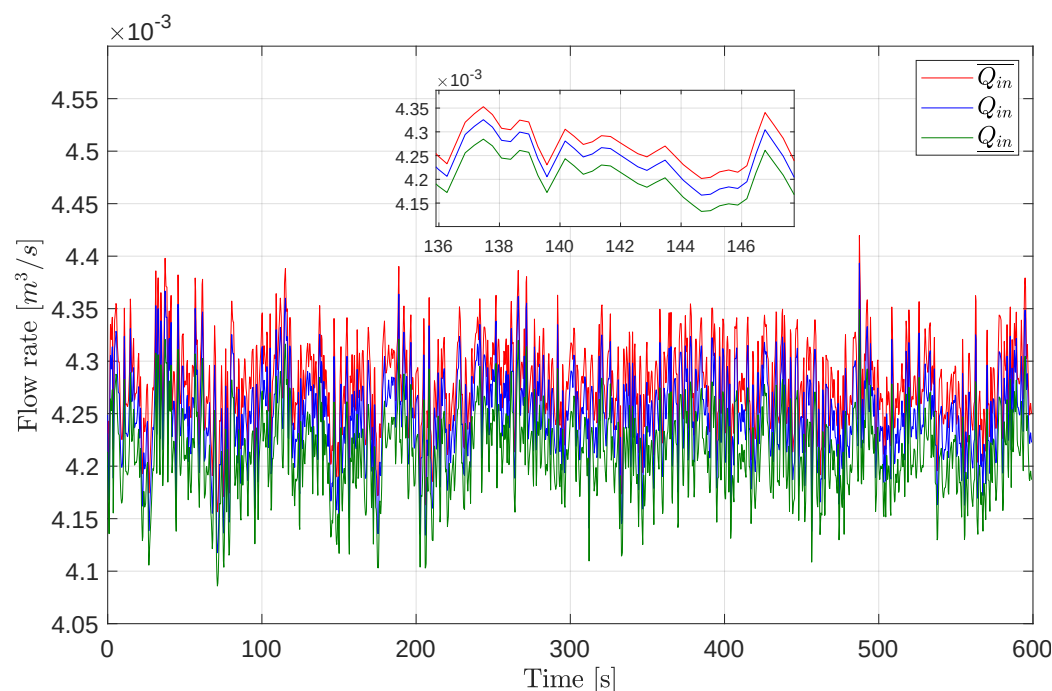


Figure 6. Inlet flow Q_{in} under fault-free conditions.

The faults can be visualized graphically in the following figures under pump conditions at 60 Hz. Due to space limitations in the article, only two graphs are shown for the types of sensor faults. In Figure 8, a fault in the inlet pressure sensor (H_{in}) is activated at time $t = 150$ s, exceeding the observer's upper threshold. In Figure 9, a fault in the inlet flow sensor (Q_{in}) is shown at time $t = 230$ s, which also exceeds the upper threshold. Finally, Figure 10 illustrates an induced leak in the pipeline at time $t = 310$ s. The upper subplot shows the increased Q_{in} due to the leak flow, while the lower subplot shows the decreased Q_{out} due to the leak. In the figures, red and green lines indicate the lower and upper thresholds, respectively.

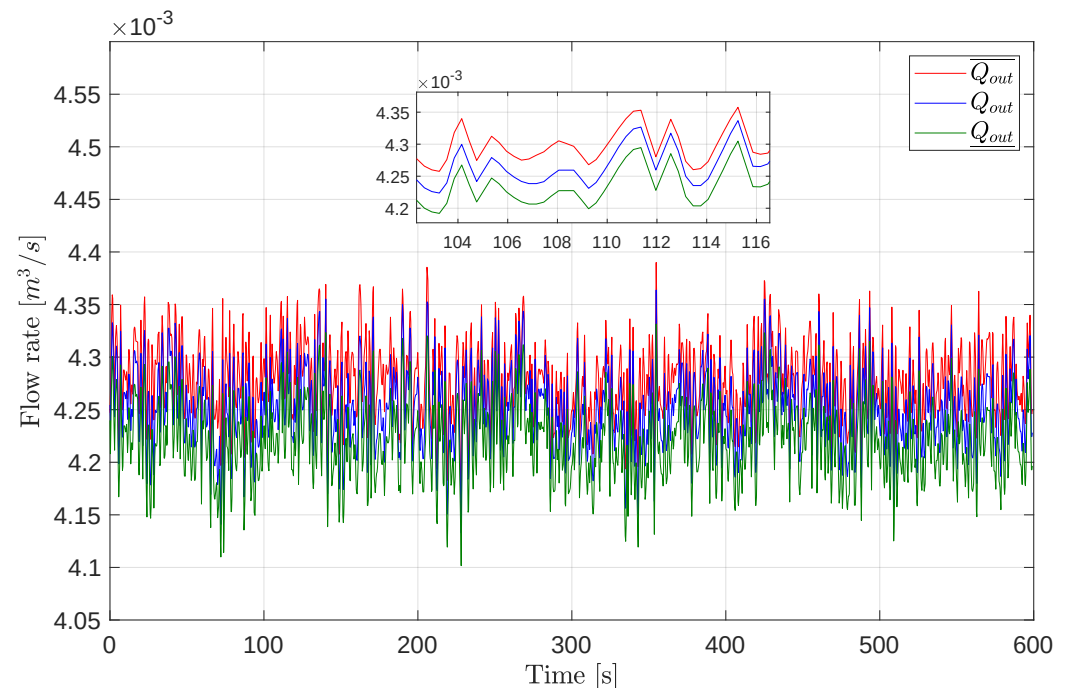


Figure 7. Outlet flow Q_{out} under fault-free conditions.

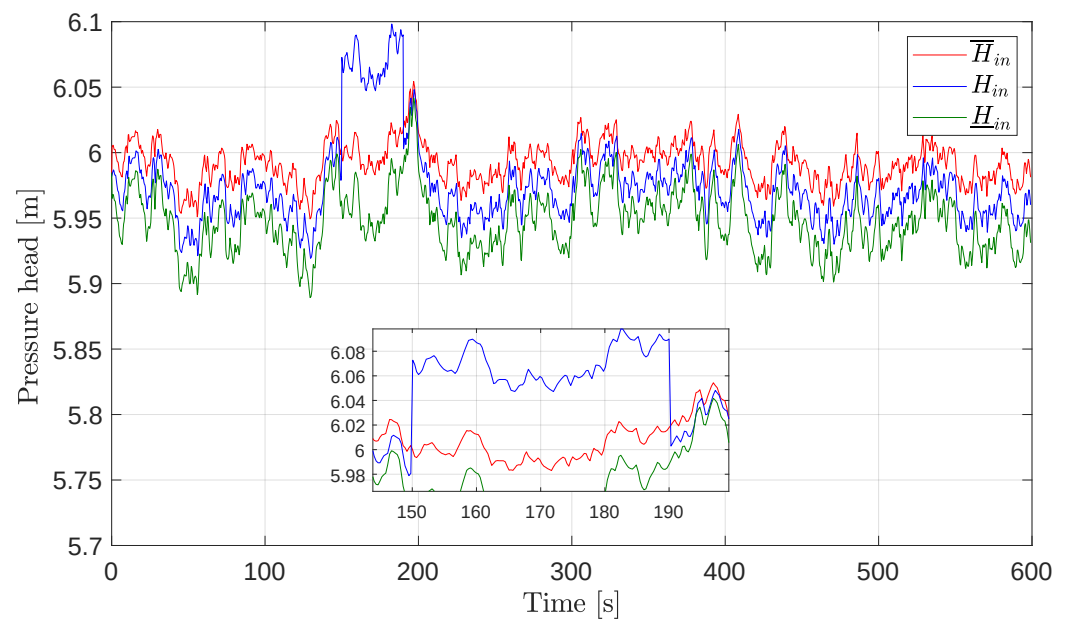


Figure 8. Fault in the H_{in} sensor at time $t = 150$ s.

Using Algorithm 1, the fault detection results are summarized in Table 5, which presents the Fault Signature Matrix (FSM). The FSM summarizes the detection results for each fault scenario. Each row corresponds to a residual r_s generated by the algorithm, while each column corresponds to a specific fault. A value of 1 indicates that the fault was detected in the respective scenario, whereas an empty cell indicates no detection. The matrix clearly shows which residuals are associated with each fault. When a leak occurs, all residuals are activated. Each fault creates a unique signature, allowing for accurate fault diagnosis.

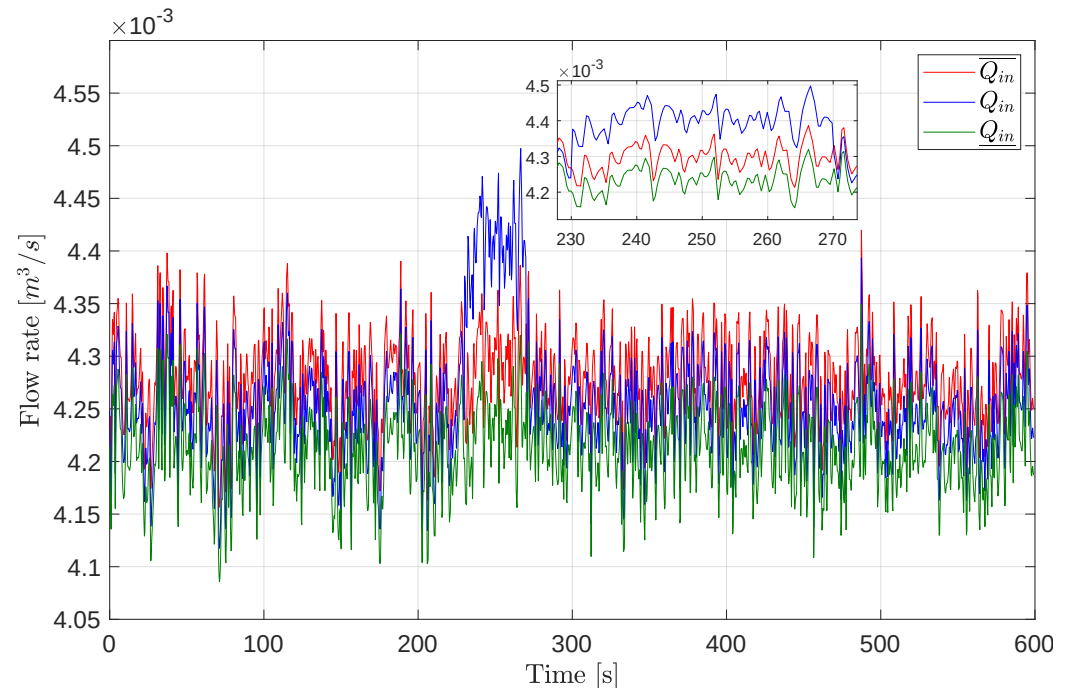


Figure 9. Fault in the Q_{in} sensor at time $t = 230$ s.

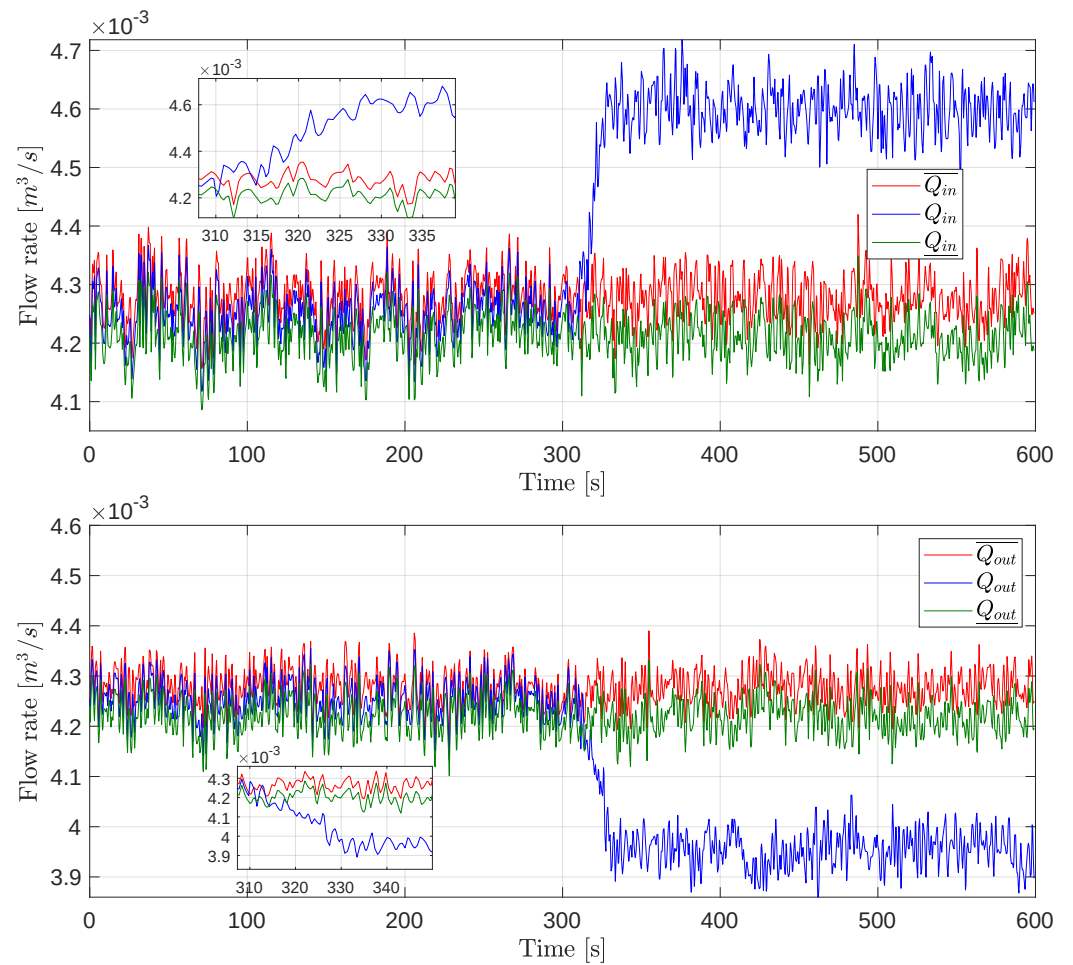


Figure 10. Leak in the pipeline at time $t = 310$ s.

Table 5. Fault signature matrix (FSM).

Residual	Fault 1	Fault 2	Fault 3	Fault 4	Fault 5
r_1	1				1
r_2		1			1
r_3			1		1
r_4				1	1

To evaluate the effectiveness and robustness of the proposed method across all fault cases and at four different pump operating points, the following metrics were used [35].

- Precision: The percentage of true positives over the total number of positive predictions made by the model, calculated as follows:

$$\text{Precision} = \frac{TP}{TP + FP}. \quad (48)$$

- Recall (Detection Rate): The percentage of true positives over the total number of actual positives (the model's ability to correctly identify faults), calculated as follows:

$$\text{Recall} = \frac{TP}{TP + FN}. \quad (49)$$

- False Positive Rate: The percentage of incorrect positive predictions (false positives) over the total number of actual negatives, calculated as follows:

$$\text{False Positive Rate} = \frac{FP}{FP + TN}. \quad (50)$$

These metrics were evaluated across all fault cases and at four different pump operating points, with TP representing the true positives, FP representing the false positives, FN representing the false negatives, and TN representing the true negatives.

The performance metrics obtained at 60 Hz indicate high precision and recall (detection rate) and low false positive rates across five fault types (see Table 6). For Fault 1, the system demonstrated a precision of 98.44% and a recall of 98.72%, with a false positive rate of 1.56%; this reflects a high identification capability, although there is room to reduce false alerts. In the case of Fault 2, the results were outstanding, with a precision of 99.24% and a detection rate of 99.12% along with a very low false positive rate of 0.76%. This indicates extreme accuracy and reliability. Fault 3 showed a precision of 97.51% and a detection rate of 97.04%, but with a false positive rate of 2.48%. Despite the high precision and recall, the higher false positive rate indicates a greater number of incorrect alerts. For Fault 4, the results showed a precision of 98.82% and a detection rate of 98.68%, with a false positive rate of 1.18%. This reflects excellent performance, with a good balance between precision and low false alarms. Fault 5 demonstrated outstanding performance, with a precision of 99.18%, a detection rate of 99.20%, and a false positive rate of only 0.82%.

Table 6. Results obtained with the pump at 60 Hz using the proposed fault diagnosis method.

Fault	Precision	Recall (Detection Rate)	False Positive Rate
1	98.44%	98.72%	1.56%
2	99.24%	99.12%	0.76%
3	97.51%	97.04%	2.48%
4	98.82%	98.68%	1.18%
5	99.18%	99.20%	0.82%

At 50 Hz, the outcomes were highly favorable, showcasing high precision and detection rates across the five faults, as shown in Table 7. For Fault 1, a precision of 97.60% and a detection rate of 97.40% were achieved, with a false positive rate of 2.40%. Despite the high effectiveness, there is potential to minimize incorrect alerts. Fault 2 recorded a precision of 98.06% and a detection rate of 97.98%, with a false positive rate of 1.94%, indicating robust system reliability. For Fault 3, the precision was 97.70%, and the detection rate was 97.58%, with a false positive rate of 2.30%. Although these figures are impressive, the false positive rate could still be improved. Fault 4 showed a precision of 98.48% and a detection rate of 98.78%, with a false positive rate of 1.52%, reflecting a good balance between precision and low false alarms. Fault 5 demonstrated a precision of 98.86% and a detection rate of 99.00%, with a false positive rate of only 1.14%.

Table 7. Results obtained with the pump at 50 Hz using the proposed fault diagnosis method.

Fault	Precision	Recall (Detection Rate)	False Positive Rate
1	97.60%	97.40%	2.40%
2	98.06%	97.98%	1.94%
3	97.70%	97.58%	2.30%
4	98.48%	98.78%	1.52%
5	98.86%	99.00%	1.14%

At 40 Hz, the results exhibited exceptional performance, displaying high precision and detection rates, as shown in Table 8. For Fault 1, a precision of 98.29% and a detection rate of 98.70% were attained, with a false positive rate of 1.72%. This signifies a high level of effectiveness in identifying faults. For Fault 2, the system achieved an impressive precision of 99.02% and a detection rate of 99.16%, with a low false positive rate of 0.98%. This confirms the system's high reliability. Fault 3 achieved a precision of 97.39% and a detection rate of 97.70%, with a false positive rate of 2.62%. Despite the high values, there is potential to reduce the false positive rate. Fault 4 exhibited a precision of 98.48% and a detection rate of 98.46%, with a false positive rate of 1.52%, demonstrating a good balance between precision and minimizing false alarms. Fault 5 showed excellent performance, with a precision of 98.92%, a detection rate of 99.02%, and a false positive rate of just 1.08%.

Table 8. Results obtained with the pump at 40 Hz using the proposed fault diagnosis method.

Fault	Precision	Recall (Detection Rate)	False Positive Rate
1	98.29%	98.70%	1.72%
2	99.02%	99.16%	0.98%
3	97.39%	97.70%	2.62%
4	98.48%	98.46%	1.52%
5	98.92%	99.02%	1.08%

At 30 Hz, the results consistently showed high precision and detection rates for the five faults (see Table 9). For Fault 1, a precision of 97.61% and a detection rate of 97.96% were achieved, with a false positive rate of 2.40%. This indicates high effectiveness, although there is room to reduce false alarms. In Fault 2, the results showed a precision of 98.24% and a detection rate of 98.24%, with a false positive rate of 1.76%. These results suggest high system reliability. Fault 3 presented a precision of 97.79% and a detection rate of 97.48%, with a false positive rate of 2.20%. Although the figures are high, the false positive rate could be improved. For Fault 4, the results showed a precision of 98.82% and a detection rate of 98.68%, with a false positive rate of 1.18%. This reflects an excellent balance between

precision and low false alarms. Fault 5 showed a precision of 98.68% and a detection rate of 98.64%, with a false positive rate of only 1.32%.

Table 9. Results obtained with the pump at 30 Hz using the proposed fault diagnosis method.

Fault	Precision	Recall (Detection Rate)	False Positive Rate
1	97.61%	97.96%	2.40%
2	98.24%	98.24%	1.76%
3	97.79%	97.48%	2.20%
4	98.82%	98.68%	1.18%
5	98.68%	98.64%	1.32%

To provide a comprehensive overview of the fault diagnosis system's performance at different frequencies, the results obtained at four operational frequencies (60 Hz, 50 Hz, 40 Hz, and 30 Hz) were averaged. The averaged results are presented in Table 10:

Table 10. Averaged results obtained at different operational frequencies.

Fault	Averaged Precision	Averaged Recall	Averaged False Positive Rate
1	98.00%	98.20%	2.02%
2	98.89%	98.63%	1.36%
3	97.60%	97.45%	2.40%
4	98.65%	98.65%	1.35%
5	98.91%	98.97%	1.09%

The averaged results reflect very consistent and robust performance of the fault diagnosis system under various operational conditions.

For Fault 1, an averaged precision of 98.00% and a detection rate of 98.20% were observed, with a false positive rate of 2.02%. Although the values are high, there is room to improve the false positive rate. Fault 2 showed an averaged precision of 98.89% and a detection rate of 98.63%, with a low false positive rate of 1.36%, suggesting high system reliability for this fault.

In the case of Fault 3, the averaged precision achieved is 97.60%, with a detection rate of 97.45% and a false positive rate of 2.40%. Although the figures are high, the false positive rate is the highest, indicating areas for improvement. For Fault 4, both the averaged precision and detection rate are 98.65%, and the false positive rate is 1.35%, indicating an excellent balance and system performance.

Finally, Fault 5 shows an average precision of 98.91% and a detection rate of 98.97%, with the lowest false positive rate of 1.09%, underscoring the high effectiveness and reliability of the diagnosis method. The fault diagnosis system has demonstrated high effectiveness and reliability across different operational frequencies, with consistently high precision and detection rates and low false positive rates. These results validate the robustness of the proposed method and its applicability under various operating conditions.

Table 11 compares various leak diagnosis methods and their respective accuracies. In [36], the authors utilized support vector machines (SVM) and achieved an accuracy of 90%. The approach by [37] combined SVM, decision trees, and Naive-Bayes, achieving 98.25%. In [38], the authors employed a deep neural network classifier, yielding an accuracy of 97.89%. In [39] deep learning with convolutional autoencoder networks was applied, achieving 96.67%. The proposed method integrating ANFIS and ZKF outperformed these methods with an accuracy of 98.41%. These results demonstrate that hybrid approaches

offer superior performance compared to traditional machine learning techniques such as SVM.

Table 11. Comparative analysis of leak diagnosis methods.

Reference	Method Used	Accuracy of Leak Diagnosis (%)
Sun et al., 2016 [36]	Support Vector Machines	90.00
El-Zahab et al., 2018 [37]	SVM, Decision Trees, and Naive-Bayes	98.25
Zadkarami et al., 2020 [38]	Deep Neural Network Classifier	97.89
Ahmad et al., 2022 [39]	Deep Learning Based on Convolutional Autoencoder Networks	96.67
Proposed Method	ANFIS and ZKF	98.41

5. Conclusions

Results obtained from the proposed data-driven method for fault and leak detection in hydraulic pipeline systems using neuro-fuzzy ZKF demonstrate high effectiveness and robustness. The proposed method combines the strengths of data-driven techniques with neuro-fuzzy observers, providing a robust solution for detecting anomalies in complex hydraulic systems.

Unlike other machine learning techniques, this method offers the notable advantage of relying exclusively on fault-free data. This makes the data collection process easier by eliminating the requirement for prior knowledge of faults. Additionally, it enhances the system's adaptability to various operating conditions. The proposed approach began with identifying the nonlinear pipeline system using an adaptive ANFIS. This system generates Takagi–Sugeno fuzzy models based on measured pressure and flow data, effectively capturing the nonlinear dynamics of the pipeline. These models are then integrated into a neuro-fuzzy ZKF framework, allowing for adaptive thresholding and enhanced fault detection capabilities. The proposed method takes into account parametric uncertainty due to hydraulic friction and fluid turbulence as well as measurement noise, ensuring robustness in real-world applications. Fault isolation is achieved by comparing zonotopic sets and evaluating a matrix of fault signatures, enabling precise identification of various fault types.

The experimental validation of the proposed method demonstrated high precision (up to 99.24%) and recall (up to 99.20%) across various fault scenarios and operational points, indicating its capability to accurately identify sensor faults and pipeline leaks. Additionally, with false positive rates as low as 0.76%, the method minimized incorrect alerts, enhancing its reliability for practical applications.

Moreover, the method's successful application across different operational frequencies (60 Hz, 50 Hz, 40 Hz, and 30 Hz) demonstrated its scalability and applicability under various operating conditions. This versatility is crucial for practical deployment, ensuring that the method can be adapted to different scenarios without significant modifications. Our method's ability to scale across different operational conditions further enhances its practical utility, making it a valuable contribution to the field of fault detection and diagnosis in hydraulic systems.

Although the proposed method shows significant promise, several avenues for future research have been identified. Extending the method to detect and isolate other types of faults in hydraulic systems, such as valve malfunctions or pump failures, would broaden its applicability. Further research could also focus on optimizing the neuro-fuzzy models to improve the accuracy and speed of fault detection and isolation. Exploring advanced data fusion techniques might enhance the method's ability to handle complex and noisy datasets, potentially improving detection performance.

By pursuing these future directions, the capabilities and applicability of the proposed fault detection method could be further enhanced, ensuring its relevance and effectiveness in a wide range of practical scenarios.

Author Contributions: Conceptualization, E.-J.P.-P. and I.S.-R.; methodology, E.-J.P.-P.; software, Y.G.-B.; validation, E.-J.P.-P., I.S.-R. and Y.G.-B.; formal analysis, E.-J.P.-P.; investigation, J.-A.F.-M.; data curation, J.-A.G.-R.; writing—original draft preparation, E.-J.P.-P.; writing—review and editing, I.S.-R.; visualization, Y.G.-B. and J.-A.F.-M.; funding acquisition, I.S.-R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Tecnológico Nacional de México (TecNM) grant number 20212.24-P and Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCYT) in Mexico.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Acknowledgments: The authors thank the RICCA (Red Internacional de Control y Cómputo Aplicados) research network for their invaluable scientific support.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Qian, J.; Song, Z.; Yao, Y.; Zhu, Z.; Zhang, X. A review on autoencoder based representation learning for fault detection and diagnosis in industrial processes. *Chemom. Intell. Lab. Syst.* **2022**, *231*, 104711. [[CrossRef](#)]
2. Cheded, L.; Doraiswami, R. A novel integrated framework for fault diagnosis with application to process safety. *Process Saf. Environ. Prot.* **2021**, *154*, 168–188. [[CrossRef](#)]
3. Datta, S.; Sarkar, S. A review on different pipeline fault detection methods. *J. Loss Prev. Process Ind.* **2016**, *41*, 97–106. [[CrossRef](#)]
4. Jul-Jørgensen, I.; Facco, P.; Gernaey, K.; Barolo, M.; Hundahl, C. Data fusion of Raman spectra in MSPC for fault detection and diagnosis in pharmaceutical manufacturing. *Comput. Chem. Eng.* **2024**, *184*, 108647. [[CrossRef](#)]
5. Eskandari, A.; Milimonfared, J.; Aghaei, M. Fault detection and classification for photovoltaic systems based on hierarchical classification and machine learning technique. *IEEE Trans. Ind. Electron.* **2020**, *68*, 12750–12759. [[CrossRef](#)]
6. Fawwaz, D.Z.; Chung, S.H. Real-time and robust hydraulic system fault detection via edge computing. *Appl. Sci.* **2020**, *10*, 5933. [[CrossRef](#)]
7. Askari, B.; Carli, R.; Cavone, G.; Dotoli, M. Data-driven fault diagnosis in a complex hydraulic system based on early classification. *IFAC-PapersOnLine* **2022**, *55*, 187–192. [[CrossRef](#)]
8. Xue, P.; Jiang, Y.; Zhou, Z.; Chen, X.; Fang, X.; Liu, J. Machine learning-based leakage fault detection for district heating networks. *Energy Build.* **2020**, *223*, 110161. [[CrossRef](#)]
9. Kim, D.; Heo, T.Y. Anomaly detection with feature extraction based on machine learning using hydraulic system IoT sensor data. *Sensors* **2022**, *22*, 2479. [[CrossRef](#)]
10. Kościelny, J.M.; Bartyś, M. A new method of diagnostic row reasoning based on trivalent residuals. *Expert Syst. Appl.* **2023**, *214*, 119116. [[CrossRef](#)]
11. Meseguer, J.; Puig, V.; Escobet, T. Fault diagnosis using a timed discrete-event approach based on interval observers: Application to sewer networks. *IEEE Trans. Syst. Man Cybern. A Syst. Hum.* **2010**, *40*, 900–916. [[CrossRef](#)]
12. Tahraoui, S.; Meghebbar, A.; Boubekur, D.; Boumediene, A. Fault Detection in a Five Tank Hydraulic System. *Electroteh. Electron. Autom.* **2015**, *63*, 51.
13. Youssef, T.; Chadli, M.; Karimi, H.R.; Wang, R. Actuator and sensor faults estimation based on proportional integral observer for TS fuzzy model. *J. Frankl. Inst.* **2017**, *354*, 2524–2542. [[CrossRef](#)]
14. Jafari, R.; Razvarz, S.; Vargas-Jarillo, C.; Gegov, A. Blockage detection in pipeline based on the extended Kalman filter observer. *Electronics* **2020**, *9*, 91. [[CrossRef](#)]
15. Tao, H.; Jia, P.; Wang, X.; Wang, L. Real-Time Fault Diagnosis for Hydraulic System Based on Multi-Sensor Convolutional Neural Network. *Sensors* **2024**, *24*, 353. [[CrossRef](#)]
16. Rousseau, P.; Laubscher, R. A Condition-Monitoring Methodology Using Deep Learning-Based Surrogate Models and Parameter Identification Applied to Heat Pumps. *Math. Comput. Appl.* **2024**, *29*, 52. [[CrossRef](#)]
17. Qiu, S.; Cui, X.; Ping, Z.; Shan, N.; Li, Z.; Bao, X.; Xu, X. Deep learning techniques in intelligent fault diagnosis and prognosis for industrial systems: A review. *Sensors* **2023**, *23*, 1305. [[CrossRef](#)] [[PubMed](#)]

18. Frausto-Solís, J.; Galicia-González, J.C.d.J.; González-Barbosa, J.J.; Castilla-Valdez, G.; Sánchez-Hernández, J.P. SSA-Deep Learning Forecasting Methodology with SMA and KF Filters and Residual Analysis. *Math. Comput. Appl.* **2024**, *29*, 19. [\[CrossRef\]](#)
19. Robles-Velasco, A.; Cortés, P.; Muñuzuri, J.; De Baets, B. Prediction of pipe failures in water supply networks for longer time periods through multi-label classification. *Expert Syst. Appl.* **2023**, *213*, 119050. [\[CrossRef\]](#)
20. Askari, B.; Bozza, A.; Cavone, G.; Carli, R.; Dotoli, M. An adaptive constrained clustering approach for real-time fault detection of industrial systems. *Eur. J. Control* **2023**, *74*, 100858. [\[CrossRef\]](#)
21. Barakat, N.; Hajeir, L.; Alattal, S.; Hussein, Z.; Awad, M. Air leaks fault detection in maintenance using machine learning. *J. Qual. Maint. Eng.* **2024**, *30*, 391–408. [\[CrossRef\]](#)
22. Le, V.T.H.; Stoica, C.; Alamo, T.; Camacho, E.F.; Dumur, D. *Zonotopes: From Guaranteed State-Estimation to Control*; John Wiley & Sons: Hoboken, NJ, USA, 2013.
23. Combastel, C. A state bounding observer based on zonotopes. In Proceedings of the 2003 European Control Conference (ECC), Cambridge, UK, 1–4 September 2003; IEEE: Cambridge, UK, 2003; pp. 2589–2594.
24. Combastel, C. A state bounding observer for uncertain non-linear continuous-time systems based on zonotopes. In Proceedings of the 44th IEEE Conference on Decision and Control, Seville, Spain, 12–15 December 2005; IEEE: Seville, Spain, 2005; pp. 7228–7234.
25. Combastel, C. Zonotopes and Kalman observers: Gain optimality under distinct uncertainty paradigms and robust convergence. *Automatica* **2015**, *55*, 265–273. [\[CrossRef\]](#)
26. Haznedar, B.; Kalinli, A. Training ANFIS structure using simulated annealing algorithm for dynamic systems identification. *Neurocomputing* **2018**, *302*, 66–74. [\[CrossRef\]](#)
27. Zhang, W.; Wang, Z.; Guo, S.; Shen, Y. Interval estimation of sensor fault based on zonotopic Kalman filter. *Int. J. Control* **2021**, *94*, 1641–1650. [\[CrossRef\]](#)
28. Chaudhry, M.H. *Applied Hydraulic Transients*; Springer: Berlin/Heidelberg, Germany, 2014; Volume 415.
29. White, F.M.; Ng, C.; Saimek, S. *Fluid Mechanics*; McGraw-Hill, Cop.: New York, NY, USA, 2011.
30. Zeghadnia, L.; Robert, J.L.; Achour, B. Explicit solutions for turbulent flow friction factor: A review, assessment and approaches classification. *Ain Shams Eng. J.* **2019**, *10*, 243–252. [\[CrossRef\]](#)
31. Santos-Ruiz, I.; Bermúdez, J.R.; López-Estrada, F.R.; Puig, V.; Torres, L.; Delgado-Aguíñaga, J. Online leak diagnosis in pipelines using an EKF-based and steady-state mixed approach. *Control Eng. Pract.* **2018**, *81*, 55–64. [\[CrossRef\]](#)
32. Takagi, T.; Sugeno, M. Fuzzy identification of systems and its applications to modeling and control. *IEEE Trans. Syst. Man Cybern.* **1985**, *SMC-15*, 116–132. [\[CrossRef\]](#)
33. Chen, J.; Patton, R.J. *Robust Model-Based Fault Diagnosis for Dynamic Systems*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012; Volume 3.
34. Alamo, T.; Bravo, J.M.; Camacho, E.F. Guaranteed state estimation by zonotopes. *Automatica* **2005**, *41*, 1035–1043. [\[CrossRef\]](#)
35. Savva, M.; Ioannou, I.; Vassiliou, V. Fuzzy-logic based IDS for detecting jamming attacks in wireless mesh IoT networks. In Proceedings of the 2022 20th Mediterranean Communication and Computer Networking Conference (MedComNet), Pafos, Cyprus, 1–3 June 2022; IEEE: Pafos, Cyprus, 2022; pp. 54–63.
36. Sun, J.; Xiao, Q.; Wen, J.; Zhang, Y. Natural gas pipeline leak aperture identification and location based on local mean decomposition analysis. *Measurement* **2016**, *79*, 147–157. [\[CrossRef\]](#)
37. El-Zahab, S.; Abdelkader, E.M.; Zayed, T. An accelerometer-based leak detection system. *Mech. Syst. Signal Process.* **2018**, *108*, 276–291. [\[CrossRef\]](#)
38. Zadkarami, M.; Safavi, A.A.; Taheri, M.; Salimi, F.F. Data driven leakage diagnosis for oil pipelines: An integrated approach of factor analysis and deep neural network classifier. *Trans. Inst. Meas. Control* **2020**, *42*, 2708–2718. [\[CrossRef\]](#)
39. Ahmad, S.; Ahmad, Z.; Kim, C.H.; Kim, J.M. A method for pipeline leak detection based on acoustic imaging and deep learning. *Sensors* **2022**, *22*, 1562. [\[CrossRef\]](#) [\[PubMed\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.