Available online at www.sciencedirect.com

ScienceDirect

IFAC-PapersOnline 58-4 (2024) i–vii

International Federation of Automatic Control

12th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes SAFEPROCESS 2024

Ferrara, Italy, June 4 – 7, 2024

PROCEEDINGS

Edited by
Louise Travé-Massuyès
LAAS-CNRS, University of Toulouse, France

ELSEVIER

IFAC Proceedings Volumes (IFAC-PapersOnline) — ISSN 2405-8963

IFAC PapersOnline — ISSN 2405-8963

Published by:
International Federation of Automatic Control (IFAC)
Hosting by Elsevier Ltd.

Available online at

www.sciencedirect.com

Publication date
June 2024

Copyright conditions

12th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes SAFEPROCESS 2024

Sponsored by

International Federation of Automatic Control (IFAC)
- Fault Detection, Supervision & Safety of Techn. Processes-SAFEPROCESS, TC 6.4.

Co-Sponsored by

International Federation of Automatic Control (IFAC)
- TC 1.1. Modelling, Identification and Signal Processing
- TC 1.3. Discrete Event and Hybrid Systems
- TC 4.2. Mechatronic Systems
- TC 6.1. Chemical Process Control
- TC 6.2. Mining, Mineral and Metal Processing
- TC 6.3. Power and Energy Systems
- TC 7.2. Marine Systems
- TC 7.3. Aerospace
- TC 7.4. Transportation Systems
- TC 7.5. Intelligent Autonomous Vehicles

Supporting Organizations

- ANIPLA – National Italian Association of Automation
- AIMAN – Italian Association of Maintenance
- SIDRA – Italian Control Systems Society
- UNIFE – The University of Ferrara
- EnDiF – Engineering Department of Ferrara
- MathWorks
- The Municipality of Ferrara

International Programme Committee

Cristina Verde, MX (IPC Chair)
Neng Eva Wu, USA (IPC Co-Chair)
Vasso Reppa, NL (IPC Vice-Chair from Industry)
Louise Travé-Massuyès, FR (Editor)
Alexandra Grancharova, BG, and Ali Zolghadri, FR (Publicity Chairs)

Ahmad Al-Dabbagh,CA
Christophe Berenguer,FR
Gildas Besancon,FR
Gautam Biswas,USA
Mogens Blanke,DK
Jozsef Bokor,HU
Richard Braatz,USA
Joao M. F. Calado,PT
Alberto Cardoso,PT
Alessandro Casavola,IT
Paolo Castaldi,IT
Mariela Cerrada,EC
Steven X. Ding,DE
Christopher Edwards,UK
Riccardo M.G. Ferrari,NL
Erik Frisk,SE
Alessandro Giua,IT
Philippe Goupil,FR
Alexandra Grancharova,BG
Biao Huang,CA
Bin Jiang,CN
Jing Jiang,CA
Michel Kinnaert,BE
Jozef Korbicz,PL
Jan Koscielny,PL
Zdzislaw Kowalczuk,PL
Orestes Llanes-Santiago,CU
Rami S. Mangoubi,USA
Didier Maquin,FR
John J. Martinez Molina,FR
Henrik Niemann,DK
Peter Fogh Odgaard,DK
Ron J. Patton,UK
Gustavo Pérez Zuñiga,PE
Vincent Cocquempot,FR
Vicenç Puig,ES
Ivo Puncochar,CZ
Dominique D.J. Sauter,FR
Jakob Stoustrup,DK
Audine Subias,FR
André M.H. Teixeira,SE
Didier Theilliol,FR
Louise Travé-Massuyès,FR
Michel Verhaegen,NL
Marcin Witczak,PL
Ping Zhang,DE
Qinghua Zhang,FR
Youmin Zhang,CA
Alexey N. Zhirabok,RU
Donghua Zhou,CN
Ali Zolghadri, FR

National Organizing Committee

Silvio Simani, IT (NOC Chair)
Alessandro Casavola, IT (NOC Co-Chair)
Paolo Pasini, IT (NOC Vice-Chair from Industry)

Tomas Parisini, IT
Marcello Bonfè, IT
Elena Zattoni, IT
Saverio Farsoni, IT
Margherita Franzoni, IT
Alessandro Giua, IT
Elena Mainardi, IT

FOREWORD

The IFAC SAFEPROCESS 2024 is continuing the successful series of symposia held in Baden-Baden (Germany, 1991), Helsinki (Finland, 1994), Hull (UK, 1997), Budapest (Hungary, 2000), Washington DC (USA, 2003), Beijing (China, 2006), Barcelona (Spain, 2009), Mexico City (Mexico, 2012), Paris (France, 2015), Warsaw (2018), and Paphos (Cyprus, 2022). The Department of Engineering of the University of Ferrara, Italy, organized the IFAC SAFEPROCESS 2024 in Ferrara, Italy, for the first time since its first edition, on June 4–7, 2024.

The theory and practice of control and technical diagnostics are facing big problems as the complexity of modern industrial systems and processes keeps growing. The need for greater reliability in their operation, control quality, and security is also growing. Early detection and diagnosis of faults and cyberattacks are critical to avoid performance degradation and damage to machinery or human life.

The SAFEPROCESS symposium is a triennial IFAC meeting and a major international gathering of leading academic and industry experts from all over the world. It aims at strengthening the contact between academia and industry to build up new networks and cultivate existing relations. High-level speakers have given talks on a wide spectrum of topics related to fault diagnosis, process supervision, safety monitoring, fault-tolerant control, cyber-security, as well as state-of-the-art applications and emerging research directions. The symposium has also served as a forum for young researchers, giving them the opportunity to present their scientific ambitions and work to an audience consisting of international technical diagnostics and control communities.

Fault diagnosis and fault-tolerant control have developed into major research areas at the intersection of system and control engineering, computer science, applied mathematics and statistics, or soft computing, as well as application fields such as mechanical, electrical, chemical, and aerospace engineering. IFAC is recognised as playing a crucial role in this aspect by launching a triennial symposium dedicated to this subject.

The SAFEPROCESS 2024 program can be accessed at https://www.safeprocess2024.eu/#. The program consisted of nineteen regular and five invited sessions on three parallel tracks. It also contained six keynote talks prepared by outstanding academics and industrials who introduced advanced results on fault diagnosis, fault-tolerant control, root cause analysis, and cyberattack prevention. In particular, Prof. Ron J. Patton (UK), from the Univ. of Hull, delivered the speech entitled “Offshore Wind Turbine Rotor Imbalance, a Fault-Tolerant Control Problem”; Prof. Christophe Combastel (FR), from the Univ. of Bordeaux, delivered the speech entitled “ Reachability and Filtering for Safe Processes: From Zonotopes to Functional Sets with Typed Symbols”; Prof. Roger Dixon (UK), from the Univ. of Birmingham, gave the talk “Fault Tolerance in Railways: “The Evolution of a Radical Next Generation Track Switch”; Dr. Steinert Olof (SE) from Scania, delivered the speech “Harnessing Data for Predictive Maintenance and Collaboration, Boost Innovation”; Prof. Biao Huang (CA), from the Univ. of Alberta, talked about “Advancing Causal Analysis for Fault Detection and Root Cause Analysis in Process Systems Engineering”; and Prof. Ping Zhang (DE), from the Univ. of Kaiserslautern-Landau, gave the talk “Detection and Avoidance of Cyber Attacks on Industrial Control Systems”.

The symposium received one hundred seventy-eight submissions, divided into three sets: one hundred forty regular papers, thirty-one invited papers, and seven invited sessions. The rejection rate for submissions was 23%. The symposium had one hundred eighty-seven participants, including one hundred twenty-two academics and sixty-five students. Regarding the statistics, we identified an average of 3.6 authors per paper. The number of participating countries was forty-eight. Regrettably, we failed to achieve a satisfactory balance between geographical regions. The countries with the most papers, in decreasing order, are China, France, Germany, Italy, Spain, Sweden, the United Kingdom, the United States of America, the Netherlands, and Mexico.

One pre-symposium tutorial, a roundtable, and a benchmark competition were also included in the technical program. As a result, Vasso Reppa from Delf University of Technology, Mayank S. Jha, and Didier Theilliol from the University of Lorraine organized the roundtable titled "Gnosis for Maintenance: From Diagnosis to Prognosis and Health-Aware Control." The meeting was very active, with comments and questions from more than fifty attendees in a two-hour session. Additionally, Eric Frisk, Daniel Jung, and Mattias Krysander from Linköping University organised a competition on fault detection and isolation techniques with incomplete data. The airflow system of an internal combustion engine was considered an industrial benchmark. The competition was intriguing and a good motivating example for young researchers. A two-hour special session presented the results of the six participants. The young researchers Nicolas Anselmi, Andrea Arici, Francesco Corrini, and Mirko Mazzolen from the University of Bergamo, Italy, took first place in the competition, and the two next classified also obtained a diploma.

Regarding the pre-symposium tutorials, three proposals were received; however, only one met the minimum registration quota defined by the organizers. Linlin Li from the University of Science and Technology Beijing, Zhiwen Chen from Central South University, and Steven X. Ding from the University of Duisburg-Essen integrated the tutorial, entitled "Control Theory-Informed Machine Learning for Fault Diagnosis in Dynamic Control Systems." The tutorial was free for students, and there were thirty-three attendees.

The symposium recognized three awards: the *Paul M. Frank Theory Paper Award* given to Louis Goupil, Louise Travé-Massuyès, Elodie Chanthery, Thibault Kohler, Sébastien Delautier for the paper entitled “Tree-Based Diagnosis Enhanced with Meta Knowledge Applied to Dynamic Systems”; the *IFAC Young Author Award* given to Henrik Sebastian Steude*, Lukas Moddemann, Alexander Diedrich, Jonas Ehrhardt, Oliver Niggemann for the paper entitled “Diagnosis Driven Anomaly Detection for Cyber-Physical Systems”; finally, the *Best Application Paper Award* was given to Andrea Mattioni, Lucas José da Silva Moreira, Herve Yves Guy Bernard Louis Roustan, Gildas Besancon, Mirko Fiacchini for the paper entitled “A step towards implementation of state observers in industrial aluminium smelters”.

SAFEPROCESS 2024 was the first IFAC SAFEPROCESS symposium to be streamed thanks to the University of Ferrara YouTube channel, enabling researchers and practitioners to participate either physically or online. As a result, the sessions are still accessible through the complete playlist at www.youtube.com/playlist?list=PLL80i9P61J-O-4-Y79u-KKkoHybZw9d0k. The presentations provided participants with an invaluable opportunity to learn from the knowledge and experiences of world-renowned scientists and experts. Covering a range of exciting topics, these sessions generated ideas, concepts, and methods that will make future industrial systems and processes more efficient and safer.

As International Programme Committee Chair and General Chair, we are filled with immense pride and joy as we reflect on the success of this remarkable event. The hard work, dedication, and collaborative spirit of everyone involved have truly paid off, creating an unforgettable experience for all participants. We extend our heartfelt gratitude to all who contributed, and we look forward to many more successful IFAC SAFEPROCESS symposia in the future.

Cristina Verde
International Program Committee Chair

Silvio Simani
General Chair

Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 58-4 (2024) 408–413

Leak detection for household pipelines based on a smart valve with single pressure and flow sensors

D. Cesani * **M. Mazzoleni** * **F. Previdi** *

* *Department of Management, Information and Production engineering University of Bergamo, via Marconi 5, 24044 Dalmine (BG), Italy.*

Abstract: Household pipelines leaks are a major source of water wastage, and several leak detection techniques have been developed. As a result, manufacturers started to produce smart valves, able to autonomously detect water leaks in house plumbing networks. A smart valve is an electronically controllable valve provided with some sensors, that must be closed if a leak is detected. The valve closure is controlled by an embedded computational unit that analyses the valve sensors data with a Leak Detection system composed by a set of algorithms. This work proposes a Leak Detection system to monitor a household plumbing network both when an house utility is opened and when all utilities are closed, considering a single flow sensor and single pressure sensor in the smart valve. The proposed Leak Detection system consists of three algorithms: (1) an `if-then-else` construct (2) a fault detection algorithm based on a logistic regression model, and (3) an anomaly detection algorithm based on the multivariate Gaussian distribution. The effectiveness of the Leak Detection system is experimentally evaluated on a laboratory house plumbing network, considering different utilities.

Keywords: Leak detection; Fault Detection; Anomaly Detection; Water Networks.

1. INTRODUCTION

Leaks in pipelines are a well-known problem that causes waste of resources and often environmental catastrophes. Chemical, oil industries and distributing companies have spurred the researchers to find effective solutions. Over the last 40 years, the researchers answered the call developing several leak detection techniques, ranging from simple physical inspection and acoustic methods to infrared thermography, optic sensors located along the entire pipelines and transient techniques (Verde and Torres, 2004). Recently, statistical and artificial intelligence based approaches have flanked the existing techniques to improve leak detection accuracy on all levels (El-Zahab and Zayed, 2019; Mazzoleni et al., 2022a).

However, despite the household leaks problem is a major source of wastage, the greatest efforts have been directed to solve the leak problem on large-scale liquids and gasses transport systems, e.g., pipelines for petroleum products or city water networks (Billmann and Isermann, 1987; Da Silva et al., 2005; El-Zahab and Zayed, 2019). Only in the last decade, the research began to focus on the development of leak detection techniques for housing contexts (Daadoo et al., 2017; Sakti et al., 2021; Rosli et al., 2018). As a result, many plumbing valves manufacturers started to develop *smart valves*, able to detect a flow leakage in house plumbing network. A smart valve is an electronically controllable valve, typically provided with a set of sensors and an embedded computational unit (ECU). To reduce costs, the number of sensors installed on the smart valves must be limited as much as possible. The ECU analyses the raw sensors data with a Leak Detection (LD) system, that consists of a set of algorithms. If a leak is detected, the smart valve must be closed by the ECU. A smart valve can be installed into several household utilities, as sinks, radiators, or on the connection pipeline to the water supply network. Since each plumbing network has its own piping-dependent characteristics, a smart valve installation includes the calibration of the LD system algorithms to best fit the plumbing network of interest. This is done performing some calibration tests.

In this work, we propose a LD system for *housing contexts* where a smart valve is placed on the connection pipeline between the plumbing network and the water supply network. The aim is to *monitor on-line* the plumbing network both when it is in use, i.e., an utility is opened, and when it is not i.e., no opened utilities. In particular, we consider the case where the smart valve is provided only with *a single flow and pressure sensor*, as well as the embedded ECU. The proposed on-line LD system relies on a recursive least-squares (RLS) estimation of a dynamic model characterizing the relation between the monitored pressure and flow. The LD system is composed of three algorithms: (1) an `if-then-else` construct, (2) a Fault Detection (FD) algorithm based on a logistic regression model, and (3) an Anomaly Detection (AD) algorithm employing a multivariate Gaussian distribution. The `if-then-else` construct enables the leaks detection when the plumbing network is not in use. The remaining algorithms employ features computed on the parameters estimated by the RLS algorithm, and are devoted to detecting leaks when the plumbing network is in use. The proposed LD system does not tackle the isolation of the leak type, i.e. which utility presents a leak in its pipe.

10.1016/j.ifacol.2024.07.252

We note that FD and AD algorithms are widely employed in industry, with the former requiring experimental tests for data collection both on the healthy and the faulty plumbing network [1] (Valceschini et al., 2022), and the latter requiring experiments only on the healthy plumbing network (Mazzoleni et al., 2022b).

The remainder of the paper is organised as follows. Section 2 presents the theoretical fundamentals of the algorithms in the proposed LD system. Section 3 and Section 4 define the proposed LD system and show its effectiveness on experimental data, respectively. Section 5 is devoted to concluding remarks and future directions.

2. PRELIMINARIES

This section reviews the main theoretical foundations of the algorithms that compose the proposed LD system.

2.1 Review of Recursive Least Squares system identification

Recursive Least Squares (RLS) is a recursive identification approach commonly employed when it is necessary to on-line update the model of a dynamical system while in operation (Ljung, 1999). The measured input/output data are processed recursively as they become available and, as a result, the model parameters estimations are adjusted recursively over time. Consider a generic SISO ARX($n_a, n_b, k = 1$) model, with $t \in \mathbb{N}_{>0}$ the discrete-time index:

$$\begin{aligned} y(t) &= a_1 y(t-1) + \cdots + a_{n_a} y(t-n_a) + \\ &+ b_1 u(t-1) + \cdots + b_{n_b} u(t-n_b-1) + e(t), \\ &= \boldsymbol{\varphi}^\top(t)\boldsymbol{\theta} + e(t), \end{aligned} \tag{1}$$

where $e(t) \sim \mathrm{WN}(0, \lambda^2)$ is a white noise with variance λ^2, $\boldsymbol{\varphi}(t) = [y(t-1)\ \ldots\ y(t-n_a)\ u(t-1)\ \ldots\ u(t-n_b-1)]^\top \in \mathbb{R}^d$ is the regressors vector, and the model parameters are collected in the vector $\boldsymbol{\theta} = [a_1 \ldots a_{n_a} b_1 \ldots b_{n_b}]^\top \in \mathbb{R}^d$.

The RLS method provides the optimal (in a least-square sense) parameters estimate $\hat{\boldsymbol{\theta}}(t)$ of $\boldsymbol{\theta}(t)$ each time a new datum is available, by updating the previous estimate $\hat{\boldsymbol{\theta}}(t-1)$ as

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + K(t) \cdot \varepsilon(t), \tag{2}$$

with the *algorithm gain*

$$K(t) = P(t)\boldsymbol{\varphi}(t) = \frac{P(t-1)\boldsymbol{\varphi}(t)}{\mu + \boldsymbol{\varphi}^\top(t)P(t-1)\boldsymbol{\varphi}(t)}, \tag{3}$$

and

$$\begin{aligned} P(t) &= S(t)^{-1} \\ &= \frac{1}{\mu}\left(S(t-1)^{-1} - \frac{S(t-1)^{-1}\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^\top(t)S(t-1)^{-1}}{\mu + \boldsymbol{\varphi}^\top(t)P(t-1)\boldsymbol{\varphi}(t)}\right), \end{aligned} \tag{4}$$

where $\mu \in [0, 1]$ is the *forgetting factor* that keeps the estimation reactive to parameters variations. The quantity $\varepsilon(t) = y(t) - \boldsymbol{\varphi}^\top(t)\hat{\boldsymbol{\theta}}(t-1)$ is the *prediction error* at time t and $S(t) = \sum_{i=1}^{t} \boldsymbol{\varphi}(i)\boldsymbol{\varphi}^\top(i)$ is an auxiliary matrix. Assuming that, at certain t_0, the matrix $S(t_0)$ is invertible, an appropriate initialisation of the algorithm is given by

$$P(t_0) = S(t_0)^{-1} = \left(\sum_{i=1}^{t_0} \boldsymbol{\varphi}(i)\boldsymbol{\varphi}^\top(i)\right)^{-1} \tag{5}$$

$$\hat{\boldsymbol{\theta}}(t_0) = K(t_0)\varepsilon(t_0) = P(t_0)\sum_{i=1}^{t_0} \boldsymbol{\varphi}(i)y(i) \tag{6}$$

2.2 Review of logistic regression

Logistic regression is a well-known model for binary classification tasks. Let $\mathcal{C}_1$ and $\mathcal{C}_2$ the two considered classes and define the logistic sigmoid function $\sigma(\cdot)$ as

$$\sigma(a) = \frac{1}{1 + e^{-a}}, \tag{7}$$

with $a \in \mathbb{R}$ a generic function argument. The probability of a regressors vector $\boldsymbol{x} = [x_0\ x_1 \ldots x_{d-1}]^\top \in \mathbb{R}^d$ to belong on the class $\mathcal{C}_1$, where $x_i, i = 0, \ldots, d-1$ are the individual features with $x_0 = 1$, can be expressed as a linear combination of the features in $\boldsymbol{x}$ with weights vector $\boldsymbol{w} \in \mathbb{R}^d$ evaluated in the logistic sigmoid function (7), so that

$$p(\mathcal{C}_1|\boldsymbol{x}) = \sigma(\boldsymbol{x}^\top\boldsymbol{w}), \tag{8}$$

with $p(\mathcal{C}_2|\boldsymbol{x}) = 1 - p(\mathcal{C}_1|\boldsymbol{x})$. The model (8) is known as logistic regression model where the vector $\boldsymbol{w} = [w_1 \ldots w_d]^\top \in \mathbb{R}^d$ represents its parameters. The logistic regression model defines a linear decision boundary in the features space separating the two classes: if $p(\mathcal{C}_1|\boldsymbol{x}) \geq 0.5$ then $\boldsymbol{x} \in \mathcal{C}_1$, else $\boldsymbol{x} \in \mathcal{C}_2$. However, the parameters in $\boldsymbol{w}$ are unknown. Using a numerical method, it is possible to solve the maximum likelihood estimation problem, so as to obtain a parameters estimation $\hat{\boldsymbol{w}}$ (Bishop, 2016).

2.3 Review of multivariate Gaussian distributions

Let $\boldsymbol{z} \in \mathbb{R}^n$ a n-dimensional vector of random variables.

Suppose that $\boldsymbol{z}$ has a multivariate Gaussian distribution with mean $\boldsymbol{\mu} \in \mathbb{R}^n$ and positive definite covariance matrix $\Sigma \in \mathbb{R}^{n\times n}$, i.e., $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$. The probability density function (pdf) of $\boldsymbol{z}$ is

$$p(\boldsymbol{z}; \boldsymbol{\mu}, \Sigma) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(\boldsymbol{z}-\boldsymbol{\mu})^\top\Sigma(\boldsymbol{z}-\boldsymbol{\mu})}. \tag{9}$$

However, the parameters $\boldsymbol{\mu}$ and Σ might be unknown and so it is not possible to compute the value of $p(\boldsymbol{z}; \boldsymbol{\mu}, \Sigma)$. Given a set of N realisations of $\boldsymbol{z}$, $\{\boldsymbol{z}(1), \ldots, \boldsymbol{z}(N)\}$, estimates $\hat{\boldsymbol{\mu}}$, $\hat{\Sigma}$ of $\boldsymbol{\mu}$ and Σ, respectively, can be computed by the Expectation-Maximization (EM) algorithm, that also generalizes to the case of a pdf expressed as a mixture of Gaussian distributions (Bishop, 2016).

3. LEAK DETECTION SYSTEM FOR HOUSEHOLD PIPELINE NETWORKS

This section presents the proposed LD system for household pipelines based on a smart valve with single pressure and flow sensors. The cases where all water network utilities are closed and where at least one utility is used are considered. An experimental evaluation is presented next.

3.1 Generalities of house networks and smart valves

We consider a house plumbing network composed by N_U utilities connected with pipes.The entire system is assumed connected to the water supply network via a smart valve,

[1] With healthy and faulty plumbing network we mean a network of pipes without any leak and with a leak, respectively.

equipped with a flow sensor to measure the flow $\mathcal{F}$ [ℓ/min] through the plumbing system, and a pressure sensor to measure the water pressure $\mathcal{P}$ [bar]. Flow and pressure measurements are acquired with a fixed sampling time $\mathcal{T}_s$ [s]. In practical smart valves applications, such values are typically stored at the last positions of a buffer $\mathcal{B}_f$ for the flow measurements and a buffer $\mathcal{B}_p$ for the pressure ones. The two buffers have the same dimension N_B, that is defined in the design phase of the LD system. The buffers are managed as a First In First Out (FIFO) queues.

Let T_o be a threshold on the amount of flow through the plumbing network. At a time t, if $\mathcal{F}(t) < T_o$, all utilities are considered closed. Typically, T_o is set during the design phase of a LD system, to a value very close to 0 ℓ/min, e.g. the flow measurement corresponding to a dripping taps. Let $\mathcal{O}_{\%,i} \in [0, 100]_{\%},\ i = 1, ..., N_U$, be the opening percentage of the i-th utility. At a time instant t, a house plumbing network can be in one of two conditions:

(1) *Stationary.* The flow through the system is such that $0 < \mathcal{F}(t) < T_o$, and all the utilities are closed, $\mathcal{O}_{\%,i} = 0\%,\ i = 1, ..., N_U$.
(2) *In use.* There is a flow $\mathcal{F}(t) \geq T_o \neq 0$ through the system and there is one or more utilities opened, $\exists i : \mathcal{O}_{\%,i} \neq 0\%,\ i = 1, ..., N_U$.

These two conditions act as the skeleton of the proposed Leak Detection (LD) system, that is discussed next.

3.2 Stationary plumbing network

Considering the *stationary* plumbing network condition, the proposed LD system reduces to the `if-then-else` construct in Algorithm 1, where the parameter $\mathcal{S}$ represents the time duration over which the fault condition must be maintained before an alarm is raised, and can be set by the user of the smart valve. The counter q measures how long the fault condition has persisted, if any.

In this stationary network condition, only the last position of $\mathcal{B}_f$ is checked, i.e., the newest flow measurement. The pressure measurements are not considered. The auxiliary variable q is set to 0 when the system is starting. All flows such that $0 < \mathcal{F}(t) < T_o$ are considered as a leak.

Algorithm 1: `if-then-else` construct of the proposed LD system

```
Input: B_f
Output: faulty/healthy system classification
q = 0
if B_f(N_B) > 0 then
  | q = q + 1
else if q > 0 then
  | q = q − 1
end
if q · T_s == S then
  | faulty system classification
  | q = 0
else
  | healthy system classification
end
```

3.3 In use plumbing network

To handle the *in use* plumbing network case, we propose two different algorithms:

- a *Fault Detection* (FD) algorithm with a logistic regression model that outputs a binary decision (healthy/faulty). This algorithm requires both data without any leak (healthy data), and data with a leak in one of the N_U utilities of interest (faulty data);
- an *Anomaly Detection* (AD) algorithm employing a multivariate Gaussian distribution to model healthy data. This algorithm requires only healthy data.

Both algorithms share a raw flow and pressure measurements pre-processing phase, described hereafter. Let t_o a time instant at which $\mathcal{F}(t_o) > T_o$, i.e., an utility is opened. Starting from t_o, $N_B - r$ new flow and pressure measurements are stored in their respective buffers $\mathcal{B}_f$ and $\mathcal{B}_p$ (both buffers have a length N_B values). So, the data buffers contain r measurements prior to time t_o and $N_B - r$ measurements subsequent to t_o. In this way, the transient data during the opening of an utility are stored. The opening transient contains almost all the information about the utility plumbing dynamics. The auxiliary parameter r is set in the design phase of the LD system.

The main idea of the proposed LD system during the *in use* phase of the plumbing network is to represent the dynamic relation between the pressure and flow measurements through a RLS identification algorithm with forgetting factor via (1)-(4). Variations in the model parameters are considered as symptoms of faults. Consider a simple ARX(1, 1, 1) model derived from (1):

$$y(t) = a_1 y(t-1) + b_1 u(t-1) + e(t), \tag{10a}$$

$$\boldsymbol{\varphi}(t) = [y(t-1)\ u(t-1)]^\top, \tag{10b}$$

$$\boldsymbol{\theta} = [a_1\ b_1]^\top = [\theta_1\ \theta_2]^\top, \tag{10c}$$

where, in this context, the y values are the flow measurements and the u values are the pressure ones. No further assumptions on $e(t)$ are made. The data pre-processing phase consists of two steps:

(1) *RLS estimations.* Estimation of the parameters $\boldsymbol{\theta}$ of (10) using the RLS formulation (2), (3), (4). The recursive estimation starts with the initialisations $\hat{\boldsymbol{\theta}}(0) = [0\ 0]^\top$ and $P(0) = 0$. The parameters estimation is updated for each datum stored in the buffers $\mathcal{B}_f$ and $\mathcal{B}_p$. All the N_B estimations of θ_1 and θ_2 are stored in the N_B-dimensional buffers $\mathcal{B}_{\hat{\theta}_1}$ and $\mathcal{B}_{\hat{\theta}_2}$, respectively.
(2) *Feature extraction.* Two features are extracted from the buffers $\mathcal{B}_{\hat{\theta}_1}$ and $\mathcal{B}_{\hat{\theta}_2}$ as:

$$x_1 = mean(\mathcal{B}_{\hat{\theta}_1}), \tag{11a}$$

$$x_2 = mean(\mathcal{B}_{\hat{\theta}_2}) \tag{11b}$$

with $mean(\cdot)$ denoting the averaging operation, and $x_1, x_2 \in \mathbb{R}$ are the extracted features. Note that (11) can be also calculated recursively, whenever a new parameters estimate is available.

The two proposed algorithms classify the features vector $\boldsymbol{z} = [x_1,\ x_2]^\top \in \mathbb{R}^2$ whenever an utility is opened. The considered classes are $\mathcal{C}_{healthy}$ and $\mathcal{C}_{faulty}$. If the plumbing network is healthy, $\boldsymbol{z} \in \mathcal{C}_{healthy}$. Otherwise, $\boldsymbol{z} \in \mathcal{C}_{faulty}$. To train and validate the algorithms, it is necessary to perform tests on the plumbing network to create a *training dataset* and a *testing dataset*. A test consists in the opening of an utility. This action triggers the data pre-processing phase and leads to the features vector $\boldsymbol{z}$. So, a test j is characterised by its own features vector $\boldsymbol{z}_j$. If the plumbing network is healthy, $\boldsymbol{z}_j \in \mathcal{C}_{healthy}$ and j is a *healthy* test. Otherwise, $\boldsymbol{z}_j \in \mathcal{C}_{faulty}$ and j is a *faulty* test. Note that the training and validation phases can be performed upon installation of the smart valve. In this way, the algorithms are calibrated on the plumbing network of interest.

Fault Detection algorithm The fault detection algorithm needs the training of the logistic regression model (8) to get the parameters estimates $\hat{\boldsymbol{w}}$. For the logistic regression model to detect faulty conditions, the *training dataset* must contain tests on the faulty plumbing network, as well as tests on the healthy one. Similarly, a *testing dataset* has to be created to evaluate the logistic regression model accuracy, defined as

$$\mathrm{ACC}_{\%} = \frac{\text{num. of correctly classified tests}}{\text{num. of tests}} \cdot 100. \qquad (12)$$

Algorithm 2 summarizes the classification logic of the proposed fault detection algorithm.

Algorithm 2: FD with logistic regression model

Input: $\boldsymbol{z} = [x_1,\ x_2]^\top,\ \hat{\boldsymbol{w}}$
Output: Classification of $\boldsymbol{z}$
$\boldsymbol{z} = [1,\ z^\top]^\top$
$p(\mathcal{C}_{faulty}|\boldsymbol{z}) = \sigma(\boldsymbol{z}^\top \hat{\boldsymbol{w}})$
if $p(\mathcal{C}_{faulty}|\boldsymbol{z}) \geq 0.5$ **then**
  $\boldsymbol{z} \in \mathcal{C}_{faulty}$
else
  $\boldsymbol{z} \in \mathcal{C}_{healthy}$
end

Anomaly Detection algorithm The anomaly detection algorithm consists in estimating the parameters $\boldsymbol{\mu}$, Σ of a multivariate Gaussian distribution (9) on the features $\boldsymbol{z} = [x_1,\ x_2]^\top \in \mathbb{R}^2$, so that $n = 2$. The estimates $\hat{\boldsymbol{\mu}}$, $\hat{\Sigma}$ are obtained relying *only on healthy data*, so that the estimated distribution characterizes the variation of the nominal, healthy network. This is a great advantage over the FD algorithm, as to perform fault injection on a plumbing network to get faulty data can be difficult and expensive. In the following, the accuracy of the AD algorithm is defined as in (12), where the testing dataset as for the FD algorithm is considered.

Let ε be a probability threshold close to 0. If $p(\boldsymbol{z}; \hat{\boldsymbol{\mu}}, \hat{\Sigma}) < \varepsilon$ it means that $\boldsymbol{z}$ has a close to 0 probability to belong to the multivariate gaussian distribution with mean $\hat{\boldsymbol{\mu}}$ and coveriance matrix $\hat{\Sigma}$. So, it is reasonable to consider $\boldsymbol{z}$ external to that distribution. But since the distribution describes the healthy network, the external $\boldsymbol{z}$ represents a faulty network, i.e., $\boldsymbol{z} \in \mathcal{C}_{faulty}$. So, if $p(\boldsymbol{z}; \hat{\boldsymbol{\mu}}, \hat{\Sigma}) \geq \varepsilon$, $\boldsymbol{z} \in \mathcal{C}_{healthy}$. Algorithm 3 summarizes the classification logic of the proposed anomaly detection algorithm. Figure 1 shows how the LD system is implemented in the embedded computational unit of the smart valve. When a new flow measurement is available and stored in $\mathcal{B}_f(N_B)$, it is compared with the threshold T_o to define the system condition. Then, the appropriate algorithm gives the *healthy*/*faulty* classification. The pressure measurements in $\mathcal{B}_p$ are used only in the data pre-processing phase when $\mathcal{B}_f(N_B) \geq T_o$.

Algorithm 3: AD with Gaussian distribution

Input: $\boldsymbol{z},\ \hat{\boldsymbol{\mu}},\ \hat{\Sigma},\ \varepsilon$
Output: Classification of $\boldsymbol{z}$
$p(\boldsymbol{z}; \hat{\boldsymbol{\mu}}, \hat{\Sigma}) = \dfrac{1}{(2\pi)|\hat{\Sigma}|^{1/2}} e^{-\frac{1}{2}(\boldsymbol{z}-\hat{\boldsymbol{\mu}})^\top \hat{\Sigma}(\boldsymbol{z}-\hat{\boldsymbol{\mu}})}$
if $p(\boldsymbol{z}; \hat{\boldsymbol{\mu}}, \hat{\Sigma}) < \varepsilon$ **then**
  $\boldsymbol{z} \in \mathcal{C}_{faulty}$
else
  $\boldsymbol{z} \in \mathcal{C}_{healthy}$
end

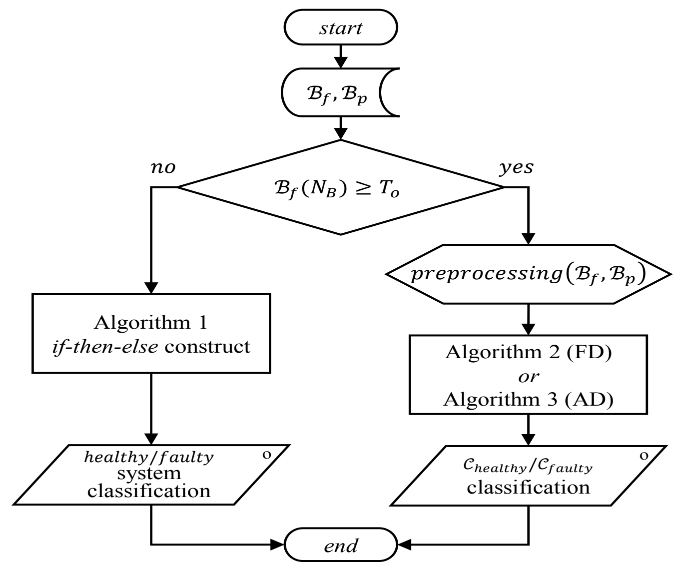

Fig. 1. Leak Detection system implementation. The flow chart uses the standard ISO 5807:1985 notation. The "o" in the parallelograms stays for "output".

4. CASE STUDY: BATHROOM PLUMBING

Experimental data are collected from a laboratory plumbing network setup to demonstrate the effectiveness of the proposed LD system. The test setup consists of $N_U = 3$ utilities: two different sinks, `sink-A`, `sink-B` and a `shower`. The utilities are connected each other with a 9 m long pipeline. The test setup has a pipeline section that can be replaced. This allows to perform fault injection using pipe sections with leak holes of different diameters. The setup is connected to the water supply network with a smart valve provided with a flow and a pressure sensor. Data are acquired from the sensors by using a PLC, connected to a computer that replaces the embedded computational unit. The LD system algorithms are implemented on the computer. Figure 2 shows the structure of the test setup.

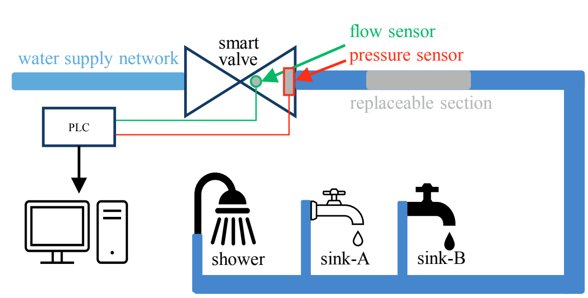

Fig. 2. Test setup. The blue color is the 9 m long pipeline. The light blue pipe is the water supply network. The grey color is the replaceable section of the pipeline.

Due to the simplicity of Algorithm 1, we omit to show its test results. Instead, we show the training results, the test results and the accuracies of Algorithm 2 and Algorithm 3. All the utilities are considered. We fix $\mathcal{T}_s = 10\,\text{ms}$, $T_o = 1.2\ \ell/\text{min}$, $N_B = 3000\ [-]$, $r = 500\ [-]$, $\varepsilon = 0.1\ [-]$ using the training dataset. The forgetting factor for the RLS is $\mu = 0.99$. Two different training datasets are created, one for each algorithm. The test dataset is the same for both the FD and AD algorithms, and aims to simulate a real utilities usage, considering different opening percentages $\mathcal{O}_{\%,i}$, $i = 1, \dots, N_U$ of the $N_U = 3$ utilities and testing both the healthy plumbing network and the faulty one. We consider two different faulty network conditions: (i) with an injected leakage flow of $1.2\ \ell/\text{min}$ and (ii) with an injected leakage flow of $0.6\ \ell/\text{min}$. The test dataset consists of 108 tests, 36 tests on the healthy network (*healthy* tests) and 72 tests on the faulty one (*faulty* tests). In the following figures, each marker identifies the feature vector $\boldsymbol{z} = [x_1\ x_2]^\top$ of a single test. Each test considers only one of the three utilities opened at a time.

4.1 Fault detection algorithm results

The logistic regression model of the fault detection algorithm is trained with a dataset that contains 27 tests. For each of the three utilities, we perform:

- 3 tests on the healthy network;
- 3 tests with a leakage flow of $1.2\ \ell/\text{min}$;
- 3 tests with a leakage flow of $0.6\ \ell/\text{min}$.

Thus, the training dataset is composed by a total of 9 healthy tests and 18 faulty ones (subdivided equally in the $N_U = 3$ utilities). The estimation of the logistic regression can be performed offline.

All the tests consider a complete opening of the utility, i.e. $\mathcal{O}_{\%,i} = 100\%$. Each test has a duration of $\approx 30\,\text{s}$. Figure 3 shows the estimated decision boundary which correctly divides the healthy tests (green markers) from the faulty ones (magenta markers). The markers color is the actual class while their position in the plan $x_1 = mean(\mathcal{B}_{\hat{\theta}_1})$, $x_2 = mean(\mathcal{B}_{\hat{\theta}_2})$ are the FD algorithm classification. The decision boundary divides the plan x_1, x_2 in two regions: a *healthy* region and a *faulty* region. If a marker is below the decision boundary (*healthy* region), the algorithm classifies it as a healthy test, i.e., $\boldsymbol{z} \in \mathcal{C}_{healthy}$. Otherwise, $\boldsymbol{z} \in \mathcal{C}_{faulty}$.

Table 1. FD algorithm confusion matrix

		actual class	
		$\mathcal{C}_{faulty}$	$\mathcal{C}_{healthy}$
predicted class	$\mathcal{C}_{faulty}$	67	0
	$\mathcal{C}_{healthy}$	5	36

Figure 4 and Table 1 show the results of the FD algorithm. The algorithm accuracy is $\text{ACC}_\% = 95.37\%$. Only 5 tests are misclassified: 4 about `sink-A` with a leakage flow of $0.6\ \ell/\text{min}$ and 1 about `shower` with a leakage flow of $1.2\ \ell/\text{min}$. These are *false-negative* misclassifications that are the most dangerous mistakes because the plumbing network is considered healthy but it is not.

4.2 Anomaly detection algorithm results

In the Anomaly Detection algorithm, we fit a multivariate Gaussian distribution on a training dataset that contains

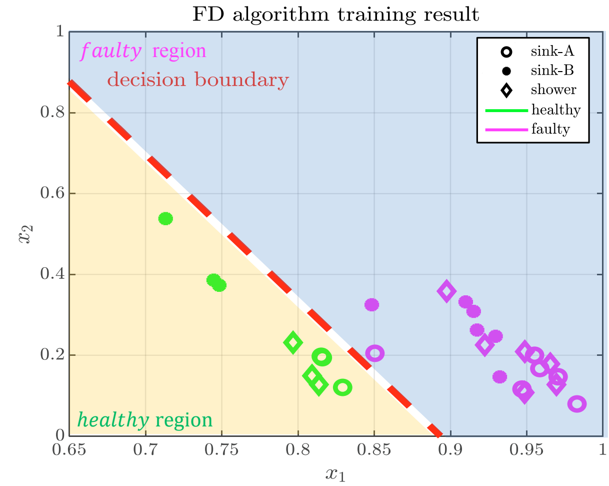

Fig. 3. Logistic Regression model training result; (green markers): healthy tests; (magenta markers): faulty tests.

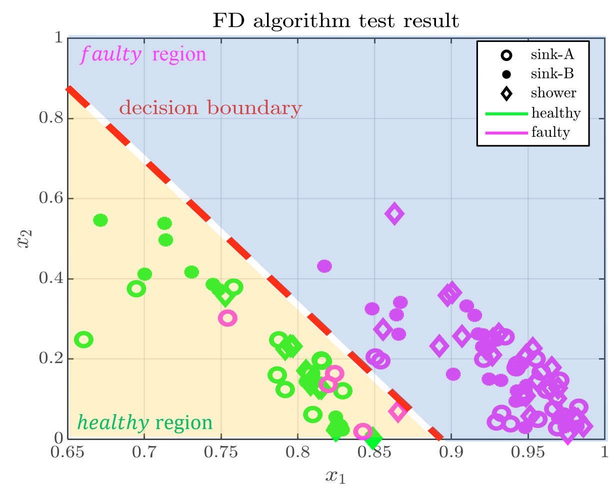

Fig. 4. Logistic Regression model test result. (green/magenta markers): actual class. The marker position on the plan represents the FD algorithm classification.

only the 9 healthy training tests, that is for each of the three utilities we perform 3 tests on the healthy network. These tests are *in addition* with respect to the healthy tests used for the FD algorithm, to provide and independent evaluation of the AD algorithm. All the tests consider a complete opening of the utility, i.e. $\mathcal{O}_{\%,i} = 100\%$. Each test has a duration of $\approx 30\,\text{s}$. The estimation of the Gaussian distribution can be performed offline.

Figure 5 shows the estimated Gaussian distribution on nominal/healthy network data. Figure 6 shows the test results. The markers color is the actual class. The colored circles around each marker are the classification results obtained with the AD algorithm. Figure 7 shows the zoom of the two boxed portions of Figure 6.

Table 2. AD algorithm confusion matrix

		actual class	
		$\mathcal{C}_{faulty}$	$\mathcal{C}_{healthy}$
predicted class	$\mathcal{C}_{faulty}$	54	1
	$\mathcal{C}_{healthy}$	18	35

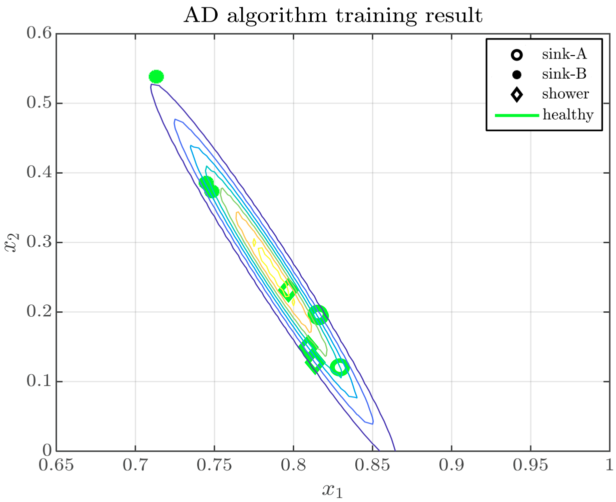

Fig. 5. Multivariate Gaussian Distribution training result. All the tests consider the healthy plumbing network.

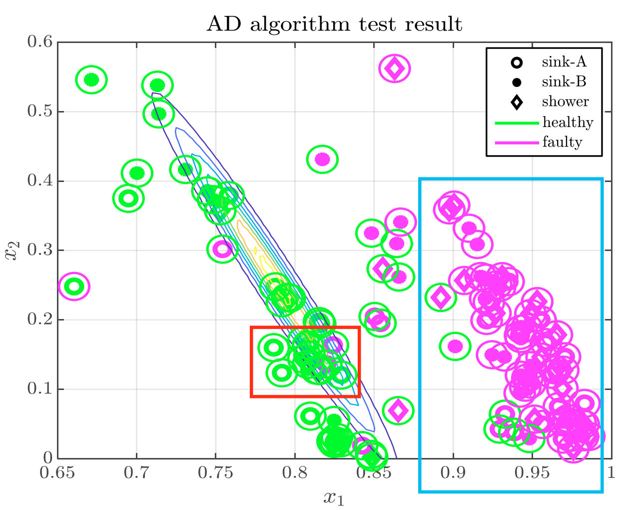

Fig. 6. Multivariate Gaussian distribution test result; (green/magenta markers): actual class; (green/magenta circles): AD algorithm classification. The red and cyan boxes are the zoomed parts in Figure 7.

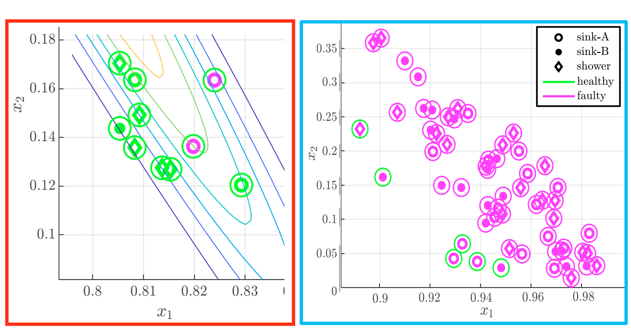

Fig. 7. Zoomed parts of Figure 6; (green/magenta markers): actual class; (green/magenta circles): AD algorithm classification.

The AD algorithm accuracy is $ACC_{\%} = 82.40\%$. Table 2 is the algorithm confusion matrix. There are 19 misclassified tests whose 18 are false-negative misclassifications and 1 is a false positive misclassification. Note that the traininig set contains only 9 tests, that however can be performed in a limited amount of time upon the smart valve installation.

5. CONCLUSIONS

The proposed LD system monitors an household plumbing network both when it is in use and when it is not, using a smart valve provided with with a single flow and pressure sensor. The proposed LD system has been shown to be effective and simple to implement. Further research is devoted to: (i) evaluate the approach considering different plumbing networks; (ii) provide the LD system with isolation capabilities.

REFERENCES

Billmann, L. and Isermann, R. (1987). Leak detection methods for pipelines. *Automatica*, 23(3), 381–385.

Bishop, C. (2016). *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer New York.

Da Silva, H.V., Morooka, C.K., Guilherme, I.R., da Fonseca, T.C., and Mendes, J.R. (2005). Leak detection in petroleum pipelines using a fuzzy system. *Journal of Petroleum Science and Engineering*, 49(3-4), 223–238.

Daadoo, M., Eleyan, A., and Eleyan, D. (2017). Optimization water leakage detection using wireless sensor networks (owld). In *Proceedings of the International Conference on Future Networks and Distributed Systems*, 1–11.

El-Zahab, S. and Zayed, T. (2019). Leak detection in water distribution networks: an introductory overview. *Smart Water*, 4(1), 1–23.

Ljung, L. (1999). *System Identification: Theory for the User*. Prentice Hall information and system sciences series. Prentice Hall PTR.

Mazzoleni, M., Scandella, M., and Previdi, F. (2022a). Evaluation of robust sensors placement schemes for leaks isolation in water distribution networks. *IFAC-PapersOnLine*, 55(33), 48–53. doi:https://doi.org/10.1016/j.ifacol.2022.11.008. 2nd IFAC Workshop on Control Methods for Water Resource Systems CMWRS 2022.

Mazzoleni, M., Sarda, K., Acernese, A., Russo, L., Manfredi, L., Glielmo, L., and Del Vecchio, C. (2022b). A fuzzy logic-based approach for fault diagnosis and condition monitoring of industry 4.0 manufacturing processes. *Engineering Applications of Artificial Intelligence*, 115, 105317. doi:https://doi.org/10.1016/j.engappai.2022.105317.

Rosli, N., Aziz, I.A., and Jaafar, N.S.M. (2018). Home underground pipeline leakage alert system based on water pressure. In *2018 IEEE Conference on Wireless Sensors (ICWiSe)*, 12–16. IEEE.

Sakti, A.N., Safirah, D., Soelistyono, A.P.H., Hadi, Y.W., and Muhaimin, H. (2021). Leak detection system design for household water supply pipe using acoustic signal analysis on frequency domain. In *2021 International Symposium on Electronics and Smart Devices (ISESD)*, 1–6. IEEE.

Valceschini, N., Mazzoleni, M., Pitturelli, L., and Previdi, F. (2022). Experimental fault detection of input gripping pliers in bottling plants. *IFAC-PapersOnLine*, 55(6), 778–783. doi:https://doi.org/10.1016/j.ifacol.2022.07.221. 11th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes SAFEPROCESS 2022.

Verde, C. and Torres, L. (2004). *Modeling and Monitoring of Pipelines and Networks*. Applied Condition Monitoring. Springer Cham.