AN INVESTIGATION OF DATA-BASED FAULT DETECTION METHODS IN PETROLEUM REFINERIES

A Thesis Submitted to
the Graduate School of
İzmir Institute of Technology
In Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE

in Chemical Engineering

by
Aslı YASMAL

May 2022
İZMİR

ACKNOWLEDGEMENTS

First of all, I would like to thank my thesis advisor, Assist Prof Dr Erdal UZUNLAR, for sharing his valuable experiences, guidance and support throughout the thesis.

I would like to thank my supervisor Gizem KUŞOĞLU in TUPRAS, who guided me throughout the project and tried to solve all my problems, and Dr. Ceylan ÇÖLMEKÇİ, who guided me in my research and always exchanged ideas. I also would like to thank my TUPRAS R&D team for their support and technical information during the project, and the operation engineers for their technical information and guidance.

Finally, I would like to express my endless thanks to my mother Candan YASMAL, my father Haydar YASMAL and my brother Anıl YASMAL for their love, encouragement and support.

ABSTRACT

AN INVESTIGATION ON DATA-BASED FAULT DETECTION METHODS IN PETROLEUM REFINERIES

The petroleum refineries are complex systems vital for energy and production sectors. During production, these complex systems might experience various faults, including fluid leaks in unit operations. The detection of leaks is important for a reliable, safe, and efficient operation. Among the possible leak detection mechanisms, data-based leak detection methods are promising in terms of low investment cost, less human intervention, ability to detect small leaks in advance and direct integration capability to distributed control systems. The aim of this study is to investigate data-based leak detection methods in a heat exchanger in a petroleum refinery. To that end, possible leaking problems in petroleum refineries are assessed, multiple leak cases from a real heat exchanger in a petroleum refinery are determined, literature studies are searched for appropriate data-based leak detection methods, applicability of a set of data-based leak detection methods is studied with a literature benchmark data set, and the real cases of heat exchanger leaks are studied with the determined leak detection methods. Data sets for multiple leak cases of a heat exchanger are obtained from a TUPRAS refinery. The benchmark data set is obtained from Tennessee Eastman Process (TEP). Discrete Wavelet Transform (DWT), Auto Encoder (AE), and Exponentially Weighted Moving Average (EWMA) are selected as the data-based leak detection methods. The selected data-based methods are first studied with TEP data set, and good fault detection capability is observed. Then, the real leak cases are studied. All three data-based methods are found successful in detecting the actual leak cases. For some of the cases, leaks are detected with data-based methods in advance of the operation engineers noticing the leak.

ÖZET

PETROL RAFİNERİLERİNDE VERİ TABANLI HATA TESPİT METOTLARI ÜZERİNE BİR İNCELEME

Petrol rafinerileri, enerji ve üretim sektörleri için hayati önem taşıyan karmaşık sistemlerdir. Bu karmaşık sistemler üretim sırasında, temel operasyonlardaki sıvı sızıntıları dahil olmak üzere, çeşitli arızalarla karşılaşabilir. Sızıntıların tespiti güvenilir, emniyetli ve verimli bir çalışma için önemlidir. Olası kaçak tespit mekanizmalarından veri tabanlı kaçak tespit yöntemleri, düşük yatırım maliyeti, daha az insan müdahalesi, küçük kaçakları önceden tespit edebilme ve dağıtık kontrol sistemlerine doğrudan entegrasyon kabiliyeti açısından umut vericidir. Bu çalışmanın amacı, bir petrol rafinerisindeki bir ısı eşanjöründe veriye dayalı kaçak tespit yöntemlerini araştırmaktır. Bu amaçla, petrol rafinerilerinde olası sızıntı problemleri değerlendirilmiş, bir petrol rafinerisindeki gerçek bir ısı eşanjöründen çoklu sızıntı durumları belirlenmiş, veriye dayalı uygun sızıntı tespit yöntemleri için literatür araştırması yapılmış, bir dizi veriye dayalı sızıntı tespit yönteminin uygulanabilirliği literatür referans veri seti ile çalışılmış ve belirlenen kaçak tespit yöntemleri ile gerçek ısı eşanjörü kaçak durumları incelenmiştir. Isı eşanjörünün çoklu kaçak durumları için veri setleri bir TÜPRAŞ rafinerisinden alınmıştır. Literatür referans veri seti Tennessee Eastman Process'ten (TEP) elde edilmiştir. Kesikli Dalgacık Dönüşümü (DWT), Otomatik Kodlayıcı (AE) ve Üstel Ağırlıklı Hareketli Ortalama (EWMA) yöntemleri, veri tabanlı kaçak tespit yöntemleri olarak seçilmiştir. Seçilen veriye dayalı yöntemler önce TEP veri seti ile çalışılmış ve iyi bir kaçak tespit kabiliyeti gözlemlenmiştir. Ardından gerçek sızıntı vakaları incelenmiştir. Her üç veriye dayalı yöntem de gerçek sızıntı vakalarını tespit etmede başarılı bulunmuştur. Bazı durumlarda sızıntılar, operasyon mühendisleri sızıntıyı fark etmeden önce veriye dayalı yöntemlerle tespit edilmiştir.

TABLE OF CONTENTS

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

Figure **Page**

Figure **Page**

CHAPTER 1

INTRODUCTION

1.1. Faults and Leaks in Petroleum Refineries

Petroleum refineries have an important role in energy market. In petroleum refineries (also known as oil refineries), crude petroleum is processed and turned into valuable products. Some of the processed petroleum products, such as diesel and gasoline, are among the major fuel sources of world. TUPRAS is one of the largest oil and gas refinery in Turkey. It continues its production in 4 different locations, as İzmir, Kırıkkale, Batman and İzmit. The petroleum refining is a complex process which require constant monitoring of process variables (Rosenfeld and Feng 2011).

In recent years, the oil and gas industry has been experiencing difficulties in meeting the demands of saving energy, trying not to harm the environment, and using resources more efficiently in production process (Clavijo et al. 2019). Since the refining process is complex, many process faults are encountered during the production. The most common faults in the petroleum refining industry are divided into two groups: equipment failure and human failure. Shutdowns caused by equipment failure cause great losses. Early detection of equipment failures and causes have great importance. Equipment such as pumps, compressors and rotating equipment can quickly fail due to wear, which can deteriorate product quality (Ohtani 2020).

It is important to detect faults for a reliable and safe production. Fault can be defined as non-admitted change of a feature of the system from appropriate and ordinary conditions (Miljković 2011). Commonly encountered faults in oil and gas industry include turbine trips, heat exchanger contamination, and leaks. Leaks are important because of the economic, health, environmental and structural problems they cause. Most leaks are found to be sourced in connectors, valves, compressors and heat exchangers by Environmental protection Agency (EPA) (Leak Detection and Repair, 2021).

Oil and gas industry consists of many complex units and pipelines. If leaks occur in visible points such as pipelines, they can be easily noticed. However, detecting leaks in invisible point such as equipment will take time and cause product losses. Therefore,

it is important to detect leaks in a timely manner. Leaks that have negative effects on the environment, human health and economy should be determined as quickly as possible. For this reason, interest in leak detection methods and applications is increasing day by day.

The records kept in the USA are the main sources to investigate the reason of leaks on all pipeline systems. The percentages of causes of leaks that have occurred in the last 20 years are shown in Figure 1.1. By looking at the chart, it can be said that excavation damages are major causes of leaks, followed by material failures and corrosion (Bolotina et al. 2017).

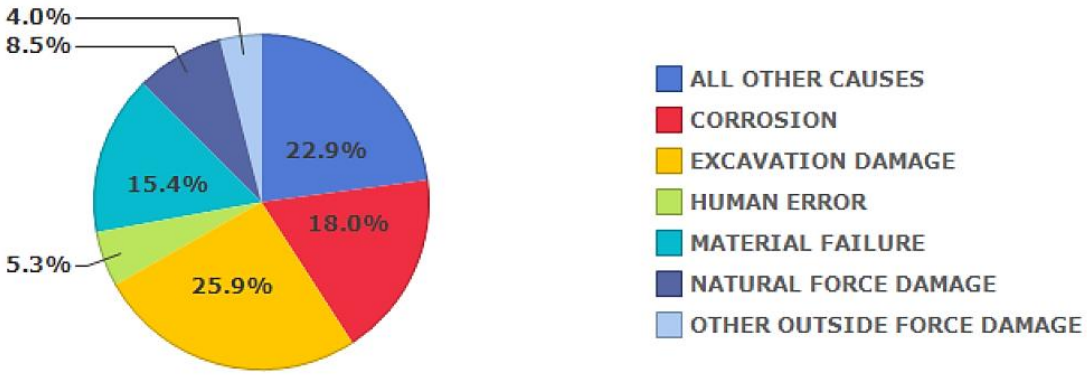

Figure 1.1. Causes of leaks
(Source: Bolotina et al., 2017)

It is also important to examine the damage caused by the leaks. The economic and health aspects of these leaks are quite painful. The best example of this is the results of the leak that occurred in the BP Alaska pipeline in March 2006. 4,800 barrels of oil were recorded as lost within 5 days. In addition, Prudhoe Bay was phased out and a fine of $20 million was also imposed (Penner et al., n.d.). To give another example, approximately 600,700 tons Volatile Organic Compound leakage (VOC) in a year is leaked from valves, compressors and connection point as reported by EPA (Leak Detection and Repair, 2021). Another event in the USA (in 1999) is undetected the chemical leaking into the groundwater for years. This situation lasted for a long time, as no leak monitoring method was implemented. It was noticed at a much later stage with the samples taken from the lake by the US Geological Survey. This situation greatly harmed the aquatic organisms

and contaminated the soil and the groundwater so leaks have to detect in a timely manner (Rosenfeld and Feng 2011).

The leaks can occur at visible points on pipelines or in places that can not be seen with the naked eye, such as underground pipelines. Considering a petroleum refinery, there are many pipeline systems through which products are transferred, and the leakage might occur due to mechanical and material related causes, such as pipe corrosion. Since the oil and gas industry is a high-risk industry, leaks should be constantly monitored by the operators and extra monitoring should be provided by detectors. According to investigations in the literature and at the TUPRAS refinery, leaks occurring at visible points such as pipes, flanges and valves are more common. They are detectable and easy to respond to the failure. Heat exchangers are in the first place in terms of leakages experienced in equipment. It has been noted in the literature that a leak occurs between the first 3-5 years of their life. It is very difficult to detect the leaks in these areas and their costs are quite high (Clover et al. 2010). Therefore, it is important to study methods to monitor leaks in petroleum refineries.

1.2. Heat Exchanger Types

Since heat exchangers are major points of leaks, a brief overview of the heat exchangers will be given in this section. Heat exchanger is the process equipment used for transferring heat between two fluids. It is one of the main process units in many industries such as oil and gas refinery, steam power station, plants of chemical processing, etc. (Zohuri 2016). There are several types of heat exchangers. These are compablock, shell and tube, plate, fluidized bed and storage type heat exchangers (Zohuri 2016; Shah 1983). In oil and gas industry, shell and tube heat exchangers are the most commonly used heat exchanger types because of their robust geometry and easy repair. These type of heat exchanger are classified into four groups within themselves as u-tube, fixed tube, floating head and kettle type heat exchangers (Kundnaney and Kushwaha 2015). A brief explanation of these types is given below.

Figure 1.2 shows a u-tube heat exchanger. This type of heat exchangers generally is not preferred in oil and gas industry because it is difficult to clean in case of any leakage.

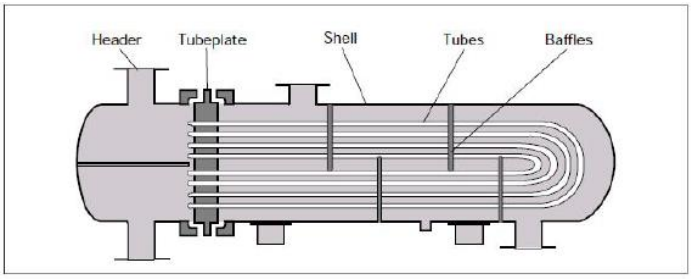

Figure 1.2. U-tube Heat Exchanger
(Source: Kundnaney & Kushwaha, 2015)

Figure 1.3 shows a fixed tube heat exchanger. This is the type of heat exchanger that is mostly preferred in refineries due to its ease of operation, low cost, and easy repair capability.

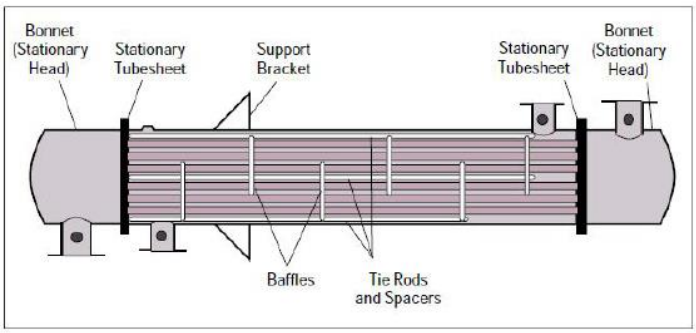

Figure 1.3. Fixed Tube Heat Exchanger
(Source: Kundnaney & Kushwaha, 2015)

Figure 1.4 shows a floating head heat exchanger. Since they have the floating head that improves the heat transfer between fluids, floating head is known as an efficient heat exchanger type for oil and gas industry. In contrast to u-tube heat exchangers, floating head heat exchangers are easier to clean.

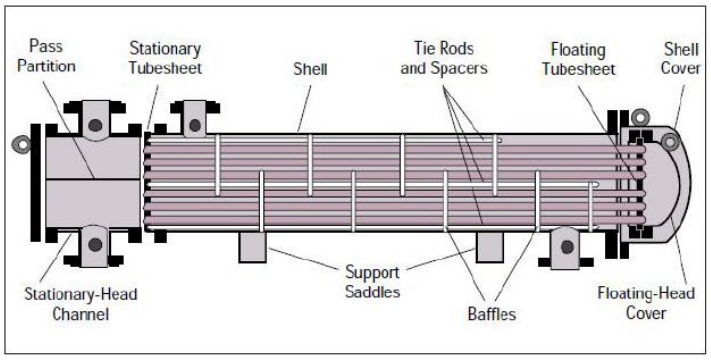

Figure 1.4. Floating Head heat Exchanger
(Source: Kundnaney & Kushwaha, 2015)

Figure 1.5 shows a kettle type heat exchanger. It is a special heat exchanger type used in cases where high pressure gases are present. Shell part is suitable to encounter expansion of gas in the system.

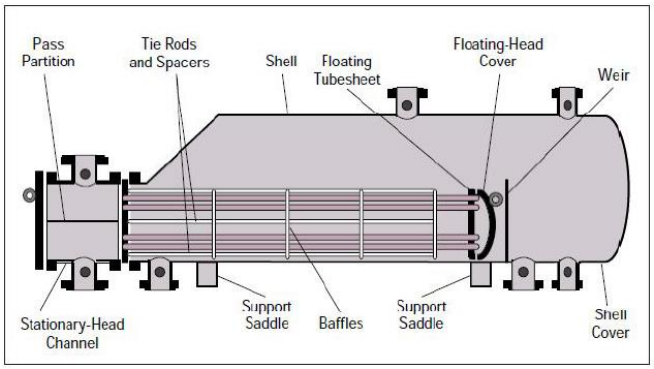

Figure 1.5. Kettle Type Heat Exchanger
(Source: Kundnaney & Kushwaha, 2015)

Now, heat exchangers, the equipment where most leaks in petroleum refineries occur, are introduced. We can continue with the leak detection methods.

1.3. Leak Detection Methods

The main purposes of leak detection are indication of leak and improvement of system reliability. These methods are mainly classified into three groups as hardware,

software and biological methods, as schematically shown in Figure 1.6. The requirement of external sensor installations is essential for detection of leaks in pipeline with hardware-based methods. Biological methods are based on the senses of humans or animals. Software-based methods are data-based methods. Software-based methods are used online, or they utilize historical data to detect leaks. While sensors and devices are important for the hardware based methods, data based and computational approaches are important for software based methods (Mujtaba et al. 2020).

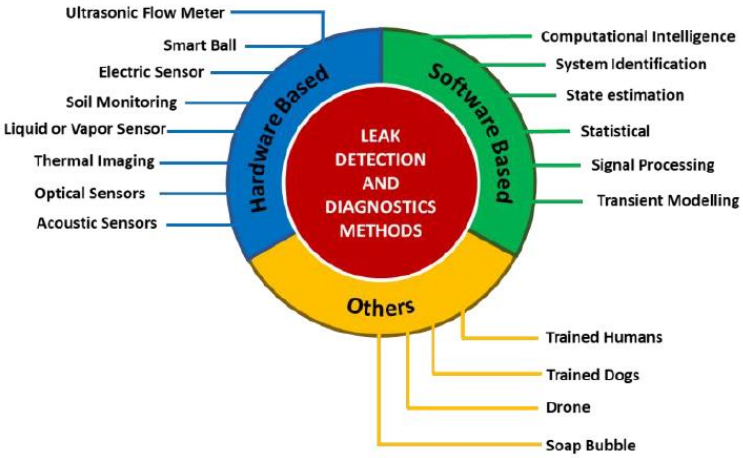

Figure 1.6. Classification of methods
(Source: Mujtaba et al., 2020)

If these methods are examined in more detail, in some sources, methods are divided into three groups as direct, indirect and external methods, as shown in Figure 1.7 (Zaman et al. 2020). While hardware-based methods are classified as a direct method, software methods are classified as an indirect method. Software-based methods are further classified as data-based and model-based methods. Data-based methods are separated based on data type, data source and technique. Based on the data type, there are supervised and unsupervised methods (Tutkan, Ganiz, and Akyokuş 2016). While supervised technique is based on the principle of training the machine with the labeled data, unsupervised technique uses unlabeled data. In addition, data based methods are classifed acording to the data source such as pressure, flow, demand and flow-pressure/demand pressure. Depending on the tehnique, data-based methods are separated into classification, prediction classification, statistical and signal processing groups.

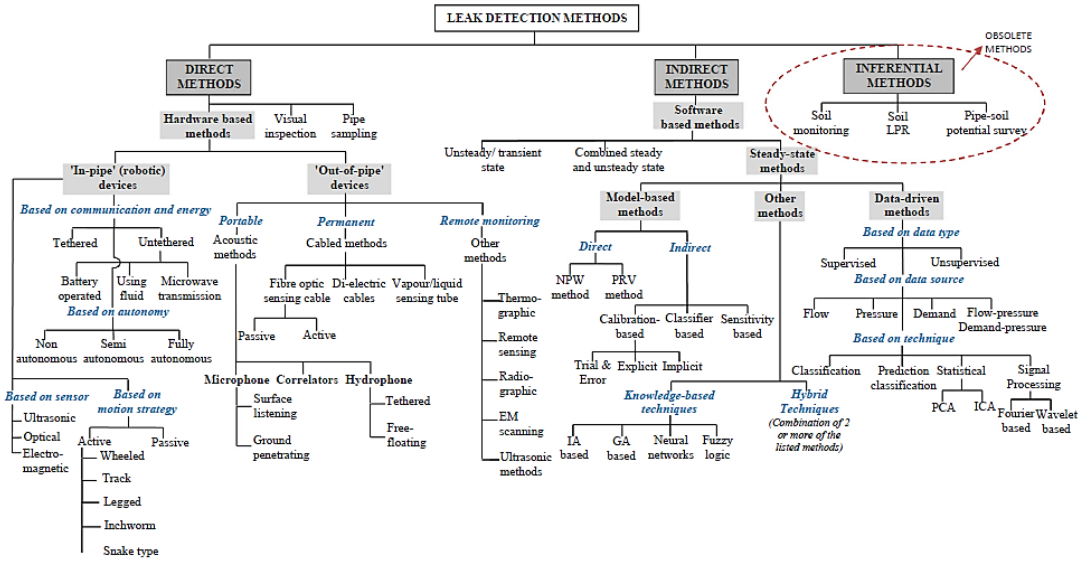

Figure 1.7. Leak Detection Methods
(Source: Zaman et al., 2020)

In the next sections, we will briefly overview the leak detection methods.

1.3.1. Hardware Based Leak Detection

Some additional and special sensors are used to detect leaks in the class of hardware-based leak detection techniques. These sensors consist of acoustic detectors, fiber optic sensors, ultrasonic technologies, infrared thermograph, and radiotracers.

1.3.1.1. Acoustic Leak Detection

This method uses acoustic detectors to detect leak and leak localization. A basic representation of the set-up is shown in Figure 1.8. Acoustic sensors are placed along the pipeline. These sensors include auscultation sticks, aqua phones, and ground microphones. These devices give the acoustic map of the system (Odusina 2008).

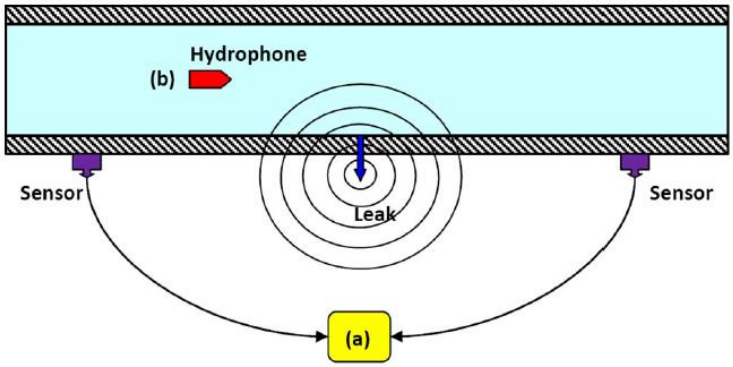

Figure 1.8. Acoustic leak detection set up (a) external (b) in-pipe measurement
(Source: Khulief et al., 2012)

1.3.1.2. Fiber Optic Sensing Leak Detection

This technique gives an information about leak based on the measurements of fiber optic probes. Figure 1.9 shows the basic set-up and representative result obtained upon a leak (Nikles et al. 2004). Temperature change gives an information about occurrence of a leak. Since these probes analyze the temperature change of leakage area, they have to be carefully placed to touch both pipe and soil.

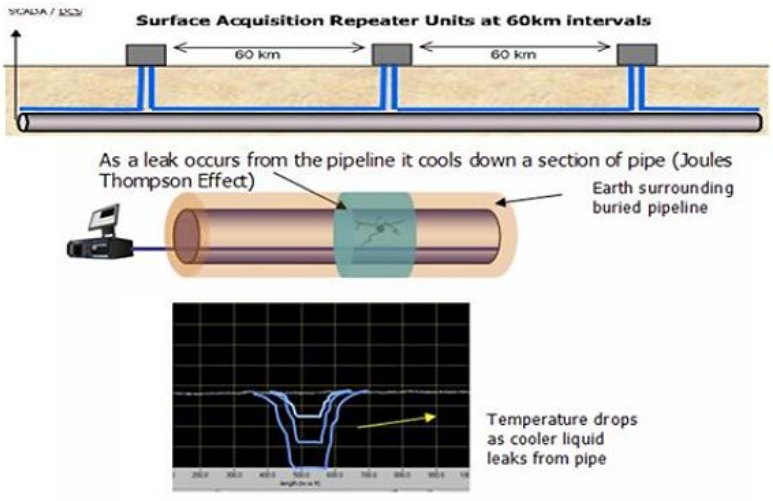

Figure 1.9. Fiber Optic Sensing Leak Detection
(Source: Nikles et al., 2004)

1.3.1.3. Infrared Thermography

Infrared Thermography (IT) method is useful for the pipeline leakage. Infrared cameras are important to notice sudden temperature change caused by leak. Experimental set up of IT is shown in Figure 1.10. Any anomaly can be detected with the color based on the warm and cool environment. This method is user friendly and has a quick response time (Adegboye, Fung, and Karnik 2019).

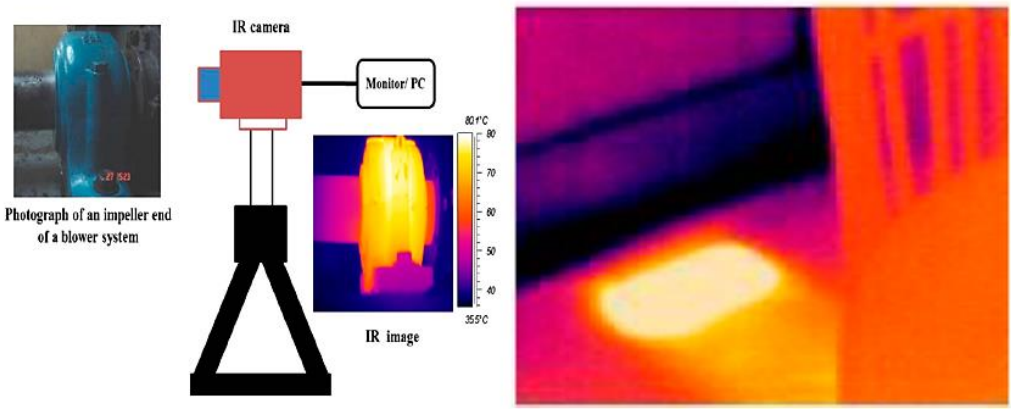

Figure 1.10. Set up of Infrared Thermography method
(Source: Adegboye et al., 2019)

1.3.1.4. Radiotracer

Radiotracer detector can be used for open and closed systems. Also, it is preferred to detect leaks both in underground pipeline network and shell and tube type heat exchangers. There are two types of detectors: injection detector known as inlet detector and leak detector known as output detector. These detectors are located at tube input and shell output, respectively, as shown in Figure 1.11 (Pipelines and Radiotracers 2009). Injection pulse is monitored using injection detector. If any leakage occurs in the heat exchanger, response peaks are observed from leak detectors,

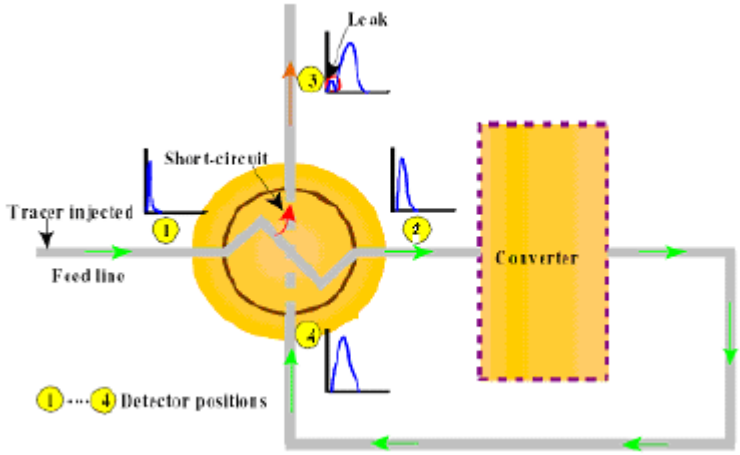

Figure 1.11. Location of Detectors
(Source: Pipelines&Radiotracers, 2009)

Table 1.1 below shows a summary of hardware-based leak detection methods. Each method has its strengths and weaknesses according to its working principles. In addition, these methods are costly because they require new hardware. This narrows their usage areas even though they have high accuracy.

Table 1.1. Summary of Hardware Based Methods
(Source: Adegboye et al., 2019 & Adedeji et al., 2017)

Methods	**Cost**	**Leak Localization**	**Principle**	**Strengths**	**Weakness**
Acoustic	High	Yes	Collect the signal from where leak occur	Easy to install and provides early detection	Affected by environmental conditions Insufficient for small leaks
Fiber Optic	High	Yes	Takes advantage of temperature changes caused by leak	Can act as sensor and transmission medium	High cost and low durability

(cont. on the next page)

Table 1.1 (cont.)

Methods	Cost	Leak Localization	Principle	Strengths	Weakness
Infrared Thermography	High	Yes	Uses infrared imaging techniques to detect temperature changes	High power to visualize images, easy to use	Not suitable for small leak detection
Radiotracer	High	No	Use the impulse and response peaks	Sensitive for leak, suitable for open and closed systems	High cost

1.3.2. Biological Based Leak Detection

Professional people and animals are required for visual or biological methods. On the pipeline, occurrence of any leaks can be detected by an experienced person. This detection can be supplied with visual and/or olfactive observation of leakage point. In addition, noise and vibration caused by leakage can give information about the presence of leak and leak location. Trained animals such as dogs and pigs also play an active role in leak detection. Strong sense of smell of dogs can sometimes give better results than human. Although this is the case, these trained dogs could be used for constant monitoring (Adegboye, Fung, and Karnik 2019).

1.3.3. Software Based Leak Detection

Software base leak detection methods are investigated into two groups as model-based and data-based methods. In software-based methods, operational parameters such as pressure, flow rate, temperature, density, volume are used to detect anomalies.

1.3.3.1. Model Based Techniques

1.3.3.1.1. Mass-Volume Balance

The principle of this method is mass conservation. This method is generally applied for pipeline leakage cases. Mass inflow and outflow values must be in balance in the absence of leaks. Any differences in flow values give an information about anomaly in the pipe. The mass balance is given in Equation 1 (Adegboye, Fung, and Karnik 2019).

$$\dot{M}_l(t) - \dot{M}_o(t) = \frac{dM_L}{dt} \qquad \text{Eqn 1.}$$

where t is time, and $\dot{M}_i(t)$ and $\dot{M}_o(t)$values show inlet and outlet mass flow rates, respectively, and M_L is the mass stored in the pipeline length L. Along the pipeline, stored mass amount changes and this changing can be represented with Equation 2.

$$\frac{dM_L}{dt} = \frac{d}{dt}\int_0^L \rho(x)A(x)dx = \int_0^L \frac{d}{dt} < \rho(x)A(x) > dx \qquad \text{Eqn 2.}$$

where A is cross-sectional area of the pipe and ρ is the density of the fluid. If ρ and A are assumed constant in Equation 2, $\frac{dM_L}{dt}$ will be zero. In that case, we obtain:

$$\dot{M}_i(t) - \dot{M}_o(t) = 0 \qquad \text{Eqn 3.}$$

Also, according to assumption that ρ is constant, we obtain as Equation 4

$$\dot{V}_i(t) - \dot{V}_o(t) = 0 \qquad \text{Eqn 4.}$$

where $\dot{V}_i(t)$ and $\dot{V}_o(t)$ are inlet and outlet volumetric flowrates, respectively.

Any imbalances in Equation 5 represented as $\dot{R}$, gives an information about existence of leak considering the threshold, $\dot{R}_{th}$.

$$\dot{R}(t) = \dot{V}_l(t) - \dot{V}_o(t) \qquad \text{Eqn 5.}$$

$$\dot{R} = \begin{cases} < \dot{R}_{th} & in\ absence\ of\ leak \\ > \dot{R}_{th} & if\ there\ is\ a\ leak \end{cases}$$

The main disadvantage of this method is inability to determine leak location. Additionally, this method is affected by random disturbance and pipeline dynamics.

1.3.3.1.2. Negative Pressure Wave

Basic representation of the setup for negative pressure wave method is shown in Figure 1.12 (Adegboye, Fung, and Karnik 2019). Occurrence of leak causes pressure drop and reduction of flow rate in pipe. These changes create a negative pressure wave (NPW) at the leakage point and this wave spread towards the ends of the pipe. Arrival time of the wave to a detector gives an information about location of leak (Sheltami, Bala, and Shakshuki 2016).

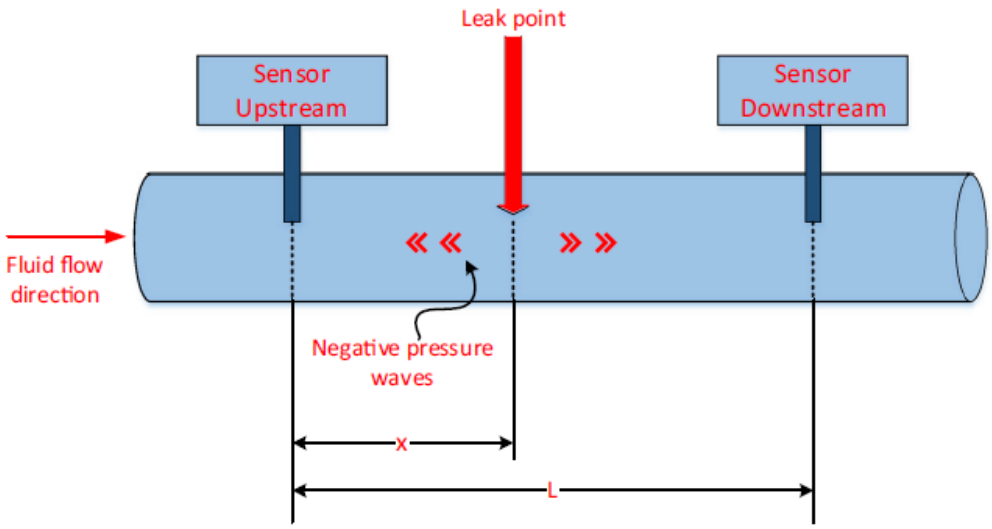

Figure 1.12. NPW representation
(Source: Sheltami et al., 2016)

The arrival time of the wave is calculated by Equations 6, Equations 7, Equations 8, and Equations 9. In these equations, t_o is the leakage time, t_1 is the wave arrival time to upstream sensor, t_2 is the wave arrival time to the downstream sensor, V is liquid

velocity, X is the distance between leak point and upstream sensor, a_x is propagation velocity of NPW, ρ is liquid density, K is liquid bulk modulus, E is elasticity modulus, C is correction factor due to constraints of pipeline, D is the diameter of pipeline, and e is the thickness of pipeline.

$$t_1 - t_o = \int_0^X \frac{1}{a_x - V} dX \quad \text{Eqn 6.}$$

$$t_2 - t_o = \int_X^L \frac{1}{a_x + V} dX \quad \text{Eqn 7.}$$

$$\Delta t = \frac{X}{a_x - V} - \frac{L - X}{a_x + V} \quad \text{Eqn 8.}$$

$$a_x = \sqrt{\frac{K/\rho}{1 + (K/E)(D/e)C}} \quad \text{Eqn 9.}$$

1.3.3.1.3. Pressure Point Analysis

The pressure point analysis (PPA) takes into consideration pressure values taken from certain points on the pipeline. If the measured value falls below the threshold values which are determined according to the average of the previous measurements (Adedeji et al. 2017) or the trend of the old measurements (Adegboye, Fung, and Karnik 2019), it indicates the leakage that has occurred on the pipeline.

This method is based on the principle of pressure drop that will occur in the event of a leak. It is one of the easy and inexpensive methods to apply. It is used to detect the leak, but it is not appropriate for detecting the leak location (Adegboye, Fung, and Karnik 2019).

1.3.3.2. Data Based Techniques

The main source of data-based methods is data acquired from sensors. These methods can be applied by using online or historical data. The complexity of the pipeline

or process does not affect the applicability of these methods. Generally, variables such as flow, pressure, temperature, vibration are used. Among the possible leak detection mechanisms, data-based leak detection methods are promising in terms of low investment cost, less human intervention, ability to detect small leaks in advance and direct integration capability to distributed control systems. As stated in the Introduction section, data-based methods are examined in three groups according to data type, data source and technique (Wu and Liu 2017), (Zaman et al. 2020). When methods are investigated according to the technique, the methods according to the classification are support vector machine, Bayesian network, rule-based and neural network methods. Statistical methods include principal component analysis (PCA), independent component analysis, exponentially weighted moving average (EWMA). Signal processing methods include fast Fourier transform (FFT) (signal based), discrete or continuous wavelet transform (DWT or CWT) (wavelet based) (Ahmed, Naser Mahmood, and Hu 2016). Autoencoder (AE) is a kind of artificial neural network (Mirsky et al. 2018). Considering the studies in the literature, DWT, AE, PCA and EWMA methods are used in this study and the detailed explanations of the methods are given in methods section (Perera, Rajapakse, and Jayasinghe 2007), (Chen et al. 2018), (Ye, Borror, and Zhang 2002).

1.4. Aim of Thesis

The aim of this thesis is to investigate data-based leak detection methods on a real unit operation in a petroleum refinery. The real leak cases are extracted from a heat exchanger in the TUPRAS Izmit Refinery. The data-based methods extracted from literature are first validated on a literature benchmark data set, and then applied on the real leak cases.

1.5. Thesis Organization

This thesis comprises of six chapters. In chapter 1, brief information is given about the leaks in refinery and leak detection methods. Also, types of heat exchangers are explained. In Chapter 2, studies using data-based anomaly detection methods in the literature and leak detection in heat exchangers are included. Detailed information about DWT, EWMA and AE is given in Chapter 3. In Chapter 4, the benchmark dataset and

chemical process are briefly explained, and the results of the data-based methods applied on benchmark dataset are provided. In Chapter 5, TUPRAS cases are explained, the methods are applied on each case and the obtained results are presented. In chapter 6, the obtained results are briefly summarized, and future work and recommendations are given.

CHAPTER 2

LITERATURE SURVEY

2.1. Literature Search Criteria

In the literature, there are many methods for detecting leaks or anomalies that may occur in heat exchangers. Most of these include additional sensors placed in the system or offline detection mechanisms. In order to focus on the aim of applying data-based methods to detect leaks in heat exchangers, the following criteria have been followed in the article selection from the literature:

- Leaks occur in heat exchanger.
- Studies are carried out online.
- Only real cases are studied which includes either operational industrial equipment or pilot scale equipment.
- Only data-based methods are studied.
- Studies in which only physics-based models (i.e., first principle-based models) are utilized in leak detection are excluded.

When the literature is searched according to the criteria, only a few studies are found. In the next section, the studies of Panday et al., Guillen et al., and Habibi et al. are summarized briefly (Panday et al. 2021), (Guillen et al. 2020), (Habbi, Kinnaert, and Zelmat 2009).

2.2. Literature Studies Satisfying the Criteria

Panday et al. carried out studies aimed to detect heat exchanger leaks in a with 300 MW coal-fired power plant. The purpose of their work was to reduce the number of unit shutdowns because of leaks in the tubes of heat exchanger. For this purpose, they carried out detection studies by applying data-based methods with the time-series data they had taken collected from the process (Panday et al. 2021).

The boiler part of the plant where data collected is shown in the Figure 2.1. Mass balance was made around the steam drum (shown in the form of a balloon in Figure 2.1) with collected the time-series data. The most important parameter for leakage was determined as the ratio of feed water mass flow rate to steam mass flow rate. Here, $\dot{m}_s$is steam mass flowrate, $\dot{m}_f$ is feed water flowrate, $\dot{m}_w$ is water mass flowrate, $\dot{m}_{bd}$ is blowdown rate, $\dot{m}_{sb}$ is sootblowing rate, $\dot{Q}_{in}$ is the heat obtained from the furnace and is used to heat the water in the boiler walls. $\dot{m}_w/\dot{m}_s$ ratio was calculated and threshold was determined based on this ratio.

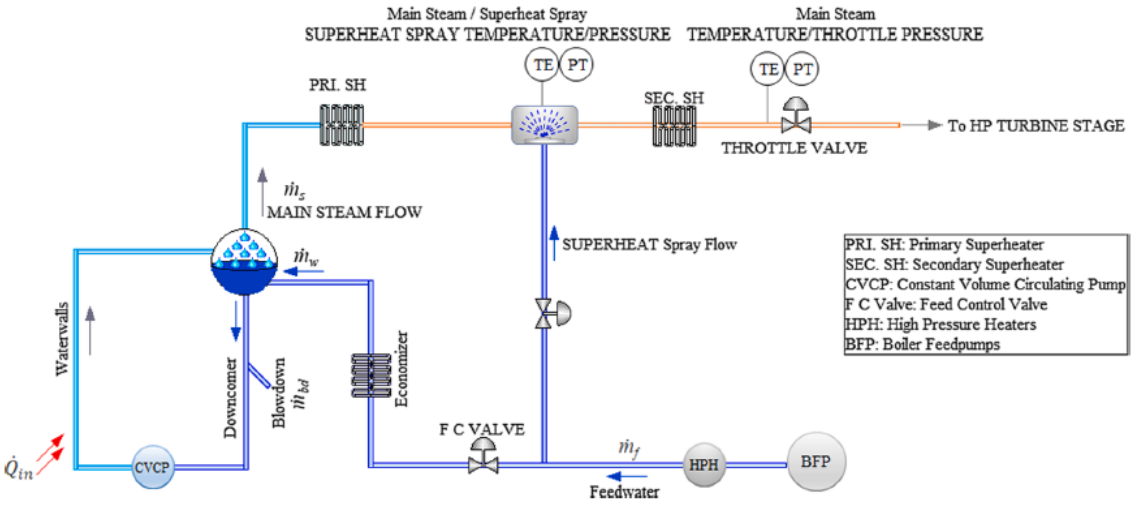

Figure 2.1. Basic representation of boiler part
(Source: Panday et al., 2021)

In this study, three different sets were examined as set A, set B, and set C. Each set had different load conditions (i.e., electrical energy output requirement conditions). The loads were calculated according to the general mass balance of the system. In Figure 2.2, set A shown with blue diamonds refers to the condition where load is changed between 50 and 99% of the full load, set B shown with green circles refers to the condition where load is changed between 48 and 100% of the full load conditions and set C shown with orange triangles to the condition where load is changed between 57 and 93% of the full load case. Here, set A and set B belong to fault free data while set C belongs to faulty data. The deviation of each set from the regression line is also shown in the Figure 2.2.

These sets deviate from the regression line at low load conditions since the power plant was designed for load conditions that require mass flowrates above 1.5 million lbm/hr. The slope of the graph is equal to the ratio $\dot{m}_w/\dot{m}_s$. When the slopes were compared for each set, the slopes of set A and set B were equal and a 1% larger slope was observed with set C. The authors interpreted this difference as the indication of leak.

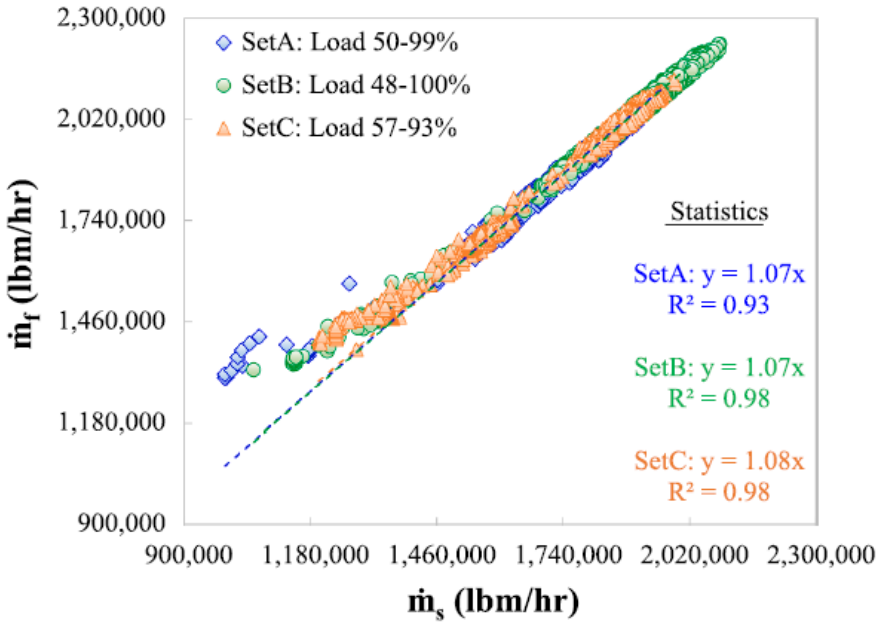

Figure 2.2. Comparison of $\dot{m}_s$ and $\dot{m}_f$
(Source: Panday et al., 2021)

Here, an optimal exponential moving average (EMA) method was proposed by Panday et al. The EMA equation is given in Equation 10 below:

$$\hat{x}(k+1) = \hat{x}(k) + \alpha\big(y(k) - \hat{x}(k)\big) \qquad \text{Eqn 10.}$$

where, $y(k)$ is the present measurement, $\hat{x}(k)$is the previously computed value, $\hat{x}(k+1)$ is the exponentially weighted mean between the $y(k)$ and $\hat{x}(k)$, and α is the smoothening constant (smaller α ignores recent data, and larger α ignores past data). In general, α is set to a value ranging between 0.05 and 0.20 in the literature. Panday et al. instead derived the optimal value of α by equating the derivative of mean squared error with respect to α to zero. This equation then becomes the Kalman filter, which is suitable

for minimizing the mean squared difference between the current value and the measured value.

In this study, Kalman filter (optimal EMA) and simple EMA filter were compared as shown in Figure 2.3. In the graph, the black line shows the Kalman based filter result, the purple line shows the EMA result, the gray line shows the actual measurement values, and the green dashed line represents the upper control limit (UCL) which was determined by using the nominal plant conditions. When the results are examined, the authors concluded that the optimal EMA (Kalman) filter responds quicker than simple EMA filter (8 hours in advance of the simple EMA filter).

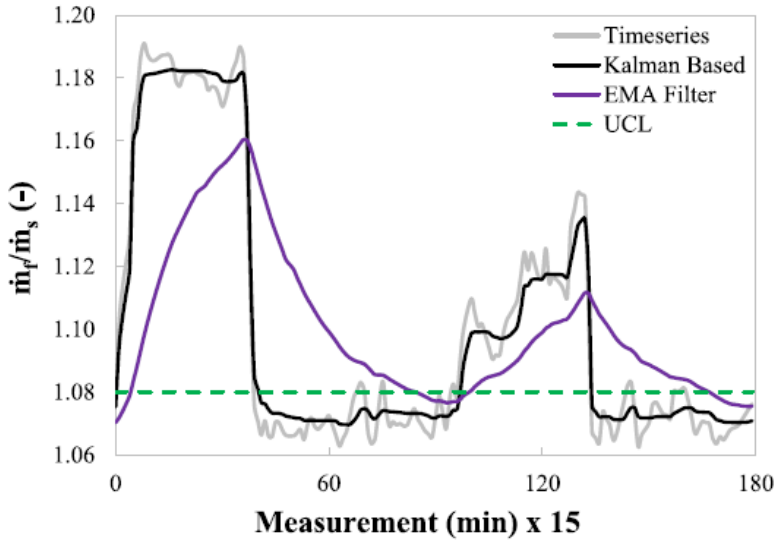

Figure 2.3. Result of proposed method and Kalman filter
(Source: Panday et al., 2021)

In another study, Guillen et al. created a model of fan coil units (FCU) in an operational nuclear power plant. The FCU unit consists of a heat exchanger and a fan, as shown in Figure 2.4. The nitrogen is cooled down by river water in the heat exchanger and fan moves the cooled nitrogen. The nitrogen is then used to cool down various parts of nuclear reactor. There were four FCUs in the nuclear power plant under investigation. FCUs are known to cause problems due to equipment failures. To detect failures, a thermal model of the FCU was created using Reactor Excursion and Leak Analysis Program (RELAP) to predict the normal operating temperatures of the fluids in FCU. For

RELAP to work, the inlet nitrogen temperatures collected by sensors were used. However, these sensors fail frequently. For that reason, the authors suggested using a long short-term memory (LSTM) method to predict inlet nitrogen temperatures using various sensor tags in the nuclear power plant. LSTM is an artificial neural network technique commonly used in machine learning applications. The anomaly in FCUs were determined by comparing the actual nitrogen outlet temperatures with those predicted by RELAP only (using measured inlet nitrogen temperature) and RELAP supported with LSTM (Guillen et al., 2020).

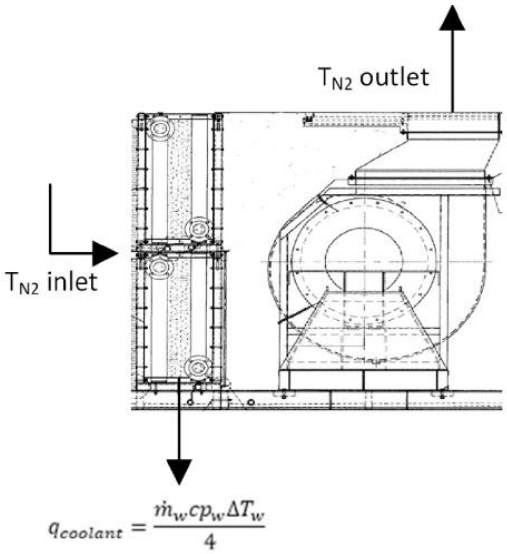

Figure 2.4. Representation of FCU (In energy balance, $\dot{m}_w$ is mass flowrate of coolant, c_p is coolant specific heat, $T_{w,inlet}$ is inlet temperature of coolant, $T_{w,outlet}$ is outlet temperature of coolant, and $q_{coolant}$ is the is the heat removed by the river water from nitrogen.) (Source: Guillen et al., 2020)

A novel contribution of the study was the implementation of the LSTM. The LSTM method was fed with a dataset consisting of 33 different variables. Each variable affects the FCU outlet temperatures that are being tried to predict. The data set was divided into three groups as training, validation, and test set. In order to decide the epoch number, the trend of the validation and training data sets and the loss function values according to the epoch number were examined. The results are shown in Figure 2.5. As the epoch number increases, the loss function decreases, and a good learning is achieved with the LSTM method. Authors determined the epoch number as 100.

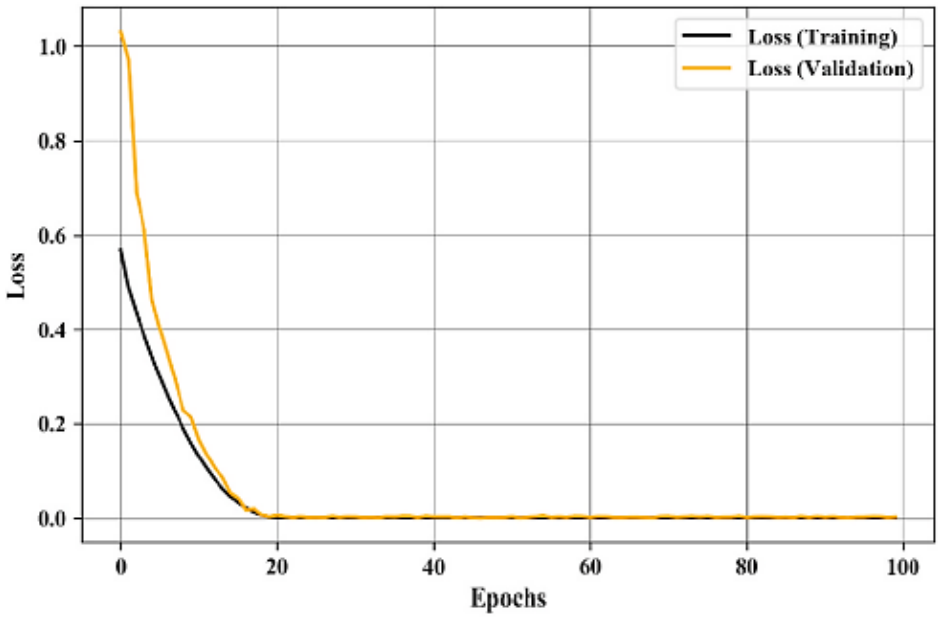

Figure 2.5. Loss function of LSTM
(Source: Guillen et al., 2020)

Comparison of measured and predicted outlet nitrogen temperatures is shown in Figure 2.6 for each FCUs. In the absence of any anomalies, the measured and predicted values are quite similar, as seen in FCU B and FCU D. However, on May 11, an anomaly occurred in FCU A and FCU C. As seen, both RELAP predictions whether using measured inlet nitrogen temperature and LSTM predicted inlet nitrogen temperature can identify the anomaly. Moreover, RELAP prediction with LSTM predicted inlet nitrogen temperature is more closely matching the measured outlet temperature, thus enabling a better representation of the actual operation and equipment failure detection.

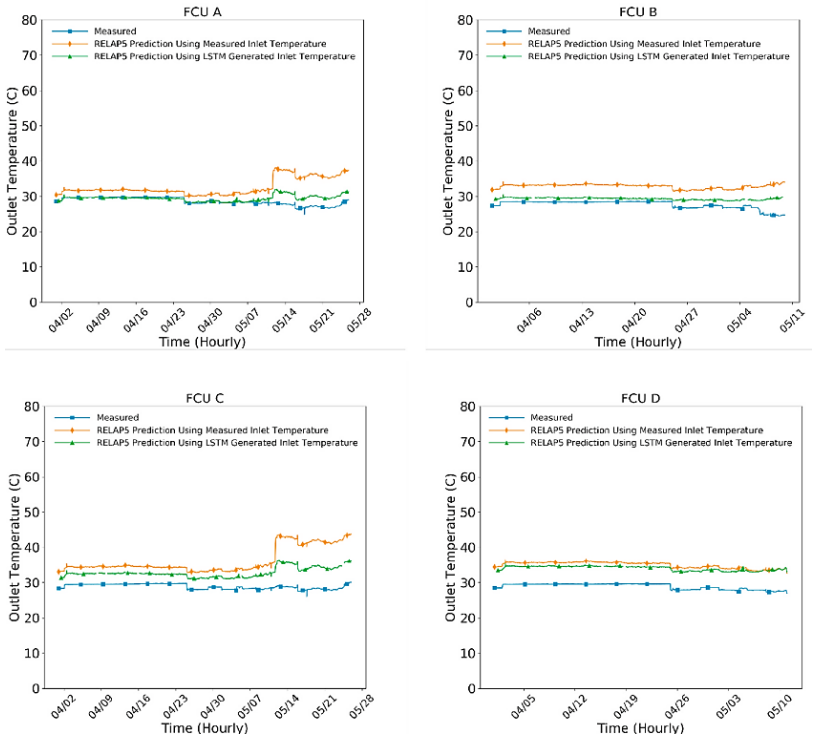

Figure 2.6. Comparison of measured and predicted outlet temperature values for each FCU (Source: Guillen et al., 2020

In another study, a fuzzy logic model-based leak detection method was developed for a pilot heat exchanger by Habbi et al. The authors discussed that first-principle based models are very complex, difficult to drive and accurate values of heat transfer coefficients are generally unknown. The pilot heat exchanger is a co-current gas-liquid heat exchanger where the water is heated using hot air that shown in Figure 2.7. The system includes electric heater E, air recycling valve V_r, and air evacuation valve V_e, and variable speed pump SP. The leaks were simulated using a bypass valve. A Takegi-Sugeno (TS) fuzzy model-based approach was used in leak detection algorithm. This model aims to develop IF-THEN rules for description of the system behavior. The relevant parameters are selected to be P, V_r, V_e, T16, and T34 (see Equation 11) for definitions of variables) based on the recommendation and process knowledge of pilot equipment operator. To develop the fuzzy model, Q and Q_a were held constant, P, V_r and V_e (assumed to be dependent on V_r) were changed in a wide range and the resulting T34 were collected without any leaks introduced in the heat exchanger (i.e., leak-free operation). A mean square error metric was considered to describe the discrepancies between model and actual measurements. The obtained TS fuzzy model for a rule i is

conceptually shown in Equation 11 where ai and bi are the rule-consequent parameters.

$$IF\ P(k)\ is\ A^i{}_1\ and\ V_r(k)\ is\ A^i{}_2\ and\ T_{16}(k)\ is\ A^i{}_3\ and\ T_{34}(k)\ is\ A^i{}_4$$
$$THEN\ T_{34}(k+1) = b^i + a^i{}_1 P(k) + a^i{}_2 V_r(k) + a^i{}_3 T_{16}(k) + a^i{}_4 T_{34}(k)$$

Eqn 11.

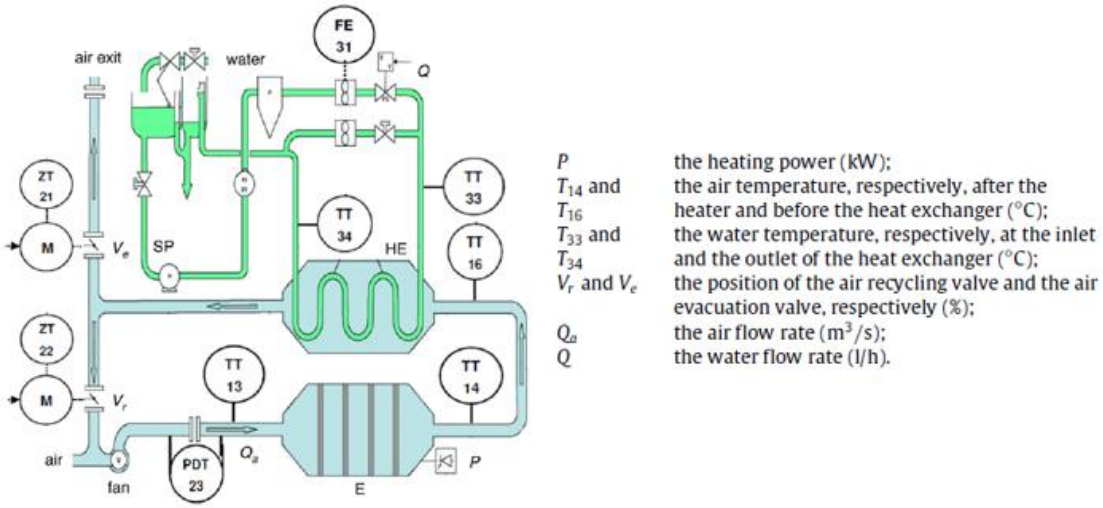

Figure 2.7. Pilot heat exchanger
(Source: Habbi et al., 2009)

The developed fuzzy model was found to perform well to describe the system behavior, as shown in Figure 2.8.

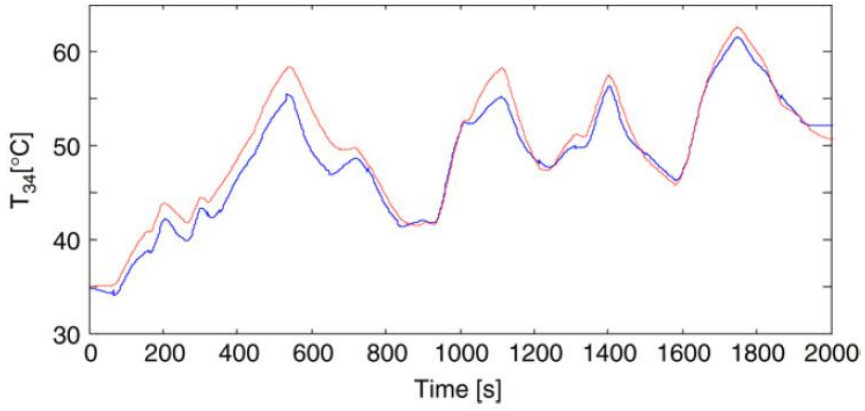

Figure 2.8.Fuzzy model performance (red: process value, blue: fuzzy model)
(Source: Habbi et al., 2009)

The leak detection was determined from the calculation of the residual of T34 as given in Equation 12 as:

$$r(k) = T_{34}(k) - \widehat{T_{34}}(k) \qquad \text{Eqn 12.}$$

where $\widehat{T_{34}}$ is the estimated value, T_{34} is the actual measured value and r is the residual at data point k. Threshold values were used to identify residuals as leaks.

The developed fuzzy model was then tested using leaks with magnitude of 25%, 30% and 40%. The residuals between the actual process values and the estimated data were examined, as shown in Figure 2.9. Figure 2.9 (a) shows the results for the fault-free situation, while Figure 2.9 (b), Figure 2.9 (c) and Figure 2.9 (d) show the results obtained in the presence of leaks with the magnitudes of 25, 30 40%, respectively. In the fault-free condition, the residual values are almost zero. Deviations in residual values were successfully captured by the fuzzy model for all leak values. The time delays for each size of leakage were 40s, 12s and 0s for leakages with 25, 30 and 40% magnitude, respectively. As the leak magnitude increased, detection became easier.

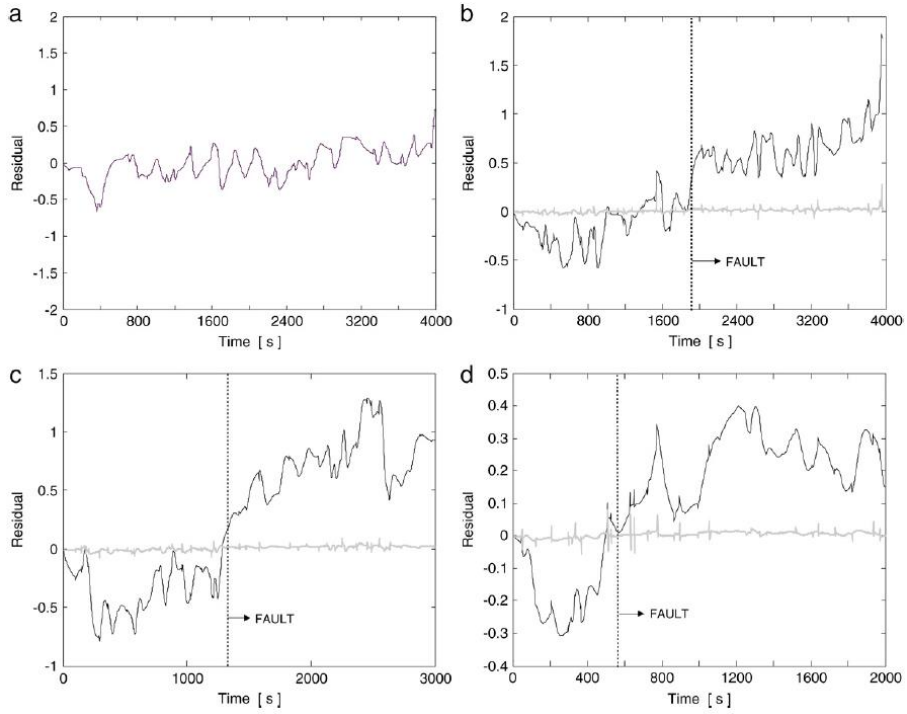

Figure 2.9. Residual behavior (a) fault-free, (b) 25% leakage, (c) 30% leakage, (d) 40% leakage

2.3. Lessons Learned from the Literature

- There are only a few studies investigating data-based leak detection in heat exchangers.
- There are various data analysis techniques to detect leaks in the literature sources.
- Data-based models are relatively easier to implement compared to physics-based models.
- Consultations with operation engineers are very valuable and necessary during the leak detection mechanism construction, and identification of relevant parameters in the system.
- For evaluating the performance of leak detection mechanism, error metrics, such as mean square error, should be used.
- Threshold values are useful to differentiate between leaks and random noises.
- It is easier to detect large leaks compared to small or slowly developing leaks.

CHAPTER 3

MATERIALS AND METHODS

3.1. Real Industry Case

3.1.1. Process Description

The leak case studied throughout the project is in the Integrated Unicracking Processing Unit (IUPU) at İzmit TUPRAS Refinery. The unit feed consists of a combination of Coker Naphtha (CN), Heavy Coker Gas Oil (HCGO), Light Coker Gas Oil (LCGO), Heavy Vacuum Gas Oil (HVGO), Light Vacuum Gas Oil (LVGO) supplied from different units throughout the refinery. IUPU unit is used for producing kerosene and diesel in accordance with certain specifications to meet the restrictions on the sulfur content in fuel oil due to its environmental effects.

The unit is designed in an integrated manner to collect the products from Hydrocracker (HCU), Naphtha Hydro-Treater (NHT) and Diesel Hydro-Treater (DHT) reactor sections in a single fractionator column and separate them into final products. The major parts of the unit consist of reactor and separator parts. Heat exchangers and furnace play a role for the heating of the feed or intermediate products in the unit.

There are three reactors in reactor part, a diolefin reactor and two hydrotreating reactors. Coker naphtha is fed into the diolefin reactor. The diolefin reactor is a single bed reactor in which dienes are converted mainly to mono-olefins. The LVGO and LCGO feeds and the heated recycle gas are combined with the naphtha which comes from the diolefin reactor, and combined feed is sent to the hydrotreating reactor. In this reactor, sulfur and nitrogen are removed from the feed. Since the reaction in the reactors is exothermic, outlet stream temperature of reactor will be higher than that at the rector inlet. Outlet stream is sent to a heat exchanger to reduce its temperature. Cooled stream is sent to the hot separators to separate the by-products from the desired main product. The distillate hot separator stream coming out of the separator is sent to the heat exchangers to be cooled. Figure 3.1 is basic representation of small part of the process. The heat

exchanger where the leakage occurs is the first heat exchanger to which the recycle gas and light hydrocarbon (HC) mixture stream is sent. This heat exchanger is circled in Figure 2.1. After the reactor effluent is cooled, it is sent to the separator where the heavy hydrocarbons are separated from recycle gas and light hydrocarbons. The top output stream of the separator (1) contains recycle gas and light hydrocarbons. It passes through the tube sections of the heat exchanger (with leakage). Liquid hydrocarbon from distillate flash drum passes through the shell part and is used to cool the flow in the tubes.

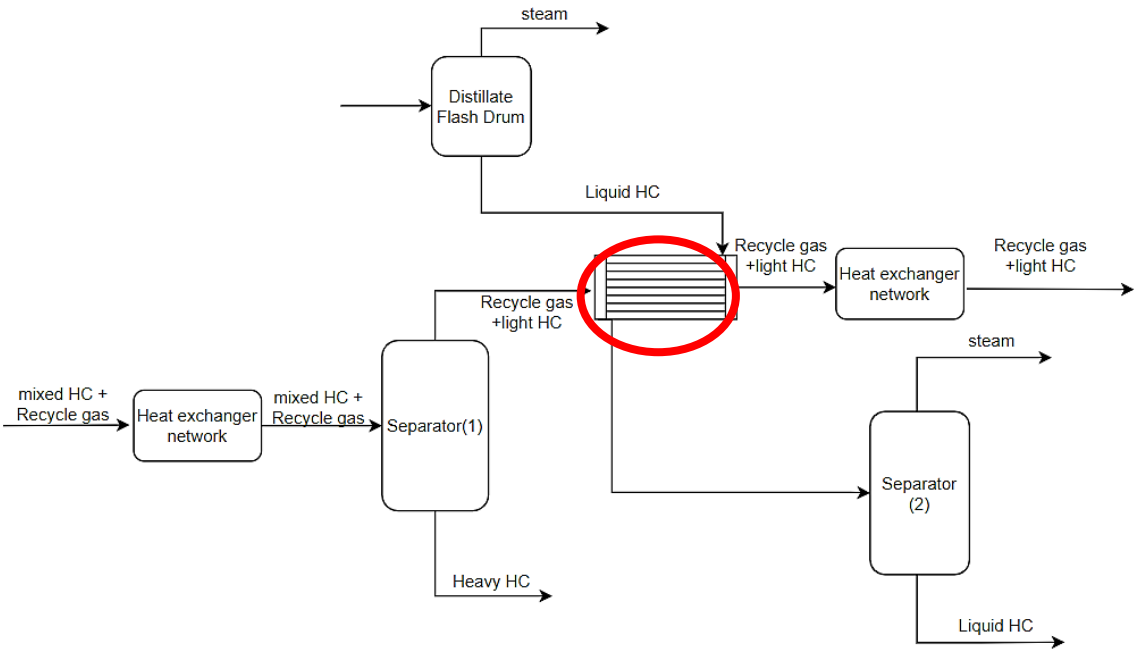

Figure 3.1. The scheme of the unit where the heat exchanger with the leak is located

3.1.2. Problem Definition

As mentioned above, it is important to detect leaks in closed systems, such as a refinery, because the leaks are generally difficult to detect and might cause high harm to the environment and human health. At the beginning of the project, cooperation was made with many units in the refinery such as process and field engineers from different units, such as instrumentation and maintenance unit, and technical safety and environment unit at the İzmit TUPRAS Refinery. Upon these collaborations, it was noted that leaks on the pipelines, valves, connection points and heat exchangers occurred frequently. Leaks were

classified as visible and invisible leaks according to where it occurs, and as gas and liquid leaks based on fluid type. Although the leakage size might be small, it is much easier to detect visible leaks than invisible leaks. When liquid leaks occur, they can be observed clearly, and actions can be taken in a short time. Gas leaks are difficult to detect and therefore there are gas detectors in the field to detect leaks. If a leak in a heat exchanger is not noticed in time, it causes a sudden and dangerous shut down of the process unit. Sudden interruption of the unit feed causes coking on the catalyst. At the same time, the sudden increase in temperature during startup of the unit also causes approximately 3-4 month decrease in catalyst lifespan. When the long-term effect is examined, the catalyst replacement means an extra 10-day downtime for the unit. Also, the downtime of a particular process unit affects other process units in the refinery.

The heat exchanger where leakage occurs is one of the shell and tube heat exchangers in the unit. While light hydrocarbon and recycle gas pass through the tube part of the heat exchanger, liquid hydrocarbon passes through the shell part. The liquid hydrocarbon passing through the shell contains H_2S which is a corrosive chemical. H_2S potentially causes formation of holes on the surface of the tubes and leads to contamination in the heat exchanger. Since the pressure of the fluid passing through the tube is high, leakage occurs from the tube side to the shell side. The increase in the pressure of separator (2) is generally accepted as a sign of leakage in the heat exchanger. The valve opening values of the valve that controls the separator (2) pressure are shown in Figure 3.2. The trend in normal operational interval and leakage is shown in the figure. As a leak occurs from the tube to the shell, an increase is observed in the valve opening.

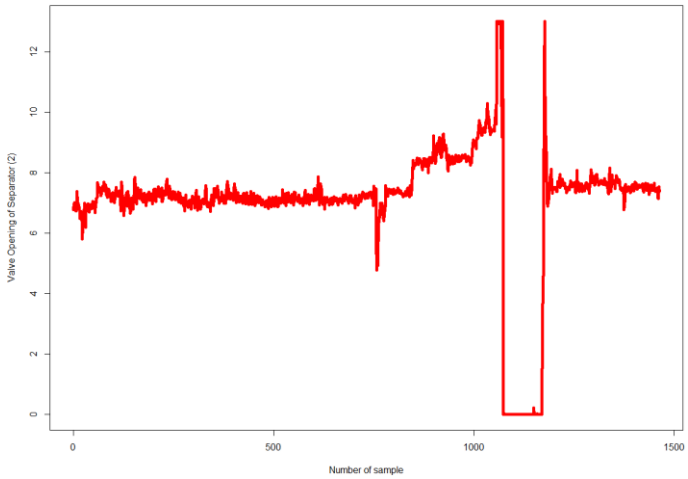

Figure 3.2. Valve opening values for real case

Numerical values are not real process values due to proprietary nature of the refinery process, instead they are symbolized in a way that does not change the trend.

3.1.3. Data Preparation

As mentioned above, the presence of leakage was noticed by the pressure increase in the separator (2). The valve opening (%) that controls the separator (2) pressure has increased during this time. Because of the complexity of refinery processes, we communicated with operation engineers throughout the study so that we could use the relevant variables for this study. There are seven variables that we can use in this case such as inlet and outlet temperatures for the tube and shell sides (°C), stripper column pressure (kg/cm^2), total flowrate of the unit (kg/h) and the valve opening (%) that controls the pressure of separator (2) (for example, valve opening % is shown in Figure 3.2). These variables have been decided upon by consultation with process engineers responsible for the related units. In addition to variables above, logarithmic mean temperature difference (LMTD) was also calculated as shown in Equation 13 (Utamura, Nikitin, and Kato 2008).

$$LMTD = \frac{[(T_h - T_c)_1 - (T_h - T_c)_2]}{\ln\left(\frac{(T_h - T_c)_1}{(T_h - T_c)_2}\right)} \qquad \text{Eqn 13.}$$

In Equation 13, T_h and T_c represent temperature value of hot and cold stream, respectively. Endpoints of heat exchanger are shown as Point 1 and 2.

There have been four different leakage cases on this heat exchanger in 3 years and this is one of the reasons why we work on this heat exchanger. In the second year, there were two different leakage cases with an interval of three months. A leakage case has been experienced in each of the other years. The data preparation part differs according to the applied methods. Since there is no model training process in applying the DWT method, 1-year data on per minute basis is taken from the TUPRAS historian database. While applying the AE and EWMA methods, the data set is divided into training and test datasets. Since the training data set will be used to train the model, it was tried to choose the date range when the unit normally operates. In this process of dataset selection, the support of operation engineers was received. The test data set includes other dates include

anomaly in a year except the training set interval. For the application of AE and EWMA, 1-year data on per minute basis is taken from the TUPRAS historian database.

3.2. Applied Methods

3.1.4. PCA

Statistical methods are in the class of Statistical Process Control (SPC) which are important for safe and reliable operation in industry (Ahsan, Mashuri, Kuswanto, and Prastyo 2018). These methods can show the trend of the process and indicate an anomaly. Principal Component Analysis (PCA) which is used to reduce dimensionality of data and detect the anomalies and Exponentially Weighted Moving Average (EWMA) which is used to detect the anomalies are examples of such statistical methods.

The PCA method is a statistical method applied for data classification and compression. In some studies, PCA is also used as an anomaly detection method. Outlier data (anomaly) can be determined by classification (Huang et al. 2007). In some studies PCA is applied as a preprocessing method (Ahsan, Mashuri, Kuswanto, Prastyo, et al. 2018).

In this thesis study, the PCA method is used to create a small number of principal components (PC) with high variation among all variables. The datasets which are used in this study contain many variables. PCA is used to select the variables that have the most impact on PCs.

In PCA, the first step is to calculate the mean value of the data. Given X is the dataset, X_1, X_2, …, X_N are individual data samples and N is the number of data samples, mean value of data samples is calculated in Equation 14 as:

$$\bar{X} = \frac{X_1+X_2+\cdots+X_N}{N} \qquad \text{Eqn 14.}$$

The first and second principal component (Y_1 and Y_2) shown in Equation 15 & Equation 16 is defined by combining variables X_1, X_2.. and X_N linearly:

$$Y_1 = a_{11}X_1 + a_{21}X_2 + \cdots a_{N1}X_N \qquad \text{Eqn 15.}$$

$$Y_2 = a_{12}X_1 + a_{22}X_2 + \cdots a_{N2}X_N \quad \text{Eqn 16.}$$

All principal components (PCs) are calculated in a similar manner. There are as many PCs as the total number of variables. Therefore, the properties of the data are not lost. Generally the conversion of original data to PC is shown below.

$$Y = AX \quad \text{Eqn 17.}$$

Here, A is the eigenvector matrix, X is the variable vector. Each row of the A gives the a_{ij} values which are loadings for PC in the Equation 17. These values explain the effect of each variable on PCs. Each loading is calculated by using Equation 18.

$$a^2{}_{11} + a^2{}_{12} + \cdots + a^2{}_{1N} = 1 \quad \text{Eqn 18.}$$

Higher values mean stronger interaction. Finally, covariance matrix of the PC is calculated by using Equation 19.

$$C_Y = AC_XA^T \quad \text{Eqn 19.}$$

C_X in Equation 19 is calculated by the Equation 20 below:

$$C_X = \frac{1}{N-1}(X - \bar{X})(X - \bar{X})^T \quad \text{Eqn 20.}$$

Here, C_x is the covariance matrix of the original data, X^T is the transpose of the X, C_y is the covariance matrix that gives an information about variances and covariance of the PCs. Results are given visually using a scree plot where percentage of variance (%) is plotted against PC number (Jollife and Cadima 2016). In this thesis study, these calculations were carried out by usin R programming language.

3.1.5. Discrete Wavelet Transform

Discrete wavelet transform (DWT) is a signal processing method. It is generally preferred to detect small changes in dataset and to reduce noise in the data (Jiang and Liu

2011), (Xu and Huang 2008). The signal is decomposed into levels. Each level consists of a series of coefficients that describe the time evolution of the signal. These coefficients correspond to a certain frequency band. In this method, a mother wavelet known as basis function is used for decomposition of signal into different frequency bands known as multi-level analysis. This method is used to clearly observe the sudden changes in the signal. It provides information in both the time domain and the frequency level.

The method for which the mathematical representation is given below is actually a linear transformation created by shifting and scaling the mother wavelet. It is important to choose the mother wavelet so that it most closely resembles the data signal being studied. Common mother wavelet types are shown in Figure 3.3 (Faust et al. 2015).

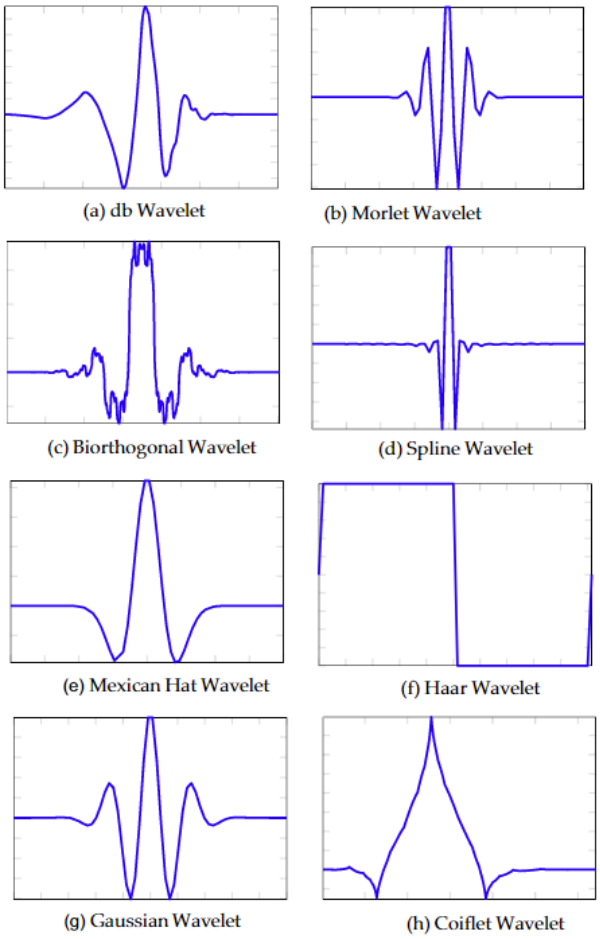

Figure 3.3. Mother wavelet types
(Source: Faust et al., 2015)

Various metrics have been used in the literature for mother wavelet selection. These are peak signal to noise ratio (PSNR) in Equation 21, mean square error (MSE) in Equation 22, mean absolute error (MAE) in Equation 23 and cross correlation (Galya Georgieva-Tsaneva 2014). The high PSNR value indicates the suitability of the selected wavelet and ensures a well reconstructed signal. It is also expected that the MSE and MAE values will be low for a suitable mother wavelet (Kricha, Kricha, and Sakly 2018).

$$PSNR = 10\log_{10}\frac{255}{\frac{1}{N}\sum_{n=1}^{N}|s(n)-\tilde{s}(n)|} \quad \text{Eqn 21.}$$

$$MSE = \frac{1}{N}\sum_{n=0}^{N-1}\big(s(n)-\tilde{s}(n)\big)^2 \quad \text{Eqn 22.}$$

$$MAE = \frac{1}{N}\sum_{n=1}^{N}|s(n)-\tilde{s}(n)| \quad \text{Eqn 23.}$$

In these equations, $s(n)$ represents the original signal, $\tilde{s}(n)$ represents the denoised signal and N represents the number of data samples.

DWT is applied in this study to separate the signal into different frequency levels in order to see the information contained in the signal clearly. For a multi resolution analysis, the signal is passed through high and low pass filters. These filters can be thought of as a means to process the signals. Calculation of DWT is carried out with Mallat-tree decomposition which is shown in Figure 3.4 for a 3-level decomposition (Laaksonen 2013). The g and h represent high pass and low pass filters, respectively. The time and frequency resolution of the signal changes at each level. $d_1[n]$, $d_2[n]$, $d_3[n]$ are called detail coefficients and $a_1[n]$, $a_2[n]$, $a_3[n]$ are called approximation coefficients.

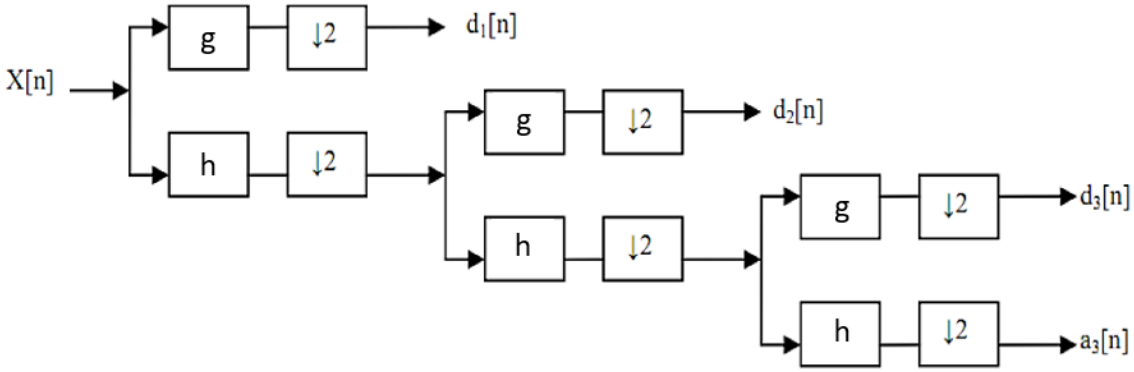

Figure 3.4. Three-level decomposition tree
(Source: Baghbidi, 2011)

Frequency band with respect to level shown in Figure 3.5. In figure, fs is the sampled frequency (Amolins, Zhang, and Dare 2007). It can be seen that increase of level decreases the frequency.

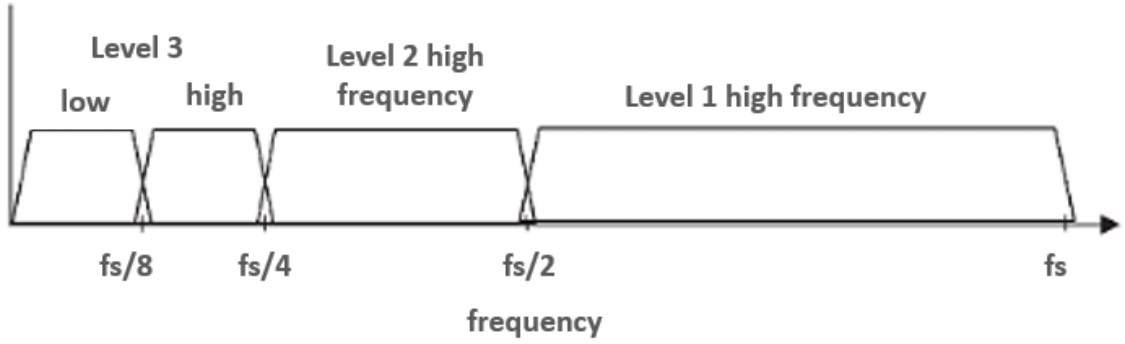

Figure 3.5. Frequency band based on each level
(Source: Amolins et al., 2007)

In Equation 24 and Equation 25, g and h represent high pass filter and low pass filter, respectively, as mentioned above. In the filters, detail and approximation coefficients are calculated by down sampling by two "2 and down arrow" in Figure 3.4 represents this "downsampling by 2" operation. Down sampling means reducing the sampling rate or removing some samples from the signal (such as dropping the middle sample between among three samples). With down sampling, the scale of the signal is changed. Low pass filter cleans the high frequency components from the signal and the

output of the low pass filter give an approximation of the original signal. High pass filter cleans the low frequency component from signal, and it gives detail information about the signal. The high and low pass filter equations are given in equations as:

$$a[n] = y_{low}[n] = \sum_n x(k)\, g[2n - k] \quad \text{Eqn 24.}$$

$$d[n] = y_{high}[n] = \sum_n x(k)\, h[2n - k] \quad \text{Eqn 25.}$$

where n is the number of samples, $x(k)$ is the original signal, k is sampled digit which is in the range of $0< k < n/2\text{-}1$ or $n/2< k < n\text{-}1$, y_{high} is output of the high pass filter and y_{low} is output of low pass filter (Orea-flores, Gallegos-funes, and Arellano-reynoso 2019).

The digital representation of DWT is shown in Equation 26 below. Here, a is the scaling factor, b is the translation factor, k refers to the number of samples in the signal $g^*(k)$ refers to mother wavelet. m and n are the integer parameters of a ($a = {a_0}^m$) and b ($b = n{b_0}{a_0}^m$). Mother wavelet is scaled and shifted along the signal (Barros, Diego, and De Apraiz 2012).

$$DWT_{(m,k)} = \frac{1}{\sqrt{a}} \sum_n x(n) g^* \left(\frac{k-b}{a}\right) \quad \text{Eqn 26.}$$

Original signal can be reconstructed with the summation of all detailed coefficients ($d_1[n]$, $d_2[n]$, $d_3[n]$) and last level approximated coefficient ($a_3[n]$) (Souza, Cruz, and Pereira 2000). This process is called as inverse DWT (Emmanuel 2012).

Wavelet toolbox used for this method in MATLAB2021a. The procedure of DWT is presented given below as:

- Data is taken from TUPRAS Historian Database.
- The mother wavelet is selected which is suitable for the data set we use.
    - For this, data is denoised using different mother wavelets and evaluation metrics are calculated (Equation 21, Equation 22 and Equation 23).
- Then, the data is decomposed, and the detail and approximation coefficients are calculated. Reconstructed data is obtained with the sum of these coefficients.

3.1.6. Auto Encoder

Auto encoder (AE) is an artificial neural network technique that is used to reduce dimensionality of dataset (Sakurada and Yairi 2014). It is an unsupervised technique which uses unlabeled input and output data meaning that there is no labeling that give information about what the anomaly is (Tutkan, Ganiz, and Akyokuş 2016). As this technique is used for size reduction, it can be used to extract features, denoise and recognize images (Sakurada and Yairi 2014). The neural network of AE is shown in Figure 3.6. AE has five hyperparameters in total, which are bottleneck (code layer), encoder, decoder, loss function and epoch number. First, input is sent to encoder and compressed here (Sublime and Kalinicheva 2019). The compressed data is stored in the bottleneck. Decoder reconstructs the data comes from bottleneck (Mirsky et al. 2018). Bottleneck is the important part of the AE, compressed data is taken place here. Each part includes hidden layers, and each layer consists of nodes. These nodes represent features and are connected with other nodes. Epoch number represents how many times the algorithm will run. When deciding on the epoch number, a graph of MSE with respect to epoch number is plotted. This graph is called the learning curve (Elbattah et al. 2021).

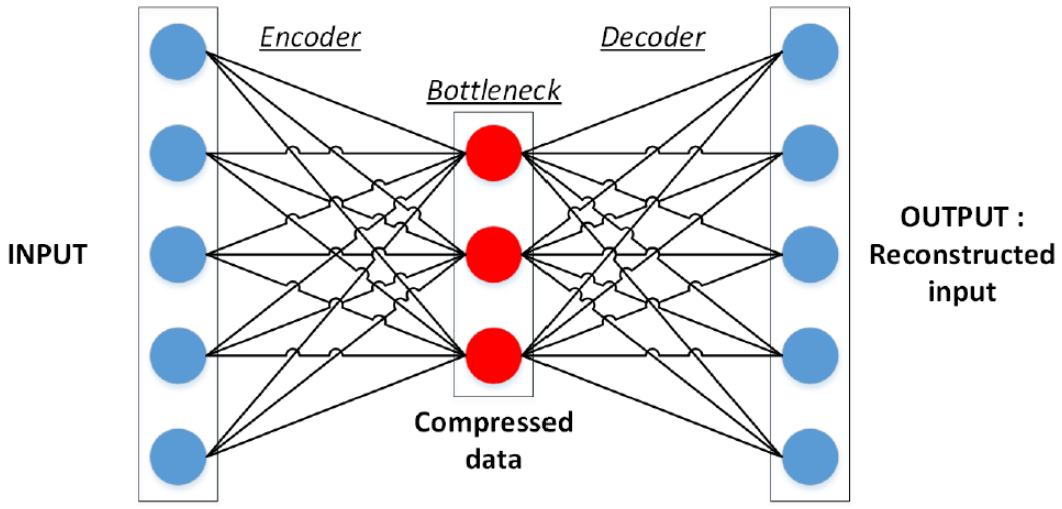

Figure 3.6. Architecture of AE
(Source: Sublime & Kalinicheva, 2019)

Mathematical representation of data compression is given in equations below. In the Equation 27, x is the input, W is the weight matrix, b and b' are bias vectors, z is the bottleneck dimension, σ and σ' are the activation functions , and x' is the output (Sagheer and Kotb 2019). Equation 27 represents the encoding network. Equation 28 belong to decoding network of AE. This calculation is repeated in each node.

$$z = \sigma(Wx + b) \quad \text{Eqn 27.}$$

$$x' = \sigma'(W'z + b') \quad \text{Eqn 28.}$$

AE layers are created considering the structure used in a study in the literature (Tavakoli et al. 2020). The encoding part consists of 3 layers and the layers are ordered according to the decreasing number of nodes. These layers are created by decreasing node number as 100, 50, 20 in accordance with the purpose of AE and can compress the data while preserving the important features. The code layer can contain as many nodes as our variable number. In order to decide on the number of nodes, training accuracy values are examined, and the results are given in the Figure 3.7. Maximum accuracy is achieved with 6 nodes. Decoder network is a mirror image of encoder part, and layers are incrementally built up to 20, 50, and 100 nodes. The input and output layers also consist of an equal number of nodes and are equal to the number of variables.

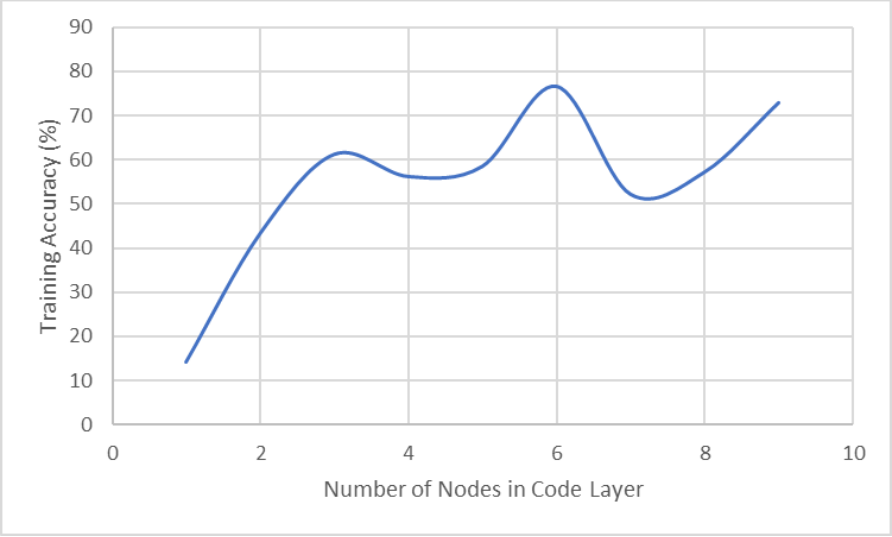

Figure 3.7. Training Performance based on number of nodes in code layer

Another important part is the activation function selection. The activation function is responsible for transmitting the sum of the weights calculated in the node to the other node. Basic representation for one node is shown in Figure 3.8.

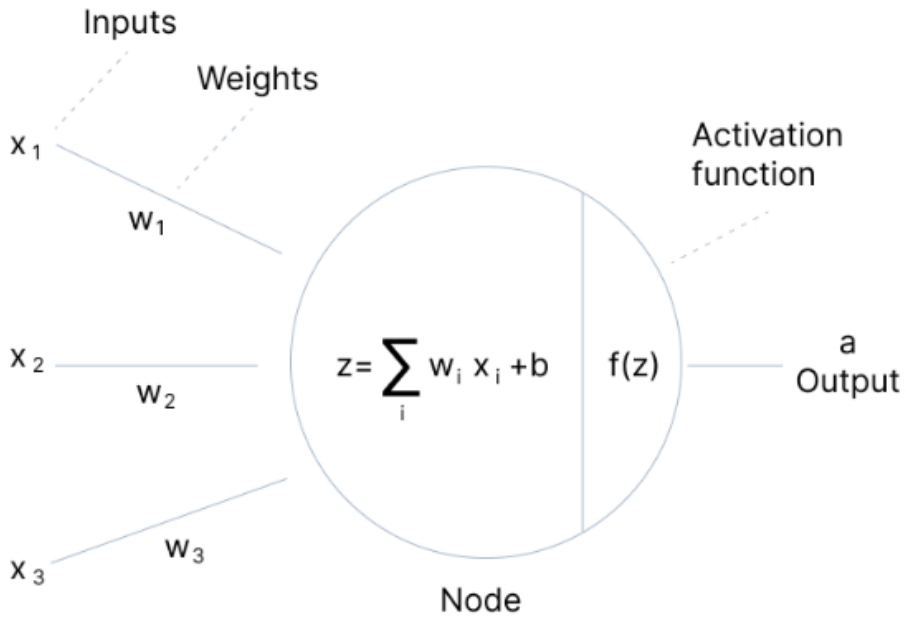

Figure 3.8. The role of activation function
(Source: Theodoridis, 2020)

Non-linear functions are generally preferred for AE networks because nonlinearity overcomes backpropagation problem. The meaning of backpropagation is adjusting of weights repeatedly in order to minimize the loss function. Because of that linear activation functions are not preferred for AE. Two kinds of non-linear activation functions are generally used for AE. One of them is Rectified Linear Unit (ReLU) and the other is sigmoid. ReLU is usually selected as a default function (Kumaresan et al. 2021). Sigmoid function has a disappearing gradient problem, The meaning of this problem that loss function closes to zero. It causes reducing the training performance of network. ReLU can overcome this problem and provides easy learning. In this study, ReLU is used as an activation function (Theodoridis 2020).

Finally, reconstruction error is calculated as a loss function which is used as the outlier score. Mathematical representation of reconstruction error is shown in Equation 29. (An et al., 2015).

$$Reconstruction\ Error = \sqrt{\frac{\sum_{i=1}^{n}(x_i - x'_i)^2}{n}} \quad \text{Eqn 29.}$$

where n refers to number of samples, x_i is the original data sample and x'_i is the reconstructed data sample.

The procedure of AE is given below as:

- Data is taken from TUPRAS Historian Database.
- LMTD values are calculated by using temperature values.
- The data is split as training and test dataset (training set: normally operated time period; test set: the dataset which has a probability of including anomalies).
- Model hyperparameters are determined and the model is trained.
- Test data set is fed to the model as an input.
- Reconstruction error is calculated using x and x'.

When another dataset containing anomaly is given to the model, the difference between the input and the reconstructed data will be high, since the trend of the data will be different from the training dataset. While this error is small for normal data samples, it is expected to be high for data with anomaly.

3.1.7. Exponentially Weighted Moving Average

EWMA is a statistical anomaly detection method. EWMA analyzes historical data to determine any deviation in data. This method uses three control limits, namely upper control limit (UCL), center line (CL) and lower control limit (LCL). UCL and LCL play an important role to determine the control region. EWMA method detects the anomalies based on these limits. EWMA can be given as in Equation 30:

$$Z_i = \lambda X_i + (1 - \lambda) Z_{i-1} \quad \text{Eqn 30.}$$

where Z_i is the EWMA at time i, λ is the weighting factor ($0 < \lambda < 1$) and X_i is the residual of the variable to be predicted at time i. Residuals are differences between the

predicted and measured values. In order to predict desired variable and to consider effect of other variables, Multiple Linear Regression (MLR) method is used and it is shown in Equation 31 (Zhao, Wang, and Xiao 2013).

$$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi} + \varepsilon \quad \text{Eqn 31.}$$

Here, p is the number of independent variables, $x_{1i}, x_{2i}, \ldots, x_{pi}$ are independent variables, $\beta_0, \beta_1, \ldots, \beta_p$ are the regression coefficients, ε is the random error (residual) term, and y_i is the predicted variable. Residuals of the model was used to detect small changes in dataset (Zhou and Tang 2016).

λ is important to consider for detecting anomalies in time series information. λ can be calculated with the equation of λ=1-θ. The θ is the coefficient of the Autoregressive Integrated Moving Average (ARIMA) model. ARIMA is a model used for time series forecasting (Ye, Borror, and Zhang 2002). The general notation of ARIMA is ARIMA (p,q,d). The p,q,d are parameters of ARIMA where p is the number of autoregressive, d is the degree of differencing, and q is the order of the moving average. ARIMA (0,1,1) is proposed in the literature for detecting small shifting in dataset (Kandananond 2014). Also, selection of weighting factor depends on user. A λ value of 1 indicates that the weights of the last measurements are more dominant. Conversely, a λ value close to zero means that the old data has a weight (Zhou and Tang 2016).

Referring back to UCL and LCL, the control limits are defined in Equation 32, Equation 33, and Equation 34 as:

$$LCL_i = \mu_0 - L\sigma\sqrt{\frac{\lambda}{(2-\lambda)}[1-(1-\lambda)^{2i}]} \quad \text{Eqn 32.}$$

$$CL = \mu_0 \quad \text{Eqn 33.}$$

$$UCL_i = \mu_0 + L\sigma\sqrt{\frac{\lambda}{(2-\lambda)}[1-(1-\lambda)^{2i}]} \quad \text{Eqn 34.}$$

where σ is the standard deviation of data, μ_0 is the target value and L is the number of standard deviations from the CL.

The procedure of EWMA is presented below as:

- Data is taken from TUPRAS Historian Database
- The data set is divided into two groups as training and test set.
- UCL, CL and LCL values are determined using the training set.
- MLR is applied to predict variables.
- Residual values are calculated.
- ARIMA coefficient θ and weighting factor λ are determined.
- The EWMA control chart is created.
- The area between both UCL and LCL is defined as the control limits.
- Values outside these two lines are defined as anomaly.

R programming language was used to implement this method.

CHAPTER 4

BENCHMARKING OF THE METHODS USING TEP

4.1. General Overview of Benchmarking Studies

In this chapter, the methods were applied on the Tennessee Eastman Process (TEP) benchmark data series given in the literature in order to check the applicability of the three methods explained in Chapter 3, namely DWT, AE and EWMA. Data is divided into two groups as training and test datasets. In Section 4.2, the flow diagram of the chemical process and the variables are explained. For the system with 73 process variables, the PCA method is applied to determine the variables that may be related to the fault. The results of the applied DWT, AE and EWMA methods are given in Chapter 3.

4.2. Benchmark Dataset

Tennessee Eastman Process (TEP) benchmark dataset is an important data source for the field of fault detection and diagnosis, alarm management or control loops. It has been published by Down and Vogel in 1993 (Ricker 1996). The basic process diagram of TEP is shown in the Figure 4.1 below. As it can be seen in Figure 4.1, this proses consists of 5 main units, namely, a reactor, a condenser, a stripper, a separator, and a compressor. There are 73 process variables (PVs) such as volumetric flowrate (F), pressure (P), temperature (T), level (L) and concentration (A). In addition to PVs, 12 manipulated variables (MVs) are given. Description of PVs and MVs are shown in Table 4.1 & Table 4.2, respectively (Gianluca Manca, n.d.), (Reinartz, Kulahci, and Ravn 2021), (Kiss, Genge, and Haller 2015). The data set of TEP is shared as open source. This data set consists of faulty free and faulty data set.

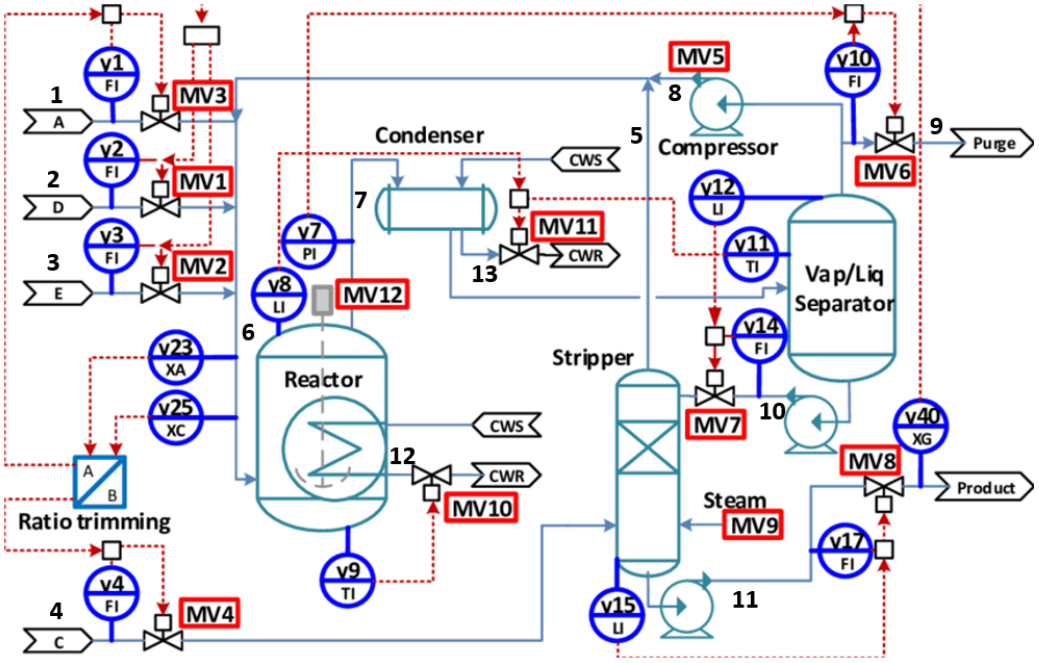

Figure 4.1.TEP model
(Source: Kiss et al., 2015)

CWS: Condenser water supply, CWR: condenser water return

Table 4.1. Description of Process Variables
(Source: Reinarts et al., 2021)

Variable	Description	unit	Variable	Description	unit
XMEAS(1)	A Feed (stream 1)	kscmh	XMEAS(12)	Product Separator Level	%
XMEAS(2)	D Feed (stream 2)	kg/hr	XMEAS(13)	Product Separator Pressure	kPa gauge
XMEAS(3)	E Feed (stream 3)	kg/hr	XMEAS(14)	Product Separator Underflow (stream 10)	m^3/hr
XMEAS(4)	A and C Feed (stream 4)	kscmh	XMEAS(15)	Stripper Level	%

(cont. on the next page)

Table 4.1 (cont.)

Variable	Description	unit	Variable	Description	unit
XMEAS(5)	Recycle Flow (stream 8)	kscmh	XMEAS(16)	Stripper Pressure	kPa gauge
XMEAS(6)	Reactor Feed Rate (stream 6)	kscmh	XMEAS(17)	Stripper Underflow (stream 11)	m^3/hr
XMEAS(7)	Reactor Pressure	kPa gauge	XMEAS(18)	Stripper Temperature	deg C
XMEAS(8)	Reactor Level	%	XMEAS(19)	Stripper Steam Flow	kg/hr
XMEAS(9)	Reactor Temperature	deg C	XMEAS(20)	Compressor Work	kW
XMEAS(10)	Purge Rate (stream 9)	kscmh	XMEAS(21)	Reactor Cooling Water Outlet Temperature	deg C
XMEAS(11)	Product Sep Temp	deg C	XMEAS(22)	Separator Cooling Water Outlet Temperature	deg C

(kscmh: kilo standard cubic meter per hour)

Table 4.2. Description of Manipulated Variables
(Source: Reinarts et al., 2021)

Variable	Description	Variable	Description
XMV(1)	D Feed Flowrate (stream 2)	XMV(7)	Separator Pot Liquid Flow (stream 10)
XMV(2)	E Feed Flowrate (stream 3)	XMV(8)	Stripper Liquid Product Flow (stream 11)
XMV(3)	A Feed Flowrate (stream 1)	XMV(9)	Stripper Steam Valve (%)
XMV(4)	A & C Feed Flowrate (stream 4)	XMV(10)	Reactor Cooling Water Flowrate
XMV(5)	Compressor Recycle Valve (%)	XMV(11)	Condenser Cooling Water Flowrate
XMV(6)	Purge Valve (stream 9)	XMV(12)	Agitator Speed

In this simulated process, A, C, D, & E components (in gas phase) are called as feed for the system. Recycle stream of the system and A, D, E are fed to the reactor and liquid G and H are obtained. The reactions taking place in the reactor are shown below (Park et al. 2019). The reactions are exothermic, irreversible reactions with first order reaction kinetics with respect to concentration of reactants.

$$A_{(g)} + C_{(g)} + D_{(g)} \rightarrow G_{(liq)}, Product\ 1 \quad \text{Eqn 35.}$$

$$A_{(g)} + C_{(g)} + E_{(g)} \rightarrow H_{(liq)}, Product\ 2 \quad \text{Eqn 36.}$$

$$A_{(g)} + E_{(g)} \rightarrow F_{(liq)},\ Byproduct \quad \text{Eqn 37.}$$

$$3D_{(g)} \rightarrow 2F_{(liq)},\ Byproduct \quad \text{Eqn 38.}$$

After the reactions occur, liquid product stream composed of G, H and F is fed to the condenser to cool down and then fed to a vapor-liquid separator. While condensed products are fed to the stripper column, uncondensed product is fed back to the reactor.

Product mixture of G and H is separated from each other in the stripper. The inert (B) and byproduct (F) are removed from the process (Park et al. 2019).

In the simulation, various disturbances are created in order to study the faulty characteristics of TEP. Created disturbance types are shown in Table 4.3. The disturbances are numbered and listed based on the type of disturbance as step, random variation, slow step and sticking. Sticking fault is encountered on sticking valves. This type of fault is noticed on the sudden change in the values of sticking of valves (Reinartz, Kulahci, and Ravn 2021).

Table 4.3. Process Disturbances in TEP
(Source: Park et al., 2019)

Variable	Description	Type	Variable	Description	Type
IDV(1)	A/C Feed Ratio, B Composition Constant (stream 4)	Step	IDV(11)	Reactor Cooling Water Inlet Temperature	Random Variation
IDV(2)	B Composition, A/C Ratio Constant (stream 4)	Step	IDV(12)	Condenser Cooling Water Inlet Temperature	Random Variation
IDV(3)	D Feed Temperature (stream 2)	Step	IDV(13)	Reaction Kinetics	Slow Drift
IDV(4)	Reactor Cooling Water Inlet Temperature	Step	IDV(14)	Reactor Cooling Water Valve	Sticking
IDV(5)	Condenser Cooling Water Temperature	Step	IDV(15)	Condenser Cooling Water Valve	Sticking

(cont. on the next page)

Table 4.3. (cont.)

Variable	Description	Type	Variable	Description	Type
IDV(6)	A Feed Loss (Stream 1)	Step	IDV(16)	Unknown	Unknown
IDV(7)	C Header Pressure Loss - Reduced Availability (Stream 4)	Step	IDV(17)	Unknown	Unknown
IDV(8)	A, B, C Feed Composition (stream 4)	Random Variation	IDV(18)	Unknown	Unknown
IDV(9)	D Feed Temperature (Stream 2)	Random Variation	IDV(19)	Unknown	Unknown
IDV(10)	C Feed Temperature (Stream 4)	Random Variation	IDV(20)	Unknown	Unknown

As shown in Table 4.3, 20 different disturbances had described by manipulated variables in simulation. Among these disturbances, to continuously see the effect of the disturbance on process variables, random variation type disturbance has been chosen. At the same time, IDV (12) was chosen as the variation in cooling water temperature among the 20 disturbances in order to be parallel to the case studied at TUPRAS. The data set in the selected disturbance was obtained by creating sudden changes in the cooling water temperature. In our study, measured process values and manipulated variables are taken into account in the PCA method. The PCA method is applied to determine the variables most affected by the disturbance. Composition values (XMEAS_23 to XMEAS_41) related to the measured product composition are considered as laboratory values (they are not continuous process measurements) and are not taken into account in PCA calculations. PCA method is implemented using the process variables such as XMEAS_1 to XMEAS_22 and manipulated variables such as XMV_1 to XMV_11. R programming

language is used for the application of PCA. The obtained results are shown in Figure 4.2. The dots denoted by dark navy blue and orange colors show the high positively and negatively correlated variables with each other, respectively.

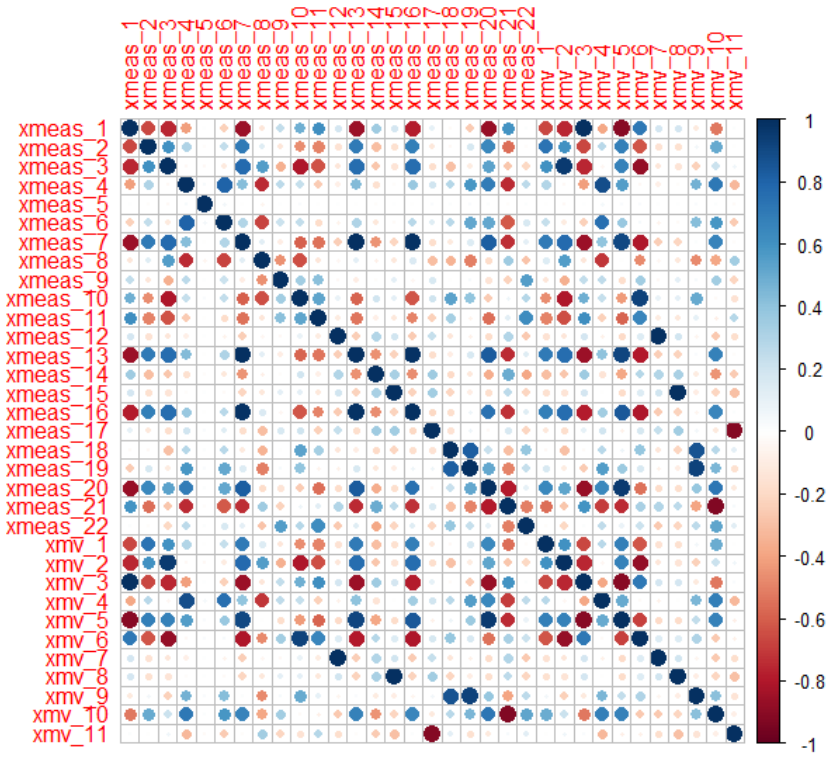

Figure 4.2. Correlation matrix for variables

In addition, percentages of variation for each principal component (PC) are shown in Figure 4.3. As seen, the PC1 and PC2 are the principal components with the two highest percentage variances with PC1 being the PC showing the highest variance, as expected.

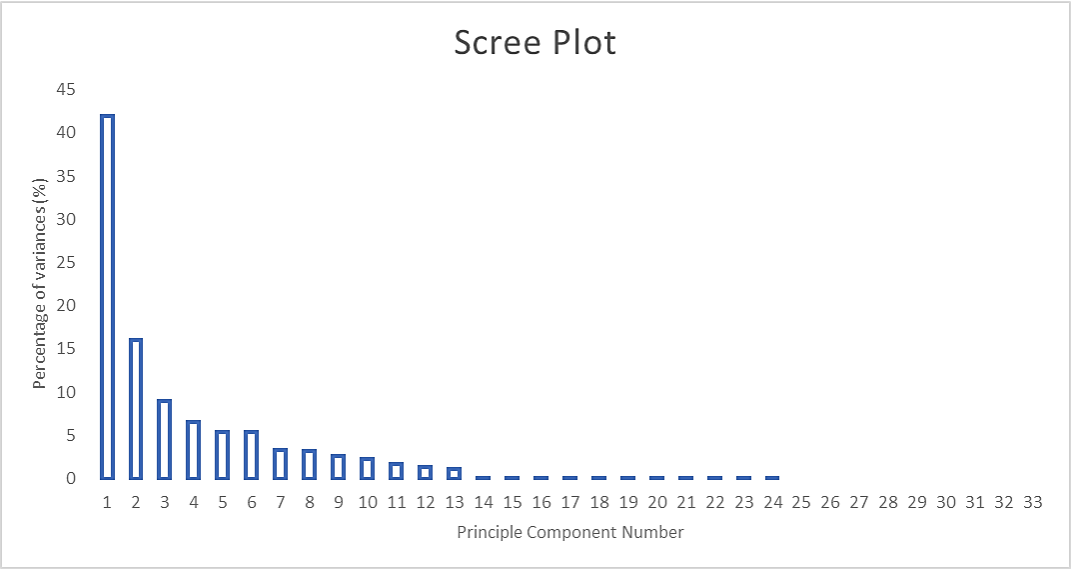

Figure 4.3. PCA cumulative variance plot

Table 4.4 contains the standard deviation and proportion of variance values of PC1 and PC2. A total of 58% variation was observed with both components. Selection of PC has high variance is important because PCs with high variance can represent the pattern of data (Cureton and D'Agostino 2019).

Table 4.4. Standard deviation and proportion of variance of PC1 and PC2

	Principle Component 1 (PC1)	Principle Component 2 (PC2)
Standard Deviation	4.698	2.907
Proportion of variance	0.42	0.16

Figure 4.4 and Figure 4.5 show the most effective variables on PC1 and PC2 that have high variation. XMEAS_6 to XMEAS_11 (Reactor Feed Rate (stream 6), Reactor Pressure, Reactor Level, Reactor Temperature, Purge Rate (stream 9), Product Separator Temperature, respectively), XMEAS_13 (Product Separator Pressure), XMEAS_15 (Stripper Level), XMEAS_16 (Stripper Pressure), XMEAS_18 (Stripper Temperature), XMEAS_19 (Stripper Steam Flow), XMEAS_21 (Reactor Cooling Water Outlet Temperature) & XMEAS_22 (Separator Cooling Water Outlet Temperature), and XMV_1 to XMV_6 (D Feed Flow (stream 2), E Feed Flow (stream 3), A Feed Flow

(stream 1) , A and C Feed Flow, Compressor Recycle Valve, Purge Valve (stream 9), respectively) were chosen as variables with high impact.

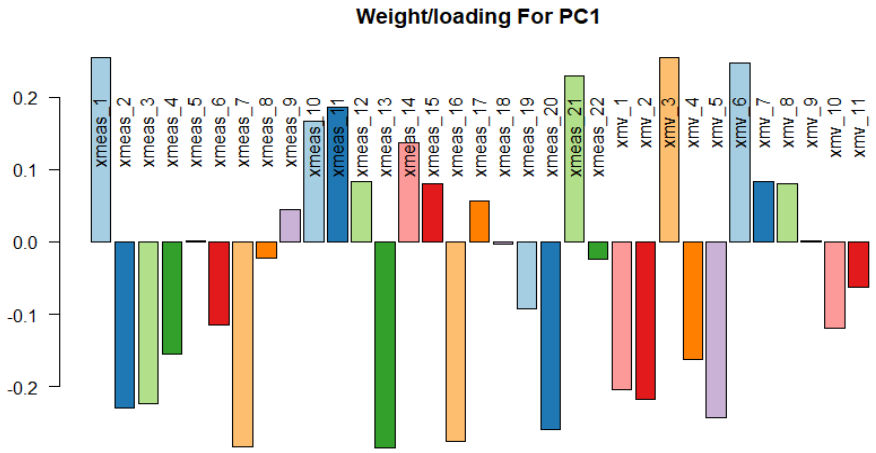

Figure 4.4. Weight/loading for PC1

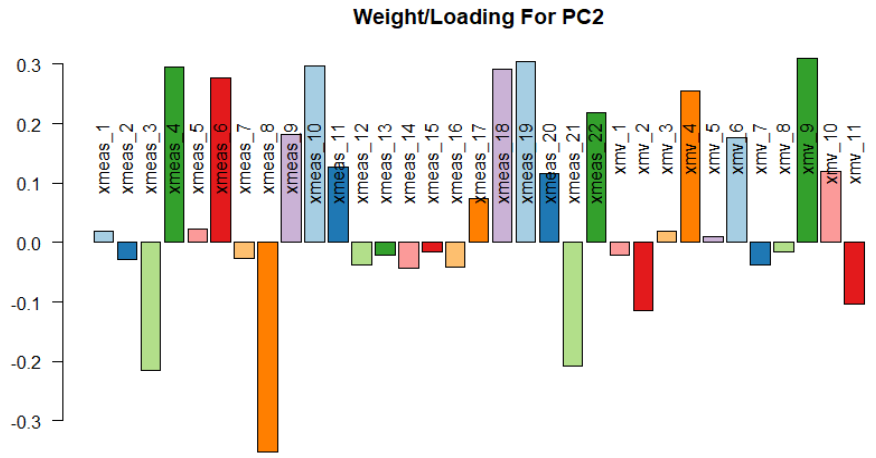

Figure 4.5. Weight/loading for PC2

A dataset consisting of 82,500 data samples was created by taking the 30,000 data samples from the training data set and 52,500 data samples from the test data set. The graphs for XMV_6 (purge valve), XMEAS_6 (reactor feed rate), XMEAS_7 (reactor

pressure) and XMEAS_22 (separator cooling water outlet temperature) variables are as shown in the Figure 4.6. In each plot, x axis shows the number of samples and y axis shows the value of process variables. The anomaly caused an increase in the data amplitude TEP benchmark is the data set which is known to consist of faulty data. As mentioned above, a faulty data set was created with 20 different disturbance types.

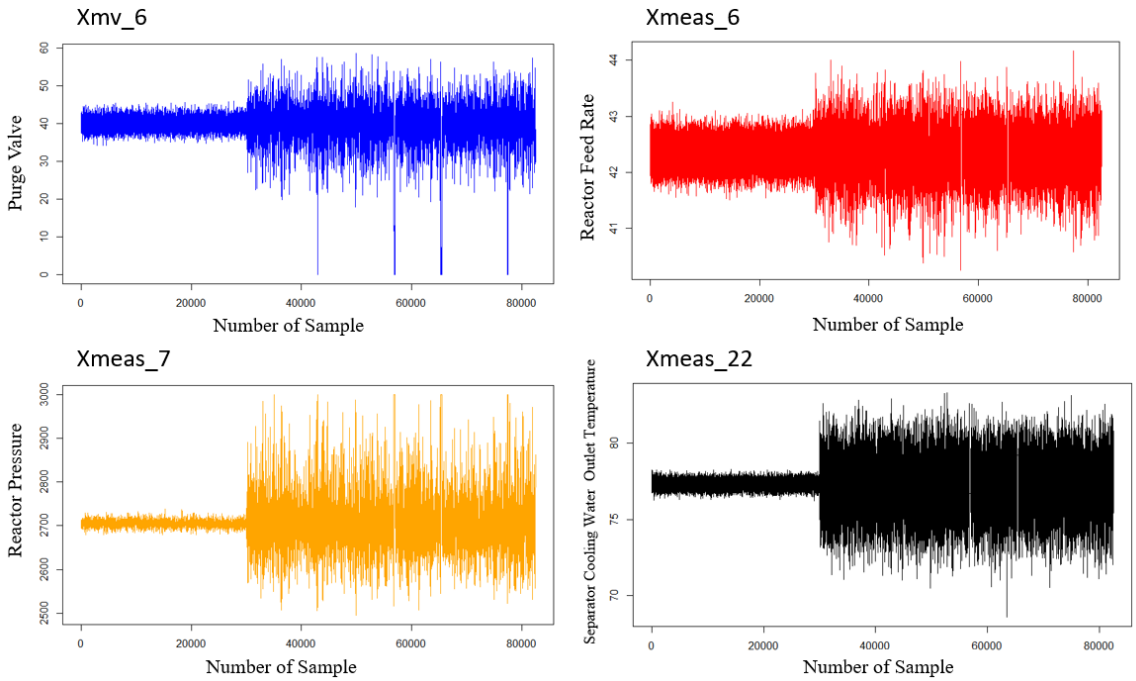

Figure 4.6. Process Variables (Blue: XMV_6, red: XMEAS_6, yellow: XMEAS_7, black: XMEAS_22)

In this study, the sudden changes in the sensor read values due to the disturbance in the cooling water temperature that cools the separator are called anomalies. The applicability of the methods was checked in the data set with known anomalies. Three different anomaly detection methods were applied on this data set. These are DWT, AE and EWMA methods. In the DWT method, the data sets without and with anomaly were combined. While applying the AE and EWMA method, the data set was grouped as training and test data sets, and the models were trained with the training set. Detailed explanations and the obtained results are given in the following sections.

4.2.1. DWT

The DWT method is preferred in most studies to reduce the noise in the signals and thus to obtain clearer information from the data. An important point in DWT is the selection of the appropriate wavelet for the dataset. These wavelets are haar, sym 4-12, Db 4-12 as mentioned above (section 3.3.2). Selected wavelets are determined as wavelets with good reconstruction feature in the literature (Emmanuel 2012). This wavelet selection is evaluated according to the PSNR, MSE, MAE and cross correlation metrics. Results of the PSNR, MAE and MSE values with respect to wavelet type are shown in Table 4.5 While applying the AE and EWMA method, the data set was grouped as training and test data sets, and the models were trained with the training set. Detailed explanations and the obtained results are given in the following sections (Horé and Ziou 2010).

Table 4.5. Performance metrics for wavelet type

Wavelet Type	**MSE**	**MAE**	**PSNR**	**Cross correlation**
Haar	2.988	1.359	30.620	0.948
Sym4	2.310	1.209	31.734	0.959
Sym6	2.274	1.201	31.808	0.960
Sym8	2.264	1.198	31.825	0.960
Db4	2.280	1.202	31.795	0.960
Db8	2.269	1.195	31.816	0.960
Db12	**2.264**	**1.194**	**31.825**	**0.960**

When the values in the table are examined and the studies in the literature are taken into consideration, we can say that the most suitable wavelet type for the TEP benchmark data set is Db 12 with the highest PSNR and lowest MSE & MAE (Payan and Antonini 2006), (Kricha, Kricha, and Sakly 2018), (Galya Georgieva-Tsaneva 2014).

In order to decide the number of levels, the same metrics (PSNR, MAE, MSE) were calculated for each level as shown in the Table 4.6. These metrics are used for the

comparison of the reconstruction performance of the DWT. The level selection was done by considering the high PSNR and low error values, which corresponds to the fifth level.

Table 4.6. Performance metrics for each level

Number of Level	MSE	MAE	PSNR	Cross correlation
Level 1	2.264	1.193	31.825	0.960
Level 2	1.878	1.086	32.637	0.967
Level 3	1.517	0.975	33.565	0.973
Level 4	1.009	0.795	35.334	0.982
Level 5	**0.356**	**0.465**	**39.861**	**0.993**
Level 6	0.356	0.521	39.755	0.990

Figure 4.7 shows the spectra of data with respect to level number. Here, x axis represents the number of samples and y axis represent the level number of DWT. The first 30,000 samples contain anomaly free data. The next 52,500 data samples belong to faulty data. There is a decreasing frequency from level 1 to level 5 (from bottom to top). It can be seen that it is easier to understand information from time series data on the low frequency layer. Wan et al. also carried out anomaly detection studies using the DWT method (Wang et al. 2018). Visually, it can be observed in the Figure 4.7 that level 5 gives clearer information about the presence of anomaly. As the number of levels increased, the presence of faulty data becomes more obvious. This also supports the performance metrics of level 5 in Table 4.6. It can be also said that with the increasing decomposition level (from level 1 to level 5), frequency resolution increase, a situation in agreement with uncertainty principle (Vošvrda and Schürrer 2015).

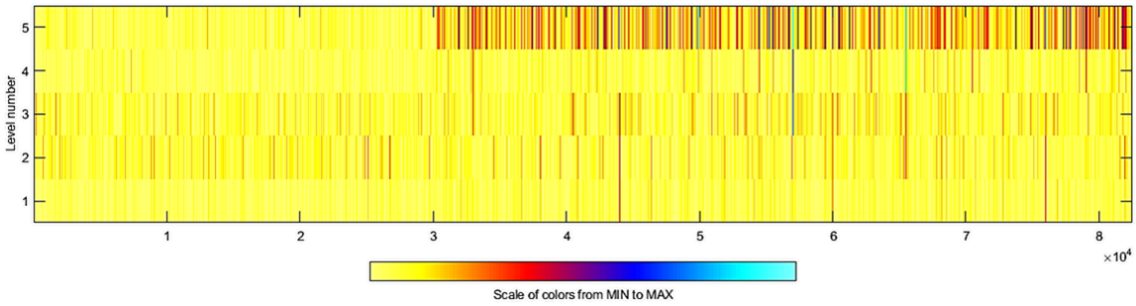

Figure 4.7. Spectra of data based on level number

After the selection of level and wavelet type, multilevel decomposition was performed. The multilevel decomposition results are shown in Figure 4.8. In Figure 4.8, d_1, d_2, … d_n are the detail coefficients and a_n is the approximation coefficient in the highest level. (Souza, Cruz, and Pereira 2000). As seen in Figure 4.7, we can say that the peaks occurring from anomalies are visually observed clearly at the fifth level.

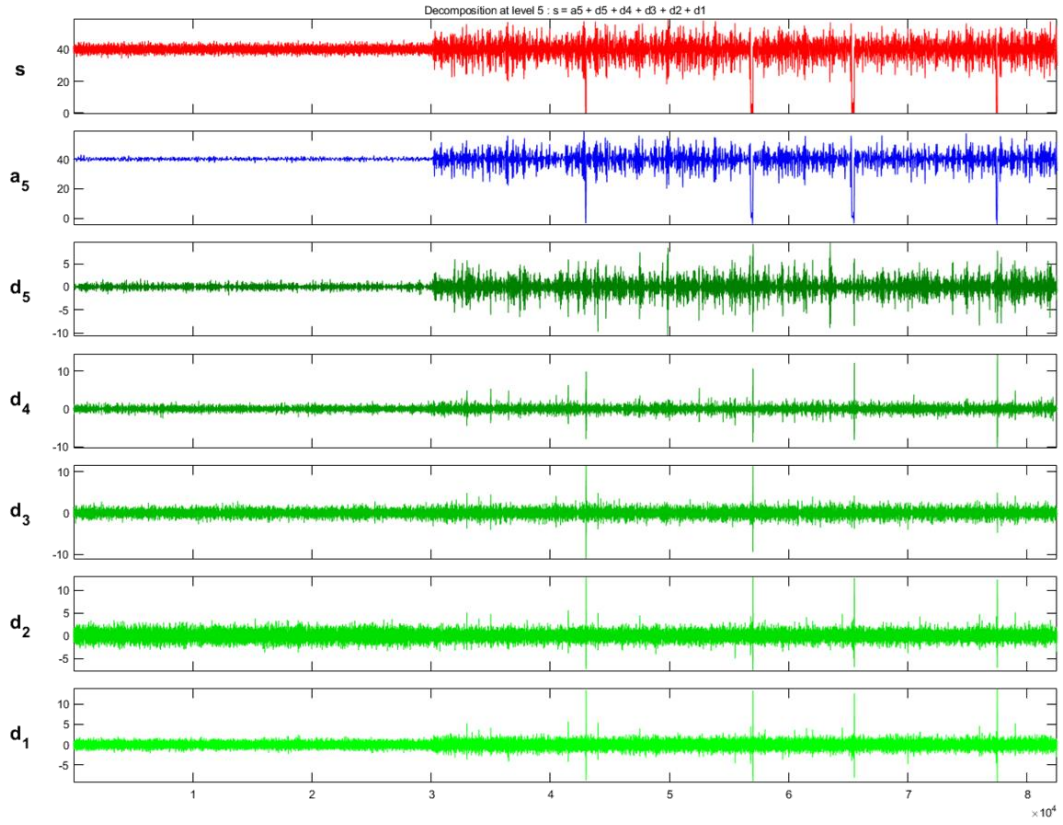

Figure 4.8. Five level decomposition with MATLAB

4.2.2 AE

The next implemented method is AE. For this method, the data set was handled as two distinct sets: a training set without any anomaly and a test set containing anomalies, and the steps are explained in Section 3.2.3 are followed.

Learning curve for the TEP dataset is shown in Figure 4.9. Here, x axis refers to epoch number and y axis refers to MSE value. Best training performance with the lowest MSE values is obtained for epoch number 500.

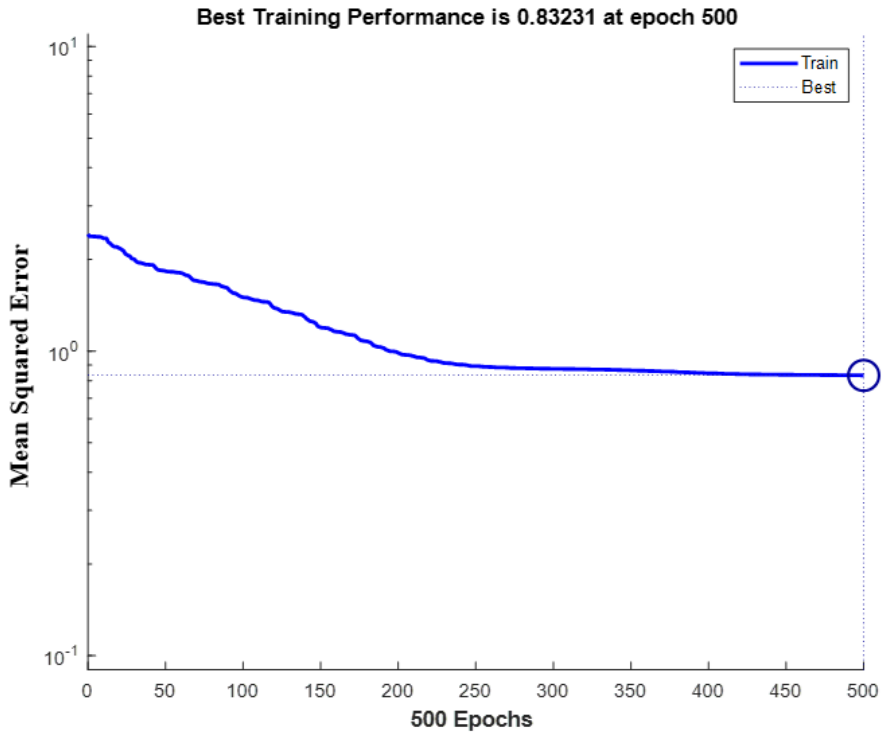

Figure 4.9. Learning curve for TEP dataset

For this method, 20,000 data samples from TEP are used as a training dataset to train the model. Faulty dataset contains 52,500 data samples from TEP. Here, it is worthwhile to mention that we use historical data set while working on TUPRAS case for which the data can be obtained continuously. On the other hand, in TEP benchmark, the data set is shared separately as fault-free and faulty dataset. Because of that, the TEP data set is turned in to a continuous data set by combining the fault-free and faulty data sets. The purpose of this combination is also to see the difference between the data set with

and without anomalies. With this dataset created for the purpose of anomaly detection, our aim is to detect an anomaly whose existence is known and to control the workability of the model.

Result of the AE is shown in Figure 4.10. Red line splits the training data from test data. Top plot shows the predicted and measured process values, and bottom plot shows the reconstruction error due to the difference between measured and predicted values. With the occurrence of anomaly, reconstructed MSE values started to increase. This indicates that AE method can be used to detect the anomalies in the dataset.

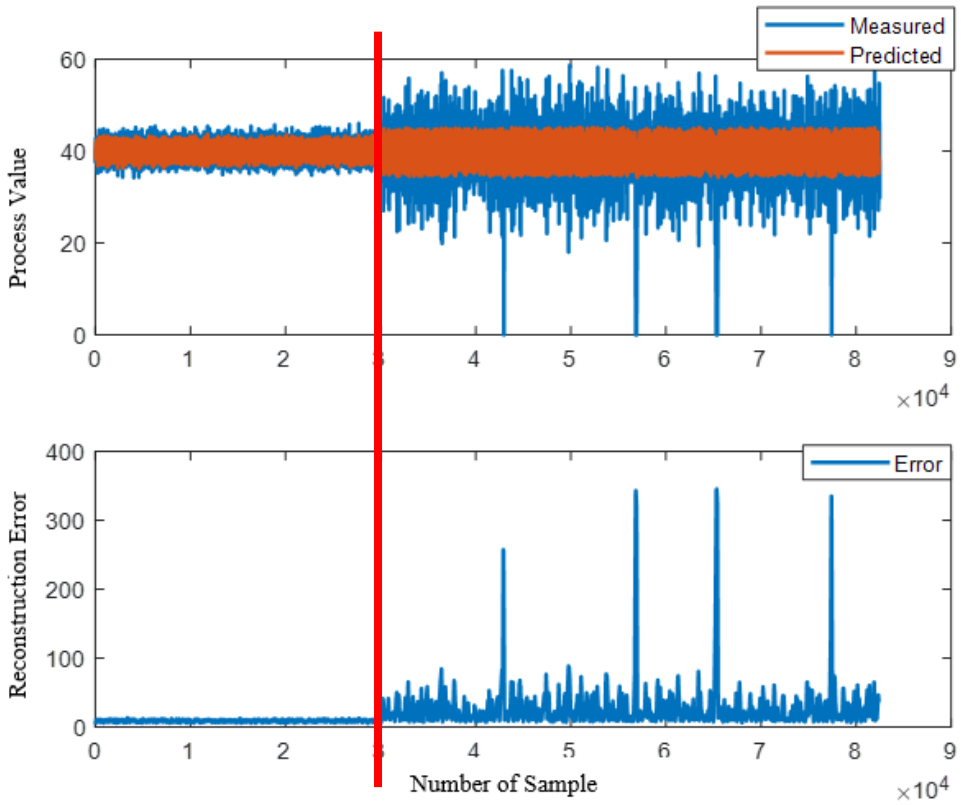

Figure 4.10. Measured and predicted process value (top), Reconstruction error (bottom) for TEP benchmark

4.2.3 EWMA

EWMA is a statistical method, unlike DWT and AE. EWMA method is generally used to detect small changes in the process (Ye, Borror, and Zhang 2002). Data is

separated as training and test datasets similar to AE. Residuals were obtained by using MLR.

The variables which had the largest impact on the IDV 12 disturbance are selected using PCA. At the same time, as mentioned above (Section 3.1.7), the value of valve opening is predicted with MLR by taking the effect of other variables into account, in order to create a similar point for the TUPRAS case. From the data that is separated as training and test data set, the training set is used for MLR modeling, and the test data is used to calculate residual of the model. Upper and lower control limits are calculated. The data set outside the control limits was defined as anomaly. UCL, UCL and LCL are calculated based on the training dataset as 3.5, -3.5 and 0, respectively. Firstly, θ was calculated with ARIMA and λ is calculated as 0.96. EWMA of the residuals are shown in Figure 4.11. Control limits and center line are shown with colored lines. The red dots over the UCL and LCL lines are defined as anomaly. As seen, EWMA can detect the anomalies in TEP benchmark data set.

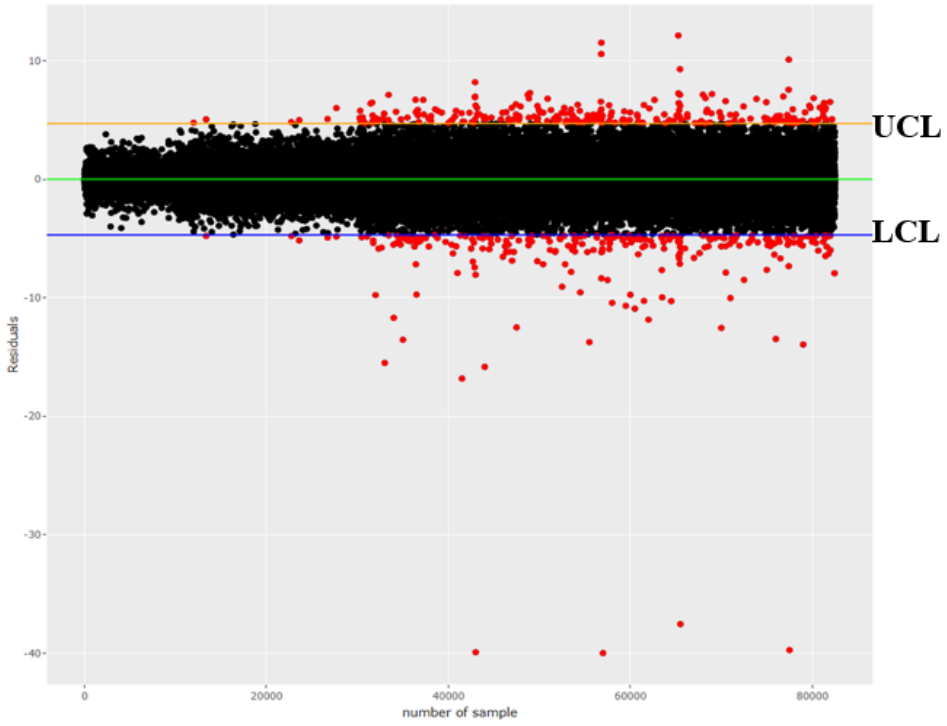

Figure 4.11. EWMA of the residual for TEP benchmark (UCL=3.5, LCL=-3.5, λ =0.96)

CHAPTER 5

RESULTS AND DISCUSSION

5.1. General Overview of Results and Discussion Section

In this thesis, detection of leaks in the heat exchanger is studied using data-based methods. This chapter includes the studies on the heat exchanger in TUPRAS Izmit Refinery, where more than one case was experienced. Three different data-based methods are applied, namely, DWT, AE, and EWMA. Data is taken in per minute basis from the TUPRAS Historian Database. For DWT, the data is not divided into separate training and test data sets and is used in a continuous manner. Data is divided into training and test data sets to apply AE and EWMA methods. To determine the relevant variables in the process, technical support is received from operation engineers. The relationship between the variables provided by the operation engineers is understood and presented in section 5.2. In Section 5.3, the results of the DWT, AE and EWMA methods applied for the first leak case are given. In Section 5.4, the results of two different leakage cases experienced in the same year are provided. In Section 5.5, the results of the last case are represented. In summary, in Section 5.6, all the results of each case are interpreted.

5.2. PCA

First, PCA is carried out in order to examine the correlation between the variables. PCA is applied to the four-year data set containing all the leak cases. As stated in the method section 3.2.2, the variables that may be relevant are determined by the operation engineers. In addition, the presence of leak in each case was understood by increase in the percentage of valve opening. PCA method is applied to understand the relation of valve opening with other variables. Valve opening can be accepted as an indicator for the leak because the response of the leak on the valve opening is higher than LMTD as can be seen from the Figure 5.1.

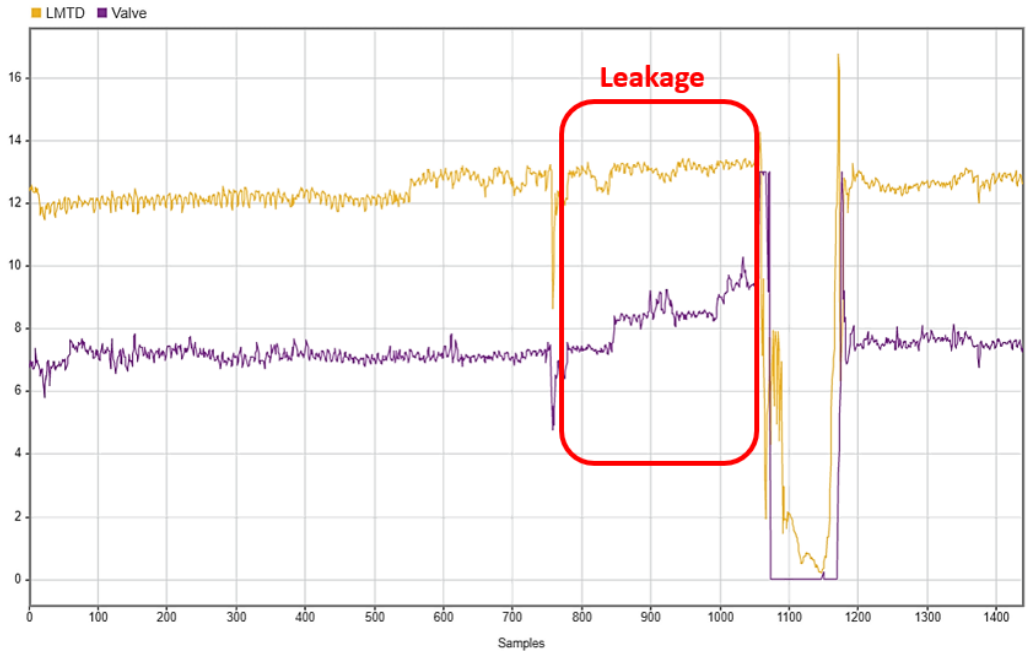

Figure 5.1. Comparison leakage response on LMTD and valve opening

Correlation matrix was formed by following the steps described in section 3.3.1. The obtained correlation matrix is shown in Figure 5.2. In the correlation matrix, the legend shows the strength of the correlation, i.e., the darker the color, the stronger the correlation. While strong correlation is observed between LMTD and tube inlet temperature, high correlation is observed between tube and shell outlet temperatures, LMTD and shell inlet temperature, valve opening and LMTD, and valve opening and shell inlet temperature. In addition, a negative correlation is observed between the shell outlet temperature and the valve opening. This is expected since the leak occurs from the tube to the shell, and in case of leakage, the temperature of the shell fluid will decrease because the temperature of the tube fluid is lower compared to that of shell fluid. Based on the PCA analysis and consultation with operation engineers, the input for the fault analysis using DWT is selected as the valve opening, and the inputs for the fault analysis using AE and EWMA are selected as LMTD along with all other variables listed in Figure 5.2. Compared to TEP benchmark dataset, real industry case data is more complex which makes it more difficult to determine any anomaly. For that reason, the DWT, AE and EWMA methods are applied for each of the leak cases studied. In the following sections, the obtained results are presented.

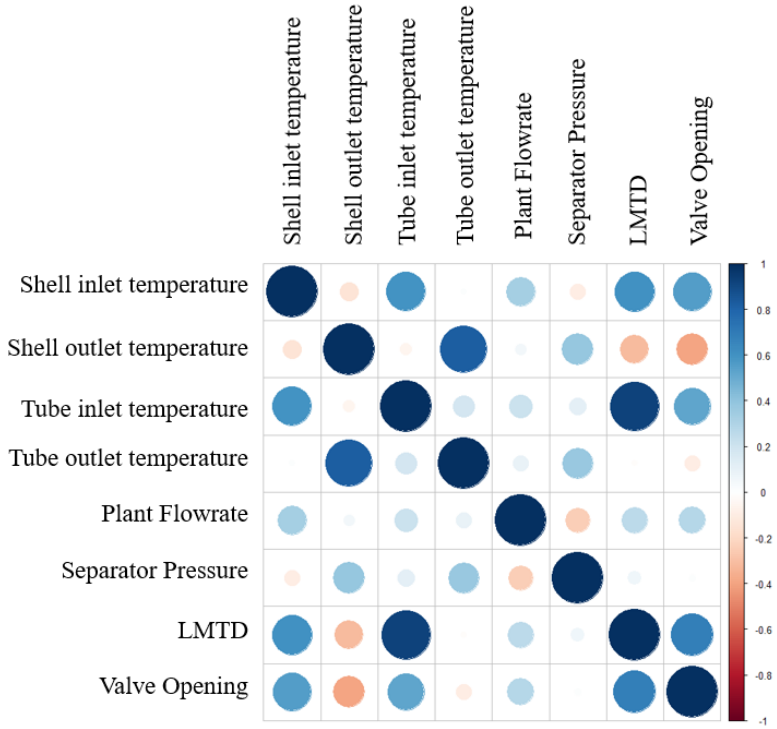

Figure 5.2. Correlation matrix for real case

5.3. CASE 1

5.3.1. DWT

Data was taken from TUPRAS historian database. First, mother wavelet selection was done using metrics such as PSNR, MAE, and MSE values shown in Table 5.1. Db4 was selected suitable wavelet type with the highest PSNR, and lowest MSE and MAE values.

Table 5.1. Performance metrics for wavelet selection (TUPRAS Case 1)

Wavelet Type	MSE	MAE	PSNR	Cross correlation
Haar	1.783	0.879	37.49	0.993
Sym4	1.105	0.736	33.908	0.995
Db4	**0.129**	**0.282**	**48.903**	**0.999**

(cont. on the next page)

Table 5.1 (cont.)

Wavelet Type	MSE	MAE	PSNR	Cross correlation
Db8	1.935	0.929	31.474	0.993
Db12	2.036	0.952	36.912	0.993

Similar comparisons were done for the level selection as shown in Table 5.2. As in TEP benchmark case, level 5 was selected as the level for anomaly detection because this level has higher PSNR value with lower MAE and MSE.

Table 5.2. Performance selection for level selection (TUPRAS Case 1)

Number of Level	MSE	MAE	PSNR	Cross correlation
Level 1	1.779	0.948	37.498	0.993
Level 2	1.517	0.856	38.190	0.994
Level 3	1.168	0.755	39.325	0.996
Level 4	0.626	0.532	42.034	0.998
Level 5	**0.294**	**0.349**	**45.312**	**0.999**
Level 6	0.468	0.450	43.985	0.998
Level 7	0.493	0.586	42.025	0.996

These results are also visually supported by the spectra shown in Figure 5.3. In this figure, the horizontal axis represents the number of data samples, and the vertical axis represents the level (i.e., frequency band; the lower the level, the wider the frequency band and vice versa). The X and Y points indicated in the figure represent the operational change and leakage in the unit, respectively. Operational changes can generally be flowrate changes, and equipment (i.e. valve, pump) replacements. Leakage can be due to corrosion in the tubes of the heat exchanger under investigation, as explained in section 3.1.2. There has also been a flowrate change in the areas marked with green, which can only be observed at level 5 As seen, the method successfully captured operational changes as well as leakage. Color bar below the figure represents the frequency scale. In cases where there is an anomaly, the color gets the corresponding color of the highest value on

the frequency scale. These results obtained here are similar to the results of change point detection in study by Lima et al (de Lima, 2020). They observed the point where the change occurred in the time dependent data set in the highest level which corresponds to lowest frequency band.

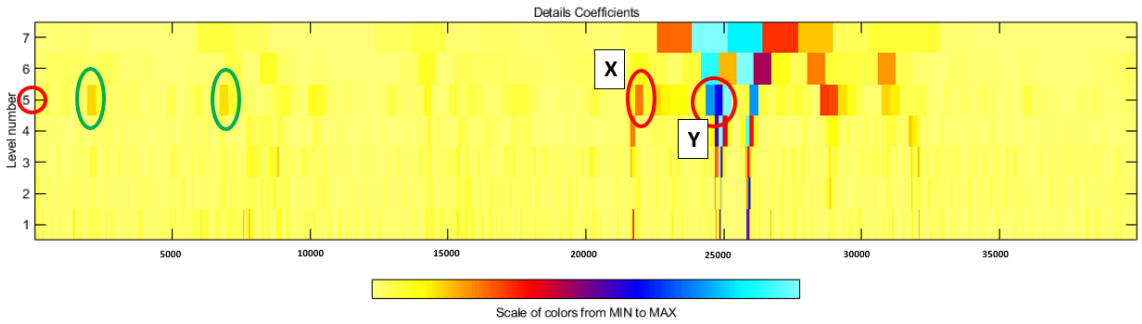

Figure 5.3. Spectra of data based on level number

In Figure 5.3, Level 5 belongs to lowest frequency band. These results can be understood based on the uncertainty principle that given in Equation 39. Here, t is time and ω is the angular frequency ($\omega = 2\pi f$). With the increasing of level, frequency decreases and in turn $\Delta\omega$ decreases. This in turn increases the Δt in Equation 39. The meaning of large Δt is bad time resolution and good frequency resolution (Vošvrda and Schürrer 2015).

$$\Delta t \Delta \omega \geq \frac{1}{2} \qquad \text{Eqn 39.}$$

The data sampling rate from TUPRAS Historian Database was one sample per minute which corresponds to a sampling frequency of 0.017 samples/sec. Table 5.3 shows the frequency bands of each level. Frequency bands belongs to each level in Figure 5.3.

Table 5.3. Frequency band which corresponds to each level

Number of Level	Frequency Band
First detail level	0.0085-0.017 Hz
Second detail level	0.00425-0.0085 Hz
Third detail level	0.00213-0.00425 Hz
Fourth detail level	0.0011-0.00213 Hz
Fifth detail level	0.0005-0.0011 Hz

The 5-level multiple decomposition results are shown in Figure 5.4. Here, s is the original signal, a_5 is the approximation coefficient of the last level and d_i is the detail coefficient at level i (i=1,2,3,4,5). Changes in the nominal values of the process were investigated at different levels. Level 5 corresponds to the low frequency band. At this level, sharp peaks showing operational changes and leakage are successfully observed.

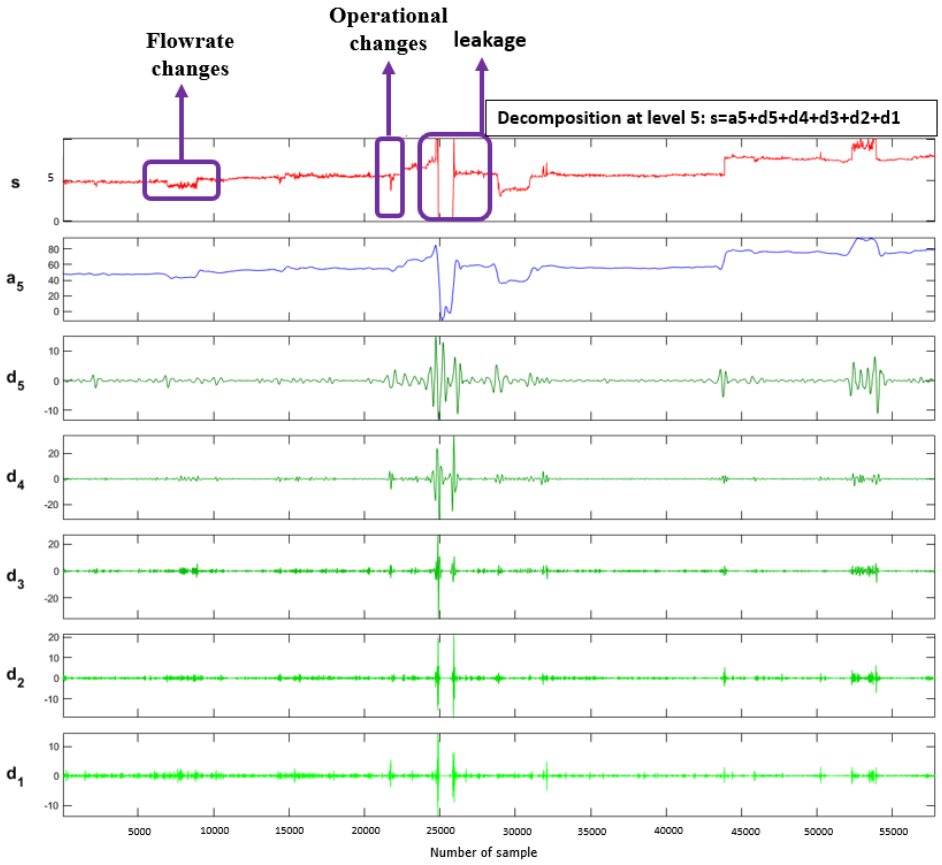

Figure 5.4. 5 level DWT decomposition for Case1

Figure 5.5 gives the original signal and the reconstructed signal comparison. Here, red line represents the original signal and black line represents the reconstructed signal. By comparing the two signals, it can be said that obtained DWT coefficients (detail and approximation) can successfully reconstruct the data.

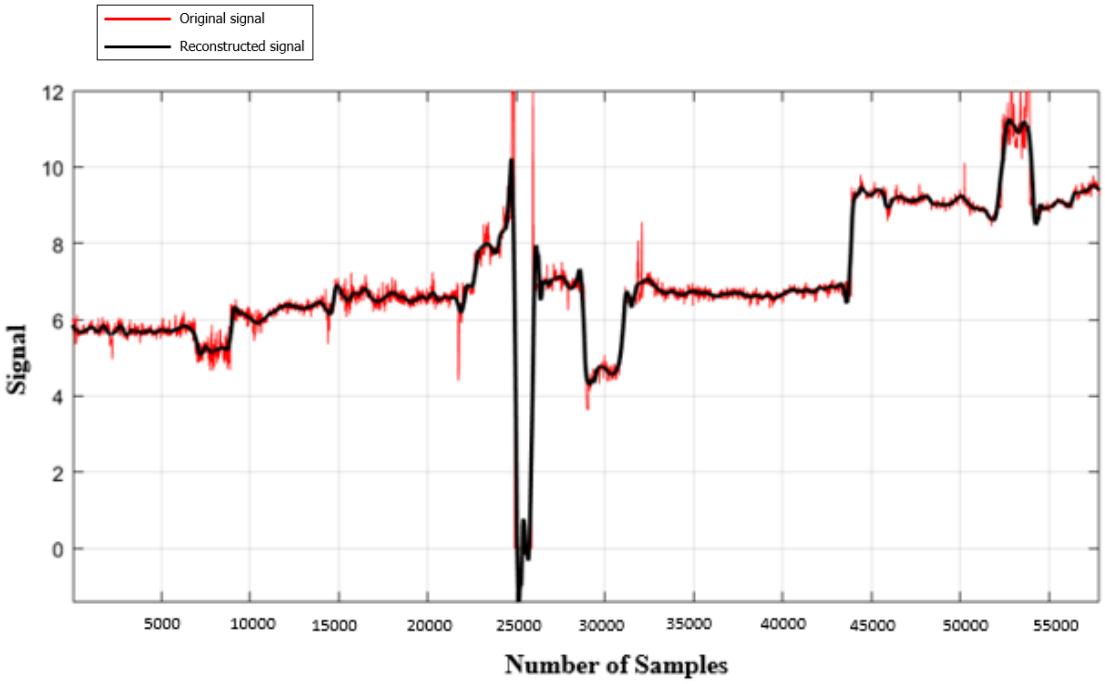

Figure 5.5. Comparison of original and reconstruction signal for Case 1

5.3.2. AE

The leak case is also studied using AE method. It is noted that it is more difficult to determine the training and test datasets in the real case, unlike the TEP benchmark dataset. The data in first two months of a year on minute basis is determined as the training set for the leak that took place in the fourth month of the year. For determining the data set, operation engineers are consulted. While the first two-month data is determined as the training set, the data until the shutdown is determined as the test set.

First, behavior of the MSE with respect to epoch number is obtained as shown in Figure 5.6. The optimum epoch number was determined as 37 based on the MSE value

of 3.76e-06 with MATLAB. After a satisfactory level of error is obtained, MATLAB automatically stops the iteration in order to prevent overfitting.

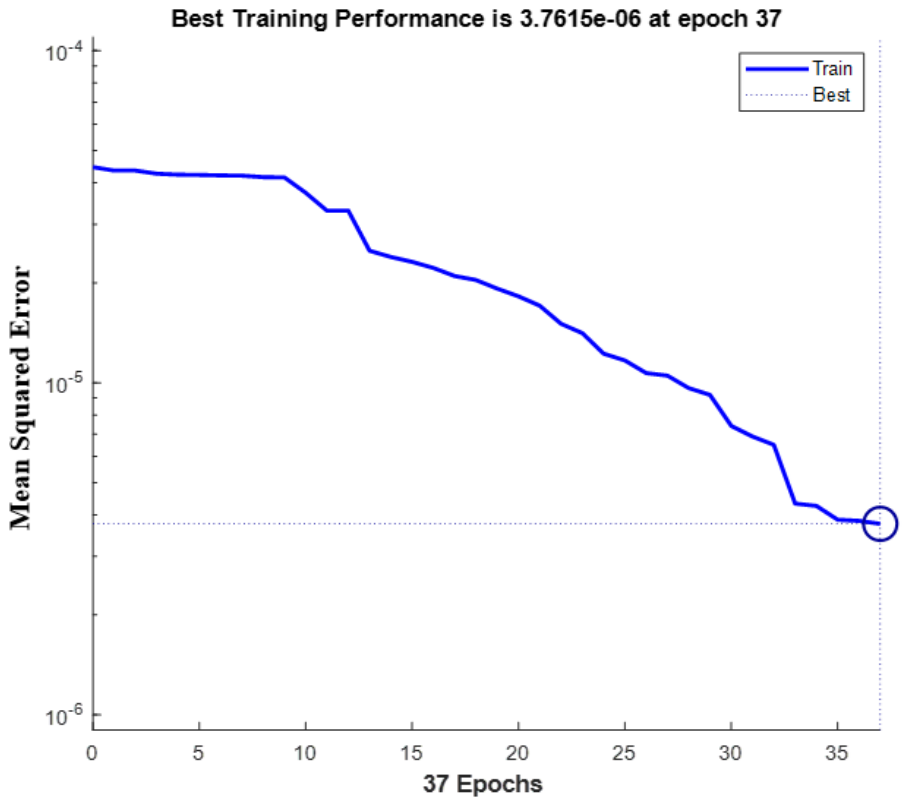

Figure 5.6. Learning curve of AE for Case 1

Then, the test set is fed to the model as an input. The obtained model results are compared to the model values, as shown in Figure 5.7. The estimated value according to the trained data set and the measured process data is given in the top graph, and RMSE values calculated using predicted and measured values are given in the bottom graph. The sudden changes observed in the top graph marked with red circles show the flowrate change and pump replacement. The operation engineers noticed the leak at the time indicated by the red star on the top graph and an immediate shutdown was done. In fact, it is observed that the actual process values deviate from the model predictions and RMSE values increases even before the operation engineers noticed the anomaly. Based on this analysis, it is possible to state that the leak started with a small amount approximately 6 days ago and grew afterwards until the anomaly is noticed by the operation.

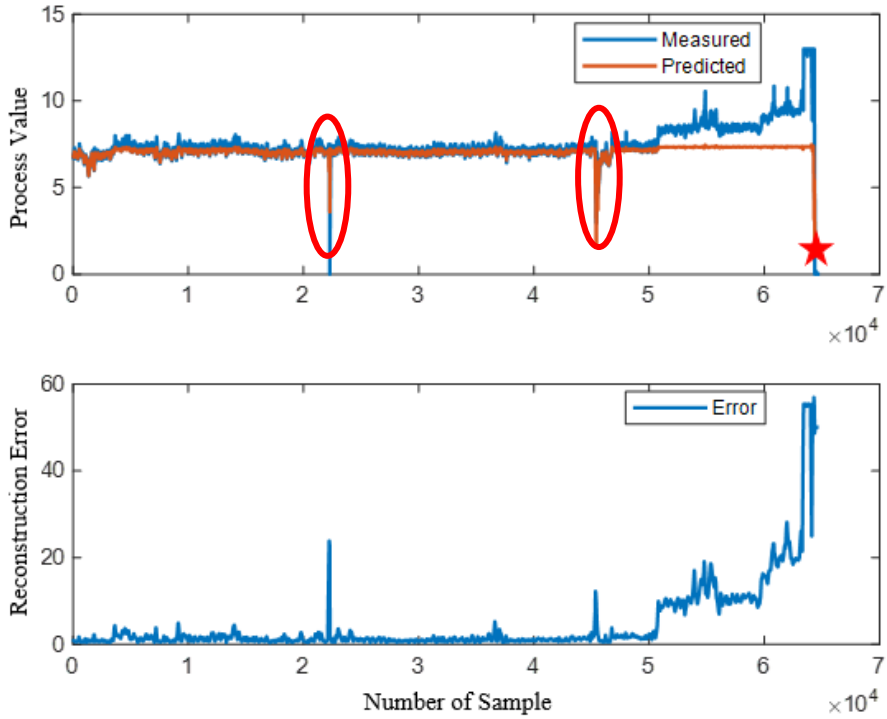

Figure 5.7. Measured and predicted process value (top), Reconstruction error (bottom) for Case 1

5.3.3. EWMA

Next, EWMA method is implemented for the leak case. For the implementation, a two-month data training set is selected, and the model is trained. Residual calculation was made using MLR method. UCL, LCL and CL values are calculated by using Equation 32, Equation 33 and Equation 34 in method section 3.3.4 as 9, -9, 2 according to the training set. λ is calculated as 0.97. λ is close to 1 as expected because weights of last measurement are dominant. The obtained residual versus time graph is shown in Figure 5.8. The region between the UCL and LCL is the control region. Data which is out of the control region give us the information about existence of the anomaly in the system. The moment when the leak is first noticed by operation engineers is shown on the figure with a blue star. The leak probably started earlier than the operation engineers noticed it as the purple rectangle in Figure 5.8 shows deviation out of the control region. It is seen that EWMA can successfully indicate the leak in the system. In fact, with this method, a leak-related early warning mechanism can be possible to implement.

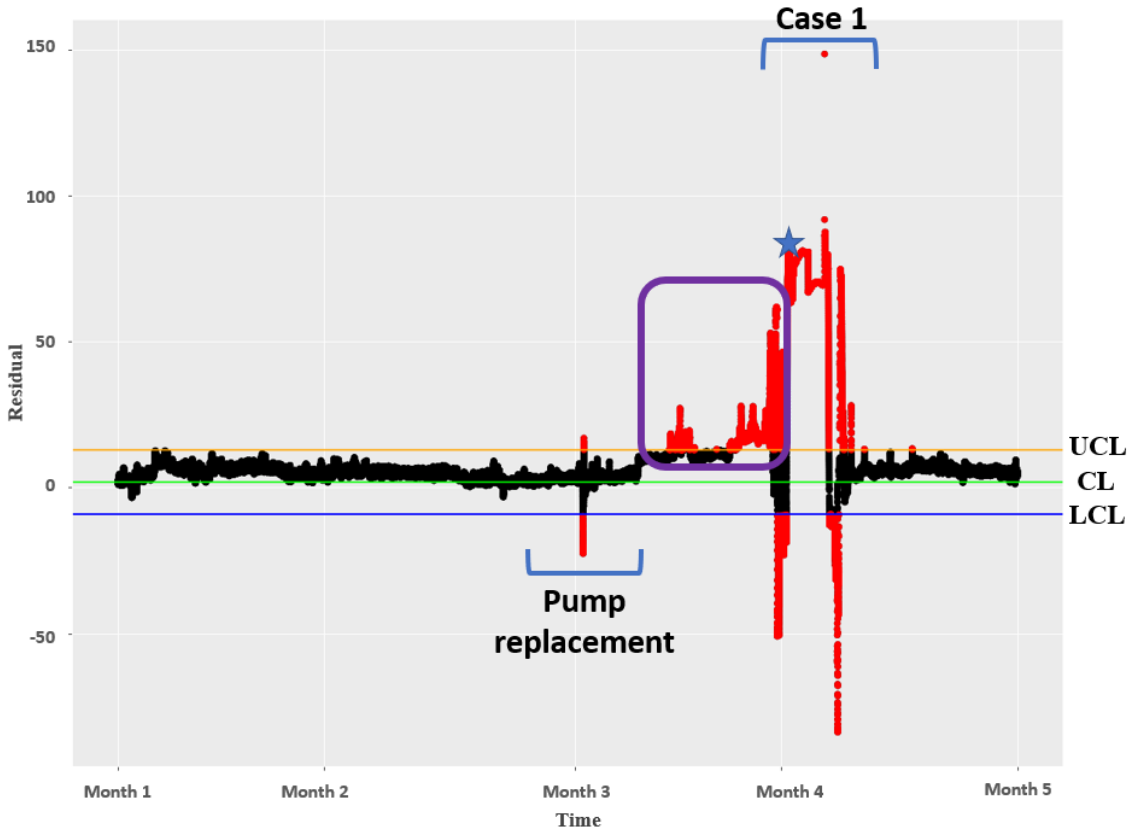

Figure 5.8. EWMA of the residual for Case 1(UCL=9, LCL=-9, CL=2, λ =0.97)

5.4. CASE 2 & CASE 3

Case 2 and Case 3 are two different leak cases that took place in the same year. There is a 3-month timeframe between the two leak cases. Therefore, leaks are tried to be detected using the same training dataset and results are shared in a single graph. As in Case 1, the valve opening is used as the input for DWT, and LTMD along with all other inputs are used as inputs for AE and EWMA. Like other leak cases, the leaks in Case 2 and Case 3 are noticed by the increase in the pressure of the separator (2) and thereby the increased opening of the valve that controls the pressure, as outlined in section 3.2.1. As in Case 1, DWT, AE and EWMA methods are implemented for the Case 2 and 3.

5.4.1. DWT

For the Case 2 and Case 3, evaluation metrics are calculated and shown on Table 5.4. Sym 4 is selected as a mother wavelet with respect to PSNR, MAE and MSE values.

Table 5.4. Performance metrics for wavelet selection (TUPRAS Case 2 & Case 3)

Wavelet Type	**MSE**	**MAE**	**PSNR**	**Cross correlation**
Haar	1.102	0.644	39.576	0.997
Sym4	**0.921**	**0.701**	**40.142**	**0.999**
Db4	1.296	0.707	38.874	0.997
Db8	1.515	0.750	38.196	0.997
Sym8	1.324	0.723	38.779	0.997
Sym3	1.217	0.692	39.147	0.998
Sym2	1.159	0.668	39.359	0.997

The 5-level DWT graph obtained with sym4 mother wavelet is shown in Figure 5.9. The x-axis represents the sample number, while the y-axis represents the detail coefficients, the approximation coefficient at the highest level, and the reconstructed signal (from bottom to top). Pattern changes such as unit current changes, downtimes for short-term maintenance, and leaks are more prominently observed at level 5, as in the DWT result for Case 1. It is seen that the DWT method successfully detects the anomalies of both Case 2 and Case 3 in a single run.

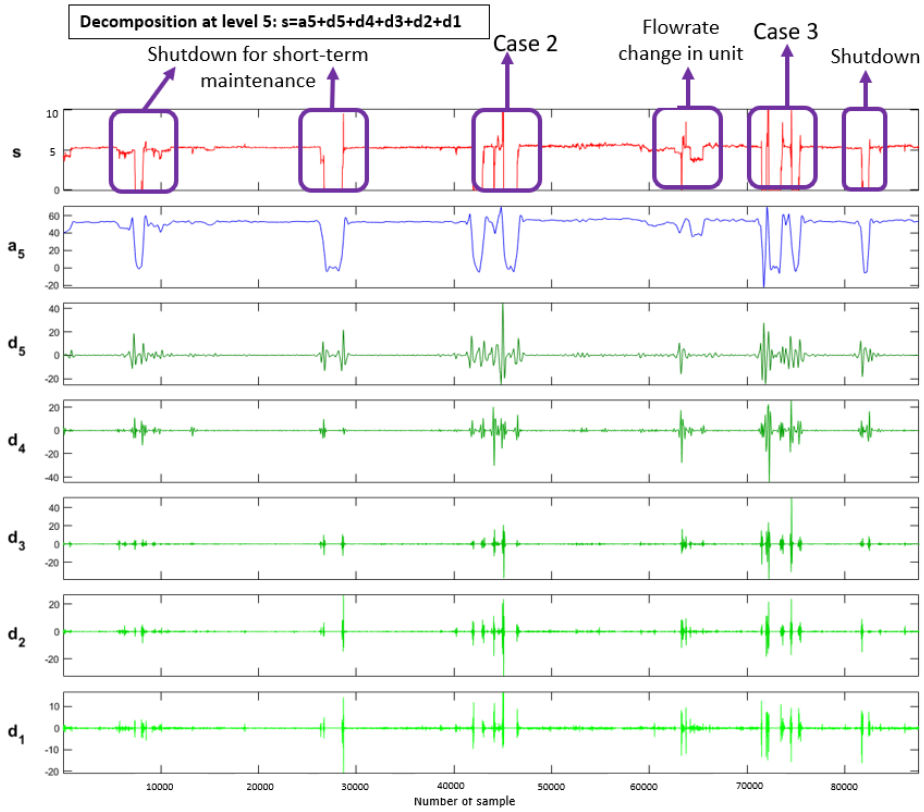

Figure 5.9. 5-level DWT decomposition for Case 2 and 3

5.4.2. AE

It was difficult to choose a training data set, as there were too many shutdowns and flowrate changes during the year. The normal operation values of the unit are decided upon consultation with the operation engineers and the model was trained accordingly. The MSE value with respect to the epoch number is shown in Figure 5.10 below. The most suitable epoch number for training this model was determined as 500 with the MSE value of 0.0012. This value is quite high compared to Case 1 due to using a dataset with the shorter time interval to train the model for both Case 2 and Case 3.

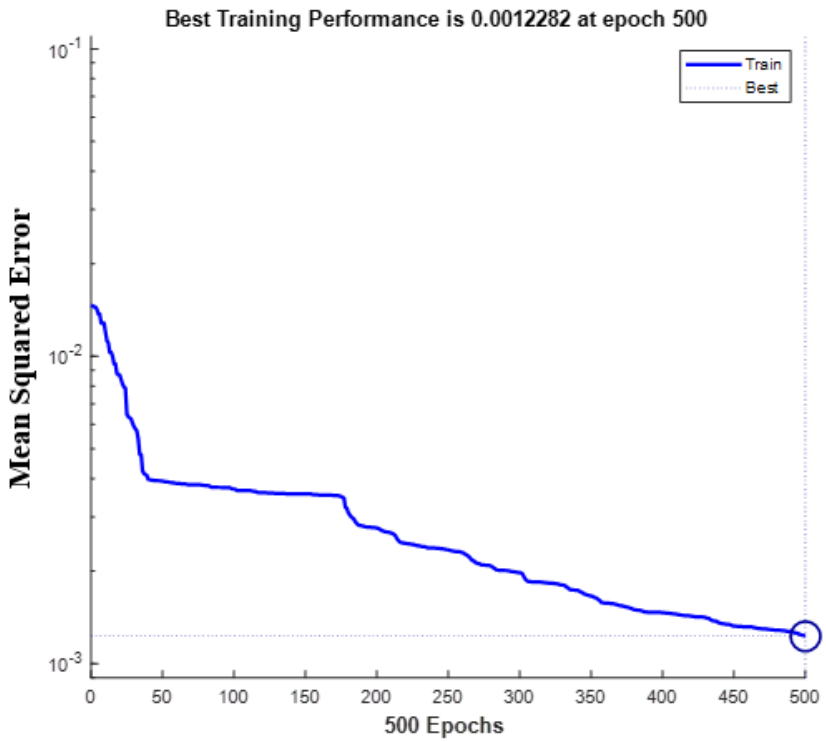

Figure 5.10. Learning curve for Case 2 &3

It was observed that the variables changed a lot because of the operational changes during the year in which the Case 2 and Case 3 took place. This situation is discussed in detail with the operation engineers. While determining the training dataset, we try to focus on more stable time periods where the change in data is small. The AE results are as shown in Figure 5.11. Although the change in the variables is very high during the year, these changes are not as much as the variation in the signal before the leak. As seen in Figure 5.11, it is possible to detect the leaks in both Case 2 and Case 3 in a single run using AE.

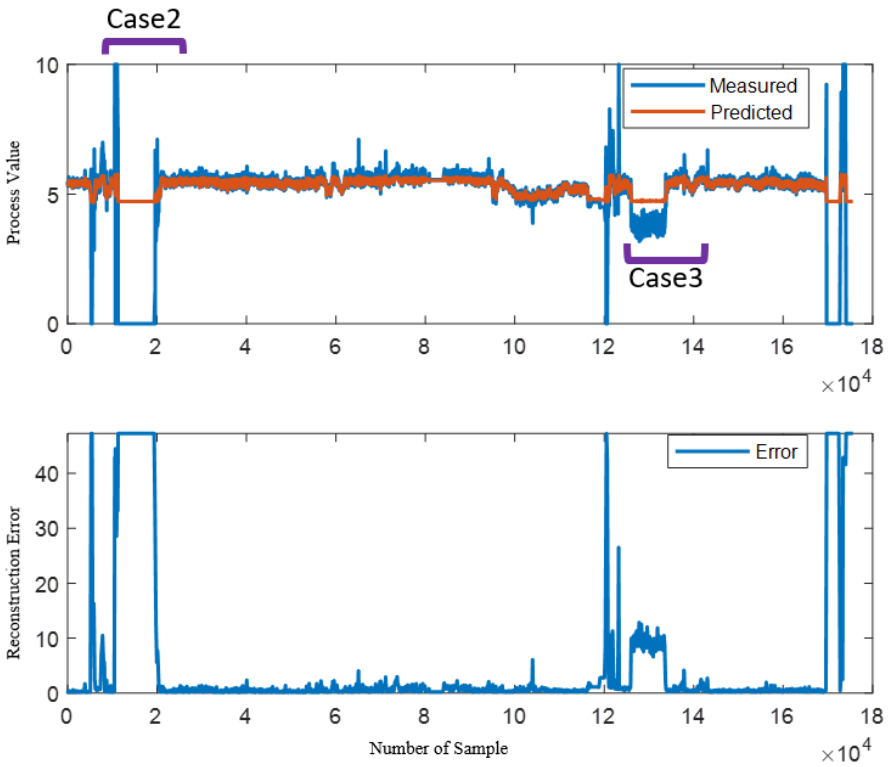

Figure 5.11. Measured and predicted process value (top), Reconstruction error (bottom) for Case 2 & 3

5.4.3. EWMA

Finally, the EWMA method is applied. The model is trained with the selected training set and the results are shown in the Figure 5.12 below. UCL, CL and LCL values were calculated as 7.23, 0 and -7.23, respectively. λ value was determined as 0.968. This value is close to 1 and indicates that the last measurements are dominant. This behavior is similar to Case 1. Accordingly, the dataset we train the model with belongs to the days just before the test dataset. A control region was determined with UCL and LCL lines, and points outside this area are accepted as anomalies or shutdown. It is seen that the leaks in both Case 2 and Case 3 are easily detected using EWMA in a single run.

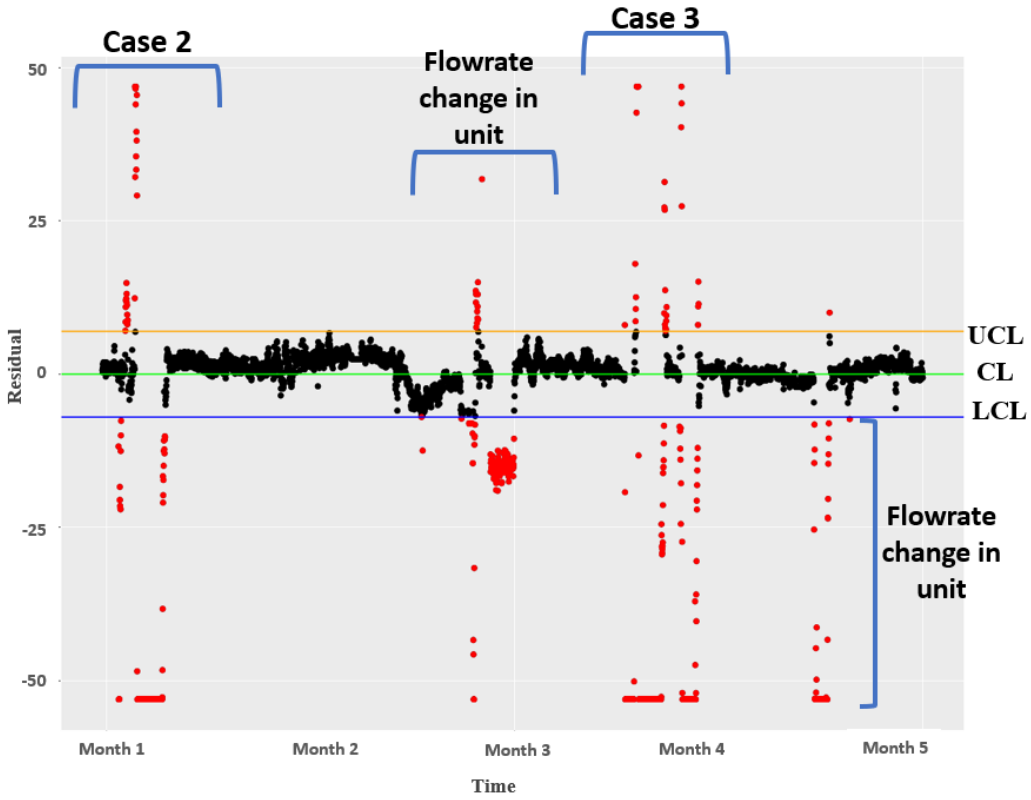

Figure 5.12. EWMA of the residual for case 2&3 (UCL=7.23, LCL=-7.23, CL=0, λ=0.968)

While applying all these methods, datasets corresponding to shutdown dates could be filtered; however, filtering was not done in this study. The valve opening can increase and reach 100% levels both in leakage and shutdown situations. Operation engineers and operators will be able to make the distinction between shutdown and leak cases easily because they already have control over the shutdown.

5.5. CASE 4

Finally, a different leakage case occurred in the same heat exchanger. This leakage was also noticed by the increase in the valve opening that controls the pressure of separator (2), as outlined in section 3.2.1. DWT, AE and EWMA methods that are applied on the above cases are also applied for this case. As in the other cases analyzed so far, the valve opening is used as the input for DWT, and LMTD along with all other inputs are used as inputs for AE and EWMA.

5.5.1. DWT

The DWT results for Case 4 are shown in Figure 5.13 A 5-level decomposition is applied and sym 3 mother wavelet is used. Evaluation metrics are calculated and shown in Table 5.5.

Table 5.5. Performance metrics for wavelet selection (TUPRAS Case 4)

Wavelet Type	**MSE**	**MAE**	**PSNR**	**Cross correlation**
Haar	1.147	0.716	39.406	0.996
Sym4	1.223	0.735	39.127	0.996
Db4	1.241	0.734	39.061	0.995
Db8	1.402	0.761	38.531	0.995
Sym8	1.246	0.738	39.043	0.996
Sym3	**1.210**	**0.731**	**39.423**	**0.997**
Sym2	1.071	0.79	39.924	0.998

Changes in the variables occurred frequently this year. At the same time, the shutdown that occurs in different units also affect the variables in this unit because the unit feed consists of a combination of products from several units. We can detect the leak in Case 4 with the DWT method.

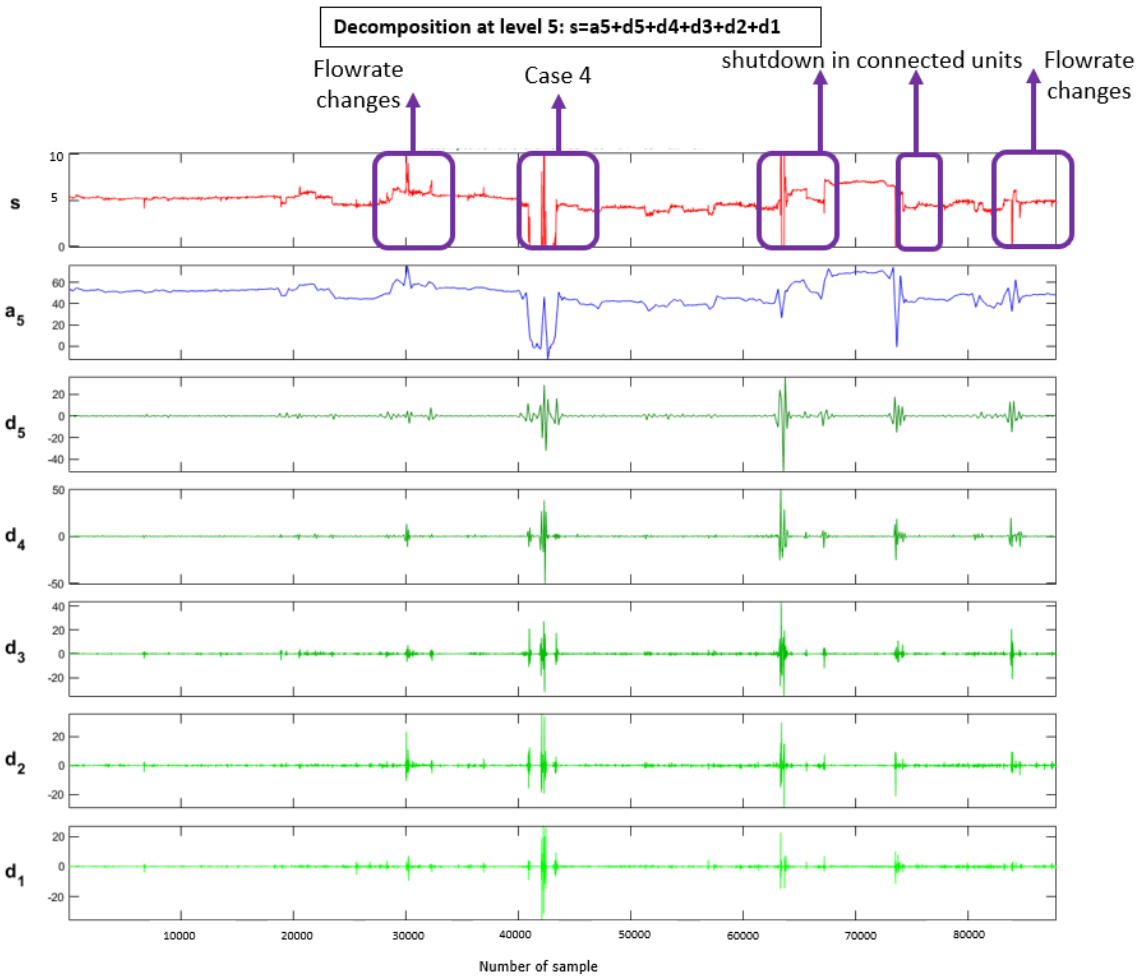

Figure 5.13. 5 level DWT decomposition for Case 4

5.5.2. AE

Due to the complexity of the unit, the training set selection for this case is made together with the operation engineers. The MSE values are shown in the Figure 5.14 below and the epoch number is set to 500 with the MSE value of 0.0045. The AE method has been applied and the resulting plots are shown in the Figure 5.15. The predicted values and the actual measurements are shown in the top, and the reconstruction error graph based on the method result is shown at the bottom. The region on the right marked in red belongs to the leak case. There is no indication of leakage before the leak is noticed and the unit is shutdown by operation engineers. The reason for this behavior might be the volume of leakage being large. Although it may seem difficult to distinguish between leakage and flowrate changes here, it will not be a problem since the flowrate change is already known and controlled by engineers and operators.

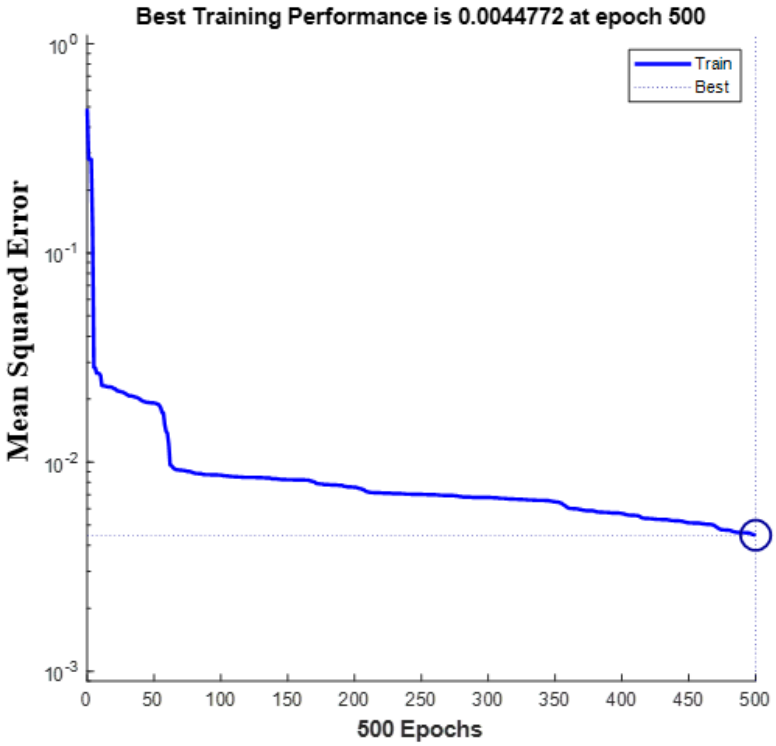

Figure 5.14. Learning curve for Case 4

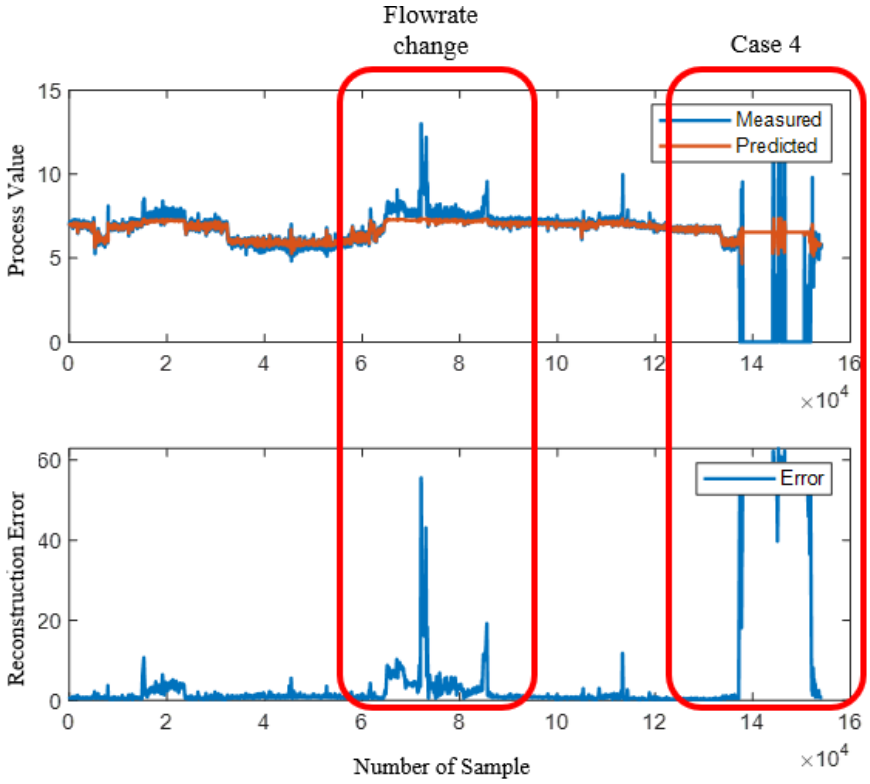

Figure 5.15. AE results for Case 4 (Measured and predicted process value (top), and reconstruction error (bottom))

5.5.3. EWMA

In order to reach a decision using a single method is not generally acceptable. It should be interpreted with the results of other methods. For that reason, EWMA is also implemented on the Case 4. UCL, CL and LCL values are determined as 4.72, 0 and -4.72, respectively. λ is calculated as 0.96. Points outside the control area are designated as anomalies and these are identified as leaks or operational changes and are shown in the Figure 5.16. It is seen that EWMA detects the leak successfully in Case 4.

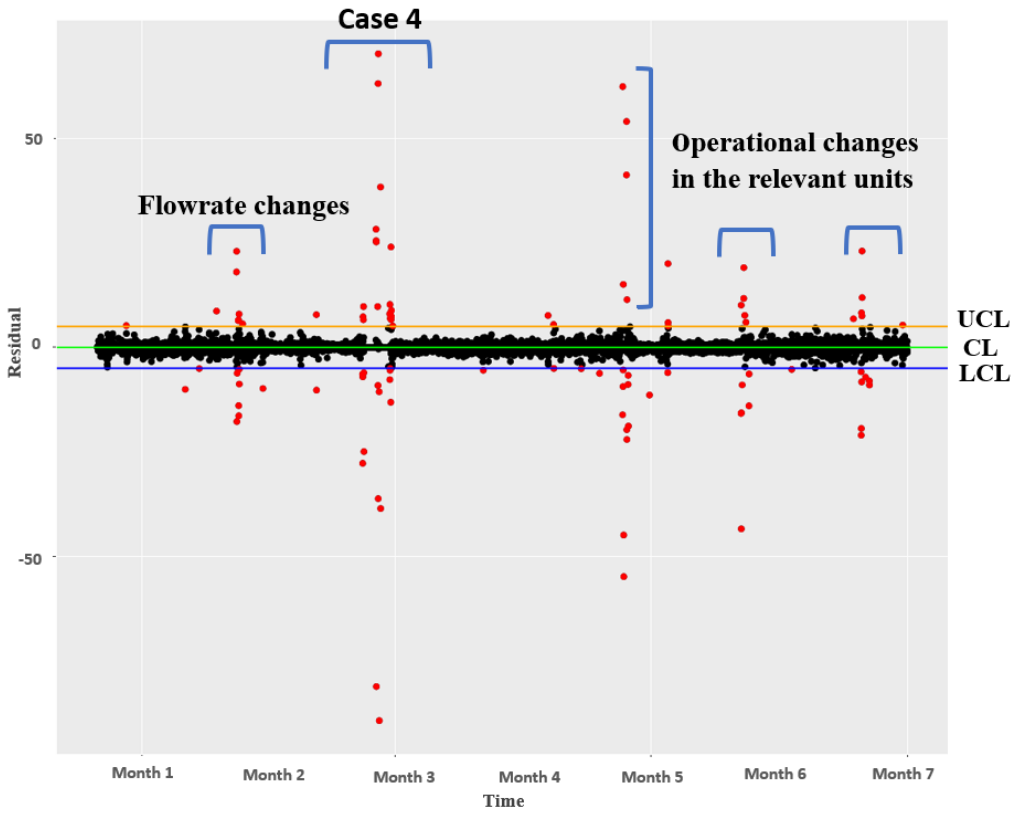

Figure 5.16. EWMA of the residual for Case 4 (UCL=4.72, LCL=-4.72, CL=0, λ =0.96)

5.6. Summary of Results

The methods to be used on the real process data from TUPRAS Historian database are tested on a TEP benchmark dataset known to contain anomalies. Training and test data sets are shared separately on the web for TEP. Two data sets are combined at the

beginning of the study in order to clearly see the difference when there is an anomaly. Since there are too many variables in the TEP benchmark dataset, PCA method is applied to select the variables that are highly associated with the selected anomaly type. DWT, AE and EWMA methods are applied. In the DWT method, the transition between normal and anomaly data sets are observed sharply. The 5th-level is the level where leak is the most obvious. In the AE method, the selection of the training dataset to be used in model training is easy, as the studied TEP dataset had already separate training and test datasets. As expected, the difference between the input data and the reconstructed data started to increase with the onset of the anomaly. Likewise, the EWMA method also indicated the anomaly clearly. Overall, anomaly could be detected with each method applied in TEP benchmark dataset.

The same methods are applied for the leakage cases in TUPRAS. There are four different leakage cases in total. All leaks were noticed by the increase in the valve opening that controls the pressure of the separator (2) located after the heat exchanger and the operation is stopped, as outlined in section 3.1.1. The operator, process and operation engineers who are responsible for the unit do not have information about the start time-date of the leak. Since the unit under investigation is large and complex, the variables that might be related to leakage are decided by the operation engineers and the data regarding for those variables is obtained from TUPRAS Historian database. The PCA method is applied to understand the relationships between the variables suggested by operation engineers. It was seen that there is an opposite relation between valve opening and the shell outlet temperature and positive relation between the valve opening and other variables.

First, Case 1 is studied. For DWT method, evaluation metrics as MAE, MSE and PSNR are considered for selection of mother wavelet. Db4 is selected as the mother wavelet with the highest PSNR and lowest MSE and MAE values. The meaning of high PSNR is a good frequency resolution. Similar to the results obtained with the TEP benchmark, the 5th-level is the level where the anomaly is clearly observed. The changes that can be observed on variables such as leak detection, flowrate change, and equipment replacement can also be detected by the method. In AE, the training dataset is selected in the normal operating range of the unit. The difference between the test data and reconstructed data is examined. An increase in error values is observed with the presence of the anomaly. In the EWMA method, data is predicted using the MLR method and the residual is calculated by taking the difference between the test and prediction data.

Control limits are determined using the training data set, and the points out of these limits are indicated as anomalies. All three methods are found to be able to detect anomalies.

Case 2 and Case 3 are observed in the same year. Therefore, the results are shown on the same graphs. With the DWT method, operational changes and leaks can be detected, but there was no indication of the presence of the leak a few days ago before the leak as opposed to Case 1. While applying the AE and EWMA methods, it is difficult to select the training dataset because lots of flowrate change, shutdown and equipment replacement situations are encountered during the year. Two different leakage cases occurred, and downtimes were experienced due to them. The normal leak-free operating time of the unit is limited. With the AE method, an increase in error values was observed before the moment when both Case 2 and Case 3 leaks are noticed by the operation engineers. In the EWMA method, the detection of the leak can not be noticed beforehand.

For Case 4, first the DWT method is applied. Similar results are obtained in for Case 4 as in other cases. It is difficult to determine the training data set for AE in Case 4 as in Case 2 and Case 3, since there are operational changes in the timeframe of the dataset. Therefore, no leak indications are observed a few days before the leak using the AE and EWMA methods. The reason for these is interpreted as the leakage started with high volume and is immediately noticed by operation engineers and unit is shutdown.

Engineers would prefer to receive alerts when data exceeds a certain threshold. Among these methods implemented in this study, we can say that EWMA and AE methods are more suitable for their requirements. Operational changes in TUPRAS cases are also observed as anomalies in method results. Since these changes are already known and controlled by the operator and operation engineers, they will not be considered as an anomaly.

CHAPTER 6

CONCLUSION

In this thesis, it has been investigated whether data-based methods can detect leaks in heat exchangers. First, the methods used for leak detection in heat exchangers were searched for in the literature. The use of hardware-based methods is common in the literature, and they are usually offline detection systems. The aim of this thesis is to create a data-based online leak detection mechanism with real process data. For this purpose, data-based anomaly detection methods were investigated and DWT, AE and EWMA methods are applied. Before working with real data, methods were validated on the TEP benchmark dataset to detect anomalies. Then, the validated methods are applied on the real process data. Obtained results are shared below.

- Leakage is one of the most common anomalies or faults in the refinery.
- While the leaks occurring on the pipelines can be easily noticed, it is very difficult to detect the leaks on the equipment.
- In the literature, data-based studies for the detection of leaks in heat exchanger are very limited.
- DWT, EWMA and AE methods were able to detect leak both in TEP data and real process data.
- The choice of mother wavelet and level are important when applying the DWT method.
- The selection of training datasets is important when training the model for EWMA and AE methods.
- EWMA and AE methods take into account the values of other variables determined in the unit.
- Working with real data increases complexity. The results obtained with the TEP benchmark dataset are clearer.
- Since the real case is complex, it is safer and more conclusive to implement several different methods simultaneously.

The recommendations for further studies can be listed as follows:

- Different data-based methods can be used.
- The methods can be merged with each other.
- Cross validation can be applied to determine training and test datasets.
- Results can be supported with different heat exchanger data to test a wider window of applicability and accuracy of the methods.

ABBREVIATIONS

EPA	Environmental Protection Agency
VOC	Volatile Organic Compound
FFT	Fast Fourier Transform
SPC	Statistical Process Control
PCA	Principal Component Analysis
RCV	Remote Control Vehicles
IT	Infrared Thermography
NPW	Negative Pressure Wave
PPA	Pressure Point Analysis
PCA	Principle Component Analysis
PC	Principle Component
CL	Control Limit
UCL	Upper Control Limit
LCL	Lower Control Limit
IUPU	Integrated Unicracking Processing Unit
CN	Coker Naphtha
HCGO	Heavy Coker Gas Oil
LCGO	Light Coker Gas Oil
HVGO	Heavy Vacuum Gas Oil
LVGO	Light Vacuum Gas Oil
HCU	Hydrocracker Unit
HC	Hydrocarbon
NHT	Naphtha Hydro-Treater
DHT	Diesel Hydro-Treater
LMTD	Log Mean Temperature Difference
PSNR	Peak Signal to Noise Ratio
MAE	Mean Absolute Error
MSE	Mean Square Error
TEP	Tennessee Eastman Process

PV	Process Variable
MV	Manipulated Variable
IDV	Disturbance Variable
EWMA	Exponentially Weighted Moving Average
Db	Daubechies
Sym	Symlet
ReLU	Rectified Linear Unit
EMA	Exponential Moving Average
FCU	Fan Coil Units
RELAP	Reactor Excursion and Leak Analysis Program
MLR	Multiple Linear Regression

REFERENCES

Adedeji, Kazeem B., Yskandar Hamam, Bolanle Tolulope Abe, and Adnan M. Abu-Mahfouz. 2017. “Towards Achieving a Reliable Leakage Detection and Localization Algorithm for Application in Water Piping Networks: An Overview.” *IEEE Access* 5: 20272–85. https://doi.org/10.1109/ACCESS.2017.2752802.

Adegboye, Mutiu Adesina, Wai Keung Fung, and Aditya Karnik. 2019. “Recent Advances in Pipeline Monitoring and Oil Leakage Detection Technologies: Principles and Approaches.” *Sensors (Switzerland)* 19 (11): 1–32. https://doi.org/10.3390/s19112548.

Ahmed, Mohiuddin, Abdun Naser Mahmood, and Jiankun Hu. 2016. “A Survey of Network Anomaly Detection Techniques.” *Journal of Network and Computer Applications* 60: 19–31. https://doi.org/10.1016/j.jnca.2015.11.016.

Ahsan, Muhammad, Muhammad Mashuri, Heri Kuswanto, and Dedy Dwi Prastyo. 2018. “Intrusion Detection System Using Multivariate Control Chart Hotelling’s T2 Based on PCA.” *International Journal on Advanced Science, Engineering and Information Technology* 8 (5): 1905–11. https://doi.org/10.18517/ijaseit.8.5.3421.

Ahsan, Muhammad, Muhammad Mashuri, Heri Kuswanto, Dedy Dwi Prastyo, and Hidayatul Khusna. 2018. “Multivariate Control Chart Based on PCA Mix for Variable and Attribute Quality Characteristics.” *Production and Manufacturing Research* 6 (1): 364–84. https://doi.org/10.1080/21693277.2018.1517055.

Amolins, Krista, Yun Zhang, and Peter Dare. 2007. “Wavelet Based Image Fusion Techniques - An Introduction, Review and Comparison.” *ISPRS Journal of Photogrammetry and Remote Sensing* 62 (4): 249–63. https://doi.org/10.1016/j.isprsjprs.2007.05.009.

Barros, Julio, Ramón I. Diego, and Matilde De Apraiz. 2012. “Applications of Wavelet Transform for Analysis of Harmonic Distortion in Power Systems: A Review.” *IEEE Transactions on Instrumentation and Measurement* 61 (10): 2604–11. https://doi.org/10.1109/TIM.2012.2199194.

Bolotina, Irina, Valeriy Borikov, Veronica Ivanova, Kseniya Mertins, and Sergey

Uchaikin. 2017. “Application of Phased Antenna Arrays for Pipeline Leak Detection.” *Journal of Petroleum Science and Engineering* 161: 497–505. https://doi.org/10.1016/j.petrol.2017.10.059.

Chen, Zhaomin, Chai Kiat Yeo, Bu Sung Lee, and Chiew Tong Lau. 2018. “Autoencoder-Based Network Anomaly Detection.” *Wireless Telecommunications Symposium* 2018-April: 1–5. https://doi.org/10.1109/WTS.2018.8363930.

Clavijo, Nayher, Afrânio Melo, Maurício M. Câmara, Thiago Feital, Thiago K. Anzai, Fabio C. Diehl, Pedro H. Thompson, and José Carlos Pinto. 2019. “Development and Application of a Data-Driven System for Sensor Fault Diagnosis in an Oil Processing Plant.” *Processes* 7 (7). https://doi.org/10.3390/pr7070436.

Clover, David, Advanced Sealing, Supply Company, David Reeves, and National Petrochemical. 2010. “All Sealing Problems Can Be Fixed Case 1 : Heat Exchanger Leakage,” 1–15.

Cureton, Edward E., and Ralph B. D’Agostino. 2019. “Component Analysis.” *Factor Analysis*, no. April: 296–338. https://doi.org/10.4324/9781315799476-12.

Elbattah, Mahmoud, Colm Loughnane, Jean Luc Guérin, Romuald Carette, Federica Cilia, and Gilles Dequen. 2021. “Variational Autoencoder for Image-Based Augmentation of Eye-Tracking Data.” *Journal of Imaging* 7 (5). https://doi.org/10.3390/jimaging7050083.

Emmanuel, Babatunde S. 2012. “Discrete Wavelet Mathematical Transformation Method for Non-Stationary Heart Sounds Signal Analysis.” *ARPN Journal of Engineering and Applied Sciences* 7 (8): 1021–28.

Faust, Oliver, U. Rajendra Acharya, Hojjat Adeli, and Amir Adeli. 2015. “Wavelet-Based EEG Processing for Computer-Aided Seizure Detection and Epilepsy Diagnosis.” *Seizure* 26: 56–64. https://doi.org/10.1016/j.seizure.2015.01.012.

Galya Georgieva-Tsaneva, Krassimir Tcheshmedjiev. 2014. “Denoising Of Electrocardiogram Data With Wavelet Transform & Thresholding.” *International Journal of Scientific & Engineering Research* 5 (June): 9–16. http://www.ijser.org.

Gianluca Manca. n.d. “‘Tennessee-Eastman-Process’ Alarm Management Dataset-Technical Report.”

Guillen, D. P., N. Anderson, C. Krome, R. Boza, L. M. Griffel, J. Zouabe, and A. Y. Al Rashdan. 2020. “A RELAP5-3D/LSTM Model for the Analysis of Drywell Cooling Fan Failure.” *Progress in Nuclear Energy* 130 (October): 103540. https://doi.org/10.1016/j.pnucene.2020.103540.

Habbi, Hacene, Michel Kinnaert, and Mimoun Zelmat. 2009. “A Complete Procedure for Leak Detection and Diagnosis in a Complex Heat Exchanger Using Data-Driven Fuzzy Models.” *ISA Transactions* 48 (3): 354–61. https://doi.org/10.1016/j.isatra.2009.01.004.

Horé, Alain, and Djemel Ziou. 2010. “Image Quality Metrics: PSNR vs. SSIM.” *Proceedings - International Conference on Pattern Recognition*, 2366–69. https://doi.org/10.1109/ICPR.2010.579.

Huang, Ling, Xuan Long Nguyen, Minos Garofalakis, Michael I. Jordan, Anthony Joseph, and Nina Taft. 2007. “In-Network PCA and Anomaly Detection.” *Advances in Neural Information Processing Systems*, 617–24. https://doi.org/10.7551/mitpress/7503.003.0082.

Jiang, Dongxiang, and Chao Liu. 2011. “Machine Condition Classification Using Deterioration Feature Extraction and Anomaly Determination.” *IEEE Transactions on Reliability* 60 (1): 41–48. https://doi.org/10.1109/TR.2011.2104433.

Jollife, Ian T., and Jorge Cadima. 2016. “Principal Component Analysis: A Review and Recent Developments.” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374 (2065). https://doi.org/10.1098/rsta.2015.0202.

Kandananond, Karin. 2014. “Guidelines for Applying Statistical Quality Control Method to Monitor Autocorrelated Prcoesses.” *Procedia Engineering* 69: 1449–58. https://doi.org/10.1016/j.proeng.2014.03.141.

Kiss, István, Béla Genge, and Piroska Haller. 2015. “Behavior-Based Critical Cyber Asset Identification in Process Control Systems under Cyber Attacks.” *Proceedings of the 2015 16th International Carpathian Control Conference, ICCC 2015*, no. June: 196–201. https://doi.org/10.1109/CarpathianCC.2015.7145073.

Kricha, Zied, Anis Kricha, and Anis Sakly. 2018. “A Robust Watermarking Scheme Based on the Mean Modulation of DWT Coefficients.” *Security and*

Communication Networks 2018. https://doi.org/10.1155/2018/1254081.

Kumaresan, S. Prabha, Chee Keong Tan, Yin Hoe Ng, and Chee Keong Tan. 2021. "Extreme Learning Machine (ELM) for Fast User Clustering in Downlink Non-Orthogonal Multiple Access (NOMA) 5G Networks." *IEEE Access* 9: 130884–94. https://doi.org/10.1109/ACCESS.2021.3114619.

Kundnaney, Nikhil Deepak, and Deepak Kumar Kushwaha. 2015. "A Critical Review on Heat Exchangers Used in Oil Refinery." *Afro-Asian International Conference on Science, Engineering & Technology*, no. March 2015: 1–5. https://www.researchgate.net/publication/290437509_A_Critical_Review_on_Heat_Exchangers_used_in_Oil_Refinery.

Laaksonen, H. 2013. "Novel Wavelet Transform Based Islanding Detection Algorithms." *International Review of Electrical Engineering* 8 (6): 1796–1805.

Miljković, Dubravko. 2011. "Fault Detection Methods: A Literature Survey." *MIPRO 2011 - 34th International Convention on Information and Communication Technology, Electronics and Microelectronics - Proceedings*, 750–55.

Mirsky, Yisroel, Tomer Doitshman, Yuval Elovici, and Asaf Shabtai. 2018. "Kitsune: An Ensemble of Autoencoders for Online Network Intrusion Detection," no. February: 18–21. https://doi.org/10.14722/ndss.2018.23204.

Mujtaba, Syed Muhammad, Tamiru Alemu Lemma, Syed Ali Ammar Taqvi, Titus Ntow Ofei, and Seshu Kumar Vandrangi. 2020. "Leak Detection in Gas Mixture Pipelines under Transient Conditions Using Hammerstein Model and Adaptive Thresholds." *Processes* 8 (4). https://doi.org/10.3390/PR8040474.

Nikles, Marc, Bernhard H. Vogel, Fabien Briffod, Stephan Grosswig, Florian Sauser, Steffen Luebbecke, Andre Bals, and Thomas Pfeiffer. 2004. "Leakage Detection Using Fiber Optics Distributed Temperature Monitoring." *Smart Structures and Materials 2004: Smart Sensor Technology and Measurement Systems* 5384 (May 2014): 18. https://doi.org/10.1117/12.540270.

Odusina, Elijah. 2008. "Pipeline Leak Presented."

Ohtani, Tetsuya. 2020. "Application of AI to Oil Refineries and Petrochemical Plants." *Yokogawa Technical Report English Edition* 63 (1): 7–10.

Orea-flores, Izlian Y, Francisco J Gallegos-funes, and Alfonso Arellano-reynoso. 2019. "In Wavelet Domain for Magnetic Resonance."

Panday, Rupen, Natarianto Indrawan, Lawrence J. Shadle, and Richard W. Vesel. 2021. "Leak Detection in a Subcritical Boiler." *Applied Thermal Engineering* 185 (October 2020): 116371. https://doi.org/10.1016/j.applthermaleng.2020.116371.

Park, Pangun, Piergiuseppe Di Marco, Hyejeon Shin, and Junseong Bang. 2019. "Fault Detection and Diagnosis Using Combined Autoencoder and Long Short-Term Memory Network." *Sensors (Switzerland)* 19 (21): 1–17. https://doi.org/10.3390/s19214612.

Payan, Frédéric, and Marc Antonini. 2006. "Mean Square Error Approximation for Wavelet-Based Semiregular Mesh Compression." *IEEE Transactions on Visualization and Computer Graphics* 12 (4): 649–56. https://doi.org/10.1109/TVCG.2006.73.

Penner, Eric, Josh Stephens, Elijah Odusina, James Akingbola, David Mannel, and Miguel Bagajewicz. n.d. "Economic Comparison of a Simulator-Based GLR Method for Pipeline Leak Detection with Other Methods," 1–29. https://pdfs.semanticscholar.org/df2d/f9185d95dcfc990649d068af29b4d14ba75b.pdf.

Perera, N, a D Rajapakse, and R P Jayasinghe. 2007. "On-Line Discrete Wavelet Transform in EMTP Environment and Applications in Protection Relaying." *Time*, no. 1.

Pipelines, Underground, and Using Radiotracers. 2009. "Leak Detection in Heat Exchangers and Underground Pipelines Using Radiotracers."

Reinartz, Christopher, Murat Kulahci, and Ole Ravn. 2021. "An Extended Tennessee Eastman Simulation Dataset for Fault-Detection and Decision Support Systems." *Computers and Chemical Engineering* 149 (June). https://doi.org/10.1016/j.compchemeng.2021.107281.

Ricker, N. Lawrence. 1996. "Decentralized Control of the Tennessee Eastman Challenge Process." *Journal of Process Control* 6 (4): 205–21. https://doi.org/10.1016/0959-1524(96)00031-5.

Rosenfeld, P.E, and L.G.H Feng. 2011. “The Petroleum Industry. Risks of Hazardous Wastes,” 57–71. https://doi.org/10.1016/B978-1-4377-7842-7.00005-2.

Sagheer, Alaa, and Mostafa Kotb. 2019. “Unsupervised Pre-Training of a Deep LSTM-Based Stacked Autoencoder for Multivariate Time Series Forecasting Problems.” *Scientific Reports* 9 (1): 1–17. https://doi.org/10.1038/s41598-019-55320-6.

Sakurada, Mayu, and Takehisa Yairi. 2014. “Anomaly Detection Using Autoencoders with Nonlinear Dimensionality Reduction.” *ACM International Conference Proceeding Series* 02-Decembe: 4–11. https://doi.org/10.1145/2689746.2689747.

Shah, R. K. 1983. “Classification of Heat Exchangers.,” no. (eds.), Washington, U.S.A., Hemisphere Publishing Corp., 1983, pp.9-14. (ISBN 0-89116-254-2): 1–77. https://doi.org/10.1201/9780429469862-1.

Sheltami, Tarek R., Abubakar Bala, and Elhadi M. Shakshuki. 2016. “Wireless Sensor Networks for Leak Detection in Pipelines: A Survey.” *Journal of Ambient Intelligence and Humanized Computing* 7 (3): 347–56. https://doi.org/10.1007/s12652-016-0362-7.

Souza, A. L., S. L. Cruz, and J. F.R. Pereira. 2000. “Leak Detection in Pipelines through Spectral Analysis of Pressure Signals.” *Brazilian Journal of Chemical Engineering* 17 (4): 557–63. https://doi.org/10.1590/s0104-66322000000400020.

Sublime, Jérémie, and Ekaterina Kalinicheva. 2019. “Automatic Post-Disaster Damage Mapping Using Deep-Learning Techniques for Change Detection: Case Study of the Tohoku Tsunami.” *Remote Sensing* 11 (9). https://doi.org/10.3390/rs11091123.

Tavakoli, Neda, Sima Siami-Namini, Mahdi Adl Khanghah, Fahimeh Mirza Soltani, and Akbar Siami Namin. 2020. “An Autoencoder-Based Deep Learning Approach for Clustering Time Series Data.” *SN Applied Sciences* 2 (5): 1–25. https://doi.org/10.1007/s42452-020-2584-8.

Theodoridis, Sergios. 2020. *Neural Networks and Deep Learning*. *Machine Learning*. https://doi.org/10.1016/b978-0-12-818803-3.00030-1.

Tutkan, Melike, Murat Can Ganiz, and Selim Akyokuş. 2016. “Helmholtz Principle Based Supervised and Unsupervised Feature Selection Methods for Text Mining.” *Information Processing and Management* 52 (5): 885–910.

https://doi.org/10.1016/j.ipm.2016.03.007.

Utamura, Motoaki, Konstantin Nikitin, and Yasuyoshi Kato. 2008. “A Generalised Mean Temperature Difference Method for Thermal Design of Heat Exchangers.” *International Journal of Nuclear Energy Science and Technology* 4 (1): 11–31. https://doi.org/10.1504/IJNEST.2008.017545.

Vošvrda, Miloslav, and Jaroslav Schürrer. 2015. “Wavelet Coefficients Energy Redistribution and Heisenberg Principle of Uncertainty.” *Institute of Information Theory and Automation* 2 (1): 1–6. http://library.utia.cas.cz/separaty/2015/E/vosvrda-0449775.pdf.

Wang, Jingyuan, Ze Wang, Jianfeng Li, and Junjie Wu. 2018. “Multilevel Wavelet Decomposition Network for Interpretable Time Series Analysis.” *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2437–46. https://doi.org/10.1145/3219819.3220060.

Wu, Yipeng, and Shuming Liu. 2017. “A Review of Data-Driven Approaches for Burst Detection in Water Distribution Systems.” *Urban Water Journal* 14 (9): 972–83. https://doi.org/10.1080/1573062X.2017.1279191.

Xu, Liang, and Changcheng Huang. 2008. “Wavelet-Based SNR Analysis in Building Satellite Terminal Fault Identification System.” *IEEE International Conference on Communications*, no. January: 1942–46. https://doi.org/10.1109/ICC.2008.372.

Ye, Nong, Connie Borror, and Yebin Zhang. 2002. “EWMA Techniques for Computer Intrusion Detection through Anomalous Changes in Event Intensity.” *Quality and Reliability Engineering International* 18 (6): 443–51. https://doi.org/10.1002/qre.493.

Zaman, Dina, Manoj Kumar Tiwari, Ashok Kumar Gupta, and Dhrubjyoti Sen. 2020. “A Review of Leakage Detection Strategies for Pressurised Pipeline in Steady-State.” *Engineering Failure Analysis* 109 (February): 104264. https://doi.org/10.1016/j.engfailanal.2019.104264.

Zhao, Yang, Shengwei Wang, and Fu Xiao. 2013. “A Statistical Fault Detection and Diagnosis Method for Centrifugal Chillers Based on Exponentially-Weighted Moving Average Control Charts and Support Vector Regression.” *Applied Thermal Engineering* 51 (1–2): 560–72.

https://doi.org/10.1016/j.applthermaleng.2012.09.030.

Zhou, Zeng Guang, and Ping Tang. 2016. “Improving Time Series Anomaly Detection Based on Exponentially Weighted Moving Average (EWMA) of Season-Trend Model Residuals.” *International Geoscience and Remote Sensing Symposium (IGARSS)* 2016-Novem (January): 3414–17. https://doi.org/10.1109/IGARSS.2016.7729882.

Zohuri, Bahman. 2016. *Compact Heat Exchangers: Selection, Application, Design and Evaluation*. *Compact Heat Exchangers: Selection, Application, Design and Evaluation*. https://doi.org/10.1007/978-3-319-29835-1.