DATA-DRIVEN FAULT DETECTION APPROACH FOR NONLINEAR THREE-TANK SYSTEM

TEE CHIN JIAN

BACHELOR OF MECHATRONICS ENGINEERING WITH HONOURS

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2020/2021

DATA-DRIVEN FAULT DETECTION APPROACH
FOR NONLINEAR THREE-TANK SYSTEM

TEE CHIN JIAN

A report submitted
in partial fulfillment of the requirements for the degree of
Bachelor of Mechatronics Engineering with Honours

Faculty of Electrical Engineering

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2020/2021

DECLARATION

I declare that this study entitled "DATA-DRIVEN FAULT DETECTION APPROACH FOR NONLINEAR THREE-TANK SYSTEM" is the result of my own except as specified in the references and acknowledgements and that the original work contained herein have not been undertaken or done by unspecified sources or persons.

Signature : *Chinjian*

Name : **Tee Chin Jian**

Date : **5 July 2021**

APPROVAL

I hereby declare that I have checked this report entitled ”DATA-DRIVEN FAULT DETECTION APPROACH FOR NONLINEAR THREE-TANK SYSTEM” and in my opinion, this study it complies the partial fulfillment for awarding the award of the degree of Bachelor of Mechatronics Engineering with Honours

Signature : _____

Supervisor Name :Nur Maisarah binti Mohd Sobran

Date : 5 July 2021

ACKNOWLEDGEMENTS

The project is carried out as part of the Universiti Teknikal Malaysia Melaka (UTeM) course under Final Year Project (FYP) 2 program. This research is only possible to be completed from the help and inspiration of my main project supervisor, Miss Nur Maisarah binti Mohd Sobran. I would like to show my gratitude and thanks to my project supervisor as the completion of my project would not possible without the guidance and the techinacal advices. She has been continuously patient in guiding me despite my first exposure in handling training data with the MATLAB Simulink.

ABSTRACT

The data-driven is a progrress which in the formed of data instead of intuition by the people. It has capabilty to show of a process operating system and transfer the information to the moniring system in the form of data. Besides that, the fault detection is to indentify or detect the faults when it occurred. It also can be pointed the type of faults and its location of the three-tank system. The faults that may happen such as leakage fault, sensor failure, actuator failure, abrupt disturbance fault and so on. The main objectives of this research are to generate faulty and non-faulty data by using block diagram of MATLAB Simulink, to develop a principle componenet analysis (PCA) of data-driven fault detection method for three-tank system and to evaluate the performance of the method of data-driven fault detection by using confusion matric. The results for non-faulty event are 100% in accuraccy and specificity and below the threshold value which mean the the predicted non-faulty and actual non-faulty too. For the abrupt disturbance fault at tank 1, accuracy and sensitivity of the tank 1 dropped to 6.2% only. For leakage fault of tank 3, the accuracy and sensitivity are dropped to 97.2% and prediction is 100%. This shows that the predicted faulty and actual faulty too. For the sensor failure at tank 2, the accuracy, precision and sensitivity are 100% as the prediction made for this faulty is fully correct. Hence, the fault alarm will be ON when the system detected the occurance of the fault as the data is above the threshol value.

Table of Contents

List of Tables

List of Figures

CHAPTER 1
INTRODUCTION

1.1 INTRODUCTION

In this twenty-first century, there are varies two of control systemin the industrial sections. For example, open loop system and closed loop system. The open loop control system is an any physical system does not automatically correct the output result or the output quantity does not affect upon the input quantity. Therefore, open loop system does not feedback function. The closed loop system is the output quantity affected upon the input quantity in order to maintain or get the desired output values. Hence, it has automictically change the result for correction. The three-tank system is one of the examples of closed loop control system.

For three-tank system, the level of the control plant is important as it is to provide the desired specification of the users. This is because it related to the efficiency and accuracy of the result. Next, the fault detection of the three-tank system played the important role in industrial system. The fault detection method is to discover the failure of the system. It is help to detect the typically log time that the error occurred or recorded in the monitoring system. Hence, this help the user to detect when, where and what of failure system for manual intervention or initiate automatic recovery.

Moreover, the faults occurred are due to the failure of the system such as leakage of the tank, sensor failure and actuator failure. The leakage of the tank is leak of the valve below the tanks, while the sensor failure is the sensor failed to detect events or changes in its environment and send the information to monitoring system. The actuator failure is uncontrollable of the regulating valve of the water pump.

Lastly, the data-driven method is a popular among the engineering control. This is due to their dependence on the measurable process data and great application such as water treatment, chemical and petrochemical plants, oil and gas systems, food industry and so on. The data-driven approach can let the industrial to examine, analysis and interpretation their

data result easily. Therefore, the data driven fault detection approach is played important role to research the three-tank system and improve the system until it reached in terms of safety and reliability of the process.

1.2 PROJECT BACKGROUND

This research of paper is about to develop a data-driven fault detection method for three-tank system. The software MATLAB Simulink is used to modeling the three-tank system with faulty and non-faulty events. In order to complete the research, various sources had been through in order to satisfy the purpose of the project and identify each designs of modeling. Before starting, the understanding of the graph patterns, used of Simulink and modeling design of three-tank system are been through, so that the flow of the system is been done well.

Furthermore, fault detection approach can be classified into 2 categories which are model-based and signal processing-based. For model-based fault detection, it used to decide about the occurrence of the faults. This method can in the form of mathematical to generate the faulty events. For signal processing-based fault detection, it is used of some mathematical or statistical operations on measurements or neutral network to extract the information about faulty and non-faulty events.

On the other hands, there are many divisions of fault detection methods such as data methods and signal model, process model-based methods and knowledge-based method. Data method and signal model included limit checking and trend checking, data analysis (PCA), spectrum analysis and parametric. For process model-based methods, it contained parity equations, state observers, parameter estimation and nonlinear models. For knowledge-based method, it only included expert system and fuzzy logic. These methods are existing in control engineering section.

1.3 MOTIVATION

In this rapidly growth of modern industrial process, the data-driven fault detection method is attracted attention to use for control system as it has capability of extracting process operating process information completely in the monitoring system in form of data. The fault detection approach bring varies of advantages to industrial engineering. For example, fault detection method can ensure safety of plant or personal, improve the efficiency in manufacturing plant resources and so on.

Firstly, the fault detection approach can ensure safety of plant and user[1], [2]. For example, the chemical industrial used the three-tank control system for manufacturing process. Without the fault detection approach, the system may not able to detect and realize the fault such as leakage of the tank. The chemical leakage can harm to the plants and the users, since the chemical substances is corrosive and toxic. This also may cause pollution of the environment too. Hence, the fault detection strategy can be illustrated and cases studies on the three-tank system as the problems are formulated by modeling of control systems with fault-free and faulty cases.

Furthermore, fault detection approach can improve the efficiency in plant resources result in increment of productivity and financial. To ensure the fault detection can fulfil the demand of industrial section, the method such as model-based and data-driven are introduced[3]. The model-based is suitable for control system which known precise dynamic, while data-driven method has independence on measurable process data. For difference demand of industrial can use difference method of fault detection to improve the efficiency and performance of the control system.

1.4 PROBLEM STATEMENT

From the previous research, there are many problems faced during the analysis. One of the problems is to find the most suitable method of data-driven fault detection for three-tank system to detect faulty and non-faulty data. This is to improve the accuracy of the fault detection approach on the system. Moreover, the problem of abnormal of the system does not been detected instantly at industrial section. Hence, this causes inaccuracy results and decrement of productivity.

1.5 OBJECTIVES

The objectives of this research are:

- To generate faulty and non-faulty data by using block diagram of MATLAB Simulink.
- To develop a principle component analysis (PCA) of data-driven fault detection method for three-tank system.
- To evaluate the performance of the method of data-driven fault detection by using confusion matrices.

1.6 SCOPES

The scopes of this research to achieve the objectives above are:

- The function of Simulink is used to operate and design the block diagram for nonlinear control law in fault-free or fault events.
- The principle component analysis (PCA) is used to distinguish between faulty and non-faulty event.
- The performance of faulty and non-faulty are analyzed and compared.

CHAPTER 2
LITERATURE REVIEW

2.1 Introduction to data-driven fault detection method

For this rapidly growth of development, some of the industries section are used fault detection and isolation (FDI) and fault-tolerant control (FTC) in the control engineering[1]–[3]. This method of process included the advanced control system which may increase requirement safety of the plant and users even reliability[1], [3]. It has the high capability to process operating information and shown in the monitoring system. Recently, the researchers are given a big effort on fault detection for nonlinear system such as three-tank system[4], [5]. It is to ensure the closed loop control system remains it sensitivity to the faults. Therefore, fault detection method is classified into model-based and data driven.

2.1.1 Principle Component Analysis

PCA is a data processing method which mainly used to reduce data dimensionality in fault diagnosis. It also popular in various data driven approaches and multivariate statistical, as it is efficient use of input and output data.[4], [6]. Next, it is available in the linear system, non-linear system and actual system[4]. Therefore, it is a simple and effective way of fault detection method for non-linear process application such as three-tank system. The equations below is the basic mean matrix and standard deviation matrix of the data sample matrix.

$$X = [x_1 \quad x_2 \quad x_3 \dots x_m] \quad where\ m = 1,2,3,4 \dots$$

$$S = [s_1 \quad s_2 \quad s_3 \dots s_m] \quad where\ m = 1,2,3,4 \dots$$

$$vector, \bar{x_i} = \frac{1}{n}\sqrt{\sum_{j=1}^{n}(x_{ji})}$$

$$Variance, s_i = \frac{1}{n-1}\sqrt{\sum_{j=1}^{n}(x_{ji} - \bar{x_i})^2}$$

Besides that, PCA is also can be shown as an eigenvector of the data. Hence, it is really simplest method of the true eigenvector-based multivariate and factor analysis.

2.1.2 Fuzzy Kernel representations

This data-driven method of fault detection is an ease of implementation, relative simplicity and practically compared to other methods. Therefore, the rate of using this method is increasing by years as it is for high reliability and safety in the industrial process such as automated control engineering. Therefore, it has capability of extracting process operating information completely from the data[5], [7], [8].

Furthermore, data-driven observers are frequently used to train with the process of input and output history data under normal conditions. This is used to estimate the process future outputs, as the outputs is compared with the process real measurements. Therefore, the errors signal can be calculated and compared. These errors signals are detected by the fault detection of the stimulation and isolated faults by using a threshold values and statistical analysis[7].

The equations below is the consider non-linear systems[5], [8]:

$$E^{x0}: \begin{cases} x(k+1) = g\big(x(k), u(k)\big) + v(k) \\ y(k) = f\big(x(k), u(k)\big) + w(k) \end{cases}$$

Where $x(k) \in \mathcal{R}^{k_x}, y(k) \in \mathcal{R}^{k_y}$, $and\ u(k) \in \mathcal{R}^{k_u}$ denote the system state, the measurable output, and the control input. $v(k) \in \mathcal{R}^{k_x}\ and\ w(k) \in \mathcal{R}^{k_y}$ indicate the noise sequences on the process and the measurement.

For a noise-free and fault-free cases, let the $f(k) = 0, n_1(k) = 0, n_2 = 0$, then the equation below will be shown[5], [8].

$$K(z) \begin{bmatrix} u(z) \\ y(z) \end{bmatrix} = 0$$

Therefore, data-driven realization of fuzzy kernel representation for non-linear system is:

$$r(z) = K(z) \begin{bmatrix} u(z) \\ y(z) \end{bmatrix}$$

Then, the data-driven realization of kernel representation at the ith working point as defined as follows:

$$r(k) = K_{i,l} \begin{bmatrix} u_l(k) \\ y_l(k) \end{bmatrix}$$

The data-driven fault detection detected the errors efficiency. The stiffness of the signal is higher which mean it detected immediately and given the results.

2.1.3 Clustering technique

This clustering method is classified into two groups which are automatic and non-automatic techniques[7]. While automatic clustering approach, several number of the subsets of the dataset is determined without any prior knowledge of data. For non-automatic clustering, the number of the clusters must be defined in prior for algorithm.

Besides that, the initial dataset of the automatic clustering is assumed as to be normal or non-faulty condition at the beginning of the operation. It does not need to have the knowledge of occurrence the number of clusters such as faulty and non-faulty conditions. There are two ways to apply this method. Initially, the variable of the process is measured included the inlet flow rates and water level of the three tanks. Next, the errors of the result are recorded and used as features for clustering. Therefore, the results will be shown in it accuracy and identified number of clusters.

Furthermore, non-automatic clustering approach has very high accuracy for detecting and diagnosing the faulty condition. This is due to it is compared and combined the results with initial dataset and targeted dataset based on the process measured variables and error of the signals.

A fault detection and diagnosis (FDD) of nonlinear process is suggested in which automatic and non-automatic clustering methods are exploited. Hence, these approaches are to estimate the expected number and types of fault that may affect the process and system based on initial data recorded and collected from operation.

2.1.4 Just-in-time-learning (JITL)

The just-in-time-learning (JITL) is provided an effective tool for output prediction of non-linear system by measuring the input and output data[4]. It is a data-based fault detection method design to solve the problem of model-based. JITL has high performance of designed method for nonlinear system and wastewater treatment system benchmark. This is because it has significant advantages such as provided a potential tool to fault detection of complex industrial systemin data-based manner, inherent online adaptation, simple algorithm and easy implementation. It also owns a strong adaptability and general usage which can applied in actual industrial systems.

Furthermore, JITL can be represented in a simple mathematical model of the current input data for approximate prediction. The equation is ARX model described equation:

$$\hat{y}(k) = z^T(k-1)\aleph$$

where $\hat{y}(k)$ is the predicted output of the kth input sample, z^T is a regression vector of input sample and $\aleph$ is the local model parameter. By comparing the description of model, the equation will become as

$$\widehat{y_q} = {x_q}^T(P^TP)^{-1}P^Tv$$

$$\aleph = (P^TP)^{-1}P^Tv$$

$$P = W\Phi$$

$$v = WY$$

where W is represented the weight parameter, Φ and Y are the matrixes corresponding to current input x_q and output y. These are the basic equations of the JITL approach used to prediction purposes.

2.1.5 Moving Window Principal Component Analysis (MWPCA)

MWPCA is not same as PCA because it does not need to compute the principle component space and eigenvectors of the entire test data[6]. It can method can avoid the impact of the data before the steady state on the data of the fault. It also can make the fault characteristics more obvious when occurrence of fault.

Next, MWPCA used the general matrix as shown as below and where f is the window length and k is window matrix.

$$\begin{bmatrix} x_{k1} & x_{k2} & \dots & x_{km} \\ x_{k+1,1} & x_{k+1,2} & \dots & x_{k+1,m} \\ \dots & \dots & \dots & \dots \\ x_{k+f-1,1} & x_{k+f-1,2} & \dots & x_{k+f-1,m} \end{bmatrix}$$

After that, the window matrices are processed by PCA algorithm and T^2 and square prediction error (SPE) are calculated.

$$T^2{}_{ki} = x_{k\iota} P_k \lambda^{-1} \dot{P^T}{}_k \dot{X^T}{}_{k\iota}$$

For the performance, MWPCA has been improved significantly of detection rate compared to PCA. It is still high rate of false positives since the threshold selection is obey the multivariate statistical rules.

2.2 Selection of fault detection

2.2.1 Comparison of fault detection method

Table 1 Comparison method of fault detection

No	Technique	Engineering Aspect	Advantages
1	Clustering [7]	▪ To improve the accuracy of estimation number and types of fault	▪ Non-automatic cluster has relative-accuracy of 52.2% and absolute-accuracy of 47.2% ▪ Automatic cluster has absolute accuracy of 66% and 99.4%
2	Just-in-time learning (JITL)[4]	▪ To estimate the system output at difference events	▪ The strong applicability, high accuracy and practical significance are mentioned

3	Fuzzy Kernel representations[11]	▪ To achieve the systematic construction of fuzzy model ▪ To estimate the modelling error ▪ To test this application design scheme-based fault detection system	▪ Effectively detect fault in a timely manner as this technique is together with residual evaluation and threshold computation
4	Principle Component Analysis (PCA)[2]	▪ To diagnosis abnormal operating faulty and non-faulty conditions	▪ It able to handle the high dimensionality of the data ▪ It is more efficient and less process specific ▪ It is useful for modelling and controlling process plant with different events.
5	Moving Window Principle Component Analysis (MWPCA)[6]	▪ It can avoid the impact of the data before the steady state on the data of the failure period ▪ It makes fault data characteristics more obvious when fault occurred.	▪ The fault detection rate of MWPCA is higher compared to principle component analysis (PCA). ▪ MWPCA is more sensitive to fault data compared to PCA

For research on fault detection methods, the Principle Component Analysis (PCA) is more suitable to the three-tank system. This is due to it can reduce high dimension in the data sets and save operating cost by increasing the energy efficiency. It also can detect more than one faults at the same time of linear system. From the data produced by the PCA, it can be classified into faulty and non-faulty data. Therefore, PCA approach is a good for modelling and controlling of three-tank system.

2.3 Overview of Three tank system

The three-tank system is the three identical cylindrical tanks which connected with same cross-section areas of the cylindrical pipes. There have 2 actuator pumps located by above

tank 1 and tank 3. The water will flow through the pumps according to the control from a user. The figure 1 above is layout of the three-tank system[6].

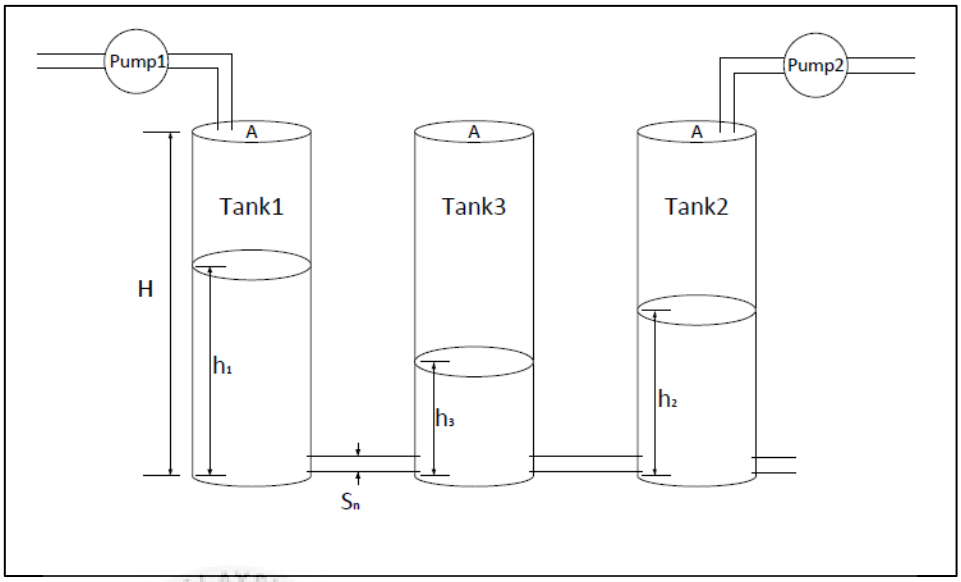

Figure 1 The layout of the three-tank system [6].

The three-tank system can be described by the following mass balance equations[6], [9]:

$$S\frac{dh}{dt} = sum\ of\ all\ flux$$

$$S\frac{dh_1}{dt} = Q_1 - Q_{13}$$

$$S\frac{dh_3}{dt} = Q_{13} - Q_{32}$$

$$S\frac{dh_2}{dt} = Q_2 + Q_{32} - Q_{20}$$

Q_{ij} represents the flow of the water between two tanks as shown as Figure 2.1.

2.3.1 Three tank system mathematical modelling

Model of three-tank system in State Space Equation

$$\dot{X}(t) = AX(t) + BU(t) \qquad (1)$$

$$Y(t) = CX(t) \qquad (2)$$

Where the X(t) is state variable, Y(t) is output variable, U(t) is input variable and A, B, C are state matrices.

An observer is used because it is a control system modeling. Therefore, new equation 3 and 4 are formed as below.

$$\hat{X}(t) = A\hat{X}(t) + BU(t) + \dot{L}\left[Y(t) - C\hat{X}(t)\right] \qquad (3)$$

$$\hat{Y}(t) = C\hat{X}(t) \qquad (4)$$

Where L is the gain matrix of the observer

Then, a residual r is generated to predict the observer output and the actual measurement.

$$r = Y(t) - \hat{Y}(t) = Y(t) - C\hat{X}(t) \qquad (5)$$

From the theoretically study, r = 0 means no fault occurred and r is not zero means a fault occurred

Moreover, a tolerant threshold, δ is used so that the fault detection can be carried out with equation 4

$$\begin{cases} |r| \leq \delta & Healthy \\ |r| \leq \delta & Faulty \end{cases} \qquad (6)$$

According to the linearized model, the equation below is given as follow[12]:

$$q_{i1} - q_1 = S_1 \frac{d}{dt} h_1 = q_{i1} - \frac{h_1 - h_2}{R_1} \qquad (7)$$

$$q_{i2} - q_3 = S_3 \frac{d}{dt} h_3 = q_{i2} - \frac{h_3 - h_2}{R_3} \qquad (8)$$

$$q_1 + q_3 - q_2 = S_2 \frac{d}{dt} h_2 = \frac{h_1 - h_2}{R_1} + \frac{h_3 - h_2}{R_3} - \frac{h_2}{R_2} \qquad (9)$$

Where i = 1, 2, 3…. and S is the cross-section areas of the three-tanks.

From the state space equation 1 and 2, there have unknown variable A, B and C. Them can represent in form of matrices as below.

$$A = \begin{bmatrix} -\frac{1}{S_1 R_1} & \frac{1}{S_1 R_1} & 0 \\ \frac{1}{S_2 R_1} & -\frac{1}{S_1}\left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}\right) & \frac{1}{S_2 R_3} \\ 0 & \frac{1}{S_3 R_3} & -\frac{1}{S_3 R_3} \end{bmatrix} \qquad (10)$$

$$B = \begin{bmatrix} \frac{1}{S_1} & 0 \\ 0 & 0 \\ 0 & \frac{1}{S_3} \end{bmatrix} \qquad (11)$$

$$C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad (12)$$

2.3.2 Fault detection of the three-tank system

The fault detection of the industrial process is becoming more and more important in the areas of the petroleum chemical industry, water conservancy, electricity, light industry, textile, pharmaceutical and building materials[6]. Therefore, the fault detection is one of the tools approaches give effective to solve the problem. Data-driven fault detection application is provided more and more attention in the recent years. This is due to the because their dependence in just measurable process data, not analytical model and great applicability in practices[5], [6].

The main point of the fault detection is to able to detect the faults when processing the control systems. This is to prevent the loss of the product and reduced the unnecessary wasted. Fault detection method enhance the safe operability of the process, lack of attention to a suitable level of process monitoring. This method will obviously lead to an environmental and health issues[2]. Therefore, the fault detection not only help in industries sections but also in decrease the rate of water pollution of environment.

The fault detection is an important research area in system control due to the improvements that it can be reached in the terms of safety and reliability of the process. This is because to detect actuator and components faults around during the operation of the process of the industries. The system is also to estimate such as an eventual actuator fault and proposed online. It is quite important as it able to carry out the faulty with the fault detection and estimation before the fault effect the system which may decrease the system performance[3].

On the other hands, the most common faults which may detect by the fault detection are leakage fault, sensor fault and actuators fault[4]–[6].For this fault detection, there is only few methods of data based in the three-tank system such as principle component analysis (PCA). This is because PCA is a relatively mature data-based fault diagnosis method[6].

2.3.3 The different cases of the failure

There are 3 different types of failure of the three-tank system which are sensor failure, actuator failure and water tanks leaked [6]. The sensor failure which mean the it does not sense the water level of the three-tanks system accurately. It can separately in 2 types which are partially failure and completely failure. The partially failure of the sensor is involved of inaccurate of the output and feedback of the water level data to the controller. The completely failure of the sensor mean there is no output feedback level of the water level tank.

Furthermore, actuator failure of the three-tanks system is mostly referred to the regulating valve of the water pumps above tank 1 and 2 as shown as figure 2.1. It can divide