Fault detection in a continuous stirring tank heater case study

Ruomu Tan *

* *Centre for Process Systems Engineering, Imperial College London, London, U.K., (e-mail: r.tan@imperial.ac.uk).*

Abstract: Previously in her secondment at the University of Alberta, Nina created the continuous stirring tank heater case study. This case study can be used for testing techniques in various aspects of process systems engineering, such as system identification, fault detection and diagnosis, and controller design. With my research being about data-driven fault detection algorithms for process monitoring, this case study is of great interest to me. Although many previous works have already used this case study for testing the fault detection and diagnosis algorithms, there is potential for more practical faulty scenarios. Inspired by the discussions with Nina, I try to create two fault scenarios. The first one develops gradually over time and the second one is to simulate a fault that may be physically induced in the experiment. The data from both healthy and faulty scenarios are collected Canonical Variate Analysis, which is a multivariate approach for modelling linear and dynamic processes, is applied to the data for fault detection. This test can be the starting point towards fault detection and prognosis for nonlinear and dynamic processes with multiple operating modes and transition periods.

Keywords: Fault detection, process monitoring, CSTH case study, benchmark simulation.

1. INTRODUCTION

In June 2015, I first met Nina at the University of Alberta, Edmonton, Canada, where I was studying my Master of Science (MSc) in Process Control. During Nina's visit, we had a great group discussion regarding the ongoing researches in the Process Control group and Nina kindly offered her insights and opinions to the questions my colleagues and I had. What I will always remember is that, between the meetings, Nina asked me what the plan after my study was. If I were told that I will be doing my PhD with Nina being my supervisor by then, I would not have believed. Luckily enough, various unforeseen and foreseeable events have made it happen. It is absolutely great to work with Nina.

Long before I started my MSc study, Nina used a Continuous Stirring Tank Heater (CSTH) test rig, which was located at the University of Alberta, for experiment and data collection. The first principles model of this rig was derived by combining the measured data and the first principles relationships, including the volumetric and the energy balances, the heat transfer and the characteristics of actuators and sensors. Afterwards, the CSTH case study has been set up in Simulink. The case study bridges the physical process, the simulated model and the data analytic algorithms. Although one may not have access to the test rig, the simulated model can produce data for testing data analytic algorithms in various scenarios. The fact that the noise and disturbance data were also generated by experiment adds on to the practical value. The case study has been described in Thornhill et al. (2008) while a website (http://personal-pages.ps.ic.ac.uk/∼nina/CSTHSimulation/index.htm) provides the Matlab and Simulink files.

In this work I try to use this case study for faulty data generation and fault detection. One of the suggestions for applications in Section 8.1 in the original paper is that the CSTH case study may be used for testing of fault detection and diagnosis algorithms. Many works have already done so. Most of them considered faulty scenarios such as increased measurement noise (Yu et al., 2016), sensor bias (Shang et al., 2019), step changes in Set Points (SPs) (Tong et al., 2014), or varying operating modes (Ge et al., 2016). Other tested faulty scenarios include sticking valves and increasing valve demands (Tong et al., 2014; Amin et al., 2019). The motivation for me is that, according to discussions with Nina, there are possibilities for physically inducing faults to the process. Therefore, I would like to test faulty scenarios which are more practical in the process. Moreover, as there is a growing interest towards fault prognosis, it may also be interesting to observe the behaviour of process variables with a gradually developing fault especially in a process under closed-loop control. In the CSTH case study, there are saturation units for setting the minimum and maximum flow rates through the valves, the maximum water temperature and the full tank level. The ability of the control loops to compensate process faults is therefore limited, which is true to the behaviour of the control loops in real-life processes.

A coincidence is that the original paper presenting this case study belongs to the Festschrift, which is a special issue in the Journal of Process Control, honouring the 65th birthday of Professor Dale Seborg. While Professor Seborg was the awardee of the prestigious Nordic Process Control Award in 2018, Nina receives this award in 2019. It is great to see that the exploration inspired and led by great researchers continues while the great researchers are being honoured. I hope that Nina will enjoy reading this small piece of work.

The remainder of this article is organized as follows. Section 2 briefly introduces the CSTH model. The test case for fault detection is presented in Section 3, which also introduces how the healthy and faulty data were generated. The fault detection results are presented in Section 4. In Section 5, there are several remarks, based on my own experience, that might be useful for others who also would like to use this case study. Section 6 draws the conclusions from the test.

2. THE CSTH MODEL

In the CSTH test rig, hot and cold water flows are mixed in a stirred water tank, where the mixed water is heated by hot steam via a heating coil. The mixed and heated water flows out of the tank through a draining pipeline at the bottom of the tank. For brevity, the P&ID of the test rig and details for the instrumentation and the tested conditions are omitted here.

The first principles model of the tank using energy and volumetric balance is built in Simulink. In addition, the models for the sensors and the valves in the rig were calibrated using experimental data. Two control loops exist in the Simulink model. One loop controls the water temperature in the tank. Given that the tank is well mixed, the measurement is taken at the outflow pipeline. The manipulated variable here is the steam valve opening. The water level in the tank is controlled by a cascade control loop where the inner loop controls the inlet cold water flow rate and the outer loop controls the tank level. The inlet hot water flow rate and the outflow rate, which are not controlled variables, may also be set by adjusting the corresponding valve opening. The tank level, the water temperature and the cold water flow are controlled using PI controllers. The open-loop and closed-loop behaviour of the test rig and the Simulink model compared in the original paper proves that the Simulink model can be a reliable digital twin of the test rig.

When running the simulation, the SPs for the level and the temperature can be specified. Since the original process model is nonlinear, linearized models can be realized at a steady state by fixing the SPs. The disturbance sequence in the level and the temperature and the noise sequence in the temperature, collected from experiment, were used as the inputs to the process to excite the process dynamics and to facilitate closed-loop identification.

3. TEST SETUP

This section introduces the test setup in the Simulink model for generating healthy and faulty data.

3.1 Healthy data generation

In the test, the process is running at the Standard operating condition 2. The level SP is 12 mA, the temperature SP is 10.5 mA and the hot water valve opening is set to be 5.5 mA. The training data set comprises 500 samples which are generated by using the first 500 samples of the cold water disturbance sequence, the level disturbance

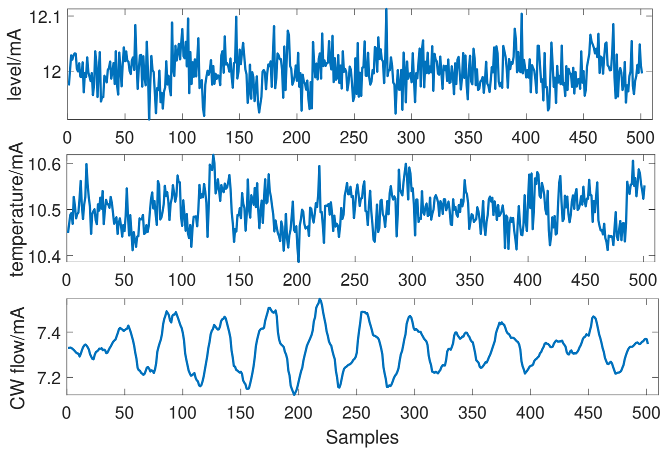

Fig. 1. Process data: healthy

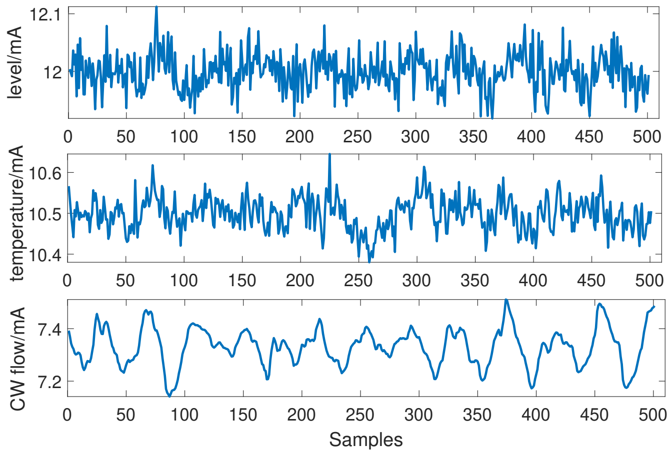

Fig. 2. Process data: validation

sequence and the temperature noise sequence as the inputs to the process. These noise and disturbance sequences are used as inputs to the simulation in all tested scenarios. The measurements taken in fault detection are the cold water flow rate, the tank level and the outflow temperature measured in mA, as Fig. 1 shows. Since that it is not necessary for a monitoring model to represent an input/output relationship, the inputs to the simulation are not included in the test data.

To test the robustness of the fault detection algorithm, a cross-validation data set is generated by using Sample 501 to Sample 1000 of the disturbance and noise sequences as the inputs. The validation data are plotted in Fig. 2.

3.2 Tested scenarios

The following faulty scenarios have been induced to the simulation model. The first fault is a developing fault which may cause conflicting behaviour of the controllers. The second fault aims to simulate a behaviour that may physically induce faults to the test rig.

Too much hot water inlet The objective of creating this faulty scenario is to challenge closed-loop control effect. The existing control loops for the temperature and the tank level can already cope with a variety of disturbances or drifting that may occur in the process.

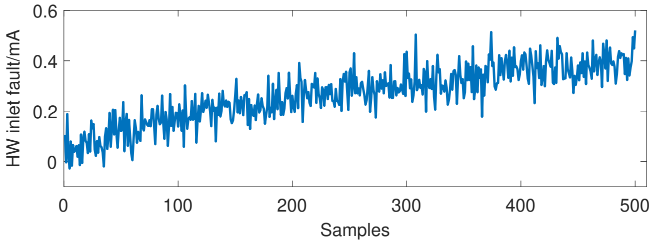

Fig. 3. Test sequence in the hot water input

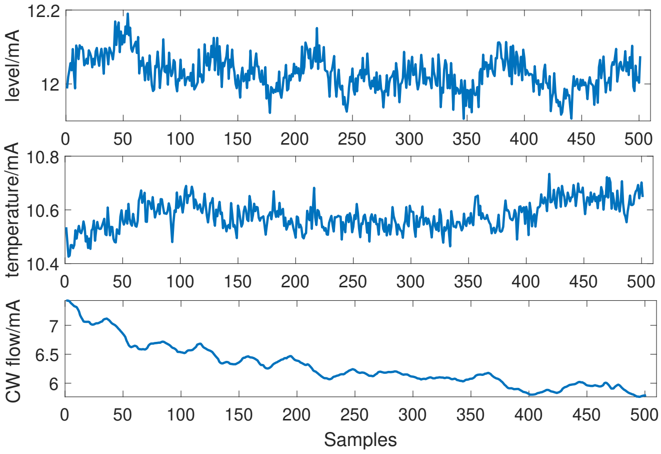

Fig. 4. Process data: Fault scenario 1

For example, if a fault is induced by reducing the hot water inlet, the controllers may compensate the influence on the temperature and the tank level by simultaneously increasing the steam flow rate and the cold water flow rate. On the other hand, the increasing hot water flow rate may not be easily managed since the cold water flow will be reduced to keep the tank level constant while it may be difficult for the steam heater to react to the extra enthalpy brought by the hot water inlet.

Fig. 3 shows the test sequence with 500 samples, which results in the hot water inlet gradually increasing. For this test, the input sequence comprises Sample 1001 to Sample 1500 of the noise and disturbance sequences.

The process data generated in this faulty scenario are visualized in Fig. 4. It can be seen that the inlet cold water flow reduces in reaction to the increasing inlet hot water flow. In the meantime, when compared with the temperature in the healthy case, the temperature in this faulty case begins to increase slowly especially at the end of this test, demonstrating that temperature controller may not be able to maintain the tank temperature. The controller output in Fig. 5 shows the effort of the temperature controller to reduce the steam flow rate. The controller output is not used for fault detection.

Outlet pipeline blockage In practice, randomly stepping on the outflow hosepipe may create blockages in the outflow. To produce this faulty scenario in the simulation, the following multiplicative fault model has been adopted to the outflow section:

$$f_{\text{out}}^{\text{faulty}} = \alpha f_{\text{out}}. \tag{1}$$

In the simulation, f_{out}, the outflow of the tank, is an empirical function of the tank level. To induce the fault,

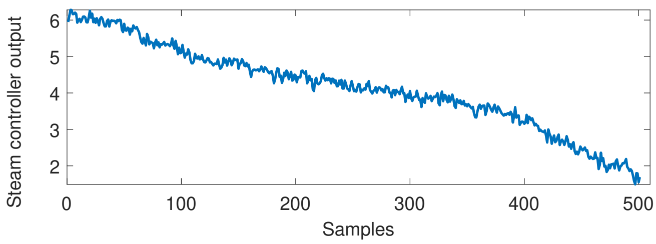

Fig. 5. Steam controller output

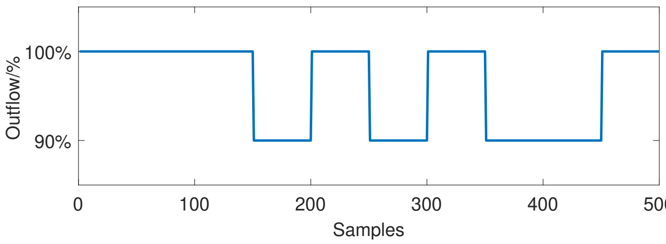

Fig. 6. α sequence in the outlet flow

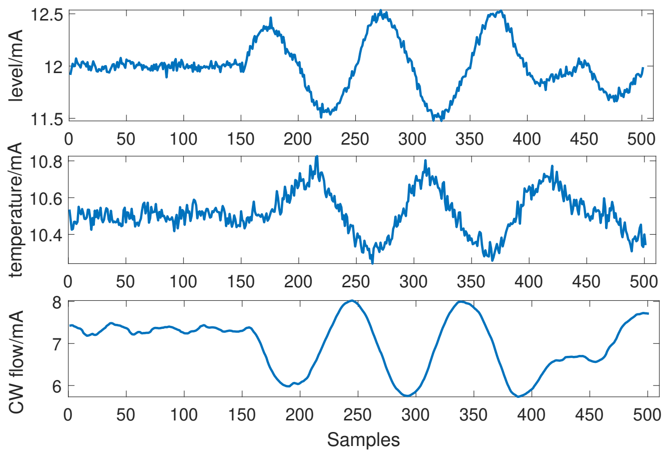

Fig. 7. Process data: Fault scenario 2

α is a random binary sequence which may take the value of 90% or 100%. In the fault-free situation, α is constantly 100%, indicating that there is no blockage in the outflow. α being 90% is to simulate the case when one steps on the outlet hosepipe, causing the blockage. When the blockage is removed, α returns to 100%.

Fig. 6 shows the binary sequence of α. Again, 500 samples are generated using this sequence. The inputs are Sample 1501 to Sample 2000 in the original input sequences. The measured data are plotted in Fig. 7. It can be seen that the disturbance in the physical layout of the process propagates to other process variables. During the propagation, the profile of the disturbance may have changed. For example, the step changes in the outflow result in the sine wave-like profiles in the tank level, the temperature and the cold water flow measurements. The influence of the blockage is immediately visible in the flow and level measurements while there exists a time delay for the tank temperature to react. If the duration of the blockage changes, the reaction in the process variables may be different with respect to the time constants.

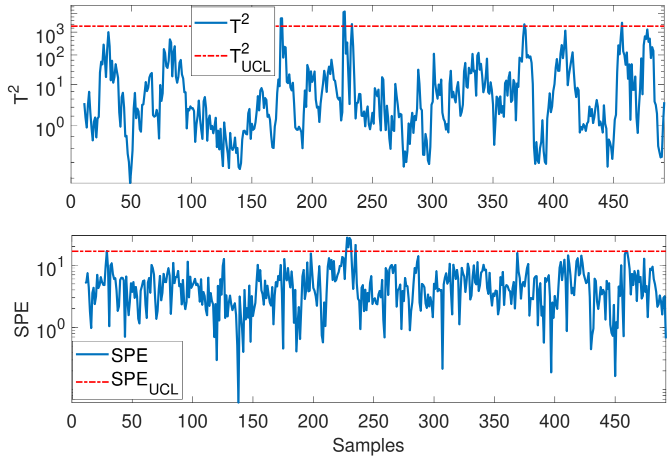

Fig. 8. Monitoring result: Validation

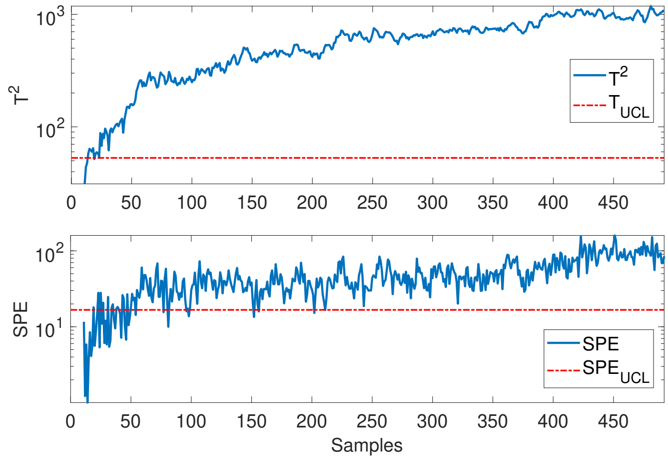

Fig. 9. Monitoring result: Fault scenario 1

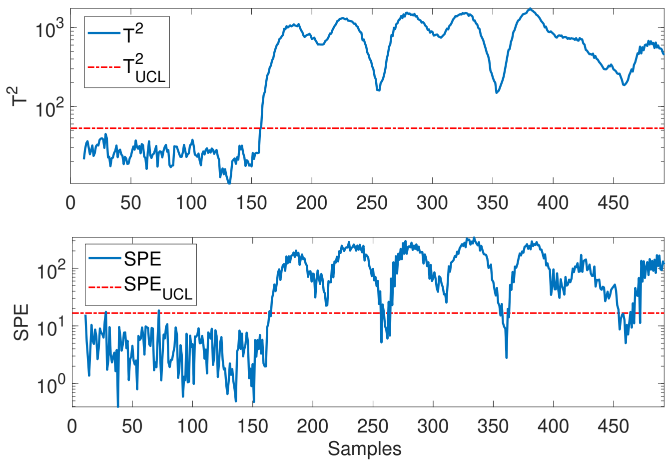

Fig. 10. Monitoring result: Fault scenario 2

4. RESULTS

By visually inspecting the data, the process variables in the faulty scenarios behave differently from the variables in the healthy scenario. However, the fault may not always be visible in every process variable, making the multivariate statistical approach necessary. Given that the process model is linear and dynamic, the Canonical Variate Analysis (CVA) is adopted. Since the methodology is not the focus of this work, the introduction has been omitted. Ruiz-Cárcel et al. (2015) have adopted the CVA algorithm for fault detection and details of the algorithm can be found there.

The monitoring statistics T^2 and SPE obtained using the validation set, the fault scenario 1 and the fault scenario 2 are visualized in Figs 8-10, respectively. Since the time delay in CVA is taken as 10 samples, the monitoring statistics are available starting from the 11th sample. For better visibility, the vertical axis is the logarithmic value of the monitoring statistics. Fig. 8 shows that the algorithm is robust to the randomness in the input sequences and does not trigger many false alarms on the validation data. By comparing the fault sequence in Fig. 3 and the T^2 statistic in Fig. 9, it can be seen that the trend of the fault sequence and the trend of T^2 are similar. Therefore, T^2 may be a good indicator not only for the fault existence but also for the fault development. For Fig. 10, the fault detection happens as soon as the blockage is induced for the first time. Alarms are still triggered in the periods when the blockage has been removed. This may be caused by the delay in the temperature, indicating that, even if the fault has been removed from the process, some of the process variables may still be abnormal and may take time to fully return to the healthy condition. Additionally, when developing dynamic algorithms, the optimal time delay for different process variables may also be different.

5. DISCUSSIONS

The CSTH website and the original paper have given a self-contained description of the case study and the tests were well-documented. Nevertheless, practical issues may exist when using this case study for various purposes, particularly when there are changes to be made to the original simulation model. This section focuses on the remarks that might be useful when using this case study.

5.1 The units

In the simulation model, the SPs and measurements are all taken in the unit of mA and the range for all variables is 4 20 mA. Although the unified unit may facilitate data pre-processing and visualization, extra attention needs to be paid when making changes to the original model. Particularly when inducing fault sequences, which may have physical meanings, the sequences should be designed such that the magnitude is in accordance with the units of the variables where the sequence has been induced. For example, the disturbances used for exciting the system were recorded in the unit of mA even though these disturbances were in different process variables with different units. When inducing the *too much hot water* fault, the fault sequence was generated in mA because it is inserted as an additive sequence to the hot water valve set point which is taken in mA. As for the *outlet pipeline blockage* fault, the sequence was designed in percentage because it simulates a multiplicative fault to the empirical equation of the outflow with respect to the tank level.

5.2 *The disturbances and the faults*

While noise and disturbances always exist in real-life processes, the disturbance used in the simulation model can excite the dynamic effects in the components of the model such that the simulated dataset can be used for model identification. Similarly in this test, the disturbances are used as the excitation of the process, making it possible to use the process measurements to train the monitoring model. Additional faulty sequences, which may simulate the degradation in some process components and may result in anomalies in process data, are induced to generate faulty data that can be used to test the fault detection performance.

5.3 *Variable selection*

The Simulink model enables users to measure the variables and the states of the process, some of which may not be accessible in practice. The variables being collected should be specified properly according to the purpose of the. When using this case study to generate data for system identification, the inputs and outputs should be specified properly. For example, the closed-loop models for the tank level and the tank temperature can be built using the SPs of the tank level and temperature as the inputs and the measurements of these two variables as the outputs. When using the case study for testing fault detection algorithms, it may not be necessary to specify the input and output variables. Nevertheless, it is necessary to select the variables that are influenced by the fault being tested and are easy to measure. For example, for the *too much hot water inlet* fault, if the cold water level is removed from the data set, it may take longer for the algorithm to detect the fault as the other two variables are under control. On the other hand, this fault may be identified easily if the inlet hot water flow is measured. However, in practice the root cause of the fault may not always be measured.

6. CONCLUSIONS

In this work, two faulty scenarios were tested for the CSTH case study. The aim was to generate faulty scenarios which are gradually developing and are able to be physically induced to the process. The CVA algorithm was applied to the data sets generated from the simulation at the standard operating mode with and without the tested faulty scenarios for training the monitoring model and for fault detection, respectively. The fault detection results demonstrated that the CVA algorithm can detect both faults whilst being robust to the randomness in the input sequence. Moreover, the T^2 statistic can also tract the development of the fault severity.

There are two directions for the future work. The first one is to consider the fault prognosis in Fault Scenario 1. This can be implemented either by building a time sequence model for the T^2 obtained by CVA or by building a predictive model of the process variables using CVA directly. The second one is to extend the test to nonlinear and dynamic process monitoring. Given that the original CSTH case study is nonlinear and the linear models were obtained by linearization at standard operating modes, it is possible to create a test case where the input sequence covers the whole operating space. Algorithms for nonlinear dynamic process monitoring can then be adopted for learning a global monitoring model which accounts for the varying operating modes and the transition periods between them.

ACKNOWLEDGEMENTS

This work is dedicated to Nina. I am sincerely thankful for having Nina as my PhD supervisor. But more importantly, she is the role model that I, as a researcher, always look up to. Just like the golden standard in statistics, her enthusiasm about the technical topics, her solid and comprehensive knowledge of a variety of aspects in process automation and control, her willingness to listen and to discuss, and her concise way of writing (especially her attitude towards waffles) will always be the benchmark for me to evaluate my performance as a researcher.

Moreover, Nina's inspiration on young researchers is worldwide and is not limited to the people who have directly worked with her. In the DYCOPS conference held in Brazil this year, a PhD student from a local university presenting his work on oscillation detection told me that Nina has also been his role model and his work was inspired by Nina's paper on oscillation detection. I can relate to the feeling and I think that reading papers written by Nina lays a good foundation for one's research.

I would like to offer my warmest congratulations to Nina for the great and long-lasting success, while having fun all the way, in her career. Thank you Nina for training, inspiring and motivating me and many other researchers in the field of process systems engineering. I wish you the very best for your retirement.

REFERENCES

Amin, M.T., Khan, F., and Imtiaz, S. (2019). Fault detection and pathway analysis using a dynamic bayesian network. *Chemical Engineering Science*, 195, 777 – 790.

Ge, Z., Zhong, S., and Zhang, Y. (2016). Semisupervised kernel learning for FDA model and its application for fault classification in industrial processes. *IEEE Transactions on Industrial Informatics*, 12(4), 1403–1411.

Ruiz-Cárcel, C., Cao, Y., Mba, D., Lao, L., and Samuel, R. (2015). Statistical process monitoring of a multiphase flow facility. *Control Engineering Practice*, 42, 74–88.

Shang, J., Zhou, D., Chen, M., Ji, H., and Zhang, H. (2019). Incipient sensor fault diagnosis in multimode processes using conditionally independent bayesian learning based recursive transformed component statistical analysis. *Journal of Process Control*, 77, 7 – 19.

Thornhill, N.F., Patwardhan, S.C., and Shah, S.L. (2008). A continuous stirred tank heater simulation model with applications. *Journal of Process Control*, 18(3), 347 – 360. Festschrift honouring Professor Dale Seborg.

Tong, C., El-Farra, N.H., Palazoglu, A., and Yan, X. (2014). Fault detection and isolation in hybrid process systems using a combined data-driven and observer-design methodology. *AIChE Journal*, 60(8), 2805–2814.

Yu, H., Khan, F., and Garaniya, V. (2016). A sparse pca for nonlinear fault diagnosis and robust feature discovery of industrial processes. *AIChE Journal*, 62(5), 1494–1513.