Preprints, 10th IFAC International Symposium on
Advanced Control of Chemical Processes
Shenyang, Liaoning, China, July 25-27, 2018

Analytics of Heterogeneous Process Data: Multiphase Flow Facility Case Study

Anna Stief*, Ruomu Tan, Yi Cao**, James R. Ottewill***

*ABB Corporate Research Center, ul. Starowislna 13a, 31-038 Krakow, Poland
Emails: anna.stief@pl.abb.com; james.ottewill@pl.abb.com
** School of Water, Energy and Environment Cranfield University, Cranfield, UK MK43 0AL
Email: R.Tan@cranfield.ac.uk;
***College of Chemical and Biological Engineering, Zhejiang University, Hangzhou, 310027, China
Email: Yi4Cao2@gmail.com

Abstract: Improvements in sensing, connectivity and computing technologies mean that industrial processes now generate a vast amount of data from a variety of disparate sources. Data may take a number of different forms, from different time-domain signals, sampled at different rates using various types of sensors, through to more disparate sources such as alarm and event logs. New process and condition monitoring techniques are needed to be developed to tackle the new challenges of big and heterogeneous data. Although there are a few publicly available benchmark studies, e.g. the Tennessee Eastman process plant (Ricker, 1995), a multiphase flow benchmark case for statistical process monitoring (Ruiz-Cárcel et al., 2015), they provide only standard process data. This work presents a benchmark case on an industrial scale multiphase flow facility. Various operational conditions were tested under normal operating modes as well as with seeded faults. Heterogeneous data was collected from various sources, including process data, alarm data and high frequency ultrasonic and pressure data. Two different fault detection algorithms are applied to the data, a multivariate PCA-enhanced Canonical Variate Analysis (CVA) and a probabilistic Bayesian method. This benchmark case study with data from disparate sources can be used for algorithm development and validation for fault detection, fault identification, fault classification, fault severity detection, monitoring of fault evolution and prognostics.

Keywords: Process monitoring, Condition monitoring, Data analytics, Experiments, Fault detection

1. INTRODUCTION

In today's process industry fail-safe operation without any deviation from schedule is essential to ensuring plant productivity, profitability and sustainability. To achieve such desirable operation, process monitoring, early fault detection and diagnosis are core necessities. Industrial plants often contain a vast array of potential sources of data, ranging from sensors used for the fundamental control and monitoring of the process, through to more disparate sources of data such as event logs or video records. Improvements in sensing, connectivity and computing technologies mean that the quantity of data that may be recorded and stored from a variety of disparate sources is increasing. Considering these multiple sources of heterogeneous data in combination potentially offers a number of opportunities for improved reliability and robustness of monitoring algorithms (Lu et al., 2014), as complex system interactions can be modelled, strengths of different sensing approaches can be leveraged and weaknesses may be mitigated (Hou and Bergmann, 2012). However, in order to realize these opportunities new analytics approaches are required in order to manage, fuse and process the heterogeneous data so that meaningful and actionable insight might be extracted.

Data may take a number of diverse forms, from different time-domain signals, sampled at different rates using various types of sensors, through to more disparate sources such as alarm and event logs. The decreasing limitations in storage capacity make it possible to collect sensor data at higher sampling rates, which can increase the sensitivity of detecting mild changes in the system. For example, Ruiz-Cárcel et al. showed that combining process and vibration data, the latter sampled at a much higher rate, improves the performance of fault detection for a wider range of diagnosable faults (Ruiz-Cárcel et al., 2016). Industrial-scale benchmark studies like the Tennessee Eastman process plant (Ricker, 1995), or a multiphase flow benchmark case for statistical process monitoring (Ruiz-Cárcel et al., 2015) only include standard process data. Therefore there is a need for a benchmark study which includes more heterogeneous data to support the development and validation of advanced process and condition monitoring techniques.

This work aims to fill this gap by describing a benchmark case based on an industrial scale multiphase flow facility with data available from several sources including process data, alarm data and high frequency ultrasonic and pressure data. Various operational conditions were considered under normal operating modes as well as with seeded faults to generate a multi-rate, multimodal data set from disparate sources. A multivariate PCA-enhanced CVA algorithm is applied to the ultrasonic and process data for fault detection, whilst a probabilistic Bayesian method is used to show that the high frequency pressure data itself can be used for fault classification. These examples illustrate how the data from

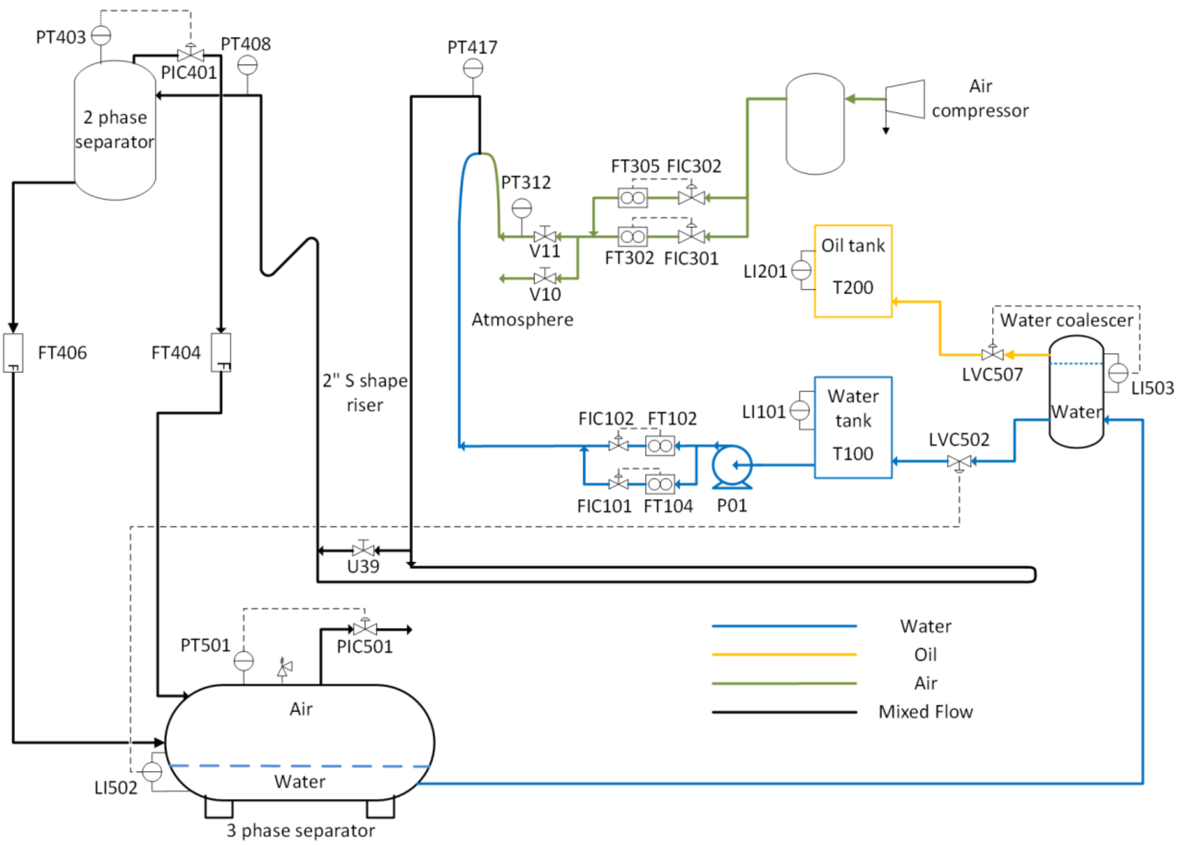

Figure 1. Schematic of the multiphase flow facility

disparate sources can be used for algorithm development and validation.

The rest of the paper is organized as follows: firstly, a general overview of the experimental facility is given, then the available data is described, followed by a summary of the tested normal and faulty scenarios. Initial results are presented with a discussion of the challenges using disparate data. Finally, the paper concludes by highlighting the applicability of this benchmark case study and planned future uses of the acquired data. This is a preliminary study of a subset of the benchmark data set. The full data set is still under review and verification. It will be made accessible to public when a comprehensive study of the data set is published.

2. EXPERIMENTAL FACILITY

2.1 Multiphase flow facility

The benchmark case study is conducted on a multiphase flow facility in the Process System Engineering lab of Cranfield University, which has previously been studied based on process data (Ruiz-Cárcel et al., 2015). The test facility was originally designed for the investigation of the transportation, measurement and control of multiphase flow which, in off-shore oil and gas operations, usually is comprised of water, air and oil. In this experiment, water and air are mixed to flow through the horizontal section and then separated. A number of different testing scenarios are implemented in order to generate data for both normal and faulty states. It is planned that this data will subsequently be used for the development of data analysis and process monitoring approaches.

2.2 Measurement specification

A sets of sensors measuring different variables with different sampling rates have been installed on the test facility in order to obtain real-time data for both on-line monitoring and further off-line data analysis. Sensor readings are collected and stored along with their time stamps and alarm information during the experiment. Table 1 summarizes the data recorded in this case study.

Table 1. Data availability through the experiment

Measured variable	Sampling rate	Availability	Platform
Process variables	1 Hz	Continuous	DeltaV
Alarm, event, change logs	Event driven	Discrete event	DeltaV
Doppler ultrasonic sensor	10 kHz	60 s	LabView
High frequency pressure sensors	5 kHz	60 s	LabView
Videos	-	30-60 s	Camera

2.3 Facility layout

Figure 1 provides an overview of the configuration of pipelines and instrumentations of the test facility. The feed air and water flows are the inputs to the facility and their flow rates are controlled for implementing different operating conditions. Input flows are mixed in the mixing zone and directed through a horizontal pipeline to the 2” vertical riser, which has an S shape section implanted half way. After reaching the riser top, the mixed flow is separated by two separators in sequence; the water flow returns to the storage tank and the air flow is exhausted to atmosphere after separation. The facility is instrumented with various pressure, flow rate, temperature and density sensors; all of the 17 process variables recorded in the tests are shown in Figure 1.

The high frequency pressure sensors are distributed along the pipelines from the mixing zone to the riser top while the ultrasonic sensor is located at the riser top, as shown in Figure 2. Additionally, two transparent sections, displayed in this zoomed-in diagram, are installed on the riser bottom and top for observation of the flow regime.

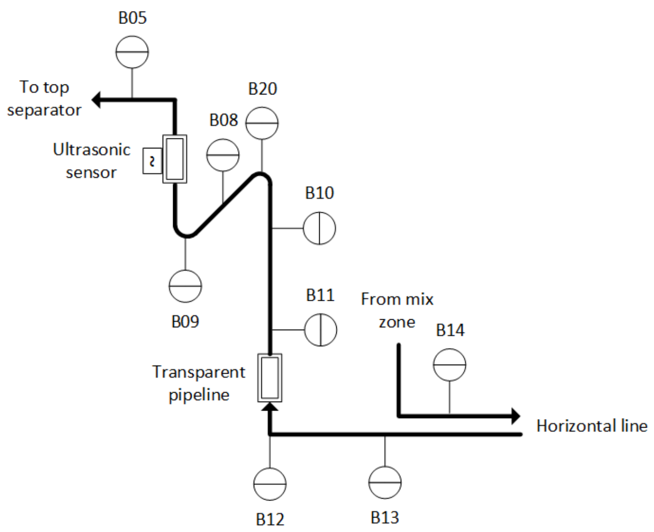

Figure 2. High frequency measurement zone

3. DATA TYPES

In this section an overview is given about the data types available throughout the experiment, a summary of which can be found at Table 1.

3.1 Process data

Process data was collected using DeltaV, which is a SCADA system provided by Emerson Process Management. It is responsible for the control of the process and provides a way to retrieve process data. DeltaV samples all the connected sensors at 1 Hz. The selected process variables are listed in Table 2 with their corresponding tag and unit.

3.2 Alarms, events and changes data

Alarm, event and change data was logged during the whole operation of the process. The alarm data consists of module alarms, which indicate the health state of the sensors and their connectivity to the DeltaV system. The event data consist of logs of sensor failure events. Change data consist of a log of changes that the user made to the system or to the process, such as adjusting a valve position, acknowledging an alarm, changing the set point of a control valve and changing the critical value for a sensor.

3.3 Doppler ultrasonic data

A Continuous Wave Doppler Ultrasound non-invasive, clamp-on sensor was used for the experiment. It is based on the Doppler Effect, namely that the frequency of an ultrasonic wave reflected from the scatterers of a moving medium is shifted in proportion to the velocity of the medium. This principles makes ultrasonic sensors suitable as flow meters for multiphase flows. (Lynnworth, 1989).

The sensor has two piezoelectric crystal transducers. One of the transducers emits an ultrasonic signal at 500 kHz, while the other transducer receives the reflected signal from the multiphase flow. The sensor provides the Doppler frequency shift in the form of an output voltage signal. The Doppler ultrasonic data was recorded with LabView at a 10 kHz sampling rate for 60 seconds for all of the tested scenarios in steady-state. The recordings were manually synchronized with the process data from DeltaV.

Table 2. Process variables

Sensor tag	Measured process variable	Unit
FT305/302	Inlet air flow rate	Sm^3/h
FT305	Inlet air temperature	°C
PT312	Air delivery pressure	barg
FT102/104	Inlet water flow rate	kg/s
FT102	Inlet water temperature	°C
FT102	Inlet water density	kg/m^3
PT417	Pressure in the mixing zone	barg
PT408	Pressure at the riser top	barg
PT403	Pressure in the top separator	barg
FT404	Top separator output air flow rate	m^3/h
FT406	Top separator output water flow rate	kg/s
PT501	Pressure in the 3-phase separator	barg
PIC501	Air outlet valve 3-phase separator	%
LI502	Water-oil 3-level phase separator	%
LI503	Water coalescer level	%
LVC502-SR	Water coalescer outlet valve	%
LI101	Water tank Level	m

3.4 High frequency pressure data

For all tested scenarios, high frequency pressure data was recorded using LabView at a 5 kHz sampling rate for 60 seconds during steady-state operation. The location of the 9 pressure sensors is given in Figure 2, while the sensors are listed in Table 3. The measurements are given in units of barg. The recordings were manually synchronized with the process data from DeltaV. As there are no process measurements through the horizontal pipeline, vertical riser and S-shape riser in the DeltaV system, the high frequency pressure data provides a better insight to the pressure fluctuations at the horizontal pipeline and the riser.

Table 3. Pressure variables

Sensor tag	Measured pressure variable
B14	Before horizontal line
B13	After horizontal line, before riser base
B12	Riser base
B11	Vertical riser after transparent pipe beyond riser base
B10	The middle of vertical riser, before S shape
B20	Top of S shape
B08	Middle of inclining part of S shape
B09	Bottom of S shape
B05	After S shape riser and riser top

3.5 Videos

The 2” pipeline has a transparent section at both the top and the bottom of the riser. Videos were taken for a period of 30-60 seconds during different tested scenarios, for educational purposes and to provide a better understanding of the flow regimes.

4. TESTED SCENARIOS

The experiment was conducted under 5 different scenarios. In addition to normal operating conditions, 4 different incipient or intermittent faults were tested. These faults were designed to simulate real process malfunctions, such as leakage, blockage or incorrect operation of the system. In the following section a detailed description is given for each tested scenario.

4.1 Normal operating conditions

In case of algorithm development for process monitoring, a representative normal training dataset is essential for the success of the algorithms. In this benchmark case study normal data was collected with a continuous, stable flow regime. The corresponding air and water flow rate combinations are shown in Table 4. 13 normal datasets were recorded. The high frequency measurements were taken once the flow stabilized. Videos are recorded from the transparent section at the riser top.

4.2 Slugging

Slugging is an intermittent fault that happens in multiphase risers, when the gas and liquid flow rates are relatively low. It is an unwanted condition in the offshore oil and gas production. The liquid builds up at the bottom of the riser blocking the gas flow. The pressure increases at the riser bottom till it is sufficient to push up the air and water slug to the riser top, then the water falls down and the cycle starts again. This phenomena results in unwanted oscillations in the pressure, in the flow rate and in the density through the riser causing faster deterioration of the equipment (Jansen et al., 1996). Slugging was achieved by manipulating the input air and water flow rates until the fault mode was observed: 7 slugging datasets were collected through the experiment, the set points of which are shown in Table 4. Videos are recorded from the transparent section at the riser top.

Table 4. Operating conditions

		Water flow rate (kg/s)				
		0,1	0,5	1	2	3,5
Air flow rate (Sm^3/h)	20	slugging	slugging	slugging	slugging	normal
	50	slugging	slugging	slugging	normal	normal
	100	normal	normal	normal	normal	normal
	120	normal & faults				
	150		normal & faults			
	200	normal	normal	normal		

Figure 3 shows time-series plots of selected data from different sources for the slugging case of 20 Sm^3/h air, 0,1 kg/s water flow rate. The data is scaled to have values between zero and 1 to represent the data trends. The top three signals are process variables sampled at 1 Hz. The fourth signal shows the ultrasonic sensor sampled at 10 kHz, the fifth signal is a module alarm, which is triggered twice during the 60 second measurement period. The last signal shows a pressure signal sampled at 5 kHz.

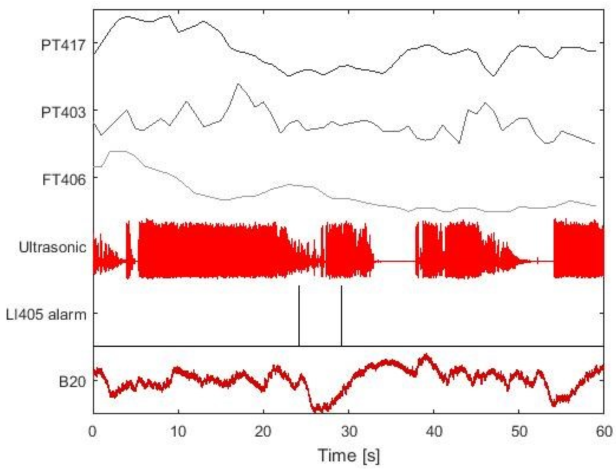

Figure 3. Heterogeneous data for a slugging case

4.3 Air leakage

The aim of this seeded fault scenario is to simulate a gradually developing air leakage in the input air pipeline. By opening valve V10 manually the air is partially leaked out to the atmosphere. The valve was opened gradually simulating an incipient leakage with the valve positions being recorded along with the data. The set points of the air and water flow rates chosen for the air leakage tests are shown in Table 4 ("normal & faults"). Two tests were carried out, with high frequency measurements being taken once the flow had stabilized at its given set point for the selected valve position.

At the beginning of each test with a fixed air and water flow rate the valve was fully closed resulting in normal flow conditions. With the developing of the air leakage, the flow regime in the riser shifts from normal to slugging. As the amount of input air is reduced an intermittent cyclic behaviour appears: first there is normal flow present, then the flow disappears, then water appears with big air bubbles, then the cycle starts again. This behaviour is very similar to severe slugging, although the reason behind it is not only due to the reduced air flow rate but also due to the pressure drop caused by the leakage. Once all of the input air leaked out, the pressure drops and only the water stays, leading to a continuous liquid only flow regime. Videos were recorded from the transparent section at the riser top to study this phenomenon.

4.4 Air blockage

The aim of this fault scenario is to simulate a gradually developing blockage in the input airline by manually closing valve V11. Initially for each set point with a fixed air and water flow rate the valve was fully open with normal flow conditions. Then the valve was gradually closed. Similarly to the air leakage fault, the set points are shown in Table 4. The high frequency measurements were taken once the flow stabilized at its given set point for the selected valve position.

The flow regime for air blockage is different from the leakage case: as there is no pressure drop in the system the flow remains continuous, although mild slugging is observed.

4.5 Diverted flow

The aim of this fault scenario is to simulate a case of a mixed diverted flow. In real processes this fault could be caused by the incorrect operation of the system. The U39 bypass valve was gradually opened. The mixed flow is partially led straight to the riser and partially led into the horizontal pipeline before joining the riser. The set points are the same, as for the previous seeded faults, as shown in Table 4. The high frequency measurements were taken once the flow stabilized at a certain set point with a certain valve position.

This seeded fault is different from the previous two faults, as there is no change in the flow regime in the riser and it is not observable at the riser top. However at the riser bottom, at the transparent section the change is visible. With increasing the opening of U39 the air disappears from the transparent section showing that most of the air flow is directed straight to the riser. As the diverted flow fault is introduced between the beginning of the horizontal pipeline and the riser base, the high frequency pressure sensors are able to detect the change in the flow. Videos were recorded from the transparent section at the riser bottom.

5. RESULTS AND DISCUSSION

5.1 PCA-enhanced CVA for fault detection

To demonstrate the feasibility of this case study in validating data-driven process monitoring algorithms, PCA-enhanced CVA is selected as a candidate. This algorithm has also been proposed and applied to a previous benchmark case study for fault detection using regular process measurements (Tan and Cao, 2017)

In order to incorporate the ultrasonic information in PCA-enhanced CVA method, a representative feature at a lower sampling rate is extracted from ultrasonic data by calculating the sample variance within a 1 second time window. It is included as an additional variable by aligning with other regular process measurements during 60-second segments in both normal and faulty scenarios. For slugging detection, the monitoring model and corresponding control limits of the monitoring index are trained on different data segments taken from normal operation. For other seeded faults, the models are trained using normal data segments recorded under flow conditions marked as "normal & faults" in Table 4. To evaluate the monitoring performance, the Detection Rate for the faulty sets and the False Alarm Rate for the healthy sets are calculated. The quantified performance of detection and false alarms on normal and faulty data sets are presented in Table 5.

It may be observed that the PCA-enhanced CVA algorithm has a satisfactory performance in slugging detection owing to the availability of additional ultrasonic sensor data. Nevertheless, there is still scope for improvement in other faulty scenarios; in particular, the air blockage fault is promising as an example of incipient faults as it develops gradually over time and the detection performance of the current PCA-enhanced CVA algorithm remains unsatisfactory. In the future, improved approaches for incipient fault detection, which make use of the additional available data, can be developed and validated using this case study.

Table 5. Fault detection by PCA-enhanced CVA

	Scenarios	T^2	Q
False alarm rate (%)	Normal	3,97	3,05
Detection rate (%)	Blockage	99,32	64,38
	Leakage	42,77	25,11
	Diverted	72,28	61,25
	Slugging	84,29	66,53

5.2 A probabilistic Bayesian method for fault classification using high frequency pressure data

The high frequency pressure data contains valuable information which can be used to detect the different fault scenarios. The pressure sensors are sampled at 5 kHz for 60 seconds. For the analysis, frequency domain features were extracted from data recorded from each of the nine sensors. The maximum amplitude at each 100 Hz frequency window was saved as a feature, resulting in 25 features per second per pressure sensor. The features were normalized and then divided into normal, leakage, blockage, diverted flow and slugging sets accordingly.

The algorithm used for fault classification on the pressure data is a probabilistic method based on Bayesian statistics. For further details on the Bayesian method see (Stief et al., 2017). A single stage Bayesian inference approach was implemented. The feature data for each fault scenario was randomly split into 70 % training set and 30 % test set. Thresholds were set at the upper and lower 1% of the normal operation range.

Table 6. Results of the Bayesian fault classification

		Diagnosed condition				
	%	Normal	Blockage	Leakage	Diverted	Slugging
Actual condition	Normal	82,93	9,50	1,34	0,03	6,21
	Blockage	19,05	75,56	0,42	0,02	4,94
	Leakage	24,15	5,00	64,88	3,41	2,55
	Diverted	0,59	0,20	2,53	96,08	0,60
	Slugging	32,25	1,90	3,66	0,19	62,00

The results in Table 6 show how the algorithm performed during the fault classification: the columns show what the algorithm diagnosed, while the rows show the actual conditions. For example, for the normal condition, the algorithm successfully detected the actual normal condition in 82,93% of the test cases, while it falsely predicted blockage in 9,5%. The correct classification rate was highest with 96,08% for the diverted flow, while it was lowest 62% for the slugging condition. The reason behind the high misclassification of blockage and leakage as normal is that valves are non-linear: small valve adjustments cause minor changes in the flow regime at the beginning of the experiment. The algorithm in its current form is suitable for fault classification using the normalized pressure data, however the multimodality issue is yet to be addressed. In the future using multivariate statistical methods like PCA on the process data, as a pre-processing step before the Bayesian method, could solve the above mentioned issue, while keeping the probabilistic outcome of the Bayesian method.

6. CONCLUSION

In this paper we describe a benchmark case study which utilizes a multiphase facility for the development and validation of monitoring algorithms. A variety of data, including regular process measurements, high frequency signals, alarm and event logs, and videos, were collected. A multivariate statistical process monitoring algorithm and a Bayesian method for condition monitoring are given as illustrative examples demonstrating how the benchmark data may be used for validating monitoring algorithms. The following data analysis and algorithm validation tasks are anticipated in the future: validation of process and condition monitoring algorithms; integration of data from disparate sources for a synthesized monitoring framework; application of signal processing and pattern recognition techniques to the dataset; development and validation of fault detection, fault identification, fault classification, fault severity detection, monitoring of fault evolution and prognostics methods using heterogeneous data.

ACKNOWLEDGEMENT

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 675215, for which the authors are grateful. The authors are thankful for Liyun Lao, Stan Collins, Henry Tandoh, Prafull Sharma and Godfrey Nnabuife for providing helpful insight to the design and the implementation of the experiment.

REFERENCES

Hou, L., Bergmann, N.W., 2012. Novel Industrial Wireless Sensor Networks for Machine Condition Monitoring and Fault Diagnosis. IEEE Trans. Instrum. Meas. 61, 2787–2798. doi:10.1109/TIM.2012.2200817

Jansen, F.E., Shoham, O., Taitel, Y., 1996. The elimination of severe slugging - Experiments and modeling. Int. J. Multiph. Flow 22, 1055–1072. doi:10.1016/0301-9322(96)00027-4

Lu, Z.-J., Xiang, Q., Xu, L., 2014. An Application Case Study on Multi-sensor Data fusion System for Intelligent Process Monitoring. Procedia CIRP 17, 721–725. doi:10.1016/J.PROCIR.2014.01.122

Lynnworth, L.C., 1989. Ultrasonic Measurements for Process Control. doi:10.1016/B978-0-12-460585-5.50012-6

Ricker, N.L., 1995. Optimal steady-state operation of the Tennessee Eastman challenge process. Comput. Chem. Eng. 19, 949–959. doi:16/0098-1354(94)00043-N

Ruiz-Cárcel, C., Cao, Y., Mba, D., Lao, L., Samuel, R.T., 2015. Statistical process monitoring of a multiphase flow facility. Control Eng. Pract. 42, 74–88. doi:10.1016/j.conengprac.2015.04.012

Ruiz-Cárcel, C., Jaramillo, V.H., Mba, D., Ottewill, J.R., Cao, Y., 2016. Combination of process and vibration data for improved condition monitoring of industrial systems working under variable operating conditions. Mech. Syst. Signal Process. 66–67, 699–714. doi:10.1016/j.ymssp.2015.05.018

Stief, A., Ottewill, J.R., Orkisz, M., Baranowski, J., 2017. Two stage data fusion of acoustic, electric and vibration signals for diagnosing faults in induction motors, in: Elektronika Ir Elektrotechnika, to Be Published.

Tan, R., Cao, Y., 2017. Contribution Plots based Fault Diagnosis of a Multiphase Flow Facility with PCA-enhanced Canonical Variate Analysis, in: The 23rd IEEE International Conference on Automation and Computing.