Ensemble learning in the estimation of flow types and velocities of individual phases in multiphase flow using non-intrusive accelerometers' and process pressure data

Yan, Ru¹; Viumdal, Håkon¹; Fjalestad, Kjetil²; Mylvaganam, Saba¹ ¹Department of Electrical engineering, Information Technology and Cybernetics, University of South-Eastern Norway ²Specialist Production Technology TPI OG SUB RPE PE, Equinor, Norway

This is an Accepted Manuscript of an article published by IEEE in 2022 IEEE Sensors Applications Symposium (SAS) on September 12, 2022, available online: <u>https://doi.org/10.1109/SAS54819.2022.9881352</u>

Yan, R., Viumdal, H., Fjalestad, K. & Mylvaganam, S. (2022, August 1-3). Ensemble learning in the estimation of flow types and velocities of individual phases in multiphase flow using non-intrusive accelerometers' and process pressure data [Conference presentation]. 2022 IEEE Sensors Applications Symposium (SAS), Sundsvall. <u>https://doi.org/10.1109/SAS54819.2022.9881352</u>

© 2022 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Ensemble learning in the estimation of flow types and velocities of individual phases in multiphase flow–using non-intrusive accelerometers' and process pressure data

Ru Yan Department of Electrical engineering, Information Technology and Cybernetics University of South-Eastern Norway Porsgrunn, Norway ru.yan@usn.no

Saba Mylvaganam Department of Electrical engineering, Information Technology and Cybernetics University of South-Eastern Norway Porsgrunn, Norway saba.mylvaganam@usn.no

Abstract-Multiphase flows with oil/ gas/ water are common in oil and gas industries. Accurately identifying flow types and estimating flow velocities of the individual phases are crucial for different purposes, such as observing the process status and providing inputs to control systems. This paper presents a solution for identifying flow contents and estimating flow rates in single-phase or each phase in multiphase flows by using pressure measurements and pipe vibrations caused by the flows. The necessary experiments were performed using the multiphase flow rig with three-inch diameter pipelines transporting natural gas, water, and crude oil in a closed loop with a separator tank as source and sink. A series of tree-based ensemble machine learning models have been developed and tested with the data collected from accelerometers, differential pressure transmitters, and upstream- and downstream pressure transmitters. With these inputs, the developed models can identify volume ratios of individual phases (such as water cut) and can estimate the flow velocity of each phase in the flow loop, including the open/close status of the choke valve. After describing briefly, the P&ID diagram of the multiphase flow rig, the paper focuses on exploratory data analysis of the data from three accelerometers and three pressure sensors using three submodels cascaded to perform ensemble learning.

Keywords—multiphase flow measurement, phase volumes, classification, flow velocity, accelerometer, pipe vibration, pressure, data fusion, machine learning, ensemble learning

I. INTRODUCTION

University of South-Eastern Norway (USN) has developed and tested different data fusion methods based on data from many different multiphase flow studies based on experimental campaigns at USN and Equinor. The work at USN using AI/ML work has been going on for more than a decade. The results here are presented as part of the contribution from the team at USN for the Self Adapting Model-based system for Process Autonomy (SAM) project.

Usage of non-invasive and non-intrusive sensors in process measurements have been discussed in pneumatic transport of particles, detection, and quantification of sand in

Håkon Viumdal Department of Electrical engineering, Information Technology and Cybernetics University of South-Eastern Norway Porsgrunn, Norway hakon.viumdal@usn.no Kjetil Fjalestad Specialist Production Technology TPI OG SUB RPE PE Equinor Porsgrunn, Norway kfja@equinor.com

oil and gas flow, [1-6] and even in water flow measurements in flexible fire hose used by fire fighters, as reported by experts from Engineering Laboratory of National Institute of Standards and Technology (NIST) in the USA, [7].

Since the development of sand detectors in the 1990s [1], clamp-on acoustic emission-based sensors have been used by many industrial actors for monitoring flow of different phases in different processes. Due to a plethora of clamp-on acoustic emission sensors in the oil and gas industries for detecting and quantifying formation sand in flow of oil and gas, there is an increasing focus on exploiting this technique for use in estimating the flow of different phases present in multiphase flow encountered in the production and transport of oil and gas. Based on this trend, Equinor has different activities exploring the possibilities of using acoustic emission and accelerometers in the estimation of flow of gas, oil, water, and their mixed/multiphase flow.

The models presented in this paper are based on flow induced vibration monitored by accelerometers and pressure data from Equinor's multiphase flow rig in Porsgrunn, Norway, and have the following as the main objectives: identifying flow types; estimating flow rates of individual phases and volume ratios.

Generally, the piezoelectric accelerometer is a noninvasive transducer widely used in various industrial fields. Usage of accelerometer data to study the velocity of singlephase flow is described in Medeiros et al. [8] and Penttine et al. [9] employing models with deep neural networks.

The data used for analysis and results presented in this paper are from extensive tests carried out using the multiphase flow rig at Equinor (Porsgrunn, Norway). The multiphase rig is built to generate and emulate single-phase and multiphase flows encountered during oil and the production and transport of oil and gas to facilitate new solutions relevant to the production and transport of oil and gas. Equinor's multiphase flow rig is run by one of the world's leading research centres in the field [10]. The rig includes modules that can be combined to enable the possibility to conduct tests and collect data according to specific requirements emulating real-world conditions.

II. METHODOLOGY

A. Experiments setup and data logging

The datasets used for developing models were logged from the tests conducted at the Equinor's multiphase flow rig. The inner diameter of the flow loop is three-inch. Natural gas, crude oil and water were the materials used. Among all the tests, the temperatures on upstream and downstream sides varied from about 65°C to 94°C. Total mass flow changed from approximately 1 ton per hour to 64 ton per hour.

Fig.1 is a simplified piping & instrumentation diagram (P&ID) of the test setup at the rig indicating flow direction, locations of choke valves and the sensors and instruments, from which measurements are selected as inputs to the data fusion models used in this study. As illustrated in Fig.1, the three explosion-proof piezoelectric accelerometers("ACC"), with a measuring range $\pm 2~000$ m/s², are installed along with the flow direction: accelerometers #1 and #2 are placed at the second and fourth bends; #3 is placed close to the choke valve at the outlet side. Differential pressure ("DP") measurements come from two Venturi meters installed in the second riser section and the last flow loop section (horizontal pipe). "PT1" and "PT2" are the upstream and downstream pressures.

The accelerometer sampling frequency is set at \sim 50 kHz. The acceleration data were logged continuously for about 5 minutes before the next series started during each run. Thus, there are about 15M data points in one acceleration measurement (raw data) from each test.

Other measurements, such as differential pressure and upstream and downstream pressure data, were logged as discrete process data. Thus, for all measurands excluding the accelerometers, only their average values were saved. Therefore, except for the data from the three accelerometers, each other measurements has only a single value per test in the datasets.



Fig. 1. Simplified P&ID of test setup at the multiphase flow rig, indicating the location of sensors used, choke valve and flow direction. "ACC": accelerometer, "PT": pressure, "DP": differential pressure, "HIC": choke valve position. Two pairs of DP cells are placed at the ends of venturi meters. Note the location of accelerometers at bends.

The whole datasets are logged from 117 test runs from 10 different types of choked or unchoked single-phase and multiphase flows: single-phase: Gas ("G"), Oil ("O"), Water ("W"); multiphase: Gas/ Oil ("GO"), Oil/ Water ("OW"), Gas/ Water ("GW"); choked: Oil/

Water choked ("OWC"), Gas/ Oil choked ("GOC"), .Oil choked ("OC").

B. Exploratory data analysis

While analysing the vibration data from the accelerometers, it is natural to study the distribution of their major frequency components. Fig.2 (a) presents the frequency distributions from the accelerometer #1 measurements among all the tests excluding the oil test ("O"). Due to the significantly shorter measurement duration, and fewer data points in the oil test, the results from "O" are presented separately.

The spectrums indicate that, in all types of choked flows, the primary vibration energy is located below \sim 1kHz. In oil choked "OC" flow, a part of high energy vibration is also situated between 2kHz - and 4kHz. For the unchoked flows, the main frequency components are below 100Hz, and a few others are located at around 500Hz, as shown in the "zoomed in" plot of Fig. 2(a) shown in Fig.2 (b).

In Fig. 3, the upper plots (a) present the difference between saved averaged differential pressure data "DP 2-1" and "DP 2-2" through the test. The lower plot (b) is the total volume flow rates at the liquid phase: oil (yellow) and water (blue). Comparison of these two plots indicates that liquid rates are highly correlated with the pressure differences between "DP 2-1" and "DP 2-2".



Fig. 2. (a) Full (0-20 kHz) and (b) zoom-in (0-1kHz) normalized frequency spectrums from the accelerometer #1 measurements. The x-axis presents frequency (Hz), and the y-axis presents the tests that were named, such as "Gxx", "Wxx", "GOWxx", etc., where "xx" represents test numbers. The white-coloured texts on the left side indicate the spectrum corresponding flow types with the different combinations of phases: gas ("G"), oil ("O") and water ("W") with or without choked ("C"). The colour represents the amplitude of a particular frequency

Besides the test from Gas flow ("G"), the two plots have very similar variation patterns during the tests. As a result, the pressure difference between "DP 2-1" and "DP 2-2" is selected as one of the inputs in the data fusion algorithms used to estimate liquid flow rates.

C. Extraction of main features

Referring to the frequency distribution shown in Figs 2(a) and 2(b), low frequency data from the accelerometers will be analysed as the starting point in our analysis.

Moreover, to let the model achieve a high tolerance, the overall values (root mean square (RMS) values) of the filtered data are used instead of finding the spectrum peaks that could be too sensitive to the conditions prevailing in the test environment.



Fig. 3. (a) Difference between the differential pressure "DP 2-1" and "DP 2-2" vs (b) liquid phase (oil and water) volume flow rate. Oil yellow and water blue. The y-axis presents the tests that were named, such as "Gxx", "Wxx", "GOWxx", etc. (note, due to the limited space, not all the test names are visible in the figure)

The results shown in Fig. 4 are the overall values from 1kHz filtered data from all three accelerometers, indicating the high possibility of using the overall value to identify flow types. While, at some clusters, the data from different flow types overlapped. Therefore, to avoid overlapping flow types in the plots generated, other process data, such as upstream and downstream pressure, are needed and will be used in our analysis.



Fig. 4. Overall values (RMS values) of filtered acceleration data (unit in G) from the three accelerometers. The dot colours vary according to their flow types. Each dot presented in the plot represents the RMS value from a 1-second acceleration measurement.

The raw data from three accelerometers are filtered six times separately using suitable cut-off frequencies to extract more independent features in the model presented here. By doing this, the raw accelerometer data can be divided into different sets corresponding to different frequency zones. And then, the overall values are calculated for each set.

D. Model Development

The model to be developed is expected to identify ten different flow types (with or without choked) and estimate the flow rates of gas, oil, and water separately using the accelerometer data and other possible measurements taken from the 117 tests. Comparing the number of expected model outputs with the possible number of inputs and total available training & test data samples gives one of the biggest challenges in this work, which is: too small data (Volume in ML/AI jargon)

The data resolution between the acceleration data and the other discrete process measurements is uneven. Therefore, a solution has been figured out to mitigate the challenge: Divide one acceleration measurement into several segments, i.e., one second per segment. Then, use one segment and other process measurements from the same test as one data sample. Thus, the saved available data volume for the training and testing of models will increase significantly. In addition, if more independent features can be extracted from acceleration data and used as inputs for our models, it could also help to improve the accuracy of our model even when operating with a smaller size of data samples.

However, two issues must be addressed: One is the similarity between each acceleration segment data. In the same test, the vibration/acceleration will have repetitions (similarities) during the whole measurement period due to its physical property. The idea is to increase the number of training data samples and improve the model accuracy. The remaining data samples from the same tests should not be used for model test purposes. (section II-D-(2) below explains how the data were prepared for training, validation, and testing). Another is the influence of the applied discrete process measurements. In many of these 'new' data samples, discrete process data remain the same.

1) Overview of model structure

From the data-exploratory observations, a model design strategy is proposed, as illustrated in Fig.5, showing the ensemble learning modules. In the complete model, there are three submodels:

- Submodel 1 (S1) is to identify flow types. Its output will be one of the inputs to submodel 2 (S2) and submodel 3 (S3).
- Submodel 2 (S2) is to estimate the gas and liquid rates. Its output will be one of the inputs to submodel 3 (S3).
- Submodel 3 (S3) to estimate the oil and water rates.



Fig. 5. Ensemble learning model design strategy as a cascade of three submodels, S1, S2, and S3. These submodels will be trained separately. Outputs from the total model – flow type, flow rates, individual phases, i.e., gas, oil and water.

With all three submodels in cascade as shown in Fig. 5, the total model's outputs will be the identified flow types and flow rate of each phase. In this work, submodels will be trained separately.

2) Setting up development and test sets

The whole dataset is split into three parts: Training set, Dev (development) set, and Test set. In the Dev set, there are two parts also:

- Part 1 (p1) is the remaining acceleration segments from the same test as in the Training set. This part of the data is used to find an optimised amount of training samples to ensure the model will have the same performance during the whole test period. In addition, this optimised number also will indicate that an idea training data set should contain the data from how many seconds or minutes measurements.
- Part 2 (p2) is for the data from the test that is 'unseen' during the model training process. This part of the dev data will be used to optimise the model hyperparameters.

All data in the Test-set are never "seen" by the trained model or involved in the model development process.

Fig. 6 gives an example of how the data from the oil/water choked ("OWC") flow is split into different data sets. As a result, 10 - 20 % of total tests per flow type will be selected for the Dev set (p2). Furthermore, it will ensure at least one test sample from each flow type will be picked for the Dev set (p2) since there are only i.e. five tests available for gas/water flow. For the flow type with more than ten tests, one test will be picked for the test set.



Fig. 6. Example of splitting "OWC" flow data into Training, Dev, and Test sets. The x-axis represents time. The y-axis shows all the tests from the same type of flow, here "OWC" Test 1 (t1) to Test 19 (t19).

As a result, the datasets from 99 off tests are for training, the data sets from 13 off tests are for Dev set (P2), and the remaining datasets from 5 off tests are for Test-set.

The acceleration data are not used for the two submodels, S2 and S3, that estimate flow rates. Instead, a numeric number representing the flow types will be used as input. One reason for doing this is because S2 and S3 are the regression models. Therefore, only one reference flow rate data was saved for each phase in each test. Therefore, applying the same idea for flow type identification will not benefit the model.

3) Development environment

The programming language used is Python 3.8. The main applied software tool packages are *scikit-learn* for model development, *matplotlib* and *pyecharts* for plotting.

4) Development of submodel 1 (S1) to identify flow types *a*) Model structure

The block diagram in Fig. 7 illustrates the structure of the submodel (S1). A random forest classifier (RFC) algorithm is applied to identify ten different flow types. Besides, S1 also has a "Feature extractor" to extract features explained in sections II-B and C.

a) Optimising the length of the training set and the number of estimators in the Random Forest Classifier

For the flow type identifier, a set of continuous acceleration measurements are used for training. For instance, the first 30 seconds are used to train the remaining data to validate the model accuracy from the same test, which is the first part of the Dev set (p1). In submodel S1, the length of the data set used in training is set to 75 data samples (Equivalent to 75 seconds of continuous measurements), and the number of estimators in the RFC is set to120.



Fig. 7. Structure of submodel (S1) - identify flow types, 10 types shown in the last block in the diagram. Inputs are the time-series data from the three accelerometers ("ACC #1", "ACC #2" and "ACC #3"), upstream pressure ("Pupstream") and differential between upstream- and downstream pressure ("Pupstream - Pdownstream"). Refer to Fig. 1 for the details. "GOW", "OWC", etc., are flow types presented in Section II-A.

5) Development of submodel 2 (S2) to estimate the gas rate and the liquid rate

The block diagram in Fig. 8 illustrates the structure of the submodel (S2). A random forest regressor (RFR) algorithm is applied to estimate the flow rates. Similar to S1, S2 has its own "Feature extractor" to extract features from differential pressures. In addition, the output from S1- identified flow type will be input to S2. In the S2, the number of estimators in the RFR is set to140.



Fig. 8. Structure of submodel (S2) – Estimating flow rates of gas and liquid. Inputs are the flow type (the output from the submodel S1) and data from the differential pressure readings ("DP1-1", " DP1-2", " DP2-1" and " DP2-2").

6) Development of submodel 3 (S3) to estimate the flow rates of oil and water

Fig. 9 shows the inputs and the output for the submodel (S3). Gas flow rate, liquid flow rate, differential pressure measurements, and upstream and downstream pressure will be used as inputs. Since the liquid rate is a known value, the model output is set to the ratio of water rate to liquid rate to reduce the complexity of model computation.

In all flow types, there are mainly two (or three, if choked flow counted) flow types we need to take into consideration: "GOW" and "OW" (including "OWC"). Therefore, the S3 submodel has two base models: a Bagging regressor for "GOW" flow; a Voting regressor for "OW" flow. The selected base regressor is the Decision tree regressor in the bagging regressor. In addition, the Voting regressor contains a random forest regressor and a gradient boosting regressor as the base estimator. In S3, the number of estimators in the Bagging regressor is set to 300.



Fig. 9. Structure of submodel (S3) - Estimating the oil rate and the water rate. Inputs are the flow rates of gas and liquid (the outputs from the submodel S2), data from the differential pressure readings ("DP1-1", "DP1-2", "DP2-1" and "DP2-2") and upstream- and downstream pressure ("P_{upstream}", "P_{downstream}",)

III. RESULTS

A. Results of flow type identification (from model S1)

The results of flow type identification from development data (p1 and p2) and test data are given in the confusion matrixes shown in Fig.10 (a), (b), and (c) separately. In these confusion matrixes, the targets (ground truth) flow types are given in the direction of the y- axis; the submodel, S1, outputs (identification results) are given the direction of the x-axis. The numbers shown in each grid at the confusion matrixes are the corresponding number of classified samples.

From the results, the identification of most types of flow is 100% correct. Only 2 out of 299 samples (less than 0.7%) from gas/water flow were misclassified as gas flow.

B. Results of estimate the gas flow rate and the liquid phase flow rate (from model S2)

The correlation between estimated and actual rates for gas and liquid (water and oil) is given in Fig. 11 and Fig. 12. The correlation coefficient, r, for gas rate estimating is 0.98 with 95% confidence intervals (CI95%) between 0.98 and 0.99. Correspondingly, for estimating the liquid flow rate, r is 0.99, and CI95% is between 0.99 and 1.

The root mean square error (RMSE) values between actual and estimated gas flow rates are: 7.51 on training data, 14.84 on development data and 1.63 on test data (with unit m^3/h). For liquid estimating: 1.69 on training data, 3.75 on development data and 1.58 on test data (with unit m^3/h).

C. Results of estimate the water-cut in oil/water and gas/oil/water flow (from model S3)

The correlation between estimated and actual water-cuts in oil/water and gas/oil/water flow are shown separately in Fig. 13 and Fig. 14. The correlation coefficient, r, for watercut at oil/ water flow is 0.98 with 95% confidence intervals (CI95%) between 0.91 and 0.98. For the water-cut in gas/ oil/ water flow, r is 0.95, and CI95% is 0.89 and 0.98.

The root mean square error (RMSE) values between actual and estimated water-cut in oil/ water flow are: 4.15 on training data, 1.8 on development data and 8.89 on test data (with unit





Fig. 10. Results of identification of flow types based on phases present (Submodel S1)



Fig. 11. Regression plot of estimating flow rate using submodel S2 for gas (single phase). The correlation coefficient, *r*, is 0.98



Fig. 13. Regression plot of estimating water-cut in oil/ water (two phase) flow using submodel, S3. The correlation coefficient, *r*, is 0.96

IV. CONCLUSION

A cascaded (serial) tree-based model is developed and tested using the saved pipe vibration data (acceleration data), differential pressures, and upstream and downstream pressure for the ensemble learning techniques used in the exploratory data analysis presented in this paper. The complete model consists of three submodels connected sequentially that the output(s) from the previous submodel is used as input to current submodels. From the above-presented results, it has been shown that the model has an excellent possibility to identify flow types (flow contents and if choked or not) and estimate volume flow rates of gas, oil and water using only one-second acceleration data together with discredited pressure data, as mentioned above. However, the reproducibility and generalisation ability of the model needs to be further studied by performing new tests with more variations in phase fractions and their flow rates. Further to study the influence of the discrete process data (pressure measurements), a higher data logging frequency for these data is expected while performing new tests.

ACKNOWLEDGEMENT

Multiphase tests with applied sensors and instrumentations, including the accelerometers, pressure transmitters, etc., were taken out at the Equinor test site (Plab) at Herøya Industrial Park (Porsgrunn, Norway). The tests campaign was partly financed by the SAM project supported by the Research Council of Norway and the process industries (Project No. 295945).

REFERENCES

 T. Folkestad, and K. S. Mylvaganam, "Acoustic measurements detect sand in North Sea flow lines," Oil & Gas Journal (USA), vol.: 88:35, pp. 33-39, August 1990.



Fig. 12. Regression plot of estimating flow rate using submodel S2 for liquid phase (oil and water, two phases). The correlation coefficient, r, is 0.99



- Fig. 14. Regression plot of estimating water-cut in gas/ oil/ water (all three phases) flow using submodel S3. The correlation coefficient, *r*, is 0.95
- [2] Y. K. Kim and Y. H. Kim, "A three accelerometer method for the measurement of flow rate in pipe," Journal of the acoustical society of America, vol.100, no.2, pp.717 – 726, August 1996.
- [3] R. Hou, A. Hunt, and R. A. Williams, "Acoustic monitoring of pipeline flows: particulate slurries," Powder Technology, vol. 106 (1-2), pp30-36, November 1999.
- [4] L. Fabbiano, G. Vacca, G. Dinardo, "Fluid Flow Rate Estimation using Acceleration Sensors," Proceedings of the International Conference on Sensing Technology (ICST), pp 221–5, December 2013.
- [5] S. Al-lababidi, A. Addali, H. Yeung and F. Khan, "Gas Void Fraction Measurement in Two-Phase Gas/Liquid Slug Flow Using Acoustic Emission Technology," Journal of Vibration and Acoustics, vol. 131 (6): 064501, November 2009, DOI:10.1115/1.4000463
- [6] K. A. R. Medeiros, C. R. H. Barbosa and E. C. de Oliveira, "Flow measurement by piezoelectric accelerometers: Application in the oil industry." Petroleum Science and Technology, vol.33(13–14), pp.1402–1409, September 2015
- [7] C. U. Brown, G.W. Vogl and W.C. Tam, "Measuring Water Flow Rate for a Fire Hose Using a Wireless Sensor Network for Smart Fire Fighting," Fire Technol vol. 57, pp. 3125–3150, November 2021, https://doi.org/10.1007/s10694-020-01054-1
- [8] K. A. R. Medeiros, C. R. H. Barbosa, É. C. de Oliveira, "Flow Measurement by Piezoelectric Accelerometers: Application in the Oil Industry," Petroleum Science and Technology, vol. 33, Issue 13-14, pp. 1402-1409, Sep 2015. [Online]. Available: <u>http://dx.doi.org/10.1080/10916466.2015.1044613</u> [Accessed 6 June 2022].
- [9] O. Penttinen et al, "Towards flow measurement with passive accelerometers," Flow Measurement and Instrumentation, Vol. 80, 2021, 101992, ISSN 0955-5986, [Online]. Available: <u>https://doi.org/10.1016/j.flowmeasinst.2021.101992</u>. [Accessed on 06 June 2022]
- [10] O. B. Ulsnæs, The multiphase rig Statoil research facility flagship at Herøya, March. 2018. [Online]. Available: <u>https://eng.heroyaindustripark.no/latest-news/the-multiphase-rig-statoil-researchfacility-flagship-at-heroeya</u> [Accessed 02 April 2022].