Classification of Gases and Estimation of Gas Flow Rate Based on Unsupervised and Supervised Learning Respectively

Maths Halstensen[1]   Simen S. Halstensen[2]

[1]Department of Electrical Engineering, IT and Cybernetics, University of South-Eastern Norway, maths.halstensen@usn.no
[2]Department of Electrical Engineering, IT and Cybernetics, University of South-Eastern Norway, simenshalstensen@gmail.com

Abstract

In this research project, acoustic chemometrics was assessed as a new method for both classification and prediction of flow rate of five selected gas types. The gas types were selected to span different densities as much as possible while at the same time being relatively safe to use. The five gas types were Argon, Helium, Carbon dioxide, Nitrogen and Air. The research questions were 1) Can measurements of the vibrations in a gas control valve in combination with signal processing and unsupervised learning be used to classify the five gases mentioned above? 2) Can the vibrations in the gas valve in combination with supervised learning be used to determine the flow rate of the five gases? 3) Can a simple low cost piezo disk provide signals comparable to that of an industrial accelerometer? The results show that it is possible to classify the five gas types based on principal component analysis with three components. The gas flow rate could also be predicted for all five gases based on partial least squares regression with an average error of 2-5%. The Piezo disk could not be used for gas classification, but for prediction of gas flow rates it was comparable to the accelerometer. All the prediction models were validated based on independent data.

Keywords: *supervised learning, unsupervised learning, partial least squares regression, classification, principal component analysis, accelerometer*

1 Introduction

Acoustic chemometrics (Halstensen, 2006, 2010; Esbensen, 1999; Arvoh, 2012) is a relatively new general process analytic approach for on-line monitoring of industrial processes. Acoustic chemometrics concerns capturing characteristic system vibrations generated by e.g. a transportation or manufacturing process.

The vibrations -or acoustic signatures -can easily be measured by a non-intrusive sensor (accelerometer).

Previous research has shown that these acoustic signatures carry embedded information about physical and chemical parameters, such as composition (oil, fat, ammonia, buttermilk, glycol, and ethanol), mixing progress, fiber length, flow, density, temperature - as well as system state. For extraction and quantification of these types of specific analytes and parameters of interest, domain transforms (Fast Fourier Transform) (Ifeachor & Jervis, 1993) is essential prior to multivariate analysis and modelling.

Unsupervised methods such as Principal Component Analysis (PCA) (Esbensen and Swarbrick, 2018) can be used to e.g. classify products. Supervised methods such as Partial Least Squares (PLS) (Martens & Næs, 1989) regression based on empirical input-response data are used to extract relevant information and to calibrate multivariate regression models. The models can then be used either to classify or to predict the parameter of interest based on mew independent acoustic spectra. As with all multivariate models, the role of proper validation is critical (Esbensen and Geladi, 2010).

Advantages of acoustic chemometrics:
1. clamp on non-intrusive sensor
2. No process modifications necessary
3. Several predictions from the same measurement

The fact that almost all processes produce some kind of acoustic emission opens up for many potential applications, all of which depend on multivariate calibration. The only requirement is that the acoustic emission must contain relevant information, which can be correlated with the parameter(s) of interest. In this work, we test the feasibility of acoustic chemometrics to both classify gas type and predict gas flow rate. The method is demonstrated on five different gas types: Argon, Helium, Carbon dioxide, Nitrogen and Air, which is a mixture.

The proposed method cannot be directly compared to the accuracy of e.g. Coriolis type flow meters (LaNasa & Upp, 2014), but the advantage of this method is that it is non-intrusive, inexpensive, easy to implement on exiting equipment (no interference with existing

equipment is necessary). The industry often prefers non-invasive or non-intrusive measurement methods since parts sticking into the process pipelines can cause critical fouling and clogging which is problematic but also leads to faulty measurements.

The main objective in this research project is to test a new method for gas flow metering and show that it works on different gas types. The second objective is to show that the same method also can be used to classify the gas type directly from the same raw measurement as is used for the flow rate estimation. The last objective is to validate the method based on independent experimental data to ensure optimal model complexity and estimation of the uncertainty of the method.

Very little causal interpretation of the acoustic spectra is reported in literature. In a study from 2008 (Kupyna et al., 2008] tried to explain the physical theory behind the flow of liquids through an orifice plate and the vibrations that occurred. The experimental results agreed with theory that they found, but it is difficult to see how it can be directly used to improve acoustic chemometrics. We have not attempted to explain the physical theory behind the vibrations that occur in the valve from the different flowing gases in this study.

2 Materials and methods

The experiments were carried out on an experimental gas flow rig in the chemometrics laboratory at University of South-Eastern Norway, Porsgrunn, Norway.
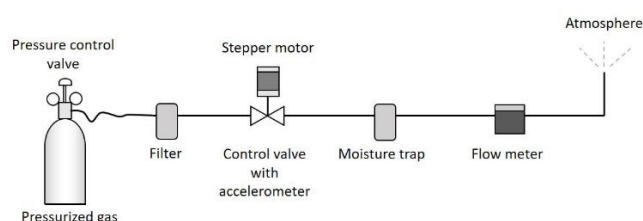
2.1 The experimental gas flow rig

Figure 1. Experimental gas flow rig at USN.

All the experiments with gas flow were carried out in the chemometrics laboratory at the University of South-Eastern Norway, Porsgrunn, Norway (USN).

The experimental rig used to develop the method is shown in Figure 1 above. The rig consists of a pressurized gas bottle, a filter, a gas control valve, a moisture trap and a gas flow meter. All the elements are connected with hoses and pipes, which were pressure-tested to avoid gas leaks in the system prior to the gas flow experiments. The filter is there to avoid getting particles into the control valve. These particles can come from connecting and disconnecting the hoses and clamps at around the gas bottle.

The gas control valve is a simple rotary gas valve used in compressed air systems. A stepper motor is

connected directly to the valve to be able to control the gas flow rate automatically. The control loop was implemented in LABVIEW 2017 from National Instruments. A large moisture trap/filter normally used for spray painting is located right in front of the gas flow meter to avoid moisture entering the flowmeter. Moisture can damage the gas flow meter or cause false readings.

The gas flow meter Mass-Stream D-6300 from BronkhorstTM has a gas flow range up to 200 l/min of air. Since we here have other gas types a conversion factor was necessary to get the correct flow rates for all other gases than air. The accuracy of the flow meter is +/- 1 %.

The gas leaving the gas flow meter was injected into the ventilation system and finally disposed to the atmosphere. All the elements mentioned above was connected with hose with inner diameter = 10mm. On the gas control valve shown in Figure 2 the two sensors (accelerometer and piezo disk) were glued on to the main metal body of the valve.

Figure 2. Gas control valve with the stepper motor connected from top.

2.2 Gas types

Five different gases were selected: air, argon, carbon dioxide, nitrogen and helium. These gases were selected because we wanted to span the densities as much as possible but at the same time, we wanted to use gases, which are relatively safe. Air is a mixture containing mostly nitrogen (78%) and oxygen (21%) so in the gas classification it should behave similar to nitrogen.

2.3 Acoustic sensors S1 and S2

S1 is an industrial standard accelerometer with a frequency range of 0-40kHz, S2 is a simple 0.5mm thick piezo electric element. Figure 3 shows a picture of the two vibrational sensors S1 and S2.

Figure 3. Vibrational sensors, S1 (left) is an accelerometer from Endevco 7259B, S2 (right) is a simple low cost ($1) piezo electric disk with resonant frequency 4 kHz.

Two different sensors were mounted in order to test if the results from a low cost piezo element (S2) would be comparable to the results based on the much more expensive accelerometer (S1).

The measurement systems also consist of a Signal Amplifying Module (SAM) as shown in figure 4 and a data acquisition system from National Instruments (USB-6351) connected to a computer. The SAM unit is a special experimental equipment developed and built at USN.

Figure 4. Signal Amplifying Module (SAM)

The signal from the accelerometer was amplified 100 times, while the signal from the piezo disk was only amplified 40 times.

2.4 Signal processing

The signals from the sensors S1 and S2 were converted from time to frequency domain using Fourier transformation (Ifeachor & Jervis, 1993) prior to analysis and model calibration. The Discrete Fourier transform (DFT) can be expressed as

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi kn/N} \quad k = 0, \dots, N-1 \quad (1)$$

A more efficient implementation of the DFT is the Fast Fourier Transform (FFT), which in this work has been implemented in LabVIEW 2017 for fast real time calculation of the Fourier spectrum.

Each measurement, which was a time series of 4096 samples recorded with a sampling frequency of 300 kHz, were transformed using Fast Fourier Transform. The resulting spectrum contains 2048 frequencies.

2.5 Principal Component Analysis

Principal component analysis (PCA) is a standard multivariate data analysis method (Esbensen and Swarbrick, 2018), PCA uses an orthogonal transformation to convert a set of observations often described by many correlated variables into a few linearly uncorrelated latent variables called principal components. Principal Component Analysis (PCA) which is a so-called unsupervised method can be used for classification of observations. It is called unsupervised because of the lack of reference data to guide the decomposition. The decomposition of a multivariate matrix X is shown in equation 2.

$$X = T \, P^T + E \quad (2)$$

The T matrix contains the score vectors, P contains the loading vectors and E is the residual.

In this work, the NIPALS algorithm (Ergon et al., 2009) was used because of one of its advantages, which is that it is possible to calculate only a relevant number of components. Principal component analysis can alternatively be carried out based on singular value decomposition (SVD). The advantage of using SVD is that the code is simple, the disadvantage is that all possible components always is calculated so if X is large (I.e. is made up of a high number of variables), SVD is time consuming. NIPALS also works on matrices with missing values, in which case SVD will fail.

2.6 Partial Least Squares Regression

Partial least squares regression PLS-R was used to calibrate multivariate prediction models for flow rate of all five gas types. The reference measurements from the flow meter in the experimental flow loop was used as the reference during calibration.

Partial Least Squares Regression (PLS-R) is a supervised modelling approach which is well explained in literature (Martens & Næs, 1989; Ergon et al., 2009) thus only a short introduction is given here.

PLS-R relies on representative training data for two variable blocks, often called X and Y respectively. In the present study, the X data matrix contains the acoustic frequency spectra, and Y is a vector containing the gas flow rates from the gas flow meter used as a reference. The NIPALS algorithm is the most widely used algorithm in PLS regression. In this algorithm, the intention is to model both X and y simultaneously, make the error as small as possible and at the same time extract as much useful information from the X matrix in order to describe the y response variable. A simplified version of the NIPALS algorithm is presented below (Ergon et al., 2009).

A is the optimal number of components in the model.

1. Let $X_0 = X$. For a = 1, 2,..., A perform steps 2 to 6
2. $w_a = X_{a-1}^T y / \|X_{a-1}^T y\|$ (with length 1)
3. $t_a = X_{a-1} w_a$
4. $q_a = y^T t_a (t_a^T t_a)^{-1}$
5. $p_a = X_{a-1}^T t_a (t_a^T t_a)^{-1}$
6. Compute the residual $X_a = X_{a-1} - t_a p_a^T$

$$X = T_w P^T W W^T + E \qquad (3)$$
$$y = T_w q_w + f \qquad (4)$$

where the score matrix $T_w = [t_1 \quad t_2 \cdots \quad t_A]$ is orthogonal, loadings matrix $P = [p_1 \quad p_2 \cdots \quad p_A]$, $q_w = [q_1 \quad q_2 \cdots \quad q_A]$ and the loading weight matrix $W = [w_1 \quad w_2 \cdots \quad w_A]$

The loading matrix, P, is calculated as
$$P = X^T T (T^T T)^{-1} \qquad (5)$$

The prediction vector for $y = Xb + f$ corresponds to:
$$\hat{b} = W(W^T X^T X W)^{-1} W^T X^T y \qquad (6)$$

The response vector $\hat{y} = X\hat{b}$ \qquad (7)

In evaluating the regression model, the root mean squared error of prediction RMSEP offset, slope and correlation coefficient are commonly used. The root mean squared error of prediction is calculated as:

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_{i,\text{predicted}} - y_{i,\text{reference}})^2}{n}} \qquad (8)$$

3 Experimental

The gas bottle was connected to a constant pressure valve, which were connected to the experimental gas flow rig. The constant pressure valve was adjusted to four bar. The set point of the gas control valve was entered and the gas started to flow. When the gas flow rate became stabile after a few seconds, the acoustic measurement was started. The gas flow rate was simultaneously measured and recorded by the flow meter in figure 1. The gas flow rate levels were chosen randomly until all the 21 discrete levels from 1 to 100 l/min were covered. The discrete gas flow levels were 1, 5, 10, 15, 20, 25,..., 100 [l/min]. The experiment was repeated following the same strategy as in the first round to get independent data for validation of the models. This was the experimental procedure followed for all the five gas types.

In order for the gas flow meter to give reliable flow rate measurements for the gases different from air, a set of conversion factors were necessary to scale the signal to the correct level. The conversion factors were provided by the manufacturer of the flow meter. The conversion factors used can be seen in table 1.

Table 1. Gas flow rate conversion factors

Conversion factors for the five gas types measured by the flowmeter:	
Air:	1.00
Argon (Ar):	2.01
Carbon dioxide (CO2):	1.20
Helium (He):	0.24
Nitrogen (N2):	1.00

Seven replicates of acoustic spectra were recorded for each gas and flow rate. Each of these replicates was an average of 200 individual spectra. The averaging is necessary to reduce white noise and get a precise measurement. Each batch of seven replicates took approximately 35 seconds to record.

During the experiments on CO2, a significant amount of ice was building up around the valve on the gas bottle. This is normally not a big problem but in our case, it was problematic because we used hoses to transport the gases to the flow rig and these hoses became quite fragile at this low temperature. Fortunately, no hoses broke during the experiments.

4 Results and discussion

4.1 Classification of gas types (S1)

The score plot t1-t2 in Figure 5 shows that principal component 1 describes gas flow rate for all the five gas types, the gas flow rate increases from right to left in the score plot. The loading vector (Figure 5) for principal component 1 which indicates that all the frequencies in the acoustic frequency spectra increase with increasing gas flow rate since all frequencies have the similar loading value. All the results in Figure 5 is based on data from the accelerometer only.
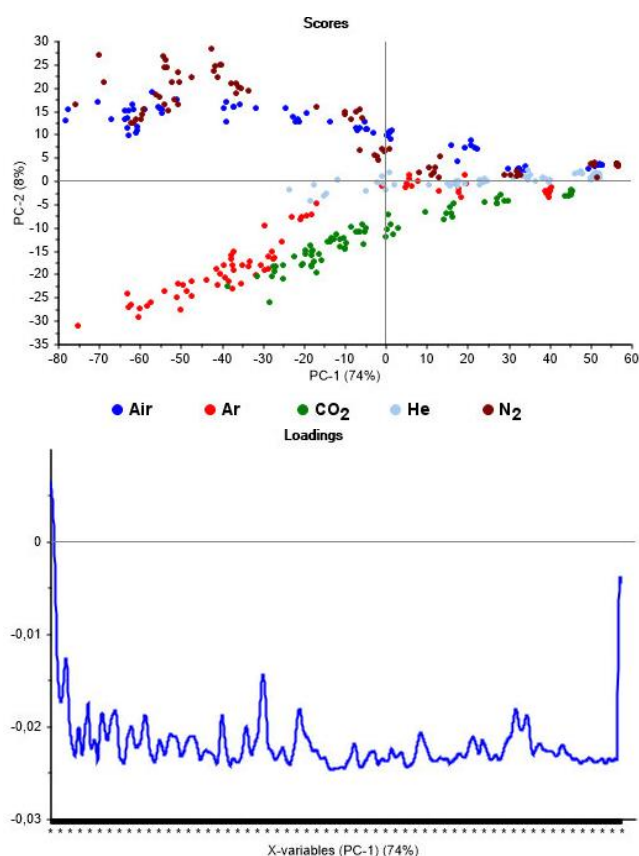
Figure 5. Upper: Score plot t1-t2, Lower: loading vector p1.

Figure 6 shows the classification of the five different gas types. The score plot in Figure 6 shows all the five gas types with different colors.

The score plot in Figure 6 above shows that the gas groups/clusters are located in a star pattern. It is obvious why it has to be exactly like this because "no flow" sounds exactly the same for any gas, namely no sound at all! The zero flow (0 l/min) is therefore the center of the "classification star". As expected, the nitrogen (N2) group overlaps significantly with the cluster for air because air contains 78% of nitrogen.

The Loadings for the first two components shows that there are relevant information in almost all the frequencies in the spectra.

The explained variance plot shows that most of the variance in the data is caused by variation in airflow rate as PC-1 explains 78% of the variance. PC-2 and PC-3, which is used for classification of the gases, only explains 8% and 4% respectively.

The PCA classification model was based on data from two independent experiments separated in time. An alternative strategy would be to determine the PCA classification model based on one of the experiments and used the other one for testing. Based on the results shown in the score plot in Figure 6 we have here only shown that classification of these five gases based on

vibration analysis is promising. Further validation of the concept is needed to verify reliability over time.
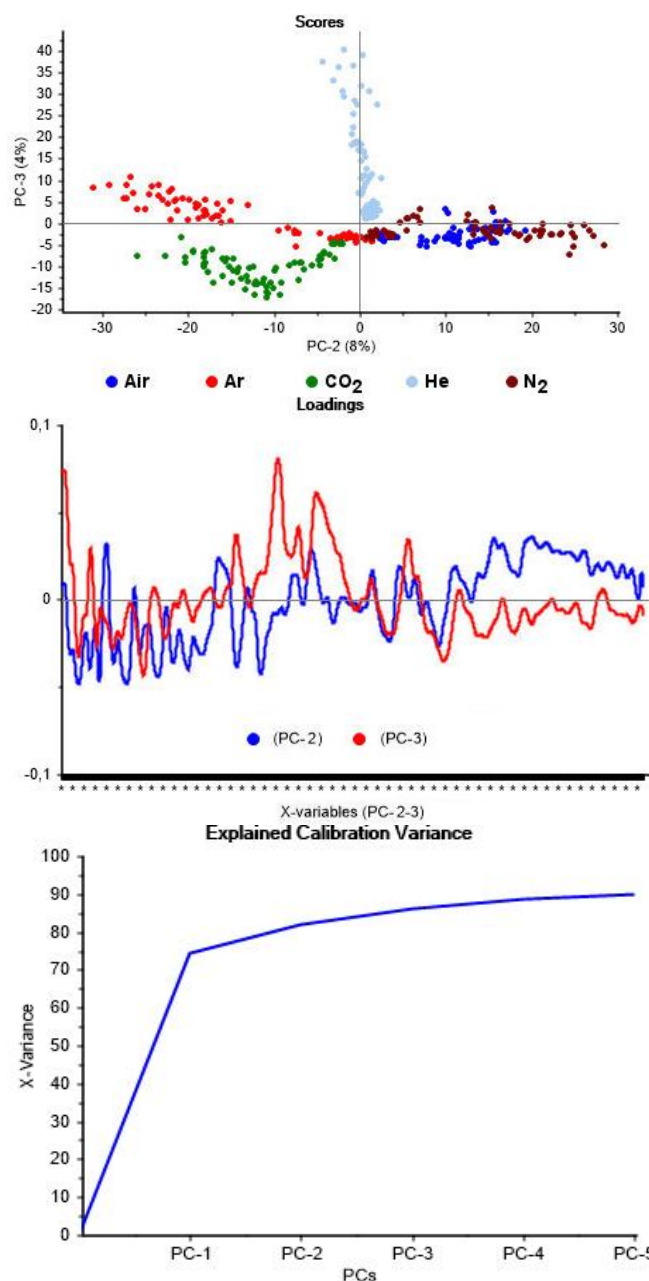

Figure 6. Classification of gases based on the accelerometer data, Upper: Score plot t2-t3, Middle: Loading vectors p2 and p3. Lower: Explained variance.

4.2 Classification of gas types (S2)

From the analysis of the data from the low cost piezo disk it became clear that it was not possible to completely separate and thus classify the five gases. Figure 7 shows the score plot t2-t3 of the piezo disk data, which is directly comparable to the score plot in figure 6 representing the accelerometer. As can be seen in figure 7 the data is more spread (noisy) leading to overlap between gas types. The PCA classification model was also here based on data from two

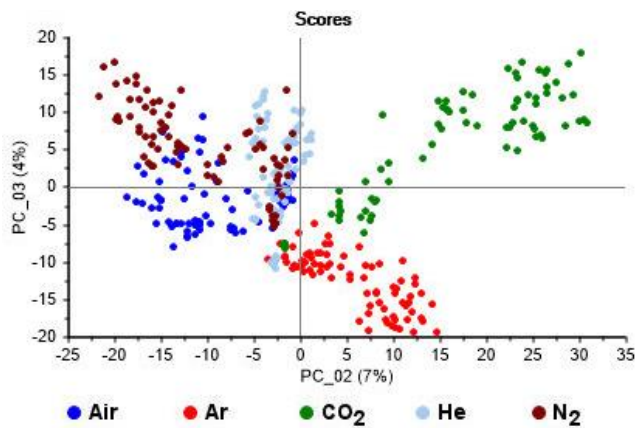independent experiments separated in time as described in Section 4.1

Figure 7. Classification of gases based on the piezo disk data, Score plot t2-t3.

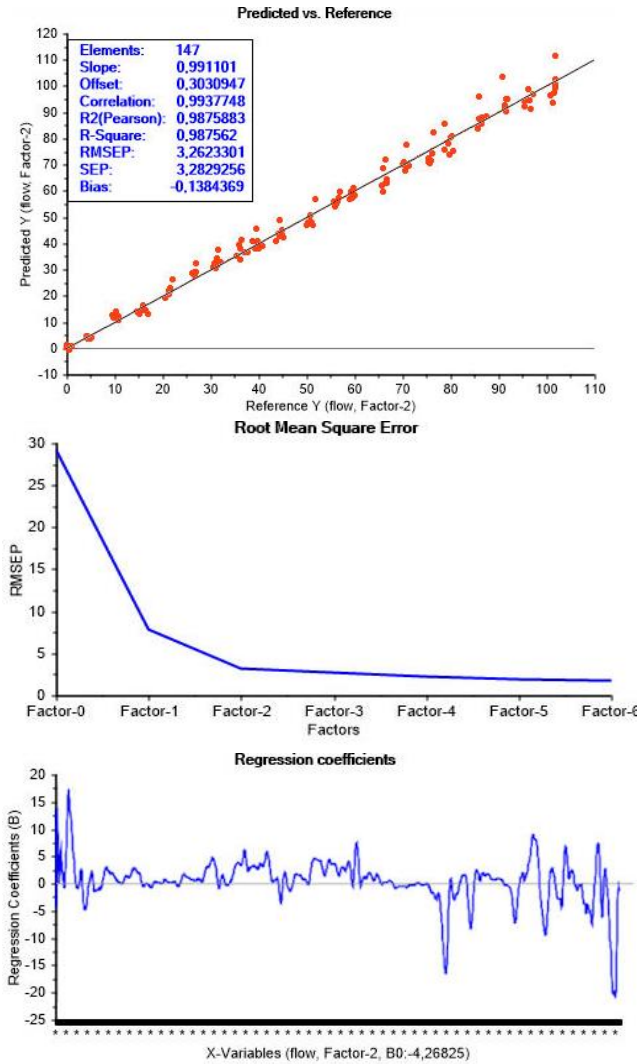## 4.3 Estimation of gas flow rate

Figure 8. PLS regression model for prediction of gas flow rate of Argon, 2 components was optimal and no outliers were deleted.

Five PLS-R models were calibrated for flow rate of the five different gas types using the acoustic data (spectra) as the independent X-matrix and the reference Y-vector was the flow rates from the gas flow meter.

The prediction performance of the five models were similar with comparable RMSEP values. Only the full model result for Argon is shown in Figure 8 while the results for all the 5 gas types has been summarized in table 2 for comparison. The model for argon flow rate in figure 8 is based on both sensors (Accelerometer and Piezo disk) combined.

Table 2 below shows a comparison of all the models for gas flowrate of all five gases, and all combinations of sensors.

Table 2. Comparison of gas flow rate models for Air, N2, CO2, He and Argon. The lowest RMSEP values of the models based on single sensors are indicated with bold numbers.

Gas type	Both sensors (4096 variables)			Accelerometer (2048 variables)			Piezo disk (2048 variables)		
	RMSEC	RMSEP	#Comp.	RMSEC	RMSEP	#Comp.	RMSEC	RMSEP	#Comp.
Argon	3.1583	3.2623	2	3.5545	3.8124	2	3.5275	**3.4026**	2
Air	2.8109	4.0350	2	3.1184	**4.1794**	2	3.0684	4.2792	2
CO2	3.3641	3.3641	3	3.8272	5.0107	3	3.8713	**4.6002**	3
He	2.2870	2.5896	4	1.6909	**2.0514**	3	4.8856	4.9125	4
N2	1.8420	3.5744	2	2.3368	3.6871	2	1.7049	**3.6106**	2

A comparison of the RMSEP values of the models based on single sensors shows that the sensors performs very similar and both can be used for prediction of gas flow rate. However, if classification of gas type is of interest it is only the accelerometer, which can be used. The loading plots in Figure 5 and 6 and the regression coefficients in Figure 8 shows that there is information in all frequencies in the acoustic spectra, an attempt to use variable selection where some of the variables are omitted in the model will fail because of this. In order to improve the classification or prediction of gas flowrate, the most effective method is to use moving average either in the frequency direction or in the time direction. Filtering in the time direction will often result in a lower RMSEP but since we introduce a low-pass filter, the model will not be able to detect rapid changes of the parameter of interest.

5 Conclusion

A new method for non-intrusive gas classification and flow rate estimation is proposed. The research objectives was 1) assess acoustic chemometrics potential to classify 5 selected gases, 2) predict the flow rate of each gas in the range 0-100 l/min and 3) compare results from a standard accelerometer and a simple low cost piezo disk. The results in section 4 shows a successful classification of the gas types based on PCA with three components on the measurements from the accelerometer. The results also show that it is possible to predict gas flowrate based on vibrations in the gas

flow rate valve. The RMSEP values shows an average error between 2 and 5% for the gas flow rate models when tested on independent data. From the comparison of the low cost piezo disk and the accelerometer, it was clear that the piezo disk could not be used for gas classification, but for prediction of gas flow rates it was comparable to the accelerometer.

References

B. K. Arvoh, S. Asdahl, K. Rabe, R. Ergon, and M. Halstensen. Online estimation of reject gas and liquid flow rates in compact flotation units for produced water treatment. *Flow Meas. Instrum.*, 24: 63–70, 2012. doi:10.1016/j.flowmeasinst.2012.03.008

R. Ergon. Re-interpretation of NIPALS results solves PLSR inconsistency problems. *J Chemometrics,* 23: 72-75, 2009. doi:10.1002/cem.1180

K. H. Esbensen and B. Swarbrick, Multivariate Data *Analysis – An introduction to Multivariate Analysis, Process Analytical Technology and Quality by design*, 6th Ed., *CAMO Publishing*, 2018. ISBN 078-82-691104-0-1

K. H. Esbensen and P. Geladi. Principles of Proper Validation: use and abuse of re-sampling for validation. *Journal of Chemometrics,* 24: 168-187, 2010. DOI: 10.1002/cem.1310

K. H. Esbensen, B. Hope, T. T. Lied, M. Halstensen, T. Gravermoen, and K. Sundberg. Acoustic chemometrics for fluid flow quantifications: II. A small constriction will go a long way. *Journal of Chemometrics,* 13 (27): 209–236, 1999. doi: 10.1002/ (SICI) 1099-128X (199905/08)13:3/4b209::AID-CEM553>3.0.CO;2–5.

M. Halstensen and K. H. Esbensen. Acoustic chemometric monitoring of industrial production processes in: K.A. Bakeev, *Process Analytical Technology,* Wiley, Chichester, United Kingdom, 2, 281–302, 2010. doi:10.1002/9780470689592, ISBN: 978-0-470-72207-7

M. Halstensen, P. de Bakker, and K. H. Esbensen. Acoustic chemometric monitoring of an industrial granulation production process—a PAT feasibility study. *Chemometr. Intell. Lab Syst.,* 84: 88–97, 2006. doi:10.1016/j.chemolab.2006.05.012

E. Ifeachor and B. Jervis, *Digital Signal Processing, a Practical Approach*, *Addison-Wesley Publishing,* 1993. ISBN-13: 978-0201596199

A. Kupyna, R. B. Schüller, E-O. Rukke, and T. Isaksson. Acoustic chemometrics on liquid flow: Shift in the frequency spectra and its relationship to the physical properties of the liquid and the pipe. *Chemometr. Intell. Lab. Syst.,* 91 (2): 151-163, 2008. doi:10.1016/j.chemolab.2007.10.010

E. L. Upp and P. J. LaNasa. *Fluid Flow Measurement: A Practical Guide to Accurate Flow Measurement. Gulf Professional Publishing,* 2002. **ISBN:** 9780884157588

H. Martens and T. Næs. *Multivariate Calibration*, *Wiley*, Chichester, UK, 1989 reprint 1994. ISBN-10: 0471930474, ISBN-13: 978-0471930471