

Dimensionality Reduction and Pattern Recognition of Flow Regime Using Acoustic Data

Nafiseh Vahabi

Department of Electronic
and Electrical Engineering
University College London
UK, London
Email: uceevah@ucl.ac.uk

Dr. David R. Selviah

Department of Electronic
and Electrical Engineering
University College London
UK, London
Email: d.selviah@ucl.ac.uk

Abstract—In this study we investigated the novel application of Principle Component Analysis (PCA) in order to reduce the dimensionality of acoustic data. The acoustic data are recorded by fibre optic distributed acoustic sensors which are attached along 4000 (m) pipe with frequency of 10 (kHz) and for a period of 24 hours. Data are collected from distributed acoustic sensors are very large and we need to identify the area that contains the most informative signals. The algorithm is applied to water, oil and gas dataset. We aimed to form the smaller dataset which preserves the pattern of original dataset but it is more efficient and less expensive for further analysis. The result of this study will lead to automation of multiphase flow pattern recognition for oil and gas industry application.

Index Terms—Feature extraction, Principle component analysis (PCA), Signal processing, Dimension reduction

I. INTRODUCTION

Monitoring in-well flow was a fundamental subject for oil industry to manage the oil field. Flow surveillance can help identify the region of unusual condition in the pipes to optimise and maximise the oil production [1]. Furthermore, effective flow monitoring results in low cost intervention which means material cost savings are achievable by the oil companies. Flow regime recognition techniques and devices became even more desirable for oil and gas industries as they need to separate unprocessed multiphase flow materials [2]. In addition, receiving real time flow information in downhole flow for injection wells enables the changing of the injection rate appropriately which improves the reservoir life time and efficiency of oil producers. The other potential applications are leak detection, monitoring the condition of flow control devices in downhole and identifying the gas breakthrough [3].

Many investigations performed and techniques developed by experts to overcome the challenge exists to process oil and gas data [4]. The ultrasonic techniques are one of the popular methods for measuring the flow meter and to extract the required information from the flow in the oil industry [4]. These methods have the advantage of inexpensive, non-invasive and robust although the signal needs to calibrate before any measurement is taken. Artificial Neural Network (ANN) is proposed as an alternative solution to overcome the difficulties of prior methods because it improves the precision

of flow measurement and in some cases no pre-processing is required [6], [7]. Ultrasonic attenuation data fed into the neural network for training and learning the multiphase flow pattern. In recent work [8] the use of Artificial Neural network with dual energy fan-beam gamma-ray attenuation technique increased the accuracy of oil, gas and water classification with mean absolute error of less than 5.68% . They used radial basis function to train the ANN and classify the flow regime by selecting just four features. The methodology was developed to recognise the pattern in gamma-ray pulse height distributions using ANN to identify the flow regime in multiphase flow [9]. This method has also been developed using ANN to describe the details of two phase flow regimes [10]. The network trained by Levenberg-Marquardt algorithm is based on the competing forces occurring within the multiphase flow. A similar procedure was followed by another research group [11] when they developed a new multiphase flow metering device for real time multi phase flow classification. Their device used multilayer Neural Network to process the signals which the neural network trained by using the parameters in physical models of multiphase fluid.

Although the few challenges for oil and gas processing has been addressed by mainly developing new sensors and applying Artificial Intelligence algorithms for pattern recognition in multiphase flow, but these techniques have their drawbacks. The sensors produce a lot of data, Big Data, which is sometimes complex and requires an expensive storage and process resources.

The big data complexity refers to its Volume, Variety, Veracity and Velocity [12]. There are key challenges for processing big data such as; data quality, data validation, data cleansing, feature engineering, high dimensionality, data reduction, data representation, distributed data sources, data sampling, real time analysis, scalability of algorithms, data visualisation, parallel and distributed data processing, decision making and data discovery. Therefore, the different methodologies developed to handle the complexity of Big data depend on the application [13]. The number of technique are increasing and the demand for processing and storing Big Data arises Cloud computing, quantum computing and bio-inspired computing are just a few examples of these technologies [13], [14].

TABLE I
THE PROPERTIES OF PIPES

Well Type	Packer True Vertical Depth(m)	Length Optical (m)
Oil	2600	3000
Gas	2700	3500
Water	2400	3200

Our data collection starts by running an optical fibre down a well alongside a pipe carrying a fluid, and it is attached to that pipe at several points. Then a laser pulse is sent down the optical fibre and it reflects from non-uniformities in the glass [5] from every point along the fibre. As the fluid mixture travels through the pipe its motion is very turbulent and it generates a noisy sound. This sound affects the light pulse so that when it is received at the surface again, the sound can be recorded from every depth as if it was an array of thousands of microphones about a meter or less apart, all along the well pipe over a distance of thousands of meters. The pipe acts like an acoustic wave guide tending to guide sound waves up and down the pipe. The sound is recorded continuously at each microphone and this results in a very large amount of data recorded each day, typically 48 Terabytes.

In this study the novel application of principle component analysis is presented. The focus on this work is investigating the effect of reducing the dimensionality of the acoustic data prior to any other further analysis. The result of this study will lead to the automation the multiphase flow pattern recognition for oil and gas industry application.

II. EXPERIMENTAL DATA

The dataset was collected from distributed acoustic sensors which attached along 4235 (m) of pipes with one meter apart. Data is collected with 10 kHz spatial frequency from oil, water and gas pipes. Table I provides the details of each pipeline such as the length of optical fibre for each pipe [3]. The Packer True Vertical Depth is the vertical distance from a final depth to a point at the surface which is usually less than the optical fibre length. The other parameters of the pipes such as their diameter and pipe material are not considered in this study. Also, the optical characteristics (single or multiphase optical fibre), the position of the optical fibre and sensor's parameters are not included in this research.

III. METHODOLOGY AND RESULT

1) *Principle Component Analysis*: Principle Component Analysis (PCA) is an unsupervised technique which reduce the dimensionality of data by finding the new variable that are linear combination of the variable in the dataset. Therefore the algorithm maximises the variance of uncorrelated variables called principle components [15]. PCA simplifies the data structure by selecting the significant data points [17]. In our application we aim to find out which part of the data set best describes the information so PCA reduces the dimension of acoustic data. Fig. ?? shows three dimensional data with distance and time in x and y axis respectively and acoustic energy in z axis. Therefore, the PCA applied on our high

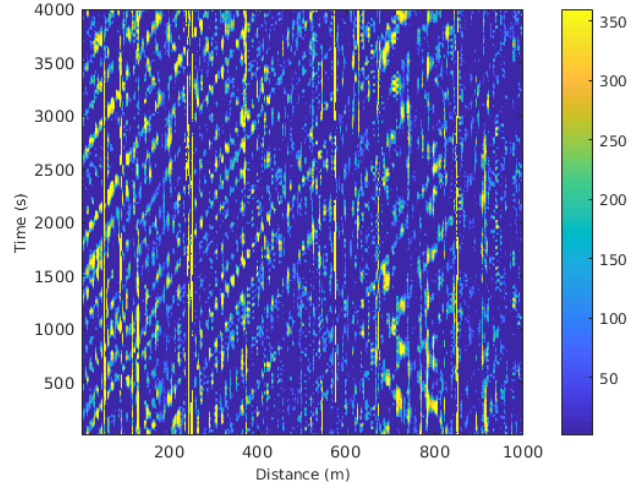


Fig. 1. Time vs distance data set.

dimensional dataset both with regards to distance and also with regards to time. The first step in PCA algorithm to standardize the original dataset by using Eq. 1.

$$X = \frac{x - \mu}{\sigma} \quad (1)$$

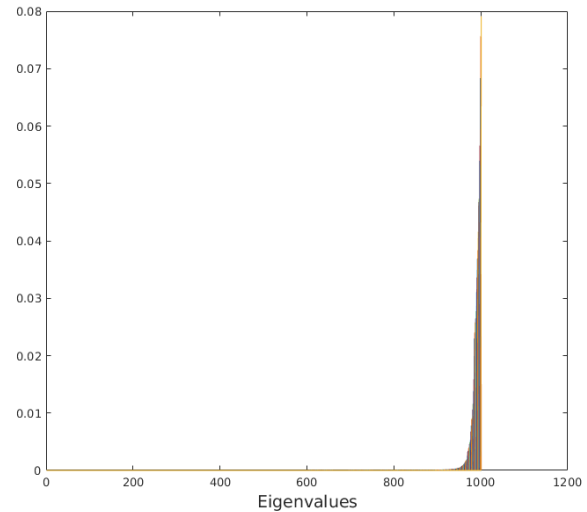
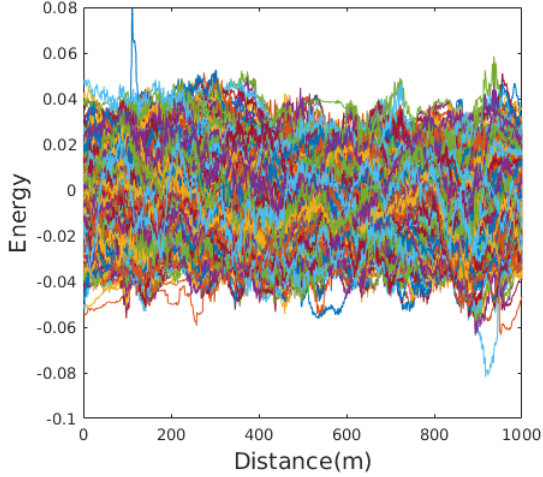
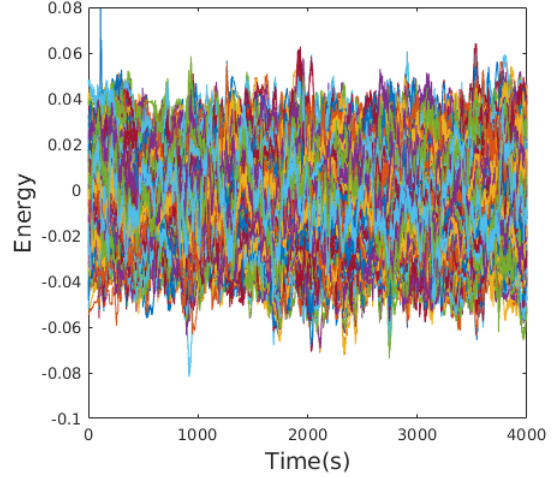


Fig. 2. Eigenvalues. The strenght of 4000 eigenvalues which each corresponded with 4000 feature vectors is shown.

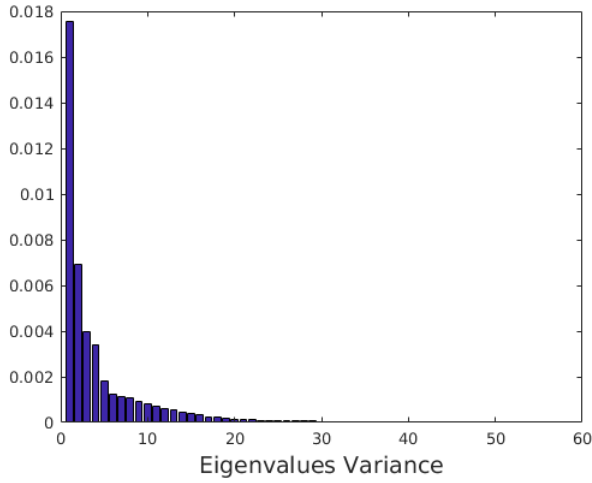
Where X , x , μ and σ are normalised signal, input signal, mean of all data set and standard deviation of data set respectively. This follows by calculating the Covariance matrix (Eq. 2) and finding the eigenvalues and eigenvectors of covariance matrix by solving the Eq. 3. If the dataset is n dimension the covariance matrix is $n \times n$. The Eq. 2 shows the covariance matrix for 2 dimensional data set [16].



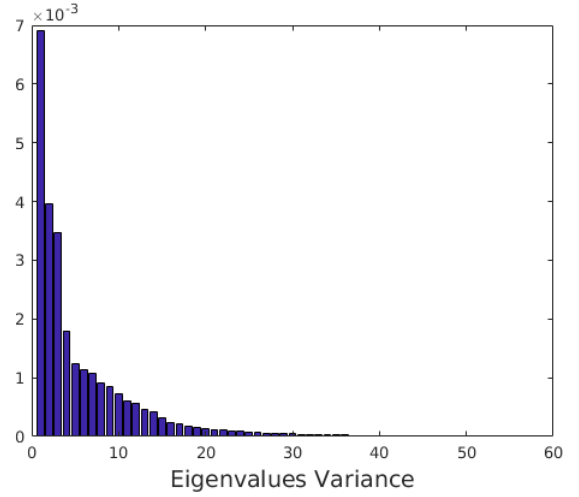
(a) Acoustic energy verces distance (water pipe)



(a) Acoustic energy vs time (water pipe)



(b) Variance of eigenvalue



(b) Variance of eigenvalue

Fig. 3. PCA result for water dataset with regards to distance. PCA applied to water dataset to reduce the dimensionality with regards to distance (a). The first fifty eigenvalues (b) plotted, however the eigenvalues are very small after the 23rd one.

Fig. 4. PCA result for water dataset with regards to time. PCA applied to water dataset to reduce the dimensionality with regards to distance (a). The first fifty eigenvalues (b) plotted, however the eigenvalues are very small after the 20th one.

$$A = \begin{bmatrix} Var(X_1) & Cov(X_1, X_2) \\ Cov(X_2, X_1) & Var(X_2) \end{bmatrix} \quad (2)$$

Where A is the covariance matrix, X_1 and X_2 are dataset belong to first and second dimension of dataset.

$$det(\delta I - A) = 0 \quad (3)$$

Where A and I have the same dimension and I is an identity matrix. The condition to solve the Eq. 3 is $(\delta I - A)v = 0$ where δ is eigenvalue and v is eigenvector. The larger eigenvalues related to more significant component of the data or principle component. Therefore we can form our features vector by choosing the number the principle component which in our case we selected the first 50 (Eq. 4). The eigen values

after 50 are very small, approximately zero, as it is evident in Fig.2 and so these values are ignored.

$$FeatureVector = (eig_1, eig_2, \dots, eig_n) \quad (4)$$

Where n is dataset dimension. The eigenvectors explain the pattern of the data and it forms the feature vector (Eq.4). The last step is using the feature vector to form the new dataset (5). The transpose of the feature vector multiplies by transpose of scaled data to achieve the lower dimension data.

$$NewDataset = Transpose(FeatureVector) \times Transpose(ScaledData) \quad (5)$$

The first dataset is from water pipe during 4000 (s) and within 1000 (m), so the dimension of the data is 4000×1000 as it is shown in Fig. 1. PCA is applied to this data set with regards to distance to reduce the dimensionality whilst preserving the features. After calculating the high dimensional covariance matrix, we have 4000 eigenvalues and 4000 eigen vector associated to those eigen value (Eq. 2). Matlab software used for all the computation for this analysis. The same procedure applied to oil and gas data set within the same distance and during the same amount of time.

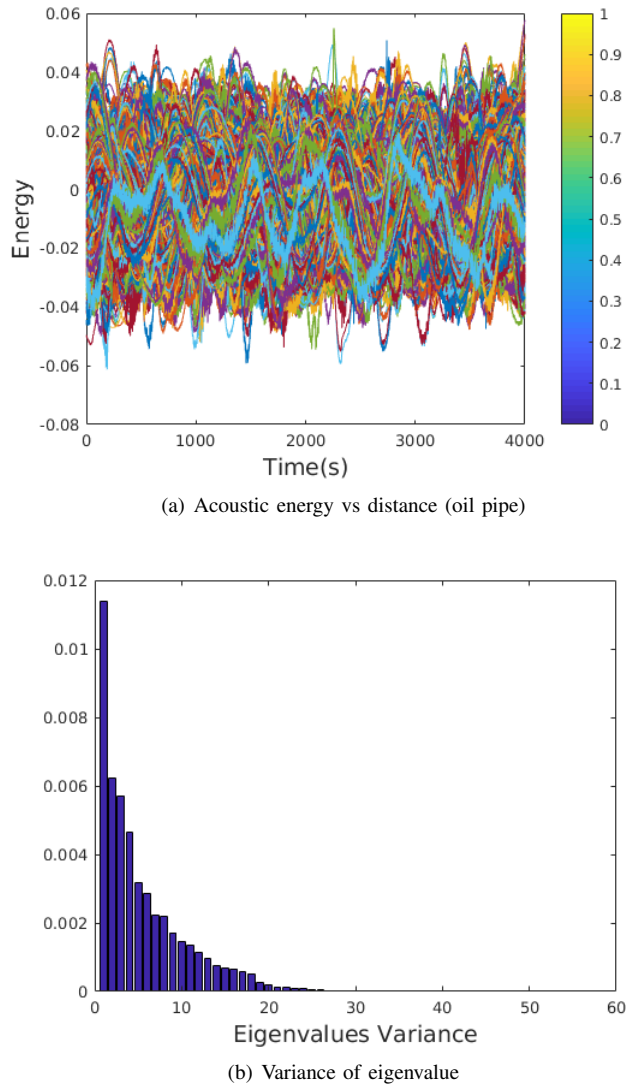


Fig. 5. PCA result for water dataset with regards to distance. PCA applied to oil dataset to reduce the dimensionality with regards to distance (a). The first fifty eigenvalues (b) plotted, however the eigenvalues are very small after the 20st one.

The result of applying PCA on water, oil and gas dataset with regards to distance are shown in Fig.3 (b), Fig.5 (b) and 7 (b) respectively. The original data in all three cases (water in Fig. 3 (a), oil in Fig.5 (a) and gas in Fig.7 (a)) has very large dimension therefore, the first 50 largest eigenvector selected to form the feature vector. However, after first 20 eigenvalues the

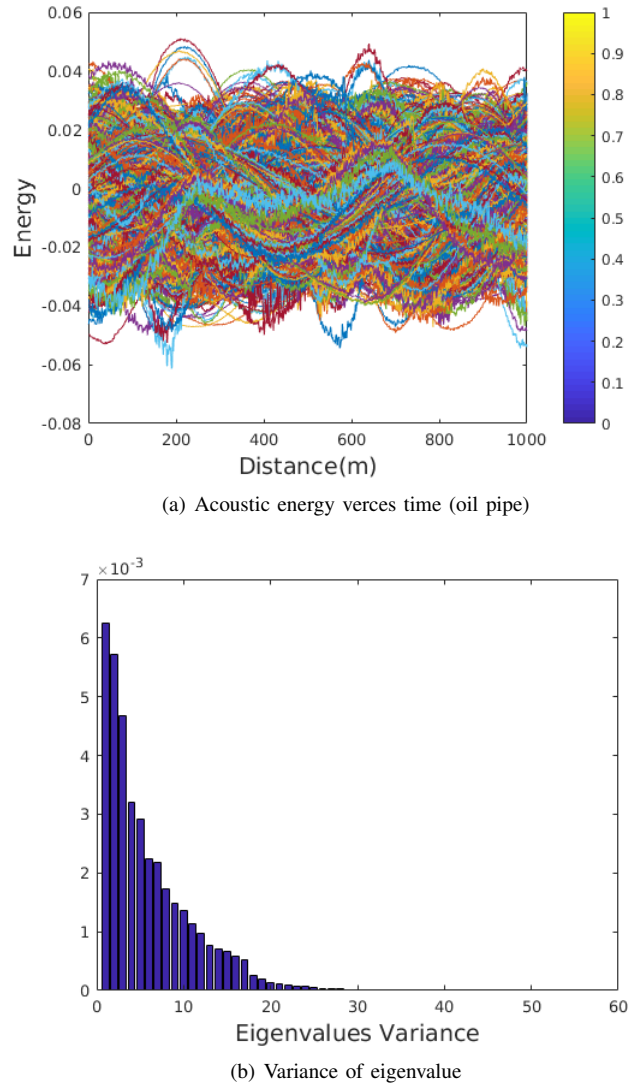
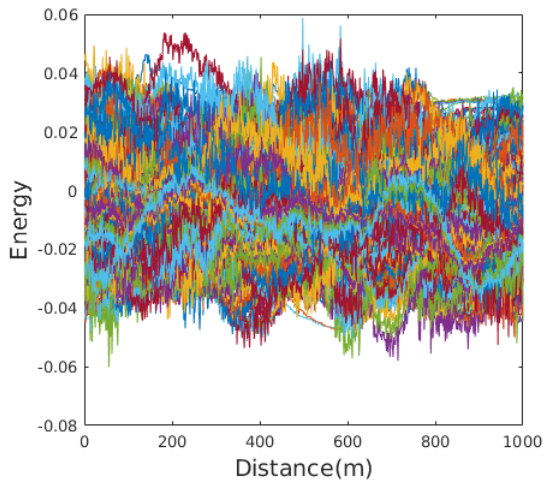


Fig. 6. PCA result for water dataset with regards to time. PCA applied to oil dataset to reduce the dimensionality with regards to distance (a). The first fifty eigenvalues (b) plotted, however the eigenvalues are very small after the 20st one.

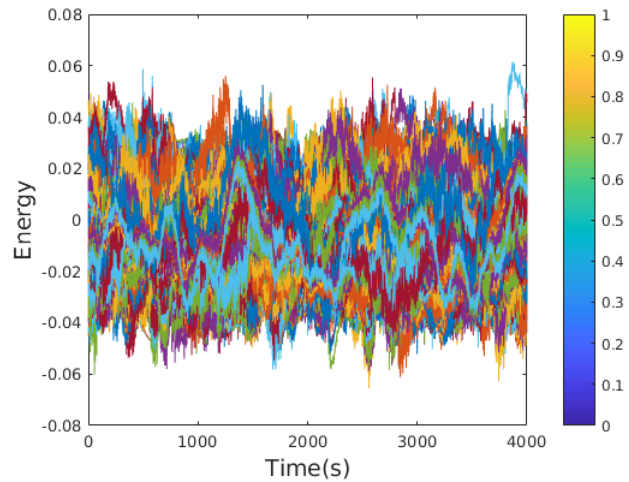
variance of eigenvalue become very small and close to zero (3 (b), Fig.5 (b) and 7 (b)). The result is slightly different when PCA is applied to the datasets with regards to time as it can be seen in Fig. 4 (b), Fig.6 (b) and 8 (b) in which there are non-zero feature vectors even after first 30 eigenvalues. Looking at the result of PCA on distance vector (Fig.3 (b), Fig.5 (b) and 7 (b)) and comparing with the result on time vectors (Fig.4 (b), Fig.6 (b) and 8 (b)) we can see the variance of eigenvectors in PCA with distance are higher than PCA with time.

IV. CONCLUSION

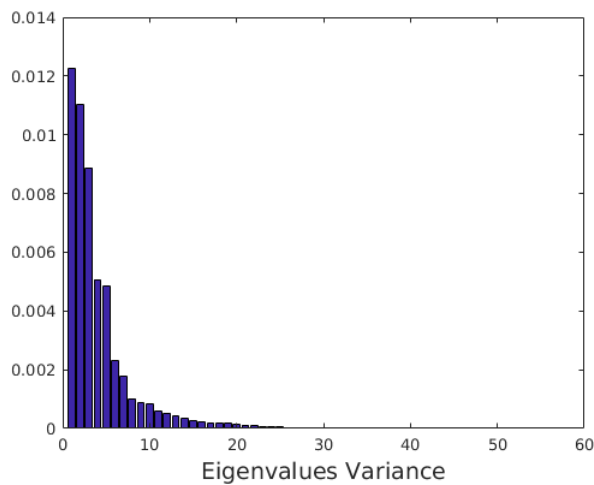
In this research Principal Component Analysis (PCA) is applied on three high dimensional acoustic dataset which are water, oil and gas to reduce the dimensionality of the high volume data. PCA applied to each dataset with regards to both distance and time by creating new uncorrelated variables



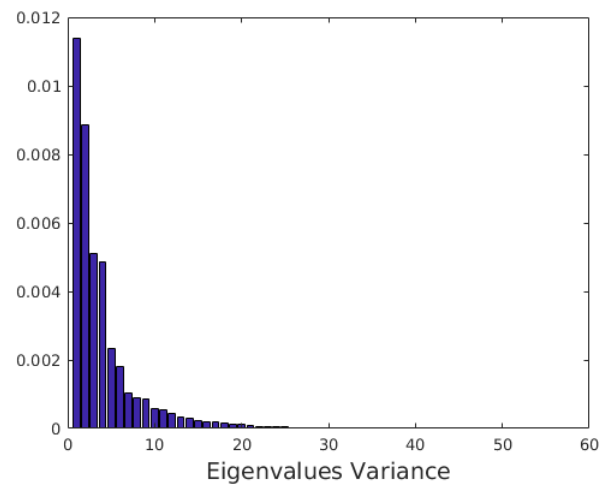
(a) Acoustic energy verces distance (gas pipe)



(a) Acoustic energy verces time (gas pipe)



(b) Variance of eigenvalue



(b) Variance of eigenvalue

Fig. 7. PCA result for gas dataset with regards to distance. PCA applied to gas dataset to reduce the dimensionality with regards to distance (a). The first fifty eigenvalues (b) plotted, however the eigenvelues are very small after the 20st one.

Fig. 8. PCA result for gas dataset with regards to time. PCA applied to gas dataset to reduce the dimensionality with regards to distance (a). The first fifty eigenvalues (b) plotted, however the eigenvelues are very small after the 20st one.

that successively maximize variance to form the feature vector. Therefore the new dataset can be processed more efficiently, faster and less expensive. The result of this study can be used for further analysis and flow measurements in oil and gas industry.

ACKNOWLEDGMENT

The author thanks K. Johannessen of Statoil for permitting the data recorded in their wells to be used in this paper. The authors thank Mahmoud Farhadiroushan, Tom Parker, Sergey Shatalin, Mladen Todorov of Silixa for use of their iDAS recording system. The author Nafiseh Vahabi also thanks the UK Government Research council EPSRC for funding studentship.

REFERENCES

- [1] Xiao, J. J., Farhadiroushan, M., Clarke, A., Abdalmohsen, R. A., Alyan, E., Parker, T. R. and Milne, H. C., *Intelligent Distributed Acoustic Sensing for In-well Monitoring*. In SPE Saudi Arabia Section Technical Symposium and Exhibition, Society of Petroleum Engineers, 2014.
- [2] Roshani, G.H., Fegghi, S.A.H., Mahmoudi-Aznavah, A., Nazemi, E. and Adineh-Vand, A., *Precise volume fraction prediction in oil/water/gas multiphase flows by means of gamma-ray attenuation and artificial neural networks using one detector*. Measurement, vol.51, 2014, 34-41.
- [3] Johannessen, K., Drakeley, B. K. and Farhadiroushan, M., *Distributed Acoustic Sensing-a new way of listening to your well/reservoir*. In SPE Intelligent Energy International, Society of Petroleum Engineers, 2012.
- [4] Figueiredo, MMF, Goncalves, JL, Nakashima, AMV, Fileti, AMF, Carvalho, RDM., *The use of an ultrasonic technique and neural networks for identification of the flow pattern and measurement of the gas volume fraction in multiphase flows*. Experimental Thermal and Fluid Science, vol.70, 2016, 29-50.
- [5] Mielke, A., Elam, K. and Sung, C., *Rayleigh scattering diagnostic for measurement of temperature, velocity, and density fluctuation*. In

- 44th AIAA Aerospace Sciences Meeting and Exhibit, IEEE Transactions on, 2006, 9-12.
- [6] Bishop, C.M., *Novelty detection and neural network validation*. IEE Proceedings-Vision, Image and Signal processing, vol.144, no.4 1994, 217-222.
- [7] Al-Naser, M., Elshafei, M. and Al-Sarkhi, A., *Artificial neural network application for multiphase flow patterns detection: A new approach*. Journal of Petroleum Science and Engineering, vol.145, no.4 2016, 548-564.
- [8] Roshani, GH., Nazemi, E. and Roshani, MM., *Intelligent recognition of gas-oil-water three-phase flow regime and determination of volume fraction using radial basis function*. Flow Measurement and Instrumentation, vol.54, 2017, 39-45.
- [9] C.M. Salgado, C.M.N.A. Pereira, R. Schirru, L.E.B. Brandao., *Flow regime identification and volume fraction prediction in multiphase flows by means of gamma-ray attenuation and artificial neural networks*. Progress in Nuclear Energy, vol.54, no.6 2010, 555-562.
- [10] E.H. Inoue, R.D.M. Carvalho, V. Estevam, A.C. Bannwart, A.M.F. Fileti., *Development of a Neural Network for the Identification of Multiphase Flow Patterns*. 14th International Conference on Intelligent Systems and Control (ISC 2013) 2013.
- [11] Meribout, M., Al-Rawahi, N. Z., Ahmed M., Al-Bimani, A., Al-Busaidi, K., Meribout, A., *A multisensor intelligent device for real-time multiphase flow metering in oil fields*. IEEE Transactions on Instrumentation and Measurement, vol.59, no.6 2010, 1507-1519.
- [12] Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., and Muharemagic, E., *Deep learning applications and challenges in big data analytics*. Journal of Big Data, vol.1, no.1 2015, .
- [13] Chen, C.P. and Zhang, C.Y., *Data-intensive applications, challenges, techniques and technologies: A survey on Big Data*. Information Sciences 275, vol.1 2014, 314-347.
- [14] Poria, S., Cambria, E., Hussain, A. and Huang, G.B., *Towards an intelligent framework for multimodal affective data analysis*. Neural Networks, vol.63 2015, 104-116.
- [15] Jolliffe, I.T. and Cadima, J., *Principal component analysis: a review and recent developments*. Phil. Trans. R. Soc. A, 374(2065), p.20150202, 2016.
- [16] Bro, R. and Smilde, A.K., *Principal component analysis*. Analytical Methods, vol.6, no.9 2014.
- [17] Li, K., Hu, C., Liu, G. and Xue, W., *Building's electricity consumption prediction using optimized artificial neural networks and principal component analysis*. Energy and Buildings, vol.108, 106-113, 2015.