A new recognition method for oil pipeline leakage using PCA and SOM neural networks

Beilei Ji¹, Hong Zhang¹, Shen Liu¹, Kezheng Zhang², Wei Zhang³, Dong Zhang¹, and Xiaoben Liu*¹

¹National Engineering Laboratory for Pipeline Safety/MOE Key Laboratory of Petroleum Engineering/Beijing Key Laboratory of Urban Oil and Gas Distribution Technology, China University of Petroleum-Beijing, 102249 Beijing, China
²CNOOC Gas and Power Group Ltd, 100000 Beijing Chaoyang, China
³PetroChina Tarim Oilfield Company, 841000, Korla, China
*Corresponding author’s e-mail: xiaobenliu@cup.edu.cn

Abstract. Oils are mainly transported by pipe in long distance for its high efficiency. While oil pipe leakage will cause serious social and environmental consequences, e.g. fire even life lost, water and soil pollution. Thus it is important to recognize pipe leakage at initial stage in engineering practice. In this research, a negative pressure wave based detection method was established for pipeline leakage recognition. Suitable parameters of negative pressure wave signals with significant difference for different working conditions were selected. Principal Component Analysis (PCA) method was conducted to reduce the dimensions of the negative pressure wave vector. Self-organizing map (SOM) Neural network was finally adopted to identify the signals for different working conditions. The proposed method was validated by experimental data, which shows that the methodology gives a high recognition rate, which can be referenced in pipe monitoring in engineering practice.

1. Introduction
Oil has been an important energy resource for over 100 years. Pipelines are commonly considered as the safest and most economical method for long distance transportation of the natural resources[1-2]. However, there is no doubt that, accidents are always inevitable which may cause pipe leakage and eventually leading to fire or explosion and serious environmental pollution[3]. For the complicated chemical components of oils, especially crude oil, if they have flown into river or soil, it will lead to sustained environmental pollution, which is intractable to handle technically. So prevent severe consequences, efficient and accurate pipeline leakage detection method should be developed.

In recent years, a series of methodologies for leakage detection of pipeline have been proposed[2]. Liang[3] used the fuzzy theory for monitoring the leakage of oil pipeline. Wang[4] established a modal identification model for negative pressure wave of pipeline based on singular value feature vectors. Lu[5] used the wavelet transform theory in pipeline leakage detection. Yi[6] studied the noise smoothing method for structural vibration test signals using an improved wavelet thresholding technique. With the advantage of quick response speeds, the negative pressure wave method is the most widely used leak location technology[7]. So it is of great importance to distinguish the negative pressure wave modal of pipeline leakage from other common pipeline working conditions such as pump stopping and so on. In this paper, an attempt was made to identify the negative pressure wave of pipeline leakage efficiently.
Firstly, the typical features of negative pressure wave of different working conditions of pipeline were extracted. Secondly, the dimensionality of the negative pressure wave data was reduced by principal component analysis method. Finally, in order to obtain an accurate recognition of pipeline leakage, the SOM neural networks were trained. The experimental data obtained by a 140-meter-long pipe was used to validate the efficiency of the proposed methodology. Results show that the method is very efficient and the detection results are reliable.

2. Methodology

2.1. Typical parameters of pipeline negative pressure wave

Researches show that negative pressure waves in pipeline are low-frequency signals, which are suitable to be analysed in time domain. Liang[3] summarized that typical parameters describing the features of negative pressure wave signal of pipeline can be divided into four categories:

(1) Parameters describing the variation size of negative pressure wave: maximum value $X_1$, peak to peak value $X_2$, average amplitude $X_3$, variance $X_4$, standard deviation $X_5$ and square root of amplitude $X_6$.

\[
X_2 = x_{\text{max}} - x_{\text{min}}, \quad X_1 = \frac{1}{N} \sum_{i=1}^{N} |X_i|, \quad X_4 = \frac{1}{N} \sum_{i=1}^{N} \left(x_i - \frac{\sum_{j=1}^{N} x_j}{N}\right)^2
\]

(2) Parameters describing the variation of negative pressure wave shape: the shape factor $X_7$.

\[
X_7 = \frac{X_3}{X_5}
\]

(3) Characteristic parameters describing the abrupt change of signals: peak factor $X_8$, impulse factor $X_9$, clearance factor $X_{10}$.

\[
X_8 = X_1 / X_3, \quad X_9 = X_1 / X_4, \quad X_{10} = X_1 / X_5
\]

(4) Parameters describing the amplitude distribution of signals: kurtosis $X_{11}$ and kurtosis factor $X_{12}$.

\[
X_{11} = \left(\frac{1}{N} \sum_{i=1}^{N} s_i^4\right)^{1/2}, \quad X_{12} = X_{11} / X_5
\]

When the working conditions of pipelines change, the above-mentioned parameters will change accordingly. But it should be mentioned that the sensibility of the parameters differs in different cases. So it is important to select appropriate parameters in the signal recognition process. The derivative reflects the trend of the pressure wave signal. As the main difference of the wave signal of four working conditions is the changing trend, derivatives are used to identify the signals in this study. When useful parameters are determined, a vector $P$ representing the negative pressure wave derivative can be obtained as equation (5):

\[
P = [X_1', X_2', \ldots, X_p']
\]

where $X$ is the parameter of pipeline negative pressure wave, and $p$ is the number of parameters selected to analysis.

2.2. Principal component analysis

Using principal component analysis method, the derived negative pressure wave vector $P$ can be translated to $F$ according to equation (6):

\[
\begin{align*}
F_1 &= a_{11}X_1' + a_{12}X_2' + \cdots + a_{1p}X_p' \\
F_2 &= a_{21}X_1' + a_{22}X_2' + \cdots + a_{2p}X_p' \\
&\vdots \\
F_p &= a_{p1}X_1' + a_{p2}X_2' + \cdots + a_{pp}X_p'
\end{align*}
\]

where $F_i$ and $F_j$ are irrelevant; $a_{ki}^2 + a_{kj}^2 + \cdots + a_{pp}^2 = 1, \quad k = 1-p$. If $i>j$, then $F_i > F_j$. 
The covariance matrix of variable matrix \( Y = (x_1, x_2, \ldots, x_p)' \) has eigenvalue \( \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0 \), and its eigenvector is \( a_i = (a_{i1}, a_{i2}, \cdots, a_{ip})' \). The principal component gained is \( F_i = a_i'X \) \((i=1, 2, \ldots, p)\). The contribution rate of principal component \( F_i \) is \( \frac{\lambda_i}{\sum_{i=1}^{p} \lambda_i} \). In addition, the accumulated contribution rate of the first \( n \) principal component is \( \frac{\sum_{i=1}^{n} \lambda_i}{\sum_{i=1}^{p} \lambda_i} \). Generally, the accumulated contribution rate of principal component should be no less than 85% in order to represent the original data reliably[8].

After principal component analysis, the principal component vector \( F \) of the negative pressure wave vector \( P \) can be derived.

### 2.3. Self-organizing map neural network

The network structure of SOM is divided into two parts: output layer and input layer. The network can efficiently represent the similarity between the data with feature mapping approach in space[9].

The algorithm used in this study is as follows:

1. Dividing the principal component vectors of negative pressure waves of different pipeline working conditions into two sets. One for network training, the other for network validation;
2. For the training set, setting random numbers of the weight vector of the output layer and normalizing it to be \( \hat{w}_j = (1, 2, \ldots, m) \). Assigning the initial value of initial winning clinical domain \( N_j(0) \) and learning rate \( \eta \), where \( m \) is the neuron’s number on output layer;
3. Choosing an input vector randomly from the training data and normalizing it as \( \hat{X} = (1, 2, \ldots, p) \) \( X \). Where \( n \) is the number of the neurons of the input layer;
4. Computing the Euclidean distance \( d_j = \left( \sum_{j=1}^{m} (X - \hat{w}_j)^2 \right)^{1/2} \) between \( \hat{X} \) and \( \hat{w}_j \), finding the minimum distance to ensure winner neuron \( j' \);
5. Adjusting the connection weight in winning clinical domain:
   \[
   w_j(t+1) = \begin{cases} 
   w_j(t) + \alpha(t, N)[x_i^p(t) - w_j(t)] & j \in N_j(t) \\
   w_j(t) & j \notin N_j(t)
   \end{cases}
   \] (7)
6. Inputting next data until learning is all complete.

When the training of a SOM network is accomplished, the trained network can be used to recognize the input negative pressure wave to identify the pipeline is leak or not.

### 3. Results and discussion

#### 3.1. Experiment data

The experimental data was obtained by a negative pressure wave simulation device designed by the fault diagnosis laboratory of China University of petroleum-Beijing, as shown in figure 1[10]. Typical working conditions of pipelines can be simulated by the device including: adjusting valves, pump stopping, pipeline leakage and pump starting.
36 groups of negative pressure waves of the mentioned four typical working conditions provided by Hu\(^4\) were used in this paper. Hu de-noised and normalized the signals and selected the part which differs obviously to conduct the signal recognition. Figure 2 illustrates the normalized signals for the typical four working conditions.

![Normalized pipeline pressure wave signals for the working condition of adjusting valves, pump stopping, pipeline leakage and pump starting](image)

3.2. Typical parameters selected for pipeline negative pressure wave recognition

As mentioned in typical parameters of pipeline negative pressure wave, the derivatives of all the signals were used for signal recognition. All parameters of the 36 groups of signals are illustrated in figure 3. It can be obtained that, four parameters, \(X_5', X_9', X_{10}', X_{12}'\), have no obvious difference for the four working conditions. So only the other eight parameters were selected for the signal recognition of pipeline leakage.

![Typical parameters for the 36 negative pressure wave signals](image)
3.3. Principal component analysis of negative pressure wave signal data

As obtained in the previous section, the negative pressure wave signals used for recognition have eight parameters for each group, which is interminable. In this section, principal component analysis method was adopted to reduce the dimension of signals. A program coded by the commercial numerical computation software MATLAB was used to conduct this analysis. Results show that the contribution rates of the first and second principal components were 59.6008% and 27.1772%, respectively. The cumulative contribution rate of these two components is 86.7780%, which is more than 85%. Therefore, these two principal components can be used to represent the main features of the original data. The coefficients of the eight parameters for the first and second principal components in principal component analysis are listed in table 1.

The distribution of the first two components for all the 36 groups of signals is illustrated in figure 4. It can be obtained that four working conditions were divided effectively by using the two principal components.

| Parameters | The first principal component coefficient | The Second principal component coefficient | Parameters | The first principal component coefficient | The Second principal component coefficient |
|------------|----------------------------------------|------------------------------------------|------------|----------------------------------------|------------------------------------------|
| X1'        | 0.369                                  | 0.386                                    | X6'        | 0.389                                  | -0.061                                   |
| X2'        | 0.456                                  | -0.015                                   | X7'        | 0.314                                  | -0.030                                   |
| X3'        | -0.074                                 | 0.656                                    | X8'        | -0.071                                 | 0.642                                    |
| X4'        | 0.452                                  | 0.037                                    | X12'       | 0.438                                  | -0.057                                   |

3.4. SOM network training and recognition for principal signal components

Four groups of signal for each working condition were selected randomly to training the SOM network, which was established by the neural network toolbox provided by the commercial numerical computation software MATLAB. The size of its layer dimension was set to be 2×3, and the topology function was set to be 'hextop'. The network training was completed after 30 times' training, and then it can be used for signal recognition. It can be found that it needs more than 100 times to complete the training procedure if we use SOM network to train all the eight parameters directly. It turns out that the calculation time of the network has been greatly reduced, and the convergence speed of the network has been improved after the principal component analysis.

After training the self-organizing map neural network, the other five groups of signals for each working condition were used to validate the accuracy of the trained network. The recognition rates of the four working conditions were all 100%, which reflects that this proposed method can provide high recognition rate for negative pressure waves.
4. Conclusions
A negative pressure wave based pipeline leakage detection method was proposed. With the combined use of principle component analysis method and SOM neural networks in this study, conclusions can be drawn as follows: The derivative reflects the trend of the negative pressure wave signal. As the main difference of the wave signals of four working conditions is the changing trend, derivatives are suitable to be selected to identify the signals. Principal component analysis is a simple and convenient method to reduce the dimension of data, which can extract the features of negative pressure wave signals effectively and reduce the complexity of the data greatly. This proposed combined method was validated by experimental data, and results show that it can efficiently reduce the complexity of the negative pressure wave signals and give high accuracy recognition for the typical working conditions of pipelines.

Acknowledgments
This research has been co-financed by National Science Foundation of China (Grant No. 52004314), Tianshan Youth Program (Grant No. 2019Q088), the Open Project Program of Beijing Key Laboratory of Pipeline Critical Technology and Equipment for Deepwater Oil & Gas Development (Grant No. BIPT2020005), Science Foundation of China University of Petroleum, Beijing (No. 2462018YJRC019, No. 2462020YXZZ045).

References
[1] S. TOPRAK, E. NACAROGLU, V. HELVA. (2017) Design charts for pipelines under longitudinal soil loading. Journal of Environmental Protection & Ecology, 18(1), 63-70.
[2] Liu X B, Zhang H, Li M, Xia M Y, Zheng W, Wu K, Han Y S. (2016) Effects of steel properties on the local buckling response of high strength pipelines subjected to reverse faulting. Journal of Natural Gas Science & Engineering, 33, 378-387.
[3] S. GLISOVIC, D. J. PESIC, D. ZIGAR, I. ANGHEL. (2016) Assessing the environmental impact of accidents in natural gas metering-regulating stations. Journal of Environmental Protection & Ecology, 17(2), 532-540.
[4] R. SILVA, C. BUIATTI, S. CRUZ, J. A. F. R. PEREIRA. (1996) Pressure wave behavior and leak detection in pipelines. Computers & Chemical Engineering, 20, S491-S496.
[5] Lu N X, Sang N Y., Zhang N J, Fan N Y. (2006) A pipeline leakage detection technology based on wavelet transform theory. Proceedings of IEEE International Conference on Information Acquisition, Shandong, China.
[6] Wang M, Zhang L, Liang W. (2007) Pipeline pressure wave identification method based on singular value feature. Chinese Petroleum Machinery, 37, 68-71.
[7] Yi T H, Li H N, Zhao X Y. (2012). Noise smoothing for structural vibration test signals using an improved wavelet thresholding technique. Sensors, 12, 11205-11220.
[8] Tian H, Yan J, Huang J, Wang Y. (2012) Negative pressure wave based pipeline Leak Detection: Challenges and algorithms. IEEE International Conference on Service Operations and Logistics, Suzhou, China.
[9] Guo Z, Zhang H. (2012) The gas pipeline risk assessment base on principal component analysis and BP neural network. Advances in Electronic Engineering, Communication and Management 1, 661-665.
[10] Liang W, Zhang L, Wang Z. (2007) Study of oil pipeline leakage detection based on fuzzy methods. Chinese Petroleum Machinery, 35, 42-45.