Fisher Discriminant Analysis (FDA), a supervised feature reduction method in seismic object detection

Mohammad Ghasem Fakhari, Hosein Hashemi*

Institute of Geophysics, University of Tehran, Tehran, Iran
*Corresponding author, e–mail: hashemy@ut.ac.ir

Abstract

Automatic processes on seismic data by using pattern recognition is one of the interesting fields in geophysical data interpretation. One part is the seismic object detection using different supervised classification methods that finally has an output as a probability cube. Object detection process starts by generating a pickset of two classes; labeled as object and non–object then selecting a set of attributes that are inputs to a classifier. As a significant step before classification, it is better to perform a feature extraction algorithm to transfer data from input space to feature space; which results can be used to reduce the dimensions by eliminating less important features in the new feature space. In this paper, our goal is to investigate and propose a proper feature extraction method for the seismic object detection process. For this purpose, we propose the Fisher Discriminant Analysis (FDA) feature extraction as a suitable method for seismic object detecting and compared this method with the Principal Component Analysis (PCA) that is very popularly used method for feature extraction. The seismic object in this study is fluid migration pathways in the North Sea and Support Vector Machine (SVM) classifier with Radial Basis Function (RBF) kernel used for classification. Finally, the obtained results show that the minimum average classification error after 30 repeats by PCA is 0.152 with 3 features but for FDA it is very smaller (.078) with only 1 dimension. The second and most important result is a posterior probability in the physical domain obtained by FDA is more interpretable than PCA. In conclusion, based on the results of the FDA for seismic object detection in this study, it is recommended that the FDA as a feature extraction method (FDA in this case automatically performs feature selection and reduces the dimension) be performed before classification in seismic object detection.

Keywords: *Seismic Object Detection; Gas Chimney; Fisher Discriminant Analysis; Fluid Migration; Gas Pockets*

Introduction

Seismic objects such as salt bodies, fluid migration paths, faults and stratigraphic features are defined here as spatial elements with an observable size and orientation and with a different seismic response with respect to their surroundings.

Meldahl *et al*., (1999) developed a pattern recognition technique to improve object detection. This method has been developed to detect objects using a supervised neural network. Inputs of the neural network are a set of attributes extracted at locations labeled by object and non–object by a user.

The neural network is applied to the input seismic attributes and trained on pre–selected object and non–object locations. Based on this training, the network makes a classification of all samples in the seismic volume into object and non–object classes. The final neural network output can be regarded as object probability volume.

The first step in seismic object detection is picking positions in two classes: object and non–object. Hashemi (2010) discussed considerations related to generating a pickset. The next step after selecting pickset calculating and identifying a set of

seismic attributes that distinguish between object and non–object. To get an optimum object image, we have to use the information from several attributes simultaneously. Then classifier will be trained on attributes extracted at locations picked by the interpreter to create a classification model for two Classes object and non–object.

The start of seismic object detection technique (Meldahl *et al*., 1999; Heggland *et al*., 1999; Tingdahl *et al*., 2001) was based on a pattern recognition method which features are seismic attributes. In the Meldahl method, Inputs of the neural network are a set of attributes. Then, Hashemi used regularized discriminant analysis (RDA) technique to rank the relevant importance of each seismic attribute in the classification. In RDA, new features are not produced, but also the importance of each attribute in classification is ranked, which this rating can use in feature selection (select of attributes). Feature extraction as a preprocessing step in pattern recognition is effective in eliminating irrelevant and redundant data, increasing accuracy and improving result comprehensibility. So, in pattern recognition before classification, it is better to run one feature

extraction method. Feature extraction is a dimensionality reduction technique that extracts a subset of new features from the original set by means of some functional mapping keeping as much information in the data as possible (Fukunaga 1990).

In this research, our goal is to investigate and propose a proper feature extraction method for the seismic object detection process, So that this additional step can be proposed for the other seismic object detection. For this purpose, we compared two feature extraction methods of Principal Component Analysis (PCA) which is very popularly used method for feature extraction and Fisher Discriminant Analysis (FDA) for seismic object detection (in this study our object is fluid migration paths).

Classifier which is selected in this study is SVC. Support Vector Machine is a supervised machine learning algorithm which can be used for both classification and regression problems. The main idea is to identify the optimal separating hyperplane which maximizes the margin of the training data (Corres & Vapnik, 1995). The detailed formalization of SVC is explained (Corres & Varpnik, 1995). In general, the SVC is a linear classifier that uses a kernel trick for nonlinear classification. We use Support Vector Machine (SVM) classifier with Radial Basis Function (RBF) kernel. The radial basis function kernel, or RBF kernel, is a popular kernel function used in numerous kernelized learning algorithms. Here, we used k–fold cross–validation method to choose optimal values for SVM parameter (penalty function) and RBF kernel parameter.

Finally, we propose the FDA feature extraction as a suitable method for seismic object detecting.

Methodology
Seismic attributes quantify properties of seismic data and each attribute highlights a special information relativly to the propagated wavefront. From a pattern recognition point of view, each computed seismic attribute is called a 'feature'. Clearly, adding each new seismic attribute can reveal additional information that will help us to object detection. But a large number of attributes in pattern recognition makes data analysis difficult and reduces the effectiveness of learning methods such as classification.

In addition, in machine learning as the dimensionality of the data rises, the amount of data required to provide a reliable analysis grows exponentially. To overcome this problem, it is necessary to find a way to reduce the number of features in consideration. The most well–known dimensionality reduction algorithms are *principal component analysis* (PCA) and Fisher discriminant analysis (FDA).

Principal Component Analysis (PCA)
There exist various techniques for dimensionality reduction which Principal Component Analysis (PCA) is one of the oldest and most commonly used. The Principal Component Analysis (PCA) reduces a large set of seismic attributes to highlight variations in the data by a linear combination of attributes and find new reduced features.

Principal Component Analysis (PCA) transforms the original set of features into a smaller subset of linear combinations that account for most of the variance of the original set (Jolliffe 1986). PCA is a linear and unsupervised dimensionality reduction technique which projects higher dimension data into a lower dimension, while it preserves the significant information in the original input data set. Consider the input data set matrix is $X \in R^{n \times m}$, where m represents original variables and n is the number of observations for each variable (the number of data samples).

The standard PCA algorithm is briefly described as follows:

Step 1: Normalize the columns of X (variables) by subtracting the mean of each column and divide it with the standard deviation of each column in such a way that mean and variance of each column becomes equal to 0 and 1, respectively.

Step 2: use the following expression to compute covariance matrix C

$$C = \frac{1}{n-1} X^T X \tag{1}$$

Step 3: Obtain the Eigenvectors and Eigenvalues from the covariance matrix C (or correlation matrix), by performing EigenValue decomposition (EVD) on covariance matrix C

$$C = \frac{1}{n-1} X^T X = V A V^T \tag{2}$$

Where *V* and A are the Eigenvectors and Eigenvalues of covariance matrix C respectively
That $A = diag(\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_m \geq 0)$

All eigenvectors are arranged according to their eigenvalues in descending order. In this step feature extraction by PCA is completed and the new features are extracted as follows:

$$T = XV \tag{3}$$

Where T is $n \times n$ transformed data in PCA domain.

Step 4: Now we have to decide how many eigenvectors to preserve. Determining the number of principal components by another analysis method and divide V into score and residual space

$$A = \begin{bmatrix} A_{pc} & 0 \\ 0 & A_{res} \end{bmatrix}, A_{pc} = diag(\lambda_1, \lambda_2, ..., \lambda_a), A_{res} = diag(\lambda_{a+1}, \lambda_{a+2}, ..., \lambda_m)$$

$$V = \begin{bmatrix} V_{pc} & V_{res} \end{bmatrix}, V_{pc} \in R^{m \times a}, V_{res} \in R^{m \times (m-a)}$$

Where V_{pc} and V_{res} indicates score and residual spaces, respectively.

Step 5: Compute the projection matrix T by the following expression

$$T = XV_{pc} \tag{4}$$

Where T is $n \times a$ reduced transformed data in PCA domain.

Fisher discriminant analysis (FDA)

PCA is a simple, well known and useful linear transformation technique that is used in various applications, but it does not consider the information (discrimination) between different classes, while calculating its transformed matrix T (PCA ignores class labels). This problem is incorporated by Fisher Discriminant Analysis (FDA) technique. The transformation matrix of FDA contains vectors which maximize the separation between different classes, while minimizing the scatter within classes. In general, FDA can be regarded, as supervised PCA.

Consider input data set matrix is $X \in R^{n \times m}$ by m variables and n observations for each variable, p is the number of classes and n_j number of observations in j^{th} class. Let x_i represents the transpose i^{th} of row of matrix X. The standard FDA algorithm is briefly formulated as follows:

Step 1: Compute total–scatter matrix S_t

$$S_t = \sum_{i=1}^{n}(X_i - \overline{X})(X_i - \overline{X})^T \tag{5}$$

Where $\overline{X}$ represents the total mean vector.

Step 2: Compute within–class scatter matrix S_w

$$S_w = \sum_{j=1}^{p} S_j \tag{6}$$

Where

$$S_j = \sum_{X_i \in X_j}^{n} (X_i - \overline{X_j})(X_i - \overline{X_j})^T \tag{7}$$

Where $\overline{X_j}$ is the mean of jth class.

Step 3: The between–class–scatter matrix S_b

$$S_b = \sum_{j=1}^{p} n_j (\overline{X_j} - \overline{X})(\overline{X_j} - \overline{X})^T \tag{8}$$

$$S_t = S_w + S_b \tag{9}$$

Where n_j is the number of samples in j^{th} class

Step 4: FDA vectors are obtained by solving the generalized eigenvalue problem solving following expression.

$$S_b W = \lambda S_w W \tag{10}$$

Step 5: Since rank(S_b) < p, there exist at most p−1 eigenvectors which are associated with non–zero eigenvalues (Fukunaga, 1990). This implies that FDA can find at most p−1 meaningful features (the residual features created by FDA are arbitrary). This is an essential restriction of FDA in dimensionality reduction that is greatly restrictive in application

This is a fundamental restriction of FDA in dimensionality reduction and is greatly restrictive in the application. Let k represents the number of non–zero eigenvalues, then:

$$W_k = \begin{bmatrix} w_1 & w_2 & ... & w_k \end{bmatrix} \tag{11}$$

Step 6: FDA transformation vectors are calculated by the following equation:

$$Z_i = W_k^T X_i \tag{12}$$

FDA considers the information between various classes and calculates the transformation vectors, while PCA does not. Therefore, it is expected that FDA transformation vectors have more discriminant power rather than PCA for Multi–class classification problems.

Case study

In this study, we decided to recognize a seismic object 'fluid migration paths' in the seismic dataset from the F3 block in the Dutch sector of the North Sea. In this area, hydrocarbon migration phenomenon and the presence of gas seepage is already discussed in the direct measurements e.g. headspace gas analysis (Schroot, 2005). Furthermore, many studies have been conducted to identify the gas chimneys, shallow gas and hydrocarbon migration in this area by Heggland *et al.*, 2000; Meldahl *et al.*, 2001; Aminzadeh and Connolly 2002; Connolly *et al.*, 2002; Hashemi *et al.*, 2008.

The first step in seismic object detection is to pick a representative pickset. The most important

part of the supervised seismic pattern recognition practice is choosing a well–defined pickset.

Some features on seismic sections, such as gas chimneys or gas clouds, have been taken as an indication of fluid migration (Buhrig 1989; Hunt 1996; Heggland 1998; Parnell 2002). Vertical hydrocarbon migration paths are recognized in seismic data as a vertical zone of chaotic low amplitude (and low energy) seismic data called gas chimneys. Here, we use this definition for picking fluid migration paths (gas chimneys is one of the fluid migration paths). We introduced 600 representative pick locations, with an equal number in object class and non–object class. In Fig. 1, the position of these picked locations in seismic section inline 130 and in Fig. 2, the location of these picked samples in the seismic cube is displayed.
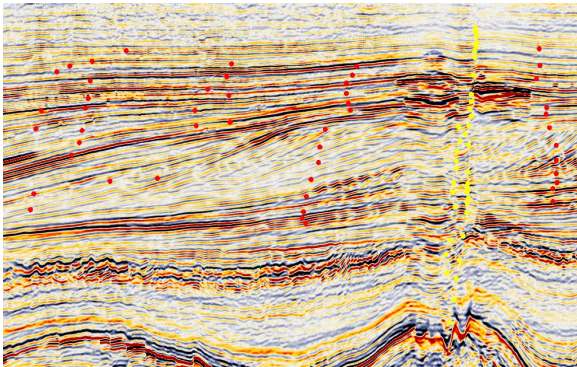

Figure 1. A section from in-line 130 of F3 seismic data. Yellow picks are"migration pathways" and red picks are "non-migration pathways" locations.
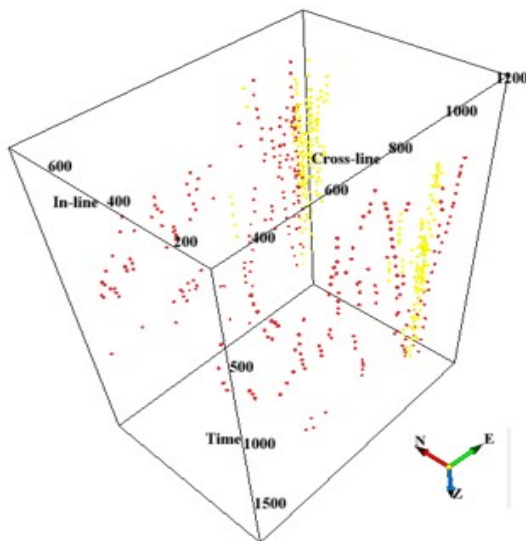

Figure 2. The spatial position of pick locations in the F3 seismic cube. Yellow picks are"migration pathways" and red picks are "non- migration pathways" locations.

After picking, we should select a set of attributes which have well distinguish between two classes. In this study, the criteria for the selection of attributes includes 5 steps. In the first step, according to the fluid migration characteristics in the seismic sections, seismic attributes such as energy, similarity, dip, curvature, signal to noise ratio and etc, were selected. Tingdahl *et al.*, (2001) introduced a set of attributes for chimney detection. At this step, based on the interpreter's knowledge of fluid migration in the seismic section, a large initial set of seismic attributes was selected.

In the next step, the initial attributes were compared qualitatively and visually, then a subset of them was selected.

Cross–plot is an effective visualization tool which can be used to predict visually the correlation between attributes and also the power of each attribute to separate two classes. Then, Cross plot of selected attributes was plotted and selected a subset of attributes with more separation power.

In the next step, the density distribution of the selected attributes (from the previous steps) was plotted for both classes. This method, like cross–plot, can provide a qualitative criterion for the separation of two classes for each attribute.

Finally, used a quantitative feature selection method, decision trees. A tree consists of a number of branches, one root, a number of nodes and a number of leaves. Individual branch is a series of nodes from root to the leaf, and each node includes one attribute. The position of an attribute (node) in a tree provides information about the importance of the associated attribute in classification. Eventually, after accomplishing the steps mentioned, we have selected 8 attributes (namely: Similarity Parallel, Similarity Diagonal, Similarity All direct, RMS, Noise, Filter residual, Signal/Noise, Variance polar dip) that number of them is shown in Fig. 3.

The values of these seismic attributes calculated at selected points in both classes, provides input data for detecting pattern and seismic object detection. Seismic attributes' values are extracted at all 600 presented locations, finally our dataset is created which includes 600 samples in 8 dimensions and 2 class migration path and non–migration path.

In Fig. 4, the cross–plots of selected attributes are shown. Usually, in pattern recognition, after selecting attributes, a statistical algorithm is used to feature extraction and feature reduction. In this

research, after determining and selecting the appropriate attributes as the primary features, in the next stage, two statistical algorithms were used: Principal Component Analysis (PCA) and Fisher Discriminant Analysis (FDA), to extract new features from a set of basic features and further reduce dimensions.

To see how separate two classes in the new features extracted space, The cross–plot features extracted from methods PCA and FDA are shown in Figures 5 and 6, respectively.
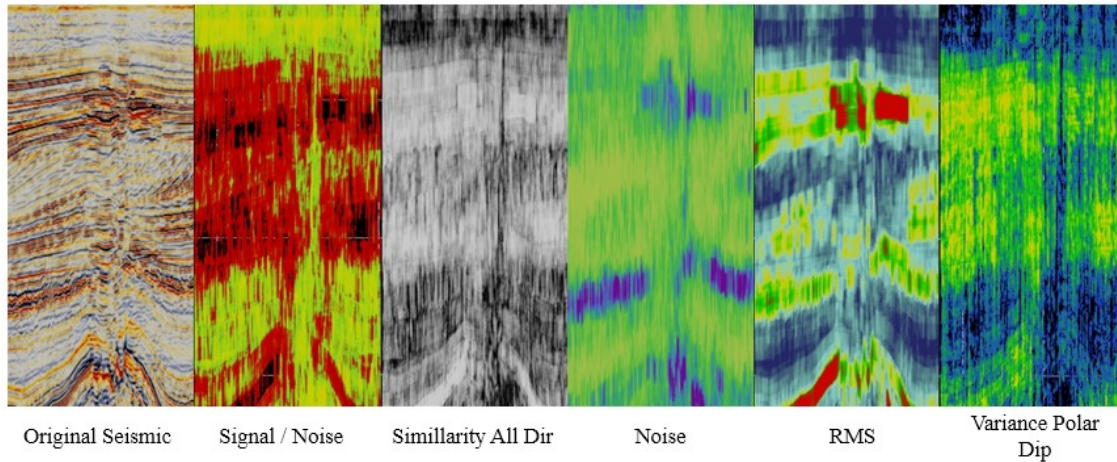

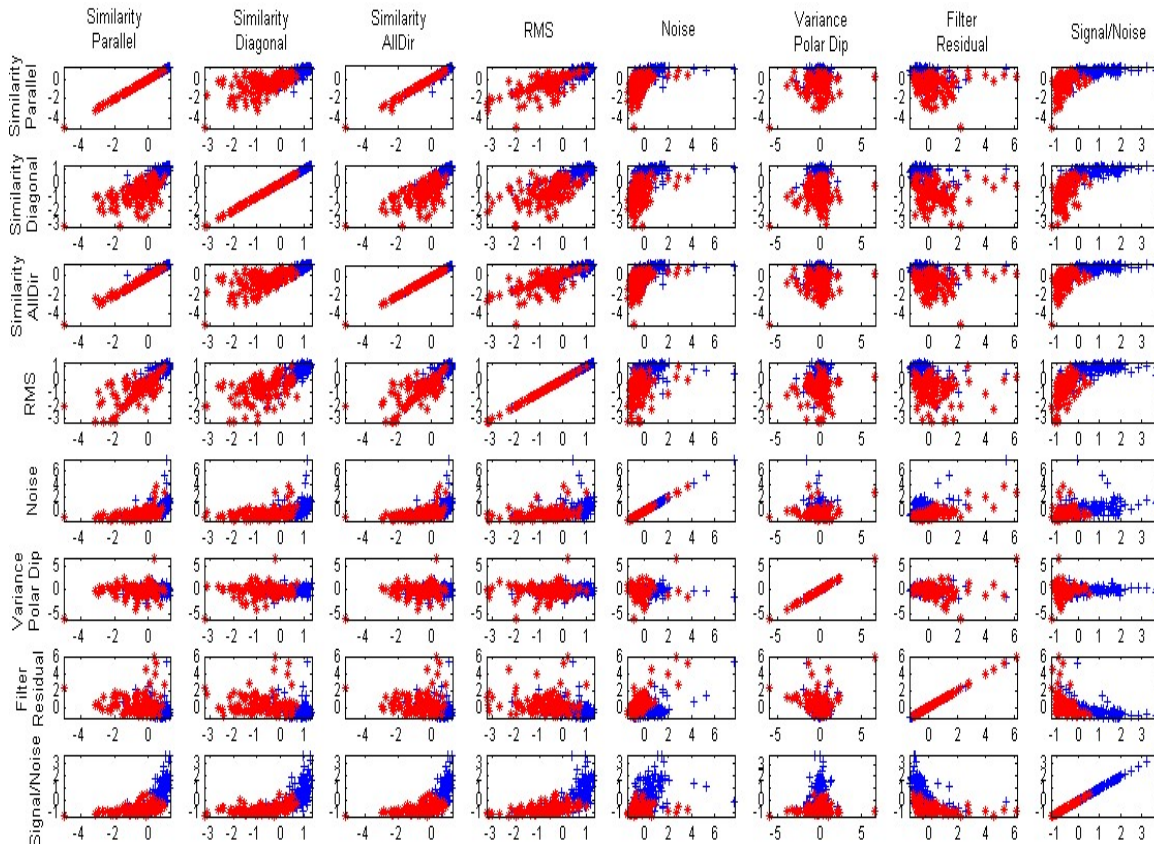Figure 3. Zoomed section of seismic data and selected attributes that include the migration pathway

Figure 4. Cross-plots of selected 8 important attributes for fluid migration paths detection. Two class of migration paths (red points) and non-migration paths (blue points) are shown.
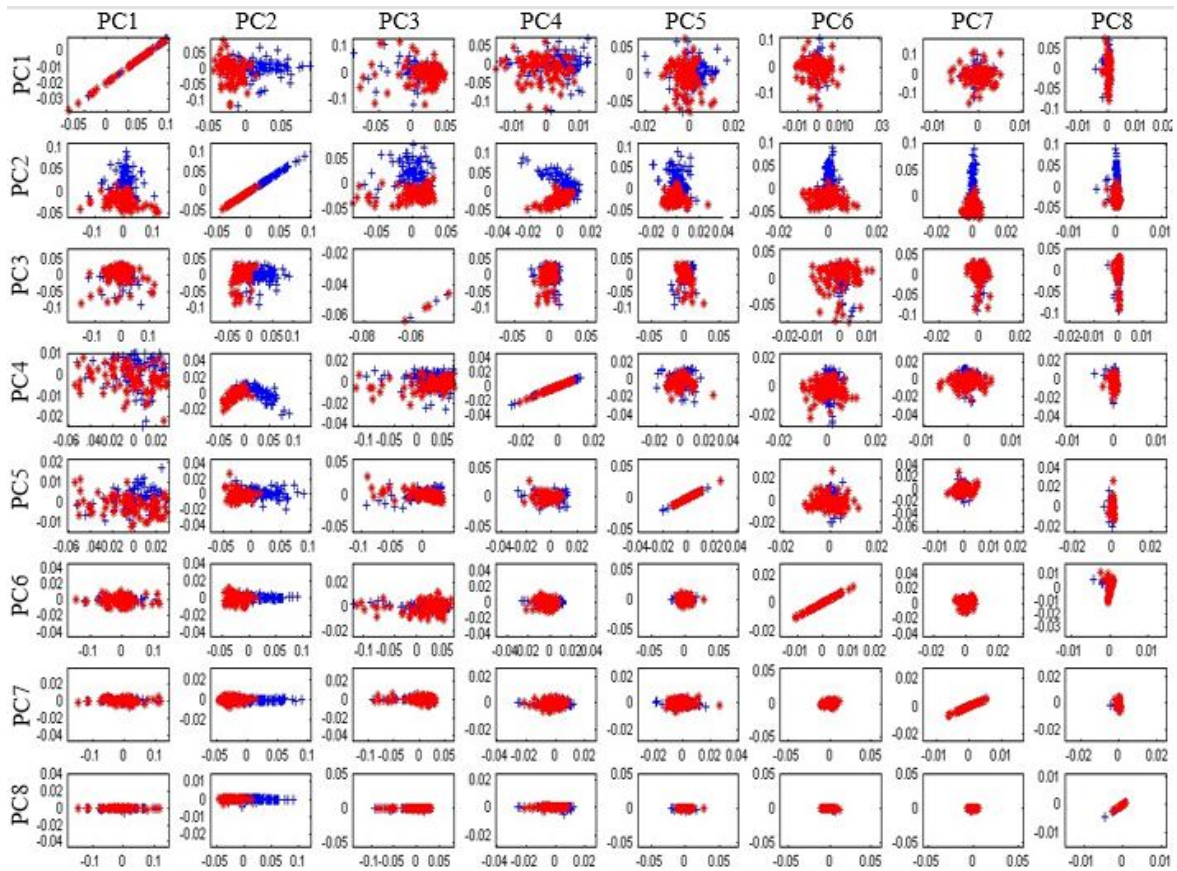
Figure 5. Cross-plots of PCA extracted features. Two class of migration paths (red points) and non-migration paths (blue points) are shown.
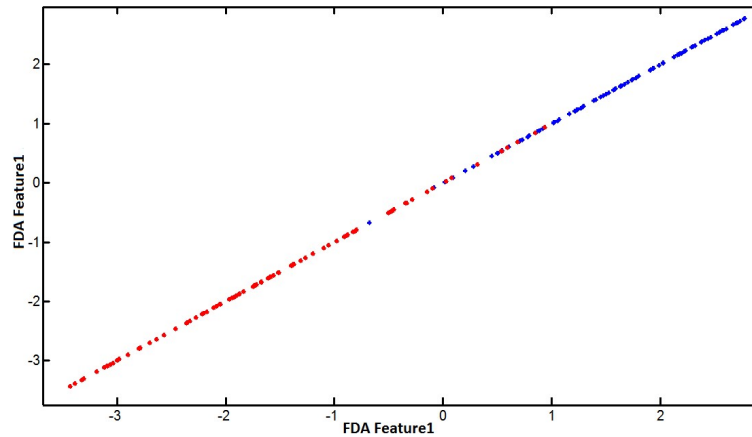

Figure 6. Cross-plot of FDA extracted feature. Since seismic object detection is a two-class problem, only one attribute of the FDA method is extracted. Red points are "migration pathways" and blue points are "non- migration pathways".

For classification, dataset should be divided into two parts; training data and test data. Then Support Vector Machine (SVM) classifier with Radial Basis Function (RBF) kernel was trained by the training data. Misclassification error is obtained by comparing the classifier output for test data and true label. Eventually, as the output of the detection process, a posterior probability cube is obtained that highlights fluid migration pathways.

Here to compare PCA and FDA as well as the original space (attributes); first, the classification is done on attributes. Here the classification is done with 30 repetitions and average classification error is calculated for all 8 dimensions.Then the

classification is performed on the extracted features from PCA. And finally, the classification is performed on 1 dimension extracted from FDA (in two class problem, FDA only can find one meaningful feature).

Finally, the average classification error after 30 repeats for every 3 ways is shown in Fig. 7.

Finally, as a classification output and its results in the physical domain, a posterior probability (named objectCube) is obtained for each way which shows the probability of fluid migration paths in the seismic cube.

The posterior probability of the 'migration pathways' class which is fixed on section inline 130 for every 3 methods (Attributes, PCA, FDA) is shown in Fig. 8. Posterior probability shows the possibility of migration paths with the three colors; Yellow (high probability), green (average probability) and blue (low probability) and areas where there is no possibility of migration path, are shown without color and it is only shown by seismic section.

Discussion

Commonly object detection performance is analyzed from two aspects. First is average classification error and the second which is more important is the posterior probability in the physical domain. Average classification error in fig. 7 shows that minimum average error after 30 repeats for attributes obtained 0.152 with 7 attributes while for PCA it is 0.152 with 3 features; but for FDA it is

.078 with only 1 dimension. In Addition to the significant less error in FDA rather than PCA, dimension reduction to one feature, is one of the advantages of FDA to PCA.

Another important aspect to assess performance, is the meaning of posterior probabilities in the physical domain and ability to interpret their consistency with the direct measurement experiments and other petroleum systems. The posterior probability of the migration pathways class fixed on the section for every 3 methods is shown in Fig. 8. In this section, migration pathways include 3 major areas where the shallowest part is shallow gases (gas packet). And the second area which is a vertical zone from bottom to the surface is a gas chimney, and the third part in the bottom section is a dewatering zone.

Posterior probability shows the possibility of migration paths with the three colors; Yellow (high probability), green (average probability) and blue (low probability) and areas where there is no Possibility of migration path, are shown without color and it is only shown by seismic section.

According to the results obtained from FDA in Fig. 4, areas of migration paths have shown correctly in yellow (high probability), as well as areas that should be classified in the class of non–migration path. In PCA and attributes, migration paths with low probability observed with blue are incorrectly classified, while in FDA this region is correctly classified in non–migration paths class with no color.
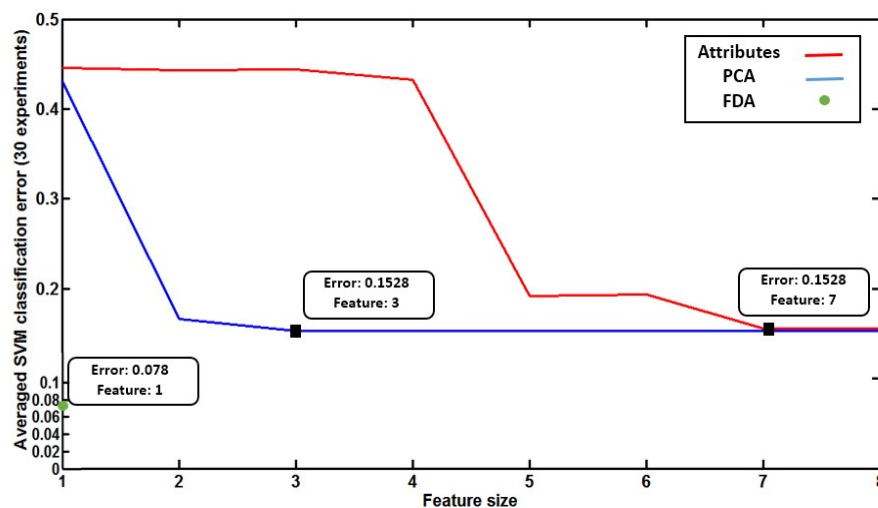

Figure 7. Averaged classification error by 30 cross-validation experiments for attribute space, PCA and FDA versus the number of features.
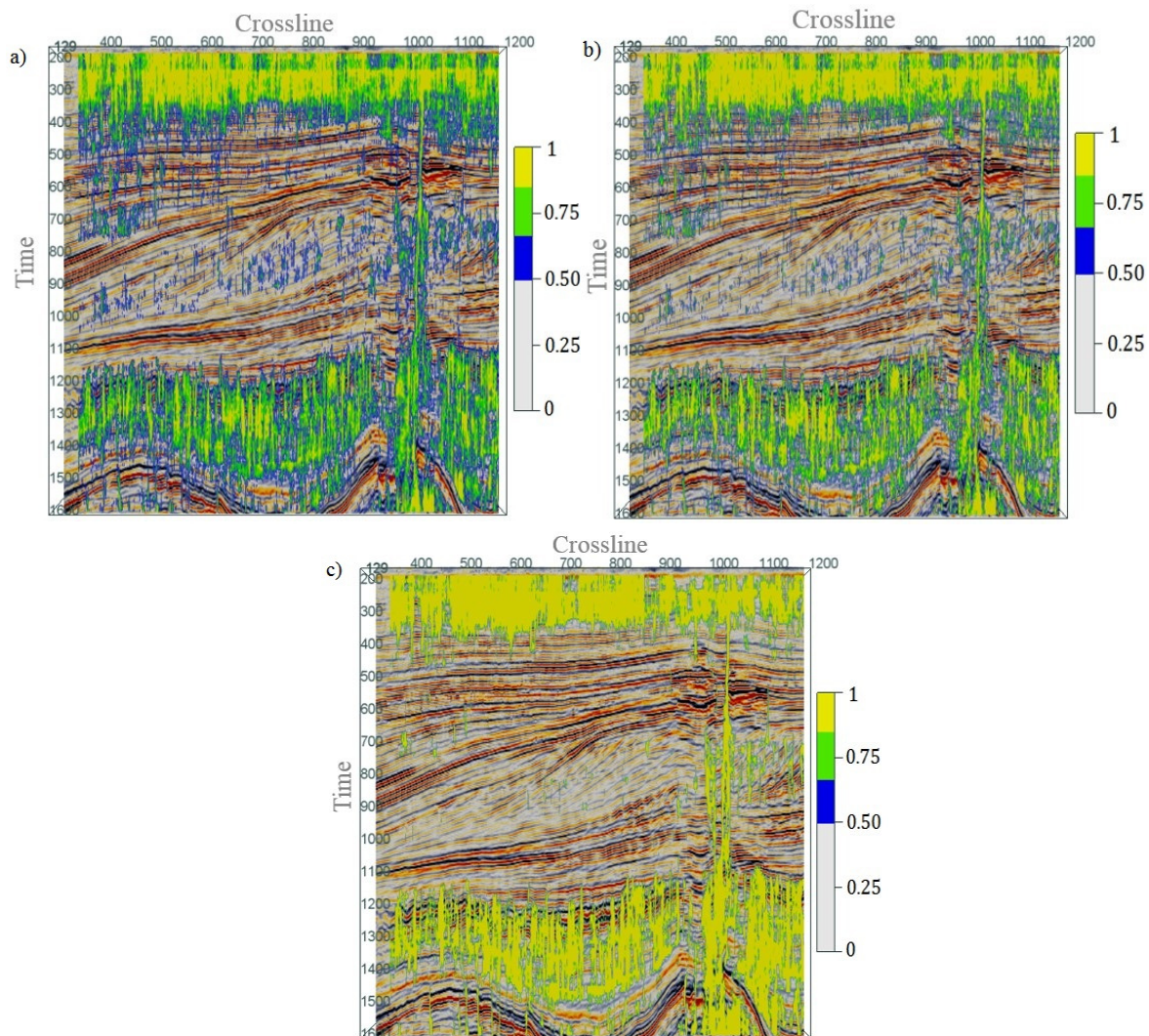
Figure 8. Posterior probability of "migration pathways" on seismic section Inline 130 obtained from a) attributes, b) PCA with 3 features and c) FDA with 1 feature.

Conclusions

In the seismic object detection process, each step requires considerations which can affect the results. For example, generating the pickset, selecting seismic attributes and choosing the appropriate classifier should be done with the necessary considerations.

In this study, two feature extraction methods, PCA and FDA, in seismic object detection were compared. The object in this study is fluid migration pathways and used SVM classifier for classification.

Object detection performance is evaluated from two aspects; the first is average classification error and the second which is more important is a posterior probability in the physical domain. Finally, the results show that in FDA, average classification error is less than PCA. The second and most important result is the posterior probability in the physical domain where the FDA approach is better and more interpretable than PCA. As a generalization of the results in the process of automatic seismic object detection, applying the FDA feature extraction method prior to final classification not only reduces the initial number of selected features to not only one dimension but also is more interpretable in the physical domain.

References

Aminzadeh, F., Connolly, D., 2002. Looking for gas chimneys and faults. AAPG Explorer, 23(12): 20–21.

Buhrig, C., 1989. Geopressured Jurassic reservoirs in the Viking Graben: modelling and geological significance. Marine and Petroleum Geology, 6(1): 31–48.

David, L. C., Sawyer, R. K., Aminzadeh, F., de Groot, P. F., Herald, J., 2002. Gas Chimney Processing as a New Exploration Tool: A West Africa Example. In AAPG Annual Meeting 2002.

Fukunaga, K., 1990. Introduction to Statistical Pattern Recognition. Academic Press, London.

Hashemi, H., 2010. Logical considerations in applying pattern recognition techniques on seismic data: Precise ruling, realistic solutions. Cseg Recorder, 35(4): 47–50.

Hashemi, H., Tax, D. M. J., Duin, R. P. W., Javaherian, A., Groot, P. D., 2008. Gas chimney detection based on improving the performance of combined multilayer perceptron and support vector classifier. Nonlinear Processes in Geophysics, 15(6): 863–871.

Heggland, R., Meldahl, P., Bril, B., de Groot, P., 1999. The chimney cube, an example of semi–automated detection of seismic objects by directive attributes and neural networks: Part II; interpretation. In SEG Technical Program Expanded Abstracts 1999: 935–940.

Heggland, R., Meldahl, P., De Groot, P., & Aminzadeh, F., 2000. New Technology Enhances Seismic Chimney Interpretation. American Oil and Gas Reporter, 43(2): 78–83.

Hunt, J. M., 1996. Petroleum geochemistry and geology. New York: WH Freeman, 2: 1–743

Jolliffe, I.T., 1986. Principal Component Analysis. Springer, New York, NY.

Ligtenberg, J. H., 2005. Detection of fluid migration pathways in seismic data: implications for fault seal analysis. Basin Research, 17(1): 141–153.

Meldahl, P., Heggland, R., Bril, B., de Groot, P., 1999. The chimney cube, an example of semi–automated detection of seismic objects by directive attributes and neural networks: Part I; methodology. In SEG Technical Program Expanded Abstracts 1999: 931–934.

Schroot, B. M., 2003. Surface and subsurface expressions of shallow gas accumulations in the southern North Sea. Search and Discovery, article, 40090.

Schroot, B. M., Schüttenhelm, R. T., 2003. Shallow gas and gas seepage: Expressions on seismic and other acoustic data from the Netherlands North Sea. Journal of Geochemical Exploration, 78: 305–309.

Schroot, B. M., Klaver, G. T., Schüttenhelm, R. T., 2005. Surface and subsurface expressions of gas seepage to the seabed—examples from the Southern North Sea. Marine and Petroleum Geology, 22(4): 499–515.

Xie, X., Li, S., He, H., Liu, X., 2003. Seismic evidence for fluid migration pathways from an overpressured system in the South China Sea. Geofluids, 3(4): 245–253.