A Feature Extraction Technique Based on Factor Analysis for Pulsed Eddy Current Defects Categorization

Nurul A’in Ahmad Latif1,3, Ilham Mukriz Zainal Abidin1, Noorhazleena Azaman1, Nordin Jamaludin2 and Ainul Akmar Mokhtar3

1Leading Edge Non Destructive Testing Group, Industrial Technology Division, Malaysian Nuclear Agency, 43000 Bangi Selangor, Malaysia
2Department of Mechanical and Materials Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia,43600 Bangi, Selangor, Malaysia
3Mechanical Engineering Department, Universiti Teknologi Petronas, 32610 Bandar Seri Iskandar, Perak, Malaysia

Corresponding author, email: nurul_ain@nm.gov.my

Abstract. Pulsed eddy current (PEC) is an advance Non-destructive testing that is widely used in multiple industries for surface and subsurface defect detection. Normally, the conventional features extractions are conducted based on time to peak and peak value response. However, signals extraction could be confusing for a thin specimen because times to peak for both surface and subsurface defects are occurred at almost similar time. This paper introduced the application of confirmatory Factor Analysis for defect categorizing for stainless steel thin plate. Through this statistical method, two categories of defects namely surface and subsurface defects are clearly to be distinguished between one to another.

1. Introduction

Pulsed eddy current (PEC) is an emerging technique and operates based on electromagnetic principle and commercially applied in multi discipline fields including aeronautical, manufacturing and petroleum industries. PEC is categorized as an advance Non-destructive testing (NDT) develops for defect quantification including surface and subsurface crack measurement, depth estimation and crack construction. In contrast to conventional eddy current that excites by single or harmonic excitation at one frequency value, PEC is excites by square waveform. The advantages of using continuous spectrums are it allowed a deeper penetration and it is thoughtful to bring information about anomalies that are deeply buried inside the specimen and higher robustness of anti-interference [1].

In PEC, the peak value and peak time are the main features used for signals interpretation. Peak values are correlated to the size of the defect whereas the latter are representing the depth of the defect. Tian and Sophian [2] proposed a new feature for defect classification termed as rising time. Rising time feature is used to identify the types of defects and overcome the lift-off issue. Since PEC system are developed from wide-band electronics device thus it may impair and prone to signal-to-noise. It may require a complex data analysis for signal interpretation compare to eddy current conventional system. Recently, it has gained an attention and requirements from industry to develop an accurate signals analysis for ensuring the integrity of tested structure.

In this research, the specimen used is stainless steel 304 with nominal thickness 0.051 mm. The specimen selection is based on the material and parameters that representing the fuel cladding of TRIGA
MARK II Nuclear Reactor. The signals interpretation and characterization for very thin specimen may have a bit confusing due to the difficulty to distinguish between the surface and subsurface defects. The possibilities for time to peak value for both defects to be occurred at the same time are relatively very high.

Predominantly, statistical analysis generates important information from research conducted including patterns, characterization and demographics. In 2002, Sophian et. al [3] introduced the application of principal components analysis (PCA) for defect classification and quantification based on PEC response acquired. In this research, a new signals analysis is introduced based on statistical method, factor analysis. Generally, PCA and FA are looks like very similar and familiar in many ways. However, they have different in the term of its fundamental that will bring a huge effect on its applications. The function of PCA is to perform data reduction and the purposes are to create one or more index from one large population measured. These created index variables are called components. Figure 1 shows the fundamental operation processes of PCA in data reducing.

![Figure 1. Fundamental operation of PCA in data reducing.](image)

Based on this figure, PCA is combining 4 measured variables (Y) into single component, C. It is clearly shown that the Y variables contribute to the components variable, C and this model can be simplified using Eq. 1.

$$C = W_1Y_1 + W_2Y_2 + W_3Y_3 + W_4Y_4$$

(1)

where $W_n$ $(n = 1, 2, 3 and 4)$ = weight or factor score coefficient

In contrast to PCA, factor analysis measures a latent variable that is cannot to be measured from one single variable. The measurement for FA is illustrated as shown in Fig. 2. Theoretically, factor analysis is a multivariate analysis method that used to identify the underlying factors that are correlated between covariation among the independence variables. The objective of its implementation is to reduce the number of variables used to explain any linkages or relationship establish among the variables. From the figure it describes that the factor, F is causing the responses on the four measured variables, Y. This model literally can be interpreted as a set of regression and can be calculated using Eq. 2.
Figure 2. Fundamental operation of Factor Analysis in data reducing.

\[ Y_1 = b_1 F \]
\[ Y_2 = b_2 F \]
\[ Y_3 = b_3 F \]
\[ Y_4 = b_4 F \]

where \( b_n \) (n=1, 2, 3 and 4) = factor loading

Factor analysis has been reported was used in multi disciplines research including construction [4], groundwater quality inspection [5] and financial analysis [6]. However, to date is never been reported about its application in any of non-destructive testing and inspections. This research used a confirmatory factor analysis approach to determine the relationships between the variables selected which are time to peak and peak value. Confirmatory approach will give a confirmation whether the selected measurement fitted the scope of the research.

The rest of the paper is organized as follows. A section 2 discusses about the factor analysis approaches. Section 3 introduces the experiment works conducted and results are briefly discuss in Section 4 and finally in Section 5 is the conclusion.

2. Application of factor analysis for experimental data

In our study, we investigate the application of factor analysis in categorizing defects into surface and subsurface based on the variables and compare this statistical results with the normal signal analysis outcome.

Factor Analysis is a commonly operates based on variable reduction technique. This multivariate statistical technique is used for three primary reasons including to reduce the number of variables, from large to small and also to establish underlying dimensions between measured variables and constructs. In addition, it is also very useful for providing construct validity evidence.

We apply a confirmatory Factor Analysis based signal processing using PEC developed system to extract the features for defect detection and classification. As described in Section 1, FA is measuring the latent variable that cannot be calculated from one single variable by itself. In this experiment, the two identified factors to be measured are surface and subsurface defects. For the defect quantification, the common signal features, time to peak and peak value are analysing in order to find the correlation between these variables and to establish the underlying factors that linked to each other. By applying the FA for a signals analysis, the defects that are identified as are factors are more easily to be distinguished based on the correlation established between selected variables.
According to Cunningham [7], the factor analysis method is utilized to determine the nature of the construct influencing a set of responses and to achieve the validity of data purposes. Thus this research followed Ramani [8] who indicated that it is the simplest method to explore the constructs. Initially, there are four guidelines to be performed in ensuring that the data are well cooperate and suitable for factor analysis statistical method. The four steps to be conducted are including The Kaiser-Meyer-Olkin (KMO) test, Bartlett’s test, variance measurement and factor rotation.

The Kaiser-Meyer-Olkin (KMO) is the test conducted to measure how suited the data to be performed using Factor Analysis and it measures the sampling adequacy for each variable. This statistical test is measuring the proportion of variance among variables that might be common variance. The accepted KMO values are equal or more than 0.50. If the KMO value acquired is less than 0.5, it means that the data is not suitable for this statistical analysis. KMO values between 0.5 and 0.7 are considered as mediocre, 0.7 and 0.8 as good, values between 0.8 and 0.9 are great and values above 0.9 as superb [9].

The second statistical test performed is Bartlett’s test. The Bartlett’s test of sphericity was used to test the significance of correlations among all factors, with 0.05 cut off point employed in determining the significance level.

Factor Rotation (loading factor) - The process of adjusting the factor axes in order to get a simpler and more significant factor solution. The rotated values should be more than 0.40. If no factor rotation appeared, this means the measurements are already significantly simple. The item(s) need to be deleted as measurement in variable if loading value is less than rotated threshold. The data can only be recognized to perform factor analysis if all the results from these four guidelines are fulfilled.

The purpose of the factor analysis rule above is to achieve data reduction or retain the nature and character of the original items, and to delete those items which had lower factor loadings and cross loading (10).

3. Case Study

3.1 Finite element model

The signals analysis conducted for approach, Factor Analysis and conventional signal analysis are performed using simulation works’ results. This finite element method work is simulates using a dedicated software, COMSOL Multphysics in 3-dimension geometry. The main reason of selecting a simulation approach instead of experiment work is to establish the ideal environment workspace thus it is expected producing an accurate result (discrepancies). Hence the results are free from any insignificants errors contributed by electronics noises, external magnetic field disturbance and human errors in which it may leads into a misleading interpretation. In this simulation works the specimen used is 0.50 mm thickness stainless steel 304 plate fabricated with eight different depths and the opening size of each defect is 2.0mm. The defects depth, \(d\) are 0.15mm, 0.20 mm, 0.25mm, 0.30mm, 0.35mm,0.40mm,0.45mm and 0.50mm as illustrated in Fig 3. The time step of the pulse waveform is 10 ms with 5 ms excitation pulse width and pulse repetition frequency is 100 Hz.

![Figure 3. Illustration of stainless steel 0.50 mm plate thickness fabricated with seven different defects depth.](image)
3.2 EC probe

PEC probe is generally robust, small and provides a high sensitivity in detection. In this simulation work, the probe has dimension of 16 mm outer diameter, 10.5 mm inner diameter, 30 mm height and made up from 300 turns of copper wire. Figure 4 illustrated the schematic diagram of PEC probe.

![Figure 4. Schematic diagram of PEC probe.](image)

For the surface detection, the probe will be positioned above the defect opening with a lift-off of 2 mm. Meanwhile for subsurface defect detection, the probe is positioned on the other side of the plate. The orientation for these finite element method works are shown in Figure 5a and 5b. All the data acquired from these simulation works will be analysed using statistical factor analysis method and conventional signals analysis approach for comparison purposes.

![Figure 5a. Finite element method work on surface defect.](image)   ![Figure 5b. Finite element method work on surface defect.](image)

4. Results

4.1 Simulation work

In this paper, we are going to categorize the defects based on their location in the specimen conducted through two different approaches, finite elements work and factor analysis. Figure 6a and 6b show the results acquired from simulation works conducted on surface and subsurface defects. From the results it shows that the deepest defect having the lowest net magnetic field and the scenario is vice versa for the shortest depth. Based on the results, the magnetic field value for each defect depths show a similarity regardless their location either on surface or subsurface. For example, Figure 7 shows that the 0.20 mm defect depth for both surface and subsurface generates magnetic field relatively same around 0.018 T. From Fig. 6a and 6b, the peak value for surface and subsurface at 0.20 mm depth are 0.01843 T and 0.01850T. Meanwhile the time to peak are 0.1742 ms and 0.1751 ms. Peak value for both types of defects are relatively same for one particular depth and the time to peak is happened at almost same time. These kind of signals condition will lead into confusing especially for interpretation and defect categorization based on their location. To overcome this problem, the peak and time to peak values for both types of defects are extracted and analysis using factor analysis.
Figure 6a Transient net magnetic field acquired from surface defects.

Figure 6b Transient net magnetic field acquired from subsurface defects.

Figure 7. Transient net magnetic field acquired from 0.20 mm surface and subsurface defect.

4.2 Factor analysis
FA is conducted to categorize the defects based on peak value and time to peak values. For each type of defect, 8 measurements are taken giving it a total of 128 measurements. According to Hatcher [11] for factor analysis approaches the number of sample size should be of at least 100 or 5 times the number of variables.
Table 1 show the factor analysis results performed using the simulation data collected. In accordance to the requirements, it is mandatory to identified either the sampling is adequate for the analysis to be conducted. Table 1 shows the KMO, Bartlett’s Test and variance for the whole set of data.

Table 1. Kaiser-Meyer-Olkin, Bartlett’s Test and variance measurements acquired from finite elements work conducted on surface and subsurface defects.

| Factor     | Component 1 | Component 2 |
|------------|-------------|-------------|
| pv1        | 0.945       |             |
| ttp1       | 0.945       |             |
| pv2        | 0.979       |             |
| ttp2       | 0.979       |             |
| % of variance | 49.573     | 43.009      |
| Kaiser-Meyer-Olkin (KMO) | 0.50       |             |
| Bartlett’s Test of Sphericity |             |             |
| Approx. Chi Square | 148.251     |             |
| df         | 6           |             |
| Significant | 0.0000      |             |

Component 1 is the subsurface defect and it represents time to peak (tt1) and peak value (pv1). Meanwhile component 2 is representing surface defect having of two variables, time to peak (tt2) and peak value (pv2).

The KMO is measuring the sample adequacy and it varies between 0 and 1. The value that closer to 1 is better and 0.50 is set as minimum value as described by Ramani [12]. From the table, the KMO (p) value calculated is 0.50 thus the sampling data is accepted to be proceeded using factor analysis. Meanwhile, a Bartlett’s Test of Sphericity is the test for null analysis used to determine whether the correlation matrix is an identity matrix. The purpose of using Bartlett’s Test of Sphericity is to identify whether the variables are load properly. According to the law, an identity matrix is where the diagonals are all 1 and the off-diagonals are all 0. There are two hypotheses assume for this analysis as stated below:

i. Null Hypothesis, H₀: There is no statistically significant interrelationship between variables and type of defects.

ii. Alternate Hypothesis H₁: There may be a statistically significant interrelationship between variables and type of defects.

Taking a 95% of level of Significance, α=0.05. From Table 1, the p-value (Sig) 0.00˂ 0.05, thus the Factor Analysis is valid. Since p˂α, therefore we reject the H₀ and accept the H₁ that there is a statistically significant interrelationship between variables and type of defects. Meanwhile the minimum value for variances as mentioned by Ramani is 50%. In Table 1, the total variance value calculated is 92.582 %. Thus the data is valid for factor analysis statistical approach.

Figure 8 shows the scree plot obtained from the simulation conducted on both surface and subsurface defects. From the scree we can determine how many numbers of the factors can be included into the model by looking at the scree plot begin to level off. Additionally, any factors with Eigenvalue less than 1 need have to be excluded from the model. The Eigenvalue used to establish the cut off of the factors involve represent the strength of each factor. Based on this figure, there are two components, component 1 and 2 with Eigenvalue more than 1. It is clearly shows that surface and subsurface defects are the variables that explain at least an average amount of the variance.
Meanwhile Figure 9 shows the components plot in rotated space using orthogonal Varimax with Kaiser Normalization and rotation converged in 3 iterations. In Varimax rotation, the values less than 0.40 had been blanked out. In fact, the higher factors loadings indicate that the particular variable is closely associated with the monitored factor. Table 2 shows the rotated component matrix for both components, surface and subsurface defects. Based on the table, times to peak and peak value for subsurface defect are highly associated and can be loaded into same group of factor. Meanwhile, times to peak and peak value for surface defect with factor value, 0.945 are highly related and can be clustered into one factor group.

![Component Plot in Rotated Space](image)

**Figure 9.** Component plot in rotated space.

| Factor | Component 1 | Component 2 |
|--------|-------------|-------------|
| pv2    | 0.979       |             |
| ttp2   | 0.978       |             |
| pv1    |             | 0.945       |
| pv2    |             | 0.945       |

### Table 2. Rotated component matrix.

5. **Conclusion**

A new feature extraction of PEC signals using Factory Analysis analytical method has been developed and investigated. The KMO measurement and Bartlett’s test show that the data are significantly related.
This statistical method focusing on the defects categorization based on the eigenvalue on the Scree plot and component plot in rotated space. These values indicate the differences of the defects classes which in this experiment both are represent surface and subsurface defects. In future, it is suggested to conduct the testing using thin ferromagnetic specimen because it has a high conductivity value.

6. Acknowledgement
The authors would like express gratitude to Ministry of Science and Technology and Malaysian Nuclear Agency for the research grant (03-03-01-SF0220) and facilities provided.

References
[1] He Y, Luo F, Pan M, Hu X, Gao J and Liu B 2010 Defect classification based on rectangular pulsed eddy current sensor in different directions Sensors and Actuators A: Physical. 157(2010):26 https://doi.org/10.1016/j.sna.2009.11.012
[2] Tian G Y and Sophian, A 2005 Defect classification using a new feature for pulsed eddy current sensors NDT & E International. 38 (2005):77-82 https://doi.org/10.1016/j.ndteint.2004.06.001
[3] Sophian A, Tian, G Y, David T and Rudlin, J 2002 A feature extraction technique based on principal component analysis for pulsed eddy current NDT NDT & E International. 36(2003):37-41 https://doi.org/10.1016/S0963-8695(02)00069-5
[4] Ocal E, M Oral, L E, Erdis E and Vural G 2007 Industry financial ratios—application of factor analysis in Turkish construction industry Building and Environment. 42(2007):385-392 https://doi.org/10.1016/j.buildenv.2005.07.023
[5] Liu C W, Lin K H and Kuo Y M 2003 Application of factor analysis in the assessment of groundwater quality in a black foot disease area in Taiwan Science of the Total Environment. 313(2003) pp 77-89 https://doi.org/10.1016/S0048-9697(02)00683-6
[6] De A, Bandyopadhyay G and Chakraborty B N 2011 Application of the Factor Analysis on the Financial Ratios and Validation of the Results by the Cluster Analysis: An Empirical Study on the Indian Cement Industry. 2(2011) pp 13-31 https://doi.org/10.1063/1.3516362
[7] Cunningham W A, Nezlek J B and Banaji M R 2004 Implicit and explicit ethnocentrism: Revisiting the ideologies of prejudice Personality and Social Psychology Bulletin. 30(2004) pp 1332–1346 https://doi.org/10.1177/0146167204264654
[8] Ramani G and Siegler RS 2008 Promoting broad and stable improvements in low-income children’s numerical knowledge through playing number board games. Child Development. 79 pp 375–394 https://doi.org/10.1111/j.1467-8624.2007.01131.x
[9] Hair, Jr J F, Black, WC, Babin B J, Anderson R E, Tatham R L 2006 Multivariate data analysis Pearson-Prentice Hall (Upper Saddle River, NJ) 6th Ed
[10] Hatcher L 1994 A step-by-step approach to using the SAS(R) system for factor analysis and structural equation modeling (Cary, NC/ SAS Institute)
[11] Kaiser H F 1974 An index of factorial simplicity Psychometrika 39(1974) pp 31–36
[12] Hutcheson G and Sofroniou N 1999 The multivariate social scientist: Introductory Statistics Using Generalized Linear Models Sage Publication (Thousand Oaks, CA) https://doi.org/10.4135/9780857028075