Electromagnetic Nondestructive Evaluation of Tubes using Data Mining Procedure

A Savin1*, N Iftimie1, P Vizureanu2, R Steigmann1 and G S Dobrescu1

1 National Institute of Research and Development for Technical Physics, Nondestructive Testing Department, Iasi, Romania
2 Gheorghe Asachi Technical University, Faculty of Materials Science and Engineering, Iasi, Romania

E-mail: asavin@phys-iasi.ro

Abstract. The fundamental issues in nondestructive evaluation consists in the identification of events corresponding to the flaws which can appear in the examined object and their extraction from noises. This is usually done by comparison with pre-established thresholds, experimentally determined by using standard samples or in the basis of the solution of the forward problem and simulations. This paper presents the features extraction using data mining procedure in the case of tubes from steam generators having different flaws. The data mining is carried on using simulated models in CIVA 9 and experimental data gathered using an inner differential sensor developed in this purpose.

1. Introduction
Electromagnetic (EM) nondestructive evaluation (NDE) is an interdisciplinary domain, which plays an essential role in the assuring components and systems reliability. NDE tests are periodically effectuated, without affecting the further use of object or material, in order to characterize the materials and locate the flaws, preventing the apparition of events that can alter the functioning of the product. NDE offers a balance between quality/price and efficiency and can be applied especially to industrial inspections. Many papers have combined ultrasonic tests with electromagnetic tests to increase the reliability and researchers continue to obtain new ways of applying physics and other disciplines to develop better NDE methods.

In NDE domain, impressive amount of data are collected and stored, with specific details and relevance for each applied inspection method and also the nature of inspected product. The amount of data and their complexity exceed the ability of human selection and analysis of data, automatic techniques being required to adapt the corresponding work sequence. For NDE methods, the data extraction from data basis implies the processing of data and identification of patterns and tendencies from these information, in order to adopt a decision for selection.

The data mining principles have been known for a long time, but with the emergence of large data bases they have gained a broad spreading.

In order to streamline NDE operations (costs/time), nowadays, the NDE procedures store large amount of data at relatively low costs (data which often do not deliver relevant information) and then are post processed by special departments. The system “Online analytical processing” (OLAP) [1] is part of the broader category of business intelligence and comprises also reports writing and data mining. Adopting new technologies to obtain data basis, which integrates data from multiple sources
into a given location, allows for quick answers. In NDE, detailed data are obtained, allowing the
establishing of a general trend of the evolution of the inspected product. Their visualization is difficult,
and therefore the data storage concept is split into components for easier understanding and use.

The paper is based on the use of database data processing methodology and some possibilities of
applying this procedure in NDE of steam generators using differential inner sensors.

2. Principles of temporal data mining

Data mining, known as knowledge discovery in databases [2,3], represents a domain which grown
rapidly, having multiple applications due the necessity of computer processing of information.
According to [4] data mining is “the process of discovering interesting knowledge, such as patterns,
associations, changes, anomalies and significant structures from large amounts of data stored in
databases, data warehouses, or other information repositories”. The simplified steps taken in process
are: collect data; filtering by eliminating errors; selecting properties that interest the analysis;
application of filter and detection/analysis of new knowledge; validation/visualization and evaluation
of results.

![Data mining principle.](image)

The data provided by eddy current equipment is sampled and digitized and the results of an
inspection can be presented as a time series

$$X = \{x_t, t = 1, 2, ..., N\}$$

is a collection of length $N$ where $x_t$ is the value of a time series $x$ at time $t$. It is assured that the length
of each time interval $[t, t+1]$ is equal for all $1 \leq t \leq N$. Each series contain data from a large number of
samples, each determination being affected by noises due to the equipment or displacement system, or
can contain information that can be considered as noises and which are in fact local modifications of
electrical permittivity or magnetic permeability, variations of lift-off, temperature fluctuations, etc.
prefiltering is required to avoid high frequency noises and the common method use filtering with the
average oven a slide window [5]. After prefiltering, few time series transformations are applied:
shifting; scaling and rank-space.

Then, the possibility to exist an event for all starting points and for different window widths is
calculated, an event in time series corresponds to region where the probability is maxim. The common
method uses the filtering with the average over a sliding window [5].

$$x_t = x_t - \frac{1}{L} \sum_{k=-\frac{L}{2}}^{\frac{L}{2}} x_{t+k}$$

(2)

where $L$ is the window length.

The shifting transform on a time series is to get a new time series by adding some real number to
each item in the old time series [5]. A shifting of $\delta$ on a time series $X$ is

$$X + \delta = \{x_t + \delta, t = 1, 2, ..., N\}$$

(3)
The scaling transform of a on a time series \( X \) is [6].

\[
X = \{ ax, t = 1, 2, \ldots, N \}
\]

(4)

such transform on a time series in to a new time series by multiplying same real number to each item in the old time series. A simple way to make a similarity measure invariant to shifting and scaling is to first normalize the time series. The normal form \( \hat{X} \) of a time series is transformed from \( X \) by shifting the time series by its standard deviation

\[
\hat{X} = \frac{X - \text{mean}(X)}{\text{std}(X)}
\]

(5)

The Pearson’s correlation coefficient between two times series is defined as follows

\[
corr(X, Y) = \frac{\text{mean}(X \cdot Y) - \text{mean}(X)\text{mean}(Y)}{\text{std}(X)\text{std}(Y)} = \frac{1}{N} \sum_{i=1}^{N} \hat{X}_i \hat{Y}_i
\]

(6)

This shifting and scaling is performed on the amplitude axis of the time series. A widely used transform in data mining is rank-space transformation. We rank each point in a time series \( T_R \) from 1 to the number of total point \( N \), where the lowest value is assigned a value of 1 and the largest is assigned a value of \( N \). This will allow us to consider a time series with a uniform distribution of noise, regardless of the noise distribution before the transformation [7].

Using the sum of the values inside any arbitrary sliding window of \( TR \), we can obtaining its p-value (the probability that the sum occurred by chance). For a starting point \( s \) and windows size \( w \), \( S(s,w) \) the sum of the ranked values inside that window, with \( 1 \leq s \leq N, 1 \leq w-1 \leq N \). The sum is defined as \( S(s,w) = r_{s+1} + r_{s+2} + \ldots + r_{s+w-1}, \) for \( r_{s+1}, \ldots, r_{s+w-1} \in X_R \). According to [7] for each pair of \( (w,N) \) exist a distinctly probability distribution, because each sum is dependent only on the number of values being added together \( (w) \) and their bounds \( 1 \leq s \leq N \).

Dividing the formed sums by the number of possible combinations for \( s \) and \( w \), and respecting the above method, the probability density defined as \( \left( \frac{N-w}{w} \right) \) can be evaluated [8]. Two problems appear: identification of event in time series and the evaluation of shape and severity of discontinuities which had generated this event. The identification of the event is made knowing that an event in time series corresponds in \( (s,w) \) plane where the probability is maximum, being characterized by two values – starting point and window width (corresponding to its end point).

The NDE of discontinuities can be carried out by clustering procedure (based on the appurtenance of the event to a cluster determining Mahalanobis distance in spatial feature associated to the possible flaws), either using a neuro-fuzzy system (trained with possible data answers and tested in time series in the region of event apparition).

The procedure of clustering requires the construction of clusters on the basis of simulated models, for a large variety of flaws, 6-10\(^{th}\) order Fourier harmonics amplitudes being chose as features. For the second procedures, the database must be obtained by simulation for the training of the network.

### 3. Methodology

Steam generator tubes from Pressurized Heavy Water Reactors, CANDU 600 type, made from Incoloy 800 having the inner diameter 18.2 mm and wall thickness 1.80 mm were taken into consideration. On these samples, artificial discontinuities were practiced using electro discharge machine, table 1.
Table 1. Dimensions and location of EDM slots.

| No | Defect type               | Dimension [mm] | Depth [mm]          |
|----|---------------------------|----------------|---------------------|
| 1  | Hole                      | 0.1            | Through hole        |
| 2  | Hole                      | 0.3            | Through hole        |
| 3  | Hole                      | 0.3            | Through hole+fretting|
| 4  | Hole                      | 0.4            | Through hole        |
| 5  | Inner circumferential slot| 1*             | 0.1                 |
| 6  | Inner circumferential slot| 0.5*           | 0.05                |
| 7  | Inner circumferential slot| 1*             | 0.2                 |
| 8  | Inner circumferential slot| 0.5*           | 0.1                 |
| 9  | Inner circumferential slot| 0.5*           | 0.2                 |

Obs * mark the opening flaw

The artificial discontinuities presented above were distributed on the tube sample having 800 mm length, similarly with those used in steam generator construction figure 2. The precision of slots execution was 0.01 mm, the distance between them being 30 mm.

In order to obtain minimum noise, the sensor developed in this purpose is inner differential type, each coil has 50 turns, from CuEm wire 0.1 mm diameter and the distance between coils is 3mm. During the measurements the sensor is maintained fixed meanwhile the tube is displaced with a motorized XY stage- Newmark type, coupled with the PC through RS 232 interface. The sensor is connected to Network/Impedance/Spectrum Analyzer 4395A Agilent USA, connected with PC through IEEE 488.2 Keithley interface. The scheme and the experimental set-up are presented in figure 3. Once the level of noise was estimated, the optimal frequency was established at 240 kHz. The measurements were effectuated with 0.01 mm displacement step.

4. Results and discussions

According to [9] the results obtained by simulation have supported a database with the simulated responses of the sensor at the main type of flaws, grouped on categories (through holes, inner circumferential slots, etc.). In order to locate the position of artificial discontinuities practiced on the tubes, for obtaining simulated results of response of the differential sensor described above in presence of discontinuities presented in table 1, software CIVA 9.2 has been employed. For Incolloy 800, the electrical conductivity is $1.1 \times 10^6$ S/m and for flaws electrical conductivity is 0 S/m. The relative magnetic permeability is 1. The database obtained is used for data mining procedure. The obtained
data were processed in order to correspond to the set-up prescribed by [10]. These types of discontinuities form clusters, their appurtenance at a cluster being given by Mahalanobis distance.

The real and imaginary components of the signal obtained at the scanning over the discontinuities through holes #1 – 0.1 mm diameter and #3 – 0.3 mm respectively, as well as inner circumferential slot #5 with1x0.1 mm dimensions are presented in Figure 4a. Due to the high level of signal sampling (over 20100 for a scanning step of 0.01 mm), the quantization noise appears, so the signal must be filtered, the best choice being the use of sliding window [9]. Figure 4b present the filtered signal. The representations show the good signal to noise ratio, the discontinuities being well emphasized.

![Figure 4](image.png)

**Figure 4.** Response of sensor, real and imaginary components of signal: (a) raw signal; (b) filtered signal.

It can be shown that the quantization noise was reduced. Down sampling until signals reach 100 samples can be performed to decrease the sampling rate without deformation of signal shape. The signals invariant to shifting and scaling can be obtained by normalization (eq. (5)), figure 5a. Applying rank-space transformation to signals presented above, the uniform distribution of the noise is obtained (figure 5b).

![Figure 5](image.png)

**Figure 5.** Post processing of the signals: (a) Normalized signal; (b) Rank space transformation of the signal.

Figure 6 presents the calculated distribution probability of event apparition in the data series, in \((S,w)\) plane. \(S\) is the starting point and \(w\) is the window width, distinctive for two channels, real and imaginary, respectively.

The start point of the events and their width can be evaluated from the high density of probability zones, the flaw being defined by the starting point and its length by the window width.
The nature of flaws and its appurtenance to a cluster is carried on following the clustering procedure described in [5], function of Mahalanobis distances between the point representing the emphasized flaw ad the weighting centers of the clusters formed on the basis of simulation and experimental data.

![Image](a)

**Figure 6.** The density of probability of event apparition: (a) real; (b) imaginary.

Figure 7 shows the clusters of vector signal in amplitude plane (a) and in 3D space (b), the axes being the coefficient amplitudes of the 6th, 7th and 8th harmonics. The results of classification are presented in table 2, only for three events visible in figure 7.
Figure 7. Classification on flaws categories: (a) amplitude-phase plane; (b) 3D in 6th, 7th and 8th harmonics of Fourier spectrum.

Table 2. Probability of characterization of events I, II and III.

| Event | Through wall hole diameter [mm] | Inner circumferential slots - w×d [mm] |
|-------|---------------------------------|----------------------------------------|
|       | 0.1 | 0.3 | 0.3* | 0.4 | 1x0.1 | 0.5x0.05 | 1x0.2 | 0.5x0.1 | 0.5x0.2 |
| I     | 0.91 | 0.51 | 0.42 | 0.35 | 0.43 | 0.21 | 0.32 | 0.17 | 0.21 |
| II    | 0.65 | 0.89 | 0.69 | 0.61 | 0.31 | 0.19 | 0.37 | 0.41 | 0.52 |
| III   | 0.15 | 0.31 | 0.41 | 0.27 | 0.94 | 0.41 | 0.72 | 0.31 | 0.37 |

It can be shown that clusters containing different categories of the flaws can be constructed, even if one flaw have been detected into a certain category. The cluster formed by the signals provided by the through hole with different diameters generate smaller Mahalanobis distance than the cluster consisting in inner circumferential slots. If the number of features is increased, the classification can be carried on with more refinement, decreasing the probability of incorrect classification.

5. Conclusions
The use of an automatic classification system allows the classification of the flaws into classes or even into indications with estimated geometrical dimensions. Eddy current examination operations have two temporal series as output (one for real component of induced voltage and the second for the imaginary component). The identification and estimation of artificial flaws practiced on steam generator tubes has been performed by specific data mining procedures allowing a correct identification, classification and estimation of flaws dimension. The clusters for various flaws classes were obtained using synthetic data obtained by simulations and real data obtained by eddy current testing of samples.

References
[1] Thomsen E 2002 OLAP solutions: building multidimensional information systems. John Wiley & Sons
[2] Chen M S, Han J and Yu P S 1996 IEEE T Knowl Data En. 8 (6) 866-883
[3] Fayyad U and Piatesky-Shapiro G, Smyth P 1996 From Data Mining to Knowledge Discovery: An Overview (Massachusetts: The MIT Press).
[4] Hui S C and Jha G 2000 Inform Manage 38(1) 1-13
[5] Cosarinsky G, Ruch M, Rugirello G, Domizzi G, Grimberg R, Steigmann R, Savin A 2012 Insight 54(6) 316-321
[6] Ryan D ed., 2013 High performance discovery in time series: techniques and case studies. Springer Science & Business Media
[7] Preston D, Protopapas P, Brodley C 2009 Stat Anal Data Min 2(5-6) 396-411
[8] Andrews G E 1984 The Theory of Partition, Cambridge University Press, Cambridge
[9] Cosarinsky G, Ruch M, Rugirello G, Domizzi G, Savin A, Steigmann R, Grimberg R 2011 Eddy current testing reliability improvement; application to nuclear fuel cladding tubes, VIII Congresso Regional de Ensaios No Destructivos y Estructurales – VIII CORENDE, Campana, Buenos Aires, Argentina
[10] Code ASME, Section V, art.8, app.1 and 2., 2002

Acknowledgments
This paper is partially supported by Romanian Ministry of Research and Innovation under project PN-II-PCCA-2013-4-0656 and Nucleus Program PN 16-37 01 01.